

ESTIMACIÓN DE DATOS DIARIOS FALTANTES EN LA TEMPERATURA SUPERFICIAL DEL MAR, LABORATORIOS COSTEROS - IMARPE

ESTIMATION OF MISSING DAILY SEA SURFACE TEMPERATURE DATA, IMARPE COASTAL LABORATORIES

Carlos Quispe-Ccalluari^{1,*} 

Jorge Quispe Sánchez¹ 

¹ Instituto del Mar del Perú, Dirección General de Investigaciones Oceanográficas y Cambio Climático, Callao, Perú.

*Correspondencia. E-mail: cquispe@imarpe.gob.pe

Recibido: 08-01-2024, Aceptado: 14-05-2025, Publicado: 19-05-2025

RESUMEN

Los datos de la Temperatura Superficial del Mar (TSM) obtenidas *in situ* por los Laboratorios Costeros del IMARPE, representan uno de los registros históricos diarios más completos de variables oceanográficas en el Perú, los que actualmente son monitoreados con fines de investigación científica y para detectar la ocurrencia de eventos como El Niño y La Niña. Sin embargo, algunos de estos registros presentan vacíos que dificultan calcular la climatología en cada estación costera. El presente trabajo tiene el propósito de completar estos vacíos de datos de TSM *in situ* diario, usando como predictor la TSM satelital diaria en los modelos estadísticos lineales. Para ello, se emplearon una Regresión Múltiple y una Regresión Múltiple Segmentada (por tramos). Se concluye que se lograron reconstruir las series de tiempo diarias de TSM *in situ* usando como predictor la TSM satelital. Para las estaciones costeras de Chicama, Chimbote, Huacho, Callao e Ilo, la Regresión Múltiple Segmentada (por tramos) permitió generar con mayor robustez los datos diarios faltantes, con los cuales se han reconstruido las series continuas de datos diarios de todas las estaciones costeras.

PALABRAS CLAVE: estimación de datos diarios, Regresión Múltiple, Regresión Múltiple Segmentada

ABSTRACT

The *in situ* Sea Surface Temperature (SST) records collected by IMARPE's coastal laboratories constitute one of the most extensive daily time series of oceanographic variables available for the Peruvian coast. These datasets are critical for oceanographic research and for monitoring large-scale climate events such as El Niño and La Niña. However, gaps in the daily records persist at several stations, making the development of consistent climatologies and long-term trend analyses difficult. This study aims to estimate missing daily *in situ* SST values using satellite-derived daily SST as a predictor within a linear statistical framework. Two methods have been applied and evaluated: Multiple Linear Regression (MLR) and Segmented Multiple Linear Regression (SMLR). The analysis shows that satellite SST can effectively serve as a predictor for the reconstruction of *in situ* SST time series. Among the approaches tested, PMLR provided more robust estimates, particularly for the coastal stations of Chicama, Chimbote, Huacho, Callao and Ilo. The resulting gap-filled time series improve the continuity and reliability of historical SST datasets for climatological assessments and operational monitoring.

KEYWORDS: daily data estimation, Multiple Regression, Segmented Multiple Regression

1. INTRODUCCIÓN

La Temperatura Superficial del Mar (TSM) es un parámetro físico oceánico, y se refiere a la temperatura de la capa más superficial del océano, generalmente en los primeros metros de profundidad (MINNETT, 2014), con el registro histórico más completo que se ha obtenido en los laboratorios costeros del IMARPE. El monitoreo de la va-

1. INTRODUCTION

Sea Surface Temperature (SST) is a key oceanographic physical parameter that represents the temperature of the ocean's uppermost layer, typically within the first few meters (MINNETT, 2014). Constitutes one of the most comprehensive historical datasets collected by IMARPE's coastal laboratories. Continuous monitoring of

riación temporal de la salinidad, es importante para trabajos de investigación y también en la detección de periodos cálidos y fríos a diferentes escalas de tiempo, como, por ejemplo, El Niño y La Niña a escala interanual (YU & KIM, 2012; QUISPE-CCALLUARI *et al.*, 2018; NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION [NOAA], 2025a, b); la variación estacional asociada al ciclo anual de la radiación solar (GRADOS *et al.*, 2018); a escala intra-estacional la variación asociada a los pasos de la onda Kelvin cálida o fría (BACHELERY *et al.*, 2016; ILLIG *et al.*, 2014) y, a escala sinóptica asociada a los vientos del afloramiento costero peruano (DEMARCO & FAURE, 2000).

Sin embargo, el registro histórico de datos de TSM tiene algunos vacíos debido a problemas en la recolección por variadas razones o por pérdida de información. Estos vacíos dificultan el análisis de las series de tiempo y el cálculo de las climatologías, así como de las anomalías diarias, por lo que es necesario estimar los datos faltantes. Actualmente, hay varios tipos de productos de TSM obtenidos por diferentes fuentes observacionales como, por ejemplo, (i) el que se basa únicamente en observaciones *in situ* que incluye barcos, boyas a la deriva, puntos fijos y laboratorios costeros; (ii) los que están basados solamente en satélites y (iii) los que están basados en una combinación de observaciones tanto *in situ* como satelitales (HUANG *et al.*, 2021). Los productos que contienen, muy probablemente, datos más completos, es decir sin faltantes, son los que están basados en la combinación de observaciones tanto *in situ* como satelitales. Entre las fuentes de este tipo de producto se tiene:

(1) *Daily Optimum Interpolation Sea Surface Temperature*, versión 2.0 (DOISST v2.0), (REYNOLDS *et al.*, 2007) que contiene un registro histórico desde setiembre de 1981 con resolución espacial de 0,25°; este producto incorpora observaciones de diferentes plataformas (satélite, barcos, boyas y flotadores Argo) en una grilla regular global.

(2) OSTIA (DONLON *et al.*, 2012) que proporciona mapas diarios sin interrupciones a una resolución espacial de 0,05°, usando datos *in situ* y satelitales de radiómetros infrarrojos y de microondas, con un registro histórico que inicia desde enero 2007.

SST is essential for scientific research and for identifying warm and cold periods across various temporal scales. These include interannual phenomena such as El Niño and La Niña (YU & KIM, 2012; QUISPE-CCALLUARI *et al.*, 2018; NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION [NOAA] 2025a, b); seasonal variations driven by the annual solar radiation cycle (GRADOS *et al.*, 2018); intra-seasonal fluctuations linked to the passage of warm or cold Kelvin waves (BACHELERY *et al.*, 2016; ILLIG *et al.*, 2014); and synoptic-scale changes associated with coastal upwelling winds along the Peruvian coast (DEMARCO & FAURE, 2000).

Despite its extensive historical record, the SST dataset contains gaps due to challenges in data collection and occasional loss of information. These missing values hinder time series analysis, the construction of climatologies, and the calculation of daily anomalies, making the reconstruction of these gaps essential. Currently, there are several types of SST products available, derived from different observational platforms: (i) datasets based solely on *in situ* observations, including ships, drifting buoys, fixed stations, and coastal laboratories; (ii) satellite-only datasets; and (iii) blended datasets that integrate both *in situ* and satellite observations (HUANG *et al.*, 2021). Among these, blended products typically offer the most complete and uninterrupted records. Notable examples include:

(1) Daily Optimum Interpolation Sea Surface Temperature, Version 2.0 (DOISST v2.0; REYNOLDS *et al.*, 2007) provides daily global SST fields since September 1981 at a 0.25° spatial resolution, incorporating satellite data, ship observations, buoys, and Argo floats.

(2) OSTIA (DONLON *et al.*, 2012) provides uninterrupted daily SST maps at a spatial resolution of 0.05°, integrating both *in situ* observations and satellite data from infrared and microwave radiometers. Its historical record begins in January 2007.

(3) Multi-Scale Ultra High Resolution (MUR) offers daily SST data since June 2002 at a spatial resolution of 1 km (CHIN *et al.*, 2017). It combines

(3) *Multi-Scale Ultra High Resolution* (MUR) proporciona un conjunto de datos de TSM diaria desde junio 2002, a 1 km de resolución espacial (CHIN *et al.*, 2017) e incorpora datos procedentes de satélites (infrarrojos y microondas) y registros *in situ* de *iQuam Project* de la NOAA (*National Oceanic and Atmospheric Administration*).

Los productos basados en información satelital combinados con observaciones *in situ* de base de datos internacionales resultan ser útiles, pues casi no tienen datos faltantes y es continua en su registro, de modo que se empleó la base de datos que tiene el registro histórico más completo, es decir, *Daily Optimum Interpolation Sea Surface Temperature*, versión 2.0 (OISST v2.0) de la NOAA.

En consecuencia, el presente trabajo tiene el objetivo de estimar los datos faltantes de las series de tiempo diarias de TSM de los laboratorios costeros del IMARPE, usando la información satelital.

2. MATERIALES Y MÉTODOS

Datos: se analizaron los datos de TSM *in situ* diarios de nueve laboratorios costeros (Tabla 1, Fig. 1) información suministrada por el Área Funcional de Investigaciones en Oceanografía Física.

Se analizó la TSM satelital diaria de Nivel 4, DOISST v2.0 (*Daily Optimum Interpolation Sea Surface Temperature*) obtenido de NOAA NCDC (*National Climatic Data Center of National Oceanic and Atmospheric Administration*) con resolución espacial de 0,25° (REYNOLDS *et al.*, 2007) y del periodo 1982-2019.

satellite observations (infrared and microwave) with *in situ* measurements from NOAA's *iQuam Project* (*National Oceanic and Atmospheric Administration*).

Satellite-derived products that incorporate *in situ* data from international databases are particularly valuable for time series analyses, as they provide nearly complete, continuous records with minimal data gaps. Therefore, this study used the dataset with the most comprehensive historical coverage: the NOAA Daily Optimum Interpolation Sea Surface Temperature, version 2.0 (OISST v2.0).

Accordingly, the objective of this study is to estimate missing values in the daily SST time series recorded at IMARPE's coastal laboratories using satellite-based data.

2. MATERIALS AND METHODS

Data: this study analyzed daily *in situ* Sea Surface Temperature (SST) records from nine IMARPE's coastal laboratories (Table 1, Fig. 1), provided by the Functional Area of Physical Oceanography Research.

Satellite-derived daily SST data (Level 4) were also examined, specifically the Daily Optimum Interpolation Sea Surface Temperature Version 2.0 (DOISST v2.0), obtained from NOAA's National Climatic Data Center. This dataset offers global coverage at a spatial resolution of 0.25° and spans the period from 1982 to 2019 (REYNOLDS *et al.*, 2007).

Tabla 1.- Estaciones costeras del IMARPE

Table 1. IMARPE's Coastal Stations

Laboratorios costeros Coastal laboratories	Coordenadas geográficas Geographical coordinates
Tumbes	3°38'17,7"S - 80°36'24,8"W
Paíta	5°4'37,98"S - 81°07'13,98"W
San José	6°46'15"S - 79°58'00"W
Chicama	7°41'40"S - 79°26'19,3"W
Chimbote	09°27'19,8"S - 78°23'0,2"W
Huacho	11°07'18,83"S - 77°36'58,53"W
Callao	12°03'57,78"S - 77°09'30,54"W
Pisco	13°42'33,4"S - 76°13'18,5"W
Ilo	17°38'38,4"S - 71°20'45,6"W

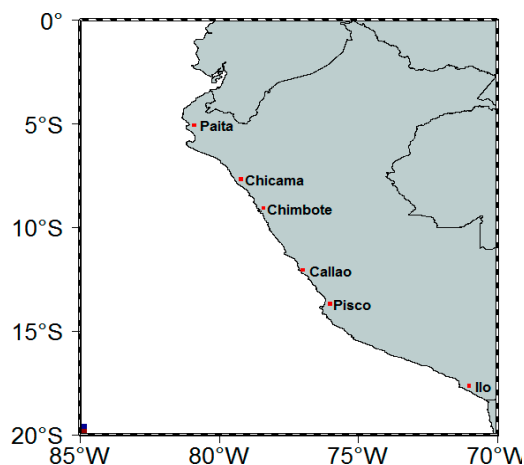


Figura 1.- Ubicación geográfica de las estaciones costeras del IMARPE

Figure 1. Geographical location of IMARPE's coastal stations

Se empleó el **Índice Térmico Costero Peruano** (ITCP) (QUISPE-CCALLUARI *et al.*, 2018) para categorizar las condiciones térmicas en la franja costera peruana. El ITCP fue clasificado en tres condiciones: frío, neutro y cálido.

Métodos: se identificó un punto grilla en la base de datos satelital de TSM, el más cercano a la ubicación de cada Laboratorio Costero (con radios menores a 13 millas náuticas), debido a que contendría información más precisa sobre la variabilidad local del Laboratorio Costero. Por su parte, se extrajeron series de tiempo de datos completos usando el programa Ferret, que es un producto de NOAA's *Pacific Marine Environmental Laboratory* (NOAA, 2014), considerando que la grilla del campo de datos de TSM es 0,25°.

Se realizó el análisis de correlación lineal entre las series de tiempo *in situ* y satelital y, para la estimación de los datos faltantes se emplearon la Regresión Lineal Clásica y la Regresión por tramos.

En la Regresión Lineal Clásica (CHAMBERS, 1992), método clásico, que de aquí en adelante llamaremos "Regresión Lineal Univariada (RLU)": se utilizó la siguiente ecuación

$$Y_i = \alpha_0 + \alpha_1 X_i + u_i, \quad \text{con } i=1,2,\dots,n$$

donde, Y_i es la serie de tiempo de TSM *in situ*; X_i es la serie de tiempo de TSM satelital; u_i es el término aleatorio; α_1 es la pendiente de la Regresión Lineal.

La estimación de los coeficientes fue realizada a través de la técnica de mínimos cuadrados para modelos lineales, los cuales minimizan la suma de cuadrados de los residuales implementado en el programa R Core Team (2019).

Asimismo, se aplicó otro modelo de Regresión Lineal Múltiple (RLM), donde se incluyó la variable categórica "condición" que corresponde a las categorías del ITCP (fría=1, neutra=2 y cálida=3). También se incluyó el efecto de la variación estacional con la variable "mes" (enero=1, febrero=2, marzo=3, abril=4, mayo=5, junio=6, julio=7, agosto=8, setiembre=9, octubre=10, noviembre=11 y diciembre=12).

$$Y_i = \alpha_0 + \alpha_1 X_i + \text{mes} + \text{condicion} + u_i, \\ \text{con } i=1,2,\dots,n$$

Additionally, the **Peruvian Coastal Thermal Index** (PCTI) (QUISPE-CCALLUARI *et al.*, 2018) was used to classify thermal conditions along the coast into three categories: cold, neutral, and warm.

Methods: for each IMARPE's coastal station, the satellite SST value was extracted from the grid point nearest to the coastal laboratory's location (within a radius of less than 13 nm), assuming it would best represent local variability. Time series with complete records were extracted using Ferret, which is a tool developed by NOAA's Pacific Marine Environmental Laboratory (NOAA, 2014), considering that the SST data field grid has a resolution of 0.25°

A linear correlation analysis was performed between the *in situ* and satellite-based time series, and two statistical approaches were used to estimate missing data: Classical Linear Regression and Segmented Regression.

In the Classical Linear Regression method (CHAMBERS, 1992), hereafter referred to as Univariate Linear Regression (ULR), the following equation was applied:

$$Y_i = \alpha_0 + \alpha_1 X_i + u_i, \quad i=1,2,\dots,n$$

Where Y_i is the *in situ* SST time series; X_i is the satellite-derived SST time series; u_i is the random error term; and α_1 is the slope of the regression.

The coefficients were estimated using the least squares method for linear models, minimizing the sum of squared residuals, as implemented in the R software (R Core Team, 2019).

Additionally, a Multiple Linear Regression (MLR) model was applied, incorporating categorical variables such as the "condition", representing the PCTI phase (cold=1, neutral=2, warm=3), and "month", to account for seasonal variation (January=1, February=2, ..., December=12).

$$Y_i = \alpha_0 + \alpha_1 X_i + \text{month} + \text{condition} + u_i, \\ i=1,2,\dots,n$$

Finalmente, según se necesitó, se empleó un modelo de Regresión Múltiple por Tramos (RMT) o conocida también como Regresión Segmentada, que consta de dos partes o tramos lineales, y que está representado de la siguiente manera:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 (X_i - X^*) D_i + \text{mes} + \text{condicion} + u_i$$

con $i=1,2,\dots,n$

Donde, Y_i es la serie de tiempo de TSM *in situ*; X_i es la serie de tiempo de TSM satelital; X^* es el valor del umbral de la TSM satelital (la cual puede presentar dificultades en su obtención); $D_i = 1$ si y solo si $X_i > X^*$ o $D_i = 0$ si y solo si $X_i < X^*$; u_i es el termino aleatorio; β_1 es la pendiente de la regresión donde $X_i < X^*$; $(\beta_1 + \beta_2)$ sería la pendiente de regresión si y solo si $D_i = 1$ y $X_i > X^*$; “condición” es la variable categórica del ITCP (fría=1, neutra=2 y cálida=3) y “mes” es la variable categórica de los meses (enero=1, febrero=2, marzo=3, abril=4, mayo=5, junio=6, julio=7, agosto=8, setiembre=9, octubre=10, noviembre=11 y diciembre=12).

Las estimaciones de los parámetros se obtuvieron a través de un procedimiento iterativo (MUGGEO, 2003) que necesita de valores iniciales para los parámetros del umbral. El algoritmo está implementado en el programa R Core Team (2019).

Para la validación de los métodos y su comparación sobre los desempeños de los modelos ajustados para la estimación de datos faltantes, se empleó la técnica de validación cruzada (MASON & BADDOUR, 2008; HYNDMAN & ATHANASOPOULOS, 2018) que consiste en dividir el total de datos en un subconjunto para el ajuste (“datos de entrenamiento”) que fue 80% y el otro 20% de los datos para la evaluación de habilidad predictiva (“datos de test”) para su posterior comparación.

Con el fin de medir el desempeño de los modelos estadísticos se calculó la Desviación Absoluta Media (DAM) o error medio absoluto que mide el promedio de las distancias absolutas entre un valor estimado y el valor real, la raíz del error cuadrático medio (RMSE) que penaliza en términos cuadráticos las mayores distancias entre un valor estimado y el valor real, y el coeficiente de determinación (R^2) que varía en un rango de 0 a 100% (valores cercanos a 0 indica mal desempeño, mientras que valores cercanos a 100% indica muy buen desempeño del modelo).

When necessary, a Segmented Multiple Linear Regression (SMLR) model—also known as Segmented Regression—was applied. This approach assumes two distinct linear segments and is represented by:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 (X_i - X^*) D_i + \text{month} + \text{condition} + u_i$$

$i=1,2,\dots,n$

Where Y_i is the *in situ* SST time series; X_i is the satellite-derived SST time series; X^* denotes the breakpoint or threshold value in the satellite SST time series (which can be challenging to identify), and; $D_i = 1$ if $X_i > X^*$ or $D_i = 0$ if $X_i < X^*$; u_i is the random error term; β_1 is the regression slope for the segment where $X_i < X^*$; $(\beta_1 + \beta_2)$ represents the regression slope only when $D_i = 1$ and $X_i > X^*$; “condition” is the categorical variable representing the PCTI phase (cold = 1, neutral = 2, warm = 3); and “month” is the categorical variable for the months (January = 1, February = 2, March = 3, April = 4, May = 5, June = 6, July = 7, August = 8, September = 9, October = 10, November = 11, December = 12).

Parameter estimates for the segmented model were obtained through an iterative procedure (MUGGEO, 2003) that requires initial values for the threshold. This algorithm is also implemented in the R software (R Core Team, 2019).

Cross-validation was applied to validate the models and compare their performance in estimating missing data (MASON & BADDOUR, 2008; HYNDMAN & ATHANASOPOULOS, 2018). The dataset was split into two subsets: 80% for model training and 20% for testing predictive skill.

The performance of the statistical models was evaluated using the Mean Absolute Deviation (MAD), which quantifies the average absolute difference between estimated and observed values; the Root Mean Square Error (RMSE), which penalizes larger deviations more heavily; and the Coefficient of Determination (R^2), which ranges from 0% (poor model performance) to 100% (excellent model performance).

$$DAM = \frac{1}{n} \sum_{i=1}^n |Yobs_i - Yest_i|$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Yobs_i - Yest_i)^2}$$

Dónde: $Yobs_i$ es el i-esimo valor TSM *in situ* observado, $Yest_i$ es el i-esimo valor TSM estimado y n es número de datos empleado en la evaluación de habilidad predictiva de la validación cruzada.

Los análisis de datos se realizaron usando el programa R *statistical language* (R Core Team, 2019).

3. RESULTADOS Y DISCUSIÓN

Comparación entre TSM diarias *in situ* y datos satelitales de TSM

Las series de tiempo de TSM *in situ* y satelitales cuyas correlaciones fueron mayores a 0,8 ($p < 0,05$) se encontraron en Tumbes, Paita y San José (zona norte), así como en Chicama y Chimbote (zona centro); mientras que, en Huacho y Callao las correlaciones se ubicaron entre 0,7 y 0,8. En Pisco e Ilo (zona centro – sur), los datos estuvieron más dispersos, razón por la cual se encontraron las correlaciones más bajas con valores menores a 0,7 ($p < 0,05$). Ante la mayor dispersión de pares ordenados de TSM *in situ* y satélite en Pisco, hubo baja correlación. Se observó que, para altos valores de TSM *in situ* y satelital en las estaciones Chicama, Chimbote, Huacho, Callao e Ilo, habría un cambio de pendiente (cambio de estructura) que estaría asociado a los meses de verano o también a la ocurrencia de eventos cálidos asociados a El Niño según el ITCP (Fig. 2).

Estimación de datos faltantes de TSM diarias en las estaciones costeras

Para la estimación de datos faltantes de temperatura superficial del mar en los diversos Laboratorios Costeros del IMARPE (nueve), se utilizó el modelo de Regresión Lineal Univariada (RLU) y el modelo de Regresión Lineal Múltiple a través de la validación cruzada, calculándose los indicadores de desempeño (Tabla 2). Las simulaciones (en líneas rojas) del modelo de Regresión Lineal Múltiple (RLM) se presentan en las Figuras 3, 4 y 5, Anexo). En la Tabla 2, se aprecia que la estación costera de Pisco presentó la mayor desviación ab-

$$DAM = \frac{1}{n} \sum_{i=1}^n |Yobs_i - Yest_i|$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Yobs_i - Yest_i)^2}$$

Where: $Yobs_i$ is the i-th observed *in situ* SST value, $Yest_i$ is the i-th value of the estimated SST, and n is the number of observations used for evaluating predictive skill in the cross-validation process.

All data analyses were conducted using the R statistical software (R Core Team, 2019).

3. RESULTS AND DISCUSSION

Comparison between *in situ* daily SST and satellite SST data

Time series of *in situ* and satellite derived SST showed strong correlations > 0.8 ($p < 0.05$) at Tumbes, Paita and San José (northern region) and Chicama and Chimbote (central region). In contrast, stations such as Huacho and Callao showed moderate correlations, with values between 0.7 and 0.8. At Pisco and Ilo (south-central region) the data were more scattered, resulting in the lowest correlations < 0.7 ($p < 0.05$). The greater dispersion of matched SST pairs in Pisco led to weaker correlations. In addition, a possible shift in the regression slope at higher SST values (both *in situ* and satellite) was observed in Chicama, Chimbote, Huacho, Callao and Ilo, suggesting a structural change probably related to the summer months or to warm anomalies associated with El Niño events, as indicated by the Peruvian Coastal Thermal Index (PCTI) (Fig. 2).

Estimation of Missing Daily SST Data at Coastal Stations

To estimate missing daily sea surface temperature (SST) data at IMARPE's nine coastal laboratories, both Univariate Linear Regression (ULR) and Multiple Linear Regression (MLR) models were applied using cross-validation, with performance indicators summarized in Table 2). Simulations generated by the MLR model (shown in red lines) are presented in Figures 3 to 5 (Annex). Table 2 shows that the

soluta media (DAM; $>1,19$ °C), es decir, en Pisco, cada predicción de TSM diaria faltante está desviada del valor real en promedio por $1,19$ °C. Los siguientes valores altos de DAM se encontraron en las estaciones costeras de Paita e Ilo.

En todos los casos, las estimaciones de RLM tuvieron mejor desempeño (menor RMSE, menor DAM y mayor R^2) que la RLU (que incluye únicamente la temperatura superficial del mar satelital), por lo que la inclusión de variables categóricas como el mes y las condiciones del ITCP (fría, neutra y cálida) fue relevante (Tabla 2). No obstante, las Regresiones Múltiples para las estaciones costeras de Chicama, Chimbote, Huacho, Callao e Ilo exhiben una subestimación de los valores de TSM durante los periodos El Niño 1982-83 y 1997-98 (Figs. 4, 5, Anexo). Cabe señalar que, el más bajo valor de coeficiente de determinación entre los valores observados y estimados de las estaciones costeras, en la validación cruzada, se obtuvo 55,8% para Ilo.

Mejoras en las estimaciones de datos faltantes de TSM diarias *in situ* de las estaciones costeras de Chicama, Chimbote, Huacho, Callao e Ilo

En la Figura 2, se muestran las correlaciones más bajas que se registraron en las estaciones costeras de Callao e Ilo; también se observa que las bajas correlaciones están asociadas a una relación diferente entre la TSM satelital y la TSM *in situ* bajo condiciones cálidas o El Niño (puntos rojos en la Fig. 2). Para resolver esta relación no lineal, se aplicó la regresión por tramos para los datos de las estaciones Callao e Ilo. Esta regresión consta de dos partes o tramos lineales, su representación matemática está en la sección de métodos.

Para el caso de la estación costera del Callao, el análisis de la regresión por tramos (Fig. 6) estimó el umbral de cambio de pendiente de la regresión en $21,15$ °C. Al comparar con el modelo de Regresión Múltiple, el coeficiente de determinación (R^2) se incrementó de 72% a 77,4% y los estimados tuvieron una mejor aproximación hacia los valores *in situ* (Fig. 7), como, por ejemplo, en comparación a lo obtenido con la Regresión Lineal Múltiple. Se observó que la subestimación de los valores de TSM se redujo durante la ocurrencia de los eventos extraordinarios El Niño (EN) 1982-83 y 1997-98, y los valores de RMSE y DAM fueron menores a los estimados con Regresión Lineal Múltiple (Tabla 3).

Pisco station exhibited the highest mean absolute deviation (MAD > 1.19 °C), meaning that each missing daily SST value estimated at this station deviated, on average, by 1.19 °C from the observed value. High MAD values were also observed at the Paita and Ilo stations.

In all cases, the MLR model outperformed the ULR model, as indicated by lower RMSE and MAD values and higher R^2 scores. This improvement highlights the relevance of including categorical variables such as month (January-December) and the thermal conditions defined by the Peruvian Coastal Thermal Index (PCTI: cold, neutral, and warm), as shown in Table 2. However, the MLR model tended to underestimate SST values during El Niño events (1982-83 and 1997-98) at the Chicama, Chimbote, Huacho, Callao, and Ilo stations (Figs. 4, 5, Annex). Notably, the lowest R^2 value observed in the cross-validation analysis was found at the Ilo station, with a coefficient of determination of 55.8%.

Improved Estimation of Missing Daily *in situ* SST Data at the Chicama, Chimbote, Huacho, Callao, and Ilo Coastal Stations

Figure 2 shows the lowest correlations between *in situ* and satellite-derived SST data observed at the Callao and Ilo stations. These low correlations appear to be associated with a different relationship between satellite and *in situ* SST under warm conditions or during El Niño events (highlighted by red dots in Fig. 2). To account for this non-linear relationship, a segmented linear regression approach was applied to the Callao and Ilo data. This method involves two linear segments, as mathematically described in the Methods section.

For the Callao station, the segmented regression analysis (Fig. 6) identified a breakpoint - marking a change in the slope of the regression - at 21.15 °C. Compared to the standard multiple linear regression model, the coefficient of determination (R^2) increased from 72% to 77.4% and the estimates were more in line with the observed *in situ* values (Fig. 7). In particular, the underestimation of SST values during the major El Niño events of 1982-83 and 1997-98 was reduced compared to the MLR model. Furthermore, the RMSE and MAD values obtained with the segmented approach were lower than those obtained with the MLR model (Table 3).

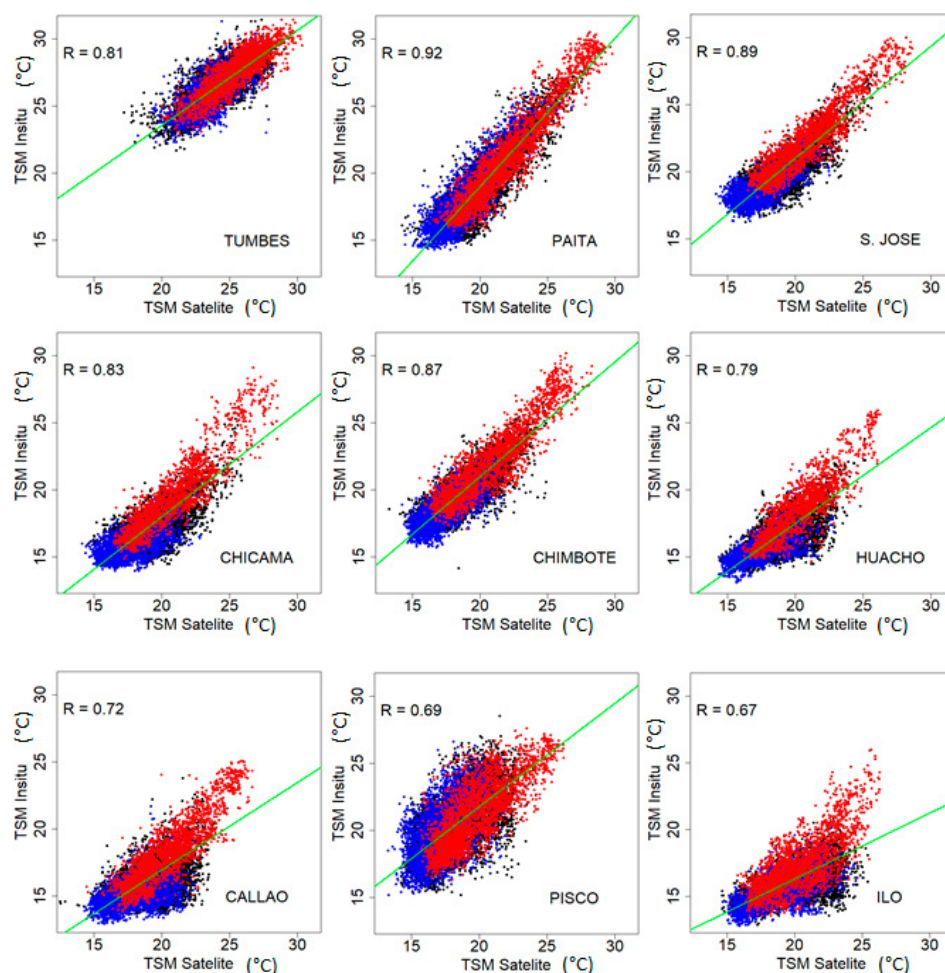


Figura 2.-Dispersión entre TSM *in-situ* y TSM satelital. Los puntos en color rojo corresponden a los periodos de condiciones cálidas del ITCP, mientras que los puntos en color azul corresponden a los periodos de condiciones frías del ITCP, y los puntos en color negro corresponden a los periodos de condiciones neutras

Figure 2. Dispersion between *in situ* SST and satellite SST. Red points correspond to periods of warm PCTI conditions, while blue points correspond to periods of cool PCTI conditions, and black points correspond to periods of neutral conditions

En la estación costera de Ilo, el análisis de la regresión por tramos (Fig. 8) estimó el umbral de cambio de pendiente de la regresión en 23,22 °C. El coeficiente de determinación (R^2) se incrementó de 55,8% a 57,9%, asimismo los estimados mostraron mejor aproximación hacia los valores *in situ* (Fig. 9). En unos casos particulares, se redujo la subestimación de los valores de TSM durante la ocurrencia de los eventos extraordinarios El Niño 1982-83 y 1997-98. Mientras que en otros casos la aplicación del método llevó a una ligera sobreestimación de la TSM en eventos EN más débiles (esto mismo ocurrió en Chicama, Huacho y Chimbote). Los valores de RMSE y DAM fueron menores a los estimados con Regresión Lineal Múltiple (Tabla 3).

For the Ilo coastal station, the segmented regression analysis (Fig. 8) identified a breakpoint - marking a change in the slope of the regression - at 23.22 °C. The coefficient of determination (R^2) increased from 55.8% to 57.9% and the estimates showed improved agreement with the *in situ* observations (Fig. 9). In some cases, the model significantly reduced the underestimation of SST values during the extreme El Niño events of 1982-83 and 1997-98. In other cases, however, the method led to a slight overestimation of SST during weaker El Niño events - a pattern also observed at Chicama, Huacho and Chimbote. Overall, both RMSE and MAD values were lower than those obtained using the standard multiple linear regression model (Table 3).

Tabla 2.- Comparación del desempeño de los modelos estadísticos (Regresión Lineal Univariada y Regresión Lineal Múltiple) en la estimación de datos faltantes. Indicadores de desempeño de la validación cruzada: coeficiente de determinación (R^2), la raíz del error cuadrático medio (RMSE) y Desviación Absoluta Media (DAM)

Table 2. Comparison of the performance of statistical models (Univariate Linear Regression and Multiple Linear Regression) for estimating missing data. Performance indicators from cross-validation include: coefficient of determination (R^2), root mean square error (RMSE), and mean absolute deviation (MAD)

Laboratorio Costero Coastal laboratory	Modelos estadísticos	R^2	RMSE	DAM
TUMBES	Regresión Lineal Univariada	65,4%	0,949	0,751
	Regresión Lineal Múltiple	72,9%	0,840	0,658
PAITA	Regresión Lineal Univariada	86,0%	1,156	0,890
	Regresión Lineal Múltiple	86,9%	1,120	0,864
SAN JOSE	Regresión Lineal Univariada	78,2%	0,984	0,784
	Regresión Lineal Múltiple	82,2%	0,889	0,704
CHICAMA	Regresión Lineal Univariada	70,4%	1,197	0,922
	Regresión Lineal Múltiple	79,4%	0,999	0,770
CHIMBOTE	Regresión Lineal Univariada	79,2%	1,047	0,821
	Regresión Lineal Múltiple	83,6%	0,929	0,718
HUACHO	Regresión Lineal Univariada	63,0%	1,116	0,834
	Regresión Lineal Múltiple	71,9%	0,973	0,736
CALLAO	Regresión Lineal Univariada	54,0%	1,344	1,030
	Regresión Lineal Múltiple	72,0%	1,049	0,774
PISCO	Regresión Lineal Univariada	50,4%	1,712	1,390
	Regresión Lineal Múltiple	61,0%	1,517	1,199
ILO	Regresión Lineal Univariada	45,9%	1,213	0,884
	Regresión Lineal Múltiple	55,8%	1,096	0,819

En el caso de la estación costera de Chicama, el análisis de la regresión por tramos (Fig. 10) estimó el umbral de cambio de pendiente de la regresión en 18,87 °C. El coeficiente de determinación (R^2) aumentó del 79,4 % al 82,4 %. Asimismo, los estimados mostraron una mejor aproximación a los valores *in situ* (Fig. 11). Por ejemplo, se observó una disminución de la subestimación de los valores de TSM durante el evento extraordinario de El Niño de 1997-1998. Al igual que en el caso de Ilo, se observó que, en otros casos, la aplicación del método llevó a una ligera sobreestimación de la TSM en eventos EN más débiles. Los valores de RMSE y DAM fueron inferiores a los estimados mediante Regresión Lineal Múltiple (Tabla 3).

En la estación costera de Huacho, el análisis de la regresión por tramos (Fig. 12) estimó el umbral de cambio de pendiente de la regresión en 21,3 °C. El coeficiente de determinación (R^2) se incrementó de 71,9% a 75,8% y los estimados mostraron mejor aproximación hacia los valores *in situ* (Fig. 13). Particularmente, se observó menor

In the case of the Chicama coastal station, the analysis of the regression by section (Fig. 10) estimated the threshold of change in the slope of the regression at 18.87 °C. The coefficient of determination (R^2) increased from 79.4% to 82.4%. The estimates also showed a better approximation to the *in situ* values (Fig. 11). For example, a decrease in the underestimation of SST values was observed during the exceptional El Niño event of 1997-1998. As in the case of Ilo, in other cases the application of the method led to a slight overestimation of SST during weaker EN events. The RMSE and DAM values were lower than those estimated by Multiple Linear Regression (Table 3).

In the Huacho coastal station, the segmented regression analysis (Fig. 12) identified a breakpoint - indicating a change in the regression slope - at 21.3 °C. The coefficient of determination (R^2) increased from 71.9% to 75.8%, and the model estimates showed a closer fit to the observed *in situ* SST values (Fig. 13). Notably, the underestimation

subestimación de los valores de TSM durante la ocurrencia del evento extraordinario El Niño 1997-98, en comparación con lo obtenido con Regresión Lineal Múltiple. También los valores de RMSE y DAM fueron menores a los estimados con Regresión Lineal Múltiple (Tabla 3).

Para el caso de la estación costera de Chimbote, el análisis de la regresión por tramos (Fig. 14) estimó el umbral de cambio de pendiente de la regresión en 20,4 °C. El coeficiente de determinación (R^2) se incrementó de 83,6% a 85,8% y los estimados también mostraron mejor aproximación hacia los valores *in situ* (Fig. 15). Es así que se observó que la subestimación de los valores de TSM fue menor durante la ocurrencia de los eventos extraordinarios El Niño 1982-83 y 1997-98, en comparación con lo obtenido con Regresión Lineal Múltiple, y en el caso de los valores de RMSE y DAM, también fueron menores a los estimados con Regresión Lineal Múltiple (Tabla 3).

of SST during the extreme 1997-98 El Niño event was reduced compared to the results obtained using the Multiple Linear Regression model. In addition, both the RMSE and MAD values were lower than those estimated with the Multiple Linear Regression approach (Table 3).

For the Chimbote coastal station, the segmented regression analysis (Fig. 14) estimated a breakpoint in the regression slope at 20.4 °C. The coefficient of determination (R^2) improved from 83.6% to 85.8%, and the model estimates showed a better alignment with the *in situ* SST observations (Fig. 15). Notably, the underestimation of SST during the major El Niño events of 1982-83 and 1997-98 was reduced compared to the results obtained using the Multiple Linear Regression model. RMSE and MAD values were also lower than those estimated with the Multiple Linear Regression approach (Table 3).

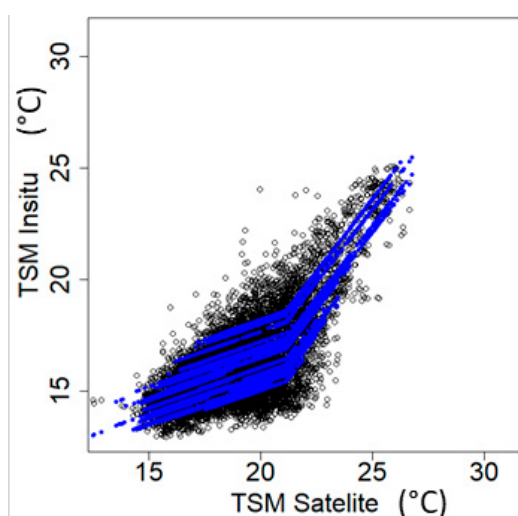


Figura 6.- Análisis de regresión por tramos de la TSM *in situ* (línea azul) del laboratorio costero de Callao y la TSM satelital

Figure 6. Segmented regression analysis of the *in situ* SST (blue line) of the coastal laboratory of Callao and satellite SST

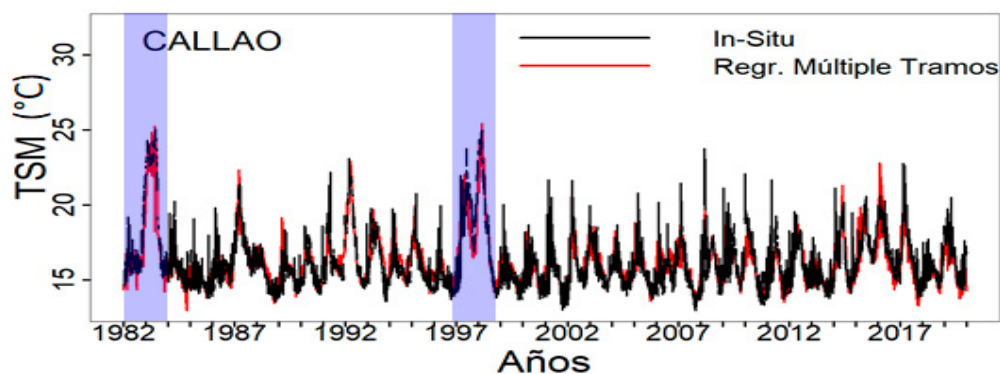


Figura 7.- Variación temporal de la TSM *in situ* (línea negra) del laboratorio costero de Callao y TSM estimado de una Regresión Múltiple por tramos (línea roja). Las áreas sombreadas de azul representan los periodos de ocurrencia de los eventos extraordinarios El Niño 1982-83 y 1997-98

Figure 7. Temporal variation of *in situ* SST (black line) of the coastal laboratory of Callao and SST estimated from a Segmented Multiple Regression (red line). The blue shaded areas represent the periods of occurrence of the extraordinary 1982-83 and 1997-98 El Niño events

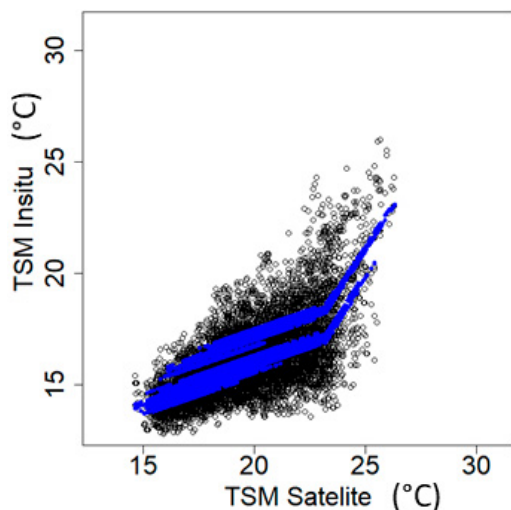


Figura 8.- Análisis de regresión por tramos de la TSM *in situ* (línea azul) del laboratorio costero de Ilo, y la TSM satelital

Figure 8. Segmented regression analysis of the *in situ* SST (blue line) of the coastal laboratory of Ilo, and satellite SST

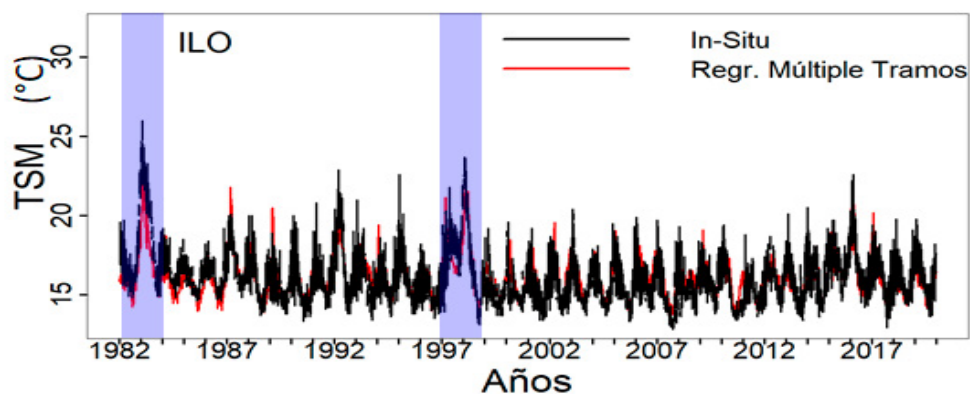


Figura 9.- Variación temporal de la TSM *in situ* (línea negra) del laboratorio costero de Ilo y TSM estimado de una Regresión Múltiple por tramos (línea roja). Las áreas sombreadas de azul representan los periodos de ocurrencia de los eventos extraordinarios El Niño 1982-83 y 1997- 98

Figure 9. Temporal variation of *in situ* SST (black line) of the Ilo coastal laboratory and SST estimated from a Multiple Regression by sections (red line). The blue shaded areas represent the periods of occurrence of the extraordinary El Niño 1982-83 and 1997-98 events

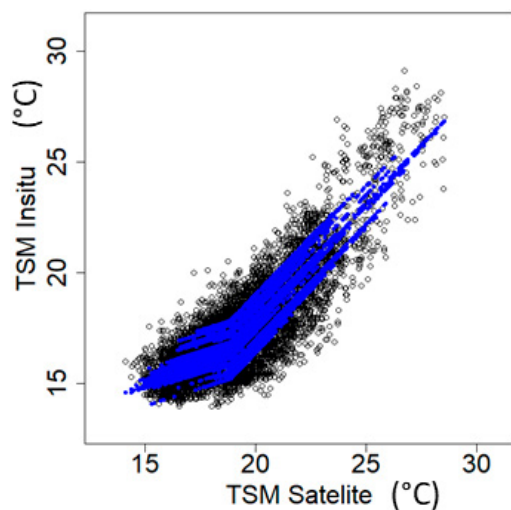


Figura 10.- Análisis de regresión por tramos de la TSM *in situ* (línea azul) del laboratorio costero de Chicama, y la TSM satelital

Figure 10. Segmented regression analysis of the *in situ* SST (blue line) of the coastal laboratory

Tabla 3.- Comparación del desempeño de los modelos estadísticos Regresión Lineal Múltiple y Regresión Múltiple por Tramos en la estimación de datos faltantes en Chicama, Chimbote, Huacho, Callao e Ilo. Indicadores de desempeño de la validación cruzada: Pvalor= 0; coeficiente de determinación (R^2), la raíz del error cuadrático medio (RMSE) y Desviación Absoluta Media (DAM)

Table 3. Comparison of the performance of statistical models—Multiple Linear Regression (MLR) and Segmented Multiple Regression (PMR)—in estimating missing data in Chicama, Chimbote, Huacho, Callao, and Ilo. Cross-validation performance indicators: p-value = 0; coefficient of determination (R^2), root mean square error (RMSE), and mean absolute deviation (MAD)

Laboratorio Costero Coastal laboratory	Modelos estadísticos Statistical model	R^2	RMSE	DAM
CHICAMA	Regresión Lineal Múltiple	79,4%	0,999	0,770
	Regresión Múltiple por Tramos	82,4%	0,922	0,717
CHIMBOTE	Regresión Lineal Múltiple	83,6%	0,929	0,718
	Regresión Múltiple por Tramos	85,8%	0,867	0,665
HUACHO	Regresión Lineal Múltiple	71,9%	0,973	0,736
	Regresión Múltiple por Tramos	75,8%	0,903	0,685
CALLAO	Regresión Lineal Múltiple	72,0%	1,049	0,774
	Regresión Múltiple por Tramos	77,4%	0,942	0,686
ILO	Regresión Lineal Múltiple	55,8%	1,096	0,819
	Regresión Múltiple por Tramos	57,9%	1,070	0,804

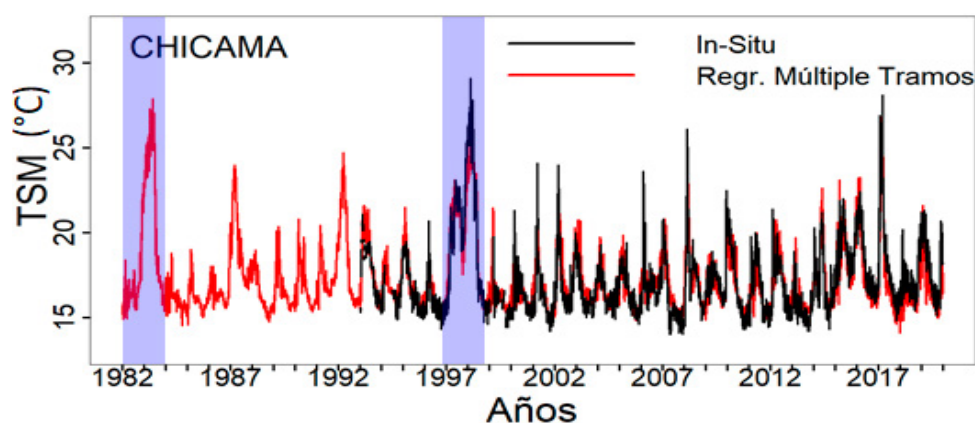


Figura 11.- Variación temporal de la TSM *in situ* (línea negra) del laboratorio costero de Chicama y TSM estimado de una Regresión Múltiple por tramos (línea roja). Las áreas sombreadas de azul representan los periodos de ocurrencia de los eventos extraordinarios El Niño 1982-83 y 1997-98

Figure 11. Temporal variation of *in situ* SST (black line) of the coastal laboratory of Chicama and SST estimated from a Segmented Multiple Regression (red line). The blue shaded areas represent the periods of occurrence of the extraordinary 1982-83 and 1997-98 El Niño events

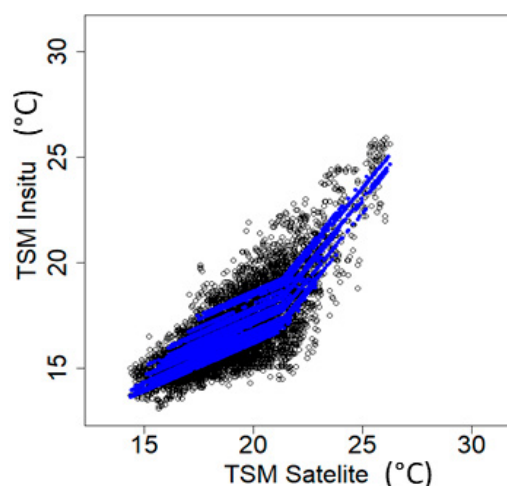


Figura 12.- Análisis de regresión por tramos de la TSM *in situ* (línea azul) del laboratorio costero de Huacho, y la TSM satelital

Figure 12. Segmented regression analysis of the *in situ* SST (blue line) of the coastal laboratory of Huacho, and satellite SST

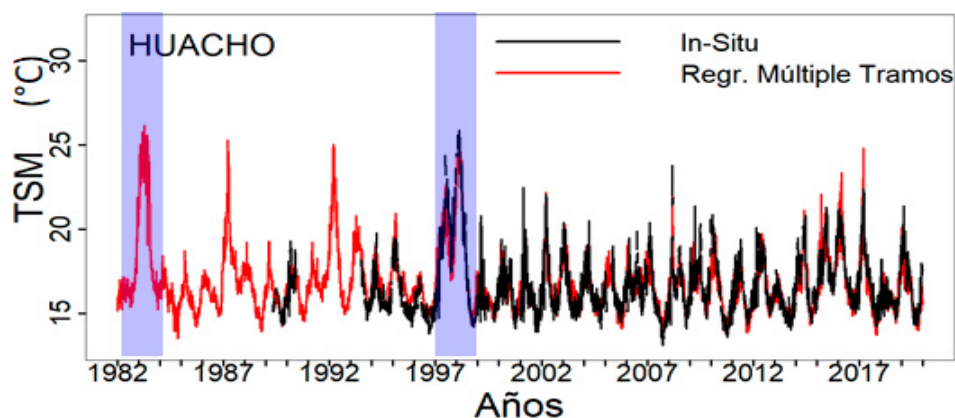


Figura 13.- Variación temporal de la TSM *in situ* (línea negra) del laboratorio costero de Huacho y TSM estimado de una Regresión Múltiple por tramos (línea roja). Las áreas sombreadas de azul representan los periodos de ocurrencia de los eventos extraordinarios El Niño 1982-83 y 1997-98

Figure 13. Temporal variation of *in situ* SST (black line) of the coastal laboratory of Huacho and SST estimated from a Segmented Multiple Regression (red line). The blue shaded areas represent the periods of occurrence of the extraordinary 1982-83 and 1997-98 El Niño events

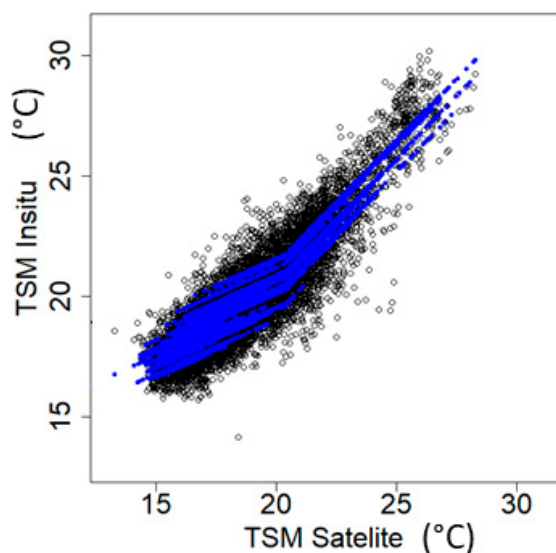


Figura 14.- Análisis de regresión por tramos de la TSM *in situ* (línea azul) del laboratorio costero de Chimbote, y la TSM satelital

Figure 14. Segmented regression analysis of the *in situ* SST (blue line) of the coastal labor

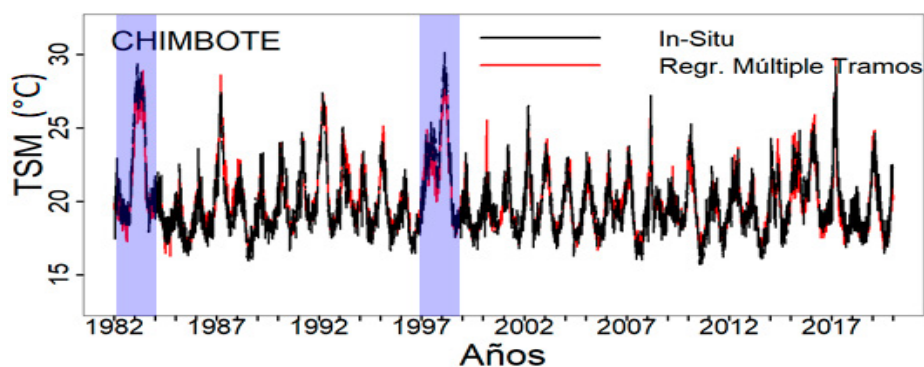


Figura 15.- Variación temporal de la TSM *in situ* (línea negra) del laboratorio costero de Chimbote y TSM estimado de una Regresión Múltiple por tramos (línea roja). Las áreas sombreadas de azul representan los periodos de ocurrencia de los eventos extraordinarios El Niño 1982-83 y 1997-98

Figure 15. Temporal variation of *in situ* SST (black line) of the coastal laboratory of Chimbote and SST estimated from a Segmented Multiple Regression (red line). The blue shaded areas represent the periods of occurrence of the extraordinary 1982-83 and 1997-98 El Niño events

Los análisis de las estaciones costeras de Chicama, Chimbote, Huacho, Callao e Ilo, señalan que los datos de TSM durante los años 1982-1983 y 1997-1998 estuvieron fuertemente influenciados por los eventos de El Niño extraordinario. Es por ello que los datos se ajustaron mejor a una Regresión Múltiple por tramos que a una Regresión Lineal Múltiple.

Fue relevante la inclusión de variables categóricas como el mes y las condiciones del ITCP para encontrar mejores modelos en las estimaciones de los valores *in situ*. La variable categórica del mes indicaría que habría diferentes contribuciones para las diferentes estaciones del año, mientras que las condiciones del ITCP, indicaría diferentes contribuciones en El Niño o La Niña.

Los datos de TSM diaria de la estación de Pisco tuvo los mayores errores debido al mayor grado de dispersión de pares ordenados entre la TSM *in situ* y satélite, así su correlación lineal fue una de las más bajas en comparación con las demás estaciones, no fue posible superar usando la regresión por tramos (RMT) debido a que no se observó un cambio de estructura (Fig. 2), similar a las estaciones de Chicama, Chimbote, Callao, Huacho e Ilo.

Por otro lado, el nivel de significancia de las correlaciones (R) y coeficientes de determinación (R^2) entre el dato *in situ* y satelital, nos indican un nivel de confianza (p valor) aceptable. Sin embargo, en la estación de Ilo, resulta tener los mayores errores, por lo que se podría sugerir que, para completar datos, en el futuro se podría hacer uso de otras bases satelitales como MUR (desde el 2003) y OSTIA (desde 2007). Cabe señalar, que el uso de la base de datos satelital del Producto de REYNOLDS *et al.* (2007) se debió a que los datos disponibles son desde 1982 en adelante.

4. CONCLUSIONES

Usando como predictor la TSM diaria satelital (AVHRR) en los modelos estadísticos lineales, se reconstruyeron las series de tiempo continuas de datos diarios en las estaciones costeras seleccionadas para el presente estudio, con coeficientes de determinación estadísticamente significativos.

The analyses for the coastal stations of Chicama, Chimbote, Huacho, Callao and Ilo indicate that the Sea Surface Temperature (SST) data for the years 1982-1983 and 1997-1998 were strongly influenced by the extreme EN events. As a result, SST values during these periods were better captured by a Segmented Multiple Regression model than by traditional Multiple Linear Regression.

The inclusion of categorical variables, such as month and PCTI, was found to be important in improving model performance in estimating *in situ* SST. The month variable is likely to reflect seasonal contributions, while the PCTI condition accounts for different effects under El Niño or La Niña phases.

The daily sea surface temperature (SST) data from the Pisco station exhibited the largest estimation errors, primarily due to the high dispersion between *in situ* and satellite-derived SST pairs. Consequently, its linear correlation was among the lowest of all stations analyzed. The segmented regression model (SRM) failed to improve the results, as no clear structural change was detected (Fig. 2), in contrast to what was observed at the Chicama, Chimbote, Callao, Huacho, and Ilo stations.

Nevertheless, the correlation coefficients (R) and coefficients of determination (R^2) between *in situ* and satellite data generally showed statistically significant relationships with acceptable confidence levels (p -values). However, the Ilo station also showed relatively high estimation errors. This suggests that future gap-filling efforts at this site could benefit from alternative satellite datasets, such as MUR (available since 2003) and OSTIA (available since 2007). It should be noted that the dataset used in this study - REYNOLDS *et al.* (2007) - was selected for its extensive historical coverage, starting in 1982.

4. CONCLUSIONS

Daily satellite-derived sea surface temperature (SST) data from AVHRR were used as predictors in linear statistical models to successfully reconstruct continuous daily time series at the selected coastal stations. These reconstructions yielded statistically significant coefficients of determination.

En la estimación de datos diarios faltantes (TSM), el modelo de Regresión Múltiple por tramos permitió obtener la reconstrucción de las series en las estaciones costeras Chicama, Chimbote, Huacho, Callao e Ilo. En las estaciones de Tumbes, Paita, San José y Pisco la reconstrucción fue con el modelo de Regresión Múltiple sin tramos.

Fue significativa la inclusión de variables categóricas como el mes y las condiciones del ITCP (cálido, frío, neutro) para encontrar los mejores modelos para las estimaciones de los valores *in situ*.

La estimación con Regresión Lineal Múltiple para las TSM de las estaciones costeras de Tumbes, Paita, San José y Pisco tuvieron como valores de Desviación Absoluta Media (DAM) o error medio absoluto a 0,65 °C; 0,86 °C; 0,70 °C y 1,19 °C, respectivamente. Las estaciones costeras de Chicama, Chimbote, Huacho, Callao e Ilo, usando Regresión Múltiple por tramos tuvieron valores de DAM igual a 0,71 °C; 0,66 °C; 0,68 °C; 0,68 °C y 0,8 °C, respectivamente. Las estimaciones diarias de valores faltantes en la estación costera de Pisco, tuvieron los mayores errores de estimación en comparación con el resto de las estaciones costeras.

La metodología empleada en este trabajo que combina el uso de datos satelitales de TSM diaria (AVHRR) y modelos estadísticos lineales adecuados, permitió reconstruir series de tiempo diarias para las estaciones costeras. Esta aproximación ha sido crucial para obtener estimaciones precisas de los datos faltantes, lo que facilita el análisis continuo de la TSM. Contar con series de tiempo diarias completas de TSM tiene importantes implicancias para el monitoreo de fenómenos climáticos y la toma de decisiones en áreas como el manejo de recursos pesqueros. De este modo, la metodología seguida no solo cumplió el objetivo de reconstruir las series de tiempo, sino que también proporciona una base para futuros estudios y aplicaciones en las ciencias marinas.

Agradecimientos

Los autores agradecen la disponibilidad de los datos de TSM de los laboratorios costeros proporcionados por el Laboratorio de Hidrofísica Marina del IMARPE y al personal del Laboratorio de Modelado Oceanográfico, Ecosistémico y del Cambio Climático por sus valiosos comentarios.

For the estimation of missing daily SST values, the segmented multiple regression model provided accurate reconstructions at the coastal stations of Chicama, Chimbote, Huacho, Callao, and Ilo. In contrast, reconstructions at the stations of Tumbes, Paita, San José, and Pisco were obtained using conventional multiple linear regression.

The inclusion of categorical variables such as month and PCTI conditions (warm, cold, neutral) significantly improved model performance, highlighting their importance in capturing the seasonal and interannual variability of *in situ* SST.

The multiple linear regression estimates of sea surface temperature (SST) at the coastal stations of Tumbes, Paita, San José and Pisco gave mean absolute deviation (MAD) values of 0.65 °C, 0.86 °C, 0.70 °C and 1.19 °C, respectively. In contrast, Chicama, Chimbote, Huacho, Callao and Ilo stations using segmented multiple regression models showed MAD values of 0.71 °C, 0.66 °C, 0.68 °C, 0.68 °C and 0.80 °C, respectively. Among all stations, the Pisco station had the highest estimation error in the reconstruction of missing daily SST data.

The methodology used in this study, which integrates daily SST satellite data (AVHRR) with appropriate linear statistical models, successfully reconstructed continuous daily time series for the coastal stations. This approach proved to be essential to obtain reliable estimates of missing values, thus allowing uninterrupted SST analysis. Having complete daily SST time series is highly valuable for monitoring climate variability and informing decision-making in areas such as fisheries management. Thus, the methodology not only achieved the goal of reconstructing the time series, but also provides a robust basis for future studies and applications in marine science.

Acknowledgments

The authors gratefully acknowledge the IMARPE's Marine Hydrophysics Laboratory for providing the SST data from the coastal monitoring stations, and thank the staff of the Oceanographic, Ecosystem, and Climate Change Modeling Laboratory for their valuable insights and comments.

5. REFERENCIAS

- BACHELERY, M.-L., ILLIG, S., & DADOU, I. (2016). Interannual variability in the South-East Atlantic Ocean, focusing on the Benguela Upwelling System: Remote versus local forcing. *Journal of Geophysical Research: Oceans*, 121, 284–310. <https://doi.org/10.1002/2015JC011168>
- CHAMBERS, J. M. (1992). Linear models. In J. M. Chambers & T. J. Hastie (Eds.), *Statistical Models* (Cap. 4). Routledge. <https://doi.org/10.1201/9780203738535>
- CHIN, T. M., VAZQUEZ-CUERVO, J. & ARMSTRONG, E. M. (2017). A multi-scale high-resolution analysis of global sea surface temperature. *Remote sensing of environment*, 200, 154–169. <https://doi.org/10.1016/j.rse.2017.07.029>
- DEMARCO, H. & FAURE V. (2000). Coastal upwelling and associated retention indices derived from satellite SST. Application to *Octopus vulgaris* recruitment. *Oceanologica Acta*, 23(4), 391–408. [https://doi.org/10.1016/S0399-1784\(00\)01113-0](https://doi.org/10.1016/S0399-1784(00)01113-0)
- DONLON, C. J., MARTIN, M., STARK, J., ROBERTS-JONES, J., FIEDLER, E. & WIMMER, W. (2012). The Operational Sea Surface Temperature and Sea Ice Analysis (OS-TIA) system. *Remote Sensing of Environment*, 116, 140–158. <https://doi.org/10.1016/j.rse.2010.10.017>
- GRADOS, C., CHAIGNEAU, A., ECHEVIN, V. & DOMINGUEZ, N. (2018). Upper ocean hydrology of the Northern Humboldt Current System at seasonal, interannual and interdecadal scales. *Progress in Oceanography*, 165, 123–144. <https://doi.org/10.1016/j.pocean.2018.05.005>
- HUANG, B., LIU, C., BANZON, V., FREEMAN, E., GRAHAM, G., HANKINS, B., SMITH, T. & ZHANG, H.-M. (2021). Improvements of the Daily Optimum Interpolation Sea Surface Temperature (DOISST) Version 2.1. *Journal of Climate*, 34, 2923–2939. <https://doi.org/10.1175/JCLI-D-20-0166.1>
- HYNDMAN, R. J. & ATHANASOPOULOS, G. (2018). *Forecasting: principles and practice* (2ª ed.). OTexts.
- ILLIG, S., DEWITTE, B., GOUBANOVA, K., CAMBON, G., BOUCHARÉL, J., MONETTI, F., ROMERO, C., PURCA, S. & FLORES, R. (2014). Forcing mechanisms of intraseasonal SST variability off Central Peru in 2000–2008. *Journal of Geophysical Research: Oceans*, 119, 3548–3573. <https://doi.org/10.1002/2013JC009779>
- MASON, S. J. & BADDOUR, O. (2008). Statistical modelling. In A. Troccoli, M. Harrison, D. L. T. Anderson & S. J. Mason (Eds.), *Seasonal climate: Forecasting and managing risk* (NATO Science Series 82, pp. 163–201). Springer. https://doi.org/10.1007/978-1-4020-6992-5_7
- MUGGEO, V. M. R. (2003). Estimating regression models with unknown break-points. *Statistics in Medicine*, 22, 3055–3071. <https://doi.org/10.1002/sim.1545>
- MINNETT, P. J. (2014). Sea Surface Temperature. In E. G. Njoku (Eds.), *Encyclopedia of Remote Sensing. Encyclopedia of Earth Sciences Series* (pp. 754–759). Springer. https://doi.org/10.1007/978-0-387-36699-9_166
- NOAA'S PACIFIC MARINE ENVIRONMENTAL LABORATORY [NOAA]. (2014). *Ferret* (Version 6.9) [Software]. National Oceanic and Atmospheric Administration. <https://ferret.pmel.noaa.gov/Ferret>
- NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION [NOAA]. (2025a). *El Niño regions*. National Weather Service Climate Prediction Center. Recuperado el 20 de marzo de 2025, de https://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/nino_regions.shtml
- NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION [NOAA]. (2025b). *Cold & Warm Episodes by Season*. National Weather Service Climate Prediction Center. Recuperado el 20 de marzo de 2025, de https://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php
- QUISPE-CCALLUARI, C., TAM, J., DEMARCO, H., CHAMORRO, A., ESPINOZA-MORRIBERÓN, D., ROMERO, C., DOMINGUEZ, N., RAMOS, J. & OLIVEROS-RAMOS, R. (2018). An index of coastal thermal effects of El Niño Southern Oscillation on the Peruvian Upwelling Ecosystem. *International Journal of Climatology*, 38(7), 3191–3201. <https://doi.org/10.1002/joc.5493>
- REYNOLDS, R. W., SMITH, T. M., LIU, C., CHELTON, D. B., CASEY, K. S. & SCHLAX, M. G. (2007). Daily high-resolution blended analyses for sea surface temperature. *Journal of Climate*, 20, 5473–5496. <https://doi.org/10.1175/2007JCLI1824.1>
- YU, J.-Y. & KIM, S. T. (2012). Identifying the types of major El Niño events since 1870. *International Journal of Climatology*, 33, 2105–2112. <https://doi.org/10.1002/joc.3575>