A COMPACT PROBABILISTIC RECURRENT OPERATOR FOR FORECASTING CLIMATE-SENSITIVE DISEASE INCIDENCE UNDER PARTIAL OBSERVABILITY

Anonymous authors
Paper under double-blind review

ABSTRACT

Public health agencies need reliable forecasts of infectious diseases to prevent epidemics with limited routinely available epidemiological and climate data. In this research, we introduce a compact probabilistic operator to forecast disease incidence under real-world information constraints. The operator encodes climate inputs separately using one-way encoding, retains recent dengue incidence with a gradually decaying memory, bounds month-wise changes in the internal disease state, and anchors the next-month forecast to the most recently observed incidence. We combined historical dengue surveillance and meteorological data covering 90 countries from 1990 to 2024. Data from 75 countries contained at least one pair of 12 months of history and next-month dengue incidence. Using a threshold of 1.5 standard deviations above the mean, we found that 4.23% of the 4,449 histories from 73 countries with at least one eligible non-outbreak history were followed by an outbreak month. We empirically measured how much variability in next-month dengue incidence can be predicted from observed disease incidence and climate data using a nearest-neighbor analysis ($k = 25$). The log-scale R^2 on the held-out set was 0.730 (country-bootstrap 95% confidence interval: 0.586–0.808). Our forecaster has 89,652 parameters. We trained it independently five times using different random initializations and report ensemble forecasts by averaging the predicted outputs of the five fitted operators. When evaluated against observations on the original observation scale, the ensemble attained $R^2 = 0.713$ on the test set. After split-conformal calibration, its nominal 90% prediction intervals achieved empirical marginal coverage of 95.6%. We also compared the operator with a matched recurrent model lacking the four constraints. The operator reduced disagreement across independently trained models by $2.78\times$ and sensitivity to small input perturbations by $1.84\times$. It also outperformed negative-binomial and SARIMAX-style models in matched assessments. Across seeds 42–46 on fold 0, the operator achieved the lowest mean SMAPE in a comparison with 34 baseline models on the data-rich cohort. Performance worsened when the historical data length was short. On 156 matched test country-months from 13 countries, all-period outbreak average precision (AP) was 0.958 for SPECTRA and 0.907 for the pooled SARIMAX-style comparator. Among 64 months following a non-outbreak month, sudden-onset AP was 0.657 and 0.695, respectively, against an event prevalence of 0.219. We also assessed model accuracy when transferring the model to countries not used during training. The model achieved positive log-scale R^2 in 12 of 13 countries included in training and in 15 of 17 previously unseen countries with at least 10 years of historical data and at least 30% of months having nonzero target values, without country-specific retraining. We showed that the operator can also be used in resource-constrained settings.

1 INTRODUCTION

Infectious disease prevention is crucial for global health security (Commission on a Global Health Risk Framework for the Future; National Academy of Medicine Secretariat, 2016; Clarke et al.,

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

2022). Public health agencies need accurate forecasts of these diseases to allocate resources and implement preventive measures such as immunization, mosquito control, and surveillance in a timely and targeted manner. The disease signals are often autoregressive as the previous burden of infections often influences the future incidence of the diseases (Wiens & Shenoy, 2018; Alahmari et al., 2024), but they are at times non-stationary due to the influence of environmental and climatic factors (Altizer et al., 2006; Fisman, 2007; Cazelles et al., 2018). With changes in host immunity, population-related and environmental drivers, entomological factors, surveillance coverage, and many other real-world factors (Mora et al., 2022; Agyarko et al., 2025), the number of affected people often rises sharply above what is anticipated by the public health authorities based on historical trends, constituting epidemics (Sophia et al., 2025).

A number of factors that determine the observed incidence of a disease, such as host immunity, vector abundance, serotypes, and surveillance quality, are latent and are not contained in the observed caseload and meteorological histories–time series data that are routinely available to forecast infectious diseases (Sahai et al., 2020; Maharaj, 2017; Redding et al., 2021). Therefore, developing accurate transmission models of an infectious disease is often not possible by using only surveillance and meteorological data, which are routinely available. The forecasting models developed using only available data are supposed to retain uncertainty associated with the unobserved information mentioned above. When outcomes remain variable conditional on the available predictors, this component of uncertainty cannot be eliminated simply by increasing the training sample size while keeping those predictors fixed (Hüllermeier & Waegeman, 2021). Since additional data required to address the uncertainty is often not available, a principled approach is required to address the problem of predicting these diseases under partial observability. Moreover, regression and time series regression models are widely used in infectious disease forecasting (Dixon et al., 2022). When used to produce only a conditional-mean point forecast, they do not provide predictive dispersion and are therefore insufficient on their own to estimate the probability of crossing a public health action threshold (Sophia et al., 2025; Redding et al., 2021), which is a functional of the predictive distribution itself (Rahman et al., 2025; Hussien, 2019).

Therefore, in this study, we examine empirically how predictable next-month incidence of a climate-sensitive infectious disease is from routinely available data of that disease and corresponding meteorological variables and develop a probabilistic approach of infectious disease forecasting under real-world data constraints. In this study, we use dengue as an instance. We also assess whether it (i) represents uncertainty, including upper-tail outcomes; (ii) can help in outbreak prediction based on a conventional threshold-based approach used in public health; (iii) improves on forecasting baselines when sufficient historical data is available; and (iv) can be reused across countries without country-specific retraining. Our primary contribution is developing a forecasting model based on dengue incidence and meteorological data using four scientifically derived structural properties: one-way meteorological encoding (Tishby & Zaslavsky, 2015; Granger, 1969), decaying incidence memory (Hawkes, 1971), bounded recurrent dynamics (Hochreiter & Schmidhuber, 1997), and persistence-anchored probabilistic output (Nix & Weigend, 1994; Salinas et al., 2020). The article introduces two approaches. First, we introduce a forecasting framework that reuses and combines existing machine learning methods to address the structural properties, and evaluate that framework using point, distributional, outbreak, and decision-oriented criteria. We name this framework SPECTRA. Second, we develop a compact recurrent operator, reducing the model complexity, that retains the properties mentioned above and compare that empirically with the SPECTRA framework.

2 Related Work

Application of general-purpose time-series forecasting in public health. Over the past decade, several general-purpose time series forecasting models have been adapted to predict the incidence of different infectious diseases at scales ranging from sub-district to national and time horizons ranging from daily to multi-decade (Leung et al., 2023; Johansson et al., 2016; Zhao et al., 2020; Jain et al., 2019; Salim et al., 2021). Broadly these models fall into the following families–statistical, gradient-boosted, recurrent, transformer, state-space, neural-ODE, diffusion, and graph-based (Leung et al., 2023; Zhao et al., 2020; Guo et al., 2017; Salim et al., 2021). The inventory is available in Appendix A. In terms of functional challenges, these models can be grouped into three categories (Leung et al., 2023; Funk et al., 2019). a) Statistical and tree-based models use historical caseloads or incidences and climate variables to predict future incidence (Johansson et al., 2016; Lowe et al., 2018; Zhao

et al., 2020; Guo et al., 2017; Jain et al., 2019; Salim et al., 2021). Examples include dengue forecasts for Mexico, Barbados, and Colombia (Johansson et al., 2016; Lowe et al., 2018; Zhao et al., 2020). These models often perform well for short-term forecasting and are relatively easy to fit (Johansson et al., 2016; Zhao et al., 2020; Guo et al., 2017; Jain et al., 2019). However, they are often location specific and return only a conditional mean or, at best, symmetric prediction intervals, and rarely test whether outbreak-risk probabilities are calibrated (Leung et al., 2023; Funk et al., 2019). b) Deep sequence models are good at learning temporal patterns (Hochreiter & Schmidhuber, 1997; Mussumeci & Codeço Coelho, 2020; Majeed et al., 2023). But a review of dengue-specific models shows that their validation is inconsistent and rarely focuses on upper-tail risk for alerts (Leung et al., 2023; Funk et al., 2019). c) Mechanistic and hybrid models are built on explicit assumptions on disease/transmission biology (Lee et al., 2021; Chowell, 2017; Oidtman et al., 2021). But they require latent states and parameters, which are difficult to identify from observed aggregated data alone (Lee et al., 2021; Chowell, 2017; King et al., 2016). In summary, across these groups, we did not find a method that meets all the four operational objectives stated above (Leung et al., 2023; Funk et al., 2019; Oidtman et al., 2021). The remainder of the paper develops one.

Inverse problems, partial observation, and structural priors. Classical inverse-problem theory suggests that when the causes of an observed signal are not fully measured, the estimate becomes unstable in the absence of domain-specific assumptions (Hadamard, 1902; Stuart, 2010). It is a well-documented issue in Markov processes and compartmental epidemiological models. In the presence of latent epidemiological states, it becomes difficult to identify parameters from aggregate incidence alone (Chowell, 2017; King et al., 2016). Separate bodies of work address parts of this problem, but mostly in isolation. For example, Granger analysis tests whether past values of one time series improve prediction of another, conditional on the included history; it does not by itself establish a causal effect (Granger, 1969). It has been applied to test whether rainfall helps predict dengue incidence in Bangladesh (Hossain et al., 2024) and whether rainfall, rainy days, relative humidity, and wind speed help predict dengue incidence in Sri Lanka (Prabodanie et al., 2020). Hawkes processes provide a statistical method to model self-excitation, where past events increase the probability of future events over a decaying time window (Hawkes, 1971). Studies show its applications in real-time prediction of the 2018–2019 Ebola outbreak in the Democratic Republic of Congo (Kelly et al., 2019) and modeling imported and local malaria cases in near-elimination settings (Unwin et al., 2021). Gated recurrent networks provide a learning method to model temporal dependence while controlling how much past information is retained or forgotten at each step (Hochreiter & Schmidhuber, 1997; Cho et al., 2014). They have been applied to infectious disease forecasting, including influenza-like illness prediction using recurrent neural networks (Volkova et al., 2017; Zhu et al., 2019) and weekly dengue incidence forecasting across 790 Brazilian cities using LSTM models (Mussumeci & Codeço Coelho, 2020). Probabilistic forecasters provide a learning method to estimate the full predictive distribution of future values rather than only a single point forecast (Gneiting & Raftery, 2007; Salinas et al., 2020). They have been applied to infectious disease forecasting, including collaborative probabilistic influenza forecasting in the United States (Reich et al., 2019) and probabilistic dengue forecasting in Vietnam (Colón-González et al., 2021). Conformal prediction helps in calibrating prediction intervals with finite-sample coverage guarantees without assuming a specific error distribution (Vovk et al., 2005; Lei et al., 2018; Angelopoulos & Bates, 2023). Researchers have used conformal-inference ideas to construct prediction intervals for county-level COVID-19 death counts in the United States (Altieri et al., 2021). Despite the presence of these methods that apparently address the different constraints of our problem, we did not find a single forecasting rule that combines all of them to address the climate-sensitive infectious disease forecasting problem.

Probabilistic forecasting and calibration. The priors discussed above determine how the forecaster should use the available information. Climate information should be used in the disease model as a compressed exogenous variable (C1). Recent incidence should be carried forward as disease memory (C2). The recurrent state to be developed should be constrained so that small perturbations do not make estimates unstable (C3). However, these three constraints are not sufficient for public health use because the decision to issue an alert is not supposed to be set only on the expected number of future cases. It depends on whether future incidence may cross an action threshold. Therefore, the output of our model must be a calibrated predictive distribution (C4). To satisfy the output prior (C4), we found five components in the forecasting literature that are directly relevant: proper scoring rules for evaluating predictive distributions, heteroscedastic likelihoods for time-varying variance, pinball losses for quantiles, adversarial/feature-matching losses for

sequence-level realism, and split-conformal prediction for calibrated coverage. Appendix B maps each component to the operational constraint it addresses. However, prior work usually uses these components separately. Climate-sensitive disease forecasting requires them to work together with C1–C3 in a single forecasting rule.

3 PROBLEM FORMULATION

Let $\mathcal{D} = \{(y^{(n)}_{1:T_n}, c^{(n)}_{1:T_n})\}_{n=1}^{N}$ denote the monthly panel, where y_t is the reported dengue incidence and c_t is the vector of climate and demographic variables, whose information is available through public sources. Here $n \in \{1, \ldots, N\}$ indexes countries, T_n is the length of country n's history, $h \in \{1, \ldots, H\}$ is the forecast horizon, κ is the alert threshold, θ are the model parameters, and k is the number of nearest neighbours used to compute the empirical forecastability diagnostic. The observed history $(y_{1:t}, c_{1:t})$ does not contain two processes that affect future incidence: the latent epidemiological state α_t and the reporting-noise process ν_t. Therefore, the forecasting problem is to learn the predictive distribution obtained after marginalizing over these unobserved processes:

$$p_\theta(y_{t+h} \mid y_{1:t}; c_{1:t}) = \iint p_\theta(y_{t+h} \mid \alpha_{t+h}, \nu_{t+h}) \cdot p_\theta(\alpha_{t+h}, \nu_{t+h} \mid y_{1:t}; c_{1:t})\, d\alpha_{t+h}\, d\nu_{t+h}, \quad h \in \{1, \ldots, H\}. \tag{1}$$

For public health use, the relevant measure is the threshold-crossing probability, given by:

$$\Pr_\theta(y_{t+h} > \kappa \mid y_{1:t}; c_{1:t}). \tag{2}$$

We evaluate it with respect to the empirical k-nearest-neighbours forecastability diagnostic of the observed panel.

4 METHOD

We present the method in two artefacts: a framework (the teacher) and a canonical operator (the artefact reported throughout the paper), shown in Figures 1 and 2. Both realise the four structural priors C1–C4 of Section 2; the operator collapses the framework into a single recurrent cell. Equations are tabulated in Appendix C; the explicit reduction map is given in Appendix D.

4.1 SPECTRA FRAMEWORK

SPECTRA learns the forward mapping of a climate manifold to a target disease signal as a first-class object to predict the disease target along with its uncertainty distribution. It has two major components: a climate encoder and a disease forecaster. SPECTRA's climate encoder is a conditional β-variational autoencoder (Cβ-VAE) adapted from the general VAE framework and the β-VAE (Kingma & Welling, 2014; Higgins et al., 2017). We introduced this encoder before training the forecasting model to represent the underlying structure of the climate force. The β-term prevents the encoder from freely assigning values to the latent variables to exactly match the climate input by penalizing the encoder when the latent distribution deviates substantially from a normal distribution. Thus, it limits the impact of noise in meteorological inputs on the construction of latent variables and allows the distribution of these variables to more accurately capture the broad and recurring patterns present in the actual climate input. From the latent variables, 1-, 3-, and 6-month lagged variables, and 3- and 6-month rolling means and standard deviations are constructed. The latent variables along with these derived features, as well as seasonal features (sine and cosine transformations of the calendar month) and linearly projected noise and country embedding, are passed as input to the disease forecaster. We constructed the disease forecaster to address the multiscale, non-Markovian, and non-stationary incidence problem. It has a generator network comprising a Causal Temporal Convolutional Network (TCN) (Bai et al., 2018), a Long Short-Term Memory (LSTM) network (Hochreiter & Schmidhuber, 1997), and a Multi-Head Self-Attention (MHSA) (Vaswani et al., 2017) layer, and a Critic that comprises an LSTM of the same configuration as that of the Generator followed by a fully connected layer. The forecaster is trained adversarially using a composite loss that combines negative log likelihood and quantile loss (Nix & Weigend, 1994; Koenker & Bassett, 1978) with a feature matching term and an adversarial loss based on the Wasserstein

Generative Adversarial Network (WGAN) gradient penalty (GP) (Salimans et al., 2016; Arjovsky et al., 2017; Gulrajani et al., 2017). The mapping of the priors to the design is given in Appendix C, Table 6.

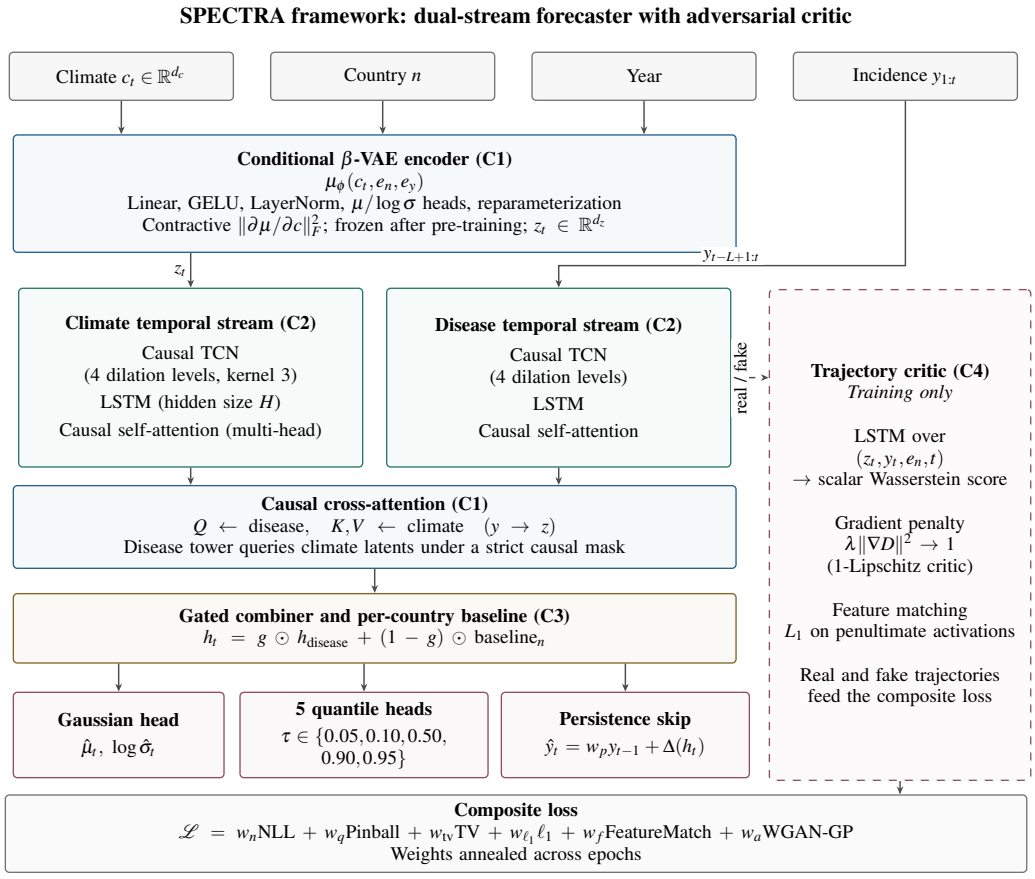


Figure 1: SPECTRA framework.

4.2 Canonical recurrent operator (SPECTRA)

The canonical operator (Figure 2) is obtained from the framework by a structure-preserving reduction: every framework module is replaced by the smallest closed-form construction that provably preserves its associated structural prior. The reduction yields a single recurrent cell, stacked L = 2 times, with 89,652 trainable parameters. C1 (separate one-way exogenous climate encoding) is preserved by retaining the frozen β-VAE and collapsing the framework's TCN–LSTM–attention climate stream into a single tanh-bounded affine projection of the precomputed latent; because the latent is fixed prior to operator training, architectural separation of the climate pathway with respect to incidence is inherited from the framework by construction. C2 (Hawkes-style decaying incidence memory) is reduced from a learned TCN over disease history to a closed-form parametric Hawkes recursion (Hawkes, 1971; Kelly et al., 2019; Unwin et al., 2021) with per-country excitation α_n and decay β_n, supplemented by an engineered multi-scale lag bank that absorbs the slow-decay tail without an additional convolutional module. C3 (bounded gated convex update) replaces the framework's LSTM with a gated convex combination of a tanh-bounded candidate and the previous state, modulated by a surge gate that opens during fast incidence rises; the convex form yields $\|h_t\|_\infty \leq 1$ and a finite Lipschitz constant in closed form, properties the LSTM attains only empirically. C4 (persistence-skip emission) anchors both the Gaussian head and the median quantile at a trainable persistence baseline, and constructs the four non-median quantiles from the median by softplus deltas; monotone ordering then holds by construction rather than as a soft penalty. The reduced operator is trained without an adversarial critic. Per-country split-conformal calibration (Vovk et al.,

2005; Lei et al., 2018; Angelopoulos & Bates, 2023) adjusts the predictive distribution post hoc to deliver marginal coverage 1 - α within each country, and the threshold-crossing probability (2) is read off the Gaussian head and corroborated against the conformalised upper interval at deployment.

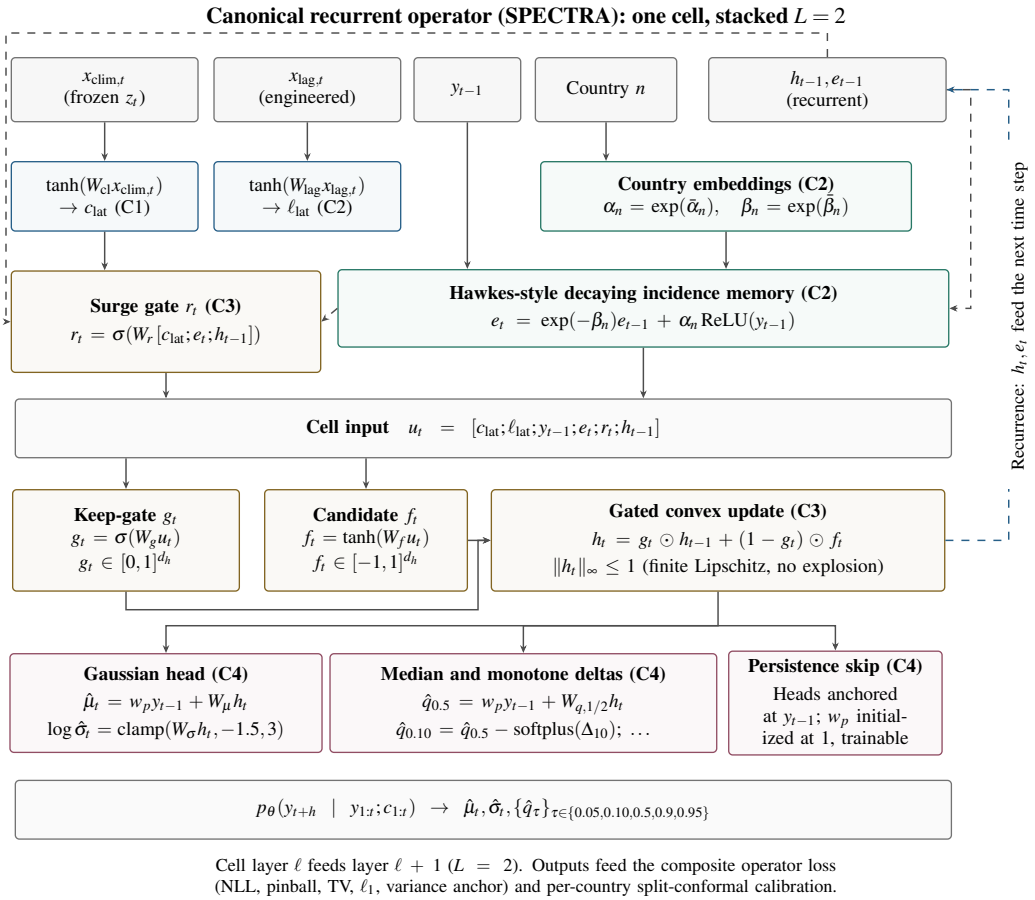


Figure 2: Canonical recurrent operator (SPECTRA).

5 EXPERIMENTS

We grouped the countries in different tiers based on the length of historical data available. The details of the same are available in Appendix E. Here, we only present the result for countries where the longest historical datasets are available. The remaining results are available in Appendices F and J. Tier A: 16 dengue-endemic countries, $\geq$ 15 yr monthly history (training set; Appendix E); per-country results on 13 countries with 12 test months each and Viet Nam with two test months in fold 0 (China, Malaysia excluded from per-country reporting; both retained in training). Setup, hyperparameters, 34-baseline taxonomy, and bootstrap audit in Appendices E–H.

5.1 EMPIRICAL FORECASTABILITY AND STRUCTURAL-PROPERTY DIAGNOSTICS

The k-nearest-neighbour diagnostic achieved log-scale $R^2 = 0.730$ on its held-out sample. Among 1,958 eligible non-outbreak histories in Tier A, 67 (3.42%) were followed by a seasonal-outbreak month. Per-prior structural-property and stability diagnostics (Appendix G, Table 12) verify C1–C4: architectural separation of the climate pathway, Hawkes Green function match within 10^{-5}, gated-update Lipschitz L = 0.358, persistence-skip emission with nominal coverage. Removing all four priors triples across-seed disagreement (2.78×) and nearly doubles Lipschitz sensitivity (1.84×, Table 1).

Table 1: Structural-property and stability diagnostics, Tier A fold 0: canonical operator vs. unconstrained (C1–C4 removed).

Diagnostic	SPECTRA	Unconstrained	Ratio
Across-seed forecast disagreement (lower is better)	0.0814	0.2265	2.78×
Empirical Lipschitz sensitivity (lower is better)	0.402	0.740	1.84×
Empirical Lipschitz constant L (impulse response)	0.358	–	–
90% conformal coverage (target 0.90)	0.956	–	–

5.2 Established forecasters

Across seeds 42–46 on fold 0, SPECTRA achieved the lowest mean SMAPE among the 34 baselines. The Friedman omnibus test over 36 models, including SPECTRA and SPECTRA+FT, indicated overall rank differences ($\chi^2 = 162.35$, $p = 2.6 \times 10^{-18}$). The full ranking is reported in Appendix F.

5.3 Predictive performance and operational alerting

The ensemble of the five models achieves $R^2 = 0.713$ on test data. Across the five separately trained fold-0 models, the average SMAPE is 43.26 ± 1.80. Headline ensemble COV90 is 0.956. On the 156-observation checkpoint evaluation, log-CRPS and the scaled 90% interval score (log scale) are 0.254 and 0.116, respectively. We did not divide this R^2 by the k-NN result because of differences in the outcome scale and the data splitting process. On 156 matched test country-months from 13 countries, all-period outbreak average precision (AP) was 0.958 for SPECTRA and 0.907 for the pooled SARIMAX-style comparator. For sudden-onset evaluation restricted to 64 eligible months containing 14 events, AP was 0.657 and 0.695, respectively. The country-bootstrap confidence intervals for both AP differences included zero. The decision value is shown in Figure 8 (Appendix H.1). In a matched assessment over 156 unseen country-months from 13 countries, we found that the operator achieved lower SMAPE, log-CRPS, and scaled 90% interval score (log scale) than pooled SARIMAX-style, negative-binomial GAM, and negative-binomial DLNM specifications (Appendix N.1, Table 22). However, the comparison was made only on pooled specifications. The result does not provide a formal basis to infer superiority over every possible SARIMAX, GAM, or DLNM model.

5.4 Operator–framework alignment

We trained the operator and the framework separately on the same fold using the same seeds, training and validation rules, and early stopping process, using the same outcomes. The operator had 89,652 parameters. We found that the operator achieved a higher R^2. The reported 95% confidence intervals for differences in log-scale R^2, SMAPE, and log-CRPS included zero; no confidence interval was reported for the coverage difference. At best, the result can be interpreted as broadly comparable performance on matched folds. The operator took 4.2 ms per country-month on an NVIDIA T4. The matched comparison is reported in Appendix M (Table 21).

6 Discussion

The results of our experiment support that the performance of models designed to forecast infectious diseases from available surveillance and meteorological data depends not just on the capacity of a model, but also on how that model constrains information flow, memory, state evolution, and uncertainty. The practical output of our research is a forecasting operator that can be trained using routinely collected data. Our operator achieved the lowest relative difference between actual values and forecasted values measured using mean SMAPE among the 34 baselines using five seeds. The forecaster can also be transferred across several countries without architectural modification, which was not the case for a vast majority of forecasting models used in climate-sensitive infectious disease research. The model produces prediction intervals as well as outbreak-risk scores along with point estimates, which are required for the model to be trusted by public health decision-makers for real-

Table 2: Headline fold-0 benchmark on the data-rich cohort (mean ± standard deviation across seeds 42–46). R^2_{ens}: EMA-ensemble single-pass; R^2_{seed}: per-seed mean; R^2_{med}: per-country median. Full ranking in Appendix F, Table 11. $\text{sIS90}_{\log}$ is defined in Appendix N.2.

Model	SMAPE ↓	R^2_{ens} ↑	R^2_{seed}	R^2_{med}	COV90 → 0.90	WIDTH90 ↓	$\text{sIS90}_{\log}$ ↓
LastValue (persistence)	46.01	–	-0.694	–	–	–	–
Seasonal naive	78.89	–	0.590	–	–	–	–
HistGB	54.45	–	0.629	–	–	–	–
RandomForest	54.17	–	0.417	–	–	–	–
NF-PatchTST	80.25±1.15	–	0.216	–	–	–	–
LSTM-tiny	81.84±2.73	–	0.103	–	–	–	–
NF-DeepAR	95.46±3.20	–	-0.366	–	–	–	–
SPECTRA (ours)	43.26±1.80	0.713	0.398±0.39	0.41	0.956	517.4	0.116
SPECTRA + per-country FT	43.27±1.33	–	0.587±0.20	0.49	0.945	1255	–

Table 3: Outbreak detection on 156 matched test observations from 13 countries, using training-only country/calendar-month mean plus 1.5 standard deviations as the outbreak threshold. Sudden-onset AP uses 64 months following a non-outbreak month, with 14 events. AP: average precision; AUC: ROC AUC. Per-country diagnostics in Appendix H.

Model	AP_{all}	$\text{AP}_{\text{sudden}}$	AUC_{all}
SPECTRA (ours)	.958	.657	.937
Pooled SARIMAX-style	.907	.695	.898

world use. Across alert thresholds of 5%–50%, both SPECTRA and the pooled SARIMAX-style comparator provided positive net benefit in the matched evaluation. For sudden outbreaks, their average precision was 0.657 and 0.695, respectively, compared with an event prevalence of 0.219. Split-conformal calibration provides a distribution-free procedure for marginal interval coverage (Vovk et al., 2005; Lei et al., 2018; Angelopoulos & Bates, 2023). A separate k-nearest-neighbour regression analysis produced $R^2 = 0.730$. However, this finite-sample result cannot be directly compared with the operator's natural-scale $R^2 = 0.713$. The remaining error may be due to changes in the outcome variations depending on other factors and/or model estimation error. In this study, we have not identified the entire residual as irreducible.

The structural priors (Granger, 1969; Hawkes, 1971; Hochreiter & Schmidhuber, 1997; Nix & Weigend, 1994; Salinas et al., 2020) that we have used to construct our framework and operator have already been well-researched in existing literature. The priors are realised in the trained operator: no gradient flows from disease to climate; incidence memory matches the Hawkes Green function within 10^{-5}; the gated update is contractive ($L = 0.358$); the output is anchored to the last observation with valid upper-tail coverage. Removing the priors triples across-seed disagreement ($2.78\times$) and nearly doubles Lipschitz sensitivity ($1.84\times$), the two empirical fingerprints of across-seed disagreement and instability in ill-posed problems (Hadamard, 1902; Stuart, 2010). Therefore, our specific contribution through this study is how the priors could be combined together to develop a problem-specific machine learning forecasting framework, which can be used in public health for epidemic prevention, preparedness, and response. More importantly, we deduced a single recurrent operator from the framework, conducting property-level validation of that operator within the same forecasting setup. The results also indicate that the structural priors we introduced through the above-mentioned principled design approach improve stability across models trained independently. They also helped in reducing the sensitivity to input perturbations while retaining competitive point and distributional performance. An important value of the same may pertain to a situation of a sudden and explosive outbreak of a disease with trivial changes in environmental factors, especially when they are not rare in an era of climate change. However, we neither claim to advance a disease-specific mechanistic or transmission model, nor do we claim that such an advantage has already been formally established through our experiments.

With 89,652 parameters, fast fitting, no Wasserstein-GP critic at training (Arjovsky et al., 2017; Gulrajani et al., 2017), we also infer that a substantive forecasting model can be deployed even

in resource-limited settings. The operator becomes less complex due to the removal of the adversarial critic required by the larger framework (Arjovsky et al., 2017; Gulrajani et al., 2017). Its computational efficiency makes ensemble inference also feasible in settings where digital public infrastructure is limited, especially in the low- and middle-income countries, where public health preparedness is crucial for epidemic prevention and better health sector readiness. Its data-readiness and uncertainty signals identify countries or forecasts requiring additional review.

Limitations. Our study does not advance a dengue-specific model. We have not compared the operator with transmission models. Our analytical approach should be treated only as an available data-based benchmark, and we do not claim to provide any inference related to the collection and aggregation of the available data. We also can't infer how much of the prediction error arises from different factors like uncertainty of the outcome, modeling errors, data limitations, and others. The structural priors we used in the study offer a principled approach for forecasting, but they neither make any underlying inverse problem formally well posed nor advance a latent data-generating process. Average precision for sudden-onset outbreaks does not establish that the remaining prediction error is caused by irreducible uncertainty from unmeasured factors. The empirical evaluation is limited to one disease, monthly country-level data, and a one-step horizon. Performance of the model worsens when historical disease data is short and discontinuous. As performance also varies across independently trained models, we report the ensemble R^2. Another notable issue is that the model's intervals were slightly too narrow for unusually high incidence before any calibrations were done. So, country-specific split-conformal calibration was a necessity to achieve the reported coverage (Vovk et al., 2005; Lei et al., 2018; Angelopoulos & Bates, 2023). Additionally, evaluation limitations include that split conformalization establishes only marginal interval coverage. The test panel comprising data from 14 countries lacks sufficient sustained variation to identify the Hawkes self-excitation effect. So, its form is not determined solely from the data (Chowell, 2017; King et al., 2016). The ablations show the effects of removing components. But they do not establish a component's unique importance. The comparison of the framework and the operator shows that the operator retained the desired properties, but the assessment does not reflect identical predictions by the two approaches.

7 Conclusion

To the best of our knowledge, SPECTRA provides the first principled approach in which uncertainty-aware representation learning is used to regularize the otherwise ill-posed task of climate-sensitive infectious disease forecasting. In practice, the primary benefit that SPECTRA offers is a fundamental change in how climate information can be used in epidemic preparedness decisions: instead of attempting to infer the unique climatic drivers of outbreaks, SPECTRA provides a way to learn what outbreak patterns are plausible in a given meteorological condition and how uncertain those risks are. Finally, the contribution of this experiment to machine learning can also be considered an engineering blueprint, rather than a new primitive, for building scalable and trustworthy early-warning systems that can span geographies. In this article, the evidence we have presented does not support any inference about whether the operator can be used to forecast other important climate-sensitive diseases like malaria, influenza, and diarrhoeal diseases. Given the current evidence, however, there is an opportunity to conduct future work to test that possibility using other disease-specific data and matched evaluation protocols.

REFERENCES

Bijaya Adhikari, Xinfeng Xu, Naren Ramakrishnan, and B. Aditya Prakash. EpiDeep: Exploiting Embeddings for Epidemic Forecasting. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 577–586, New York, NY, USA, 7 2019. ACM. ISBN 9781450362016. doi: 10.1145/3292500.3330917. URL https://dl.acm.org/doi/10.1145/3292500.3330917.

Robert Kwame D. Agyarko, Dennis Kithinji, and Kizito Bishikwabo Nsarhaza. Climate Change and the Rise of Emerging and Re-Emerging Infectious Diseases in Africa: A Literature Review. *International Journal of Environmental Research and Public Health*, 22(6):903, 6 2025. ISSN 1660-4601. doi: 10.3390/ijerph22060903. URL https://www.mdpi.com/1660-4601/22/6/903.

Ahmed A. Alahmari, Yasir Almuzaini, Fahad Alamri, Reem Alenzi, and Anas A. Khan. Strengthening global health security through health early warning systems: A literature review and case study. *Journal of Infection and Public Health*, 17:85–95, 4 2024. ISSN 18760341. doi: 10.1016/j.jiph.2024.01.019. URL https://linkinghub.elsevier.com/retrieve/pii/S187603412400025X.

Nick Altieri, Rebecca L Barter, James Duncan, Raaz Dwivedi, Karl Kumbier, Xiao Li, Robert Netzorg, Briton Park, Chandan Singh, Yan Shuo Tan, Tiffany Tang, Yu Wang, Chao Zhang, and Bin Yu. Curating a COVID-19 Data Repository and Forecasting County-Level Death Counts in the United States. *Harvard Data Science Review*, 2 2021. doi: 10.1162/99608f92.1d4e0dae. URL https://hdsr.mitpress.mit.edu/pub/p6isyf0g.

Sonia Altizer, Andrew Dobson, Parviez Hosseini, Peter Hudson, Mercedes Pascual, and Pejman Rohani. Seasonality and the dynamics of infectious diseases. *Ecology Letters*, 9(4):467–484, 4 2006. ISSN 1461-023X. doi: 10.1111/j.1461-0248.2005.00879.x.

Anastasios N. Angelopoulos and Stephen Bates. Conformal Prediction: A Gentle Introduction. *Foundations and Trends® in Machine Learning*, 16(4):494–591, 3 2023. ISSN 1935-8237. doi: 10.1561/2200000101. URL https://www.emerald.com/ftmal/article/16/4/494/1332423/Conformal-Prediction-A-Gentle-Introduction.

M Arjovsky, S Chintala, and L Bottou. Wasserstein generative adversarial networks. In *Proceedings of the 34th International Conference on Machine Learning*, pp. 214–223. PMLR, 2017. URL https://proceedings.mlr.press/v70/arjovsky17a.html.

Shaojie Bai, J. Zico Kolter, and Vladlen Koltun. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. arXiv:1803.01271, 4 2018. URL https://doi.org/10.48550/arXiv.1803.01271.

Gopi Battineni, Nalini Chintalapudi, and Francesco Amenta. Forecasting of COVID-19 epidemic size in four high hitting nations (USA, Brazil, India and Russia) by Fb-Prophet machine learning model. *Applied Computing and Informatics*, 21(1/2):2–11, 1 2025. ISSN 2634-1964. doi: 10.1108/ACI-09-2020-0059. URL http://www.emerald.com/aci/article/21/1-2/2-11/1241886.

Hannah R. Biegel and Joceline Lega. EpiCovDA: a mechanistic COVID-19 forecasting model with data assimilation. arXiv preprint arXiv:2105.05471, 10 2021. URL https://doi.org/10.48550/arXiv.2105.05471.

Johannes Bracher, Evan L. Ray, Tilmann Gneiting, and Nicholas G. Reich. Evaluating epidemic forecasts in an interval format. *PLOS Computational Biology*, 17(2):e1008618, 2 2021. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1008618. URL https://dx.plos.org/10.1371/journal.pcbi.1008618.

Anna L. Buczak, Benjamin Baugher, Linda J. Moniz, Thomas Bagley, Steven M. Babin, and Erhan Guven. Ensemble method for dengue prediction. *PLOS ONE*, 13(1):e0189988, 1 2018. ISSN 1932-6203. doi: 10.1371/journal.pone.0189988. URL https://dx.plos.org/10.1371/journal.pone.0189988.

Bernard Cazelles, Clara Champagne, and Joseph Dureau. Accounting for non-stationarity in epidemiology by embedding time-varying parameters in stochastic models. *PLOS Computational Biology*, 14(8):e1006211, 8 2018. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1006211.

Vinay Kumar Reddy Chimmula and Lei Zhang. Time series forecasting of COVID-19 transmission in Canada using LSTM networks. *Chaos, Solitons & Fractals*, 135:109864, 6 2020. ISSN 09600779. doi: 10.1016/j.chaos.2020.109864. URL https://linkinghub.elsevier.com/retrieve/pii/S0960077920302642.

Kyunghyun Cho, Bart van Merrienboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1724–1734, Stroudsburg, PA, USA, 2014. Association for Computational Linguistics. doi: 10.3115/v1/D14-1179. URL http://aclweb.org/anthology/D14-1179.

Gerardo Chowell. Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts. *Infectious Disease Modelling*, 2 (3):379–398, 8 2017. ISSN 24680427. doi: 10.1016/j.idm.2017.08.001. URL https://linkinghub.elsevier.com/retrieve/pii/S2468042717300234.

Lorcan Clarke, Edith Patouillard, Andrew J. Mirelman, Zheng Jie Marc Ho, Tessa Tan-Torres Edejer, and Nirmal Kandel. The costs of improving health emergency preparedness: A systematic review and analysis of multi-country studies. *eClinicalMedicine*, 44:101269, 2 2022. ISSN 25895370. doi: 10.1016/j.eclinm.2021.101269. URL https://linkinghub.elsevier.com/retrieve/pii/S2589537021005502.

Felipe J. Colón-González, Leonardo Soares Bastos, Barbara Hofmann, Alison Hopkin, Quillon Harpham, Tom Crocker, Rosanna Amato, Iacopo Ferrario, Francesca Moschini, Samuel James, Sajni Malde, Eleanor Ainscoe, Vu Sinh Nam, Dang Quang Tan, Nguyen Duc Khoa, Mark Harrison, Gina Tsarouchi, Darren Lumbroso, Oliver J. Brady, and Rachel Lowe. Probabilistic seasonal dengue forecasting in Vietnam: A modelling study using superensembles. *PLOS Medicine*, 18(3):e1003542, 3 2021. ISSN 1549-1676. doi: 10.1371/journal.pmed.1003542. URL https://dx.plos.org/10.1371/journal.pmed.1003542.

Commission on a Global Health Risk Framework for the Future; National Academy of Medicine Secretariat. The Case for Investing in Pandemic Preparedness. In *The Neglected Dimension of Global Security: A Framework to Counter Infectious Disease Crises*, chapter 2. National Academies Press, Washington, DC, 2016. doi: 10.17226/21891.

Estee Y. Cramer, Evan L. Ray, Velma K. Lopez, Johannes Bracher, Andrea Brennen, Alvaro J. Castro Rivadeneira, Aaron Gerding, Tilmann Gneiting, Katie H. House, Yuxin Huang, Dasuni Jayawardena, Abdul H. Kanji, Ayush Khandelwal, Khoa Le, Anja Mühlemann, Jarad Niemi, Apurv Shah, Ariane Stark, Yijin Wang, Nutcha Wattanachit, Martha W. Zorn, Youyang Gu, Sansiddh Jain, Nayana Bannur, Ayush Deva, Mihir Kulkarni, Srujana Merugu, Alpan Raval, Siddhant Shingi, Avtansh Tiwari, Jerome White, Neil F. Abernethy, Spencer Woody, Maytal Dahan, Spencer Fox, Kelly Gaither, Michael Lachmann, Lauren Ancel Meyers, James G. Scott, Mauricio Tec, Ajitesh Srivastava, Glover E. George, Jeffrey C. Cegan, Ian D. Dettwiller, William P. England, Matthew W. Farthing, Robert H. Hunter, Brandon Lafferty, Igor Linkov, Michael L. Mayo, Matthew D. Parno, Michael A. Rowland, Benjamin D. Trump, Yanli Zhang-James, Samuel Chen, Stephen V. Faraone, Jonathan Hess, Christopher P. Morley, Asif Salekin, Dongliang Wang, Sabrina M. Corsetti, Thomas M. Baer, Marisa C. Eisenberg, Karl Falb, Yitao Huang, Emily T. Martin, Ella McCauley, Robert L. Myers, Tom Schwarz, Daniel Sheldon, Graham Casey Gibson, Rose Yu, Liyao Gao, Yian Ma, Dongxia Wu, Xifeng Yan, Xiaoyong Jin, Yu-Xiang Wang, YangQuan Chen, Lihong Guo, Yanting Zhao, Quanquan Gu, Jinghui Chen, Lingxiao Wang, Pan Xu, Weitong Zhang, Difan Zou, Hannah Biegel, Joceline Lega, Steve McConnell, V. P. Nagraj, Stephanie L. Guertin, Christopher Hulme-Lowe, Stephen D. Turner, Yunfeng Shi, Xuegang Ban, Robert Walraven, Qi-Jun Hong, Stanley Kong, Axel van de Walle, James A. Turtle, Michal Ben-Nun, Steven Riley, Pete Riley, Ugur Koyluoglu, David DesRoches, Pedro Forli, Bruce Hamory, Christina Kyriakides, Helen Leis, John Milliken, Michael Moloney, James Morgan, Ninad Nirgudkar, Gokce Ozcan, Noah Piwonka, Matt Ravi, Chris Schrader, Elizabeth Shakhnovich, Daniel

Siegel, Ryan Spatz, Chris Stiefeling, Barrie Wilkinson, Alexander Wong, Sean Cavany, Guido España, Sean Moore, Rachel Oidtman, Alex Perkins, David Kraus, Andrea Kraus, Zhifeng Gao, Jiang Bian, Wei Cao, Juan Lavista Ferres, Chaozhuo Li, Tie-Yan Liu, Xing Xie, Shun Zhang, Shun Zheng, Alessandro Vespignani, Matteo Chinazzi, Jessica T. Davis, Kunpeng Mu, Ana Pastore y Piontti, Xinyue Xiong, Andrew Zheng, Jackie Baek, Vivek Farias, Andreea Georgescu, Retsef Levi, Deeksha Sinha, Joshua Wilde, Georgia Perakis, Mohammed Amine Bennouna, David Nze-Ndong, Divya Singhvi, Ioannis Spantidakis, Leann Thayaparan, Asterios Tsiourvas, Arnab Sarker, Ali Jadbabaie, Devavrat Shah, Nicolas Della Penna, Leo A. Celi, Saketh Sundar, Russ Wolfinger, Dave Osthus, Lauren Castro, Geoffrey Fairchild, Isaac Michaud, Dean Karlen, Matt Kinsey, Luke C. Mullany, Kaitlin Rainwater-Lovett, Lauren Shin, Katharine Tallaksen, Shelby Wilson, Elizabeth C. Lee, Juan Dent, Kyra H. Grantz, Alison L. Hill, Joshua Kaminsky, Kathryn Kaminsky, Lindsay T. Keegan, Stephen A. Lauer, Joseph C. Lemaitre, Justin Lessler, Hannah R. Meredith, Javier Perez-Saez, Sam Shah, Claire P. Smith, Shaun A. Truelove, Josh Wills, Maximilian Marshall, Lauren Gardner, Kristen Nixon, John C. Burant, Lily Wang, Lei Gao, Zhiling Gu, Myungjin Kim, Xinyi Li, Guannan Wang, Yueying Wang, Shan Yu, Robert C. Reiner, Ryan Barber, Emmanuela Gakidou, Simon I. Hay, Steve Lim, Chris Murray, David Pigott, Heidi L. Gurung, Prasith Baccam, Steven A. Stage, Bradley T. Suchoski, B. Aditya Prakash, Bijaya Adhikari, Jiaming Cui, Alexander Rodríguez, Anika Tabassum, Jiajia Xie, Pinar Keskinocak, John Asplund, Arden Baxter, Buse Eylul Oruc, Nicoleta Serban, Sercan O. Arik, Mike Dusenberry, Arkady Epshteyn, Elli Kanal, Long T. Le, Chun-Liang Li, Tomas Pfister, Dario Sava, Rajarishi Sinha, Thomas Tsai, Nate Yoder, Jinsung Yoon, Leyou Zhang, Sam Abbott, Nikos I. Bosse, Sebastian Funk, Joel Hellewell, Sophie R. Meakin, Katharine Sherratt, Mingyuan Zhou, Rahi Kalantari, Teresa K. Yamana, Sen Pei, Jeffrey Shaman, Michael L. Li, Dimitris Bertsimas, Omar Skali Lami, Saksham Soni, Hamza Tazi Bouardi, Turgay Ayer, Madeline Adee, Jagpreet Chhatwal, Ozden O. Dalgic, Mary A. Ladd, Benjamin P. Linas, Peter Mueller, Jade Xiao, Yuanjia Wang, Qinxia Wang, Shanghong Xie, Donglin Zeng, Alden Green, Jacob Bien, Logan Brooks, Addison J. Hu, Maria Jahja, Daniel McDonald, Balasubramanian Narasimhan, Collin Politsch, Samyak Rajanala, Aaron Rumack, Noah Simon, Ryan J. Tibshirani, Rob Tibshirani, Valerie Ventura, Larry Wasserman, Eamon B. O'Dea, John M. Drake, Robert Pagano, Quoc T. Tran, Lam Si Tung Ho, Huong Huynh, Jo W. Walker, Rachel B. Slayton, Michael A. Johansson, Matthew Biggerstaff, and Nicholas G. Reich. Evaluation of individual and ensemble probabilistic forecasts of COVID-19 mortality in the United States. *Proceedings of the National Academy of Sciences*, 119(15):e2113561119, 4 2022. ISSN 0027-8424. doi: 10.1073/pnas.2113561119. URL https://pnas.org/doi/10.1073/pnas.2113561119.

Raj Dandekar and George Barbastathis. Quantifying the effect of quarantine control in Covid-19 infectious spread using machine learning. medRxiv 2020.04.03.20052084, 4 2020. URL http://medrxiv.org/lookup/doi/10.1101/2020.04.03.20052084.

Songgaojun Deng, Shusen Wang, Huzefa Rangwala, Lijing Wang, and Yue Ning. Cola-GNN: Cross-location Attention based Graph Neural Networks for Long-term ILI Prediction. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pp. 245–254, New York, NY, USA, 10 2020. ACM. ISBN 9781450368599. doi: 10.1145/3340531.3411975. URL https://dl.acm.org/doi/10.1145/3340531.3411975.

Samuel Dixon, Ravikiran Keshavamurthy, Daniel H. Farber, Andrew Stevens, Karl T. Pazdernik, and Lauren E. Charles. A Comparison of Infectious Disease Forecasting Methods across Locations, Diseases, and Time. *Pathogens*, 11(2):185, 1 2022. ISSN 2076-0817. doi: 10.3390/pathogens11020185.

David N. Fisman. Seasonality of Infectious Diseases. *Annual Review of Public Health*, 28(1): 127–143, 4 2007. ISSN 0163-7525. doi: 10.1146/annurev.publhealth.28.021406.144128.

Sebastian Funk, Anton Camacho, Adam J. Kucharski, Rachel Lowe, Rosalind M. Eggo, and W. John Edmunds. 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, 2 2019. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1006785. URL https://dx.plos.org/10.1371/journal.pcbi.1006785.

Junyi Gao, Rakshith Sharma, Cheng Qian, Lucas M Glass, Jeffrey Spaeder, Justin Romberg, Jimeng Sun, and Cao Xiao. STAN: spatio-temporal attention network for pandemic prediction using

real-world evidence. *Journal of the American Medical Informatics Association*, 28(4):733–743, 3 2021. ISSN 1527-974X. doi: 10.1093/jamia/ocaa322. URL https://academic.oup.com/jamia/article/28/4/733/6118380.

Tilmann Gneiting and Adrian E Raftery. Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association*, 102(477):359–378, 3 2007. ISSN 0162-1459. doi: 10.1198/016214506000001437. URL http://www.tandfonline.com/doi/abs/10.1198/016214506000001437.

C. W. J. Granger. Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, 37(3):424–438, 8 1969. ISSN 00129682. doi: 10.2307/1912791. URL https://www.jstor.org/stable/1912791?origin=crossref.

A Gu and T Dao. Mamba: Linear-time sequence modeling with selective state spaces. In *COLM*, 2024. URL https://arxiv.org/abs/2312.00752.

A Gu, K Goel, and C Ré. Efficiently Modeling Long Sequences with Structured State Spaces. In *ICLR*, 2022. URL https://hai.stanford.edu/research/efficiently-modeling-long-sequences-with-structured-state-spaces.

I Gulrajani, F Ahmed, M Arjovsky, V Dumoulin, and A Courville. Improved training of Wasserstein GANs. In *Advances in Neural Information Processing Systems 30*, 2017. URL https://papers.nips.cc/paper/7159-improved-training-of-wasserstein-gans.pdf.

Pi Guo, Tao Liu, Qin Zhang, Li Wang, Jianpeng Xiao, Qingying Zhang, Ganfeng Luo, Zhihao Li, Jianfeng He, Yonghui Zhang, and Wenjun Ma. Developing a dengue forecast model using machine learning: A case study in China. *PLOS Neglected Tropical Diseases*, 11(10):e0005973, 10 2017. ISSN 1935-2735. doi: 10.1371/journal.pntd.0005973. URL https://dx.plos.org/10.1371/journal.pntd.0005973.

J Hadamard. Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton University Bulletin*, 13:49–52, 1902. URL https://illposed.net/hadamard.pdf.

Alan G. Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90, 1971. ISSN 0006-3444. doi: 10.1093/biomet/58.1.83. URL https://academic.oup.com/biomet/article-lookup/doi/10.1093/biomet/58.1.83.

I Higgins, L Matthey, A Pal, C Burgess, X Glorot, M Botvinick, S Mohamed, and A Lerchner. β-VAE: Learning basic visual concepts with a constrained variational framework. In *International Conference on Learning Representations*, 2017. URL https://openreview.net/forum?id=Sy2fzU9gl.

Yien Ling Hii, Huaiping Zhu, Nawi Ng, Lee Ching Ng, and Joacim Rocklöv. Forecast of Dengue Incidence Using Temperature and Rainfall. *PLoS Neglected Tropical Diseases*, 6(11):e1908, 11 2012. ISSN 1935-2735. doi: 10.1371/journal.pntd.0001908. URL https://dx.plos.org/10.1371/journal.pntd.0001908.

Sepp Hochreiter and Jürgen Schmidhuber. Long Short-Term Memory. *Neural Computation*, 9(8):1735–1780, 11 1997. ISSN 0899-7667. doi: 10.1162/neco.1997.9.8.1735. URL https://direct.mit.edu/neco/article/9/8/1735-1780/6109.

Md. Jamal Hossain, Nazia Sultana, Anwesha Das, Fariea Nazim Jui, Md. Kamrul Islam, Md. Mijanoor Rahman, and Mohammad Mafizur Rahman. Analysis of effects of meteorological variables on dengue incidence in Bangladesh using VAR and Granger causality approach. *Frontiers in Public Health*, 12:1488742, 11 2024. ISSN 2296-2565. doi: 10.3389/fpubh.2024.1488742. URL https://www.frontiersin.org/articles/10.3389/fpubh.2024.1488742/full.

Eyke Hüllermeier and Willem Waegeman. Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods. *Machine Learning*, 110(3):457–506, 3 2021. ISSN 0885-6125. doi: 10.1007/s10994-021-05946-3.

Hamid H. Hussien. Malaria's association with climatic variables and an epidemic early warning system using historical data from Gezira State, Sudan. *Heliyon*, 5(3):e01375, 3 2019. ISSN 24058440. doi: 10.1016/j.heliyon.2019.e01375. URL https://linkinghub.elsevier.com/retrieve/pii/S2405844018368956.

Raghvendra Jain, Sra Sontisirikit, Sopon Iamsirithaworn, and Helmut Prendinger. Prediction of dengue outbreaks based on disease surveillance, meteorological and socio-economic data. *BMC Infectious Diseases*, 19(1):272, 12 2019. ISSN 1471-2334. doi: 10.1186/s12879-019-3874-x. URL https://bmcinfectdis.biomedcentral.com/articles/10.1186/s12879-019-3874-x.

Michael A. Johansson, Nicholas G. Reich, Aditi Hota, John S. Brownstein, and Mauricio Santillana. Evaluating the performance of infectious disease forecasts: A comparison of climate-driven and seasonal dengue forecasts for Mexico. *Scientific Reports*, 6(1):33707, 9 2016. ISSN 2045-2322. doi: 10.1038/srep33707. URL https://www.nature.com/articles/srep33707.

J. Daniel Kelly, Junhyung Park, Ryan J. Harrigan, Nicole A. Hoff, Sarita D. Lee, Rae Wannier, Bernice Selo, Mathias Mossoko, Bathe Njoloko, Emile Okitolonda-Wemakoy, Placide Mbala-Kingebeni, George W. Rutherford, Thomas B. Smith, Steve Ahuka-Mundeke, Jean Jacques Muyembe-Tamfum, Anne W. Rimoin, and Frederic Paik Schoenberg. Real-time predictions of the 2018–2019 Ebola virus disease outbreak in the Democratic Republic of the Congo using Hawkes point process models. *Epidemics*, 28:100354, 9 2019. ISSN 17554365. doi: 10.1016/j.epidem.2019.100354. URL https://linkinghub.elsevier.com/retrieve/pii/S1755436519300258.

P Kidger, J Morrill, J Foster, and T Lyons. Neural controlled differential equations for irregular time series. In *NeurIPS*, 2020. URL https://proceedings.neurips.cc/paper/2020/file/4a5876b450b45371f6cfe5047ac8cd45-Paper.pdf.

Aaron A. King, Dao Nguyen, and Edward L. Ionides. Statistical Inference for Partially Observed Markov Processes via the R Package pomp. *Journal of Statistical Software*, 69(12):1–43, 2016. ISSN 1548-7660. doi: 10.18637/jss.v069.i12. URL http://www.jstatsoft.org/v69/i12/.

D P Kingma and M Welling. Auto-encoding variational Bayes. In *International Conference on Learning Representations*, 2014. URL https://arxiv.org/abs/1312.6114.

Roger Koenker and Gilbert Bassett. Regression Quantiles. *Econometrica*, 46(1):33–50, 1 1978. ISSN 00129682. doi: 10.2307/1913643. URL https://www.jstor.org/stable/1913643?origin=crossref.

Marcel Kollovieh, Abdul Fatir Ansari, Michael Bohlke-Schneider, Jasper Zschiegner, Hao Wang, and Yuyang (Bernie) Wang. Predict, Refine, Synthesize: Self-Guiding Diffusion Models for Probabilistic Time Series Forecasting. In *Advances in Neural Information Processing Systems 36*, pp. 28341–28364, San Diego, California, USA, 2023. Neural Information Processing Systems Foundation, Inc. (NeurIPS). ISBN 978-1-7138-9911-2. doi: 10.52202/075280-1232. URL http://www.proceedings.com/075280-1232.html.

Chien-Hung Lee, Ko Chang, Yao-Mei Chen, Jinn-Tsong Tsai, Yenming J. Chen, and Wen-Hsien Ho. Epidemic prediction of dengue fever based on vector compartment model and Markov chain Monte Carlo method. *BMC Bioinformatics*, 22(S5):118, 11 2021. ISSN 1471-2105. doi: 10.1186/s12859-021-04059-x. URL https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-021-04059-x.

Jing Lei, Max G'Sell, Alessandro Rinaldo, Ryan J. Tibshirani, and Larry Wasserman. Distribution-Free Predictive Inference for Regression. *Journal of the American Statistical Association*, 113 (523):1094–1111, 7 2018. ISSN 0162-1459. doi: 10.1080/01621459.2017.1307116. URL https://www.tandfonline.com/doi/full/10.1080/01621459.2017.1307116.

Xing Yu Leung, Rakibul M. Islam, Mohammadmehdi Adhami, Dragan Ilic, Lara McDonald, Shanika Palawaththa, Basia Diug, Saif U. Munshi, and Md Nazmul Karim. A systematic review of dengue outbreak prediction models: Current scenario and future directions. *PLOS Neglected*

Tropical Diseases, 17(2):e0010631, 2 2023. ISSN 1935-2735. doi: 10.1371/journal.pntd.0010631. URL https://dx.plos.org/10.1371/journal.pntd.0010631.

Marcos Venícius Malveira de Lima and Gabriel Zorello Laporta. Evaluation of the Models for Forecasting Dengue in Brazil from 2000 to 2017: An Ecological Time-Series Study. *Insects*, 11 (11):794, 11 2020. ISSN 2075-4450. doi: 10.3390/insects11110794. URL https://www.mdpi.com/2075-4450/11/11/794.

Y Liu, T Hu, H Zhang, H Wu, S Wang, L Ma, and M Long. iTransformer: Inverted Transformers Are Effective for Time Series Forecasting. In *ICLR*, 2024. URL https://arxiv.org/abs/2310.06625.

Rachel Lowe, Antonio Gasparrini, Cédric J. Van Meerbeeck, Catherine A. Lippi, Roché Mahon, Adrian R. Trotman, Leslie Rollock, Avery Q. J. Hinds, Sadie J. Ryan, and Anna M. Stewart-Ibarra. Nonlinear and delayed impacts of climate on dengue risk in Barbados: A modelling study. *PLOS Medicine*, 15(7):e1002613, 7 2018. ISSN 1549-1676. doi: 10.1371/journal.pmed.1002613. URL https://dx.plos.org/10.1371/journal.pmed.1002613.

Rachel Lowe, Sophie A Lee, Kathleen M O'Reilly, Oliver J Brady, Leonardo Bastos, Gabriel Carrasco-Escobar, Rafael de Castro Catão, Felipe J Colón-González, Christovam Barcellos, Marilia Sá Carvalho, Marta Blangiardo, Håvard Rue, and Antonio Gasparrini. Combined effects of hydrometeorological hazards and urbanisation on dengue risk in Brazil: a spatiotemporal modelling study. *The Lancet Planetary Health*, 5(4):e209–e219, 4 2021. ISSN 25425196. doi: 10.1016/S2542-5196(20)30292-8. URL https://linkinghub.elsevier.com/retrieve/pii/S2542519620302928.

Rajendra Maharaj. Early warning systems for the detection of malaria outbreaks. *Indian Journal of Medical Research*, 146(5):560–562, 11 2017. ISSN 0971-5916. doi: 10.4103/ijmr.IJMR{_}933{_}17. URL https://ijmr.org.in/early-warning-systems-for-the-detection-of-malaria-outbreaks/.

Mokhalad A. Majeed, Helmi Z. M. Shafri, Aimrun Wayayok, and Zed Zulkafli. Prediction of dengue cases using the attention-based long short-term memory (LSTM) approach. *Geospatial Health*, 18(1), 5 2023. ISSN 1970-7096. doi: 10.4081/gh.2023.1176. URL https://www.geospatialhealth.net/index.php/gh/article/view/1176.

Edson Zangiacomi Martinez, Elisângela Aparecida Soares da Silva, and Amaury Lelis Dal Fabbro. A SARIMA forecasting model to predict the number of cases of dengue in Campinas, State of São Paulo, Brazil. *Revista da Sociedade Brasileira de Medicina Tropical*, 44(4):436–440, 8 2011. ISSN 0037-8682. doi: 10.1590/S0037-86822011000400007. URL http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0037-86822011000400007&lng=en&tlng=en.

Camilo Mora, Tristan McKenzie, Isabella M. Gaw, Jacqueline M. Dean, Hannah von Hammerstein, Tabatha A. Knudson, Renee O. Setter, Charlotte Z. Smith, Kira M. Webster, Jonathan A. Patz, and Erik C. Franklin. Over half of known human pathogenic diseases can be aggravated by climate change. *Nature Climate Change*, 12(9):869–875, 9 2022. ISSN 1758-678X. doi: 10.1038/s41558-022-01426-1. URL https://www.nature.com/articles/s41558-022-01426-1.

Elisa Mussumeci and Flávio Codeço Coelho. Large-scale multivariate forecasting models for Dengue - LSTM versus random forest regression. *Spatial and Spatio-temporal Epidemiology*, 35:100372, 11 2020. ISSN 18775845. doi: 10.1016/j.sste.2020.100372. URL https://linkinghub.elsevier.com/retrieve/pii/S1877584520300502.

Y Nie, N H Nguyen, P Sinthong, and J Kalagnanam. A Time Series is Worth 64 Words: Long-term Forecasting with Transformers. In *ICLR*, 2023. URL https://arxiv.org/abs/2211.14730.

D.A. Nix and A.S. Weigend. Estimating the mean and variance of the target probability distribution. In *Proceedings of 1994 IEEE International Conference on Neural Networks (ICNN'94)*, pp. 55–60. IEEE, 1994. ISBN 0-7803-1901-X. doi: 10.1109/ICNN.1994.374138. URL http://ieeexplore.ieee.org/document/374138/.

Rachel J. Oidtman, Elisa Omodei, Moritz U. G. Kraemer, Carlos A. Castañeda-Orjuela, Erica Cruz-Rivera, Sandra Misnaza-Castrillón, Myriam Patricia Cifuentes, Luz Emilse Rincon, Viviana Cañon, Pedro de Alarcon, Guido España, John H. Huber, Sarah C. Hill, Christopher M. Barker, Michael A. Johansson, Carrie A. Manore, Robert C. Reiner, Isabel Rodriguez-Barraquer, Amir S. Siraj, Enrique Frias-Martinez, Manuel García-Herranz, and T. Alex Perkins. Trade-offs between individual and ensemble forecasts of an emerging infectious disease. *Nature Communications*, 12(1):5379, 9 2021. ISSN 2041-1723. doi: 10.1038/s41467-021-25695-0. URL https://www.nature.com/articles/s41467-021-25695-0.

R.A. Ranga Prabodanie, Sergei Schreider, Bernard Cazelles, and Lewi Stone. Coherence of dengue incidence and climate in the wet and dry zones of Sri Lanka. *Science of The Total Environment*, 724:138269, 7 2020. ISSN 00489697. doi: 10.1016/j.scitotenv.2020.138269. URL https://linkinghub.elsevier.com/retrieve/pii/S0048969720317824.

Md. Siddikur Rahman, Miftahuzzannat Amrin, and Md. Abu Bokkor Shiddik. Dengue Early Warning System and Outbreak Prediction Tool in Bangladesh Using Interpretable Tree-Based Machine Learning Model. *Health Science Reports*, 8(5):e70726, 5 2025. ISSN 2398-8835. doi: 10.1002/hsr2.70726. URL https://onlinelibrary.wiley.com/doi/10.1002/hsr2.70726.

K Rasul, C Seward, I Schuster, and R Vollgraf. Autoregressive Denoising Diffusion Models for Multivariate Probabilistic Time Series Forecasting. In *ICML*, 2021. URL https://proceedings.mlr.press/v139/rasul21a/rasul21a.pdf.

David W. Redding, Rory Gibb, Chioma C. Dan-Nwafor, Elsie A. Ilori, Rimamdeyati Usman Yashe, Saliu H. Oladele, Michael O. Amedu, Akanimo Iniobong, Lauren A. Attfield, Christl A. Donnelly, Ibrahim Abubakar, Kate E. Jones, and Chikwe Ihekweazu. Geographical drivers and climate-linked dynamics of Lassa fever in Nigeria. *Nature Communications*, 12(1):5759, 10 2021. ISSN 2041-1723. doi: 10.1038/s41467-021-25910-y. URL https://www.nature.com/articles/s41467-021-25910-y.

Nicholas G. Reich, Logan C. Brooks, Spencer J. Fox, Sasikiran Kandula, Craig J. McGowan, Evan Moore, Dave Osthus, Evan L. Ray, Abhinav Tushar, Teresa K. Yamana, Matthew Biggerstaff, Michael A. Johansson, Roni Rosenfeld, and Jeffrey Shaman. A collaborative multiyear, multimodel assessment of seasonal influenza forecasting in the United States. *Proceedings of the National Academy of Sciences*, 116(8):3146–3154, 2 2019. ISSN 0027-8424. doi: 10.1073/pnas.1812594116. URL https://pnas.org/doi/full/10.1073/pnas.1812594116.

S Rifai, P Vincent, X Muller, X Glorot, and Y Bengio. Contractive auto-encoders: Explicit invariance during feature extraction. In *Proceedings of the 28th International Conference on Machine Learning*, 2011. URL https://icml.cc/2011/papers/455_icmlpaper.pdf.

Y Rubanova, R T Q Chen, and D Duvenaud. Latent Ordinary Differential Equations for Irregularly-Sampled Time Series. In *NeurIPS*, 2019. URL https://proceedings.neurips.cc/paper/2019/hash/42a6845a557bef704ad8ac9cb4461d43-Abstract.html.

A. K. Sahai, Raju Mandal, Susmitha Joseph, Shubhayu Saha, Pradip Awate, Somenath Dutta, Avijit Dey, Rajib Chattopadhyay, R. Phani, D. R. Pattanaik, and Sunil Despande. Development of a probabilistic early health warning system based on meteorological parameters. *Scientific Reports*, 10(1):14741, 9 2020. ISSN 2045-2322. doi: 10.1038/s41598-020-71668-6. URL https://www.nature.com/articles/s41598-020-71668-6.

Nurul Azam Mohd Salim, Yap Bee Wah, Caitlynn Reeves, Madison Smith, Wan Fairos Wan Yaacob, Rose Nani Mudin, Rahmat Dapari, Nik Nur Fatin Fatihah Sapri, and Ubydul Haque. Prediction of dengue outbreak in Selangor Malaysia using machine learning techniques. *Scientific Reports*, 11(1):939, 1 2021. ISSN 2045-2322. doi: 10.1038/s41598-020-79193-2. URL https://www.nature.com/articles/s41598-020-79193-2.

T Salimans, I Goodfellow, W Zaremba, V Cheung, A Radford, and X Chen. Improved techniques for training GANs. In *Advances in Neural Information Processing Systems 29*, 2016. URL https://papers.nips.cc/paper/6125-improved-techniques-for-training-gans.pdf.

David Salinas, Valentin Flunkert, Jan Gasthaus, and Tim Januschowski. DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, 36(3): 1181–1191, 7 2020. ISSN 01692070. doi: 10.1016/j.ijforecast.2019.07.001. URL https://linkinghub.elsevier.com/retrieve/pii/S0169207019301888.

Alessandro Sebastianelli, Dario Spiller, Raquel Carmo, James Wheeler, Artur Nowakowski, Ludmilla Viana Jacobson, Dohyung Kim, Hanoch Barlevi, Zoraya El Raiss Cordero, Felipe J Colón-González, Rachel Lowe, Silvia Liberata Ullo, and Rochelle Schneider. A reproducible ensemble machine learning approach to forecast dengue outbreaks. *Scientific Reports*, 14(1): 3807, 2 2024. ISSN 2045-2322. doi: 10.1038/s41598-024-52796-9. URL https://www.nature.com/articles/s41598-024-52796-9.

Farah Shahid, Aneela Zameer, and Muhammad Muneeb. Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM. *Chaos, Solitons & Fractals*, 140:110212, 11 2020. ISSN 09600779. doi: 10.1016/j.chaos.2020.110212. URL https://linkinghub.elsevier.com/retrieve/pii/S0960077920306081.

J T H Smith, A Warrington, and S W Linderman. Simplified State Space Layers for Sequence Modeling. In *ICLR*, 2023. URL https://github.com/lindermanlab/S5.

Yacob Sophia, Mathew Koll Roxy, Raghu Murtugudde, Anand Karipot, Amir Sapkota, Panini Dasgupta, Kalpana Baliwant, Sujata Saunik, Abhiyant Tiwari, Rajib Chattopadhyay, and Revati K. Phalkey. Dengue dynamics, predictions, and future increase under changing monsoon climate in India. *Scientific Reports*, 15(1):1637, 1 2025. ISSN 2045-2322. doi: 10.1038/s41598-025-85437-w. URL https://www.nature.com/articles/s41598-025-85437-w.

A. M. Stuart. Inverse problems: A Bayesian perspective. *Acta Numerica*, 19:451–559, 5 2010. ISSN 0962-4929. doi: 10.1017/S0962492910000061. URL https://www.cambridge.org/core/product/identifier/S0962492910000061/type/journal_article.

Yusuke Tashiro, Jiaming Song, Yang Song, and Stefano Ermon. CSDI: Conditional score-based diffusion models for probabilistic time-series imputation. In *NeurIPS*, 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/cfe8504bda37b575c70ee1a8276f3486-Paper.pdf.

Naftali Tishby and Noga Zaslavsky. Deep learning and the information bottleneck principle. In *2015 IEEE Information Theory Workshop (ITW)*, pp. 1–5. IEEE, 4 2015. ISBN 978-1-4799-5524-4. doi: 10.1109/ITW.2015.7133169. URL http://ieeexplore.ieee.org/document/7133169/.

H. Juliette T. Unwin, Isobel Routledge, Seth Flaxman, Marian-Andrei Rizoiu, Shengjie Lai, Justin Cohen, Daniel J. Weiss, Swapnil Mishra, and Samir Bhatt. Using Hawkes Processes to model imported and local malaria cases in near-elimination settings. *PLOS Computational Biology*, 17 (4):e1008830, 4 2021. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1008830. URL https://dx.plos.org/10.1371/journal.pcbi.1008830.

A Vaswani, N Shazeer, N Parmar, J Uszkoreit, L Jones, A N Gomez, L Kaiser, and I Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, 2017. URL https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html.

Svitlana Volkova, Ellyn Ayton, Katherine Porterfield, and Courtney D. Corley. Forecasting influenza-like illness dynamics for military populations using neural networks and social media. *PLOS ONE*, 12(12):e0188941, 12 2017. ISSN 1932-6203. doi: 10.1371/journal.pone.0188941. URL https://dx.plos.org/10.1371/journal.pone.0188941.

V Vovk, A Gammerman, and G Shafer. *Algorithmic Learning in a Random World*. Springer-Verlag, New York, 2005. ISBN 0-387-00152-2. doi: 10.1007/b106715. URL http://link.springer.com/10.1007/b106715.

Jenna Wiens and Erica S Shenoy. Machine Learning for Healthcare: On the Verge of a Major Shift in Healthcare Epidemiology. *Clinical Infectious Diseases*, 66(1):149–153, 1 2018. ISSN 1058-4838. doi: 10.1093/cid/cix731. URL https://academic.oup.com/cid/article/66/1/149/4085880.

Dongxia Wu, Liyao Gao, Xinyue Xiong, Matteo Chinazzi, Alessandro Vespignani, Yi-An Ma, and Rose Yu. DeepGLEAM: A hybrid mechanistic and deep learning model for COVID-19 forecasting. arXiv preprint arXiv:2102.06684, 3 2021a. URL https://doi.org/10.48550/arXiv.2102.06684.

H Wu, J Xu, J Wang, and M Long. Autoformer: Decomposition transformers with autocorrelation for long-term series forecasting. In *NeurIPS*, 2021b. URL https://proceedings.neurips.cc/paper/2021/hash/bcc0d400288793e8bdcd7c19a8ac0c2b-Abstract.html.

H Wu, T Hu, Y Liu, H Zhou, J Wang, and M Long. TimesNet: Temporal 2D-variation modeling for general time-series analysis. In *ICLR*, 2023. URL https://github.com/thuml/TimesNet.

Neo Wu, Bradley Green, Xue Ben, and Shawn O'Banion. Deep Transformer Models for Time Series Forecasting: The Influenza Prevalence Case. arXiv:2001.08317, 1 2020. URL https://doi.org/10.48550/arXiv.2001.08317.

Feng Xie, Zhong Zhang, Liang Li, Bin Zhou, and Yusong Tan. EpiGNN: Exploring Spatial Transmission with Graph Neural Network for Regional Epidemic Forecasting. In *Machine Learning and Knowledge Discovery in Databases. ECML PKDD 2022*, volume 13718, pp. 469–485. Springer, Cham, 2023. doi: 10.1007/978-3-031-26422-1{_}29. URL https://link.springer.com/10.1007/978-3-031-26422-1_29.

Alireza Yazdani, Lu Lu, Maziar Raissi, and George Em Karniadakis. Systems biology informed deep learning for inferring parameters and hidden dynamics. *PLOS Computational Biology*, 16 (11):e1007575, 11 2020. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1007575. URL https://dx.plos.org/10.1371/journal.pcbi.1007575.

Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. Are Transformers Effective for Time Series Forecasting? *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(9):11121–11128, 6 2023. ISSN 2374-3468. doi: 10.1609/aaai.v37i9.26317. URL https://ojs.aaai.org/index.php/AAAI/article/view/26317.

Y Zhang and J Yan. Crossformer: Transformer Utilizing Cross-Dimension Dependency for Multivariate Time Series Forecasting. In *ICLR*, 2023. URL https://github.com/Thinklab-SJTU/Crossformer.

Naizhuo Zhao, Katia Charland, Mabel Carabali, Elaine O. Nsoesie, Mathieu Maheu-Giroux, Erin Rees, Mengru Yuan, Cesar Garcia Balaguera, Gloria Jaramillo Ramirez, and Kate Zinszer. Machine learning and dengue forecasting: Comparing random forests and artificial neural networks for predicting dengue burden at national and sub-national scales in Colombia. *PLOS Neglected Tropical Diseases*, 14(9):e0008056, 9 2020. ISSN 1935-2735. doi: 10.1371/journal.pntd.0008056. URL https://dx.plos.org/10.1371/journal.pntd.0008056.

Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(12):11106–11115, 5 2021. ISSN 2374-3468. doi: 10.1609/aaai.v35i12.17325. URL https://ojs.aaai.org/index.php/AAAI/article/view/17325.

T Zhou, Z Ma, Q Wen, Xue Wang, Liang Sun, and Rong Jin. FEDformer: Frequency-enhanced decomposed transformer for long-term series forecasting. In *ICML*, 2022. URL https://proceedings.mlr.press/v162/zhou22g.html.

Xianglei Zhu, Bofeng Fu, Yaodong Yang, Yu Ma, Jianye Hao, Siqi Chen, Shuang Liu, Tiegang Li, Sen Liu, Weiming Guo, and Zhenyu Liao. Attention-based recurrent neural network for influenza epidemic prediction. *BMC Bioinformatics*, 20(S18):575, 11 2019. ISSN 1471-2105. doi: 10.1186/s12859-019-3131-8. URL https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-019-3131-8.

A FORECASTING METHODS AND INFECTIOUS-DISEASE APPLICATIONS, 2011–2025

Table 4: Inventory of forecasting methods and infectious-disease applications. Original method papers without an infectious-disease forecasting application are identified as general methods. Family groups by architecture; cadence is the reporting interval of the target series. Structural annotations refer to C1–C4 in Section 2 and are our assessment, not claims made by the cited authors.

Family / method	Method description	Application	Reference / lacks
Statistical SARIMA	Seasonal ARIMA	Year: 2011 Disease: Dengue Region: Brazil Cadence: Monthly	(Martinez et al., 2011) Lacks: C1,C2,C3
Statistical Poisson regression	Time-series multivariate Poisson regression	Year: 2012 Disease: Dengue Region: Singapore Cadence: Weekly	(Hii et al., 2012) Lacks: C2,C3
Statistical Holt–Winters ensemble	Seasonal smoothing, analogues, and historical models	Year: 2018 Disease: Dengue Region: Iquitos, Peru; San Juan, Puerto Rico Cadence: Weekly	(Buczak et al., 2018) Lacks: C1,C2,C3
Statistical Prophet	Additive Bayesian	Year: 2020 Disease: COVID-19 Region: Multi-country Cadence: Daily	(Battineni et al., 2025) Lacks: C1,C2,C3
Statistical TBATS	Trigonometric seasonality, Box-Cox transformation, ARMA errors, trend and seasonal components	Year: 2020 Disease: Dengue Region: Brazil, 26 states and Federal District Cadence: Monthly	(Lima & Laporta, 2020) Lacks: C1,C2,C3
Statistical DLNM-INLA	Distributed-lag INLA	Year: 2021 Disease: Dengue Region: Brazil Cadence: Monthly	(Lowe et al., 2021) Lacks: C2,C3
Statistical NB-GLMM superensemble	Bayesian negative-binomial generalized linear mixed models	Year: 2021 Disease: Dengue Region: Vietnam Cadence: Monthly	(Colón-González et al., 2021) Lacks: C2,C3
Tree GBM	Gradient-boosted regression trees	Year: 2017 Disease: Dengue Region: Guangdong, China Cadence: Weekly	(Guo et al., 2017) Lacks: C1,C2,C3,C4
Tree CatBoost	Categorical Boosting	Year: 2024 Disease: Dengue Region: Brazil; transfer evaluation in Peru Cadence: Monthly	(Sebastianelli et al., 2024) Lacks: C1,C2,C3,C4
RNN LSTM	Long Short-Term Memory	Year: 2017 Disease: Influenza Region: US Cadence: Weekly	(Volkova et al., 2017) Lacks: C1,C2
RNN LSTM	Long Short-Term Memory	Year: 2020 Disease: Dengue Region: Brazil, 790 cities Cadence: Weekly	(Mussumeci & Codeço Coelho, 2020) Lacks: C1,C2
RNN LSTM	Long Short-Term Memory	Year: 2020 Disease: COVID-19 Region: Canada Cadence: Daily	(Chimmula & Zhang, 2020) Lacks: C1,C2

Table 4 (continued)

Family / method	Method description	Application	Reference / lacks
RNN GRU	Gated Recurrent Unit	Year: 2020 Disease: COVID-19 Region: Multi-country Cadence: Daily	(Shahid et al., 2020) Lacks: C1,C2
RNN Att-MCLSTM	Attention-based multi-channel LSTM	Year: 2019 Disease: Influenza-like illness Region: Guangzhou, China Cadence: Weekly	(Zhu et al., 2019) Lacks: C1,C2
RNN ConvLSTM2D spatial-processing component	Convolutional LSTM	Year: 2024 Disease: Dengue Region: Brazil Cadence: Monthly	(Sebastianelli et al., 2024) Lacks: C1,C2
RNN Stacked LSTM	Multi-layer LSTM	Year: 2020 Disease: Dengue Region: Brazil, 790 cities Cadence: Weekly	(Mussumeci & Codeço Coelho, 2020) Lacks: C1,C2
Transformer Transformer	Vanilla attention	Year: 2020 Disease: Influenza Region: US Cadence: Weekly	(Wu et al., 2020) Lacks: C2,C3,C4
Transformer Informer	ProbSparse attention	Year: 2021 General method Energy and weather series Cadence: Subdaily	(Zhou et al., 2021) Lacks: C2,C3,C4
Transformer Autoformer	Decomposition attention	Year: 2021 Disease: ILI benchmark Region: US Cadence: Weekly	(Wu et al., 2021b) Lacks: C2,C3,C4
Transformer FEDformer	Frequency-enhanced	Year: 2022 Disease: ILI benchmark Region: US Cadence: Weekly	(Zhou et al., 2022) Lacks: C2,C3,C4
Transformer PatchTST	Patch Time-Series	Year: 2023 Disease: ILI benchmark Region: US Cadence: Weekly	(Nie et al., 2023) Lacks: C2,C3,C4
Convolutional TimesNet	Temporal 2D-variation modeling	Year: 2023 Disease: ILI benchmark Region: US Cadence: Weekly	(Wu et al., 2023) Lacks: C2,C3,C4
Transformer iTransformer	Inverted Transformer	Year: 2024 General method Multivariate forecasting Cadence: Variable	(Liu et al., 2024) Lacks: C2,C3,C4
Transformer Crossformer	Cross-dimension attention	Year: 2023 Disease: ILI benchmark Region: US Cadence: Weekly	(Zhang & Yan, 2023) Lacks: C2,C3,C4
Linear DLinear	Decomposition linear	Year: 2023 Disease: ILI benchmark Region: US Cadence: Weekly	(Zeng et al., 2023) Lacks: C2,C3,C4
State-space S4	Structured state-space model	Year: 2022 General method Sequence modeling and forecasting Cadence: Variable	(Gu et al., 2022) Lacks: C2,C4

Table 4 (continued)

Family / method	Method description	Application	Reference / lacks
State-space S5	Simplified state-space layers	Year: 2023 General method Sequence modeling Cadence: Variable	(Smith et al., 2023) Lacks: C2,C4
State-space Mamba	Selective state-space model	Year: 2024 General method Language, audio, and genomics Cadence: Variable	(Gu & Dao, 2024) Lacks: C2,C4
Neural-ODE Neural ODE	Neural Ordinary DE	Year: 2020 Disease: COVID-19 Region: Wuhan Cadence: Daily	(Dandekar & Barbastathis, 2020) Lacks: C2,C4
Neural-ODE Latent ODE	Latent neural ordinary differential equations	Year: 2019 General method Irregular time series Cadence: Variable	(Rubanova et al., 2019) Lacks: C2,C4
Neural-ODE Neural CDE	Neural controlled differential equations	Year: 2020 General method Irregular time series Cadence: Variable	(Kidger et al., 2020) Lacks: C2,C4
Diffusion TimeGrad	Autoregressive denoising diffusion	Year: 2021 General method Multivariate forecasting Cadence: Variable	(Rasul et al., 2021) Lacks: C1,C2
Diffusion CSDI	Conditional score-based diffusion imputation	Year: 2021 General method Time-series imputation Cadence: Variable	(Tashiro et al., 2021) Lacks: C1,C2
Diffusion TSDiff	Time-series diffusion	Year: 2023 General method Multivariate forecasting Cadence: Variable	(Kollovieh et al., 2023) Lacks: C1,C2
Graph Cola-GNN	Cross-location attention	Year: 2020 Disease: Influenza Region: US states Cadence: Weekly	(Deng et al., 2020) Lacks: C2,C4
Graph EpiGNN	Epidemic GNN	Year: 2022 Disease: Influenza Region: US/Japan Cadence: Weekly	(Xie et al., 2023) Lacks: C2,C4
Graph STAN	Spatio-temporal GNN	Year: 2021 Disease: COVID-19 Region: US counties Cadence: Daily	(Gao et al., 2021) Lacks: C2,C4
Deep-epi EpiDeep	Epi Deep	Year: 2019 Disease: Influenza Region: US Cadence: Weekly	(Adhikari et al., 2019) Lacks: C2
Hybrid Systems-biology-informed neural network	ODE-constrained inference	Year: 2020 General method Biochemical system dynamics Cadence: Variable	(Yazdani et al., 2020) Lacks: C4
Hybrid DeepGLEAM	GLEAM + diffusion convolutional recurrent neural network	Year: 2021 Disease: COVID-19 mortality Region: 50 US states Cadence: Weekly forecast targets; daily inputs	(Wu et al., 2021a) Lacks: C4

Table 4 (continued)

Family / method	Method description	Application	Reference / lacks
Hybrid EpiCovDA	Mechanistic model with variational data assimilation	Year: 2021 Disease: COVID-19 cases and deaths Region: US states, DC, and Puerto Rico Cadence: Daily observations; weekly evaluation horizons	(Biegel & Lega, 2021) Lacks: C4

B Mapping probabilistic forecasting tools to SPECTRA's structural priors and operational constraints

Table 5: Mapping probabilistic forecasting tools to SPECTRA's structural priors and operational constraints.

Prior	Constraint	Tool	SPECTRA use	Operational role	References
C1	Climate ambiguity	Granger framing	One-way climate input	External driver only	(Granger, 1969)
C1	Climate collinearity	Latent compression	Frozen climate latent	Stable climate signal	(Kingma & Welling, 2014; Higgins et al., 2017; Rifai et al., 2011)
C1	No feedback leakage	Directed attention	Dengue queries climate	separate one-way exogenous climate encoding	(Granger, 1969; Vaswani et al., 2017)
C2	Incidence memory	Self-excitation	Disease history tower	Cases seed cases	(Hawkes, 1971; Kelly et al., 2019; Unwin et al., 2021)
C2	Short-term dependence	Autoregression	Last incidence input	Recent-case memory	(Reich et al., 2019; Salinas et al., 2020)
C2	Multi-scale history	Temporal convolutions	Causal TCN stream	Lagged disease signal	(Bai et al., 2018)
C3	State instability	Gated recurrence	LSTM disease tower	Controlled memory	(Hochreiter & Schmidhuber, 1997; Cho et al., 2014)
C3	Regime shifts	Surge gate	Recent/long z-score	Outbreak switching	(Hochreiter & Schmidhuber, 1997)
C3	Critic instability	Gradient penalty	WGAN-GP critic	Stable adversarial loss	(Gulrajani et al., 2017)
C4	Full distribution	Proper scores	Distributional evaluation	Calibration and sharpness	(Gneiting & Raftery, 2007; Bracher et al., 2021; Cramer et al., 2022)
C4	Variable noise	Gaussian NLL	Mean–variance head	Time-varying uncertainty	(Nix & Weigend, 1994; Salinas et al., 2020)
C4	Upper-tail risk	Pinball loss	Quantile heads	Tail probability support	(Koenker & Bassett, 1978)

Table 5 (continued)

Prior	Constraint	Tool	SPECTRA use	Operational role	References
C4	Outbreak shape	Feature matching	Critic features	Realistic trajectories	(Salimans et al., 2016; Gulrajani et al., 2017)
C4	Sequence realism	Wasserstein loss	Trajectory critic	Distribution matching	(Arjovsky et al., 2017; Gulrajani et al., 2017)
C4	Valid intervals	Split conformal	Residual calibration	Finite-sample coverage	(Vovk et al., 2005; Lei et al., 2018; Angelopoulos & Bates, 2023)
C4	Alert decision	Threshold risk	Gaussian-head exceedance scores	Crossing probability	(Colón-González et al., 2021)

C EQUATIONS

Table 6: Mapping of structural priors C1–C4 to framework and operator components. Equations in Appendix C.

Prior	Mechanism	Realisation in framework	Realisation in operator
C1	separate one-way exogenous climate encoding	Frozen Cβ-VAE on climate-only inputs	Same Cβ-VAE + tanh-bounded projection
C2	Hawkes-style decaying incidence memory	Causal TCN over disease history	Closed-form recursion $e_t = e^{-\beta_n} e_{t-1} + \alpha_n \,\mathrm{ReLU}(y_{t-1})$ + lag bank
C3	Bounded gated update	LSTM with multiplicative gates	Gated convex combination with $\|h_t\|_\infty \leq 1$
C4	Persistence-skip emission	Persistence-skip Gaussian + quantile heads	Same heads + softplus-monotone quantiles

Notation: θ, ϕ, ψ – operator, encoder, critic parameters; n – country index; L – stacked operator cells.

Table 7: Equations of the SPECTRA framework and the canonical operator, organised by structural prior C1–C4 and by training/calibration role.

C1 Encoder pre-training: frozen post-training
$\mathcal{L}_{\mathrm{VAE}}(\phi) = \mathbb{E}\|c - \hat{c}\|_2^2 + \beta D_{\mathrm{KL}}(q_\phi \| \mathcal{N}(0, I)) + \lambda_J \|\partial \mu_\phi / \partial c\|_F^2$

C1 Bounded projection: operator
$c_t^{\mathrm{lat}} = \tanh(W_{\mathrm{cl}} x_t^{\mathrm{clim}}), \quad \ell_t^{\mathrm{lat}} = \tanh(W_{\mathrm{lag}} x_t^{\mathrm{lag}})$

C2 Hawkes recursion: $\alpha_n = \exp(\bar{\alpha}_n),\ \beta_n = \exp(\bar{\beta}_n)$, per-country
$e_t = e^{-\beta_n} e_{t-1} + \alpha_n \,\mathrm{ReLU}(y_{t-1})$

C3 Surge gate: local response modulator
$r_t = \sigma(W_r [c_t^{\mathrm{lat}}; e_t; h_{t-1}])$

C3 Cell input: concatenation

$u_t = [c_t^{\text{lat}}; \ell_t^{\text{lat}}; y_{t-1}; e_t; r_t; h_{t-1}]$

C3 Keep-gate; candidate: both bounded
$g_t = \sigma(W_g u_t), \quad f_t = \tanh(W_f u_t)$

C3 Hidden state: convex update; $\|h_t\|_\infty \le 1$
$h_t = g_t \odot h_{t-1} + (1 - g_t) \odot f_t$

C4 Gaussian head: persistence skip; variance floor
$\hat{\mu}_t = w_p y_{t-1} + W_\mu h_t, \quad \log \hat{\sigma}_t = \text{clamp}(W_\sigma h_t, -1.5, 3)$

C4 Median quantile: anchored
$\hat{q}_t^{0.50} = w_p y_{t-1} + W_{q,0.5} h_t$

C4 Lower deltas: monotone by construction
$\hat{q}_t^{0.10} = \hat{q}_t^{0.50} - \text{softplus}(\Delta_{10}), \quad \hat{q}_t^{0.05} = \hat{q}_t^{0.10} - \text{softplus}(\Delta_{05})$

C4 Upper deltas: monotone by construction
$\hat{q}_t^{0.90} = \hat{q}_t^{0.50} + \text{softplus}(\Delta_{90}), \quad \hat{q}_t^{0.95} = \hat{q}_t^{0.90} + \text{softplus}(\Delta_{95})$

Framework loss: weights annealed
$\mathcal{L}_{\text{fwk}}(\theta, \psi) = w_n \mathcal{L}_{\text{NLL}} + w_q \mathcal{L}_{\text{pinball}} + w_{\text{tv}} \mathcal{L}_{\text{TV}} + w_{\ell_1} \mathcal{L}_{\ell_1} + w_f \mathcal{L}_{\text{feat}} + w_a \mathcal{L}_{\text{WGAN-GP}}$

Operator loss: no critic
$\mathcal{L}_{\text{op}}(\theta) = w_n \mathcal{L}_{\text{NLL}} + w_q \mathcal{L}_{\text{pinball}} + w_{\text{tv}} \mathcal{L}_{\text{TV}} + w_{\ell_1} \mathcal{L}_{\ell_1} + w_v \mathcal{L}_{\text{var-anchor}}$

Conformal score: per country n
$\rho_n^\alpha = \text{Quantile}_{1-\alpha}(\{|y_t - \hat{\mu}_t| / \hat{\sigma}_t : t \in \mathcal{C}_n\})$

Calibrated upper quantile: finite-sample coverage
$\widetilde{q}_{t+h,n}^{1-\alpha} = \hat{\mu}_{t+h} + \rho_n^\alpha \hat{\sigma}_{t+h}$

Threshold-crossing: analytic from Gaussian head
$\Pr_\theta(y_{t+h} > \kappa \mid y_{1:t}; c_{1:t}) = 1 - \Phi((\kappa - \hat{\mu}_{t+h}) / \hat{\sigma}_{t+h})$

D REDUCTION MAP: FRAMEWORK TO CANONICAL OPERATOR

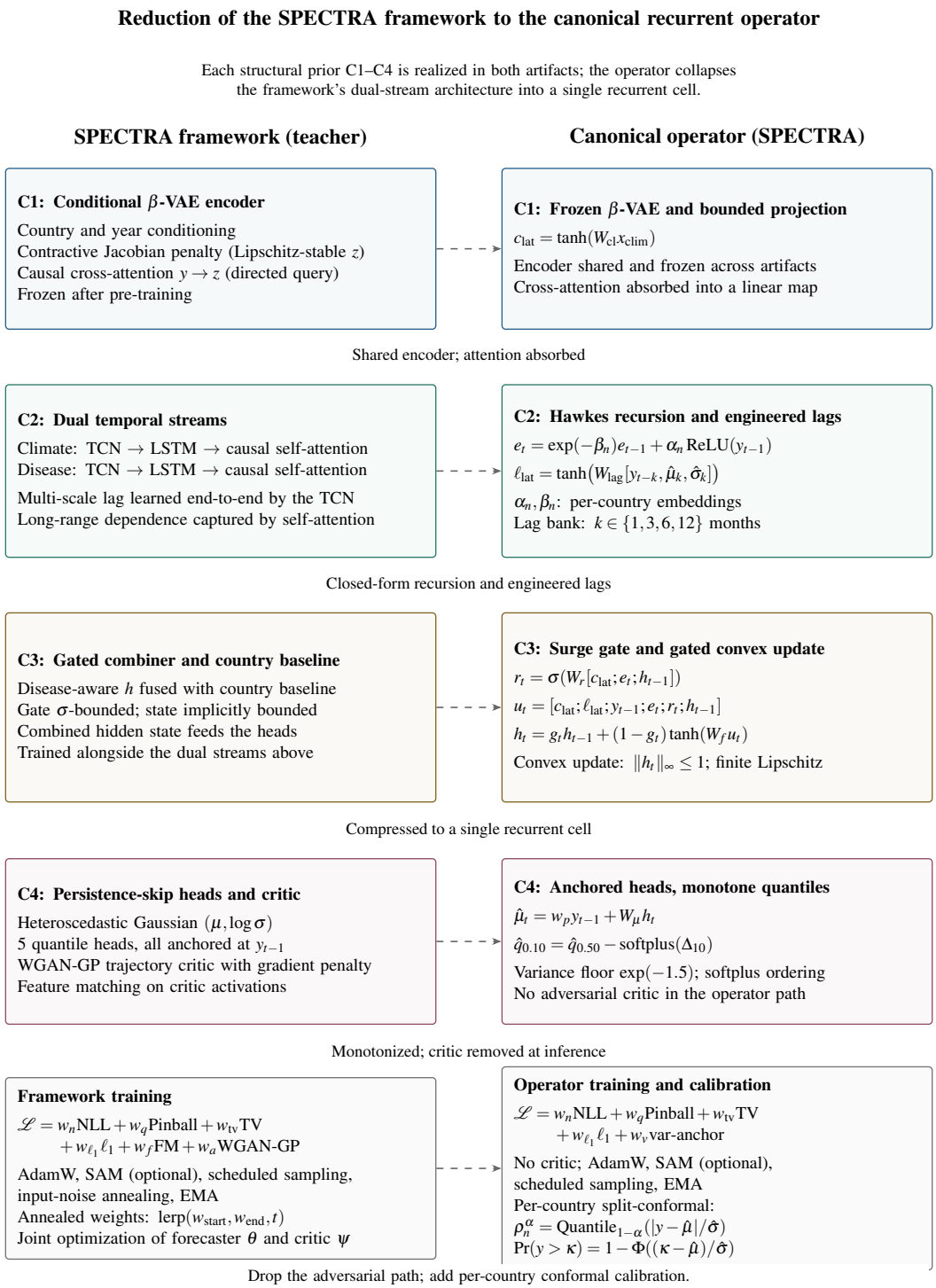


Figure 3: Framework →operator reduction. Rows: priors C1–C4. Dashed arrows: reduction operation per module. Bottom strip: training regimes (Wasserstein-GP critic vs. strictly proper scoring + split-conformal).

E SETUP, PANEL CONSTRUCTION, AND PROTOCOL

Table 8: Country tiers of the harmonised dengue panel (1990–2024 monthly). Tier A (training-time): Australia, Bangladesh, Cambodia, China, Colombia, Indonesia, Lao PDR, Malaysia, Mexico, Nicaragua, Peru, Saudi Arabia, Sri Lanka, Thailand, Venezuela, Viet Nam. Per-country reporting requires at least two eligible test observations: China has one and Malaysia has none. The pooled 24-month-window evaluation includes 159 observations; Table 10 reports 158 after excluding China's single observation.

Tier	Inclusion criterion	Countries	Use
A	At least 15 years and at least 50% nonzero incidence months	16	Training and pooled headline evaluation
B	At least 10 years and at least 30% nonzero months, but not A	22	Transfer evaluation
C	At least 5 years but not meeting A/B criteria	21	Data-scarce transfer evaluation
X	Fewer than 5 years	31	Insufficient history for model evaluation

Table 9: Operator training and calibration protocol. In the matched comparison, the framework generator uses learning rate 5×10^{-4}, weight decay 10^{-4}, and gradient clipping 1.0, rather than the operator values below.

Item	Value
Forecast horizon H	1 month
Sequence length	24 months for headline training; 60 months for the checkpoint-based transfer and delay evaluations
Validation / test windows	12 / 12 months; headline benchmark: fold 0, seeds 42–46
Seeds per (model, fold)	5
Operator hidden state d_h	128
Climate latent d_z	16
Optimizer	AdamW, lr 3×10^{-4}, wd 10^{-3}
Gradient clip	0.5
Scheduled-sampling probability	$0 \rightarrow 0.30$ linear over training
EMA decay	0.99
Variance floor $\log \sigma$	-1.5
Quantile set $\mathcal{T}$	0.05, 0.10, 0.50, 0.90, 0.95
Conformal level 1-α	0.90 (per country)
Outbreak threshold κ	Training-only country/calendar-month mean plus 1.5 standard deviations

F PER-COUNTRY AND FULL-RANKING RESULTS (TIER A)

Table 10: Per-country held-out test metrics for the canonical operator on the Tier A panel (n = test months in fold).

Country	n	SMAPE ↓	RMSE ↓	R^2 ↑	COV90
Nicaragua	12	19.55	148.23	0.982	1.000
Indonesia	12	25.59	3.40	0.728	0.917
Thailand	12	26.47	11.75	0.464	1.000
Mexico	12	28.43	254.94	0.454	1.000
Australia	12	32.35	0.60	-0.130	0.750
Viet Nam	2	32.57	29.58	-0.656	1.000
Saudi Arabia	12	33.41	2.75	0.641	1.000
Colombia	12	34.68	401.16	0.361	1.000
Cambodia	12	34.79	17.70	0.261	1.000
Sri Lanka	12	37.34	29.39	0.571	0.833
Lao PDR	12	37.36	42.38	0.714	1.000
Peru	12	54.93	661.75	0.053	1.000
Bangladesh	12	65.55	16.49	0.249	1.000
Venezuela	12	132.19	27.09	-8.270	0.917

Table 11: Ranking of the 36 models (34 baselines plus SPECTRA and SPECTRA+FT) on fold 0 of the data-rich cohort; SPECTRA values are the mean ± standard deviation across seeds 42–46.

Model	SMAPE ↓	SMAPE std	R^2
SPECTRA	43.26	1.80	0.398
SPECTRA + FT	43.27	1.33	0.587
LastValue	46.01	0.00	-0.694
RandomForest	54.17	0.52	0.417
HistGB	54.45	0.00	0.629
XGBoost	54.58	0.00	0.463
GradientBoost	57.56	0.85	0.682
LightGBM	57.74	0.00	0.549
PersistSeas	60.42	0.00	0.054
Drift	64.96	0.00	-3.196
SeasonalNaive	78.89	0.00	0.590
NF-PatchTST	80.25	1.15	0.216
LSTM-tiny	81.84	2.73	0.103
NF-NBEATS	82.23	2.11	0.005
NF-MLP	82.76	2.32	0.616
Transformer-tiny	83.38	6.90	0.089
NF-NHITS	84.06	1.63	0.144
TCN-tiny	85.48	2.90	0.220
Lasso	86.54	0.00	-508.45
ElasticNet	87.14	0.00	-547.85
Ridge	87.84	0.00	-458.91
NF-KAN	88.19	3.74	0.573
NF-DilatedRNN	88.60	2.47	0.170
NF-RNN	90.04	4.29	0.670
NF-Autoformer	91.22	8.16	0.138
NF-TCN	91.54	2.49	0.107
NF-TFT	92.37	1.30	-0.076
NF-BiTCN	92.62	2.59	0.073
NF-LSTM	92.72	4.35	0.623
NF-Informer	93.13	3.33	0.395
NF-TimesNet	94.94	9.89	0.289
NF-DeepAR	95.46	3.20	-0.366
NF-GRU	95.52	4.62	0.714
NF-NLinear	106.59	4.21	0.078
NF-DLinear	106.93	5.12	0.070
Climatology	108.23	0.00	0.280

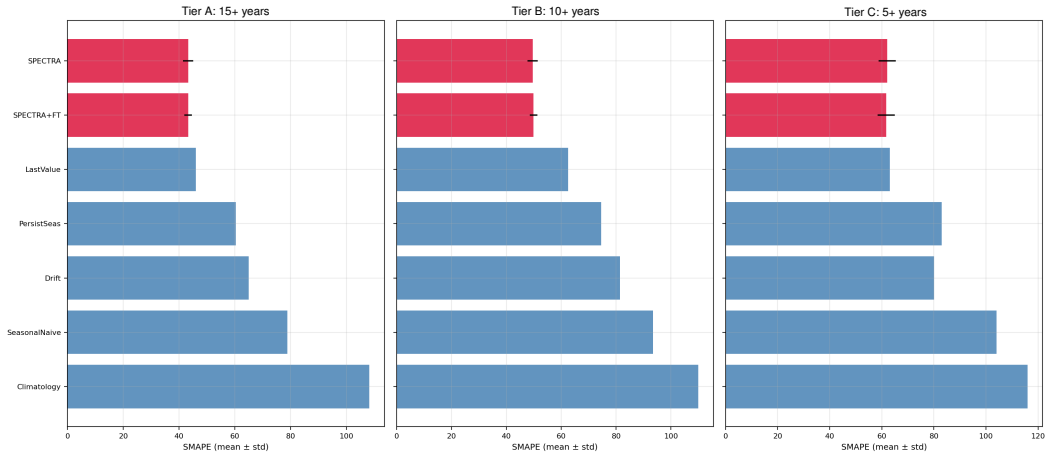


Figure 4: Selected-model SMAPE on nested eligibility cohorts: A (16 countries), A+B (38), and A+B+C (59), on fold 0. The plotted A, B, and C labels denote these nested cohorts, not mutually exclusive tiers. Error bars show standard deviation across seeds 42–46.

G STRUCTURAL-PROPERTY AND STABILITY DIAGNOSTICS

Table 12: Per-prior structural-property and stability diagnostics, canonical operator (Tier A fold 0). C1: autograd ($\partial c_t^{\text{lat}}/\partial y$ path). C2: Hawkes Green function $K(d) = \alpha e^{-\beta(d-1)}$. C3: impulse-response contraction. C4: 90% conformal coverage.

Prior	Diagnostic	Result	Quantity
C1	Architectural separation of the climate pathway: $\partial c^{\text{lat}}/\partial y$	PASS	no autograd path detected
C2	Hawkes Green function impulse response	PASS	matches closed form ($\Delta < 10^{-5}$)
C3	Bounded gated contraction (impulse decay)	PASS	empirical Lipschitz L = 0.358
C4	Persistence-skip + conformal coverage	PASS	COV90 = 0.956 at nominal 0.90
	Persistent excitation (full-rank trajectory cover)	FAIL	14-country panel under-excites a few priors

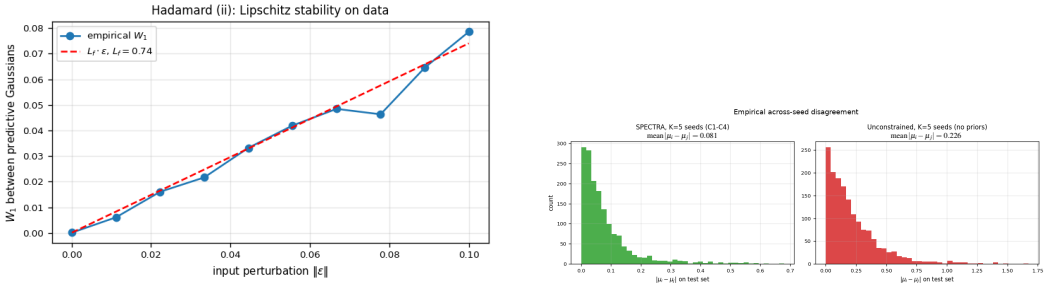


Figure 5: Empirical structural-property and stability diagnostics. Left: Lipschitz sensitivity of SPECTRA vs. the unconstrained ablation. Right: across-seed disagreement.

H OUTBREAK DETECTION

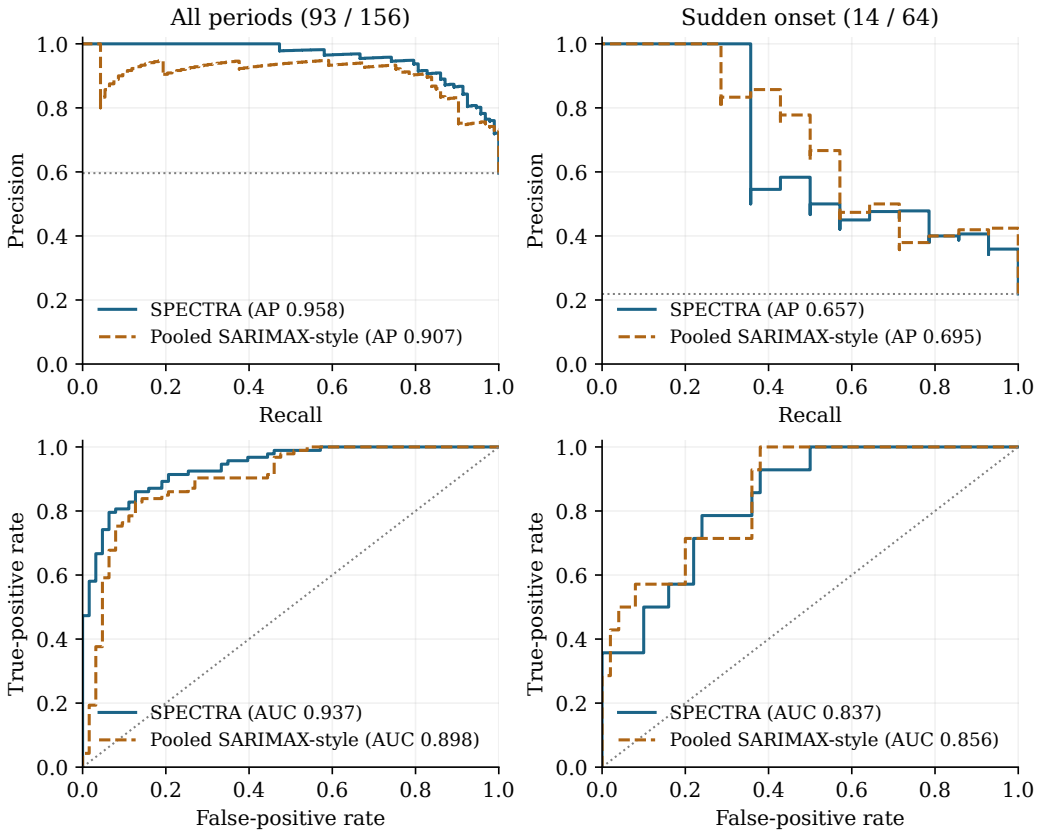


Figure 6: Precision–recall and ROC curves on 156 matched test country-months (93 all-period events). Sudden-onset evaluation uses the 64 months following a non-outbreak month (14 events). Country/calendar-month outbreak thresholds are estimated from training outcomes only.

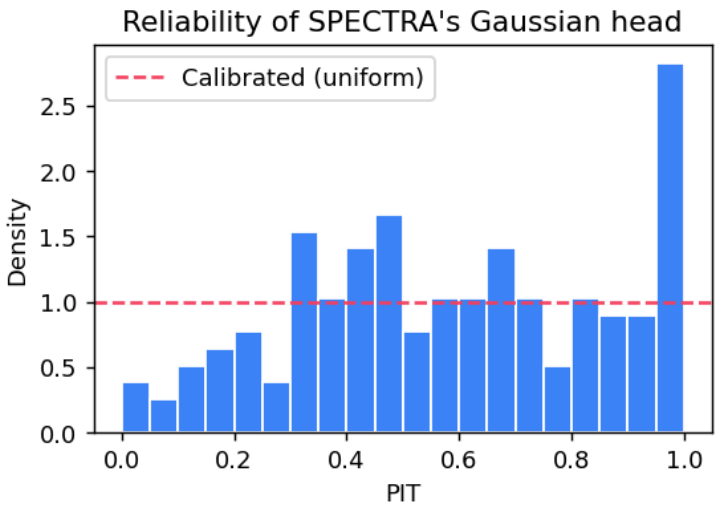


Figure 7: PIT of the operator's Gaussian head, held-out Tier A. Uniform on [0, 1] when calibrated (red dashed). Pre-conformal head over-confident at upper tail; split-conformal restores COV90 = 0.956.

H.1 DECISION-CURVE ANALYSIS

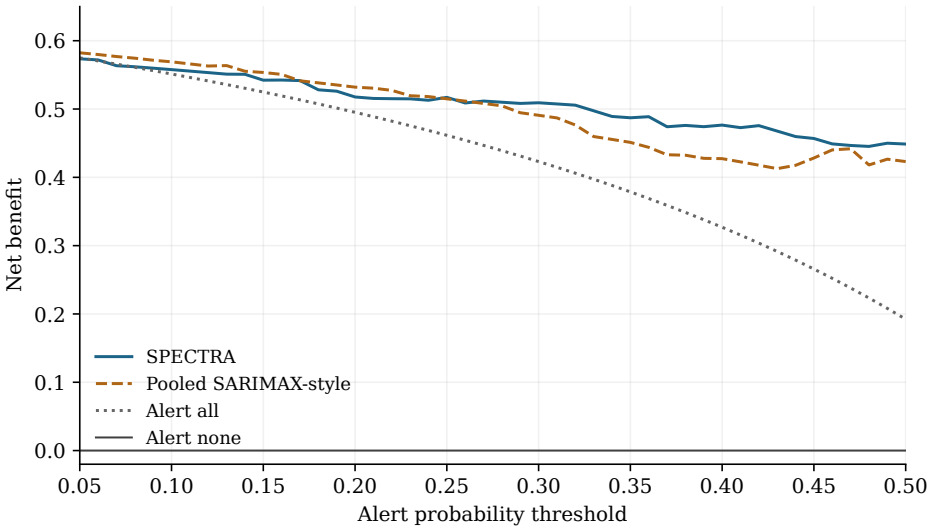


Figure 8: Decision-curve analysis on the same 156 matched test country-months. Net benefit is $\mathrm{TP}/n - (\mathrm{FP}/n)p^\star/(1-p^\star)$; outbreak labels use training-only thresholds. Both models have positive net benefit over the displayed range.

H.2 OUTBREAK-PROBABILITY DIAGNOSTICS

The Gaussian-head exceedance scores have not undergone a separate probability-calibration procedure. On the 156 matched observations, SPECTRA has Brier score 0.1103 and expected calibration error (ECE) 0.1045; the pooled SARIMAX-style comparator has Brier score 0.1268 and ECE 0.0890. ECE uses ten equal-width probability bins and sample-count weighting. The country-bootstrap 95% interval for the Brier-score improvement is $[-0.0306, 0.0661]$. Interval conformalization does not establish calibrated threshold probabilities. Country-level scores and bin counts are included in the supplementary files.

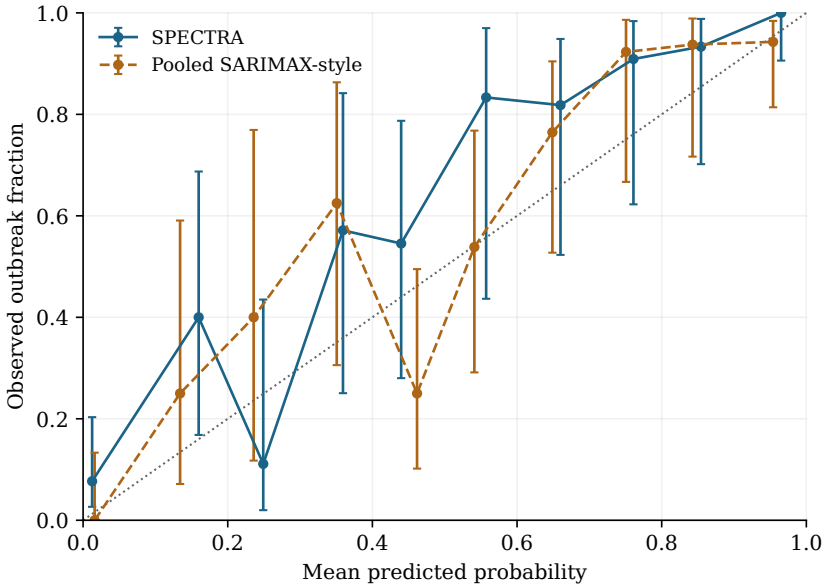


Figure 9: Reliability of outbreak-probability scores on the matched sample. Points show within-bin observed event fractions against mean predicted probabilities; bars are descriptive 95% Wilson intervals and do not account for within-country temporal dependence.

I OPERATIONAL CLIMATE-DELAY SENSITIVITY

Table 13: Climate-delay sensitivity, SPECTRA on Tier A fold 0. Trained checkpoints reused; disease history and persistence skip remain available. $R^2_{\log}$: standardised log-space (recommended diagnostic). $R^2_{\ln-\text{mean}}$: log-normal mean (unstable in raw space for high-σ countries). Retention: percent of $\Delta = 0$ skill. COV90 pre-conformal.

Δ (mo.)	0	1	2	3
SMAPE q50	42.74	47.56	52.67	58.34
SMAPE ln-mean	41.88	48.06	54.98	59.01
R^2_{q50}	+0.7156	+0.2648	+0.2245	+0.1010
$R^2_{\log}$	+0.9089	+0.8645	+0.8294	+0.7946
$R^2_{\ln-\text{mean}}$	-0.7396	+0.3220	-0.3333	-2.6490
R^2_{seed}	+0.5977	+0.2586	+0.2192	+0.0865
COV90 (pre-cf.)	0.808	0.763	0.692	0.660
CRPS log	0.2543	0.3078	0.3641	0.4114
Retention q50	100%	37%	31%	14%
Retention log	100%	95%	91%	87%

Table 14: Reading the delay table.

Quantity	Behaviour across $\Delta = 0 \to 3$
$R^2_{\log}$ (recommended)	0.91 $\to$ 0.79, -13%
R^2_{q50} (raw)	$0.7156 \to 0.1010$; retention: $100\% \to 14.1\%$
$\text{CRPS}_{\log}$	0.254 $\to$ 0.411
COV90 (pre-conformal)	0.808 $\to$ 0.660; conformal absorbs at deployment

J ZERO-SHOT GENERALISATION ACROSS THE 90-COUNTRY PANEL

Table 15: Zero-shot protocol.

Item	Setting
Training tier	Tier A (16 countries, $\geq$ 15 yr)
Eval tiers	A (in-distribution), B and C (zero-shot); tier definitions follow Table 8
ZS embedding	Mean Tier-A country embedding
ZS Hawkes (α, β)	Mean Tier-A row
Test filter	Consecutive 60-month window; 13 + 17 + 7 = 37 countries pass
Excluded	3 Tier-A + 5 Tier-B + 14 Tier-C + 31 Tier-X countries did not contribute eligible test windows
Coverage	Pre-conformal (no per-country calibration set ZS)

Table 16: Zero-shot generalisation, 90-country panel. $R^2_{\log}$ recommended (heterogeneous scales). R^2_{q50} medians dominated by catastrophic-residual countries (Venezuela, Jamaica, Dominican Republic; log-R^2 in [0.573, 0.779]).

Tier	A (in-dist.)	B (zero-shot)	C (zero-shot)
N	13	17	7
SMAPE q50 median	32.9	46.6	102.8
SMAPE IQR	[25.2, 38.7]	[32.7, 62.7]	[77.6, 115.0]
$R^2_{\log}$ median	+0.656	+0.719	-0.130
$R^2_{\log}$ IQR	[+0.45, +0.74]	[+0.57, +0.77]	[-1.95, -0.03]
COV90 (pre-cf., median)	0.833	0.833	0.400
CRPS log (median)	0.232	0.313	0.482

Table 17: Reading the generalisation table.

Finding	Evidence
Tier B $\geq$ Tier A on median $R^2_{\log}$	+0.72 vs. +0.66, tighter IQR
Climate-driven dynamics (C1–C3) generalise country-agnostically	embedding offset absorbed by conformal at deployment
Two ZS countries clear $R^2_{\log} > 0.9$	Costa Rica (+0.93), Paraguay (+0.96)
Tier C performance deteriorates	This analysis does not establish a minimum history length for deployment.

Table 18: Per-country zero-shot generalisation, fold-0 test (up to 12 held-out months; Brunei contributes five test observations). seen: country was in the operator's training set. Pre-conformal coverage. Bold: $R^2_{\log} \geq 0.80$. Italics: catastrophic raw-space residual where $R^2_{\log}$ is the honest read.

Country	Tier	seen	SMAPE_{q50}	R^2_{q50}	$R^2_{\log}$	COV90	$\mathrm{CRPS}_{\log}$
Australia	A	✓	33.5	-0.218	-0.219	0.83	0.147
Bangladesh	A	✓	64.8	+0.274	+0.726	0.67	0.400
Cambodia	A	✓	32.8	+0.323	+0.361	0.67	0.232
Colombia	A	✓	32.2	+0.479	+0.729	0.83	0.244
Indonesia	A	✓	24.8	+0.744	+0.522	0.75	0.202
Lao PDR	A	✓	36.3	+0.744	+0.656	0.83	0.258
Mexico	A	✓	25.2	+0.669	**+0.869**	0.92	0.202
Nicaragua	A	✓	20.6	+0.923	**+0.895**	1.00	0.169
Peru	A	✓	55.6	+0.104	+0.736	1.00	0.428
Saudi Arabia	A	✓	32.9	+0.665	+0.765	0.83	0.198
Sri Lanka	A	✓	38.7	+0.509	+0.445	0.75	0.285
Thailand	A	✓	24.4	+0.529	+0.451	0.67	0.177
Venezuela	A	✓	133.9	*-23.81*	+0.573	0.33	0.363
Aruba	B	n/a	118.4	+0.432	+0.269	0.67	0.906
Bolivia	B	n/a	37.5	+0.648	+0.707	0.75	0.323
Brazil	B	n/a	38.8	+0.234	**+0.834**	0.92	0.313
Costa Rica	B	n/a	21.8	+0.996	**+0.930**	0.92	0.167
Dominican Rep.	B	n/a	48.8	*-3.761*	+0.779	0.83	0.291
Ecuador	B	n/a	27.8	+0.359	+0.719	0.92	0.223
El Salvador	B	n/a	30.6	+0.654	+0.764	0.83	0.216
Guatemala	B	n/a	50.9	+0.317	+0.565	0.67	0.412
Honduras	B	n/a	53.5	+0.217	+0.692	0.67	0.432
Jamaica	B	n/a	46.6	*-35.06*	+0.601	0.83	0.321
Nepal	B	n/a	85.6	+0.104	+0.365	0.33	0.643
Panama	B	n/a	32.7	+0.477	+0.771	0.83	0.277
Paraguay	B	n/a	26.4	+0.991	**+0.957**	1.00	0.246
Philippines	B	n/a	62.7	-0.597	-0.623	0.42	0.567
Puerto Rico	B	n/a	33.7	+0.714	+0.767	1.00	0.260
Saint Lucia	B	n/a	70.5	+0.737	+0.763	0.67	0.566
USA	B	n/a	68.7	-0.485	-0.733	0.17	0.165
Brunei	C	n/a	62.8	+0.567	-0.088	0.40	0.482
Chile	C	n/a	102.8	-0.111	-0.203	0.58	0.197
Hong Kong	C	n/a	109.5	-2.435	-5.112	0.25	0.036
Oman	C	n/a	77.0	-0.088	-0.130	0.58	0.521
Suriname	C	n/a	145.7	-105.2	+0.394	0.75	0.479
Virgin Islands (US)	C	n/a	120.5	-0.297	+0.036	0.33	1.108
Yemen	C	n/a	78.3	-10.81	-3.697	0.08	0.558

K PER-PRIOR CUMULATIVE ABLATION

Table 19: Per-prior cumulative ablation, Tier A fold 0 (3 seeds, 120 epochs, cosine LR + linear warmup, no WGAN-GP critic). R^2_{ens}: median-ensemble single-pass; R^2_{seed}: per-seed median mean. Across-seed disagreement: across-seed std of $\hat{\mu}$ in log-space (lower = less across-seed disagreement). Stability: Lipschitz of $\hat{\mu}$ to 1% climate perturbation. COV90 pre-conformal.

Variant	$R^2_{\text{ens}} \uparrow$	$R^2_{\text{seed}} \uparrow$	SMAPE ↓	COV90 (pre-cf.)	ΔR^2 vs prev.	Across-seed disagreement ↓	Stability
none	-0.009	-0.010	94.7	0.76	n/a	0.111	0.0002
+C1	+0.021	+0.020	84.9	0.85	+0.030	0.180	0.0007
+C1,C2	+0.084	+0.079	82.8	0.87	+0.063	0.187	0.0006
+C1,C2,C3	+0.459	+0.452	49.5	0.53	+0.375	0.119	0.0458
All C1–C4	+0.612	+0.605	49.4	0.42	+0.153	0.098	0.0436

Table 20: Notes on the per-prior ablation.

Item	Reading
Stability column scale	single-pass 1%-climate sensitivity, short ablation; not the full impulse-response Lipschitz
Comparison with Table 1	Lipschitz 0.402 vs. 0.740 (ratio 1.84×) measured at full 5-seed × 200-epoch budget
Qualitative ordering	The short-budget climate-sensitivity diagnostic and the full-budget diagnostic have different orderings and should not be interpreted interchangeably.
1.84× stability ratio	recovered up to seed noise at full training budget
COV90 = 0.42 with all four properties	pre-conformal Gaussian-head over-confidence (PIT → 1)
Deployed COV90 = 0.956	per-country split-conformal step (Vovk et al., 2005; Lei et al., 2018; Angelopoulos & Bates, 2023), not the pre-conformal head

To complement the cumulative ablation, we removed each property separately while retaining the other three. Every condition used the same fold, seeds 42–46, training budget, validation rule, and early-stopping procedure as the matched operator experiment.

Condition	Natural R^2	Log R^2	SMAPE	Log-CRPS
All C1–C4	0.7341	0.8933	47.61	0.2875
Without C1	0.6409	0.8972	45.67	0.2763
Without C2	0.7287	0.9051	43.20	0.2597
Without C3	−0.0281	0.7627	55.76	1.2269
Without C4	0.5148	0.8750	49.27	0.2963

As shown above, C3 has the largest contribution under this intervention. When C4 is removed the performance degrades. C1 and C2 have interacting and metric-dependent effects: their removal worsens natural-scale R^2 but improves some log-scale or percentage-error metrics. Based on this evidence, we do not claim that every property independently improves every metric.

L EXPLORATORY DIAGNOSTICS ACROSS THE 90-COUNTRY PANEL

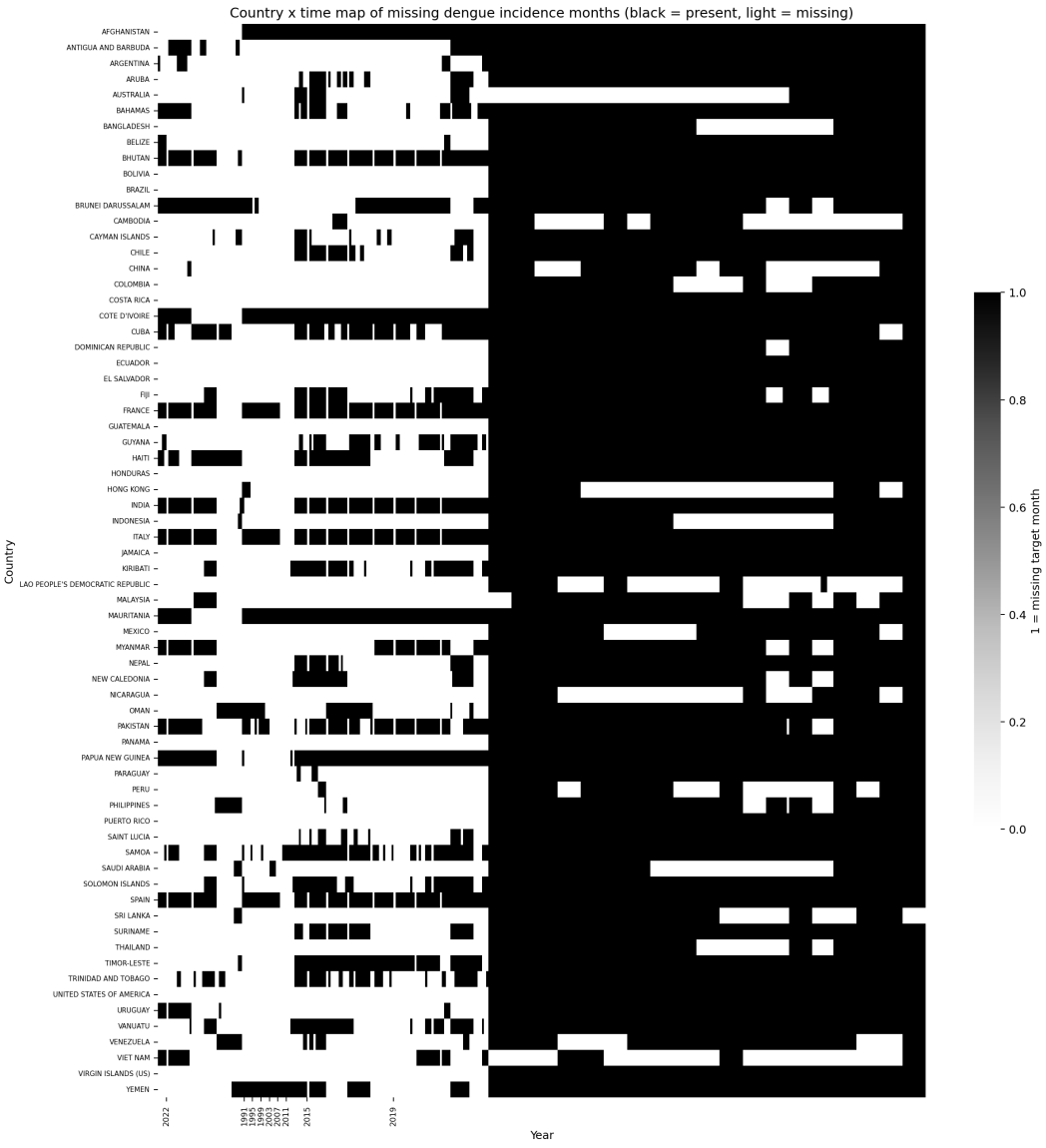


Figure 10: Country × time map of missing dengue incidence months (black = present, light = missing).

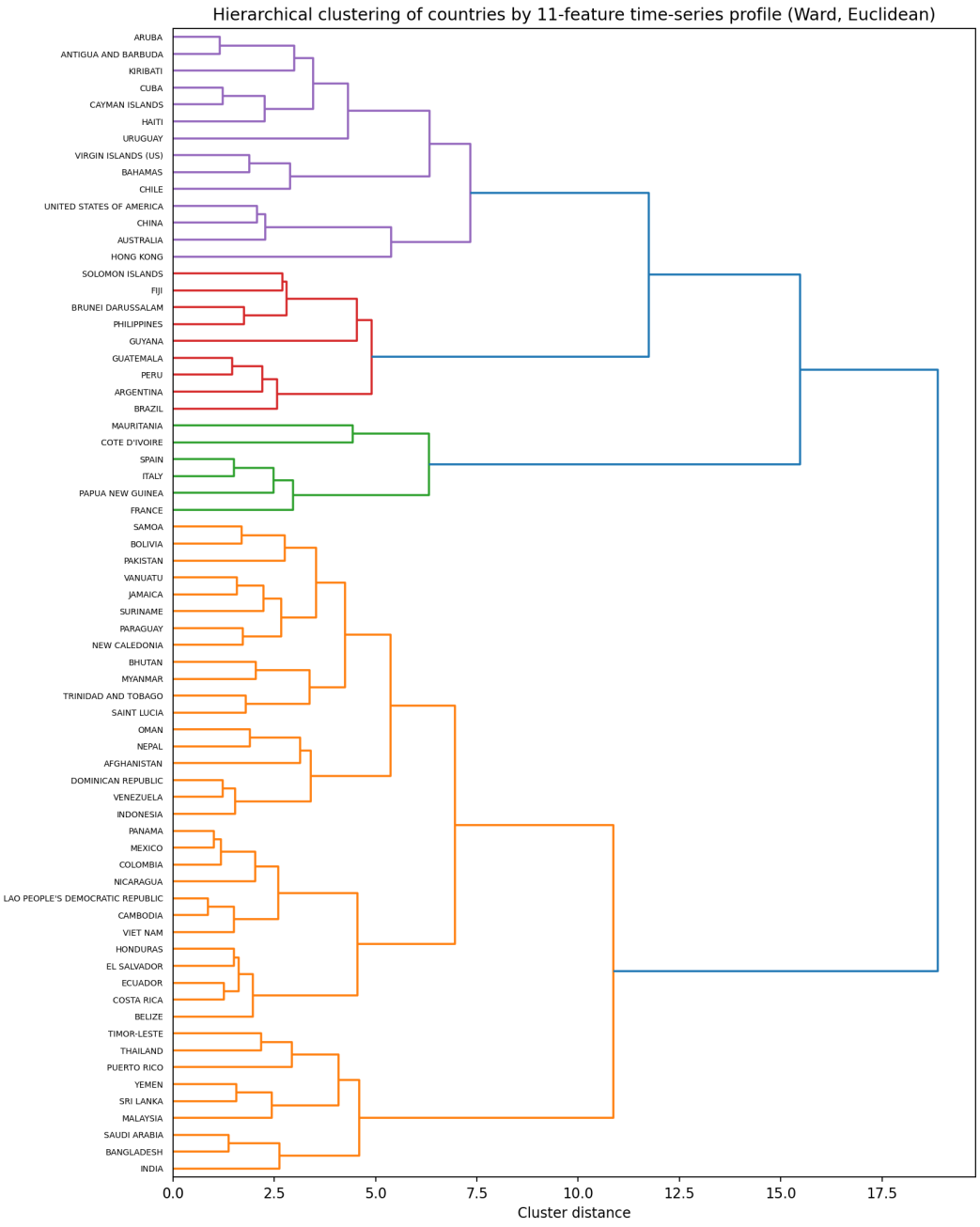


Figure 11: Country dendrogram, 11-feature time-series profile (Ward, Euclidean): log mean/std, skew, kurtosis, fraction nonzero, Hurst R/S, Higuchi FD, spectral entropy, STL trend/seasonal strength, lag-1 ACF.

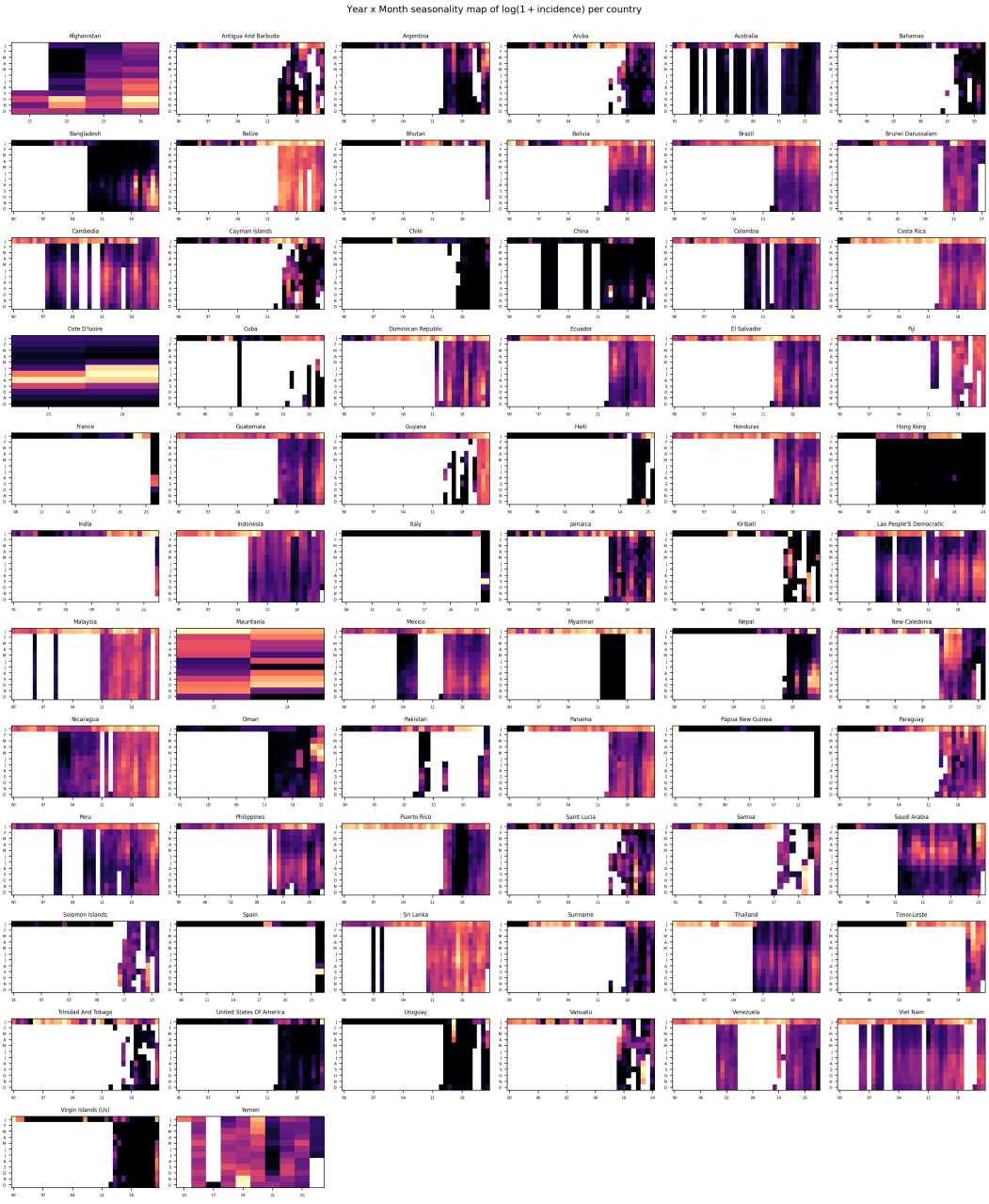


Figure 12: Year × month seasonality calendar, log(1 + incidence) per country.



Figure 13: STL of log(1 + incidence) per country (black: log y; blue: trend; green: seasonal; red: residual; 12-month period, robust loess).

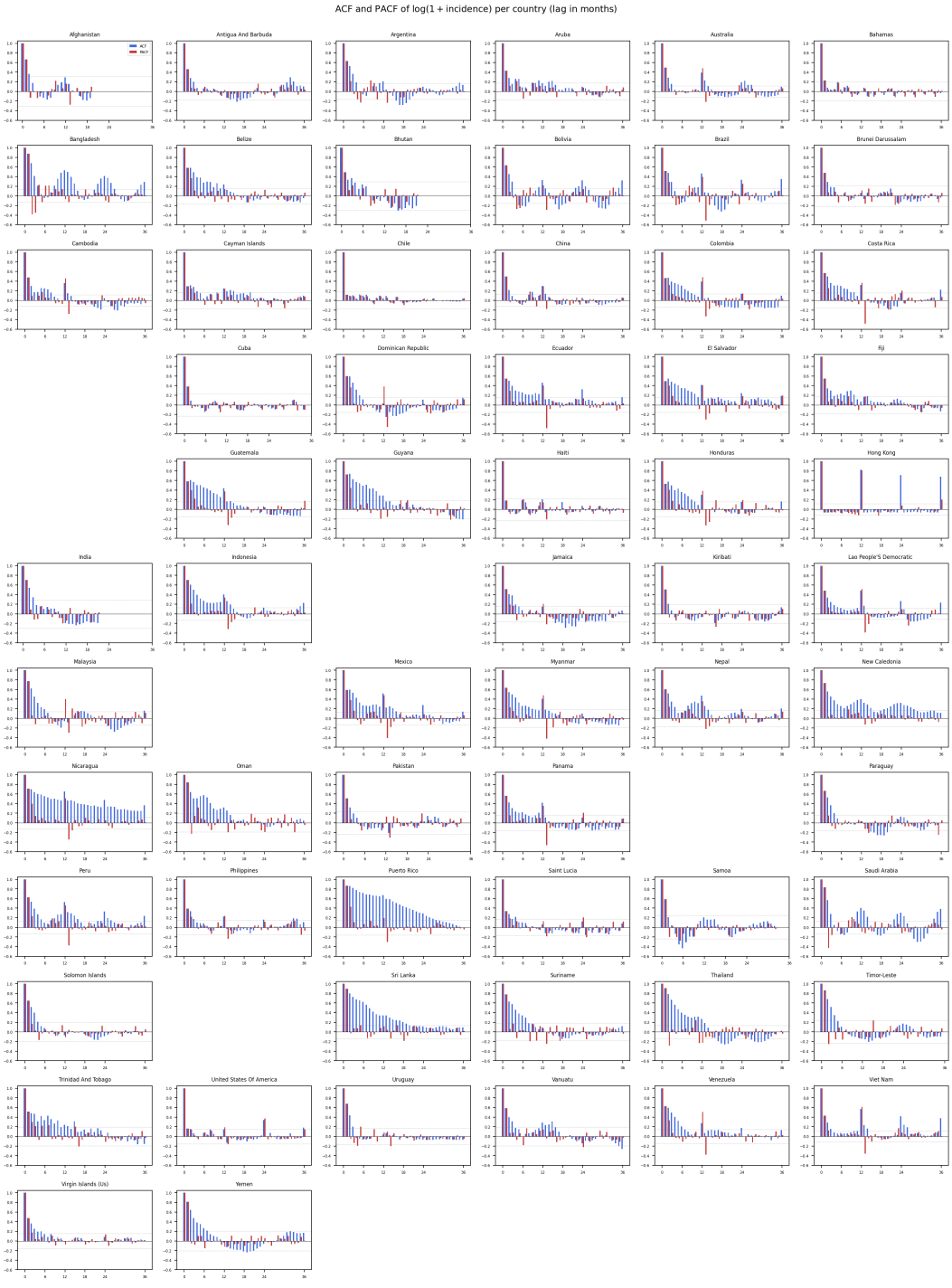


Figure 14: ACF (blue) and PACF (red) of log(1 + incidence) per country, lag in months.

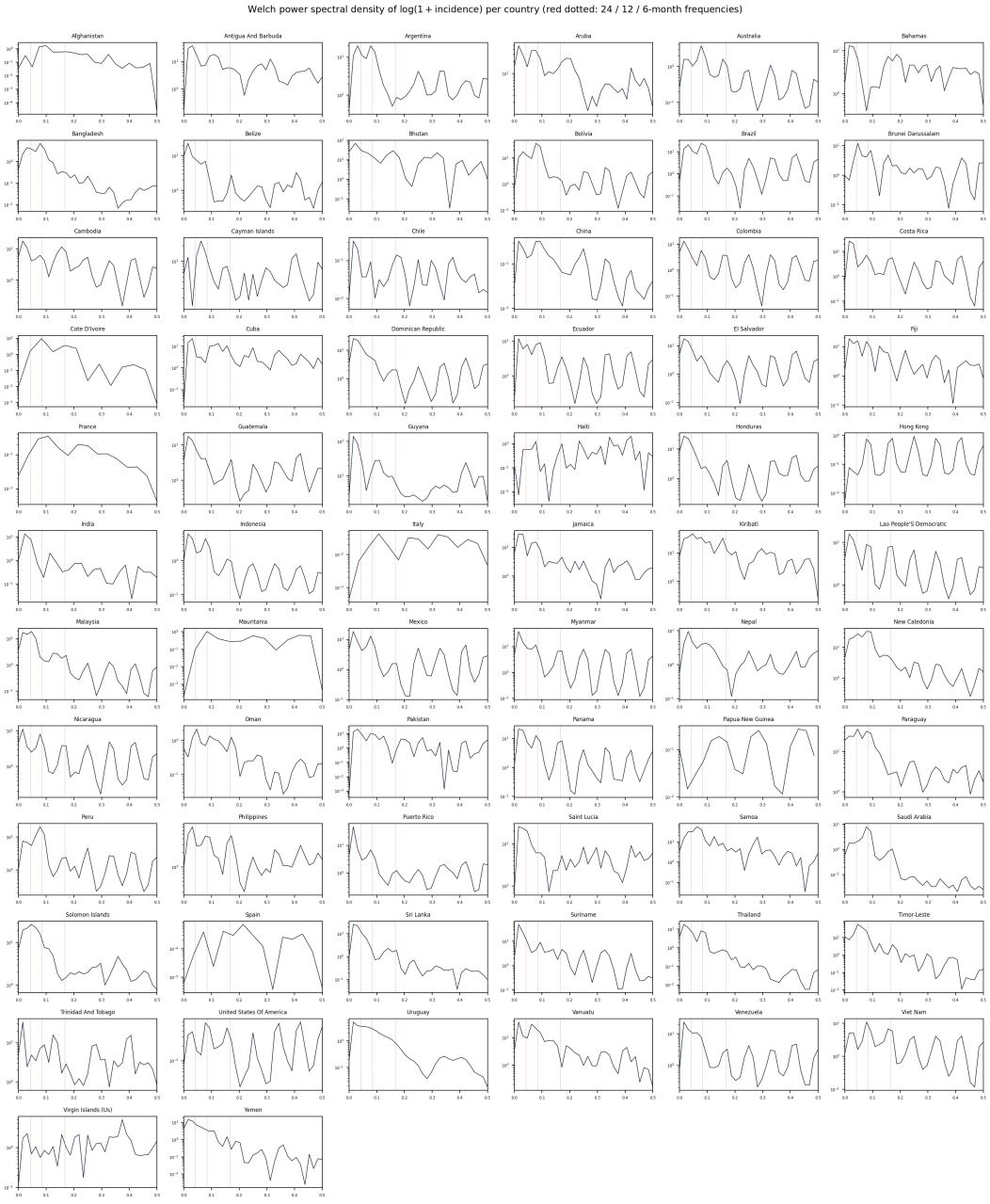


Figure 15: Welch PSD of log(1 + incidence) per country; red lines: 24, 12, 6-month frequencies.

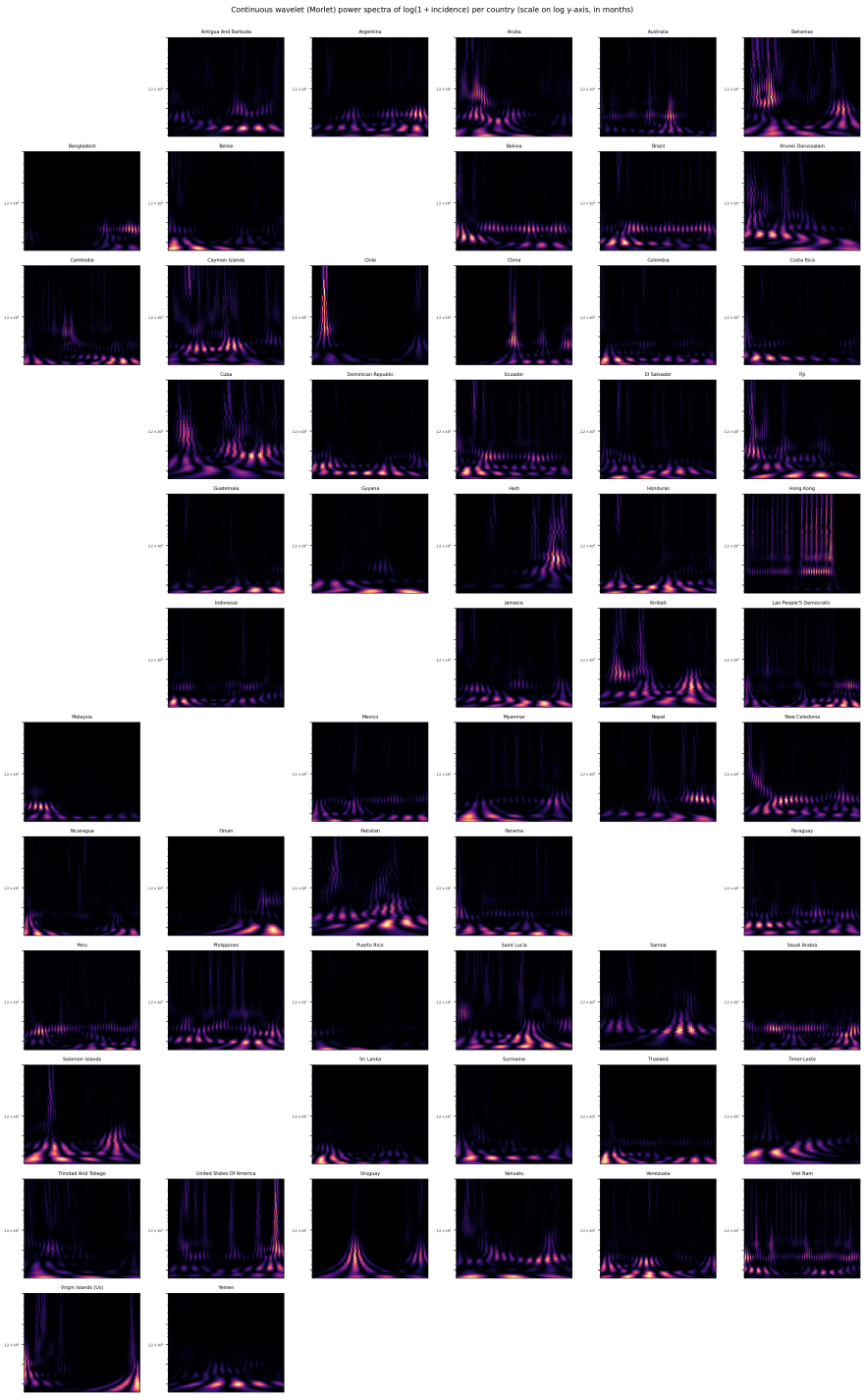


Figure 16: Morlet wavelet power spectra of log(1 + incidence) per country; vertical: scale in months (log).

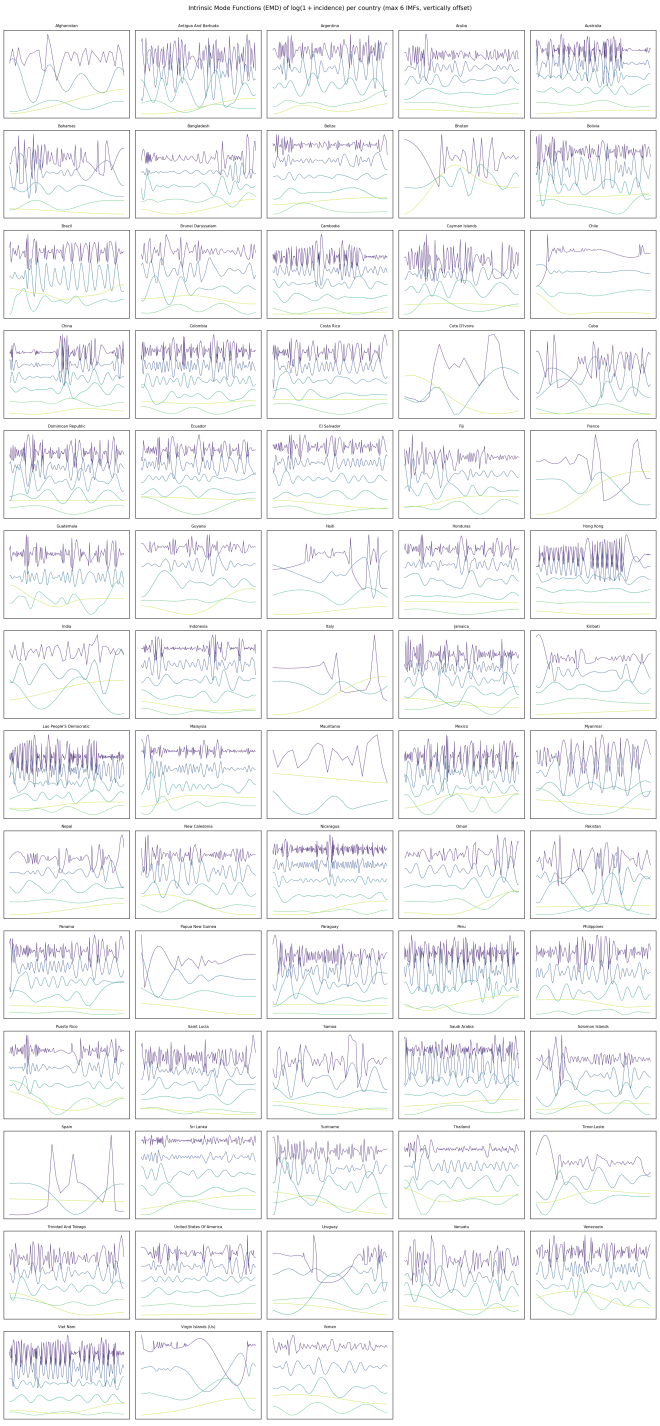


Figure 17: EMD intrinsic mode functions of log(1+incidence) per country (up to six IMFs, vertically offset).

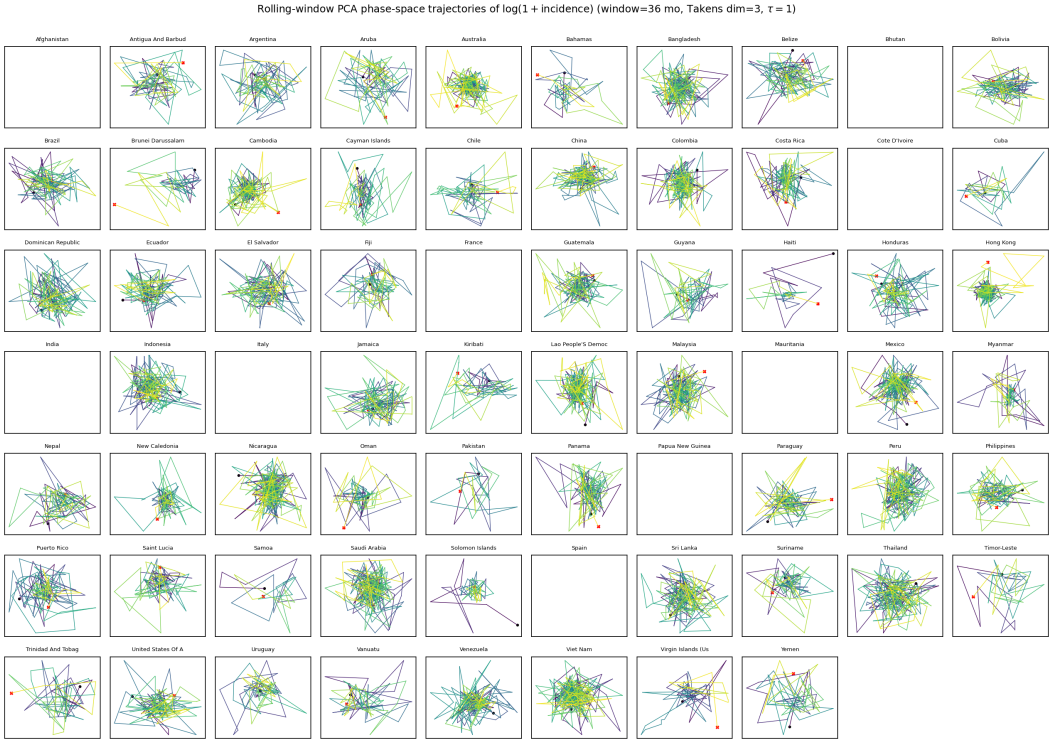


Figure 18: Rolling-window PCA trajectories per country (36-month window, Takens dim 3, $\tau = 1$); colour fades along trajectory, black dot: start, red cross: end.

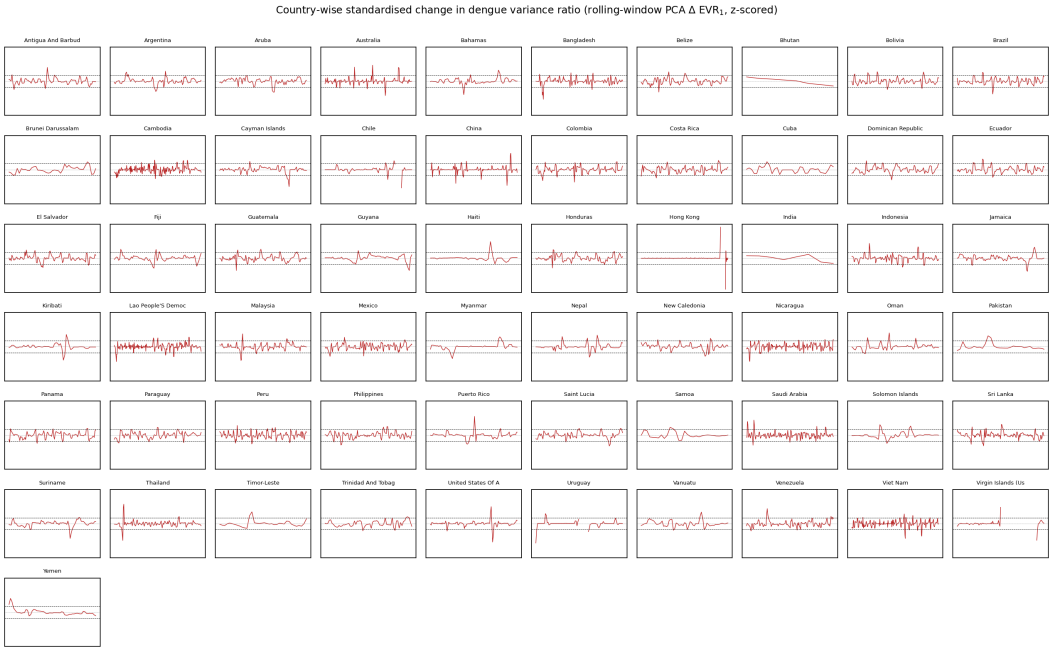


Figure 19: Per-country z-scored ΔEVR1 from rolling-window PCA; dashed lines: $\pm 2\sigma$.

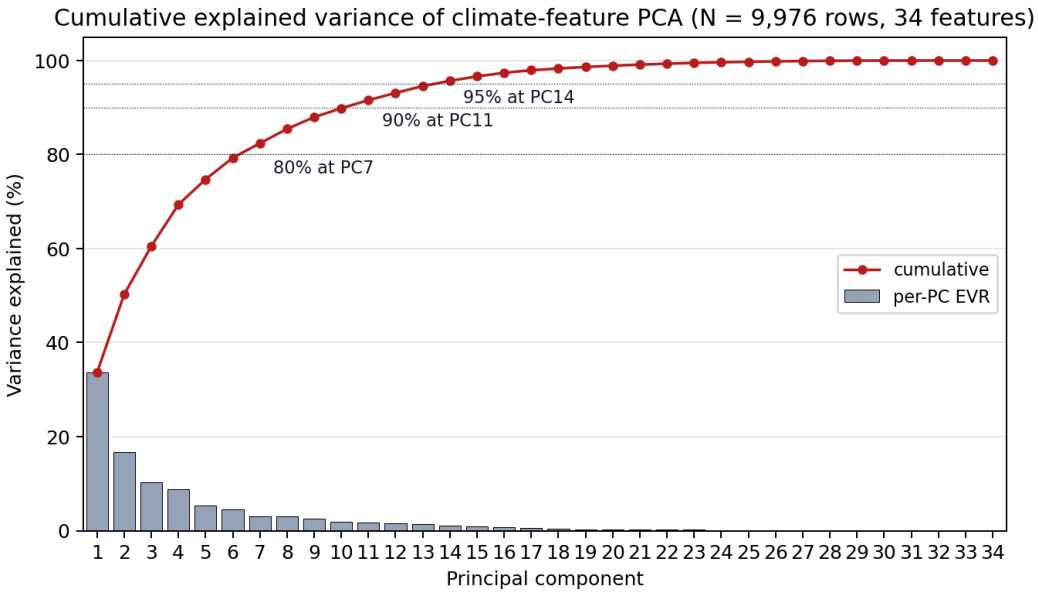


Figure 20: Per-PC and cumulative variance, climate-feature PCA; dotted lines: 80, 90, 95% cumulative thresholds.

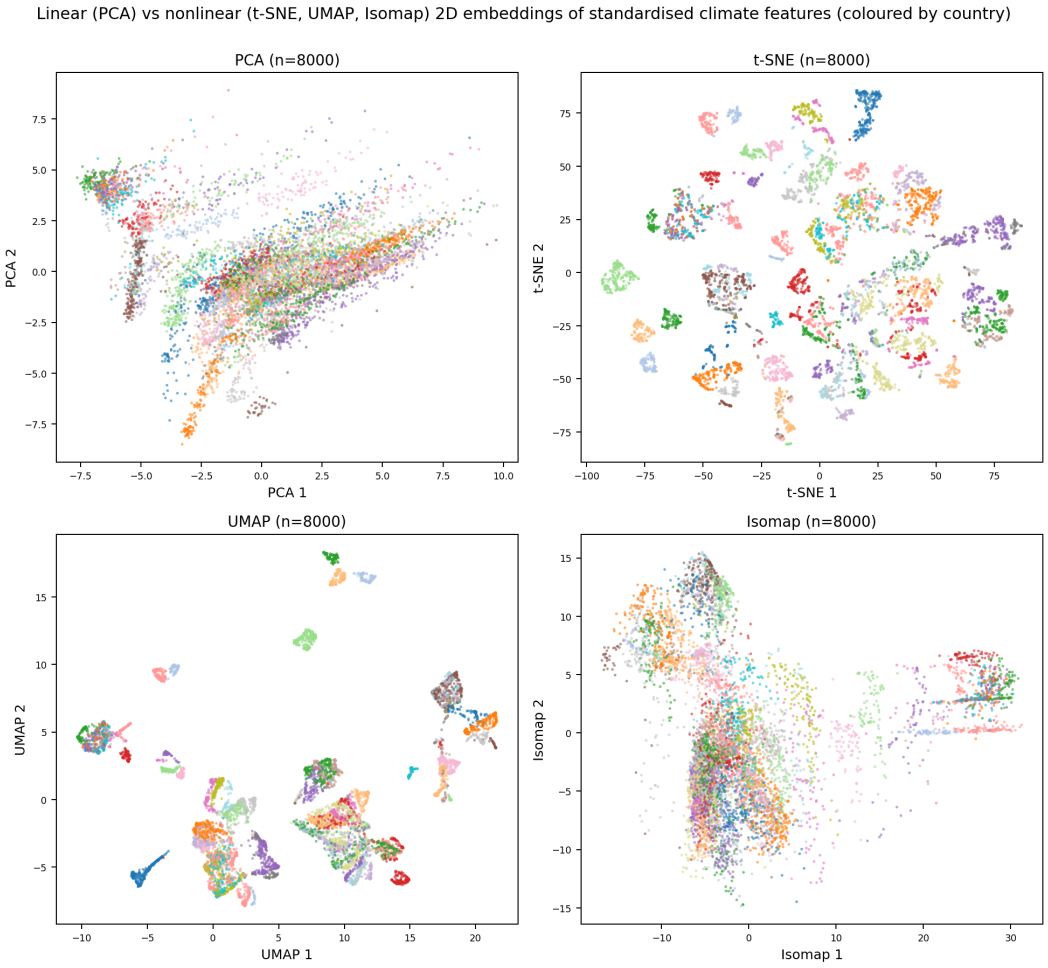


Figure 21: 2-D climate-feature embeddings: PCA, t-SNE, UMAP, Isomap; each point a country-month, colour by country.

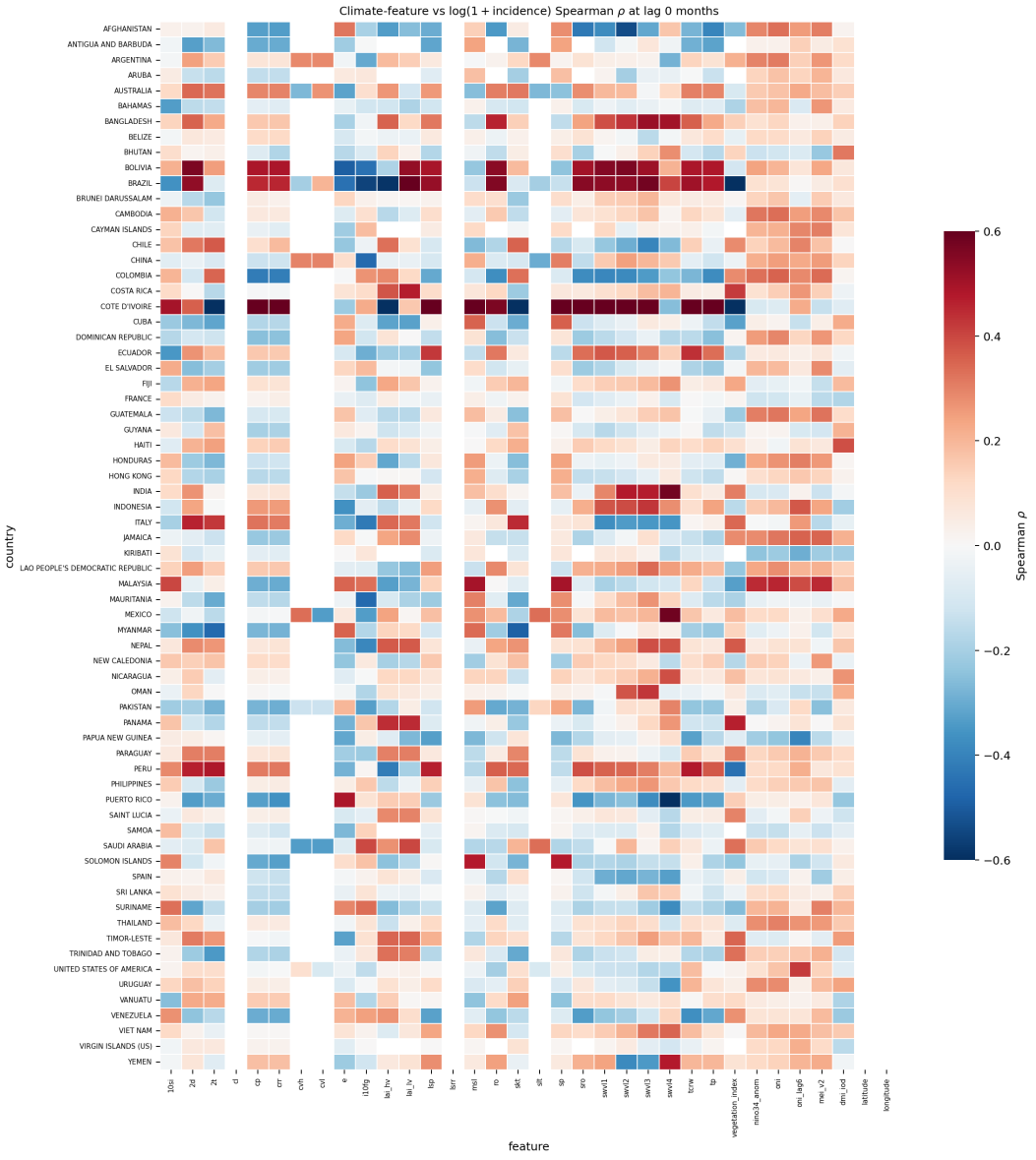


Figure 22: Spearman ρ between climate feature and log(1 + incidence) at lag 0 months, country $\times$ feature.

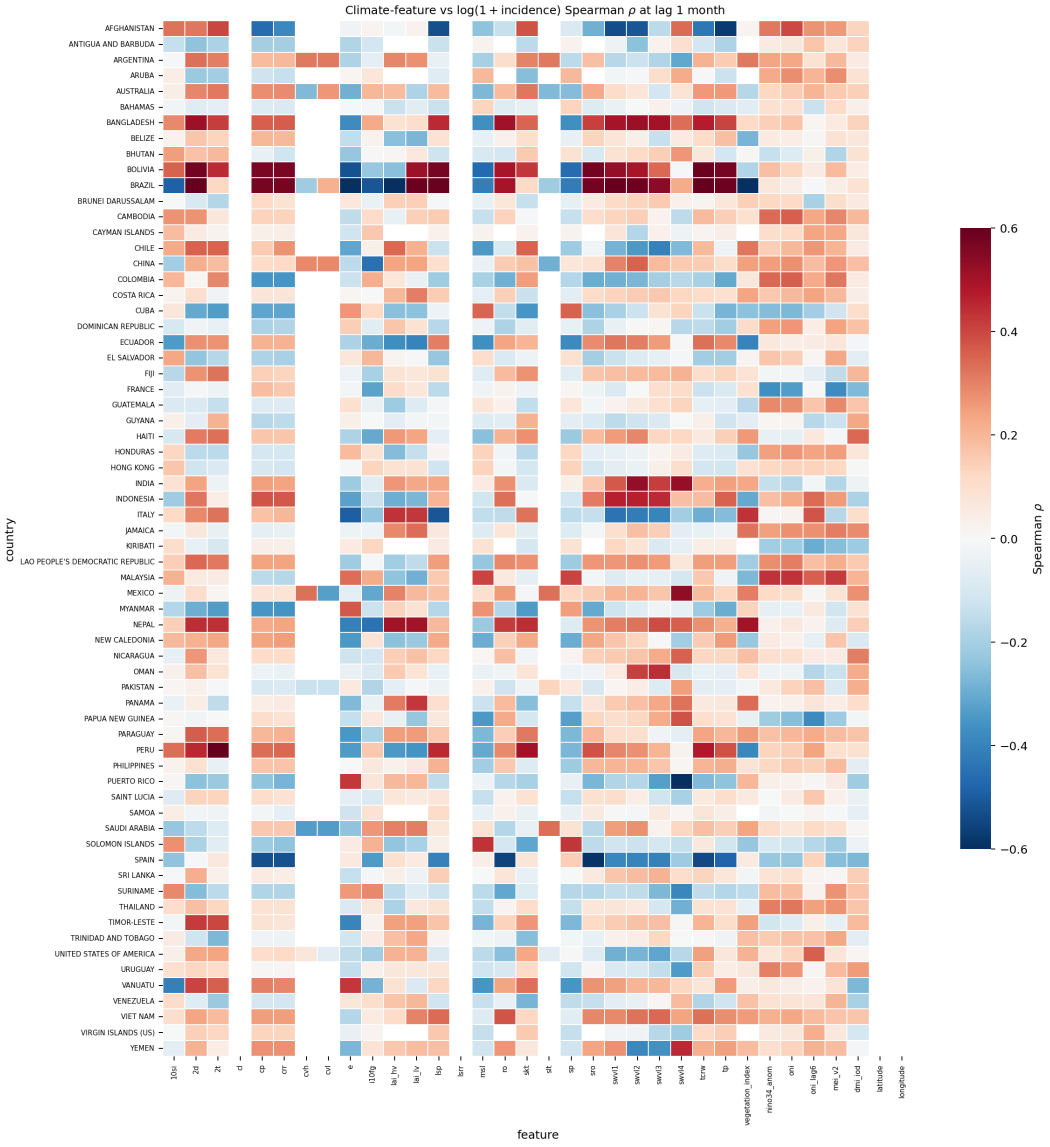


Figure 23: Spearman ρ between climate feature and log(1 + incidence) at lag 1 month.

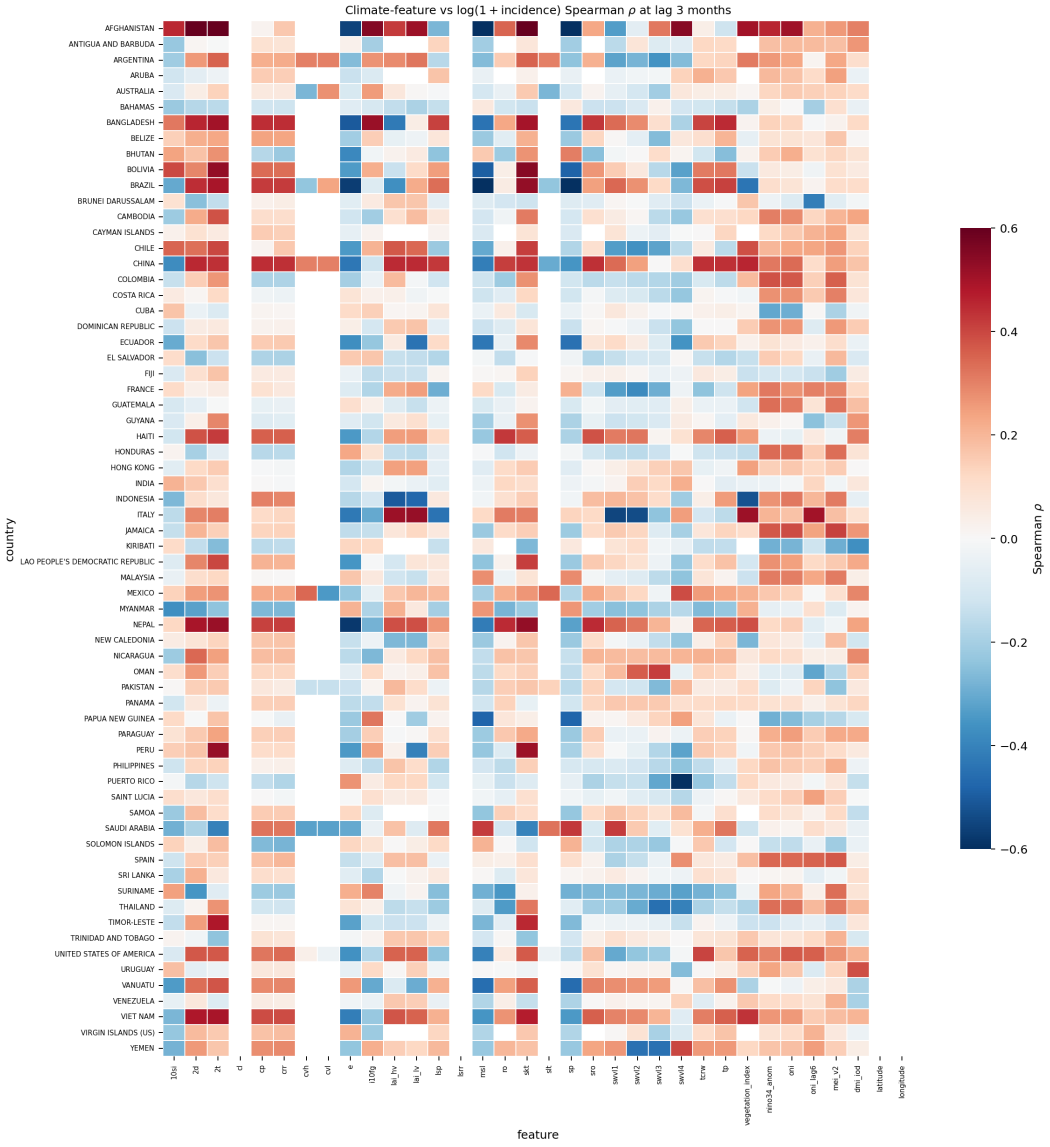


Figure 24: Spearman ρ between climate feature and log(1 + incidence) at lag 3 months.

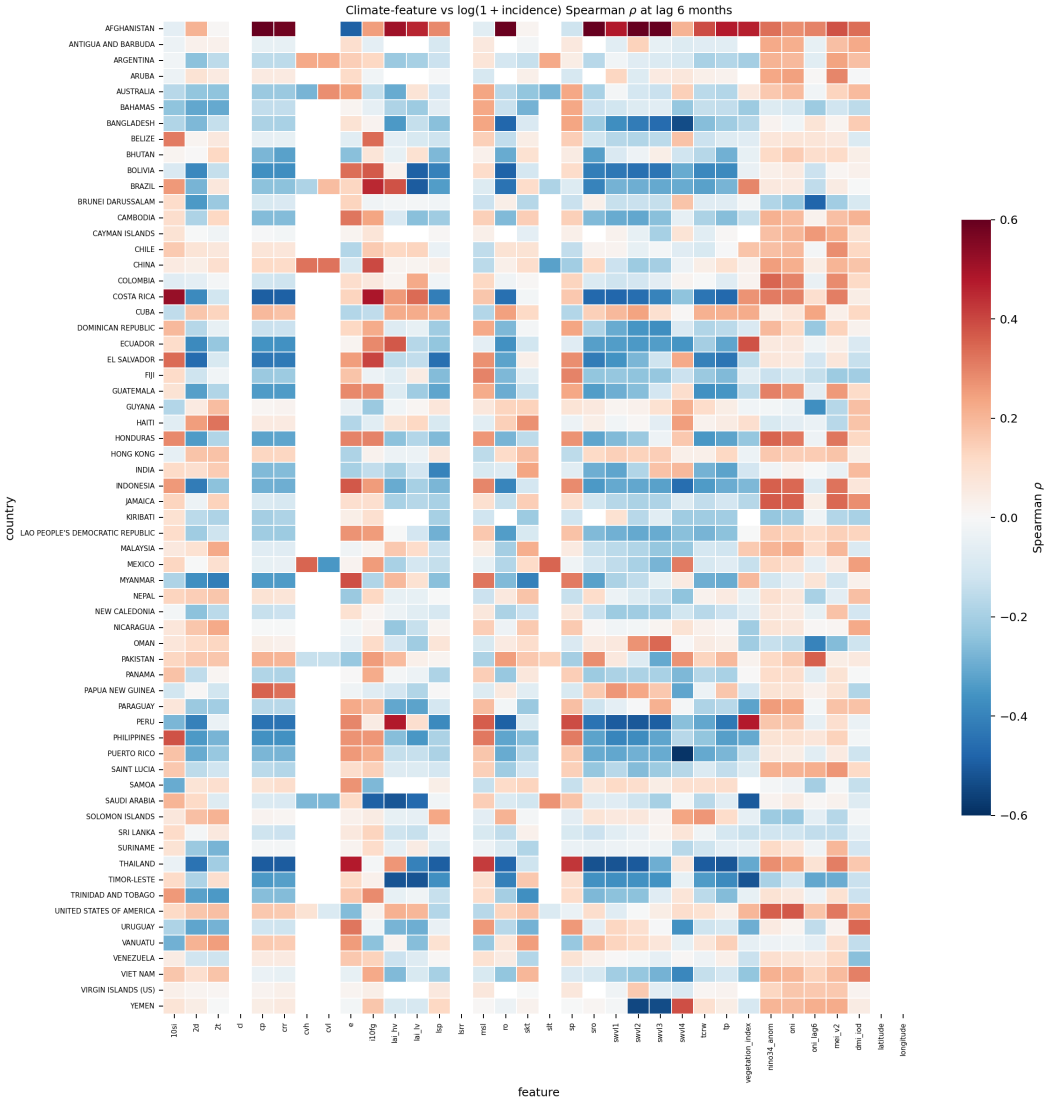


Figure 25: Spearman ρ between climate feature and log(1 + incidence) at lag 6 months.

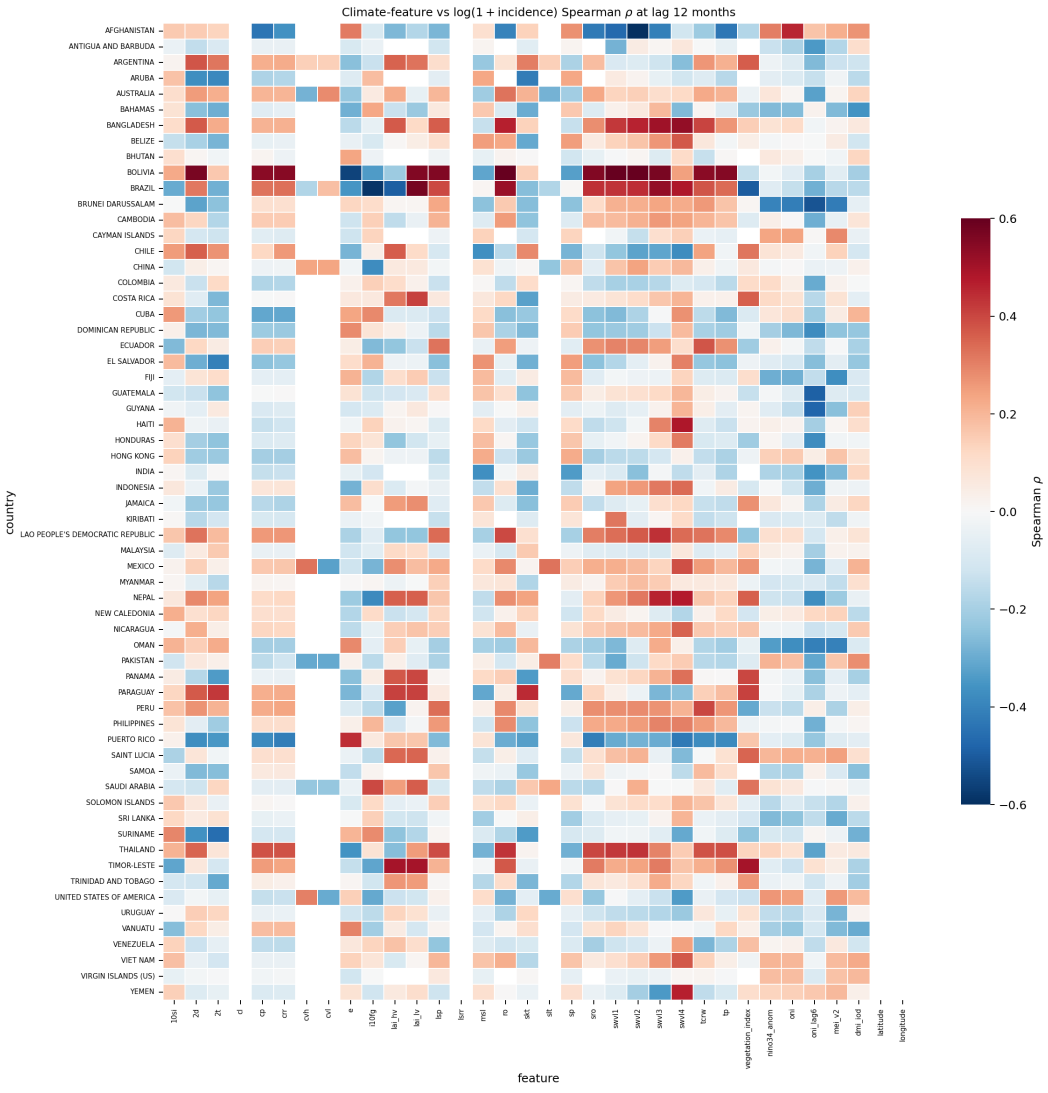


Figure 26: Spearman ρ between climate feature and log(1 + incidence) at lag 12 months.

M MATCHED OPERATOR–FRAMEWORK COMPARISON

Table 21: Matched operator–framework comparison.

Metric	Operator	Framework	Operator minus framework, 95% CI
Natural R^2	0.7341	0.5344	$+0.1997$ $[0.1534, 0.4033]$
Log R^2	0.8933	0.8888	$+0.0045$ $[-0.0474, 0.0531]$
SMAPE	47.61	43.32	$+4.29$ $[-1.16, 11.37]$
Log-CRPS	0.2875	0.2633	$+0.0242$ $[-0.0239, 0.0920]$
COV90	0.918	0.937	-0.019; 95% CI not reported

N EPIDEMIOLOGICAL BASELINE COMPARISON

N.1 POOLED SPECIFICATIONS

Table 22: Pooled baseline comparison on 156 unseen country-months from 13 countries.

Metric	SPECTRA	Pooled SARIMAX-style	NB-GAM	NB-DLNM
SMAPE	42.742	49.458	73.421	72.618
Log-CRPS	0.2543	0.3354	0.5190	0.5159
Scaled IS90 (log)	0.1163	0.1480	0.2324	0.2315

N.2 INTERVAL-SCORE DEFINITION

For $z = \log(1 + y)$ and the central 90% interval $[l, u]$ on the same scale, the reported scaled interval score is $\mathrm{sIS90}_{\log} = 0.05(u - l) + (l - z)\mathbf{1}\{z < l\} + (z - u)\mathbf{1}\{z > u\} = 0.05\,\mathrm{IS}_{0.1}(l, u; z)$. This is a scaled single-interval score, not the weighted interval score that combines the median and multiple interval levels. Lower values are better.

N.3 MATCHED DISTRIBUTIONAL COMPARISON

The matched framework/operator experiment uses 24-month histories, the same training and validation observations, seeds 42–46, and a maximum 350 epochs with early stopping. Its ensemble comparison contains 159 test observations. Distributional overlap is descriptive and does not imply identical predictive distributions. For one-dimensional Gaussian heads, the Wasserstein distance is $W_2 = \sqrt{(\mu_o - \mu_f)^2 + (\sigma_o - \sigma_f)^2}$ on the common log-incidence scale. Symmetric KL is the average of the two directed Gaussian KL divergences; interval overlap is intersection over union. The operational-tail subset contains 104 observations for which the outcome exceeds its training-only outbreak threshold or either model assigns outbreak probability at least 0.5. Alert agreement uses a probability threshold of 0.5.

Table 23: Distributional overlap in the matched experiment.

Diagnostic	All observations	Operational tail
Observations	159	104
Median W_2	0.1717	0.1874
Median symmetric KL	0.2484	0.3360
Mean interval intersection/union	0.5986	0.5848
Mean absolute tail-probability difference	0.1109	0.1415
Fraction with the same alert decision	0.8994	0.8462

N.4 SINGLE-MODEL AND ENSEMBLE VARIATION

The saved fold-0 seed results give SMAPE 43.2613 ± 1.8015 and natural-scale $R^2 = 0.3978 \pm 0.3921$ across seeds 42–46. The saved ensemble summary gives SMAPE 43.7339 and natural-scale $R^2 = 0.7133$. The ensemble improves R^2 relative to four of the five individual seeds, but not relative to the best seed ($R^2 = 0.7220$). The ensemble's SMAPE is lower than that of one individual seed. These results do not establish an advantage over equally ensembled baselines.

N.5 COUNTRY AND FORECAST-LEVEL DIAGNOSTICS

Among the 37 countries with eligible checkpoint-evaluation windows, 27 of 30 Tier A/B countries and two of seven Tier C countries have positive log-scale R^2. These are retrospective descriptive diagnostics, not an independently validated acceptance rule for deployment. On the 156 matched observations, the upper 20% of Gaussian-head standard deviations have mean absolute log error 0.6317 and log-CRPS 0.4447, compared with 0.2789 and 0.2070 in the remaining observations.

The standard-deviation cut-off (0.4621) is a test-sample quantile and is not a prospectively validated abstention threshold.

N.6 CPU INFERENCE BENCHMARK

With synthetic 60-month inputs, one CPU thread, and batch size one, measured median latency is 7.61 ms for one operator and 38.07 ms for five sequential operators. For a batch of 90 countries, the corresponding medians are 25.20 ms and 119.83 ms per batch. Five FP32 parameter sets occupy 1.71 MiB. This benchmark uses an ARM64 macOS host and PyTorch 2.14.0; it is separate from the reported single-operator NVIDIA T4 timing. The measurements do not constitute a field deployment evaluation.

N.7 FORECASTABILITY SENSITIVITY

The original $k = 25$ computation reproduces log-scale $R^2 = 0.730045$ on a random 15% holdout of 3,614 history pairs; country-bootstrap limits are $[0.5856, 0.8085]$. A grid of seven neighbour counts $(5, 10, 15, 25, 35, 50, 75)$, three distances (Euclidean, Manhattan, cosine), and two scaling rules (standard, robust) gives test R^2 from 0.5978 to 0.7752. Training-only selection chooses $k = 5$, Euclidean distance, and standard scaling, with held-out $R^2 = 0.7752$. On the same random-pair task, extra trees, histogram gradient boosting, and random forest achieve 0.8328, 0.8320, and 0.8214, respectively. These are finite-sample predictive references, not upper bounds. In a separate temporal comparison aligned to the 156 exported operator observations, training-selected k-NN models using 6, 12, 18, and 24 months of history achieve log-scale R^2 of 0.6961, 0.6816, 0.6481, and 0.6181, respectively. The operator Gaussian mean has log-scale $R^2 = 0.9089$ on these observations. History length changes the predictor information set; equality of target scale and test rows does not imply identical representations or training procedures. The supplementary notebook reports the complete grids, model selection, country-bootstrap intervals, and learning curves.

AI USE STATEMENT

AI tools (OpenAI Codex) assisted with manuscript editing and formatting, reference verification, code reconstruction and debugging, and implementation of supplementary analyses and visualizations. The authors are responsible for the manuscript, analyses, citations, and conclusions.