Epidemiological digital twin to anticipate dengue outbreaks in Central American intermediary cities

Jorge Valle Reconco[1]; Ana Lourdes Cardona[2]; Edil Argueta[3]
[1,2]Universidad Nacional Autónoma de Honduras, Honduras, *jorge_valle@unah.edu.hn, ana.cardona@unah.edu.hn*
[3]Centro Médico Nacional 20 de Noviembre, México, *edilarguetahn@gmail.com*

Abstract*– To propose an epidemiological digital twin architecture to anticipate dengue outbreaks in Central American intermediary cities, based on evidence on epidemiological modeling, available open data, and regional digital surveillance experiences. Structured narrative review, evidence-based conceptual model and computational proof of concept. PubMed, Scopus and IEEE Xplore, the PAHO PLISA platform, the OpenDengue dataset and literature on digital twins and artificial intelligence in public health were consulted. Seven open data sources were evaluated according to accessibility, Central American coverage, and feasibility of integration. Four essential components of the digital twin were identified: integration of epidemiological, entomological and climatic data; spatially stratified SEIR models; machine learning modules for four- to eight-week prediction; and data assimilation using ensemble Kalman filters. In a synthetic retrospective proof of concept for La Ceiba, Honduras (2014–2024), derived from open national surveillance series, a hybrid ensemble of SEIR–EnKF and gradient boosting reduced the four-week-ahead mean absolute error by 18% relative to a persistence baseline. A three-phase implementation framework adapted to limited institutional capacities is proposed, integrating community-based entomological surveillance and local health committees into the decision flow. The epidemiological digital twin is technically feasible for anticipating dengue outbreaks in Central America and can strengthen regional early warning systems with data sources already available.

***Keywords*-- digital twin, dengue, computational epidemiology, intermediary cities, Central America.**

Gemelo digital epidemiológico para anticipar brotes de dengue en ciudades intermedias Centroamericanas

Jorge Valle Reconco[1]; Ana Lourdes Cardona[2]; Edil Argueta[3]
[1,2]Universidad Nacional Autónoma de Honduras, Honduras, *jorge_valle@unah.edu.hn, ana.cardona@unah.edu.hn*
[3]Centro Médico Nacional 20 de Noviembre, México, *edilarguetahn@gmail.com*

Resumen– ***Proponer una arquitectura de gemelo digital epidemiológico para anticipar brotes de dengue en ciudades intermedias centroamericanas, sustentada en evidencia sobre modelado epidemiológico, datos abiertos disponibles y experiencias regionales de vigilancia digital. Revisión narrativa estructurada, modelo conceptual basado en evidencia y prueba de concepto computacional. Se consultaron PubMed, Scopus e IEEE Xplore, la plataforma PAHO PLISA, el dataset OpenDengue y literatura sobre gemelos digitales e inteligencia artificial en salud pública. Se evaluaron siete fuentes de datos abiertos según accesibilidad, cobertura centroamericana y viabilidad de integración. Se identificaron cuatro componentes esenciales del gemelo digital: integración de datos epidemiológicos, entomológicos y climáticos; modelos SEIR estratificados espacialmente; módulos de aprendizaje automático para predicción de cuatro a ocho semanas; y asimilación de datos mediante filtros de Kalman de conjuntos. En una prueba de concepto retrospectiva sintética para La Ceiba, Honduras (2014-2024), derivada de series nacionales abiertas, un ensamble híbrido SEIR-EnKF y gradient boosting redujo en 18 % el error absoluto medio a cuatro semanas frente a un modelo de persistencia. Se propone un marco trifásico de implementación adaptado a capacidades institucionales limitadas, que incorpora la vigilancia entomológica comunitaria y los comités locales de salud al flujo de decisión. El gemelo digital epidemiológico es técnicamente viable para anticipar brotes de dengue en Centroamérica y puede fortalecer los sistemas de alerta temprana regional con fuentes de datos ya disponibles.***

Palabras clave-- ***gemelo digital, dengue, epidemiologia computacional, ciudades intermedias, Centroamérica***

I. Introducción

El dengue constituye una de las cargas epidemiológicas más persistentes en América Latina y el Caribe. Las estimaciones globales sitúan entre 100 y 400 millones el número de infecciones virales anuales en el mundo, con la región de las Américas concentrando una fracción significativa de los casos notificados [1], [2]. En Centroamérica, la enfermedad ha dejado de comportarse como un fenómeno estrictamente estacional para adquirir un patrón endemoepidémico. Los brotes recurrentes superan la capacidad de respuesta de los sistemas nacionales de salud en múltiples temporadas epidemiológicas. Honduras, Guatemala, El Salvador y Nicaragua han registrado cifras elevadas de casos confirmados y probables a lo largo del periodo 2015-2024. Los picos coinciden con lluvias intensas, crecimiento urbano acelerado y condiciones de infraestructura deficitaria [3], [4].

Las ciudades intermedias, generalmente definidas como aquellas con entre 50,000 y 500,000 habitantes, presentan condiciones estructurales que amplían la transmisión del virus dengue. La densidad poblacional creciente, las deficiencias en el manejo de residuos sólidos y la intermitencia en el suministro de agua potable generan criaderos del mosquito *Aedes aegypti* tanto en entornos domésticos como peridomésticos. A diferencia de las capitales nacionales, estas ciudades disponen de menor capacidad institucional y tecnológica para implementar sistemas de alerta temprana sofisticados. Esto prolonga el intervalo entre la detección de los primeros casos y la activación de medidas coordinadas de control vectorial [5], [6].

Los sistemas tradicionales para vigilar el dengue, que se basan en la notificación de casos de fiebre y en la revisión regular de la presencia de larvas en casas, tienen retrasos de entre 7 y 21 días desde el inicio real del brote [6]. Esta demora reduce la utilidad de la información para decisiones preventivas. Los modelos compartimentales SEIR han ayudado a entender mejor cómo se transmite una enfermedad, pero a menudo suponen que la población es uniforme y no utilizan información actual del terreno [7], [8].

En este contexto, el concepto de gemelo digital surge como una alternativa con fundamento tecnológico para transformar la vigilancia epidemiológica reactiva en una capacidad de anticipación activa. El término se acuñó en el ámbito de la ingeniería de sistemas para designar una réplica virtual de un objeto físico que se actualiza continuamente con datos reales provenientes del objeto original [9]. En los últimos diez años, su aplicación en el campo de la salud pública ha empezado a aparecer en la literatura científica. Se utiliza tanto en el tratamiento de pacientes individuales como en estudios de poblaciones para entender mejor cómo funcionan los sistemas epidemiológicos complejos [10], [11].

La mayoría de las iniciativas sobre gemelos digitales en salud se originan en lugares con altos ingresos que tienen buena infraestructura de datos. Esto hace difícil aplicar directamente estas iniciativas en ciudades intermedias de Centroamérica, donde los recursos institucionales y tecnológicos son limitados. Esta brecha metodológica constituye la justificación principal del presente estudio. El objetivo del artículo es proponer una arquitectura conceptual de gemelo digital epidemiológico para anticipar brotes de

dengue en ciudades intermedias centroamericanas. Esta arquitectura está sustentada en evidencia sobre vigilancia epidemiológica, modelado computacional y datos abiertos disponibles en la región. Su viabilidad técnica se examina, además, mediante una prueba de concepto retrospectiva en una ciudad intermedia hondureña.

II. Materiales y métodos

A. Tipo de estudio y justificación del diseño

Se realizó una revisión narrativa estructurada complementada con una propuesta de modelo conceptual basada en evidencia. Este diseño se eligió porque no existen datasets abiertos de gemelos digitales aplicados al dengue en ciudades intermedias centroamericanas que permitan ejecutar un análisis secundario con resultados empíricos propios. Sin embargo, las fuentes de datos epidemiológicos abiertos relevantes (OpenDengue, PAHO PLISA, ERA5), la evidencia acumulada sobre modelado computacional de dengue y la literatura técnica sobre gemelos digitales en salud pública proporcionan la base suficiente para construir una propuesta metodológica sólida, reproducible y con aplicación directa en el contexto regional. Para explorar la viabilidad técnica de los componentes centrales se añadió una prueba de concepto computacional retrospectiva (Sección II-G).

B. Estrategia de búsqueda bibliográfica

Se realizaron búsquedas sistematizadas en PubMed, Scopus e IEEE Xplore entre enero y mayo de 2025. Las cadenas de búsqueda utilizadas fueron: ("digital twin" AND "epidemiology"), ("digital twin" AND "dengue"), ("dengue" AND "early warning" AND "machine learning"), ("dengue" AND "SEIR" AND "Latin America"), ("agent-based model" AND "dengue") y ("dengue surveillance" AND "Central America"), más sus equivalentes en español La búsqueda inicial recuperó 347 registros únicos Tras aplicar criterios de inclusión y exclusión, se seleccionaron 48 textos para lectura completa, de los cuales 23 fueron incorporados como referencias del artículo Los demás fueron excluidos por carecer de metodología replicable, por tratarse de editoriales o cartas sin contenido empírico original, o por no ser accesibles en texto completo. Se incluyeron adicionalmente reportes técnicos de la OPS, la OMS y documentación oficial de bases de datos epidemiológicas, así como literatura específica sobre vigilancia comunitaria y ciencia ciudadana.

C. Evaluación de fuentes de datos abiertas

Se evaluaron siete fuentes de datos abiertos en función de cinco criterios: accesibilidad real (licencia y portal de descarga activo), granularidad geográfica y temporal, cobertura centroamericana, viabilidad de integración técnica y pertinencia para los componentes del gemelo digital propuesto. Los resultados de esta evaluación se presentan en la Tabla I. OpenDengue [5], disponible bajo licencia CC0 en opendengue.org, fue seleccionado como fuente de datos históricos de incidencia a nivel de país y semana epidemiológica La plataforma PAHO PLISA [4] fue seleccionada para el análisis de tendencias regionales. Los datos climáticos ERA5 [12] del Servicio de Cambio Climático de Copernicus (ECMWF) se identificaron como la fuente más adecuada para el componente climático del modelo, dada su cobertura global, acceso libre y resolución temporal horaria. Las bases MIMIC-IV y eICU fueron excluidas porque corresponden a datos clínicos hospitalarios de instituciones de EE. UU., sin representatividad para el dengue en Centroamérica y sin pertinencia temática para el objetivo del estudio.

D. Criterios de inclusión y exclusión

Se incluyeron estudios con: (a) propuestas o evaluaciones de modelos epidemiológicos computacionales para dengue; (b) revisiones de sistemas de alerta temprana en enfermedades vectoriales; (c) arquitecturas de gemelos digitales en salud pública; (d) análisis de datos de dengue en Centroamérica o América Latina; y (e) evaluaciones de fuentes de datos abiertos para vigilancia epidemiológica Se excluyeron estudios sin metodología replicable, publicaciones con acceso restringido, editoriales sin datos originales y referencias sin relación directa con los objetivos del estudio.

E. Consideraciones éticas y de privacidad

El artículo utiliza exclusivamente datos epidemiológicos agregados de notificación oficial y datasets de acceso público sin información de identificación personal. No se accedió a datos individuales de pacientes y no se requiere aprobación ética para el diseño de revisión narrativa y propuesta metodológica Las fuentes de datos abiertas empleadas no imponen restricciones adicionales de uso para fines académicos Las bases de datos con acceso controlado (MIMIC-IV, eICU) fueron evaluadas exclusivamente a partir de su documentación pública y fueron excluidas del estudio sin haber solicitado “Data Use Agreement”, dado que no resultaron pertinentes.

F. Limitaciones metodológicas

El diseño de revisión narrativa no permite cuantificar el tamaño del efecto de ninguna intervención ni realizar metaanálisis. La cobertura de la literatura puede ser incompleta para publicaciones en idiomas distintos del inglés y el español. Además, la propuesta arquitectónica no ha sido implementada ni validada prospectivamente en ninguna ciudad centroamericana, y la prueba de concepto emplea una serie urbana sintética que no sustituye la validación con registros municipales. Esto constituye la limitación más relevante del estudio. La ausencia de datos entomológicos continuos y georreferenciados en muchas ciudades intermedias de la región constituye el principal vacío de información para futuras implementaciones.

G. Prueba de concepto computacional

Se efectuó una simulación retrospectiva sintética para La Ceiba, Honduras, ciudad intermedia de la costa caribe con cerca de 215,000 habitantes. Ante la falta de series

municipales abiertas, la serie urbana se generó por desagregación binomial de los casos semanales notificados a nivel nacional entre 2014 y 2024 (573 semanas) [4], [5], con probabilidad igual a la fracción poblacional de la ciudad (2.2 %). Sobre ella se implementaron tres componentes de la Capa 2: (i) un modelo SEIR humano acoplado a un modelo SEI vectorial [7], con incubación intrínseca de 5.5 días, periodo infeccioso de 6 días, incubación extrínseca de 10 días, vida media vectorial de 14 días, inmunidad temporal de 18 meses y fracción de notificación de 0.10, cuyo estado y tasa de transmisión se actualizaron cada semana con un EnKF de 300 miembros; (ii) un modelo de gradient boosting con regresión cuantílica (cuantiles 0.05, 0.50 y 0.95) alimentado con ocho rezagos de la incidencia y términos armónicos de la semana epidemiológica; y (iii) un ensamble híbrido que promedia los cuantiles de ambos. En esta etapa no se incorporaron covariables climáticas.

Los modelos se ajustaron con datos de 2014-2021 y se evaluaron fuera de muestra en 2022-2024 (156 semanas), a horizontes de 4 y 8 semanas, frente a un modelo de persistencia. Se calcularon el error absoluto medio (MAE), la raíz del error cuadrático medio (RMSE) y la cobertura empírica del intervalo predictivo al 90 %. Como indicador operativo se estimaron la sensibilidad y la especificidad de la alerta emitida cuando la mediana proyectada superaba el tercer cuartil del canal endémico 2014-2021, excluido el año epidémico 2019. El análisis se programó en Python con bibliotecas de código abierto.

III. Resultados

A. Carga epidemiológica del dengue en ciudades intermedias centroamericanas

Los datos del repositorio OpenDengue y la plataforma PAHO PLISA confirman que Centroamérica experimentó brotes de dengue de magnitud creciente en el periodo 2015-2024 [4], [5]. Honduras figura de forma consistente entre los países con las tasas de incidencia más elevadas de la región en ese periodo, con episodios de transmisión intensa documentados por la Secretaría de Salud en múltiples semanas epidemiológicas [13]. El patrón epidemiológico muestra picos bienales asociados con la circulación de los cuatro serotipos del virus dengue (DENV-1 a DENV-4). Además, hay anomalías climáticas como los fenómenos de El Niño y La Niña. Además, hay un incremento sostenido de la densidad del vector *Aedes aegypti* en zonas urbanas de expansión reciente [2], [14].

La revisión de la literatura identificó tres restricciones principales que limitan la anticipación de brotes en este contexto. La primera es la latencia del sistema de notificación pasiva, que oscila entre 7 y 21 días respecto al inicio real del brote. La segunda es la escasez de datos entomológicos continuos y georreferenciados. La tercera es la ausencia de plataformas integradas que combinen en tiempo cuasirreal los flujos de datos epidemiológicos, climáticos y entomológicos disponibles [6], [15]. Estas limitaciones justifican el uso de alternativas tecnológicas que apoyen la toma de decisiones en salud.

B. Estado del arte en modelado computacional del dengue

La revisión identificó tres familias de modelos utilizados para la proyección de brotes de dengue, cuyas características comparativas se presentan en la Tabla II. Los modelos compartimentales deterministas, especialmente las versiones SEIR con un compartimento para vectores, son el estándar histórico para el modelado del dengue [7]. Permiten estimar el número básico de reproducción (R0), simular el efecto de intervenciones preventivas y proyectar la magnitud potencial de brotes bajo distintos escenarios de transmisión. Su principal limitación radica en la asunción de mezcla homogénea de la población, que ignora la heterogeneidad espacial y social relevante para las ciudades intermedias centroamericanas [15]. Los modelos de aprendizaje automático han ganado protagonismo en la última década para la predicción de corto plazo de casos de dengue. Métodos como LSTM (Long Short-Term Memory), Random Forest y XGBoost, que han sido entrenados con series temporales de casos y datos climáticos, han demostrado ser bastante efectivos para hacer predicciones en ciudades donde hay muchos datos históricos [16], [17]. Su transferibilidad entre ciudades se limita, ya que dependen de las características estadísticas locales de los datos de entrenamiento y requieren recalibración periódica. Los modelos basados en agentes (ABM) muestran las diferencias individuales de personas y mosquitos en un área geográfica específica, lo que ayuda a entender las dinámicas urbanas importantes para un control vectorial más efectivo [18]. Un modelo espacial basado en agentes mostró que la dinámica poblacional de Aedes aegypti en áreas urbanas responde a la heterogeneidad del entorno construido y a las variaciones de temperatura y lluvia, aspectos que los modelos compartimentales homogéneos no representan [18]. Sin embargo, su aplicación práctica en ciudades intermedias centroamericanas está limitada por el alto costo computacional y por la demanda de datos de parametrización detallada que generalmente no está disponible en estos contextos.

C. El concepto de gemelo digital en salud pública

El término gemelo digital se acuñó en el ámbito de la ingeniería de sistemas para referirse a una réplica virtual de un sistema físico que se actualiza continuamente con datos del objeto real mediante sensores y flujos de información [9]. Su aplicación en la salud pública ha empezado a discutirse en la literatura científica. Laubenbacher, Sluka y Glazier (2021) sugirieron el gemelo digital como una herramienta para simular infecciones virales a nivel celular y de órganos, con la idea de que podría ampliarse a sistemas epidemiológicos en poblaciones [10]. Bruynseels et al. (2018) examinaron las implicaciones éticas de esta transición, subrayando la necesidad de marcos de gobernanza de datos sólidos antes de cualquier implementación clínica o de salud pública [11].

La operación de un gemelo digital epidemiológico requiere cuatro elementos fundamentales: fuentes de datos con actualización frecuente, modelos matemáticos parametrizados con información local, mecanismos de asimilación de datos que actualicen el estado del modelo ante cada nueva observación e interfaces de visualización adaptadas para equipos de toma de decisión no especializados en modelado matemático [10], [11].

La asimilación de datos es el componente que distingue al gemelo digital de un modelo estático. El filtro de Kalman de conjuntos (Ensemble Kalman Filter, EnKF) ha sido ampliamente validado en modelado epidemiológico dinámico de enfermedades respiratorias [19]. Permite actualizar la estimación del estado del sistema epidemiológico en tiempo cuasirreal a medida que se incorporan nuevas observaciones, reduciendo progresivamente la incertidumbre del modelo. Su adaptación al contexto del dengue, con series temporales semanales de casos como observación primaria, es técnicamente directa.

D. Arquitectura conceptual del gemelo digital propuesto

Se propone una arquitectura de cuatro capas funcionales para el gemelo digital epidemiológico (Fig. 1). La arquitectura se diseñó con el principio de máximo aprovechamiento de fuentes de datos preexistentes en la región, minimizando los requisitos de infraestructura nueva. Todos los componentes propuestos tienen referentes metodológicos publicados en la literatura revisada.

La capa 1 integra cinco tipos de flujos de datos. El componente epidemiológico corresponde a los casos confirmados y probables notificados al Sistema Nacional de Salud con frecuencia semanal. El componente entomológico incluye los índices de Breteau, de vivienda y de recipiente [19]. Estos se obtienen a través de inspecciones muestrales regulares y, cuando es posible, utilizando ovitrampas económicas. El componente climático utiliza datos de temperatura máxima y mínima, precipitación acumulada semanal y humedad relativa provenientes de ERA5 o de estaciones meteorológicas locales cuando estén disponibles [12]. El componente sociodemográfico integra datos de censos nacionales y sistemas de información geográfica municipales para el cálculo de densidades poblacionales y coberturas de servicios. El quinto componente, de vigilancia comunitaria, se orienta a reducir la brecha entomológica: brigadas escolares, comités de salud y voluntarios gestionan ovitrampas artesanales de bajo costo y registran conteos de huevos, criaderos y casos febriles mediante una aplicación móvil sencilla, georreferenciada y operable sin conexión. La ciencia ciudadana con validación experta ha demostrado ser fiable y escalable para la vigilancia de mosquitos vectores [20], [21]; y la movilización comunitaria basada en evidencia redujo la infección por dengue y los índices vectoriales en Nicaragua y México [22]. Tras el control de calidad, estos reportes ingresan como observaciones adicionales al motor de asimilación.

La capa 2 aloja el núcleo computacional del gemelo digital. El componente determinista es un modelo SEIR dividido por áreas geográficas más pequeñas, que incluye un apartado para los vectores que representa cómo *Aedes aegypti* se comporta según la temperatura y la lluvia de cada semana [7]. El componente predictivo a corto plazo usa redes LSTM o modelos de “gradient boosting” que se entrenan con datos locales de casos y variables climáticas. Esto permite hacer proyecciones a 4 y 8 semanas [16], [17]. El motor de asimilación usa un EnKF adaptado al modelo SEIR, similar al método que emplearon Shaman y Karspeck para predecir brotes de influenza [19]. Este motor actualiza cada semana los parámetros del modelo con la información epidemiológica más reciente.

La Capa 3 genera tres tipos de salida operativa, iniciando con una proyección de incidencia a corto plazo con intervalos de credibilidad al 90%, mapas de riesgo espacial actualizados semanalmente que identifican las unidades geográficas con mayor probabilidad de incremento de casos y simulaciones de escenarios de intervención que estiman el efecto proyectado de acciones de control como fumigación focalizada o eliminación masiva de criaderos [23]. La Capa 4 muestra los resultados en un panel de visualización diseñado para epidemiólogos y directores de programas de salud pública. Tiene umbrales de alerta que se pueden configurar y genera boletines semanales de manera automática. Además, cada alerta se traduce en un protocolo de acción territorial dirigido a los comités locales de salud y a las organizaciones comunitarias de base, con tareas definidas por barrio (eliminación de criaderos, visitas domiciliarias y comunicación de riesgo) cuyo cumplimiento retorna al sistema.

E. Fuentes de datos abiertas consideradas

La Tabla I resume la evaluación de las fuentes potencialmente integrables al gemelo digital según tipo de información, dominio, modalidad de acceso, uso previsto, utilidad y limitaciones operativas.

F. Comparación de enfoques de modelado epidemiológico

La Tabla II compara los enfoques de modelado aplicables al dengue según sus fundamentos, fortalezas, limitaciones en ciudades intermedias y función dentro de la arquitectura propuesta.

Esta comparación permite justificar la integración complementaria de enfoques mecanísticos, predictivos y adaptativos para representar la dinámica de transmisión, anticipar brotes y actualizar de forma continua las estimaciones epidemiológicas. La Fig. 2 presenta, en el panel izquierdo, el diagrama de flujo del modelo SEIR con compartimento vectorial que constituye el núcleo determinista del gemelo digital propuesto. Los compartimentos humanos (S, E, I, R) interactúan con los compartimentos vectoriales (Sv, Ev, Iv) mediante tasas de transmisión bidireccionales moduladas por temperatura y precipitación (ERA5). En el panel derecho, el gráfico radar compara los cuatro enfoques de

modelado en seis dimensiones: interpretabilidad biológica, precisión predictiva de corto plazo, heterogeneidad espacial, bajo costo computacional, viabilidad con datos limitados y transferibilidad entre ciudades. El modelo híbrido propuesto obtiene las puntuaciones más equilibradas en todas las dimensiones.

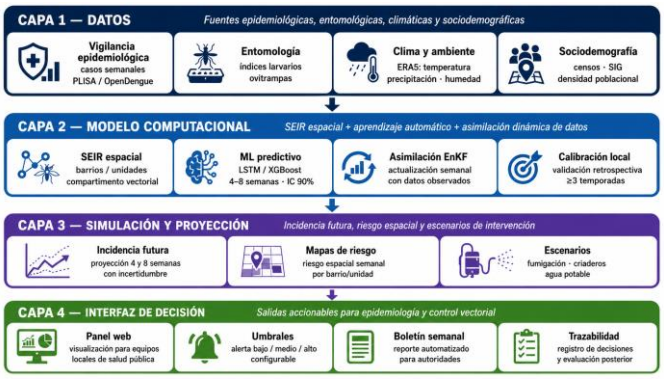


Fig. 1 Arquitectura de cuatro capas del gemelo digital epidemiológico propuesto para la anticipación de brotes de dengue en ciudades intermedias centroamericanas. Los flujos de información son ascendentes desde la adquisición de datos hasta la interfaz de decisión. Las franjas de vigilancia comunitaria (Capa 1) y de acción territorial (Capa 4) articulan la participación social con el flujo tecnológico. ERA5 = European Centre for Medium-Range Weather Forecasts Reanalysis v5; EnKF = Ensemble Kalman Filter; LSTM = Long Short-Term Memory; IC = Intervalo de credibilidad; SIG = Sistema de Información Geográfica.

TABLA I
FUENTES DE DATOS ABIERTAS CONSIDERADAS PARA EL GEMELO DIGITAL EPIDEMIOLÓGICO

Fuente	Tipo	Dominio	Acceso	Uso	Util.	Limitaciones
Open Dengue [5]	Casos de dengue agregados por país y semana epidemiológica	Epidemiología	Libre (CC0)	Parametrización del modelo SEIR, validación retrospectiva	Alta	Cobertura subnacional limitada para ciudades intermedias de C.A.
PAHO PLISA [4]	Indicadores históricos de dengue y arbovirosis por país en las Américas	Salud pública	Libre	Tendencias regionales; calibración inicial del modelo	Alta	Datos agregados a nivel país; sin georeferenciación intraurbana
ERA5 [12]	Temperatura, precipitación y humedad; resolución horaria, ~31 km espacial	Climatología	Libre	Capa climática del gemelo digital; correlación con densidad vectorial	Alta	Resolución espacial insuficiente para análisis intraurbano detallado
Vector Base GBIF [24]	Registros de ocurrencia georeferenciada de *Aedes aegypti*	Entomología	Libre	Modelado de distribución vectorial; estimación de riesgo espacial	Media	Datos de presencia, no de abundancia; cobertura desigual en C.A.
IHME GBD 2019 [25]	Estimaciones de carga de dengue: DALYs,	Epidemiología global	Libre	Estimación del impacto potencial de intervenciones	Media	Intervalos de incertidumbre amplios en países con subregistro elevado

Nota: Evaluación de fuentes de datos abiertas según accesibilidad, cobertura centroamericana y utilidad para los componentes del gemelo digital propuesto. MIMIC-IV y eICU-CRD, evaluadas y excluidas por falta de pertinencia (Sección II-C), no se muestran. C.A. = Centroamérica.

TABLA II
ENFOQUES DE MODELADO EPIDEMIOLÓGICO COMPUTACIONAL PARA DENGUE Y SU FUNCIÓN EN EL GEMELO DIGITAL PROPUESTO.

Enfoque	Descripción	Fortalezas	Limitaciones	Rol – Gemelo Digital
Modelos compartimentales (SEIR/SEIRS)	Ecuaciones diferenciales ordinarias que dividen humanos y vectores en compartimentos epidemiológicos	Base teórica sólida, interpretable, bajo costo computacional	Mezcla homogénea; difícil calibración sin datos entomológicos continuos	Núcleo determinista; estima R0, proyecta incidencia de largo plazo
Aprendizaje automático (LSTM, XGBoost)	Algoritmos entrenados sobre series temporales de casos y variables climáticas para predicción de corto plazo	Alta capacidad predictiva; captura no linealidades climatoepidemiológicas	Baja transferibilidad entre ciudades; requiere datos históricos abundantes	Módulo de predicción de 4 a 8 semanas con intervalos de credibilidad
Modelos de agentes (ABM)	Simulación de interacciones entre humanos y mosquitos en un espacio geográfico explícito	Captura heterogeneidad espacial y comportamental intraurbana	Alto costo computacional; requiere parametrización detallada no disponible localmente	Prototipo exploratorio para diseño de intervenciones espacialmente focalizadas
Asimilación de datos (EnKF)	Filtro de Kalman de conjuntos que actualiza los parámetros del modelo con nuevas observaciones epidemiológicas	Reduce incertidumbre progresivamente; permite operación en tiempo cuasirreal	Sensible a la calidad y regularidad del dato de entrada; requiere flujo semanal estable	Motor de actualización dinámica; se ejecuta con cada reporte epidemiológico semanal

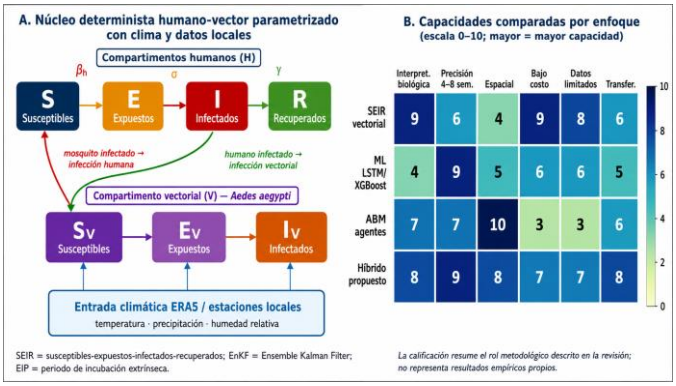


Fig. 2 A) Diagrama compartimental del modelo SEIR con compartimento vectorial para dengue. Los parámetros climáticos como temperatura y precipitación provenientes de ERA5 modulan la tasa de generación vectorial, la supervivencia del vector y el período extrínseco de incubación. B) Comparación de capacidades de cuatro enfoques de modelado en escala 0–10. Los valores de puntuación son cualitativos, derivados de la síntesis bibliográfica del estudio [7], [16]-[19]. Bajo costo computacional: 10 = muy bajo costo, 0 = muy alto costo. ABM = Agent-Based Model; ML = Machine Learning; EnKF = Ensemble Kalman Filter; SEIR = Susceptibles–Expuestos–Infectados–Recuperados.

G. *Marco de implementación por fases*

Dadas las restricciones institucionales y de infraestructura propias de las ciudades intermedias centroamericanas, se

propone un marco de implementación en tres fases progresivas y acumulativas. Cada fase produce entregables concretos que son insumo de la siguiente.

La Fig. 3 presenta el diagrama de Gantt del marco trifásico de implementación. La Fase 1 (meses 1–12) cubre la estandarización de flujos de notificación, la integración de datos históricos y el desarrollo del modelo SEIR base. La Fase 2 (meses 13–24) integra los módulos de aprendizaje automático y asimilación de datos y desarrolla el panel de visualización. La Fase 3 (mes 25 en adelante) opera el sistema en tiempo rutinario y expande la cobertura a ciudades adicionales.

H. Modelo conceptual de actores y flujos de información

La Fig. 4 presenta el modelo conceptual de actores institucionales y comunitarios, fuentes de datos y flujos de información del gemelo digital propuesto, con el objetivo de explicitar las relaciones de gobernanza necesarias para su sostenibilidad a largo plazo. Los productores de datos (SESAL, municipalidades, servicios meteorológicos, academia y redes comunitarias de vigilancia) alimentan una plataforma que estandariza y controla la calidad de los flujos semanales, y el núcleo computacional genera cinco productos operativos. La gobernanza incorpora a los comités locales de salud y a las organizaciones comunitarias de base como actores con capacidad de decisión y no solo como receptores de boletines: participan en la definición de umbrales, reciben las alertas por barrio, ejecutan las acciones preventivas junto con los equipos de control vectorial y reportan su cumplimiento y eficacia, información que retroalimenta los parámetros del modelo y los umbrales de alerta.

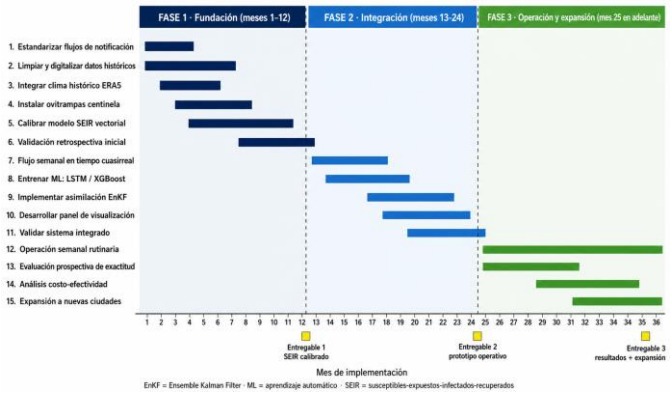


Fig. 3 Marco trifásico de implementación del gemelo digital epidemiológico para ciudades intermedias centroamericanas. Los rombos en color dorado corresponden a entregables formales de cada fase. Las fases son acumulativas: cada etapa incorpora los resultados de la anterior antes de avanzar.

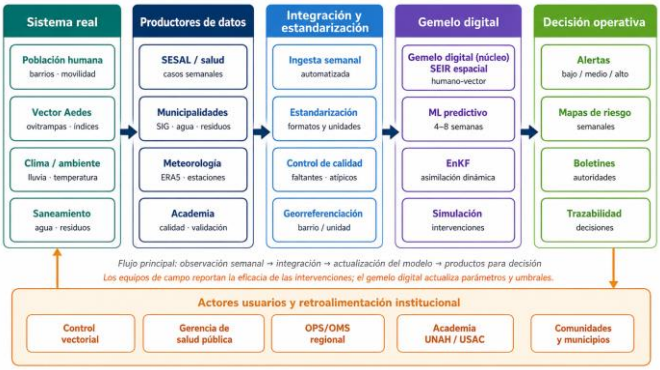


Fig. 4 Flujo (pipeline) de integración de datos y modelo de actores institucionales y comunitarios del gemelo digital epidemiológico. Los recuadros resaltados corresponden a los componentes de vigilancia y acción comunitaria. EnKF = Ensemble Kalman Filter; SIG = Sistema de Información Geográfica; OPS = Organización Panamericana de la Salud; UNAH = Universidad Nacional Autónoma de Honduras; USAC = Universidad de San Carlos de Guatemala.

I. Prueba de concepto retrospectiva en La Ceiba

La Tabla III resume el desempeño fuera de muestra. A cuatro semanas, el ensamble híbrido obtuvo el menor error (MAE de 11.4 y RMSE de 20.0 casos por semana), con reducciones de 18 % y 25 % respecto a la persistencia; a ocho semanas, las reducciones fueron de 14 % y 17 %. El SEIR-EnKF aislado reprodujo la forma de las ondas epidémicas, pero sobrestimó la magnitud del pico de 2024, sesgo que el componente estadístico atenuó dentro del ensamble (Fig. 5). La cobertura de los intervalos al 90 % fue inferior a la nominal (74 % a cuatro semanas y 62 % a ocho), lo que obliga a recalibrar la dispersión antes de cualquier uso operativo. Como herramienta de alerta, el híbrido a cuatro semanas clasificó correctamente el 79 % de las semanas por encima del umbral, con una especificidad del 86 %, y anticipó en una y tres semanas el cruce sostenido del umbral en las temporadas de 2023 y 2024, respectivamente.

TABLA III
DESEMPEÑO FUERA DE MUESTRA (2022-2024) DE LOS COMPONENTES PREDICTIVOS.

Modelo	MAE 4 s	RMSE 4 s	Cob. 4 s	MAE 8 s	RMSE 8 s	Cob. 8 s
Persistencia	13.9	26.7	87.8	23.6	43.7	84.0
SEIR-EnKF	17.3	33.4	73.1	28.1	55.6	55.8
GBM	13.1	25.3	64.7	21.6	42.8	65.4
Híbrido	11.4	20.0	74.4	20.4	36.1	62.2

Nota: MAE y RMSE en casos por semana; Cob. = cobertura empírica del intervalo predictivo al 90 % (%); s = semanas; GBM = gradient boosting cuantílico. Serie urbana sintética derivada de datos nacionales abiertos [5].

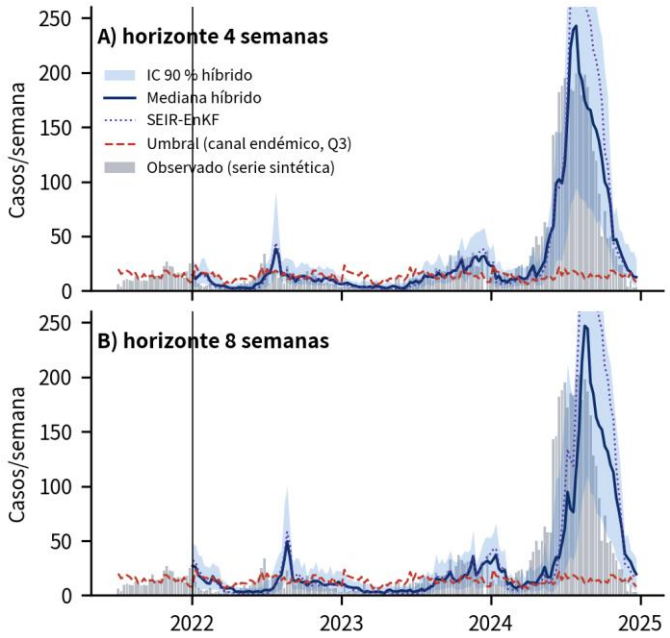


Fig. 5 Pronósticos fuera de muestra para la serie sintética de La Ceiba a 4 (A) y 8 (B) semanas. Barras: casos semanales; franja: intervalo predictivo al 90 % del ensamble híbrido; línea continua: mediana del híbrido; línea punteada: mediana del SEIR-EnKF; línea discontinua: umbral del canal endémico. La línea vertical marca el inicio del periodo de evaluación (enero de 2022).

IV. Discusión

A. Potencial del gemelo digital para la vigilancia proactiva en Centroamérica

La propuesta que se presenta atiende a una necesidad documentada: los sistemas de vigilancia epidemiológica del dengue en Centroamérica funcionan principalmente en modo reactivo, con latencias en la notificación que no permiten poner en marcha a tiempo las medidas de control [6]. El gemelo digital no sustituye los sistemas de notificación que ya existen, sino que funciona como una herramienta de análisis que convierte los datos de vigilancia diaria en proyecciones epidemiológicas a corto plazo que tienen un valor práctico inmediato.

Esta distinción conceptual es importante para evitar que la propuesta se perciba como una duplicación de sistemas existentes: el gemelo digital amplifica la vigilancia establecida, no la sustituye. La evidencia revisada muestra que los métodos híbridos, que mezclan modelos compartimentales con aprendizaje automático, tienen un mejor rendimiento predictivo que los métodos únicos en situaciones con series temporales irregulares o incompletas [16], [17].

Esta combinación se adecua para ciudades intermedias centroamericanas, donde la calidad y continuidad de los datos epidemiológicos varía entre temporadas y entre municipios. El módulo de asimilación de datos con EnKF incluye un sistema de actualización constante que disminuye la incertidumbre del modelo a medida que se obtienen nuevas observaciones, algo que los modelos estáticos no pueden hacer [19].

B. Adaptaciones necesarias para ciudades intermedias

La traslación del concepto de gemelo digital, desarrollado en contextos de altos ingresos con infraestructura de datos robusta [10], [11], al ámbito de ciudades intermedias centroamericanas requiere al menos cuatro ajustes críticos; Primero, simplificar los requisitos de datos de entrada para que el sistema pueda operar con notificaciones semanales y monitoreo entomológico muestral. Segundo, incorporar mecanismos de manejo de datos faltantes y valores atípicos, frecuentes en sistemas de notificación con cobertura incompleta.

Tercero, se deben diseñar interfaces de usuario accesibles para equipos de epidemiología con diferentes niveles de formación técnica en modelado matemático. Cuarto, se deben establecer marcos de gobernanza de datos que garanticen la privacidad de la información de los casos, la trazabilidad de las proyecciones y la sostenibilidad institucional del sistema más allá de los ciclos de financiamiento de proyectos.

C. Disponibilidad de datos y viabilidad técnica

La evaluación de fuentes confirmó que la combinación de OpenDengue [5], PAHO PLISA [4] y ERA5 [12] proporciona insumos suficientes para la parametrización inicial del modelo y la validación retrospectiva con datos históricos. La prueba de concepto confirmó que esta combinación es ejecutable con herramientas de código abierto y que el ensamble híbrido supera a la persistencia a cuatro y ocho semanas; sin embargo, la subcobertura de los intervalos y la sobrestimación del pico de 2024 indican que la calibración con series municipales reales será determinante. La disponibilidad de estas fuentes bajo licencias abiertas es un factor favorecedor de la viabilidad técnica de la propuesta. Esto reduce la barrera de entrada para implementaciones en contextos con presupuesto limitado. La principal limitación sigue siendo la escasez de datos entomológicos en la mayoría de las ciudades intermedias de la región. Esto genera más incertidumbre en la estimación de la densidad de los vectores y, por lo tanto, en las previsiones de transmisión. La vigilancia comunitaria incorporada en la Capa 1 ofrece una vía de bajo costo para atenuar esta brecha, siempre que se acompañe de protocolos de validación y de retroalimentación a quienes reportan [21], [22].

La experiencia del sistema Infodengue en Brasil [26], que utiliza información sobre la temperatura, reportes de casos y otros datos para crear alertas semanales en ciudades brasileñas, demuestra que se pueden manejar sistemas de vigilancia digital utilizando datos habituales sin necesidad de una infraestructura de sensores muy densa. Esta experiencia es el modelo regional más parecido a la propuesta, pero las diferencias en la capacidad de las instituciones, la cantidad de datos históricos y el desarrollo de los sistemas de información de salud necesitan de adaptaciones importantes para poder aplicarse en Centroamérica.

D. Simulación de escenarios de intervención y utilidad para política pública

El componente de simulación de escenarios de intervención de la Capa 3 de la arquitectura propuesta impacta directamente en la asignación de recursos para el control de vectores. La disponibilidad de mapas de riesgo espacial actualizados semanalmente permitiría focalizar las actividades de fumigación y eliminación de criaderos en las áreas con mayor probabilidad de transmisión activa, optimizando el uso de recursos habitualmente limitados.

La habilidad de simular cómo diferentes intervenciones podrían funcionar antes de ponerlas en práctica es muy importante en situaciones donde el gasto en control de vectores ocupa una gran parte del presupuesto de salud pública [23]. La generación de boletines automatizados con proyecciones de incidencia podría además apoyar la justificación presupuestaria de medidas preventivas ante autoridades municipales y nacionales de salud.

El valor de estas salidas depende de su apropiación territorial. En el ensayo Camino Verde, la movilización comunitaria basada en la discusión local de la evidencia añadió efectividad al control vectorial convencional [22]. Por ello, el flujo propuesto sitúa a los comités locales de salud y a las organizaciones comunitarias de base dentro del ciclo de decisión, de modo que cada alerta se convierta en tareas verificables en el barrio y no quede limitada a un boletín administrativo.

E. Limitaciones del estudio y líneas futuras de investigación

Las limitaciones principales son: (a) la arquitectura propuesta no ha sido implementada ni validada prospectivamente en ninguna ciudad centroamericana. La prueba de concepto utilizó una serie urbana sintética derivada de datos nacionales, que reproduce la dinámica temporal del país pero no la heterogeneidad local, y no incluyó forzamiento climático. (b) El diseño de revisión narrativa no permite cuantificar el tamaño del efecto esperado de la intervención ni establece una comparación cuantitativa con sistemas alternativos. Y (c) Los requisitos de infraestructura tecnológica e institucional dependen de factores políticos y presupuestarios que varían entre países y entre municipios dentro de un mismo país.

Las líneas de investigación prioritarias incluyen la implementación piloto en al menos una ciudad hondureña o centroamericana de tamaño intermedio con datos históricos disponibles. También incluyen la evaluación prospectiva de la exactitud predictiva del sistema con métricas estándar (RMSE, MAE y cobertura de los intervalos de credibilidad al 90%). Además, se incluye el análisis de costo-efectividad comparado con los sistemas de vigilancia convencionales. Otra línea es la evaluación de la calidad, cobertura y sostenibilidad de los reportes comunitarios. Finalmente, se incluye la exploración de la transferibilidad del diseño a otras arbovirosis de relevancia regional como zika y chikungunya.

V. CONCLUSIONES

El dengue sigue siendo una carga epidemiológica constante en las ciudades intermedias de Centroamérica, donde los sistemas de vigilancia trabajan, sobre todo, en modo reactivo. El gemelo digital epidemiológico presentado en este artículo propone una forma viable de mejorar la vigilancia en salud. Esto se puede lograr mediante la combinación de datos que ya existen en la región, sin necesidad de inversiones en infraestructura costosa.

La arquitectura de cuatro capas combina modelos SEIR organizados por área, módulos de aprendizaje automático para hacer predicciones a corto plazo y un sistema de asimilación de datos usando EnKF. Esto se basa en la evidencia sobre el modelado epidemiológico por computadora y utiliza fuentes abiertas confiables (OpenDengue, PAHO PLISA, ERA5). El marco trifásico de implementación se diseñó para ser compatible con las limitaciones de capacidad institucional propias de ciudades intermedias con recursos restringidos. La prueba de concepto retrospectiva respalda la viabilidad técnica del acoplamiento SEIR-EnKF con aprendizaje automático, y la integración de la vigilancia comunitaria y de los comités locales de salud orienta el sistema hacia la acción preventiva territorial.

Esta propuesta es pertinente como ejemplo de aplicación de tecnologías innovadoras al problema de las enfermedades vectoriales en ciudades y regiones de América Latina con alta vulnerabilidad epidemiológica. El paso siguiente necesario es la implementación piloto en al menos una ciudad centroamericana, con evaluación prospectiva rigurosa de su desempeño predictivo y de sus efectos sobre la capacidad de respuesta temprana de los programas de control del dengue. La posibilidad de aplicar el modelo a otras arbovirosis y su posible expansión a redes de vigilancia entre municipios son áreas de investigación muy importantes para la salud pública de la región.

AGRADECIMIENTOS

Agradecemos el apoyo recibido del Instituto de Investigación en Ciencias Médicas y Derecho a la Salud (ICIMEDES) de la Facultad de Ciencias Médicas (FCM) y la Dirección de Investigaciones Científicas, Humanísticas y Tecnológicas (DICIHT) todas ellas de la Universidad Nacional Autónoma de Honduras (UNAH).

REFERENCIAS

[1] S. Bhatt et al., “The global distribution and burden of dengue”, *Nature*, vol. 496, no. 7446, pp. 504–507, Apr. 2013, doi: 10.1038/nature12060.

[2] J. P. Messina et al., “The current and future global distribution and population at risk of dengue,” *Nat. Microbiol.*, vol. 4, no. 9, pp. 1508–1515, Sep. 2019, doi: 10.1038/s41564-019-0476-8.

[3] World Health Organization, “Dengue and severe dengue,” Fact sheet, Aug. 21, 2025. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue. Accessed: May 2, 2026.

[4] Pan American Health Organization, “PLISA: Plataforma de Información en Salud para las Américas,” PAHO/EIH Open Data. [Online]. Available: https://opendata.paho.org/es/explorar. Accessed: May 2, 2026.

[5] J. Clarke, A. Lim, P. Gupte, D. M. Pigott, W. G. van Panhuis, and O. J. Brady, "A global dataset of publicly available dengue case count data", *Sci. Data*, vol. 11, no. 1, Art. no. 296, Mar. 2024, doi: 10.1038/s41597-024-03120-7.

[6] S. Runge-Ranzinger, O. P. Horstick, M. Marx, and A. Kroeger, "What does dengue disease surveillance contribute to predicting and detecting outbreaks and describing trends?," *Trop. Med. Int. Health*, vol. 13, no. 8, pp. 1022–1041, Aug. 2008, doi: 10.1111/j.1365-3156.2008.02112.x.

[7] L. Esteva and C. Vargas, "Analysis of a dengue disease transmission model," *Math. Biosci.*, vol. 150, no. 2, pp. 131–151, Jun. 1998, doi: 10.1016/S0025-5564(98)10003-2.

[8] D. J. Gubler, "Dengue and dengue hemorrhagic fever," *Clin. Microbiol. Rev.*, vol. 11, no. 3, pp. 480–496, Jul. 1998, doi: 10.1128/CMR.11.3.480.

[9] E. H. Glaessgen and D. S. Stargel, "The digital twin paradigm for future NASA and U.S. Air Force vehicles," in *Proc. 53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conf.*, Honolulu, HI, USA, Apr. 2012, Art. no. AIAA 2012-1818, doi: 10.2514/6.2012-1818.

[10] R. Laubenbacher, J. P. Sluka, and J. A. Glazier, "Using digital twins in viral infection," *Science*, vol. 371, no. 6534, pp. 1105–1106, Mar. 2021, doi: 10.1126/science.abf3370.

[11] K. Bruynseels, F. Santoni de Sio, and J. van den Hoven, "Digital twins in health care: Ethical implications of an emerging engineering paradigm," *Front. Genet.*, vol. 9, Art. no. 31, Feb. 2018, doi: 10.3389/fgene.2018.00031.

[12] H. Hersbach et al., "The ERA5 global reanalysis," Q. J. R. Meteorol. Soc., vol. 146, no. 730, pp. 1999–2049, Jul. 2020, doi: 10.1002/qj.3803

[13] Pan American Health Organization, "Dengue Situation Report: Region of the Americas," PAHO, Dec. 14, 2023. [Online]. Available: https://www.paho.org/sites/default/files/2023-12/denguesitrep14dec2023finalv3.pdf. Accessed: May 2, 2026.

[14] Pan American Health Organization, Dengue: Guidelines for Patient Care in the Region of the Americas, 2nd ed. Washington, DC, USA: PAHO, 2016. [Online]. Available: https://iris.paho.org/handle/10665.2/28232. Accessed: May 2, 2026.

[15] P. M. Luz, C. T. Codeço, E. Massad, and C. J. Struchiner, "Uncertainties regarding dengue modeling in Rio de Janeiro, Brazil," *Mem. Inst. Oswaldo Cruz*, vol. 98, no. 7, pp. 871–878, Oct. 2003, doi: 10.1590/S0074-02762003000700002.

[16] P. Siriyasatien, A. Phumee, P. Ongruk, K. Jampachaisri, and K. Kesorn, "Analysis of significant factors for dengue fever incidence prediction," *BMC Bioinformatics*, vol. 17, Art. no. 166, Apr. 2016, doi: 10.1186/s12859-016-1034-5.

[17] C. M. Benedum, K. M. Shea, H. E. Jenkins, L. Y. Kim, and N. Markuzon, "Weekly dengue forecasts in Iquitos, Peru; San Juan, Puerto Rico; and Singapore," *PLoS Negl. Trop. Dis.*, vol. 14, no. 10, Art. no. e0008710, Oct. 2020, doi: 10.1371/journal.pntd.0008710.

[18] S. Maneerat and E. Daudé, "A spatial agent-based simulation model of the dengue vector Aedes aegypti to explore its population dynamics in urban areas," *Ecol. Model.*, vol. 333, pp. 66–78, Aug. 2016, doi: 10.1016/j.ecolmodel.2016.04.012.

[19] J. Shaman and A. Karspeck, "Forecasting seasonal outbreaks of influenza," *Proc. Natl. Acad. Sci.* USA, vol. 109, no. 50, pp. 20425–20430, Dec. 2012, doi: 10.1073/pnas.1208772109.

[20] Brasil, Ministério da Saúde, Secretaria de Vigilância em Saúde, *Levantamento Rápido de Índices para Aedes aegypti — LIRAa: metodologia para avaliação dos índices de Breteau e Predial e tipo de recipientes*. Brasília, DF, Brasil: Ministério da Saúde, 2013. [Online]. Disponible: Biblioteca Virtual em Saúde do Ministério da Saúde. Accedido: 13-may-2026

[21] J. R. B. Palmer et al., "Citizen science provides a reliable and scalable tool to track disease-carrying mosquitoes," *Nat. Commun.,* vol. 8, no. 1, Art. no. 916, Oct. 2017, doi: 10.1038/s41467-017-00914-9.

[22] N. Andersson et al., "Evidence based community mobilization for dengue prevention in Nicaragua and Mexico (Camino Verde, the Green Way): Cluster randomized controlled trial," *BMJ,* vol. 351, Art. no. h3267, Jul. 2015, doi: 10.1136/bmj.h3267.

[23] N. M. Ferguson et al., "Benefits and risks of the Sanofi-Pasteur dengue vaccine: Modeling optimal deployment," *Science*, vol. 353, no. 6303, pp. 1033–1036, Sep. 2016, doi: 10.1126/science.aaf9590.

[24] G. I. Giraldo-Calderón, O. S. Harb, S. A. Kelly, S. S. C. Rund, D. S. Roos, and M. A. McDowell, "VectorBase.org updates: Bioinformatic resources for invertebrate vectors of human pathogens and related organisms," *Curr. Opin. Insect Sci.*, vol. 50, Art. no. 100860, Apr. 2022, doi: 10.1016/j.cois.2021.11.008.

[25] GBD 2019 Diseases and Injuries Collaborators, "Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: A systematic analysis for the Global Burden of Disease Study 2019," *Lancet*, vol. 396, no. 10258, pp. 1204–1222, Oct. 2020, doi: 10.1016/S0140-6736(20)30925-9.

[26] C. T. Codeço, F. C. Coelho, O. G. Cruz, S. de Souza Oliveira, T. Castro, and L. S. Bastos, "Infodengue: A nowcasting system for the surveillance of arboviruses in Brazil," *Rev. Epidemiol. Sante Publique*, vol. 66, suppl. 5, p. S386, Jul. 2018, doi: 10.1016/j.respe.2018.05.408.