ISSN-e: 2737-6419 Revista Athenea
Período: octubre-diciembre de 2026 Vol.7, Número 26, (pp. 19–29)

Artículo de investigación https://doi.org/10.47460/athenea.v7i26.154

Modelización matemática y predicción de olas de calor mediante series temporales y datos satelitales

Joan Patricio Cobeña Cevallos
https://orcid.org/0000-0003-1157-0036
joanpa13.cc@gmail.com
Escuela Superior Politécnica Agropecuaria de Manabí Manuel Félix López
Calceta, Ecuador

Vicente Marlon Villa Villa
https://orcid.org/0000-0002-4292-2391
mvilla@unach.edu.ec
Universidad Nacional de Chimborazo
Riobamba, Ecuador

Marco Vinicio Parra Chávez
https://orcid.org/0009-0000-1252-0108
marcov.parra@espoch.edu.ec
Escuela Superior Politécnica de Chimborazo
Riobamba, Ecuador

Ana Carola Flores Mancheno*
https://orcid.org/0000-0003-0780-5226
acmancheno@espoch.edu.ec
Escuela Superior Politécnica de Chimborazo
Riobamba, Ecuador

Autor de correspondencia: acmancheno@espoch.edu.ec

Recibido: (14/07/2026), Aceptado: (08/10/2026)

Resumen. Las olas de calor urbanas requieren métodos capaces de representar estacionalidad, persistencia y extremos. Se desarrolló un marco retrospectivo de modelización para Guayaquil con temperatura atmosférica de reanálisis, anomalías oceánicas y temperatura superficial satelital. La estructura temporal se representó mediante regresión armónica dinámica con errores autorregresivos integrados de media móvil y se comparó con persistencia mediante validación cronológica independiente. También se evaluaron una formulación probabilística y una definición alternativa de ola de calor. La modelización temporal redujo el error de temperatura máxima de ERA5-Land respecto del referente ingenuo, mientras que las anomalías oceánicas aportaron una mejora marginal para la máxima y deterioraron la mínima. La predicción probabilística aumentó la detección de episodios, aunque generó más falsas alarmas. MODIS mostró mayor anomalía térmica superficial nocturna durante los episodios más estrictos. Los resultados separan explícitamente exactitud térmica, detección de extremos y evidencia satelital complementaria.

Palabras clave: olas de calor, series temporales, predicción probabilística, teledetección.

Mathematical Modeling and Prediction of Heatwaves Using Time Series and Satellite Data

Abstract. Urban heatwaves require methods capable of capturing seasonality, persistence, and extreme temperature events. A retrospective modeling framework was developed for Guayaquil using reanalysis-based air temperature, oceanic anomalies, and satellite-derived land surface temperature. Temporal dynamics were represented using dynamic harmonic regression with autoregressive integrated moving average (ARIMA) errors and compared against a persistence baseline through independent chronological validation. A probabilistic forecasting approach and an alternative heatwave definition were also evaluated. Temporal modeling reduced the forecasting error for ERA5-Land maximum temperature relative to the naive baseline, whereas oceanic anomalies provided only marginal improvements for maximum temperature and degraded minimum temperature predictions. Probabilistic forecasting improved heatwave event detection, although it also increased the frequency of false alarms. MODIS observations revealed greater nighttime land surface temperature anomalies during events meeting the more stringent heatwave criterion. These findings explicitly distinguish temperature forecasting accuracy, extreme event detection, and complementary satellite-based evidence.

Keywords: heatwaves, time series, probabilistic forecasting, remote sensing.

I. INTRODUCCIÓN

Las olas de calor son episodios persistentes de temperatura anormalmente elevada cuya severidad depende de la intensidad, la duración y el contexto climático local. En el capítulo dedicado a América Central y del Sur del Sexto Informe de Evaluación del IPCC, la evaluación de exposición a olas de calor utiliza como criterio un periodo de al menos dos días consecutivos en el que las temperaturas diaria máxima y mínima superan simultáneamente el percentil 95 de sus respectivas climatologías [1]. En Sudamérica, la persistencia del estrés térmico ha aumentado y presenta contrastes marcados entre regiones costeras e interiores [2]. Estas condiciones convierten la predicción de extremos urbanos en un problema temporal distinto del simple ajuste de la temperatura media.

En Ecuador, las proyecciones disponibles muestran una señal particularmente intensa en la región Costa y asociaciones entre temperaturas extremas y modos de variabilidad del Pacífico [3]. Guayaquil resulta pertinente para evaluar pronósticos de corto plazo por su localización tropical costera, su exposición al Pacífico oriental y la extensión de su superficie construida. Sin embargo, la evidencia ecuatoriana reciente se ha concentrado principalmente en proyecciones climáticas y caracterización térmica, mientras que la evaluación fuera de muestra de modelos diarios de olas de calor sigue siendo limitada.

Desde la perspectiva matemática, una serie diaria de temperatura combina un ciclo anual de periodo largo con dependencia de corto plazo y episodios persistentes. La persistencia atmosférica condiciona la predictibilidad de la temperatura y de sus extremos [4], y modelos recientes de olas de calor muestran la utilidad de representar explícitamente estados persistentes y comportamiento de cola [5]. A ello se suma la influencia potencial del Pacífico oriental: la región Niño 1+2 está estrechamente relacionada con el calentamiento costero que afecta a Perú y Ecuador [6], mientras que el producto NOAA OISST permite describir diariamente las anomalías de temperatura superficial del mar a partir de información satelital e in situ integrada [7].

La información satelital te-rrestre aporta una dimensión diferente. La temperatura superficial terrestre no equivale a la temperatura del aire, pero refleja el balance energético de la superficie y puede revelar persistencia térmica urbana. Estudios con satélites geoestacionarios y reanálisis muestran que los productos de menor resolución pueden suavizar el ciclo diurno y subrepresentar extremos urbanos [8], [9]. Por ello, resulta metodológicamente útil separar el problema predictivo atmosférico del análisis físico superficial, en lugar de asumir que ambas temperaturas son intercambiables.

El objetivo fue desarrollar y evaluar un marco de modelización matemática para pronosticar la temperatura representada por ERA5-Land y las olas de calor en Guayaquil, comparando persistencia, regresión armónica dinámica con errores ARIMA y una extensión con anomalías oceánicas. La contribución no se sustenta en proponer una nueva definición de ola de calor ni un nuevo algoritmo, sino en responder, bajo partición temporal independiente y horizontes de uno a tres días, tres preguntas verificables: si una menor pérdida continua se traduce en mejor detección de extremos; si OISST aporta valor incremental una vez representadas la estacionalidad y la autocorrelación; y si los episodios atmosféricos mantienen una señal térmica superficial coherente en MODIS. El artículo se organiza en fundamentos teóricos, metodología, resultados y su discusión, conclusiones y referencias.

II. FUNDAMENTOS TEÓRICOS

A. Modelización temporal y predicción de extremos

La regresión armónica dinámica representa la estacionalidad mediante pares seno-coseno y delega la autocorrelación residual a un proceso ARIMA. Esta formulación es parsimoniosa para frecuencias diarias porque evita una estructura estacional de cientos de parámetros. En pronóstico, la evaluación debe realizarse fuera de muestra y respetar el orden temporal; los esquemas rolling-origin son especialmente adecuados para evitar contaminación entre entrenamiento y prueba [10]. La exactitud continua y la detección de extremos no son equivalentes: un modelo puede reducir el error medio y, a la vez, suavizar los picos que definen un evento.

ERA5-Land ofrece series terrestres consistentes y de alta resolución temporal para aplicaciones ambientales [11], aunque su desempeño puede degradarse en superficies urbanas y durante extremos. Para la temperatura superficial terrestre (LST, por sus siglas en inglés), MODIS dispone de productos Terra y Aqua cuya exactitud ha sido evaluada mediante validaciones radiométricas e in situ [12], [13]. El

uso conjunto de reanálisis y observación satelital permite conservar continuidad temporal sin renunciar a una comprobación superficial independiente.

B. Información oceánica, superficie urbana e incertidumbre

La modulación oceánica puede actuar en escalas diferentes a las del pronóstico meteorológico diario. OISST integra satélite, barcos y boyas mediante interpolación óptima [7], mientras que Niño 1+2 resume el estado térmico del Pacífico oriental próximo a Ecuador [6]. En paralelo, estudios con MODIS muestran sinergias entre isla de calor superficial y episodios cálidos en grandes regiones urbanas [14]. En entornos urbanos costeros, las evaluaciones de ERA5-Land también han identificado diferencias y sesgos en la representación de la temperatura del aire, lo que justifica interpretar con cautela los extremos derivados exclusivamente del reanálisis [15]. Estas fuentes no garantizan por sí mismas una mejora predictiva; su valor debe demostrarse frente a un modelo temporal sin regresores externos.

El error absoluto medio escalado facilita comparaciones respecto de un referente ingenuo [16], la prueba de Diebold-Mariano contrasta diferencias de precisión entre modelos [17] y el Brier score cuantifica el error de probabilidades binarias [18]. Para extremos, además de precisión y F1, el recall de eventos es relevante porque una alerta puede considerarse útil aun cuando la fecha exacta de inicio presente incertidumbre. La representación explícita de la incertidumbre predictiva también puede mejorar la anticipación del inicio de olas de calor, especialmente cuando la evaluación se formula en términos probabilísticos y no únicamente mediante predicciones puntuales [19].

III. METODOLOGÍA

A. Datos, periodo y partición temporal

Se construyó una serie diaria para Guayaquil entre el 1 de enero de 2003 y el 31 de diciembre de 2025. La unidad temporal de análisis fue el día y los episodios se definieron como secuencias de días consecutivos que cumplieron el criterio térmico correspondiente. Las extracciones geoespaciales se realizaron en Google Earth Engine durante septiembre de 2026. El límite administrativo se obtuvo de `FAO/GAUL/2015/level2` y la temperatura atmosférica de `ECMWF/ERA5_LAND/DAILY_AGGR`. Se extrajeron la temperatura media, la temperatura máxima diaria (Tmax) y la temperatura mínima diaria (Tmin), todas a dos metros de altura, y se calculó la media espacial de cada variable dentro de Guayaquil. La serie final tuvo 8 401 días, sin fechas duplicadas ni huecos.

La LST se obtuvo de `MODIS/061/MOD11A1` (Terra) y `MODIS/061/MYD11A1` (Aqua). La superficie construida se derivó de `JRC/GHSL/P2023A/GHS_BUILT_S/2020` y se agregó a la malla MODIS. Tras una auditoría previa de cobertura, realizada antes de los contrastes finales, se fijó como criterio mínimo una cobertura urbana válida de 10 % para evitar que la media diaria dependiera de muy pocos píxeles; los faltantes MODIS no se imputaron. Para el océano se utilizó `NOAA/CDR/OISST/V2_1` en un sector costero de 81,5–80,5° O y 3,5–2,0° S y en Niño 1+2 (90–80° O; 10° S–0°). Un único día ausente, 14 de enero de 2025, se interpoló linealmente. Como OISST final sustituye la versión preliminar aproximadamente dos semanas después, los regresores oceánicos se desplazaron 15 días; esta decisión se fijó por disponibilidad del producto y no por desempeño predictivo.

La serie se dividió cronológicamente en entrenamiento 2003–2018, validación 2019–2022 y prueba independiente 2023–2025. Ninguna observación de la prueba intervino en la selección estructural de los modelos, en los umbrales climatológicos ni en la elección del umbral probabilístico. Los horizontes fueron uno, dos y tres días. Esta separación se mantuvo también en los análisis de sensibilidad para evitar que las decisiones metodológicas se ajustaran retrospectivamente al periodo final.

B. Definición de eventos y modelos

La definición principal (P95) siguió el criterio utilizado en la evaluación regional del IPCC para América Central y del Sur [1] y clasificó una ola de calor cuando Tmax y Tmin superaron simultáneamente sus respectivos percentiles climatológicos 95 durante al menos dos días consecutivos. Los umbrales se calcularon solo con el entrenamiento mediante una ventana circular de ± 15 días, equivalente a una vecindad climatológica de 31 días alrededor de cada fecha del calendario; la ventana se fijó para suavizar la estacionalidad diaria sin mezclar estaciones alejadas del año. Como análisis de sensibilidad se utilizó la definición alternativa (P90), que consideró Tmax por encima de su percentil climatológico

90 durante al menos tres días consecutivos. La definición alternativa se preestableció antes de comparar el desempeño final y no reemplazó al criterio principal.

El modelo M0 fue persistencia. Para un horizonte h, el pronóstico fue la última temperatura observada:

$$\widehat{T}(t+h) = T(t) \tag{1}$$

En M1 se utilizó regresión armónica dinámica con cinco pares de Fourier y errores ARIMA. La componente estacional fue:

$$H_K(t) = \sum_{k=1}^{K} \left[a_k \sin\left(\frac{2\pi kt}{365{,}25}\right) + b_k \cos\left(\frac{2\pi kt}{365{,}25}\right) \right] \tag{2}$$

El modelo completo se expresó como:

$$T(t) = \beta_0 + H_K(t) + n(t), \qquad n(t) \sim \mathrm{ARIMA}(p,d,q) \tag{3}$$

La selección mediante AICc se efectuó exclusivamente con el periodo de entrenamiento y produjo $K = 5$, ARIMA (1,0,4) para Tmax y ARIMA (2,0,4) para Tmin. Una vez elegidas, estas estructuras se mantuvieron fijas en validación y prueba. M2 conservó exactamente K y los órdenes ARIMA de M1 e incorporó únicamente las anomalías oceánicas desfasadas, de modo que la comparación M1–M2 aislara el valor incremental de OISST y no mezclara ese efecto con un cambio simultáneo de estructura:

$$\begin{aligned} T(t) = \beta_0 + H_K(t) \\ + \gamma_1 A_{\text{costa}}(t-15) \\ + \gamma_2 A_{\text{Niño 1+2}}(t-15) + n(t) \end{aligned} \tag{4}$$

C. *Validación y evaluación probabilística*

La evaluación empleó rolling-origin durante validación y prueba. Para Tmax y Tmin se calcularon error absoluto medio (MAE), raíz del error cuadrático medio (RMSE) y error absoluto medio escalado (MASE) y sesgo, siempre respecto de la serie ERA5-Land utilizada como objetivo. Las diferencias entre M1 y M2 se contrastaron con Diebold-Mariano usando pérdida cuadrática. La clasificación puntual aplicó directamente los umbrales climatológicos a las temperaturas pronosticadas. La estructura y los hiperparámetros permanecieron fijados antes de iniciar la prueba independiente.

Para M1 y M2 se calculó también la probabilidad de superar el umbral diario a partir de la media y el error estándar predictivos:

$$P[T(t) > u(t)] = 1 - \Phi\left(\frac{u(t) - \widehat{\mu}(t)}{\widehat{\sigma}(t)}\right) \tag{5}$$

Los umbrales de decisión probabilísticos se eligieron maximizando F1 únicamente en 2019–2022 y se mantuvieron fijos en 2023–2025. Para P95, dado que el evento requería la excedencia simultánea de los umbrales de Tmax y Tmin, no se multiplicaron directamente las probabilidades marginales bajo un supuesto de independencia. Se utilizó una distribución normal bivariada de los errores predictivos estandarizados, cuya dependencia contemporánea se representó mediante el coeficiente ρ. El coeficiente ρ se estimó como la correlación de Pearson entre los residuos pareados de los modelos de Tmax y Tmin obtenidos exclusivamente durante el periodo de entrenamiento 2003–2018. Este parámetro cuantificó la asociación residual entre los errores de Tmax y Tmin después del ajuste de cada modelo y se mantuvo fijo durante la evaluación fuera de muestra. Los valores utilizados fueron $\rho = 0{,}0426$ para M1 y $\rho = 0{,}0411$ para M2, lo que indica una dependencia residual positiva pero débil. Para cada fecha y horizonte, la probabilidad conjunta se obtuvo como $1 - \Phi(z_{\max}) - \Phi(z_{\min}) + \Phi_2(z_{\max}, z_{\min}; \rho)$, donde $z_{\max}$ y $z_{\min}$ representan los umbrales P95 estandarizados mediante la media y el error estándar predictivos correspondientes. Para P90, al intervenir únicamente Tmax, se utilizó la probabilidad marginal de excedencia. Se evaluaron precisión, recall, F1, tasa de falsas alarmas, Brier score, PR-AUC y ROC-

AUC. El recall de eventos consideró detectado un episodio real cuando al menos un episodio pronosticado se superpuso temporalmente con él.

D. Análisis MODIS

Para Terra día, Aqua día, Terra noche y Aqua noche se construyeron climatologías MODIS con entrenamiento y ventana ± 15 días. Cada episodio se comparó con días no clasificados como ola de calor situados dentro de ± 30 días; esta ventana local se fijó antes del contraste para reducir confusión estacional sin utilizar como controles fechas térmicamente distantes. Las diferencias por evento se resumieron mediante la mediana y su intervalo de confianza del 95 % (IC del 95 %), estimado mediante bootstrap con 5 000 remuestreos. Se aplicó Wilcoxon pareado y corrección de Holm para las cuatro comparaciones. El procedimiento se repitió con P90 como sensibilidad y no se rellenaron huecos MODIS.

E. Reproducibilidad y trazabilidad

Los archivos fuente se conservaron por separado y sin sobrescritura: ERA5-Land, MODIS Terra, MODIS Aqua y OISST. A partir de ellos se generaron una base analítica integrada, pronósticos rolling-origin, métricas, listas de eventos y datos de figuras. El procesamiento estadístico y la generación de salidas se realizaron en Python con NumPy, pandas, SciPy, statsmodels y scikit-learn.

IV. RESULTADOS

A. Frecuencia y estructura temporal de los eventos

La definición P95 identificó 54 episodios y 218 días de ola de calor. La prueba independiente concentró 36 eventos y 173 días; la definición P90 identificó 101 episodios y 536 días. La Tabla 1 resume la distribución por partición analítica. Esta concentración no se interpretó como una estimación causal de tendencia, porque el periodo de referencia se mantuvo fijo para preservar la independencia de la prueba. La sensibilidad de las métricas de ola de calor al baseline está documentada en la literatura reciente [20].

Tabla 1. Distribución de eventos y días de ola de calor

Periodo	**Eventos P95**	**Días P95**	**Eventos P90**	**Días P90**
Entrenamiento 2003–2018	13	31	38	169
Validación 2019–2022	5	14	17	73
Prueba 2023–2025	36	173	46	294
Total	**54**	**218**	**101**	**536**

Nota. P95: ola de calor definida por la excedencia simultánea del percentil 95 de Tmax y Tmin durante al menos dos días consecutivos; P90: excedencia del percentil 90 de Tmax durante al menos tres días consecutivos. Tmax y Tmin corresponden a las temperaturas máxima y mínima diarias, respectivamente.

Para complementar la distribución resumida en la Tabla 1 y visualizar la evolución interanual de los eventos, la Fig. 1 presenta el número anual de días P95, días P90 y eventos P95 registrados entre 2003 y 2025. La representación permite identificar la marcada concentración de días P95 en el tramo final de la serie.

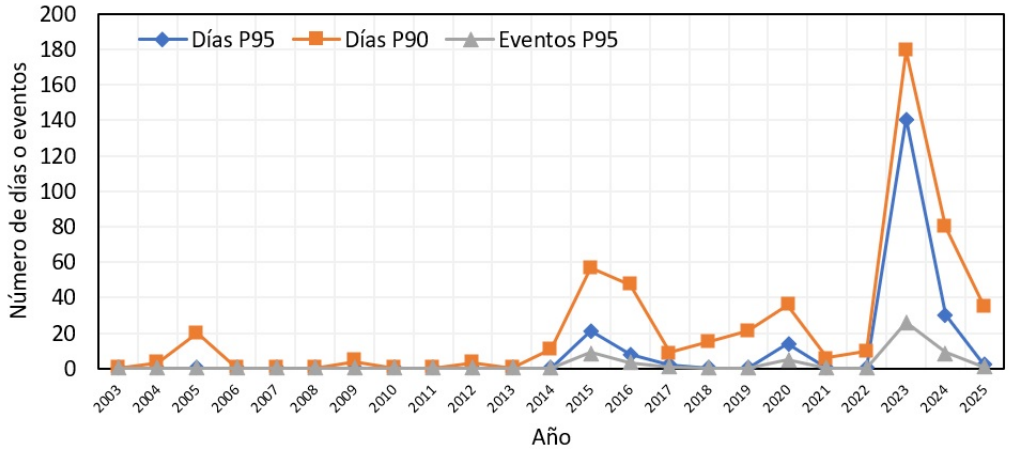


Fig. 1. Evolución anual de días P95, días P90 y número de eventos P95 entre 2003 y 2025. P95: excedencia simultánea del percentil 95 de Tmax y Tmin durante al menos dos días; P90: excedencia del percentil 90 de Tmax durante al menos tres días. Tmax y Tmin son las temperaturas máxima y mínima diarias.

La concentración de episodios recientes coincide temporalmente con años de fuerte calentamiento del Pacífico oriental, pero el diseño no permite atribuir causalidad. En Ecuador, las proyecciones regionales han mostrado una vulnerabilidad térmica elevada de la Costa y asociaciones con índices del Pacífico [3]. Por ello, el resultado debe entenderse como una característica del conjunto de prueba y no como evidencia independiente de una tendencia climática.

B. Predicción de temperatura y aporte de OISST

En 2023–2025, M1 y M2 redujeron de forma consistente el error de Tmax de ERA5-Land frente a persistencia. A un día, el RMSE pasó de 1,033 °C con M0 a 0,834 °C con M1 y 0,833 °C con M2; a tres días fue 1,161, 0,874 y 0,872 °C, respectivamente. La Tabla 2 muestra que la mejora adicional de M2 fue mínima para Tmax y que Diebold-Mariano no detectó diferencias significativas entre M1 y M2.

Tabla 2. RMSE en prueba independiente y contraste Diebold-Mariano entre M1 y M2

h	Tmax M0	Tmax M1	Tmax M2	p (DM, Tmax)	Tmin M0	Tmin M1	Tmin M2	p (DM, Tmin)
1	1,033	0,834	0,833	0,5840	0,437	0,395	0,399	0,0056
2	1,153	0,869	0,868	0,4424	0,540	0,459	0,469	0,0011
3	1,161	0,874	0,872	0,4250	0,568	0,477	0,493	0,0002

Nota. Tmax y Tmin: temperaturas máxima y mínima diarias; M0: persistencia; M1: regresión armónica dinámica con errores ARIMA; M2: M1 con anomalías oceánicas; h: horizonte de predicción en días; RMSE: raíz del error cuadrático medio, expresada en °C; DM: prueba de Diebold-Mariano entre M1 y M2; p: valor de significancia estadística del contraste.

Para complementar la comparación resumida en la Tabla 2 y examinar la dispersión de los errores más allá del RMSE, la Fig. 2 presenta la distribución del error absoluto de Tmax para los modelos M0, M1 y M2 en los horizontes de uno, dos y tres días. Esta representación permite comparar la mediana, la variabilidad y la amplitud de los errores de cada modelo en la prueba independiente.

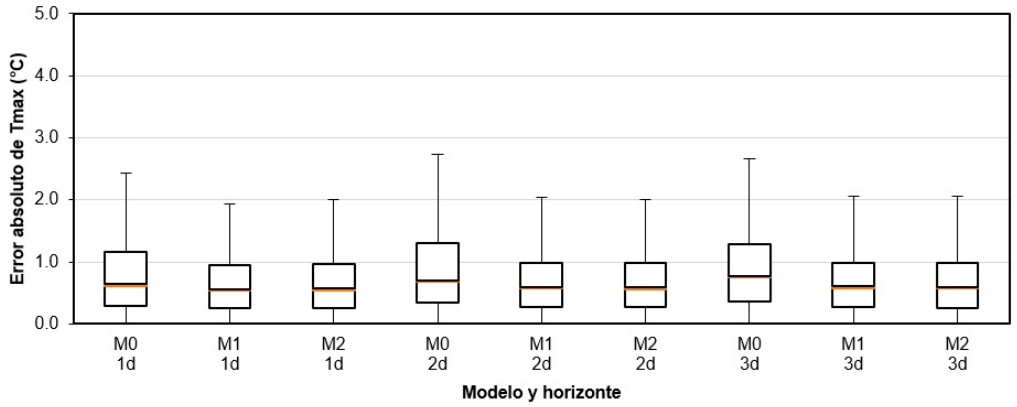


Fig. 2. Distribución del error absoluto de Tmax por modelo y horizonte en la prueba independiente. Tmax: temperatura máxima diaria; M0: persistencia; M1: regresión armónica dinámica con errores ARIMA; M2: M1 con anomalías oceánicas; d: días. Las cajas muestran el rango intercuartílico, la línea naranja la mediana y los bigotes los valores extremos no atípicos según el criterio de 1,5 veces dicho rango.

Para Tmin de ERA5-Land, M1 alcanzó RMSE de 0,395, 0,459 y 0,477 °C a uno, dos y tres días, frente a 0,399, 0,469 y 0,493 °C de M2. En este caso, Diebold-Mariano indicó deterioro significativo de M2 en los tres horizontes. En consecuencia, las anomalías OISST no demostraron un beneficio generalizable de corto plazo. Este resultado es compatible con la relevancia climática regional de Niño 1+2 [6], [7], pero delimita su valor incremental para un pronóstico urbano diario condicionado por una serie atmosférica local.

C. Detección puntual y probabilística

La exactitud continua no se trasladó directamente a la clasificación de extremos. Con P95 y un día de anticipación, persistencia obtuvo F1=0,792, mientras M1 y M2 alcanzaron 0,697 y 0,708. Los modelos temporales mostraron mayor precisión, pero menor recall, lo que evidencia suavizado de algunos extremos. La Tabla 3 compara dos niveles complementarios de evaluación para ambas definiciones: F1, calculado sobre la clasificación diaria derivada de los pronósticos puntuales, y el recall de eventos (RecEv), calculado a nivel de episodio para la clasificación probabilística de M1 y M2. Bajo P90, M2 superó ligeramente a M1 en F1 puntual en los tres horizontes, aunque persistencia siguió siendo muy competitiva.

Tabla 3. Detección de olas de calor en la prueba independiente

h	P95 F1 M0	P95 F1 M1	P95 F1 M2	P95 RecEv M1	P95 RecEv M2	P90 F1 M0	P90 F1 M1	P90 F1 M2	P90 RecEv M1	P90 RecEv M2
1	0,792	0,697	0,708	0,889	0,889	0,842	0,690	0,725	0,978	0,978
2	0,661	0,641	0,646	0,750	0,833	0,717	0,603	0,621	0,978	0,935
3	0,622	0,610	0,633	0,750	0,722	0,612	0,578	0,593	0,935	0,935

Nota. h: horizonte de predicción en días; P95: Tmax y Tmin superiores a sus percentiles 95 durante al menos dos días consecutivos; P90: Tmax superior a su percentil 90 durante al menos tres días consecutivos; M0: persistencia; M1: regresión armónica dinámica con errores ARIMA; M2: M1 con anomalías oceánicas. F1: media armónica de precisión y recall calculada por día a partir de pronósticos puntuales. RecEv: proporción de episodios observados detectados por la clasificación probabilística de M1 y M2; se considera detectado un episodio cuando existe superposición temporal con un episodio pronosticado.

La correlación residual empleada en la distribución bivariada fue baja en ambos modelos ($\rho = 0{,}0426$ en M1 y $\rho = 0{,}0411$ en M2), por lo que la dependencia remanente entre los errores de Tmax y Tmin fue limitada. La formulación probabilística modificó el equilibrio entre omisiones y falsas alarmas. Para P95, M1 detectó 32 de 36 eventos a un día y 27 a tres días; M2 detectó 32 y 26, respectivamente. Para P90, ambos modelos detectaron 45 de 46 eventos a un día, y el recall de eventos se mantuvo por encima de 0,93 hasta tres días. El costo fue un aumento de falsos positivos. La Fig. 3 muestra

que la discriminación probabilística se mantuvo alta, aunque el Brier score empeoró con el horizonte. Este patrón refuerza que un sistema de alerta debe evaluar probabilidades y eventos, no solo RMSE. La literatura reciente sobre predicción de inicio de olas de calor también destaca el papel de la incertidumbre explícita [19].

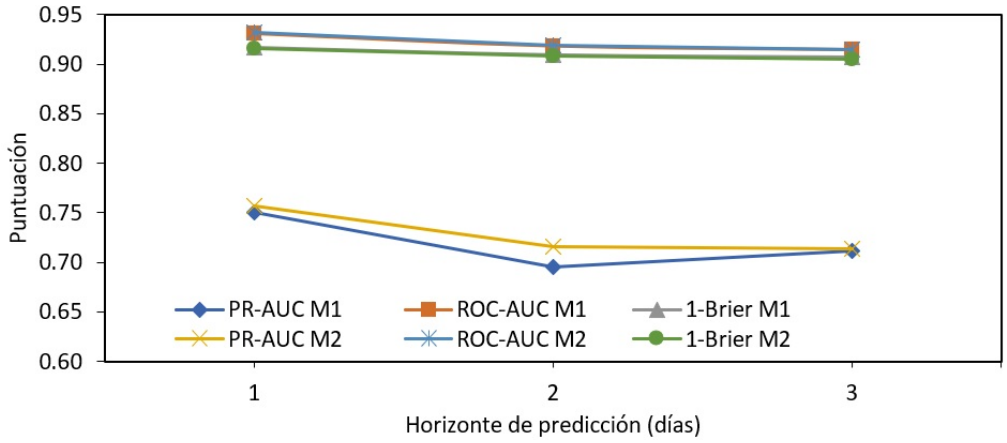


Fig. 3. Métricas probabilísticas P95 de M1 y M2 en la prueba independiente. Se muestra 1-Brier para mantener la orientación de "mayor es mejor". P95: definición de ola de calor basada en la excedencia simultánea del percentil 95 de Tmax y Tmin durante al menos dos días; M1: regresión armónica dinámica con errores ARIMA; M2: M1 con anomalías oceánicas; PR-AUC: área bajo la curva precisión-recall; ROC-AUC: área bajo la curva característica operativa del receptor; 1-Brier: complemento del puntaje de Brier.

D. Respuesta térmica superficial observada con MODIS

MODIS aportó observaciones útiles en 45 de los 54 eventos P95. Tras exigir cobertura urbana válida y comparar cada episodio con controles locales en el calendario, la señal diurna no fue concluyente. En cambio, Terra noche presentó una diferencia mediana de anomalía de +0,91 °C y Aqua noche de +1,24 °C; ambas conservaron significancia después de Holm. La Tabla 4 sintetiza el contraste. Estos resultados se interpretan como contraste físico complementario y no como validación de la temperatura del aire.

Tabla 4. Diferencias de anomalía LST entre eventos y controles temporales

Sensor	Δ P95 (°C)	IC del 95 % (P95)	p (Holm, P95)	Δ P90 (°C)	p (Holm, P90)
Terra día	-0,50	-1,99 a 2,07	0,8325	0,18	0,7400
Aqua día	0,92	-1,08 a 1,94	0,7108	0,73	0,1394
Terra noche	0,91	0,17 a 1,46	0,0359	0,43	0,3242
Aqua noche	1,24	0,23 a 2,32	0,0223	0,32	0,2067

Nota. LST: temperatura superficial terrestre; Δ: mediana de las diferencias de anomalía LST entre eventos y controles temporales; P95 y P90: definiciones de ola de calor basadas en los percentiles 95 y 90 de la temperatura del aire, respectivamente; IC del 95 %: intervalo de confianza estimado mediante bootstrap, presentado para P95; p (Holm): valor p ajustado por comparaciones múltiples mediante el método de Holm. Terra y Aqua corresponden a los sensores satelitales MODIS.

Para complementar los valores resumidos en la Tabla 4 y visualizar la variabilidad entre episodios, la Fig. 4 presenta la distribución de las diferencias de anomalía LST entre los eventos y sus controles temporales para Terra y Aqua, diferenciando observaciones diurnas y nocturnas bajo las definiciones P95 y P90.

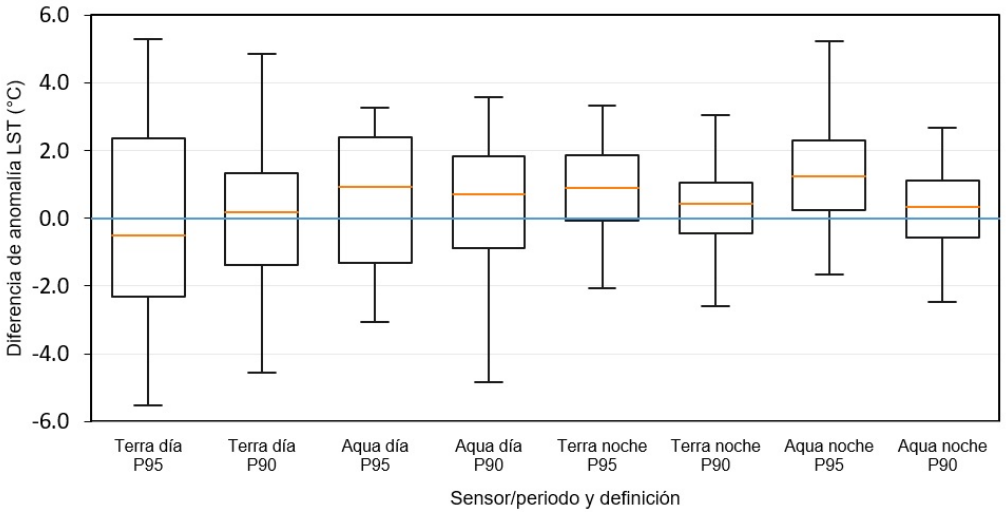


Fig. 4. Distribución de las diferencias de anomalía LST por evento, sensor, periodo y definición de ola de calor. LST: temperatura superficial terrestre; P95: olas de calor definidas por Tmax y Tmin superiores a sus percentiles 95 durante al menos dos días; P90: olas de calor definidas por Tmax superior a su percentil 90 durante al menos tres días. La línea azul indica diferencia nula de anomalía térmica entre eventos y controles.

La respuesta nocturna es físicamente coherente con una menor pérdida de calor almacenado por las superficies urbanas durante episodios intensos. Los estudios satelitales han mostrado que la LST puede revelar rasgos del ciclo diurno que los reanálisis representan de manera atenuada [8], [14], y las evaluaciones de ERA5-Land señalan limitaciones en áreas construidas y durante extremos [9], [15]. Sin embargo, la señal nocturna dejó de ser significativa con P90 después del ajuste múltiple. Por ello, MODIS se interpreta como evidencia física complementaria para los eventos más estrictos y no como un predictor continuo.

CONCLUSIONES

La regresión armónica dinámica con errores ARIMA proporcionó una representación eficaz de la Tmax de ERA5-Land a corto plazo y superó a persistencia en error continuo. Sin embargo, ese beneficio no implicó una ventaja automática para clasificar olas de calor, lo que confirma que la evaluación de extremos requiere métricas específicas además de medidas promedio de pronóstico.

Las anomalías oceánicas costeras y de la región Niño 1+2 no añadieron una mejora generalizable al modelo temporal en horizontes diarios y empeoraron el pronóstico de temperatura mínima. Su utilidad en este contexto se interpreta como información climática de fondo, no como sustituto de predictores atmosféricos de alta frecuencia.

La formulación probabilística incrementó la detección de episodios y permitió explicitar el compromiso entre omisiones y falsas alarmas. MODIS, por su parte, aportó un contraste físico independiente al mostrar mayor persistencia térmica superficial nocturna durante los episodios P95, aunque esa señal no fue robusta bajo la definición alternativa.

El marco propuesto constituye una evaluación retrospectiva y no un sistema operativo de alerta. Su extensión prioritaria debe incorporar observaciones meteorológicas in situ para validar y, si procede, corregir ERA5-Land, además de predictores atmosféricos operativos como humedad, viento, radiación y nubosidad. Mantener particiones temporales estrictas y evaluación probabilística permitirá determinar si esas variables aumentan el recall sin degradar la calibración ni incrementar excesivamente las falsas alarmas.

REFERENCIAS

[1] E. Castellanos *et al.*, "Central and south america," in *Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, H.-O. Pörtner *et al.*, Eds. Cambridge, UK and New York, NY, USA: Cambridge University Press, 2022, pp. 1689–1816.

[2] V. F. V. V. Miranda *et al.*, "Heat stress in south america over the last four decades: A bioclimatic analysis," *Theoretical and Applied Climatology*, vol. 155, no. 2, pp. 911–928, 2024.

[3] D. Portalanza *et al.*, "Projected heat waves in ecuador under climate change: Insights from hadgem-regcm4 coupled model," *Earth*, vol. 5, no. 1, pp. 90–109, 2024.

[4] E. Holmberg, S. Tietsche, and G. Messori, "Forecasting atmospheric persistence and implications for the predictability of temperature and temperature extremes," *Quarterly Journal of the Royal Meteorological Society*, vol. 150, no. 765, pp. 5518–5534, 2024.

[5] V. Gioia, G. Di Credico, and F. Pauli, "Local summer temperature dynamics: Bayesian markov-switching to forecast annual frequency and duration of heat waves," *International Journal of Forecasting*, 2026.

[6] K. R. Gomez-Tunque, E. Ingol-Blanco, R. R. Gutierrez, J. Mejia-Marcacuzco, E. Chávarri-Velarde, and E. Pino-Vargas, "Forecasting coastal enso warming in the niño 1+2 region using convlstm: Toward improved early warning in peru and ecuador," *Water Resources Research*, vol. 62, no. 8, p. e2025WR041877, 2026.

[7] B. Huang, C. Liu, V. Banzon, E. Freeman, and H.-M. Zhang, "Improvements of the daily optimum interpolation sea surface temperature (doisst) version 2.1," *Journal of Climate*, vol. 34, no. 8, pp. 2923–2939, 2021.

[8] J. Lee and A. E. Dessler, "Improved surface urban heat impact assessment using goes satellite data: A comparative study with era-5," *Geophysical Research Letters*, vol. 51, no. 1, p. e2023GL107364, 2024.

[9] J. Yu *et al.*, "Evaluating era5-land air temperature and dew point biases in built-up environments in the uk," *Urban Climate*, vol. 69, p. 103088, 2026.

[10] L. J. Tashman, "Out-of-sample tests of forecasting accuracy: An analysis and review," *International Journal of Forecasting*, vol. 16, no. 4, pp. 437–450, 2000.

[11] J. Muñoz-Sabater *et al.*, "Era5-land: A state-of-the-art global reanalysis dataset for land applications," *Earth System Science Data*, vol. 13, no. 9, pp. 4349–4383, 2021.

[12] Z. Wan, "New refinements and validation of the collection-6 modis land-surface temperature/emissivity product," *Remote Sensing of Environment*, vol. 140, pp. 36–45, 2014.

[13] S.-B. Duan *et al.*, "Validation of collection 6 modis land surface temperature product using in situ measurements," *Remote Sensing of Environment*, vol. 225, pp. 16–29, 2019.

[14] C. Wei, W. Chen, Y. Lu, T. Blaschke, J. Peng, and D. Xue, "Synergies between urban heat island and urban heat wave effects in 9 global mega-regions from 2003 to 2020," *Remote Sensing*, vol. 14, no. 1, p. 70, 2022.

[15] J. Zou *et al.*, "Performance of air temperature from era5-land reanalysis in coastal urban agglomeration of southeast china," *Science of the Total Environment*, vol. 828, p. 154459, 2022.

[16] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *International Journal of Forecasting*, vol. 22, no. 4, pp. 679–688, 2006.

[17] F. X. Diebold and R. S. Mariano, "Comparing predictive accuracy," *Journal of Business and Economic Statistics*, vol. 13, no. 3, pp. 253–263, 1995.

[18] G. W. Brier, "Verification of forecasts expressed in terms of probability," *Monthly Weather Review*, vol. 78, no. 1, pp. 1–3, 1950.

[19] Q. Zhang, M. Mu, and G. Sun, "Skillful subseasonal ensemble predictions of heat wave onsets through better representation of land surface uncertainties," *npj Climate and Atmospheric Science*, vol. 8, no. 1, p. 15, 2025.

[20] V. A. Fraser-Leonhardt, G. C. Hegerl, and J. D. P. Mollard, "Sensitivity of heatwave characteristics to the climatic baseline," *Atmospheric Science Letters*, vol. 27, no. 4, p. e70017, 2026.