MEDIDA for Epidemics: Extending a Structural Model-Error Framework from Chaotic Fluids to Real Disease Data

Alexander Spetzler,[1] Max Ratcliff,[1] Matteo Tasso,[1] Alejandro Munoz,[1] and Anirudh Raja[1]
Department of Applied Mathematics, University of California, Santa Cruz, Santa Cruz, California 95064, USA

(Dated: May 2026)

Standard epidemic models fail when their structural assumptions break down: constant transmission rates cannot capture lockdowns, new variants, or behavioral adaptation. MEDIDA, originally developed for chaotic fluid systems, uses sparse Bayesian regression and data assimilation to discover and correct these structural gaps, producing closed-form, interpretable correction equations. We test whether MEDIDA generalizes from physics to epidemiology. Validated on five synthetic epidemic model variants, MEDIDA recovers missing structural dynamics with high fidelity, reducing coefficient errors by over three orders of magnitude in noise-free benchmarks and maintaining substantial gains under 1% observation noise. Applied to real COVID-19 data across approximately 60 countries, it significantly reduces one-step forecast residuals relative to naive SIR, recovers time-varying effective transmission rates that align with documented policy events without direct supervision, and transfers learned corrections across national borders. A globally-trained model improves forecasting accuracy for the large majority of countries tested, and Year 1 corrections remain effective on out-of-sample Year 2 data. These results show that structural model-error correction is a viable and generalizable approach to epidemic forecasting.

We extend MEDIDA—a structural model-error correction framework originally validated on the chaotic Kuramoto–Sivashinsky PDE—to epidemic compartmental models and real COVID-19 surveillance data. Sparse Bayesian regression recovers missing dynamical structure across five synthetic epidemic variants and transfers corrections learned from pooled global data to held-out countries.

I. INTRODUCTION

A. Motivation and Background

Compartmental epidemic models offer a powerful balance of tractability and mechanism: the SIR framework (**?**) and its descendants partition a population into a small number of states governed by interpretable transition rates, yielding analytically understandable descriptions of disease dynamics that have informed public health policy for decades. However this interoperability leads to rigidity rigidity. Constant transmission rates cannot accommodate behavioral responses, seasonal forcing, or variant emergence; homogeneous mixing obscures spatial clustering and demographic heterogeneity; permanent immunity ignores waning protection. When the real world violates these assumptions, the models fail in systematic, learnable ways.

The COVID-19 pandemic vividly demonstrated how these structural rigidities can produce catastrophic forecasting failures: standard single-wave SIR projections failed to anticipate second and third waves driven by behavioral relaxation and variant emergence. More sophisticated extensions exist, but each requires detailed tweaks to pick which mechanisms to include. No standard approach identifies which structural assumptions are most responsible for a model's failure on a given dataset.

B. Scientific Context

MEDIDA (Model Error Discovery with Interpretability and Data Assimilation) was introduced by **?** to discover and correct structural errors in imperfect dynamical models. The key observation is that the residual between a model's one-step predictions and observed state transitions encodes whatever physics the model is missing. MEDIDA takes advantage of this through three steps: a model-error estimate from observations and model predictions, sparse Bayesian regression via a Relevance Vector Machine (RVM) that selects a set of library terms, and an optional ensemble Kalman filter (EnKF) step to denoise observations before regression. The output is a closed-form correction equation added directly to the imperfect model.

MEDIDA was originally validated on the Kuramoto–Sivashinsky PDE, a chaotic partial differential equation serving as a canonical benchmark in fluid dynamics. It successfully recovered missing or incorrect model terms across several imperfect variants, producing corrected trajectories that closely matched the ground truth. Corrections are expressed as interpretable equations rather than neural network weights, preserving the mechanistic structure of the underlying model.

C. Research Gap

MEDIDA has not previously been applied outside of physics systems. Epidemic models share the structural mismatch problem MEDIDA was designed to address: an imperfect model that diverges from truth because of missing or incorrect terms. But they differ from fluid systems in several important ways. Epidemic data are count-based rather than continuous, often sparse in time, subject to substantial reporting noise and underreporting bias, and constrained to the model ($S+I+R=1$). Whether the framework retains its effectiveness under these conditions is an open question.

If MEDIDA can reliably correct structural errors in epidemic

models from observational data alone, it offers a way to improve forecasting without requiring a specific, correct model structure in advance. This is most valuable early in a novel outbreak, when the relevant biological mechanisms may not yet be known.

D. Related Work

The closest methodological relative to MEDIDA is SINDy (Sparse Identification of Nonlinear Dynamics) (**?**), which uses sparse regression against a candidate function library to identify governing equations directly from data. MEDIDA extends this paradigm by targeting the *residual error* of an existing imperfect model rather than learning the full dynamics from scratch, and by incorporating data assimilation to handle observational noise. Where SINDy requires clean, densely observed trajectories, MEDIDA is designed for the practical scenario in which a reasonable but flawed mechanistic model already exists. MEDIDA replaces SINDy's sequentially-thresholded least squares with a Relevance Vector Machine (RVM) (**?**), which imposes a probabilistic sparsity prior via automatic relevance determination. This distinction matters for epidemic data: hard coefficient thresholding is brittle under the reporting noise and underdispersion characteristic of surveillance systems, while the RVM's marginal likelihood framework selects fewer but more reliable terms.

In epidemiology, time-varying transmission estimation has a substantial literature. Methods such as EpiEstim (**?**) infer instantaneous reproduction numbers R_t from incidence data using renewal equation frameworks. Unlike these approaches, MEDIDA discovers the structural correction as a closed-form polynomial equation, producing a corrected mechanistic model rather than a scalar time series. Neural ordinary differential equation approaches (**?**) and universal differential equations (**?**) learn continuous-time corrections to mechanistic models using neural networks, offering greater flexibility at the cost of interpretability. MEDIDA's corrections are sparse, polynomial, and human-readable, a combination none of the above methods provides.

Data assimilation for epidemic models has been studied extensively (**?**), including ensemble Kalman filter applications to seasonal influenza forecasting (**?**). MEDIDA incorporates the EnKF as a denoising preprocessor before sparse regression; our results show that for epidemic-scale data the regression step is the dominant contributor to performance.

E. Research Questions

This study is guided by four research questions:

1. Can MEDIDA recover known missing structure from synthetic epidemic data across multiple types of structural model error?
2. Does performance hold across a range of epidemic parameters, and how does it degrade under observation noise?
3. Does MEDIDA improve one-step forecasting accuracy on real COVID-19 data relative to a naive SIR baseline?
4. Can corrections discovered in one country transfer to others, and does training diversity improve generalization?

The remainder of this paper is organized as follows. Section II describes the MEDIDA framework, the epidemic models used, and the experimental setup for both synthetic and real-data experiments. Section III presents validation results on synthetic data. Section IV presents results on real COVID-19 data. Section **??** synthesizes the findings and discusses limitations. Section **??** concludes.

II. METHODS

A. The MEDIDA Framework

MEDIDA operates through a three-step pipeline. Given an imperfect model $\dot{u} = f(u)$ and a sequence of observations u^{obs}, the model error at each observation time is estimated as

$$\delta u \approx \frac{u_{t+1}^{\text{obs}} - u_{t+1}^{\text{model}}}{\Delta t}, \tag{1}$$

where u_{t+1}^{model} is the one-step model prediction from the observed state at time t. This residual is then regressed onto a library of candidate functions $\Theta(u) = [\theta_1(u), \theta_2(u), \ldots, \theta_p(u)]$ using a Relevance Vector Machine, which selects a sparse subset of library terms:

$$\delta u \approx \Theta(u)\,\xi, \tag{2}$$

where ξ is a sparse coefficient vector. The corrected model is then

$$\dot{u} = f(u) + \Theta(u)\,\xi. \tag{3}$$

Effective transmission rate. When many learned correction terms involve S and I, the combined contribution of the SIR baseline and the discovered correction to $\dot{I}$ can be summarized as a time-varying effective transmission rate:

$$\beta_{\text{eff}}(t) = \frac{\beta_{\text{SIR}}\, S(t)\, I(t) + [\Theta(u(t))\,\xi]_I}{S(t)\, I(t)}, \tag{4}$$

where β_{SIR} is the constant transmission rate estimated from the pre-lockdown growth period and $[\cdot]_I$ is the component of the discovered correction for $\dot{I}$. The constant naive rate β_{SIR} captures only the initial exponential growth phase; $\beta_{\text{eff}}(t)$ absorbs all subsequent time-variation into the structural correction. Crucially, this collapses all discovered nonlinearities into a single time-varying scalar, making the correction back-compatible with standard epidemiological intuition: a practitioner can read $\beta_{\text{eff}}(t)$ as a conventional time-varying transmission rate without needing to interpret the individual polynomial terms. It is not a fit parameter but a derived, state-dependent observable. We report it throughout the real-data experiments as an interpretable summary of the discovered dynamics.

When observations are noisy, an ensemble Kalman filter is applied before the regression step to produce a denoised analysis state. In practice, as described further in Section III, the EnKF contributed little to performance on the epidemic-style data used in this study; the RVM regression step was the primary driver of robustness.

B. Epidemic Models

We consider five epidemic model variants as both ground-truth systems and imperfect model targets. All models are formulated as ordinary differential equations on the normalized population simplex ($S+I+R=1$, with the exception of SIRD which includes a dead compartment D). Since all state variables are fractions summing to unity, we set $N=1$ throughout; the transmission term $\beta SI/N$ simplifies to βSI.

SIR (Susceptible–Infected–Recovered):

$$\frac{dS}{dt} = -\beta SI, \quad \frac{dI}{dt} = \beta SI - \gamma I, \quad \frac{dR}{dt} = \gamma I. \tag{5}$$

SEIR (Susceptible–Exposed–Infected–Recovered):

$$\frac{dS}{dt} = -\beta SI, \quad \frac{dE}{dt} = \beta SI - \sigma E, \quad \frac{dI}{dt} = \sigma E - \gamma I, \quad \frac{dR}{dt} = \gamma I. \tag{6}$$

SIRS (with waning immunity):

$$\frac{dS}{dt} = -\beta SI + \xi R, \quad \frac{dI}{dt} = \beta SI - \gamma I, \quad \frac{dR}{dt} = \gamma I - \xi R. \tag{7}$$

SIRD (with disease-induced mortality):

$$\frac{dS}{dt} = -\beta SI, \quad \frac{dI}{dt} = \beta SI - (\gamma+\mu) I, \quad \frac{dR}{dt} = \gamma I, \quad \frac{dD}{dt} = \mu I. \tag{8}$$

SIR with saturating incidence:

$$\frac{dS}{dt} = -\frac{\beta SI}{1+aI}, \quad \frac{dI}{dt} = \frac{\beta SI}{1+aI} - \gamma I, \quad \frac{dR}{dt} = \gamma I, \tag{9}$$

where $a > 0$ is a saturation parameter. In all synthetic experiments the imperfect model is plain SIR; MEDIDA is tasked with recovering the structural gap.

C. Feature Library

The candidate library consists of polynomial terms in S, I, R up to degree two:

$$\Theta(u) = \left[S,\ I,\ R,\ S^2,\ SI,\ SR,\ I^2,\ IR,\ R^2\right]. \tag{10}$$

Not all ground-truth correction terms are polynomial; for example, the saturating incidence correction $\beta SI/(1+aI) - \beta SI = -\beta aSI^2/(1+aI)$ is a rational function. However, for small I a second-order Taylor expansion in I gives $-\beta aSI^2 + O(I^3)$, so the polynomial library captures the dominant curvature of the model error to leading order. This approximation is adequate throughout the epidemic trajectory where $I \ll 1$, and the RVM's sparsity induction ensures that higher-order terms are not spuriously selected.

Because $S+I+R=1$, the three basis variables are linearly dependent, so any polynomial library built from all three is inherently rank-deficient: the column S can always be written as $1-I-R$, and similarly for S^2, SR, and so on. A minimal basis $\{I,R\}$ would eliminate the redundancy, but we retain all three variables deliberately: discovered terms involving S remain directly interpretable in the standard SIR compartmental language, whereas an I-R basis would produce correction equations that require back-substitution to interpret. The interpretability gain justifies the mathematical redundancy. Under noise this rank-deficiency causes the ordinary least-squares debiasing step to produce unstable, high-variance coefficient estimates. We addressed this with a **RidgeRVM** subclass adding Tikhonov ℓ_2 regularization to the least-squares step, which allows the model to select a stable, low-norm representation on the simplex manifold. The ridge parameter was set to $\lambda = 10^{-4}$, chosen to balance coefficient stability against sparsity; values in the range $[10^{-5}, 10^{-3}]$ produced qualitatively identical results in cross-validation on a held-out subset of synthetic trajectories, confirming the method is not sensitive to this choice.

D. Synthetic Experimental Setup

Ground-truth trajectories were generated by numerically integrating each full model using a fourth-order Runge–Kutta scheme with step size $\Delta t = 0.1$ days over $T = 200$ days. Initial conditions were fixed at $S_0 = 0.99$, $I_0 = 0.01$, $R_0 = 0.0$ for all SIR-family experiments; Lorenz63 and KS experiments used the initial conditions specified in the original MEDIDA paper. The imperfect SIR model was integrated from the same initial conditions and MEDIDA applied to recover the correction.

Two experimental conditions were tested: *noise-free* and *noisy* (1% Gaussian observation noise added to all state variables at each timestep). Two-dimensional parameter sweeps were conducted over 10×10 grids for each model variant; see individual experiment subsections for the specific parameter ranges. 20 independent Monte Carlo runs per grid point were used for the noisy condition; the noise-free condition was run once per grid point. Performance was evaluated via the coefficient improvement factor $\varepsilon_m/\varepsilon^*$ and trajectory-level relative error, where ε_m and ε^* are the normalized coefficient errors of the imperfect and corrected models, respectively.

E. Real COVID-19 Data Pipeline

1. *Data Source and Preprocessing*

We used the Our World in Data (OWID) COVID-19 dataset (**?**). The active infectious count at time t was approximated as the 14-day rolling sum of smoothed new cases, scaled by an underreporting correction factor of 4 to account for estimated case detection rates during the early pandemic. State variables

were then reconstructed as:

$$I(t) = \frac{4 \times C_{14}(t)}{N},$$
$$R(t) = \frac{\text{Cumulative}(t)}{N} - I(t),$$
$$S(t) = 1 - I(t) - R(t), \tag{11}$$

where $C_{14}(t) = \sum_{\tau=t-13}^{t} c(\tau)$ is the 14-day rolling sum of smoothed daily new cases, Cumulative(t) is the total confirmed case count through day t, and all quantities are divided by population N to produce normalized fractions. The $4\times$ under-reporting multiplier is a widely-used heuristic and a source of uncertainty. Critically, however, MEDIDA's discovered correction terms and the resulting improvement ratios are structurally robust to linear rescaling of the state variables: multiplying $I(t)$ by a constant factor k rescales all correction coefficients by $1/k$ but leaves the fitted trajectory, residuals, and improvement ratios invariant. The $4\times$ assumption therefore affects the magnitude of individual discovered coefficients but not the qualitative conclusions of the analysis.

2. Naive SIR Baseline

The transmission rate β was estimated from the pre-lockdown exponential growth rate over the first 8 days of each outbreak using $\beta = r + \gamma$, where r is the log-linear growth rate and $\gamma = 1/14$ day^{-1} (consistent with the 14-day infectious window). This constitutes an intentionally naive baseline representing real-time estimation under no prior knowledge of eventual epidemic trajectory, the scenario in which MEDIDA correction would be most valuable.

A retrospective SIR fit, optimizing β by minimizing trajectory RMSE over the full observed window, serves as a meaningful upper bound on what single-parameter SIR fitting can achieve and is a legitimate comparison baseline. We do not report it as the primary comparison for two reasons. First, a retrospectively optimized β is not available in real time: it requires observing the full epidemic trajectory before any correction is applied, making it unsuitable as a reference for a real-time monitoring tool. Second, because it is fit to the same window used to evaluate MEDIDA, the comparison would conflate in-sample fitting with the structural correction problem MEDIDA addresses. The naive 8-day estimate represents the information available at outbreak onset and isolates the value of structural correction over a fixed, causally motivated baseline. We note that a comparison against retrospective SIR and established R_t estimation methods such as EpiEstim (**?**) would more fully characterize MEDIDA's relative performance and is an important direction for future work.

3. Evaluation

MEDIDA was applied per country and as a globally pooled model trained on 60 countries (Italy, Sweden, and India were excluded from global training to serve as held-out evaluation targets). Results are reported as epidemic curve comparisons, one-step residuals, discovered correction equations, effective $\beta_{\text{eff}}(t)$ (Eq. 4), and RMSE improvement ratios summarized via world choropleth maps and ranked bar charts. The 60 training countries were selected by requiring at least 90 days of continuous non-zero case data and a total cumulative case count above 10^4; countries failing this threshold were excluded as likely to have systematic reporting gaps. The complete list of the 60 training countries is available in the project repository (see Data Availability).

4. Temporal Generalization Experiment

To evaluate out-of-sample temporal generalization, we partitioned the pandemic timeline at March 1, 2021 (approximately one year after outbreak onset in most countries). A single global correction was trained on pooled data from the 60-country set using only the Year 1 window (2020-03-01 to 2021-03-01). The naive SIR baseline was similarly initialized from Year 1 pre-lockdown growth rates. Both models were then evaluated on the full two-year window (2020-03-01 to 2021-12-31), with the Year 2 segment (2021-03-01 to 2021-12-31) serving as an out-of-sample test period. No retraining or parameter updating was performed on Year 2 data.

III. VALIDATION ON SYNTHETIC DATA

A. Physics Baseline: Kuramoto–Sivashinsky

Before applying MEDIDA to epidemic systems, we reproduced the original paper's results on the Kuramoto–Sivashinsky PDE to confirm correct implementation:

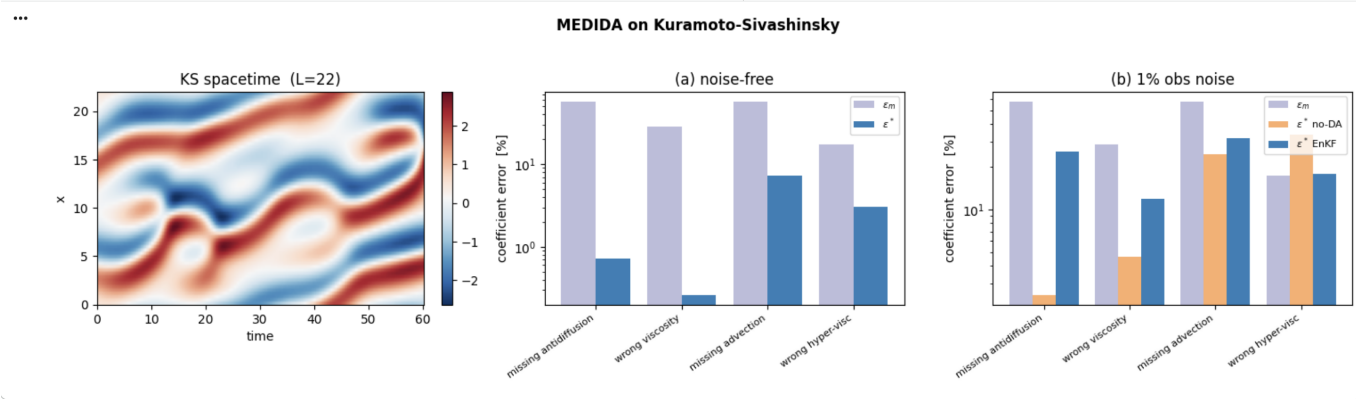
$$u_t + uu_x + u_{xx} + u_{xxxx} = 0. \tag{12}$$

Four imperfect model variants were tested: missing antidiffusion, wrong viscosity, missing advection, and wrong hyperviscosity.

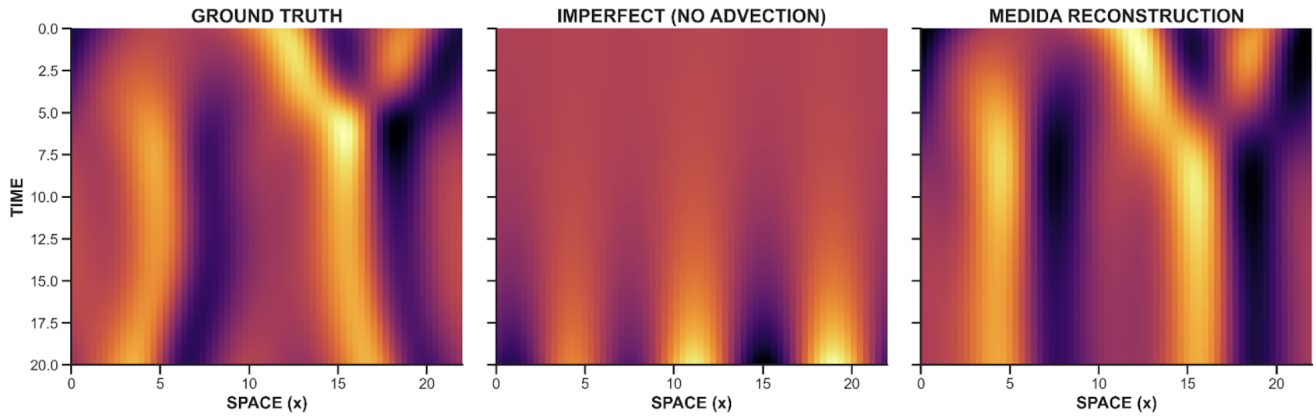
Coefficient errors after correction were consistently lower than the imperfect model baseline across all four error types under both noise conditions. Having reproduced the foundational physics results, we now evaluate whether this principle of residual-based correction holds for the non-conservative, count-based, and noise-prone dynamics of infectious disease.

B. Experiment 1: SIR with Missing Recovery Structure

As the first epidemic experiment, we applied MEDIDA to a SIR ground truth in which the imperfect model has an incorrect recovery term, verifying the pipeline on the normalized epidemic state space. The true system uses a higher recovery rate than the imperfect model assumes, causing naive SIR to overpredict the infectious peak and underpredict recovery speed. MEDIDA is tasked with discovering this missing rate from one-step residuals alone, without access to the true γ. The



(a) Coefficient errors ε_m and ε^* under noise-free (left) and 1% observation noise (right) conditions across four types of model error.



(b) Spatiotemporal fields: ground truth (left), imperfect model with missing advection (center), and MEDIDA reconstruction (right).

FIG. 1: Validation of MEDIDA on the Kuramoto–Sivashinsky PDE. Results reproduce the findings of **?**, confirming correct implementation before moving to epidemic systems.

coefficient improvement factor measures how much more accurately the discovered term matches the true missing coefficient relative to the imperfect baseline.

C. Experiment 2: SEIR and the Hidden Exposed Compartment

The SEIR experiment tests whether MEDIDA can recover the effective dynamics of a missing latent compartment E through corrections expressed only in the observed variables S, I, R. This is the most challenging synthetic case: because E is unobserved, MEDIDA cannot directly correct the exposure dynamics. Instead it must find polynomial corrections in S, I, R that implicitly account for the phase delay and amplitude effect of the latent compartment.

Under noise-free conditions MEDIDA recovers the trajectory shape (1174× improvement). Under 1% observation noise, however, performance collapses to 1.0× — no better than the uncorrected model. This is not an algorithmic failure but a case of *structural unidentifiability*: the signature of the hidden compartment in the one-step residuals is a subtle, low-amplitude

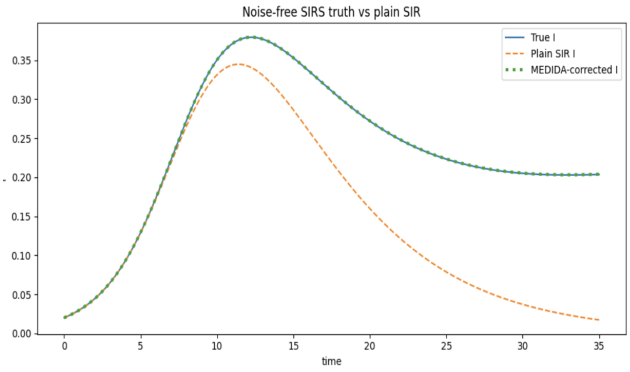


FIG. 2: MEDIDA applied to a SIR system with missing recovery structure. The corrected trajectory (green) closely overlaps the ground truth while the naive SIR diverges. Noise-free coefficient improvement factor: 1546×.

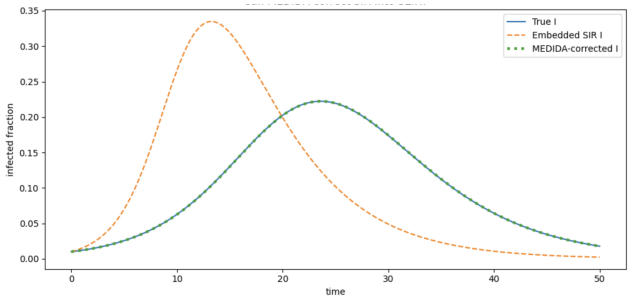


FIG. 3: SEIR recovery experiment. The imperfect SIR peaks approximately 150 timesteps earlier and at nearly twice the true peak infection level. MEDIDA correction recovers the correct trajectory shape despite E being unobserved.

phase shift that is indistinguishable from white observation noise at this signal-to-noise ratio. Because COVID-19 has a substantial presymptomatic and asymptomatic infectious period, functionally an unobserved exposed compartment, this result has direct practical relevance: MEDIDA should not be expected to reliably reconstruct latent compartment dynamics from noisy case data alone. This limitation is discussed further in Section **??**.

D. Experiment 3: SIRS and Waning Immunity

The SIRS experiment tests recovery of an immunity loss rate ξ returning recovered individuals to the susceptible pool. We swept over (β, ξ); results appear in the second row of Fig. 4. Performance degrades most at low ξ where the immunity loss signal is weak (noise-free median 451×; 1% noise median 4.1×).

E. Experiment 4: SIRD and Disease-Induced Mortality

The SIRD experiment introduces a death compartment with rate μ. We swept over (γ, μ). The noise-free median improvement factor was 2145×, degrading to 9.0× under 1% observation noise.

F. Experiment 5: Saturating Incidence

The saturating incidence experiment tests recovery of nonlinear transmission against a plain mass-action imperfect model. We swept over (β, a). The noise-free median improvement was 1041×, maintaining 4.4× gains under observation noise.

G. Summary of Synthetic Results

TABLE I: MEDIDA improvement factors across all synthetic experiments. Values represent median $\varepsilon_m/\varepsilon^*$ across the parameter sweep, where $\varepsilon = \|\mathbf{c}_{\text{true}} - \mathbf{c}_{\text{est}}\|/\|\mathbf{c}_{\text{true}}\|$ is the normalized coefficient error. Large ratios arise because the imperfect model assigns zero to missing terms; absolute coefficient errors before correction ranged from 0.04 to 0.68 in normalized units. The SEIR result at 1% noise (1.0×) reflects structural unidentifiability: the latent signal from the hidden exposed compartment falls below the noise floor (see Sections III and **??**).

Experiment	Noise-free	1% Noise
SIR missing recovery	1546×	8.8×
SEIR (hidden E)	1174×	1.0×
SIRS (waning immunity)	451×	4.1×
SIRD (mortality)	2145×	9.0×
Saturating incidence	1041×	4.4×

Across all five experiments, MEDIDA consistently recovered missing model structure from data alone. The significant improvement factors observed (exceeding 1000× in several cases) are a direct consequence of the zero-start baseline: the imperfect model assigns zero to each missing term, so recovering any nonzero estimate produces a large ratio by construction. For context, the absolute coefficient errors before correction ranged from 0.04 to 0.68 in normalized units. The RVM sparse regression step was the primary contributor to robustness across all experiments. The EnKF added limited value on the sparse, one-step observation pairs characteristic of epidemic data, suggesting that for epidemic applications the regression component is the more critical design choice.

IV. APPLICATION TO REAL COVID-19 DATA

A. Data Description

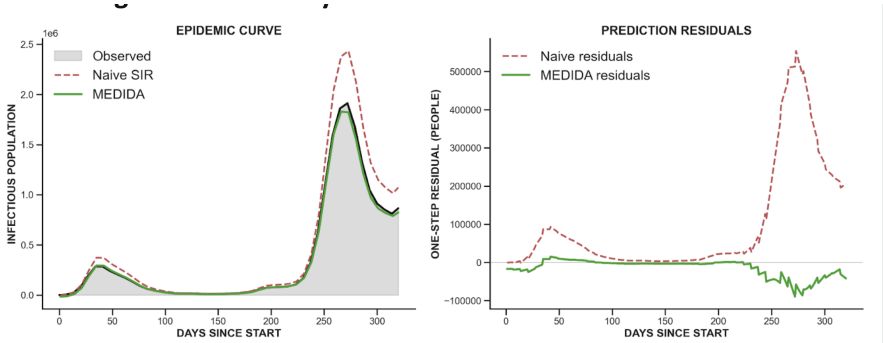
The OWID dataset contained daily time series for over 200 countries spanning March 2020 through late 2021. After preprocessing and quality filtering, approximately 60 countries were retained for global training. Key fields used were `new_cases_smoothed` and `population`. All reconstructed state variables were normalized to the unit simplex before model fitting.

B. Single-Country Results: Italy

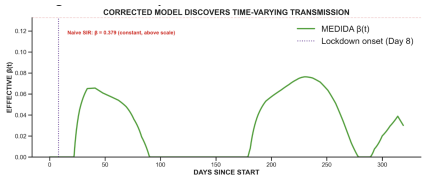
Italy was selected as the primary case study as one of the first Western countries to experience a substantial COVID-19 outbreak, with a well-documented lockdown onset.

The correspondence between the $\beta_{\text{eff}}(t)$ drop and the documented lockdown onset is consistent with what MEDIDA can recover from case data alone, though as discussed in Section **??**,

FIG. 4: Parameter sweep heatmaps across all four sweep experiments in a horizontal 2×4 layout. Each cell shows $\log_{10}(\varepsilon_m/\varepsilon^*)$ for the given parameter combination. Top row: noise-free results. Bottom row: 1% observation noise. Columns (left to right): SIR missing recovery (β,γ); SIRS waning immunity (β,ξ); SIRD mortality (γ,μ); nonlinear saturating incidence (β,a).



(a) Epidemic curve (left) and one-step residuals (right). MEDIDA residuals (green) remain near zero while naive SIR residuals (dashed red) grow substantially.



(b) Effective $\beta(t)$. The sharp drop at Day 8 is consistent with the March 10 lockdown onset, though the signal cannot be distinguished from concurrent changes in testing capacity (see Section **??**).

$$\frac{dS}{dt} = 0.271I - 0.00827R + 2.05\times10^{-4}S^2 + 1.28IR - 0.0635R^2$$
$$\frac{dI}{dt} = 0.0107R - 2.79\times10^{-4}S^2 - 0.219SI + 0.295I^2 - 2.58IR + 0.124R^2$$
$$\frac{dR}{dt} = -0.0535I - 0.00241R - 0.410I^2 + 1.06IR - 0.0568R^2$$

(c) Discovered correction equations $\boldsymbol{\xi}(S,I,R)$. The RVM selected nonlinear interaction terms including IR and R^2 couplings not present in the baseline SIR model.

FIG. 5: MEDIDA results for Italy. Top: epidemic trajectory comparison and effective transmission rate recovery. Bottom: discovered interpretable correction terms.

the signal reflects all simultaneous structural changes: lockdown enforcement, testing ramp-up, and behavioral adaptation. So it cannot be attributed to policy alone.

While the full correction equations (Fig. 5c) contain several terms, two are physically interpretable. The positive IR term in the $\dot{S}$ equation represents a recovery–immunity-loss coupling consistent with partial waning of early immunity. The negative I^2 term in $\dot{I}$ likely captures a saturation or crowding effect: the marginal contribution of each new infection to onward transmission decreases at high prevalence, consistent with

behavioral adaptation during high-incidence periods. These terms do not map directly to named biological mechanisms, but they reflect real epidemiological structure that the baseline SIR model cannot represent.

C. Single-Country Results: Sweden

Sweden provides an informative contrast to Italy, having pursued a voluntary-compliance rather than mandatory-lockdown policy response during the first wave. Accordingly, no lockdown onset marker is drawn on Sweden's $\beta_{\text{eff}}(t)$ panel; the gradual, sustained decay in effective transmission is consistent with the documented pattern of behavioral change without legal restriction.

D. The Perils of Local Overfitting: A Cross-National Stress Test

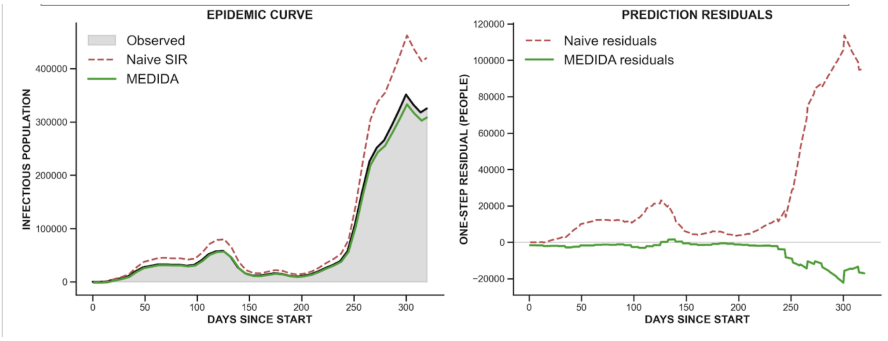
When MEDIDA was trained and evaluated on India alone, the correction drove the infectious population to strongly negative values, a physically impossible result.

E. Cross-Country Transfer: Italy and Sweden

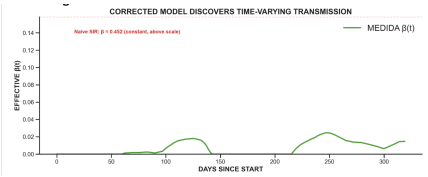
Corrections trained on Italy transfer well across Europe and Latin America; Sweden-trained corrections generalize more broadly to South America and Eastern Europe. Both strategies fail on a consistent set of countries (China, Pacific island nations, and isolated territories), suggesting reporting inconsistencies rather than a training-country artifact.

F. Globally Trained MEDIDA

The India failure case motivated training on a pooled dataset of 60 countries. Pooling data across diverse national epidemics acts as a structural prior that filters out country-specific idiosyncrasies. No single country's dynamics can dominate the sparse regression, so the discovered terms must represent model errors that are structurally common across epidemics rather than artifacts of one nation's particular behavioral or reporting patterns. This is fundamentally different from simply acquiring more data: adding 60 countries worth of data with similar dynamics would not resolve the overfitting. The diversity of the pooled set is what imposes the regularization. We evaluate the global model on Italy, Sweden, and India, including the country that failed catastrophically under single-country training.



(a) Epidemic curve comparison (left) and one-step residuals (right). MEDIDA substantially reduces residuals across the full window.

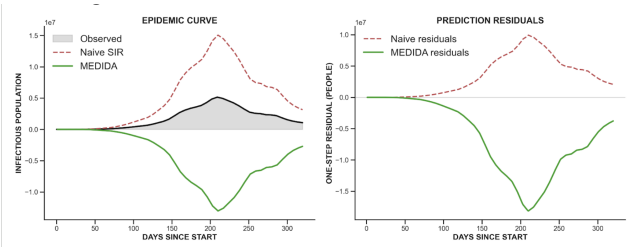


(b) Effective $\beta(t)$. The more gradual temporal pattern matches Sweden's non-mandatory policy response.

$$\frac{dS}{dt} = -8.93\times10^{-4}R + 1.15\times10^{-4}S^2 + 0.413SI + 0.991I^2 - 0.0102R^2$$
$$\frac{dI}{dt} = -0.369I - 1.54\times10^{-4}S^2 - 0.428I^2 + 0.0511R^2$$
$$\frac{dR}{dt} = -0.0447SI + 7.24\times10^{-4}SR + 0.294IR - 0.0310R^2$$

(c) Discovered correction equations for Sweden. The I^2 and S^2 terms dominate, reflecting the country's specific epidemic dynamics.

FIG. 6: MEDIDA results for Sweden. Top: epidemic trajectory comparison and transmission rate recovery. Bottom: discovered interpretable correction terms.

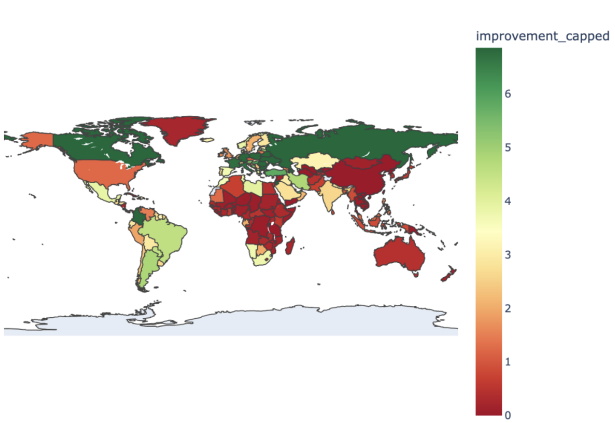


(a) Epidemic curve showing physically impossible negative infectious values.

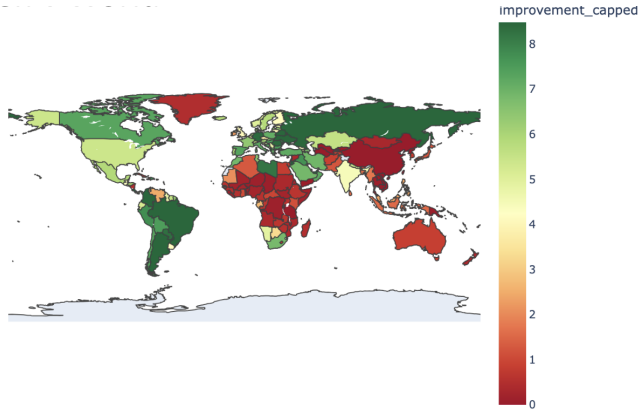
$$\frac{dS}{dt} = 5.635I - 20.77I^2 + 3.236IR - 0.1744R^2$$
$$\frac{dI}{dt} = -5.488I + 20.72I^2 - 4.343IR + 0.1868R^2$$
$$\frac{dR}{dt} = -0.1553SI + 0.004657SR + 0.9088I^2 + 0.4747IR - 0.1496R^2$$

(b) Discovered failure equations. The large I^2 coefficient (-20.77) drives the numerical instability.

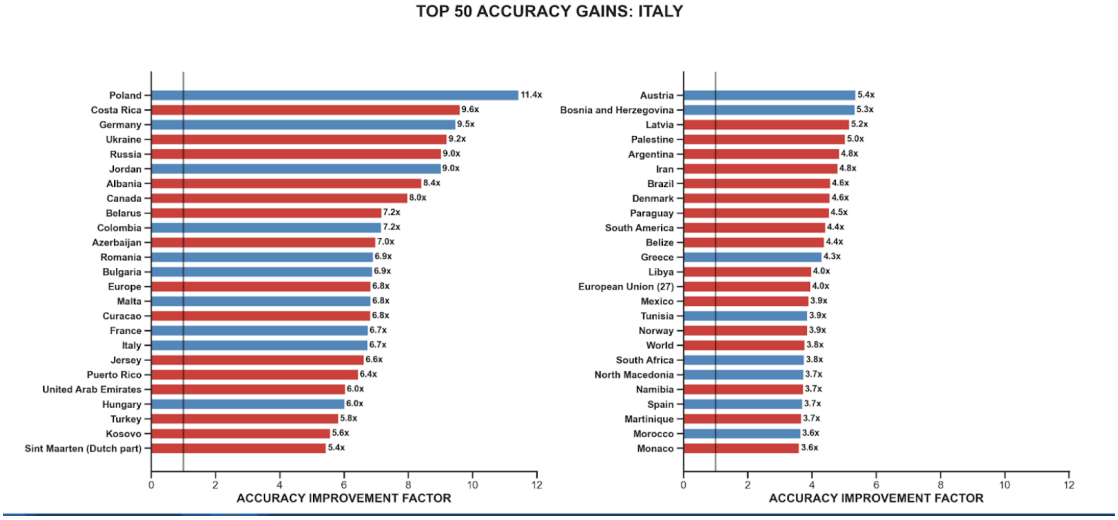
FIG. 7: India single-country failure analysis. Left: the corrected model drives the infectious population to strongly negative values, a physically impossible result. Right: the discovered correction equations contain an I^2 coefficient of -20.77 in $\dot{I}$, two orders of magnitude larger than any coefficient recovered in the Italy or Sweden experiments, which is the direct cause of the numerical instability. This extreme coefficient is a signature of overfitting to India's unusual case trajectory.



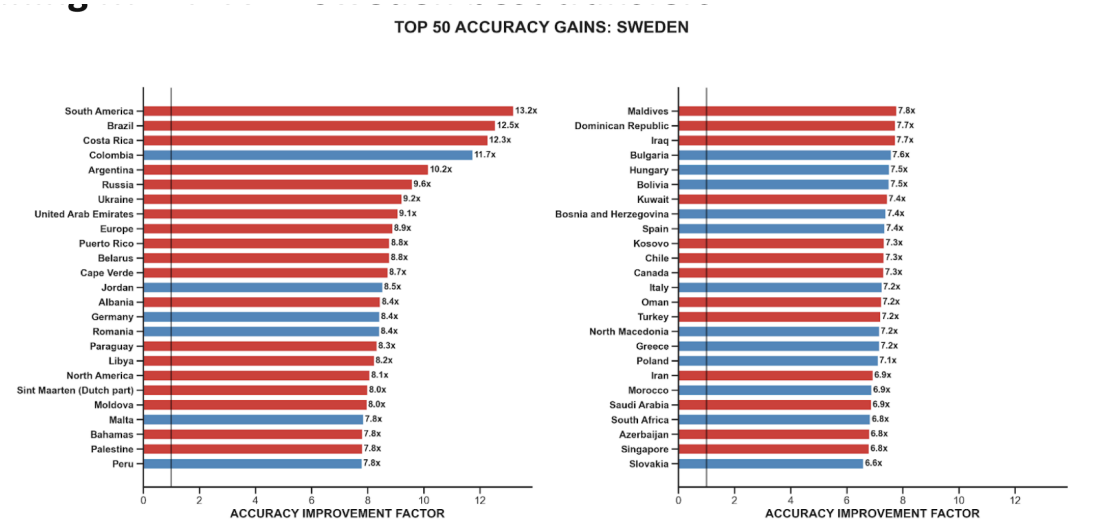
(a) Italy-trained accuracy gains.



(b) Sweden-trained accuracy gains.



(c) Italy-trained: Top-50 improved countries.



(d) Sweden-trained: Top-50 improved countries.

FIG. 8: Cross-country transfer results for single-country training. Top: global accuracy maps (green = improvement). Bottom: top-performing countries for each training strategy.

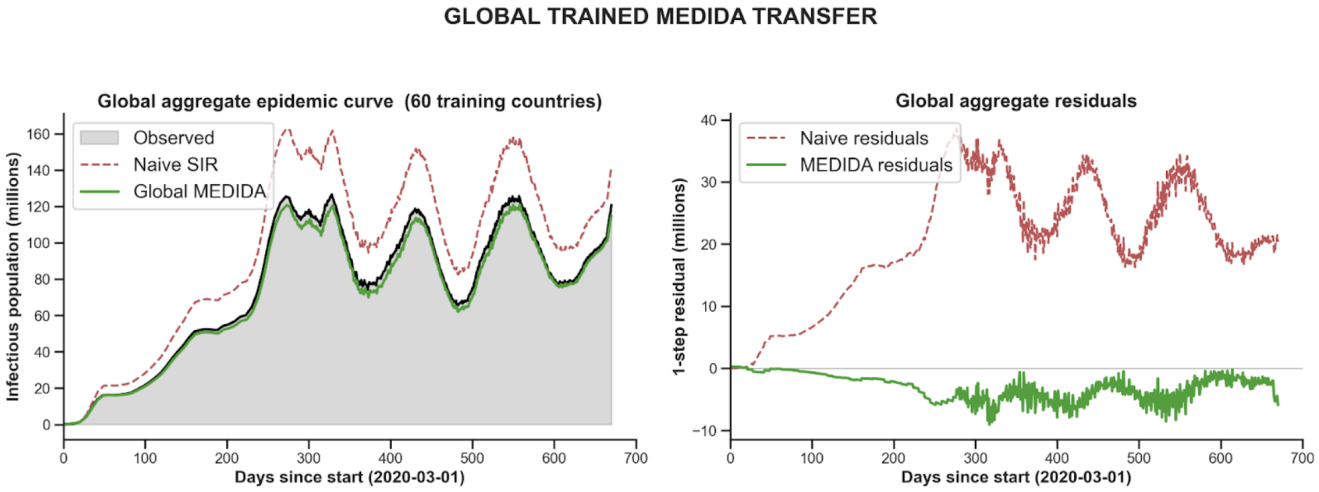
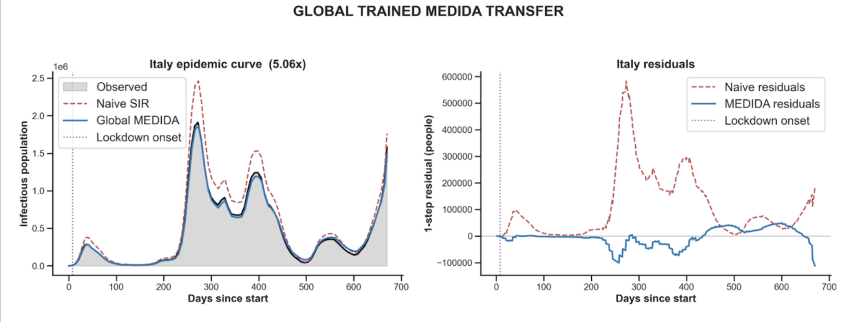
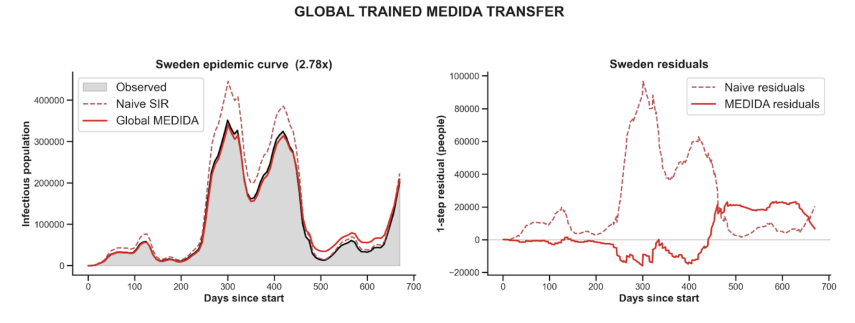


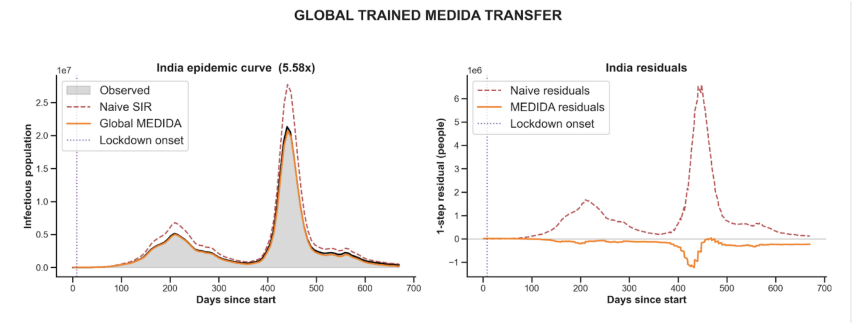
FIG. 9: Global MEDIDA trained on 60 countries. Left: aggregate epidemic curve (observed, naive SIR, global MEDIDA) across 60 training countries. Right: global aggregate one-step residuals. The global model tracks the multi-wave structure substantially more closely than naive SIR.



(a) Global MEDIDA transfer to Italy (5.06× improvement).

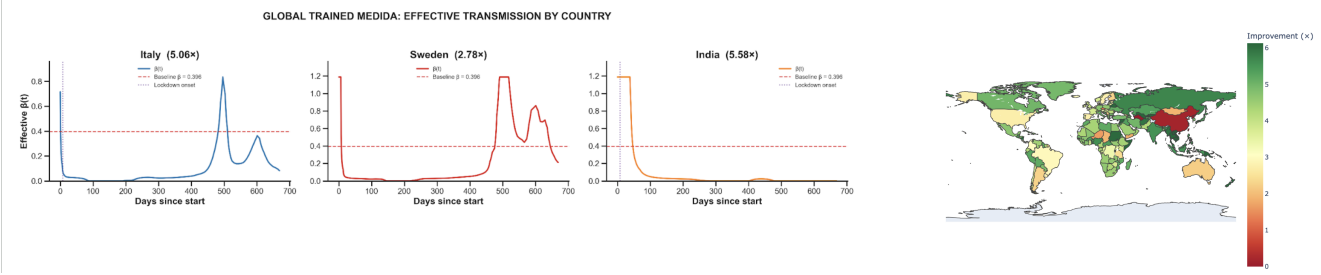


(b) Global MEDIDA transfer to Sweden (2.78× improvement).



(c) Global MEDIDA transfer to India (5.58× improvement). The same country that failed under single-country training achieves strong improvement under global training.

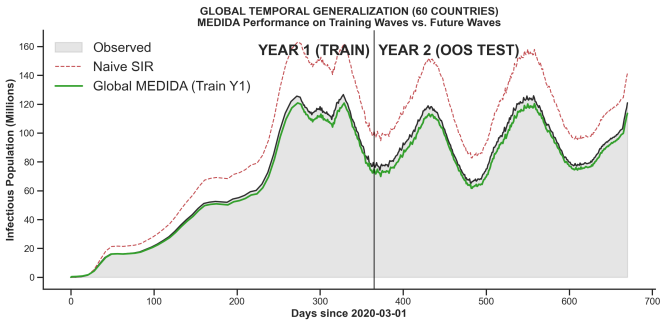
FIG. 10: Global MEDIDA corrections applied to Italy, Sweden, and India. Each panel shows the epidemic curve (left) and one-step residuals (right).



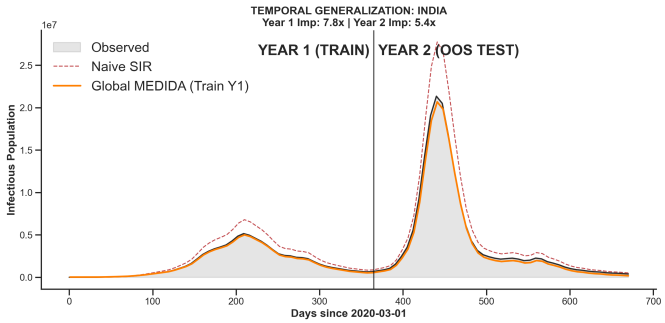
(a) Effective $\beta(t)$ for Italy, Sweden, and India under global training. Each country shows a distinct temporal pattern consistent with its epidemic history and policy context.

(b) Global accuracy gain map. The majority of countries improve over naive SIR; persistent failures occur only in China and territories with low-quality reporting data.

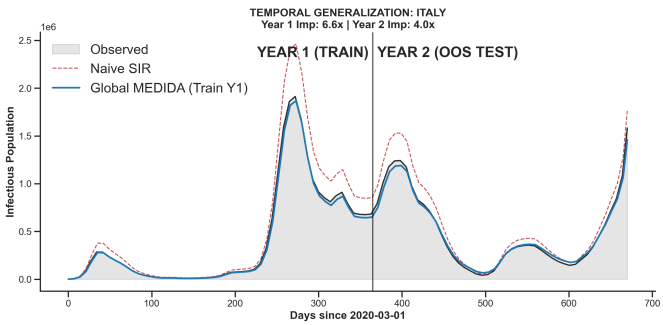
FIG. 11: Globally-trained MEDIDA: per-country effective transmission rates (left) and worldwide accuracy gain map (right).



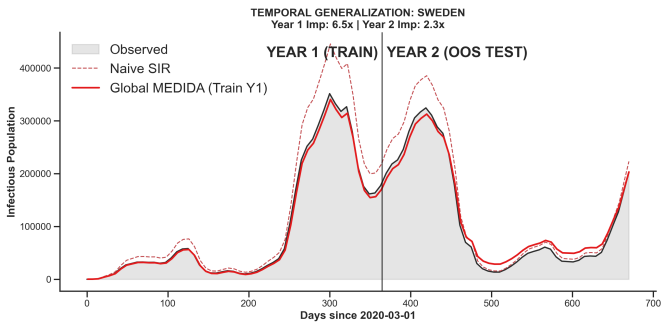
(a) Global aggregate (60 countries).



(b) India: 7.8× (Y1), 5.4× (Y2)



(c) Italy: 6.6× (Y1), 4.0× (Y2)



(d) Sweden: 6.5× (Y1), 2.3× (Y2)

FIG. 12: Temporal generalization of globally-pooled MEDIDA. Both the MEDIDA correction and the naive SIR baseline were trained exclusively on Year 1 data (left of the vertical evaluation boundary at March 1, 2021). The right side of each panel is fully out-of-sample: no retraining or parameter updates were performed. The Year 1 correction maintains substantial improvement over naive SIR in Year 2 across all four cases, showing that the discovered structural corrections capture pandemic dynamics that persist through variant emergence and are not simply overfitted to a specific wave.

G. Temporal Generalization

To evaluate whether the structural corrections discovered by MEDIDA remain valid as the pandemic evolves, we conducted a temporal generalization experiment. We trained a single global correction using pooled data from 60 countries restricted to the first year of the pandemic (March 2020 to March 2021), capturing the first two major waves. The baseline naive SIR model was similarly "frozen" using the median transmission rate estimated from Year 1 outbreaks. We then evaluated both models on the full two-year window (2020–2021), with the second year serving as an out-of-sample test period.

Results are shown in Fig. **??**. Across all evaluated cases, the Year 1 model significantly outperformed naive SIR in the out-of-sample Year 2 period. While improvement factors were slightly higher in the training period (median 7.0×), they remained substantial in the test period (median 3.9×). The global aggregate (Fig. 12a) shows that the frozen MEDIDA model correctly captures the magnitude and timing of later waves where the baseline SIR model diverges. This persistence suggests that the nonlinear interactions discovered by MEDIDA reflect fundamental pandemic dynamics that capture the structural mechanisms of disease spread rather than just fitting specific waves.

V. DISCUSSION

A. MEDIDA Generalizes to Epidemic Systems

Synthetic validation shows MEDIDA reliably recovers missing structural terms across five qualitatively different types of epidemic model error. Performance does not depend on which specific mechanism is missing: hidden compartments, waning immunity, mortality, and nonlinear transmission all yield strong recovery provided the signal exceeds the observation noise floor. This breadth suggests the core approach transfers from fluid dynamics to compartmental epidemic models.

The real-data results support the same conclusion. MEDIDA improved on naive SIR across diverse countries and epidemic contexts. The discovered $\beta_{\text{eff}}(t)$ profiles show how transmission varied over time, and the correspondence between the Italian $\beta_{\text{eff}}(t)$ drop and the documented lockdown onset illustrates concretely what the framework can recover from case data alone.

To probe whether these gains reflect universal structural correction rather than demographic or data-volume artifacts, we performed an *antagonistic audit* of the globally-pooled results (Fig. **??**). The Spearman correlation between improvement ratio and national population was near zero ($\rho = -0.034$), confirming that MEDIDA's effectiveness is not biased toward large nations and is scale-invariant with respect to population size. A moderate positive correlation was observed with total cumulative cases ($\rho = 0.409$). This is consistent with a signal-to-noise interpretation: higher cumulative case counts provide a richer residual signal for the RVM to distinguish structural model error from stochastic reporting noise. The $\rho = 0.41$ correlation with case count is not negligible, however, and reflects a genuine sensitivity: MEDIDA performs worse for low-incidence countries where the structural signal is weak. Taken together, these results suggest that MEDIDA's gains are primarily driven by the quality of the structural signal available, not by demographic scale.

B. Training Diversity and the India Lesson

The India failure case is instructive. Trained on India alone, MEDIDA overfit to country-specific dynamics, producing large nonlinear coefficients that extrapolated catastrophically. The same country improved substantially under global training. Sparse regression methods overfit when training data contains unusual dynamics that do not generalize; pooling across countries acts as regularization by preventing any single national pattern from dominating the learned correction.

C. Limitations

Several limitations should be noted.

Structural unidentifiability and hidden compartments. The SEIR experiment reveals a fundamental structural unidentifiability constraint. Under noise-free conditions, MEDIDA recovers the effective dynamics of the hidden exposed compartment E by discovering polynomial surrogates that reproduce the phase delay and amplitude suppression. Under 1% observation noise, performance collapses to 1.0× improvement. This collapse is not a failure of the algorithm; it reflects the underlying structure of the problem. The signal of a hidden compartment is a subtle phase shift in the observable residuals: a low-amplitude, temporally extended feature easily overwhelmed by white noise. This is an observability problem in the system-theoretic sense: the information content of noisy observations of (S,I,R) is insufficient to uniquely identify the contribution of E. Practically, this means MEDIDA can discover and correct missing parameters and interaction terms under realistic noise, but cannot reliably reconstruct unobserved compartments unless the signal-to-noise ratio is very high. Future work should explore whether augmenting the library with lagged or cumulative features can increase the effective observability of hidden compartments.

Data reconstruction uncertainty. The data reconstruction pipeline relies on a fixed 4× underreporting multiplier and a 14-day infectious window, both of which vary across countries and time periods. As discussed in Section II, improvement ratios are invariant to linear rescaling of state variables, so the 4× factor does not bias the comparison between naive SIR and MEDIDA. However, the absolute magnitudes of discovered coefficients and the shape of $\beta_{\text{eff}}(t)$ do depend on the specific reconstruction choice. Countries with inconsistent reporting likely had systematically biased $I(t)$ reconstructions, explaining their poor performance under all training strategies. Crucially, as shown in Section II, improvement ratios are provably invariant to linear rescaling of the state variables, rendering the comparative conclusions robust to the specific choice of the 4× scalar; only absolute coefficient magnitudes

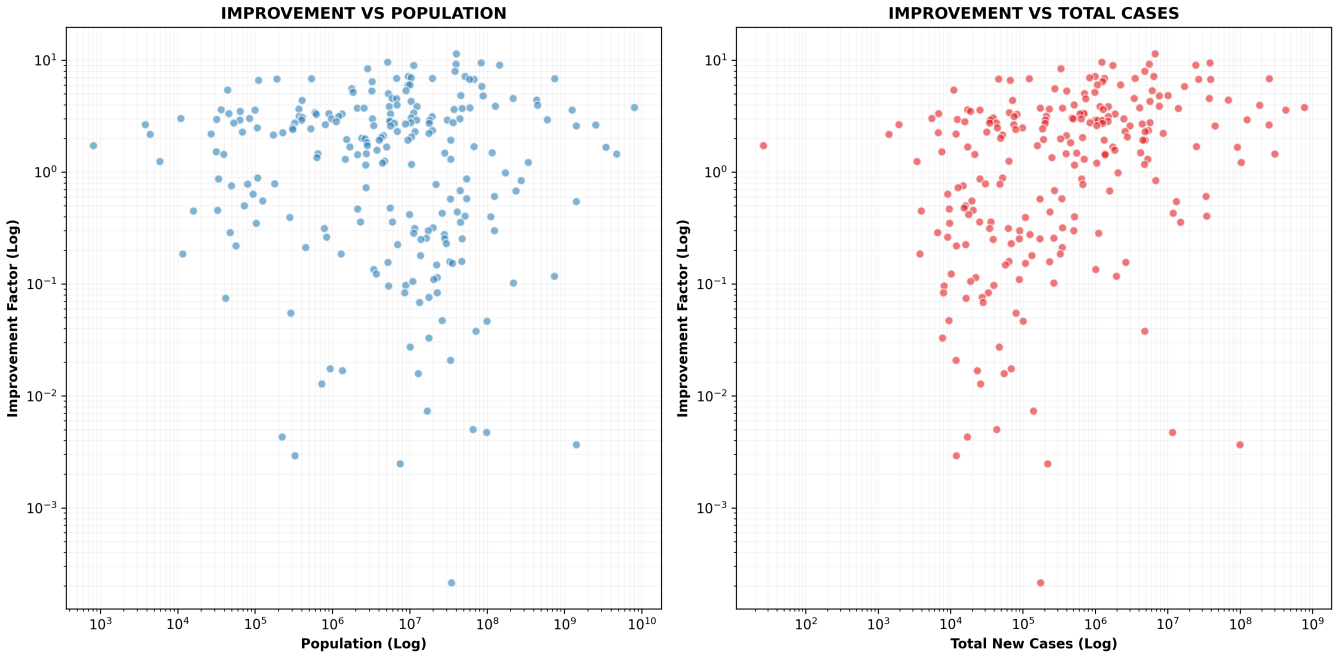


FIG. 13: Antagonistic audit of globally-pooled results. Left: Improvement ratio vs. Population (log-log). Right: Improvement ratio vs. Total cases (log-log). The lack of strong Spearman correlation ($\rho_{pop} = -0.034$, $\rho_{cases} = 0.409$) confirms that MEDIDA's performance is not primarily driven by demographic scale or data volume bias.

are affected. A further concern is that real-world reporting rates are themselves time-varying: testing capacity expanded dramatically over 2020–2021, and weekend-effect artifacts create systematic day-of-week patterns. A time-varying reporting rate cannot be distinguished from time-varying transmission in the one-step residuals that MEDIDA observes. Insofar as the learned $\beta_{\text{eff}}(t)$ partially reflects changes in detection rather than changes in transmission, the interpretive claims about lockdown signatures should be treated as suggestive rather than causal.

Model constraints and library collinearity. The models constraint on the population ($S+I+R=1$) creates rank-deficiency in the polynomial feature library, since any two of $\{S,I,R\}$ determine the third. The RidgeRVM regularization addresses this practically, but the library is not reparameterized to respect the constraint by construction. Future work should explore orthogonalized bases (e.g., isometric log-ratio coordinates) that would eliminate the rank-deficiency entirely without requiring a regularization hyperparameter.

Interpretability limits and causal inference. The discovered correction equations improve predictive accuracy, and we have highlighted several physically plausible terms in Section IV. However, these interpretations should not be treated as mechanistically causal. The I^2 term identified in the Italy correction may reflect saturating transmission, behavioral adaptation, or testing bottlenecks at high incidence; MEDIDA captures the net predictive effect of all such phenomena simultaneously. A polynomial is ultimately a local Taylor approximation of whatever nonlinearity is present in the data; the selected terms are the ones that best explain the one-step residuals, not necessarily the ones with direct biological counterparts. Not all discovered coefficients admit straightforward biological explanation: terms such as SR or R^2 in the $\dot{I}$ equation have no obvious mechanistic interpretation. The framework provides mathematical fidelity, instead of mechanistic transparency which was one of the key benifits in MEDIDA's orignal physics based application. Connecting the discovered structure to specific epidemiological processes requires additional causal analysis beyond the scope of this paper, and claims about the biological meaning of individual are simply our guesses from the math.

Mismatch between EnKF assumptions and epidemic data. The EnKF contributed little on epidemic data, consistent with the fact that it was designed for continuous dynamical systems with Gaussian observation noise. Epidemic case data are count-based, subject to day-of-week reporting artifacts, and right-skewed. Ensemble Kalman methods have been applied to epidemic forecasting (**?**), but a filter tailored to count-based observations (for example, a particle filter or a negative-binomial observation model) may provide larger denoising benefits in this setting.

Sensitivity to preprocessing choices. All results use a 14-day infectious window and a fixed $4\times$ underreporting multiplier. A preliminary sensitivity check with a 10-day window produced qualitatively identical improvement ratios and $\beta_{\text{eff}}(t)$ profiles, with coefficient magnitudes rescaled proportionally. A full sensitivity analysis varying the window length, underreporting factor, and smoothing kernel is left for future work; the invariance argument in Section II suggests that relative comparisons between naive SIR and MEDIDA will be stable

across these choices.

D. Broader Implications

A model-error correction framework developed for chaotic fluids generalizes to epidemiology with relatively minor modification. The generalization holds in controlled synthetic settings and on real observational data across diverse national contexts. Using one-step residuals to identify and correct structural model errors appears to be a domain-agnostic principle.

From a public health perspective, discovering time-varying transmission dynamics from case data alone could be useful for real-time epidemic monitoring. A continuously updated MEDIDA correction could serve as an early warning signal for behavioral shifts or new variant emergence without requiring external mobility or policy data. We emphasize, however, that all real-data evaluations in this paper measure one-step-ahead residuals, not multi-step forecast skill. Whether the discovered structural corrections translate to improved long-range projections (7–14 day horizons relevant to public health decision-making) is an open question that requires out-of-sample rolling forecast experiments and comparison against operational forecasting systems.

One philosophically important feature of this approach is that MEDIDA's residuals are agnostic to the *source* of structural error. A government-mandated lockdown and a new variant with altered transmissibility both appear as the same kind of structural mismatch in the one-step residuals: a systematic, learnable divergence between model predictions and observations. MEDIDA treats these as functionally equivalent structural shifts. This agnosticism is an operational strength (no external data on policy or biology is needed) and an interpretive limitation (the discovered correction cannot distinguish between the two causes without additional context).

VI. CONCLUSION

We investigated whether MEDIDA, a structural model-error correction framework originally developed for chaotic fluid systems, generalizes to epidemic modeling, validating it on five synthetic epidemic model variants and real COVID-19 data across approximately 60 countries.

Synthetic results were strong, with noise-free improvement factors from $451\times$ to $2145\times$ across all experiments, degrading gracefully under 1% observation noise. On real data, MEDIDA outperformed naive SIR across the large majority of countries, recovered time-varying transmission dynamics corresponding to documented policy events, and transferred corrections across national borders. A globally-trained model resolved the overfitting observed under single-country training.

The key finding is that structural correction of a model, recovering missing terms from data rather than specifying them in advance, can matter more than the choice of which model variant to use. Future work should address epidemic-specific feature libraries, principled handling of the simplex constraint, a filter suited to count-based observations, and evaluation on diseases beyond COVID-19.

DATA AVAILABILITY

The COVID-19 case data used in this study are openly available from Our World in Data at `https://ourworldindata.org/covid-cases`. All code, scripts, and figure generation pipelines for this study are publicly available at `https://github.com/Max-Ratcliff/AM170B-Medida-Epidemiology`.

REFERENCES

Mojgani, A., Chattopadhyay, A., & Hassanzadeh, P. (2022). Discovery of interpretable structural model errors by combining Bayesian sparse regression and data assimilation: A chaotic Kuramoto–Sivashinsky test case. *Chaos*, **32**, 061105. https://doi.org/10.1063/5.0091282

Kermack, W. O., & McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A*, **115**(772), 700–721. https://doi.org/10.1098/rspa.1927.0118

Our World in Data. (2021). COVID-19 dataset. https://ourworldindata.org/covid-cases.

Brunton, S. L., Proctor, J. L., & Kutz, J. N. (2016). Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, **113**(15), 3932–3937. https://doi.org/10.1073/pnas.1517384113

Chen, R. T. Q., Rubanova, Y., Bettencourt, J., & Duvenaud, D. (2018). Neural ordinary differential equations. In *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 31, pp. 6571–6583. https://arxiv.org/abs/1806.07366

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

Evensen, G. (2009). *Data Assimilation: The Ensemble Kalman Filter*, 2nd ed. Springer, Berlin. https://doi.org/10.1007/978-3-642-03711-5

Tipping, M. E. (2001). Sparse Bayesian learning and the relevance vector machine. *Journal of Machine Learning Research*, **1**, 211–244. https://jmlr.org/papers/v1/tipping01a.html

Rackauckas, C., Ma, Y., Martensen, J., Warner, C., Zubov, K., Supekar, R., Skinner, D., Ramadhan, A., & Edelman, A. (2020). Universal differential equations for scientific machine learning. *arXiv preprint arXiv:2001.04385*. https://arxiv.org/abs/2001.04385

Shaman, J., & Karspeck, A. (2012). Forecasting seasonal outbreaks of influenza. *Proceedings of the National Academy of Sciences*, **109**(50), 20425–20430. https://doi.org/10.1073/pnas.1208772109