

Pontificia Universidad Javeriana Cali
Facultad de Ingeniería y Ciencias
Maestría en Ciencia de Datos
Proyecto Aplicado

Análisis Predictivo del Impacto del Fenómeno El Niño-Oscilación
Sur (ENOS) en el Nivel del Mar en el Pacífico Colombiano,
Empleando Modelos de Series de Tiempo y Algoritmos de
Machine Learning

Laura Marcela Vásquez López
Christian Bermúdez Rivas

Director: Dra. Sandra Milena Ramírez Buelvas

Julio 2025



Agradecimientos

Los autores desean expresar su más sincero agradecimiento a la directora de este trabajo, cuya orientación académica y conocimiento especializado fueron fundamentales para la adecuada conducción y culminación de esta investigación.

De igual manera, extendemos nuestro agradecimiento a los docentes del programa de Maestría en Ciencia de Datos de la Pontificia Universidad Javeriana Cali, por su formación, sus aportes conceptuales y su disponibilidad durante todo el proceso académico. Asimismo, agradecemos a los evaluadores de este trabajo por su revisión crítica y sus observaciones constructivas, que contribuyeron al fortalecimiento del documento final.

Agradecemos de manera especial al Centro de Investigaciones Oceanográficas e Hidrográficas del Pacífico (CCCP) y a la Dirección General Marítima (Dimar), por el acceso a los datos utilizados y por el respaldo institucional que hizo posible el desarrollo de esta investigación.

Finalmente, expresamos nuestra profunda gratitud a nuestras familias y amigos, cuyo acompañamiento emocional, apoyo incondicional y paciencia fueron pilares esenciales durante cada etapa de este proceso académico.

Resumen

Este estudio evaluó el impacto del fenómeno El Niño-Oscilación del Sur (ENOS) sobre la variabilidad del nivel del mar en el litoral Pacífico colombiano, con énfasis en las zonas urbanas de Ciudad Mutis, Buenaventura y Tumaco. Se implementó una metodología basada en el estándar *CRISPD-DM*, integrando técnicas de ciencia de datos, modelado geoestadístico y algoritmos de aprendizaje profundo orientados a series de tiempo, con el objetivo de generar escenarios predictivos de sobre-elevación del nivel del mar y analizar su impacto espacial sobre ecosistemas y asentamientos humanos vulnerables.

Para alcanzar este objetivo, se integraron múltiples fuentes de datos, incluyendo series de tiempo oceanográficas y atmosféricas (nivel del mar, temperatura superficial del mar, presión media a nivel del mar, viento superficial y precipitación acumulada), productos satelitales y de reanálisis (*CMEMS*, *ERA5*, *MSWEP*), así como datos altimétricos de alta resolución obtenidos mediante sensores como el *Light Detention and Ranging - LiDAR*. Las series temporales fueron sometidas a procesos de limpieza, transformación (análisis de anomalías, suavizado, estandarización) y análisis de correlación, tanto bivariado como multivariado. Posteriormente, se implementaron modelos de redes neuronales recurrentes tipo *Long Short-Term Memory (LSTM)*, entrenados con series multivariadas para predecir anomalías del nivel del mar asociadas a diferentes fases del ENOS.

Los resultados mostraron que los modelos *LSTM* alcanzaron un desempeño sobresaliente, con coeficientes de determinación (R^2) superiores a 0.920 y errores cuadráticos medios (*RMSE*) inferiores a 0.011 m en todas las localidades. El modelo multivariado (Modelo 3), que utilizó directamente las series de tiempo de las anomalías estandarizadas de las variables ambientales, superó en precisión a los modelos basados en componentes canónicos. Se identificaron asociaciones significativas entre el nivel del mar y los índices ENOS, especialmente Niño 3.4, así como con variables locales como la presión y el viento zonal. Los eventos El Niño Canónico y Costero generaron las mayores sobre-elevaciones (hasta 0.110 m), mientras que los eventos La Niña provocaron descensos moderados (hasta -0.060 m).

En paralelo, se generaron Modelos Digitales de Elevación (*MDE*) a partir de datos *LiDAR*, utilizando técnicas geoestadísticas (*kriging*) y métodos determinísticos (interpolación lineal, cúbica y por vecinos más cercanos). La interpolación lineal demostró ser la técnica más precisa y estable, con valores de R^2 superiores a 0.999 en las tres localidades. Estos *MDE* fueron fundamentales para la construcción de escenarios de inundación, definidos como la suma de la marea máxima observada (percentil 90), el *run-up* por oleaje (estimado mediante la fórmula de *Stockdon*) y la sobre-elevación inducida por eventos El Niño.

Los mapas de inundación generados evidenciaron impactos sobre ecosistemas de manglar y áreas urbanas de alta vulnerabilidad, particularmente en Tumaco y Buenaventura. En Ciudad Mutis, se observó una afectación considerable sobre zonas ribereñas y humedales costeros. Estos hallazgos resaltaron la utilidad de los modelos *LSTM* y del análisis geoespacial de alta resolución como herramientas técnicas para la predicción de riesgos costeros, la planificación territorial y la formulación de estrategias de adaptación frente a la variabilidad climática en regiones altamente expuestas.

Índice general

1. Introducción	13
2. Contextualización del Proyecto	17
2.1. Definición del Problema	17
2.1.1. Formulación	19
2.1.2. Sistematización	19
2.2. Objetivos	19
2.2.1. Objetivo General	19
2.2.2. Objetivos Específicos	20
2.3. Marco de Referencia	20
2.3.1. Marco Teórico	20
2.3.2. Antecedentes	22
3. Exploración y preparación de datos	27
3.1. Contexto	27
3.2. Fuentes de Datos	27
3.2.1. Series de tiempo oceanográficas y atmosféricas	28
3.2.2. Obtención de los datos <i>LiDAR</i>	28
3.3. Limpieza de Datos	35
3.3.1. Tratamiento de series de tiempo oceanográficas y atmosféricas	35
3.3.2. Tratamiento de los datos <i>LiDAR</i>	46
3.4. Exploración de Datos	46
3.4.1. Exploración estadística de series de tiempo oceanográficas y atmosféricas	46
3.4.2. Exploración estadística de los datos <i>LiDAR</i>	50
3.5. Transformación de Datos	62
3.5.1. Series de tiempo oceanográficas y atmosféricas	62
3.5.2. Conversión de alturas <i>LiDAR</i> a sistemas de referencia vertical: <i>MSL</i> y <i>LAT</i>	71
4. Modelado	73
4.1. Contexto	73
4.2. Modelado de series de tiempo	74
4.2.1. Selección de variables	74
4.2.2. Selección de algoritmos	85
4.2.3. Parametrización de algoritmos	86
4.2.4. Implementación de los modelos <i>LSTM</i>	88
4.3. Modelos Digitales de Elevación	93
4.3.1. Selección de algoritmos	93

4.3.2. Parametrización de los modelos	96
4.3.3. Implementación de los modelos	101
5. Evaluación	125
5.1. Contexto	125
5.2. Métricas de evaluación de modelos	125
5.2.1. Validación de los modelos de series de tiempo	125
5.2.2. Validación de los modelos digitales de terreno	132
6. Resultados y Discusión	145
6.1. Contexto	145
6.2. Escenarios predictivos sobre la elevación del nivel del mar y su relación con el ENOS	145
6.2.1. Modelación espacial del impacto del nivel del mar sobre los principales pobla- dos del pacífico colombiano	157
7. Conclusiones y Trabajos Futuros	165
7.1. Conclusiones	165
7.2. Trabajos futuros	168
Bibliografía	169

Introducción

La oscilación climática El Niño-Oscilación Sur (ENOS) es un fenómeno de gran relevancia que genera variaciones interanuales significativas en las condiciones atmosféricas y oceánicas a escala global [1], [2], [3], [4]. Esta oscilación incluye una fase cálida, conocida como El Niño, que se caracteriza por el debilitamiento de los vientos alisios y la reducción de la surgencia costera en el Pacífico ecuatorial oriental, lo que ocasiona un calentamiento anómalo del océano [3], [5]. Por el contrario, durante la fase fría, denominada La Niña, el fortalecimiento de los vientos alisios intensifica la surgencia de aguas frías, disminuyendo la temperatura superficial del mar en el Pacífico oriental, con impactos climáticos significativos en distintas regiones del mundo [6].

A nivel global, el ENOS modifica los patrones de precipitación [6], [7], [8], alterando las condiciones climáticas en regiones distantes del Pacífico tropical. Durante El Niño, por ejemplo, el suroeste de los Estados Unidos y el sur de Sudamérica tienden a experimentar un aumento en las precipitaciones, mientras que el sudeste asiático y Australia suelen registrar sequías severas [9], [10]. Por el contrario, La Niña se asocia con un incremento de las lluvias en el sudeste asiático y Australia, y una disminución en el suroeste de los Estados Unidos y partes de América del Sur [6].

El ENOS también afecta el nivel del mar mediante mecanismos de expansión o contracción térmica del agua oceánica. Durante El Niño, el calentamiento de la superficie oceánica induce una expansión térmica, elevando el nivel del mar en diversas regiones, mientras que La Niña puede producir contracciones térmicas que reducen dicho nivel [11], [12]. Estos cambios son especialmente relevantes para regiones costeras bajas, como el Pacífico colombiano, donde se han observado impactos como inundaciones, erosión costera y salinización de acuíferos [13], [14].

El Pacífico colombiano concentra su población en tres principales núcleos urbanos: Buenaventura, Tumaco y Ciudad Mutis (Bahía Solano), donde reside aproximadamente el 90 % de la población regional [13]. Estas comunidades, en su mayoría asentadas en condiciones de pobreza, viven en viviendas palafíticas construidas sobre el mar con materiales precarios, lo que las hace altamente vulnerables a la erosión costera y al aumento del nivel del mar [15].

Este proyecto tuvo como objetivo comprender el impacto del fenómeno ENOS en la variabilidad del nivel del mar. Para ello, se emplearon modelos de series de tiempo y algoritmos de *machine learning* que permiten pronosticar el comportamiento del nivel del mar durante distintas fases del fenómeno. Esta aproximación permitió construir modelos espaciales integrados en un Sistema de Información Geográfica (SIG), útiles para identificar y proyectar riesgos de inundación y erosión costera vinculados a fluctuaciones en la cota de inundación.

La metodología contempla la recopilación y análisis de datos históricos de variables océano-atmosféricas (temperatura superficial del mar, precipitaciones, vientos, presión atmosférica, oleaje y nivel del mar), los cuales fueron procesados mediante algoritmos de *machine learning* diseñados

para modelar dinámicas temporales (series de tiempo), destacándose el uso de redes neuronales recurrentes tipo *Long Short-Term Memory (LSTM)* [16], [17], [18], por su capacidad para capturar dependencias a largo plazo en series de tiempo complejas. Se buscó identificar patrones que relacionaran estas variables con la manifestación de diferentes tipos de eventos El Niño (Canónico, Modoki y Costero) [19], [20], y analizar su influencia sobre el nivel del mar en el Pacífico colombiano.

Para cuantificar el impacto de estos cambios en el nivel del mar sobre la costa Pacífica colombiana, se desarrollaron modelos digitales de elevación (*MDE*) a partir de datos de sensores remotos como el *Light Detection and Ranging (LiDAR)*, el cual consiste en una tecnología activa de tele-detección que permite la adquisición precisa de datos geospaciales tridimensionales, mediante la emisión de pulsos de luz láser. Estos pulsos, generalmente en el espectro infrarrojo cercano, son emitidos a alta frecuencia desde una plataforma móvil, como aeronaves, vehículos terrestres o drones, y se reflejan en las superficies del terreno y objetos presentes [21]. Al medir el tiempo que tarda cada pulso (emisor) en regresar al sensor (receptor), se calcula la distancia entre el sensor y el objeto, lo que permite generar nubes de puntos tridimensionales con coordenadas georreferenciadas (x , y , z) de alta densidad y precisión con el uso de un sistema de posicionamiento global (*GNSS*) [21], [22]. La integración de estos componentes permite obtener mediciones exactas incluso en condiciones topográficas complejas o en áreas con vegetación densa [21], [22].

Los datos obtenidos mediante *LiDAR* son procesados para generar diversos productos cartográficos, como Modelos Digitales de Elevación (*MDE*), Modelos Digitales de Terreno (*MDT*) y Modelos Digitales de Superficie (*MDS*). Estos productos son fundamentales en aplicaciones que requieren un conocimiento detallado de la morfología del terreno, como la planificación urbana, la ingeniería civil, la gestión de recursos naturales y la evaluación de riesgos naturales [21], [22].

Los *MDE* derivados de *LiDAR* ofrecen una resolución espacial significativamente superior a la de otros métodos de adquisición de datos topográficos, lo cual es clave para identificar zonas vulnerables a la inundación costera, modelar escenarios de intrusión marina y estimar posibles impactos sobre infraestructura y ecosistemas. En este estudio, se abordó la problemática del ascenso del nivel del mar inducido por el ENOS en el Pacífico colombiano, integrando aprendizaje automático enfocado en series de tiempo y modelación espacial de alta resolución, desarrollando una metodología que combina la recopilación y procesamiento de variables océano-atmosféricas con el uso de redes neuronales LSTM para cuantificar las sobreelevaciones del nivel del mar durante diferentes tipos de eventos El Niño (Canónico, Modoki y Costero). Estos valores, una vez integrados con los máximos históricos de marea y el *run-up* por oleaje, permitieron estimar cotas de inundación por localidad, que fueron proyectadas sobre Modelos Digitales de Elevación derivados de datos *LiDAR*, obteniendo así escenarios centimétricos de exposición costera.

El documento se estructura conforme a la metodología CRISP-DM (*Cross Industry Standard Process for Data Mining*), abordando de forma sistemática cada una de sus fases. El Capítulo 2 corresponde a la etapa de comprensión del negocio, donde se contextualiza la problemática, se formulan los objetivos del proyecto y se presenta el marco de referencia teórico y conceptual. El Capítulo 3 desarrolla la fase de comprensión y preparación de los datos, detallando las fuentes, técnicas de limpieza, transformación y análisis exploratorio, tanto para las series oceánicas y atmosféricas como para los modelos topográficos *LiDAR*.

El Capítulo 4 aborda la fase de modelado, en la cual se implementan redes neuronales *LSTM* para la predicción del nivel del mar y se generan modelos digitales de elevación mediante geoestadística. A continuación, el Capítulo 5 presenta la fase de evaluación de los modelos predictivos y espaciales, mediante métricas estadísticas y validaciones específicas. En el Capítulo 6 se desarrolla la fase de resultados y discusión, presentando los escenarios predictivos de sobreelevación y sus efectos espaciales sobre las zonas urbanas costeras de Ciudad Mutis, Buenaventura y Tumaco. Finalmente, el Capítulo 7 expone las conclusiones generales del estudio y plantea recomendaciones y líneas de trabajo futuras orientadas a la gestión del riesgo climático y la planificación territorial adaptativa.

Contextualización del Proyecto

2.1. Definición del Problema

El fenómeno ENOS constituye una de las principales fuentes de variabilidad interanual del sistema climático global. Se caracteriza por un calentamiento anómalo de las aguas superficiales y subsuperficiales del océano Pacífico ecuatorial [2]. Según su origen y comportamiento, se reconocen distintas variantes del evento, Canónica, Modoki y Costera, cada una con particularidades que generan alteraciones significativas en el sistema océano-atmósfera, afectando variables como la temperatura superficial del mar, el nivel del mar, los vientos, la presión atmosférica y las precipitaciones [23], [24], [25], [26].

En el contexto de la costa Pacífica colombiana, estudios previos han evidenciado que estos eventos pueden provocar tanto elevaciones como descensos en el nivel del mar, dependiendo de si se trata de la fase cálida (El Niño) o fría (La Niña), comúnmente asociados con la propagación de la Onda Kelvin Ecuatorial (OKE) [27]. Dichas alteraciones impactan de forma directa las zonas costeras de la región [28], [29], [30].

Actualmente, la región enfrenta riesgos crecientes de inundación y erosión costera, agudizados por el ascenso del nivel del mar, los efectos de ENOS, las condiciones climáticas locales y otros fenómenos de distinta escala. Este problema representa una amenaza crítica para las comunidades costeras, la infraestructura y los ecosistemas estratégicos de la región, y su monitoreo resulta fundamental para la planificación de proyectos costeros y operaciones portuarias [31].

Entre los principales riesgos se encuentra la salinización de los acuíferos costeros —de los que dependen comunidades como Tumaco y Guapi ante la limitada cobertura de sistemas de acueducto—, así como la pérdida de cobertura de manglares y bosques inundables. Estos ecosistemas cumplen un rol vital para la reproducción y el ciclo de vida de recursos hidrobiológicos clave como la pian-güa, el camarón y peces pelágicos, los cuales dependen de estos hábitats en sus primeras etapas de desarrollo [32].

Frente a este panorama, se requiere un análisis exhaustivo que vincule los eventos ENOS con los cambios específicos en las condiciones océano-atmosféricas, considerando también el efecto combinado de las mareas y el oleaje local. Este proyecto, con un enfoque integrado desde la ciencia de datos, se propuso cuantificar dichos efectos y desarrollar modelos predictivos (LSTM) que permitan comprender mejor las implicaciones de El Niño en el Pacífico colombiano, evaluando particularmente sus impactos en áreas como Ciudad Mutis, Buenaventura y Tumaco (Figura 2.1) desde un enfoque geoestadístico.

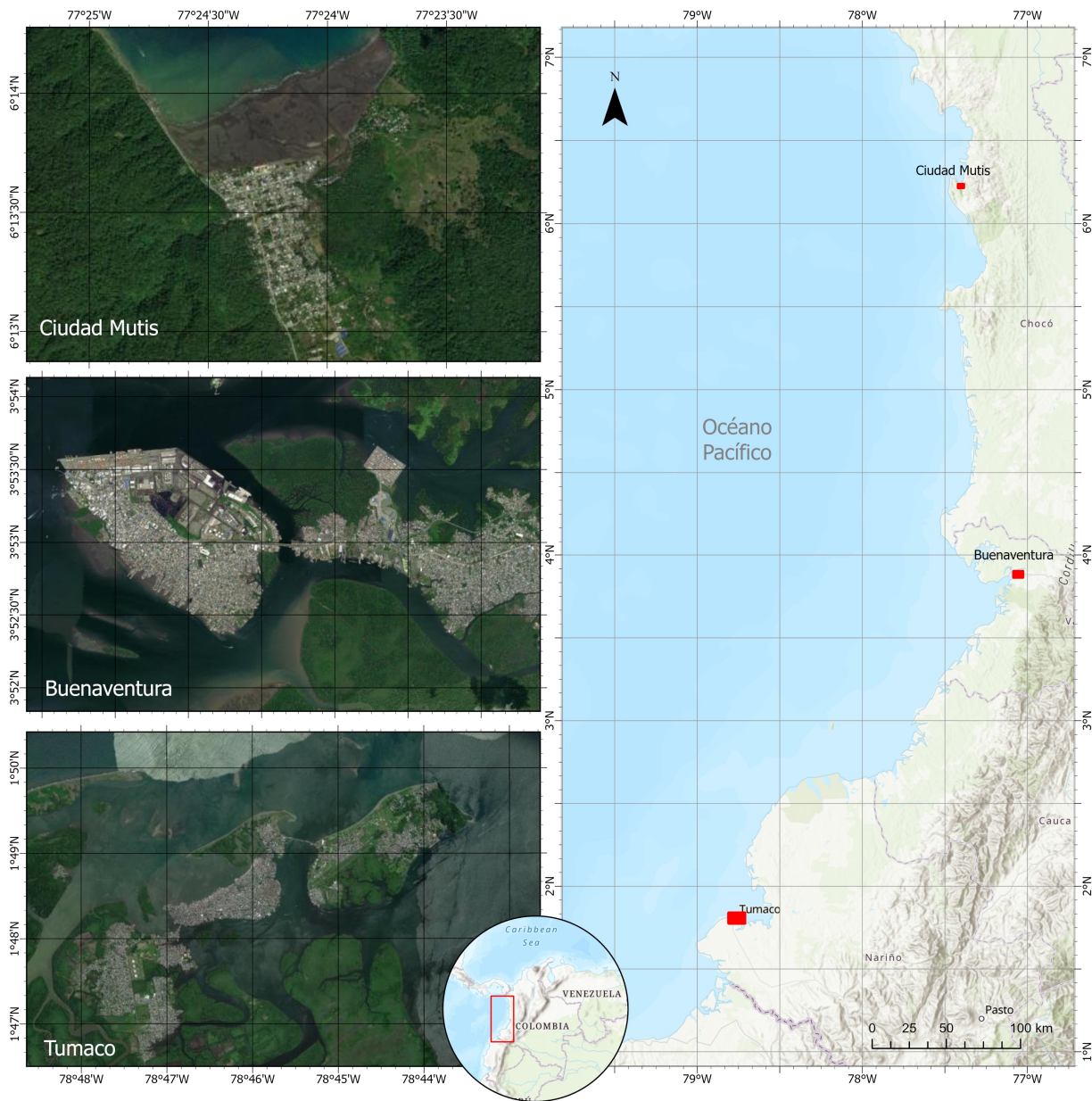


Figura 2.1: Áreas de estudio para la evaluación de los impactos del ascenso del nivel del mar (Ciudad Mutis, Buenaventura y Tumaco).

2.1.1. Formulación

La pregunta central de este proyecto fue: ¿Cómo afecta el fenómeno El Niño, en sus distintas variantes, el nivel del mar en el Pacífico colombiano, y cuáles son los impactos esperados de esta sobreelevación en combinación con las condiciones oceanográficas locales en los cascos urbanos de Ciudad Mutis, Buenaventura y Tumaco?

2.1.2. Sistematización

Para responder esta pregunta principal, se derivaron las siguientes preguntas de sistematización:

- ¿Qué patrones de asociación existen entre las variables océano-atmosféricas (temperatura superficial del mar, nivel del mar, vientos y presión atmosférica) y las diferentes génesis de El Niño (Canónico, Modoki y Costero)?
- ¿Cómo se comporta el nivel del mar bajo los diferentes escenarios ENOS (Niña y Niño) en relación con su génesis (Modoki, Canónico y Costero)?
- ¿Al relacionar el nivel del mar con otras variables se puede predecir el comportamiento del nivel del mar por influencia climática y sus efectos en la costa?
- ¿Cómo afecta cada uno de estos escenarios predictivos de la variación del nivel del mar a la costa del Pacífico colombiano (Ciudad Mutis, Buenaventura y Tumaco), en combinación con las condiciones oceanográficas locales?

Estas preguntas orientaron la estructuración del análisis y la construcción de modelos predictivos enfocados en evaluar el impacto del fenómeno El Niño sobre las tres principales localidades de la región, proporcionando así una comprensión más precisa del riesgo de inundación asociado a un evento de variabilidad climática, lo que a su vez proporciona insumos técnicos clave para la formulación de posibles estrategias de mitigación y adaptación en el territorio.

2.2. Objetivos

2.2.1. Objetivo General

Evaluar el impacto del fenómeno de variabilidad climática El Niño-Oscilación Sur (ENOS) sobre el nivel del mar, en la cuenca del Pacífico colombiano y analizar los efectos combinados de esta sobreelevación con las condiciones oceanográficas locales y sus impactos en la costa, utilizando modelos de series de tiempo y algoritmos de *machine learning*, para determinar los riesgos de inundaciones en la región.

2.2.2. Objetivos Específicos

- Identificar los patrones de asociación entre las variables océano-atmosféricas (temperatura superficial del mar, nivel del mar, vientos y presión atmosférica) y las distintas variantes del fenómeno de El Niño (Canónica, Modoki y Costera) mediante el uso de modelos de series de tiempo y algoritmos de *machine learning* orientados a las series de tiempo.
- Desarrollar escenarios predictivos sobre la elevación del nivel del mar a partir de los patrones de correlación identificados para cada tipo de El Niño, utilizando modelos de series de tiempo y algoritmos de *machine learning* orientados a las series de tiempo.
- Evaluar el impacto de los escenarios predictivos de las variaciones del nivel del mar (expansión o contracción) sobre la costa del Pacífico colombiano en Ciudad Mutis, Buenaventura y Tumaco, y su relación con las condiciones oceanográficas locales (mareas y oleaje), a partir de modelos digitales de terreno y la distribución de los ecosistemas costeros.

2.3. Marco de Referencia

El fenómeno ENOS es una de las mayores fuentes de variabilidad climática interanual, afectando significativamente las condiciones atmosféricas y oceánicas a nivel global. El objetivo de esta iniciativa de investigación se basa en evaluar y pronosticar el impacto de las distintas variantes del ENOS (Canónica, Modoki y Costera) sobre el nivel del mar en la costa del Pacífico colombiano, utilizando modelos de *machine learning* basados en series de tiempo. Dentro del marco referencial, se explorarán los principios fundamentales del ENOS, su mecanismo de acción, y cómo sus variantes afectan las variables océano-atmosféricas clave, tales como la temperatura superficial del mar, el nivel del mar, la presión atmosférica y los vientos en superficie; así como los antecedentes, en los cuales se revisarán estudios previos y casos específicos que demuestran la relevancia y los efectos de este fenómeno en la costa del Pacífico colombiano. Este enfoque permitió desarrollar modelos predictivos, que facilitaron una comprensión del riesgo al que están expuestas estas poblaciones, proporcionando herramientas técnico-científicas para la elaboración de estrategias efectivas de mitigación y adaptación frente a la sobre-elevación del nivel de la mar inducida por el ENOS.

2.3.1. Marco Teórico

La atmósfera terrestre es un sistema dinámico compuesto por una mezcla de gases que rodean nuestro planeta. Este sistema es fundamental para la distribución del calor, la humedad y otros elementos esenciales que determinan el clima global [28]. La circulación atmosférica es impulsada principalmente por el calentamiento diferencial de la superficie terrestre, entre el Ecuador y los polos. Este gradiente térmico genera diferencias de presión que impulsan las masas de aire, creando patrones de movimiento a gran escala conocidos como celdas de circulación [33].

Una contribución fundamental a la comprensión de estas dinámicas fue realizada por *Sir* Gilbert Walker, quien, al estudiar los monzones de la India, identificó la existencia de correlaciones entre presiones atmosféricas en regiones distantes del planeta, particularmente entre el océano Índico y

el Pacífico. A partir de estos hallazgos se definió la Oscilación del Sur, medida por el Índice de Oscilación del Sur (SOI), que capta la diferencia de presión entre Tahití y Darwin [2], [34]. Estas conexiones a larga distancia, conocidas como teleconexiones, son claves para entender fenómenos climáticos como el ENOS que alteran patrones globales.

En particular, la atmósfera del planeta tierra no solo presenta patrones constantes de circulación, sino también una variabilidad climática significativa que se manifiesta en múltiples escalas temporales y espaciales [2]. La variabilidad climática se refiere a las fluctuaciones naturales del clima alrededor de su estado promedio debido a factores internos y externos del sistema climático [35], [2]. Estas fluctuaciones pueden ser observadas en cambios anuales de temperatura, precipitación y patrones de viento. A nivel global, estos cambios son impulsados por variaciones en la energía solar recibida y la dinámica interna del sistema climático, que incluye la interacción entre la atmósfera, los océanos, la criósfera y la biosfera [36].

En este contexto, las oscilaciones climáticas, que son variaciones periódicas o cuasi-periódicas en el sistema climático, que pueden durar desde meses hasta décadas, juegan un papel importante. Dentro de estas se incluyen fenómenos notables como El Niño/La Niña -Oscilación del Sur (ENOS), la Oscilación del Atlántico Norte (NAO), la Oscilación Decadal del Pacífico (PDO), entre otras [2], [36]. Estas oscilaciones tienen la capacidad de modificar significativamente los patrones climáticos regionales y globales, afectando aspectos como las precipitaciones, las temperaturas y la frecuencia e intensidad de eventos extremos [37]. Por tanto, es fundamental comprender esta variabilidad climática para predecir y mitigar los impactos en diferentes regiones del mundo.

El fenómeno de variabilidad climática interanual ENOS, tiene un impacto significativo en los patrones climáticos a escala mundial. Este fenómeno se caracteriza por el calentamiento anómalo de las aguas superficiales del océano Pacífico ecuatorial central y oriental, que resulta en cambios drásticos en la circulación atmosférica y oceánica [34]. Este evento, según su génesis, puede clasificarse en tres variantes. El Niño Canónico se origina en el Pacífico oriental, frente a las costas de Perú y Ecuador, con una propagación hacia el oeste. El Niño Modoki se caracteriza por un calentamiento anómalo en el Pacífico central, mientras que en el este y oeste se presentan anomalías frías o neutras. Por último, el Niño Costero corresponde a un calentamiento localizado en el Pacífico oriental cercano a la costa de Perú y Ecuador y puede propagarse hacia el oeste. Cada una de estas modalidades presenta impactos climáticos diferenciados en términos de intensidad, extensión espacial y efectos regionales [2], [38], [39], [40].

La periodicidad del fenómeno El Niño oscila entre 2 y 7 años, con una duración promedio cercana a los 10 meses. Estos eventos, comúnmente asociados al paso de Ondas Kelvin, generan alteraciones significativas en variables ambientales como la temperatura del mar, los patrones de precipitación y la presión atmosférica. Además de sus efectos climáticos, El Niño puede inducir modificaciones temporales en la estructura biológica marina, provocar migraciones de especies y alterar la geomorfología costera mediante eventos de inundación. Asimismo, sus impactos se extienden a sectores como la agricultura, con el aumento de sequías severas, la pesca y las economías tanto locales como globales [2].

Para comprender con mayor profundidad los impactos del fenómeno El Niño sobre las variables climatológicas y oceanográficas, se recurre a una combinación de técnicas de observación remota

y mediciones directas. En este contexto, los satélites desempeñan un papel fundamental al ofrecer registros continuos y de alta resolución, tanto temporal como espacial, de parámetros clave como la temperatura superficial del mar, la altura del nivel del mar y precipitaciones [1]. Estos datos son esenciales para identificar patrones de anomalía asociados a la dinámica del ENOS y fortalecer la capacidad de monitoreo a gran escala.

Complementariamente, los cruceros oceanográficos permiten la adquisición de datos *in situ*, facilitando mediciones detalladas de temperatura, salinidad y corrientes oceánicas en diferentes niveles de profundidad mediante el uso de instrumentos como los sensores de Conductividad-Temperatura-Profundidad (CTD) [41]. A su vez, los mareógrafos constituyen una herramienta clave para registrar variaciones del nivel del mar en zonas costeras, mientras que las boyas de observación automática, como el sistema TAO/TRITON, ofrecen monitoreo en tiempo real de las condiciones oceánicas y atmosféricas en el Pacífico tropical [42], [43]. Estas observaciones son integradas con salidas de modelos numéricos y redes internacionales de intercambio de datos, garantizando una base sólida para la investigación científica y la elaboración de proyecciones robustas sobre la evolución del fenómeno ENOS.

2.3.2. Antecedentes

El estudio de la forma de la Tierra es una de las disciplinas más amplias e importantes para la planificación espacial, la evaluación de riesgos e impactos naturales y antrópicos, y la conservación del medio ambiente. Dentro de esta disciplina, la medición del nivel medio del mar proporciona un estimado crucial del comportamiento del sistema climático.

Los cambios en el nivel medio del mar pueden ser provocados por diversos factores, como el movimiento de las placas tectónicas, el oleaje, el viento, las mareas, el derretimiento de las grandes capas de hielo, los cambios en los reservorios de agua terrestres y la expansión térmica [44]. Comprender la relación entre el nivel del mar y el clima es esencial, especialmente para pronosticar los impactos que estos cambios pueden tener debido a las oscilaciones climáticas [45].

El estudio del nivel del mar ha requerido el uso de diversas técnicas a lo largo del tiempo. Desde el siglo XVIII, los mareógrafos han constituido el principal instrumento para su monitoreo [45]. No obstante, esta metodología presenta limitaciones importantes, entre ellas la inconsistencia en los registros en múltiples regiones del mundo, debido a la necesidad de mantenimiento frecuente y a la obsolescencia de algunos equipos, que impide la recopilación continua de datos a largo plazo. Además, los mareógrafos registran el nivel del mar en función de un punto fijo sobre la superficie terrestre, lo que los hace vulnerables a desplazamientos verticales de la corteza terrestre, generando posibles distorsiones en las mediciones [45], [44].

Para superar las limitaciones inherentes a los mareógrafos, se han desarrollado y adoptado tecnologías más avanzadas, entre las cuales destaca la altimetría satelital. Esta técnica permite obtener mediciones del nivel del mar con mayor precisión, coherencia y cobertura espacial a escala global. La combinación de estas observaciones modernas con los registros históricos de los mareógrafos ha permitido construir series de tiempo más robustas, proporcionando una visión integral y detallada de la evolución del nivel del mar. Desde la década de 1990, misiones espaciales como TOPEX/Poseidón

[46], Jason-1 [47], GRACE [48], Envisat [49] y OSTM/Jason-2 [50] han generado registros continuos sobre la altitud de la superficie oceánica, constituyendo una fuente clave para el análisis de los patrones y tendencias del nivel del mar a largo plazo.

El análisis de estos datos ha permitido a numerosos estudios documentar un ascenso sostenido del nivel medio del mar a nivel global, atribuido en gran parte a la expansión térmica de los océanos [51], [52], [53], [54]. No obstante, también se ha evidenciado una marcada variabilidad regional, con tasas de aumento significativamente diferentes entre distintas zonas geográficas. Por ejemplo, en el mar del Sur de China se ha registrado una tasa de ascenso de 5.5 mm/año, muy superior al promedio global [55]. Estas diferencias responden a múltiples factores, entre ellos las variaciones en la temperatura superficial del mar, los patrones de presión atmosférica y la redistribución de masas de agua oceánica [11], [56], [57], [58].

Este tipo de hallazgos resalta la importancia de continuar fortaleciendo los sistemas de observación satelital y de avanzar en el entendimiento de los procesos que explican la variabilidad del nivel del mar, tanto a escala global como regional. En este sentido, estudios realizados en Colombia han aportado evidencia sobre la relación entre el fenómeno ENOS y las fluctuaciones del nivel del mar. Por ejemplo, Poveda et al. [59] correlacionaron registros históricos de 50 ríos del país con el Índice de Oscilación del Sur (SOI), confirmando la influencia de las fases extremas del fenómeno sobre la hidrología nacional. De manera complementaria, López et al. [60] analizaron series de nivel relativo del mar y las compararon con el SOI, identificando anomalías de hasta 30 cm durante eventos de El Niño y estimando tasas de incremento y descenso en el delta del río Mira. Por su parte, Rodríguez-Rubio [61] propuso un índice multivariado para la costa occidental de Colombia, basado en series de temperatura superficial del mar, precipitación y temperatura ambiente, que se ha convertido en un referente para el seguimiento y vigilancia del ENOS en el Pacífico sur colombiano.

En los últimos años, el análisis del impacto de El Niño sobre el nivel del mar ha incorporado enfoques basados en ciencia de datos, destacándose el uso creciente de algoritmos de machine learning para mejorar la capacidad de predicción y alerta temprana. Por ejemplo, Lai et al. [62] emplearon algoritmos como *Support Vector Machines (SVM)* y *Genetic Programming (GP)* para desarrollar un sistema de alerta ante inundaciones derivadas de la variabilidad climática en Malasia, logrando predicciones altamente precisas del nivel medio del mar. Nieves et al. [63] demostraron el potencial de modelos como *Gaussian Processes (GP)* y redes neuronales recurrentes (*RNN*) para anticipar la variabilidad regional del nivel del mar en horizontes temporales de meses a años. De forma similar, Hazrin et al. [64] implementaron seis algoritmos, entre ellos regresión lineal, *SVM*, árboles de regresión, redes neuronales artificiales y regresión por procesos gaussianos, para realizar predicciones diarias en las costas de Malasia, contribuyendo al diseño de estrategias de mitigación frente a fenómenos climáticos extremos.

Muslim et al. [65], estudiaron la influencia de los parámetros meteorológicos en la exactitud sobre la predicción de nivel del mar en Sabah, Malasia, utilizando técnicas de inteligencia artificial como las *multilayer perceptron neural networks (MLP-ANN)* y un *adaptative neuro-fuzzy interference system (ANFIS)*. Jamali et al. [66] utilizó cuatro modelos de regresión: *Decision Tree Regression (DTR)*, *Decision Forest Regression (DFR)*, *Linear Regression (LR)*, y *Bayesian Linear Regression (BLR)*, para hacer predicciones mensuales sobre las oscilaciones del nivel del mar, para desarrollar

un sistema de alerta temprana. Balogun et al. [67] realizaron un estudio que integra variables oceánico-atmosféricas para predecir la variación del nivel del mar en la costa occidental de Malasia peninsular usando técnicas de *machine learning* y *deep learning*. Los resultados mostraron que la combinación de estas variables mejora la precisión de los modelos *AutoRegressive Integrated Moving Average (ARIMA)*, *Support Vector Regression (SVR)* y *Long Short-Term Memory (LSTM) Neural Network*, siendo el modelo *LSTM* el más preciso en la mayoría de las estaciones, excepto en Pulau Langkawi, donde *ARIMA* fue mejor debido a la influencia de las mareas.

Lai et al. [68] evaluaron la capacidad del algoritmo de *machine learning Reduced Support Vector Machine (RSVM)*, para predecir los cambios en el nivel del mar en la costa este de Malasia peninsular, utilizando datos mensuales de nivel del mar, temperatura superficial del mar, lluvia y cobertura de nubes. Los resultados mostraron que el modelo predijo con precisión los aumentos del nivel del mar en Kerteh, Tanjung Sedili y la isla de Tioman, con coeficientes de correlación de 0.86, 0.83 y 0.86, respectivamente, y errores muy bajos en comparación con los datos históricos, sugiriendo que el modelo RSVM es una herramienta prometedora y confiable para los tomadores de decisiones.

Bagheri et al. [69] utilizaron la *Nonlinear Autoregressive Exogenous-Neural Network (NARX-NN)* y métodos de análisis en Sistemas de Información Geográfica GIS para estimar la tasa de aumento del nivel del mar y desarrollar mapas de peligros costeros, simulando un aumento máximo de 79.26 mm/año y una tasa promedio de erosión de 1.48 m/año entre 2013 y 2020 en la costa de *Terengganu*. Las áreas más vulnerables incluyeron Kuala Nerus, Pendagan Buluh y Kuala Ibai, con Batu Rakit mostrando la mayor tasa de erosión. Kumar et al. [70] utilizaron la implementación de técnicas de inteligencia artificial para modelar y predecir el realineamiento de la costa oriental de Odisha, India. Se analizaron imágenes satelitales (Landsat) utilizando filtros de detección de bordes, principalmente Sobel y Canny, para extraer información esencial y determinar el desplazamiento de la costa a partir de modelos espaciales en un Sistema de Información Geográfica SIG. Los datos de desplazamiento de la costa, junto con datos meteorológicos y parámetros de marea y olas, se utilizaron como entradas en algoritmos de *machine learning (Artificial Neural Network (ANN), K-Nearest Neighbors (KNN) y Support Vector Machine (SVM))* para predecir el desplazamiento de la costa. El análisis mostró que la velocidad del viento y la altura de las olas eran los factores más importantes, y el modelo ANN superó a KNN y SVM con una precisión del 86 %.

Alshouny et al. [71] presentaron un marco innovador de pasos interdependientes basado en la Transformación Wavelet (WT) y algoritmos de Inteligencia Artificial (AI) para la predicción de la variación del nivel del mar. Primero, se realizó un enfoque de tiempo-frecuencia *wavelet* junto con un análisis armónico para comprender mejor el comportamiento de variación del nivel del mar. Luego, aplicaron el *Neighborhood Component Analysis (NCA)* para la selección de características. Finalmente, utilizaron un algoritmo de *Deep Learning Neural Network (DLNN)* para predecir con precisión la variación del nivel del mar basado en un nuevo conjunto de datos compactado.

Todas estas aplicaciones de técnicas de Ciencia de Datos en el análisis del comportamiento del nivel del mar en escenarios del Fenómeno de El Niño demuestran la versatilidad, aplicabilidad y desempeño de estos algoritmos. Estos avances permiten hacer predicciones precisas que no solo facilitan el desarrollo tecnológico, sino que también mejoran la administración pública y la gobernanza del territorio. Al proporcionar herramientas más fiables para la predicción de fenómenos climáticos

y sus impactos, se fortalecen las estrategias de mitigación y adaptación frente a desastres naturales. Además, estas predicciones informadas permiten tomar decisiones más acertadas en la planificación urbana, la gestión de recursos y la protección de las comunidades costeras vulnerables.

Exploración y preparación de datos

3.1. Contexto

La exploración y preparación de datos constituyeron una de las etapas más relevantes en el desarrollo de esta investigación, dado que la calidad, consistencia y pertinencia de las variables seleccionadas resultaron fundamentales para la validez y fiabilidad de los análisis posteriores y, en última instancia, para alcanzar los objetivos planteados. Este capítulo describe el proceso seguido para la recolección, integración, limpieza y transformación de los datos necesarios para modelar la respuesta del nivel del mar ante la variabilidad oceánica y atmosférica en el Pacífico colombiano.

Para lograr una mayor comprensión de las variables, se emplearon diversas fuentes, incluyendo series de tiempo provenientes de estaciones mareográficas, productos de reanálisis atmosférico y oceánico, y datos obtenidos mediante sensores remotos. Estos conjuntos de datos fueron sometidos a técnicas estadísticas y visuales, con el fin de identificar patrones inherentes, posibles valores atípicos y datos faltantes. Asimismo, se aplicaron las transformaciones necesarias para garantizar la homogeneización temporal y espacial de las series, lo que incluyó la generación de anomalías estandarizadas, con el objetivo de resaltar las fluctuaciones interanuales asociadas a la variabilidad climática y facilitar su posterior análisis comparativo entre variables.

Adicionalmente, se integraron datos *LiDAR* (*Light Detection and Ranging*, por sus siglas en inglés) de alta resolución para la generación de Modelos Digitales de Elevación (MDE), utilizados posteriormente para el análisis espacial de la exposición costera. En las secciones de este capítulo, se detallan las fuentes de datos utilizadas, los procedimientos de exploración inicial implementados y las técnicas de transformación aplicadas como paso metodológico previo a las etapas de modelado y evaluación de resultados.

3.2. Fuentes de Datos

La presente investigación integró datos de múltiples fuentes, tanto *in situ* como remotas, con el propósito de capturar de forma representativa la variabilidad espacio-temporal de las condiciones océano-atmosféricas en el Pacífico colombiano, en el contexto de los eventos asociados al fenómeno El Niño-Oscilación del Sur (ENOS). A continuación, se describen las principales fuentes de información empleadas.

3.2.1. Series de tiempo oceanográficas y atmosféricas

En cuanto a las variables oceanográficas y meteorológicas registradas en las estaciones costeras, se utilizaron series de tiempo mensuales de nivel del mar, presión media al nivel del mar, campo de viento superficial y precipitación acumulada. Estos datos fueron proporcionados por la Red de Medición de Parámetros Oceanográficos y de Meteorología Marina (Red MPOMM) de la Dirección General Marítima (DIMAR), a través del Centro de Investigaciones Oceanográficas e Hidrográficas del Pacífico (CCCP). Las mediciones son obtenidas mediante sensores automáticos instalados en estaciones meteo-mareográficas, los cuales operan bajo estándares internacionales, generando registros minutales y horarios que fueron agregados a escala mensual para esta investigación.

Adicionalmente, se integraron los índices climáticos ENOS, incluyendo **Niño 1+2**, **Niño 3**, **Niño 3.4** y **Niño 4**, obtenidos del *Climate Prediction Center (CPC)* de la *NOAA*. Estos índices representan anomalías de temperatura superficial del mar en regiones específicas del Pacífico tropical (regiones El Niño) y constituyen indicadores ampliamente utilizados para caracterizar la fase, intensidad y tipo de evento ENOS [4], [5]. Fueron empleados como series de tiempo mensuales para explorar la relación entre la variabilidad climática y el comportamiento del nivel del mar en la región.

En cuanto a las fuentes de datos espaciales, se incorporaron variables derivadas de productos globales de reanálisis y sensores remotos, seleccionados por su validez científica, disponibilidad histórica y resolución adecuada para estudios regionales. Estas incluyen:

- ***Copernicus Marine Service (CMEMS)***: Esta fuente proporciona datos de reanálisis mensuales del nivel medio del mar, temperatura superficial del mar (TSM) y oleaje (altura significativa y período medio), derivados de altimetría satelital, boyas y modelos numéricos globales, con una resolución espacial de $0.083^\circ \times \circ$. El producto de oleaje utilizado corresponde a una frecuencia temporal de 3 horas, posteriormente agregado a mensual para el análisis de escenarios extremos costeros [72].
- ***ERA5 (ECMWF)***: Esta base de datos de reanálisis atmosférico, integra múltiples fuentes de observación (satelitales, aeronáuticas y de superficie) mediante asimilación de datos y modelos dinámicos, con cobertura global y resolución de $0.25^\circ \times 0.25^\circ$ con frecuencia mensual. De este producto se extrajeron presión media al nivel del mar (MSLP) y viento superficiales (10 m) (componentes zonal - U y meridional - V) [73].
- ***MSWEP***: Este conjunto de datos híbrido, combina observaciones terrestres, estimaciones satelitales y modelos reanalizados y ponderados mediante técnicas estadísticas para generar un producto robusto de precipitación global acumulada mensual con resolución de $0.1^\circ \times 0.1^\circ$ [74].

3.2.2. Obtención de los datos *LiDAR*

Los datos *LiDAR* utilizados para los análisis de impactos costeros fueron adquiridos por la DIMAR en las zonas urbanas de Ciudad Mutis, Buenaventura y Tumaco, a través de la Subdirección

Tabla 3.1: Resumen de fuentes de datos utilizadas en la investigación

Tipo de dato	Fuente / Producto	VARIABLES	Período	Resolución temporal	Resolución espacial
Datos <i>in situ</i>	DIMAR – CCCP	Nivel del mar, Presión, Viento, Precipitación	2009–2021	Horaria	Estación Tumaco, Buenaventura, Ciudad Mutis
Índices climáticos	NOAA – CPC	Niño 1+2, Niño 3, Niño 3.4, Niño 4	1950–2024	Mensual	Por región
Reanálisis atmosférico	ERA5 (ECMWF)	Presión, Viento U/V	1980–2024	Mensual	$0.25^\circ \times 0.25^\circ$
Reanálisis oceánico	CMEMS – Copernicus	Nivel del mar, TSM	1993–2021	Mensual	$0.083^\circ \times 0.083^\circ$
Oleaje	CMEMS – Copernicus	Altura significativa, Período medio	1993–2021	3 horas	$0.083^\circ \times 0.083^\circ$
Precipitación acumulada	MSWEP	Precipitación	1979–2023	Acum. mensual	$0.10^\circ \times 0.10^\circ$

de Operaciones de Inteligencia de la Fuerza Aeroespacial de Colombia.

La adquisición de los datos brutos mediante tecnología *LiDAR* se fundamentó en sistemas de referencia geodésicos precisos y adaptados a la ubicación geográfica de cada área de estudio. Para Ciudad Mutis y Buenaventura, se empleó el *DATUM* MAGNA-SIRGAS Bogotá-Oeste, mientras que para la ciudad de Tumaco se utilizó el *DATUM* MAGNA-SIRGAS Bogotá-Oeste-Oeste. Ambos *datums*, definidos dentro del marco de MAGNA-SIRGAS para Colombia, garantizan la consistencia y la correcta referenciación horizontal de las mediciones en sus respectivas zonas de influencia. Todos estos datos tienen entre 3 y 8 puntos láser por metro cuadrado como resolución espacial.

En cuanto a la proyección cartográfica, se aplicaron las siguientes: para Ciudad Mutis y Buenaventura se usó la EPSG:3115, correspondiente a la *Transverse Mercator (TM)* origen Bogotá. Para la ciudad de Tumaco, la proyección utilizada fue la EPSG:3114, también una *Transverse Mercator (TM)* pero con un origen y parámetros específicos optimizados para esta región del occidente colombiano. Estas proyecciones cilíndricas conformes son estándares para la cartografía oficial en sus respectivas áreas, minimizando la distorsión en las cercanías de sus meridianos centrales y facilitando la realización de cálculos de distancias y áreas con alta precisión en cada zona de estudio.

Posteriormente, los datos *LiDAR* fueron sometidos a un proceso de postprocesamiento. Esta etapa se integró información proveniente de diversas fuentes de posicionamiento y navegación, incluyendo el Sistema de Posicionamiento Global (GPS), la Unidad de Medición Inercial (IMU) y el propio sensor *LiDAR*. La combinación de estas fuentes permitió una determinación de la trayectoria del sensor y la orientación de cada pulso láser, asegurando la correcta georreferenciación de los datos en los *datums* y proyecciones definidos. Adicionalmente, se implementaron protocolos de control de calidad y validación para asegurar la exactitud posicional y temática de los datos, verificando la coherencia con puntos de control terrestre y otras fuentes de información geoespacial en cada una de las áreas levantadas. La corrección geométrica del modelo de elevación se realizó conforme a la metodología de Posicionamiento Geodésico Preciso (PPP) o Posicionamiento Absoluto Preciso, eliminando errores sistemáticos y garantizando la integridad espacial de la información altimétrica con respecto al *datum* vertical implícito en MAGNA-SIRGAS, todo ello considerando las particularidades geodésicas de cada zona.

Finalmente, se procedió a la clasificación de cada uno de los retornos láser, diferenciando entre elementos del terreno, vegetación y construcciones, constituyendo un insumo fundamental para diversas aplicaciones temáticas como la generación de modelos digitales de elevación (MDE), en las zonas urbanas de Ciudad Mutis, Buenaventura y Tumaco. La totalidad de estos procedimientos se ejecutó en estricto cumplimiento de los lineamientos técnicos y requerimientos especificados por la DIMAR, y bajo los estándares establecidos en la Norma Técnica Colombiana - NTC 4611 – Metadatos Geográficos - Primera Actualización, asegurando la correcta documentación de todos los parámetros de referencia espacial y la calidad de los datos generados para cada una de las áreas de estudio.

3.2.2.1. Ciudad Mutis, Chocó

Ciudad Mutis, también conocida como Bahía Solano, está ubicada en la costa norte del departamento del Chocó, en el noroccidente de Colombia, sobre la costa del Océano Pacífico. Específi-

camente, se encuentra en la Serranía del Baudó. La cabecera municipal se llama Ciudad Mutis, en honor al sabio José Celestino Mutis.

Sus coordenadas aproximadas son 6.22, -77.24 (Tabla 3.2). La ciudad se encuentra a una altura de 152 metros sobre el nivel del mar y tiene una temperatura promedio de 25 °C.

Para el área urbana de Ciudad Mutis, se llevó a cabo un levantamiento aerotransportado con tecnología *LiDAR* entre el 11 de febrero de 2013 y el 19 de julio del mismo año. Esta adquisición de datos cubrió la extensión del casco urbano (Figura 3.1), delimitada geográficamente por las siguientes coordenadas extremas:

Tabla 3.2: Límites de las coordenadas geográficas de Ciudad Mutis

Coordenadas geográficas límite	Valor
Longitud máxima	-77.3866527
Longitud mínima	-77.4204133
Latitud máxima	6.237563
Latitud mínima	6.1894243

Bahía Solano tiene una extensión de 1 667 km². Limita al norte con el municipio de Juradó, al sur con el Parque Nacional Natural Ensenada de Utría, al oriente con la Serranía del Baudó y al occidente con el Océano Pacífico.

3.2.2.2. Buenaventura, Valle del Cauca

Buenaventura se encuentra ubicada en la costa pacífica de Colombia, dentro del departamento del Valle del Cauca (Figura 3.2). Su casco urbano se extiende principalmente sobre la Isla de Cascajal, una isla continental rodeada por la Bahía de Buenaventura. Esta bahía, de gran importancia para el puerto, está conectada al Océano Pacífico a través de una serie de canales y esteros que forman un intrincado sistema de manglares.

La ciudad se sitúa aproximadamente a 115 kilómetros al oeste de Cali, la capital del departamento, y su acceso terrestre se realiza principalmente a través de la carretera que conecta ambas ciudades. Geográficamente, Buenaventura se encuentra en una zona de transición entre la llanura costera del Pacífico y las estribaciones de la Cordillera Occidental de los Andes, lo que le confiere una topografía variada y una rica biodiversidad.

Sus coordenadas geográficas aproximadas son 3.88, -77.08 (Tabla 3.3). Esta ubicación la convierte en el puerto colombiano más cercano a países de la cuenca del Pacífico, como Panamá, Ecuador y Perú, lo que facilita el comercio marítimo con estas naciones. Además, su proximidad al Canal de Panamá le otorga una ventaja estratégica en el comercio internacional.

El levantamiento aerotransportado *LiDAR* se llevó a cabo entre el 14 de junio y el 28 de julio del año 2006. El levantamiento de datos cubrió la extensión del casco urbano de Buenaventura y el área se delimitó geográficamente por las siguientes coordenadas:



Figura 3.1: Casco Urbano de Ciudad Mutis, Chocó

Tabla 3.3: Límites de las coordenadas geográficas de Buenaventura

Coordenadas geográficas límite	Valor
Longitud máxima	-77.714184
Longitud mínima	-77.818690
Latitud máxima	3.8429602
Latitud mínima	3.6748669



Figura 3.2: Casco Urbano de Buenaventura, Valle del Cauca

3.2.2.3. Tumaco, Nariño

Tumaco es un municipio colombiano ubicado en el extremo suroccidental del departamento de Nariño, directamente sobre la costa del Océano Pacífico y muy cerca de la frontera con Ecuador. Su cabecera municipal se asienta principalmente sobre la Isla de Tumaco (Figura 3.3) y otras islas menores adyacentes, conectadas entre sí y al continente por puentes y vías terrestres elevadas sobre los esteros y manglares.

Su ubicación estratégica en la costa pacífica sur le confiere una gran importancia como puerto marítimo, aunque de menor escala en comparación con Buenaventura. Tumaco sirve como punto de conexión para el comercio y el transporte en la región fronteriza con Ecuador.

Geográficamente, Tumaco se caracteriza por ser una zona de transición entre la llanura costera del Pacífico y las estribaciones de la Cordillera Occidental. Su territorio está marcado por una extensa red de ríos, esteros, manglares y una rica biodiversidad propia del Chocó biogeográfico. Las islas sobre las que se asienta la ciudad están rodeadas por estos ecosistemas intermareales, lo que influye significativamente en su paisaje y en las actividades económicas locales, como la pesca y la extracción de recursos del manglar.

Las coordenadas geográficas aproximadas del centro urbano de Tumaco son 1.80, -77.80. Su cercanía a la línea ecuatorial le otorga un clima cálido y húmedo durante todo el año, con altas precipitaciones que alimentan la exuberante vegetación circundante.

El levantamiento de datos *LiDAR* para este casco urbano se llevó a cabo entre los meses de marzo y junio de 2012 y se enmarcó entre las siguientes coordenadas (Tabla 3.4):

Tabla 3.4: Límites de las coordenadas geográficas de Tumaco

Coordenadas geográficas límite	Valor
Longitud máxima	-78.714184
Longitud mínima	-78.818690
Latitud máxima	1.8429602
Latitud mínima	1.7486669

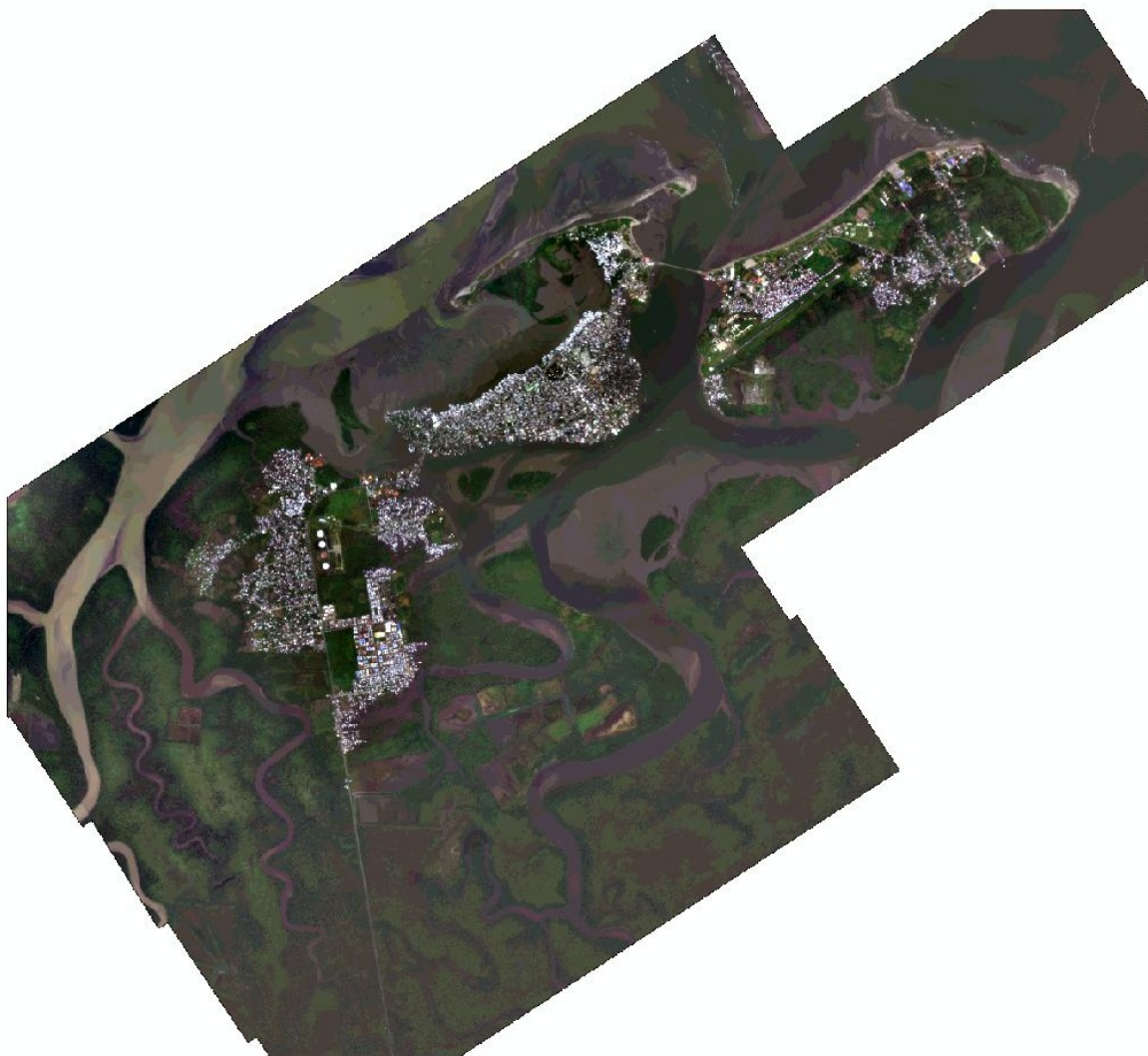


Figura 3.3: Casco Urbano de Tumaco, Nariño

3.3. Limpieza de Datos

3.3.1. Tratamiento de series de tiempo oceanográficas y atmosféricas

3.3.1.1. Series de tiempo en puntos fijos de observación

En la etapa inicial de procesamiento, se llevó a cabo una limpieza rigurosa para asegurar la calidad y confiabilidad de las series de tiempo de las variables atmosféricas: presión media al nivel del mar, velocidad y dirección del viento, y precipitación acumulada. Este proceso comprendió dos fases principales: un filtrado inicial mediante umbrales físicos, y la posterior detección de valores atípicos mediante el método del rango intercuartílico (*IQR*).

Durante la primera fase, se aplicaron umbrales físicos definidos específicamente para las condiciones típicas del Pacífico colombiano, con base en el *Atlas Climatológico de Colombia* [75] y estudios técnicos de la DIMAR [75], [76]. Los criterios aplicados por variable fueron los siguientes:

- **Presión (hPa):** Se retuvieron valores entre 980 y 1 020 hPa, consistentes con los registros promedio a nivel del mar en la región [75].
- **Velocidad del viento superficial (m/s):** Se aceptaron registros entre 0 y 12 m/s, coherentes con los rangos típicos observados en la costa pacífica bajo condiciones normales [75], [76].
- **Dirección del viento superficial (°):** Se conservaron valores entre 10° y 360°, rango suficiente para cubrir la variabilidad angular natural del viento en esta zona de compleja orografía costera [75].
- **Precipitación acumulada (mm):** Se filtraron valores en el rango de 0 a 1 000 mm. Aunque la región es una de las más lluviosas del planeta, este umbral superior permitió detectar posibles errores de medición en eventos extremos aislados [75], [76].

La Figura 3.4 muestra las series de tiempo originales, en las que se observan valores anómalos evidentes, mientras que la Figura 3.5 presenta las series resultantes tras la aplicación de los filtros por umbrales descritos anteriormente, evidenciando una mejora significativa en la coherencia de los datos.

Posteriormente, se aplicó la técnica del IQR como criterio robusto para la identificación de valores atípicos (*outliers*) en las variables continuas. Este método, ampliamente utilizado en análisis exploratorio de datos, se caracteriza por su resistencia frente a valores extremos y su independencia respecto a la forma de la distribución [77], [78]. El IQR se define como la diferencia entre el tercer y el primer cuartil ($Q_3 - Q_1$); se consideraron atípicos aquellos valores fuera del intervalo $[Q_1 - 1.5 \times \text{IQR}, Q_3 + 1.5 \times \text{IQR}]$.

Se debe precisar que esta técnica no fue aplicada a la precipitación, dado que en este caso los valores iguales a cero, frecuentes en condiciones sin lluvia, podrían clasificarse incorrectamente como *outliers*, a pesar de tener un significado físico válido.

Para analizar el efecto del filtrado mediante el IQR, se generaron diagramas de caja (*boxplots*) comparativos antes y después de la aplicación del filtro para las variables atmosféricas continuas.

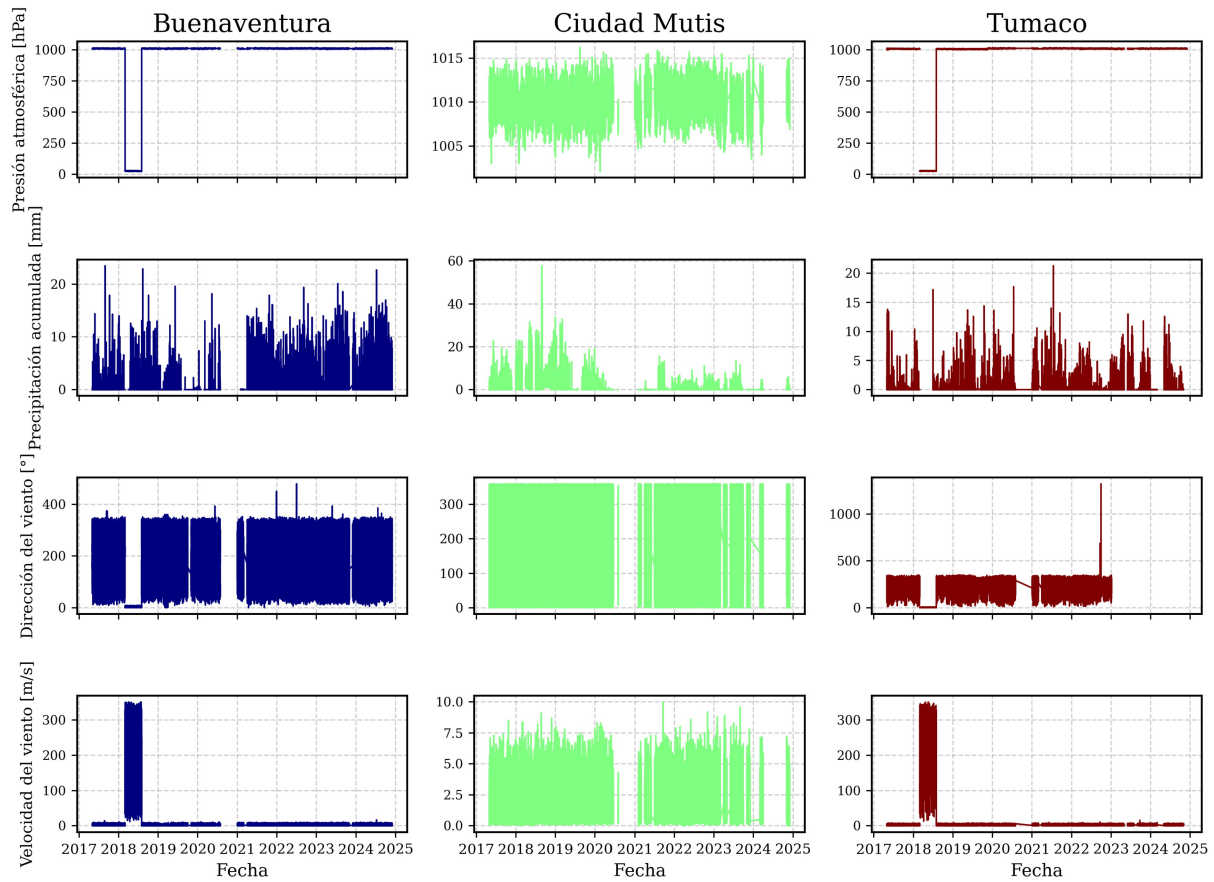


Figura 3.4: Series de tiempo originales de variables meteorológicas para Ciudad Mutis, Buenaventura y Tumaco. (2017–2024).

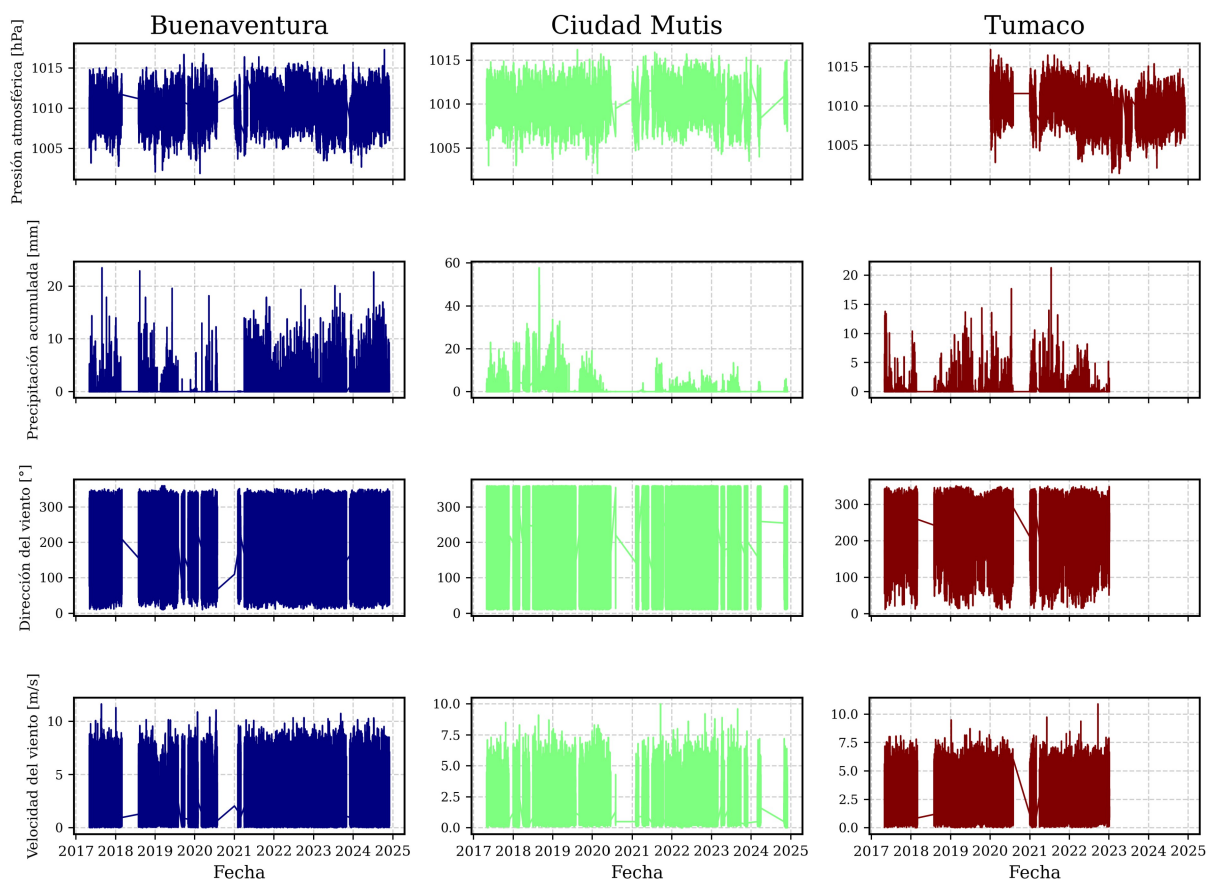


Figura 3.5: Series de tiempo de variables meteorológicas filtradas por umbrales físicos para Ciudad Mutis, Buenaventura y Tumaco. (2017–2024).

En la Figura 3.6, correspondiente a la presión media al nivel del mar, se observa una reducción significativa de los valores atípicos tras la limpieza, con una distribución más simétrica y acotada en cada una de las localidades.

De manera similar, las Figuras 3.7 y 3.8 muestran el impacto del filtrado sobre las componentes zonal (U) y meridional (V) del viento. En ambos casos, el proceso redujo considerablemente la dispersión extrema, concentrando los valores dentro de los límites intercuartílicos y facilitando una representación más realista de la variabilidad atmosférica local. Esta depuración resulta fundamental para asegurar la robustez de los análisis posteriores, especialmente en contextos predictivos donde la sensibilidad a valores extremos puede distorsionar el aprendizaje del modelo.

Es relevante señalar que, si bien el filtrado mejora la calidad estadística de los datos, se asumió un compromiso entre eliminar valores extremos plausibles y reducir la influencia de posibles errores o registros anómalos. Este procedimiento fue especialmente relevante para las variables de viento, cuyas distribuciones mostraban colas largas y asimetrías marcadas, en particular en Tumaco y Solano.

Siguiendo una metodología coherente con el tratamiento aplicado a las variables atmosféricas, se realizó también una revisión de las series de tiempo oceanográficas, específicamente del nivel del mar y la temperatura superficial del mar (TSM).

Las series de nivel del mar empleadas provienen de sensores de radar instalados en las estaciones meteo-mareográficas de Ciudad Mutis (Bahía Solano), Buenaventura y Tumaco, operadas por la DIMAR (Figura 3.9). Estas estaciones forman parte de la red de monitoreo costero del Pacífico colombiano. Inicialmente, se realizó un análisis exploratorio de las series de tiempo para identificar patrones cíclicos, semidiurnos y posibles distribuciones bimodales, características atribuibles a la dinámica mareal de la región.

Para garantizar la calidad y confiabilidad de los datos, se aplicaron pruebas de control de calidad de primer y segundo nivel, siguiendo los estándares internacionales establecidos por la Comisión Oceanográfica Intergubernamental (IOC) a través del Grupo Internacional de Intercambio de Datos Oceanográficos (IODE) [79]. Entre las pruebas aplicadas se incluyeron:

- **Prueba de rango físico:** Verificación de que los valores registrados se encuentren dentro de los límites físicos esperados para la región.
- **Prueba de variabilidad temporal:** Detección de cambios abruptos o inconsistencias en la serie de tiempo que puedan indicar errores instrumentales o de registro.
- **Prueba de valores repetidos:** Identificación de secuencias de valores idénticos que podrían sugerir fallos en la transmisión o almacenamiento de datos.

Estas pruebas permitieron la detección y eliminación de valores atípicos que no representaban el comportamiento real de la señal, asegurando así la integridad de las series de tiempo para análisis posteriores.

Después de aplicar las pruebas de control de calidad descritas anteriormente, se procedió a la imputación de las series de tiempo de nivel del mar mediante un análisis armónico. Esta técnica

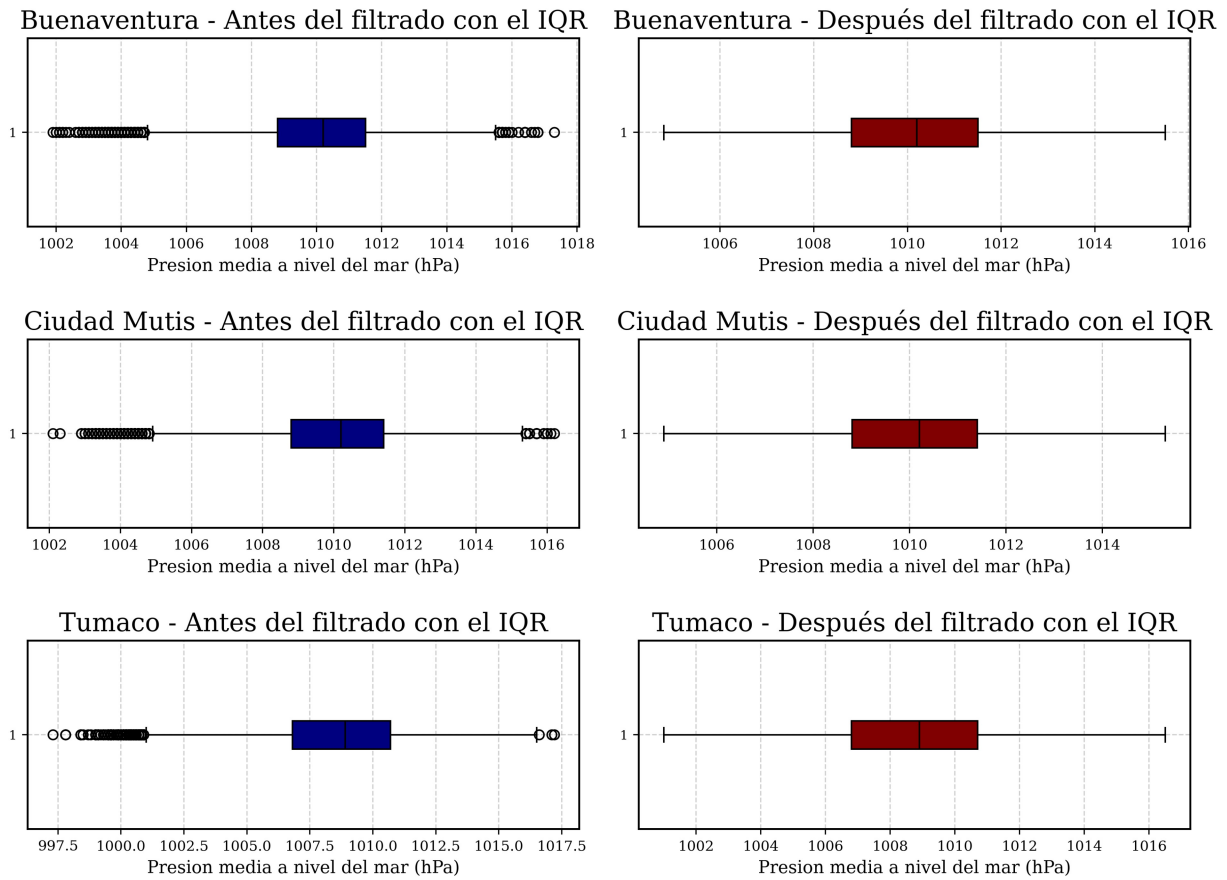


Figura 3.6: Distribución de la presión atmosférica antes (azul) y después (rojo) del filtrado con IQR para Ciudad Mutis, Buenaventura y Tumaco.

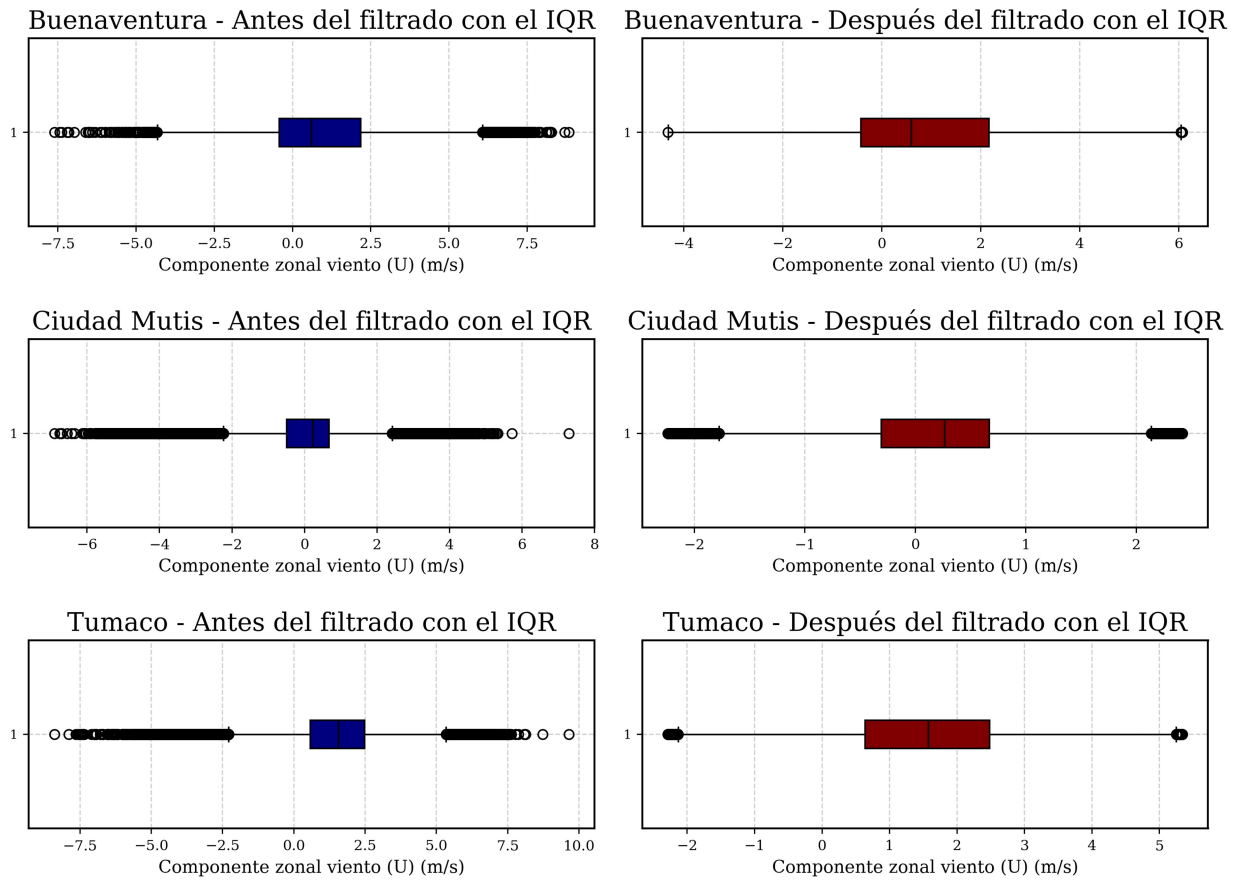


Figura 3.7: Distribución de la componente zonal del viento (U) antes (azul) y después (rojo) del filtrado con IQR para Ciudad Mutis, Buenaventura y Tumaco.

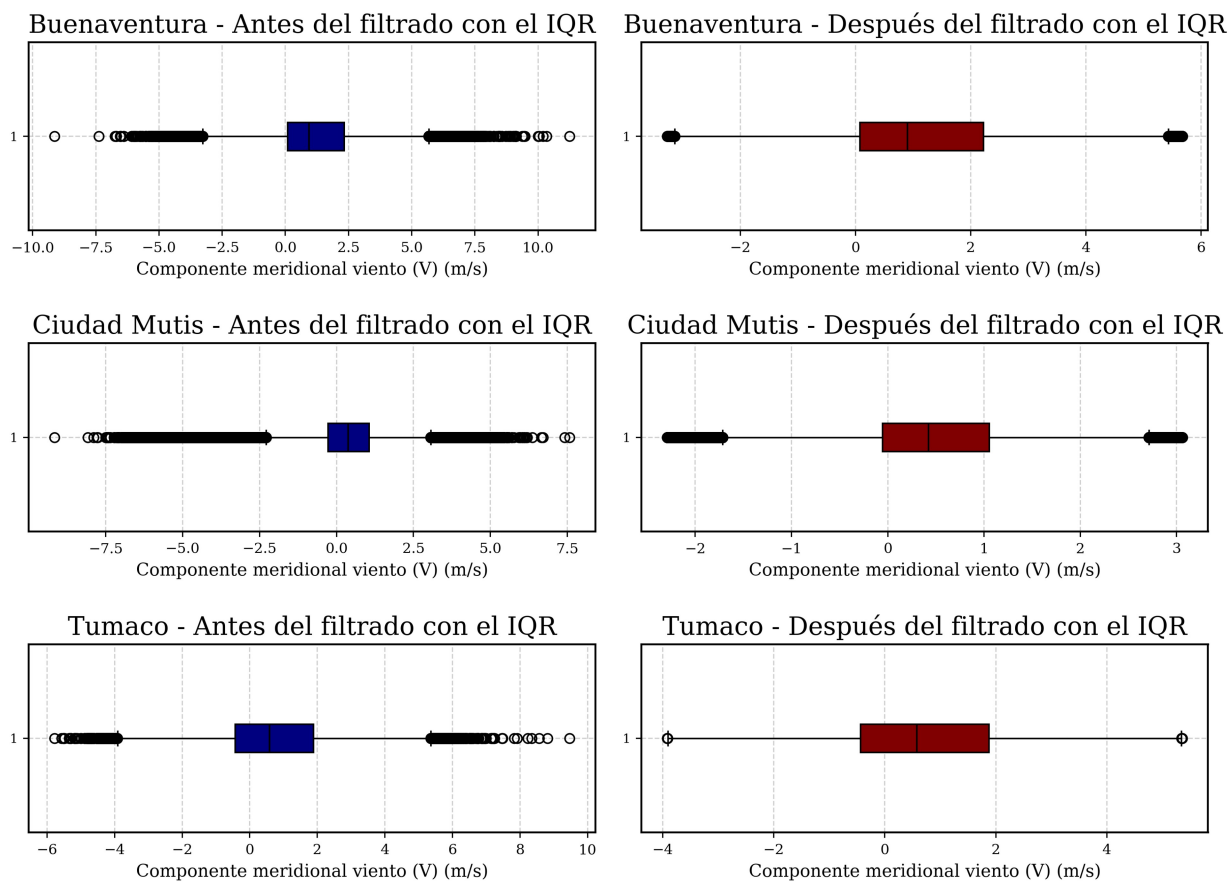


Figura 3.8: Distribución de la componente meridional del viento (V) antes (azul) y después (rojo) del filtrado con IQR para Ciudad Mutis, Buenaventura y Tumaco.

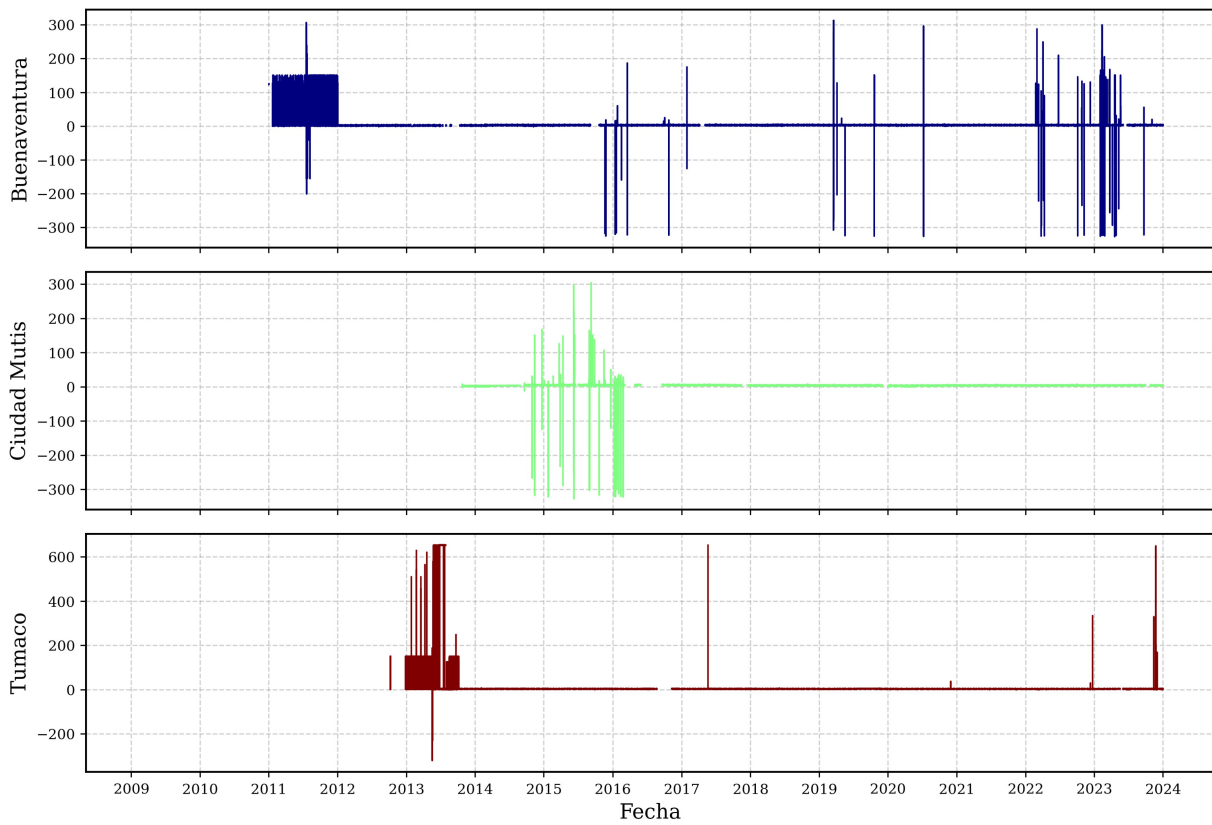


Figura 3.9: Series de tiempo originales de nivel del mar para Ciudad Mutis, Buenaventura y Tumaco. (2009-2024).

clásica que descompone la serie de nivel del mar en una suma de componentes sinusoidales de frecuencia conocida, principalmente asociadas al forzamiento astronómico de las mareas [80], permitió modelar el componente mareal de forma precisa a partir de los constituyentes armónicos principales, preservando las características cíclicas y semidiurnas observadas en la región.

Finalmente, para representar el componente meteorológico del nivel del mar, es decir, las desviaciones respecto al comportamiento mareal puramente astronómico, se utilizó la Topografía Dinámica Absoluta (*ADT*, por sus siglas en inglés) [81]. Este producto satelital integra el nivel medio del mar (*Mean Sea Surface, MSS*) y las anomalías dinámicas (*Sea Level Anomalies, SLA*), capturando así la influencia de fenómenos como ondas planetarias, circulación superficial y presión atmosférica sobre el nivel del mar [82], permitiendo su utilización como residuo meteorológico en la imputación de las series de nivel del mar en las tres estaciones costeras señaladas, dadas las condiciones específicas del Pacífico colombiano (Figura 3.10).

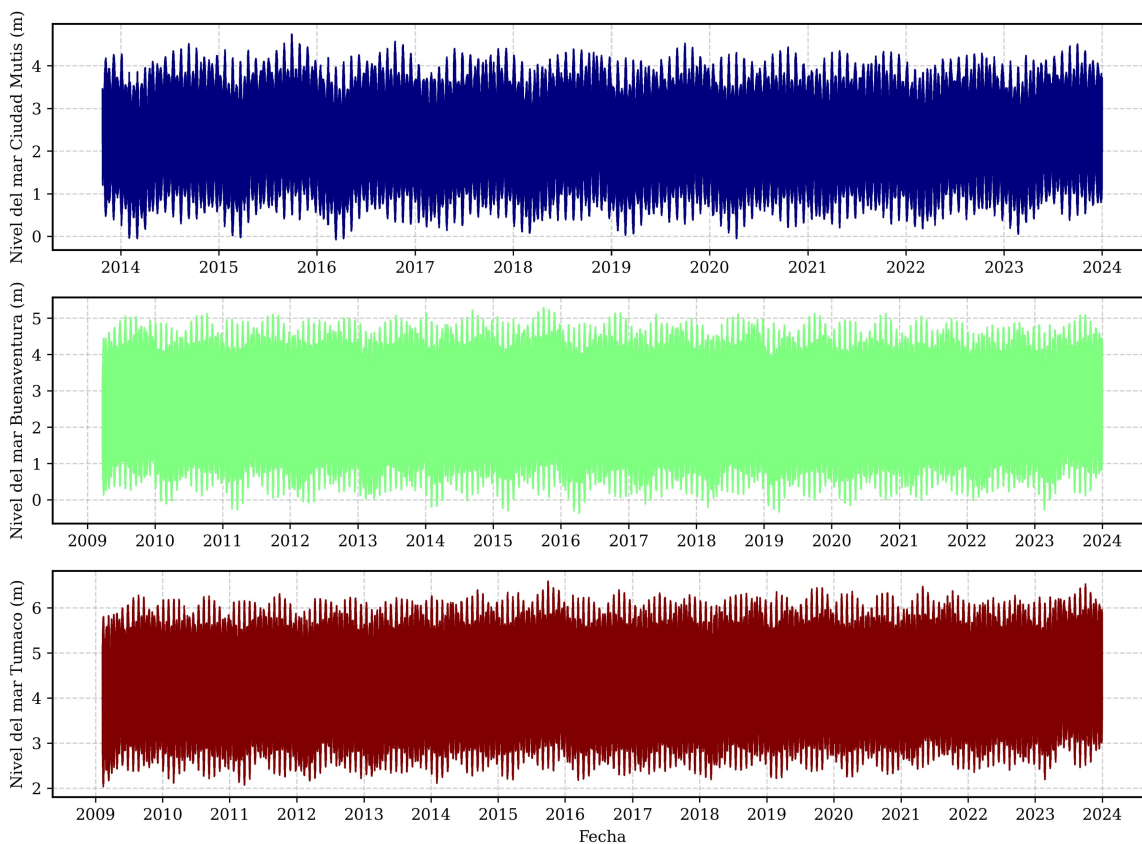


Figura 3.10: Series de tiempo filtradas de nivel del mar para Ciudad Mutis, Buenaventura y Tumaco (2009-2024).

En cuanto a la TSM, se utilizaron datos obtenidos de perfiles de Conductividad, Temperatura y Profundidad (CTD) recolectados en la estación fija costera de Tumaco (2.00, -77.78). Estos datos fueron procesados utilizando los filtros de calidad desarrollados por *Sea-Bird Scientific*, los cuales incluyen procedimientos automáticos y manuales para la detección de errores y la validación de datos [83]. Los filtros aplicados comprenden:

- **Conversión de datos:** Transformación de los datos brutos a unidades científicas mediante el módulo *Data Conversion*.
- **Edición de valores anómalos:** Identificación y marcado de datos aberrantes utilizando el módulo *Wild Edit*.
- **Filtrado de datos:** Aplicación de filtros de paso bajo para suavizar las señales y minimizar el ruido, mediante el módulo *Filter*.

Para representar adecuadamente la TSM, se extrajeron los datos correspondientes a las capas superficiales, seleccionando aquellos registros con profundidades menores a 5 metros. Esta selección se basa en la recomendación de utilizar datos cercanos a la superficie para estudios de TSM, asegurando así la representatividad de las condiciones superficiales del océano.

3.3.1.2. Series de tiempo con cobertura espacio-temporal (fuentes externas)

Las variables oceánicas y atmosféricas provenientes de productos satelitales y de reanálisis empleadas, incorporan procedimientos integrados de control de calidad que garantizan la confiabilidad y consistencia de los datos antes de su distribución (Figura 3.11).

En el caso del *Copernicus Marine Environment Monitoring Service (CMEMS)*, los productos satelitales y de reanálisis oceánico pasan por un riguroso sistema de control de calidad que incluye verificaciones automáticas y validaciones en modo diferido. Este sistema contempla la evaluación de las anomalías, la estabilidad temporal, la coherencia entre productos y la comparación con observaciones *in situ* cuando están disponibles. Además, cada producto es acompañado de informes de validación técnica con métricas estadísticas específicas y documentación sobre el rendimiento del sistema de modelado o altimetría utilizados [84].

En cuanto al producto ERA5, generado por el Centro Europeo de Previsiones Meteorológicas a Medio Plazo (ECMWF), se basa en un esquema de asimilación de datos *4D-Var* que integra millones de observaciones satelitales, aeronáuticas y de superficie. Su sistema de control de calidad automatizado incluye verificaciones físicas y espaciales de las observaciones, así como pruebas de consistencia interna durante la ingestión de datos. Los procedimientos se complementan con una validación cruzada frente a productos previos como ERA-*Interim*, garantizando así la trazabilidad y mejora continua de los reanálisis climáticos [73].

Para el caso de *MSWEP (Multi-Source Weighted-Ensemble Precipitation)*, la estimación de precipitación global combina observaciones satelitales, reanálisis y datos de más de 77 000 estaciones meteorológicas a nivel mundial. El control de calidad se basa en la corrección diaria de sesgos mediante estaciones confiables, ajuste temporal en función de la hora de observación y el uso de

ponderaciones regionales que optimizan la fusión de fuentes de acuerdo con su desempeño histórico. Esto permite mejorar sustancialmente la precisión en regiones con cobertura desigual de sensores [74].

Se debe precisar que, si bien los procedimientos descritos aplican de forma general a todos los productos distribuidos por cada sistema, existen umbrales, filtros y metodologías específicas según el tipo de variable, las cuales se encuentran detalladas en los respectivos manuales técnicos. Por esta razón, en el presente trabajo se consideró que las variables derivadas de estas fuentes ya contaban con una etapa previa de depuración y validación.

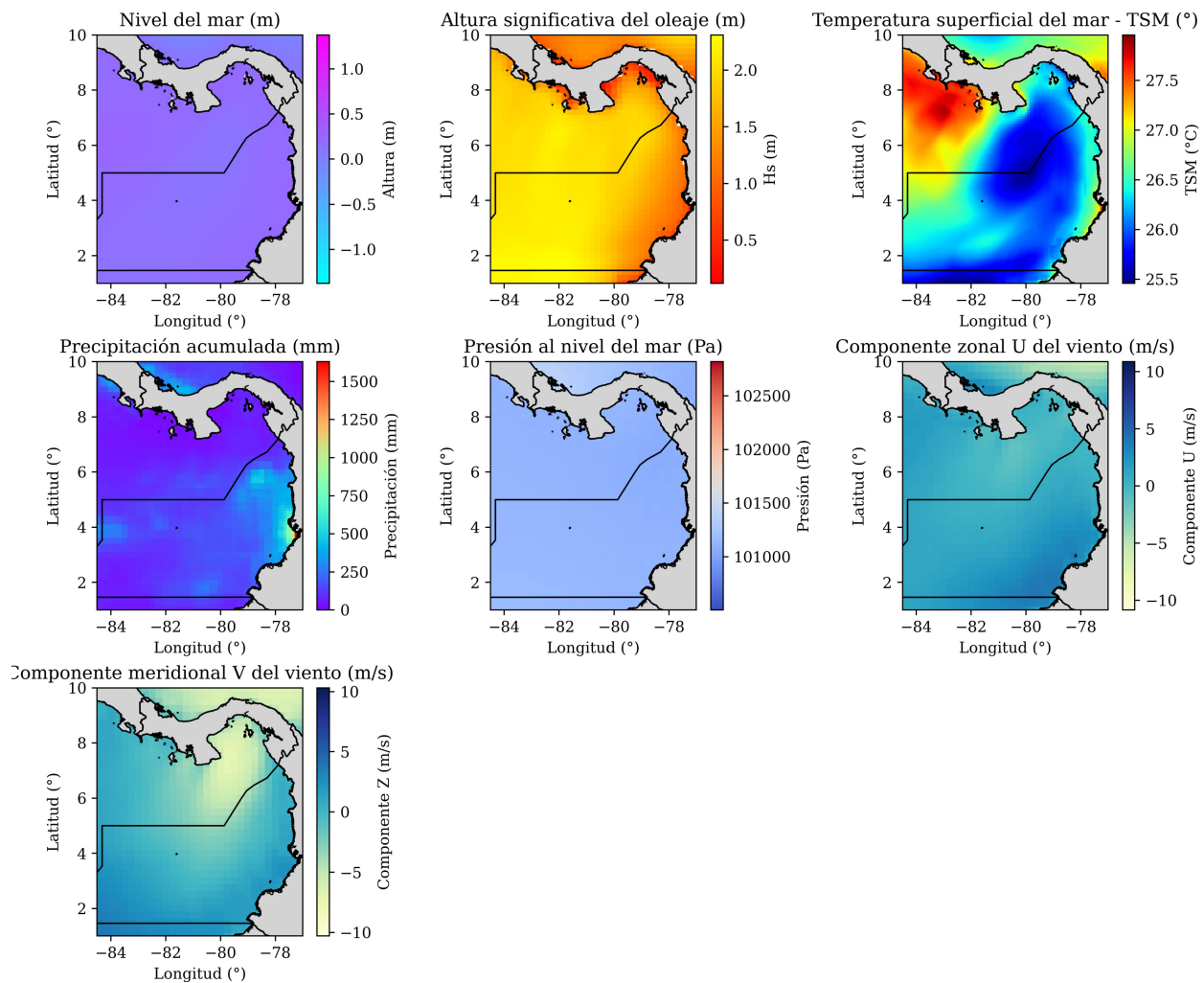


Figura 3.11: Distribución espacial de variables océano - atmosféricas en el Pacífico colombiano (Enero 2000).

3.3.2. Tratamiento de los datos *LiDAR*

Una etapa importante en el procesamiento de los datos *LiDAR* consistió en el filtrado de los retornos de suelo, un procedimiento esencial para la generación precisa de modelos digitales del terreno (MDT) o MDE que representen la superficie terrestre desnuda sin vegetación o estructuras (Figura 3.12). Este proceso se implementó utilizando las herramientas especializadas disponibles en el software ArcGIS[®] Pro 3.4.3 [85]. Dentro del software, se emplearon algoritmos de filtrado diseñados específicamente para discriminar entre los puntos que reflejan la superficie del suelo y aquellos que provienen de objetos elevados como la vegetación (árboles, arbustos), edificaciones y otras estructuras artificiales [85]. Estos algoritmos, como el *Progressive Morphological Filter (PMF)* o filtros basados en análisis de vecindad y pendiente, analizan las características tridimensionales de la nube de puntos *LiDAR* [85].

Después de esta filtración automática inicial se llevó a cabo una filtración manual con los datos que no concordaban con los patrones de suelo. A diferencia de los métodos automatizados, el filtrado manual implicó una inspección visual detallada de la nube de puntos *LiDAR* en un entorno tridimensional dentro de ArcGIS[®] Pro. Este proceso se basó en la interpretación visual de patrones. Los puntos del suelo generalmente forman una superficie continua y relativamente suave, con cambios de pendiente graduales. En contraste, los retornos de la vegetación tienden a presentarse en grupos con variaciones de altura significativas y formas irregulares. Las estructuras artificiales suelen mostrar geometrías angulares y elevaciones abruptas. Tras la finalización del filtrado manual, los puntos clasificados como suelo se utilizaron para la generación del MDE. Los puntos clasificados diferentes al retorno de suelo quedaron disponibles para otros análisis temáticos.

3.4. Exploración de Datos

3.4.1. Exploración estadística de series de tiempo oceanográficas y atmosféricas

3.4.1.1. Distribución de frecuencias de series de tiempo en puntos fijos de observación

Para caracterizar el comportamiento estadístico de las variables utilizadas en este estudio, se generaron histogramas de frecuencia y diagramas de caja (*boxplots*) para cada una de las estaciones: Ciudad Mutis, Buenaventura y Tumaco (ver Figura 3.13). Estas visualizaciones permitieron identificar patrones de distribución, presencia de valores atípicos y dispersión general.

La presión media al nivel del mar presentó una distribución aproximadamente simétrica en Buenaventura y Ciudad Mutis, y un leve sesgo hacia la derecha en Tumaco, con valores centrados entre 1 008 y 1 012 hPa, consistentes con condiciones atmosféricas normales al nivel del mar. No se evidencian valores extremos significativos, lo que indica estabilidad en la medición y calidad de los datos.

En cuanto a la precipitación acumulada, se observó una fuerte asimetría positiva con la mayoría de los valores concentrados por debajo de 10 mm/día, y algunos eventos extremos superiores a 40 mm/día, especialmente en Ciudad Mutis. Los *boxplots* reflejan esta variabilidad con múltiples valores atípicos, lo cual es esperable en zonas tropicales húmedas como el litoral Pacífico colombiano,

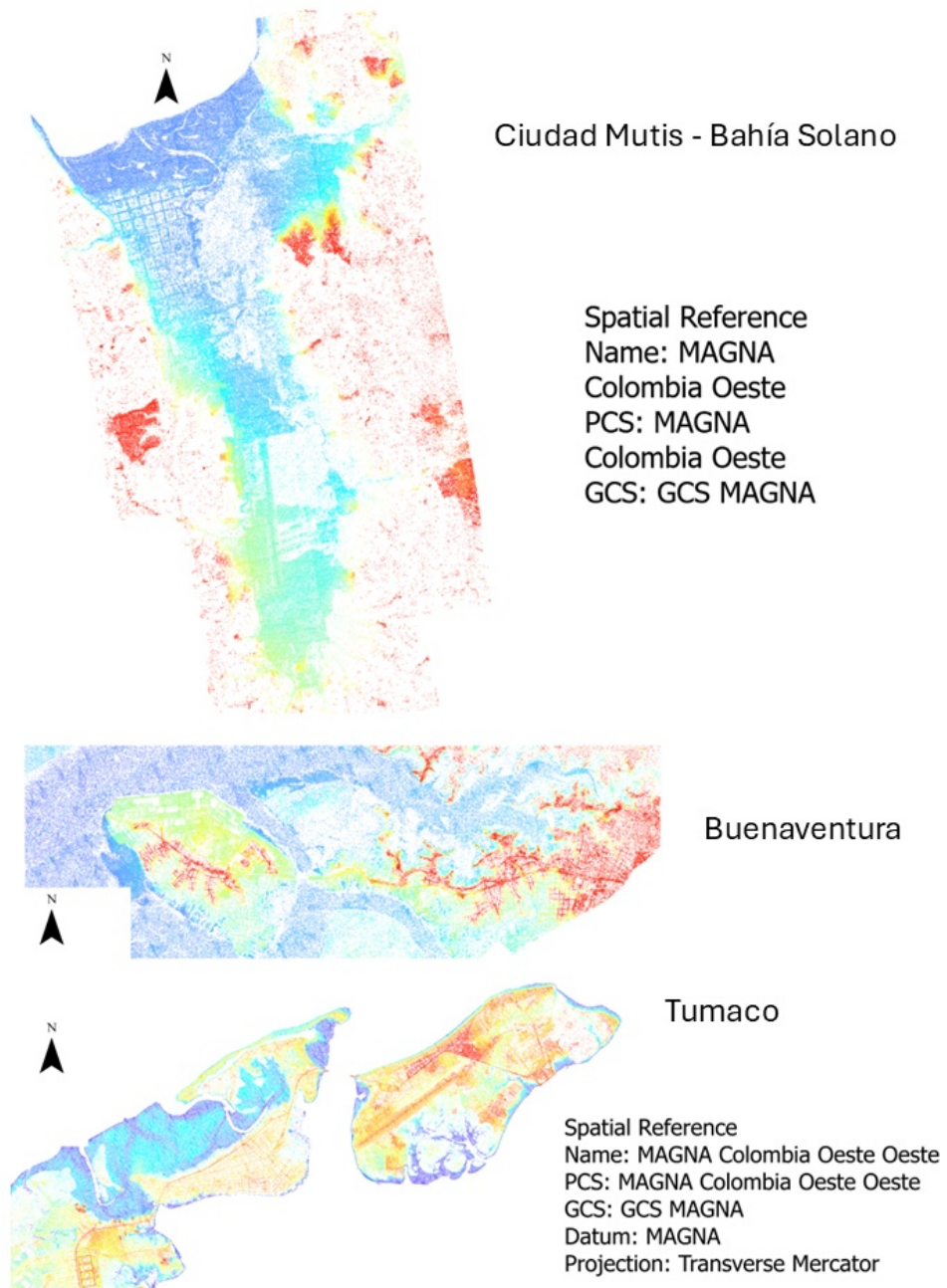


Figura 3.12: Resultado de los datos del retorno de suelo filtrado de los datos *LiDAR* en cada uno de los cascos urbanos.

donde predominan eventos de lluvia intensos pero esporádicos.

Las distribuciones de las componentes del viento (U y V) en las tres localidades muestran patrones direccionales diferenciados. En Buenaventura (primera columna), la componente U presenta una ligera asimetría hacia valores positivos, mientras que la componente V se concentra en valores negativos, indicando un predominio de vientos del suroeste. En Ciudad Mutis (segunda columna), ambas componentes muestran valores centrales claramente negativos, lo cual indica vientos con dirección predominante desde el noreste hacia el suroeste. En Tumaco (tercera columna), aunque las distribuciones aparentan mayor simetría en forma, sus medianas están también desplazadas hacia valores negativos, indicando un patrón de viento similar al de Ciudad Mutis, aunque con menor intensidad, posiblemente influenciados por procesos locales o la orografía.

En cuanto al nivel del mar, las distribuciones presentaron un patrón claramente bimodal en las tres localidades, reflejando la influencia del régimen de marea semidiurno característico del litoral Pacífico colombiano. Las frecuencias se concentraron en torno a dos rangos principales, generalmente situados entre 1.8 y 3.8 m para el caso de Ciudad Mutis y Buenaventura, y 3.5 y 6.5 m para Tumaco; correspondientes a los niveles de marea alta y baja predominantes. Es preciso aclarar que estos valores están referidos al cero del sensor.

Tabla 3.5: Resultados del *test* de normalidad D’Agostino-Pearson para cada variable y localidad. Se reporta el estadístico K^2 , el valor p , y su comparación frente a los umbrales del 1% y 0.1%.

Localidad	Estadístico	Presión	Precipitación acum.	Comp. U viento	Comp. V viento	Nivel del mar
Ciudad Mutis	D’Agostino K^2	467.743	235 860.847	5 577.390	853.078	33 174.254
	p -valor	2.70e-102	0.00e+00	0.00e+00	5.71e-186	0.00e+00
	$p > 0,01$	2.70e-102	0.00e+00	0.00e+00	5.71e-186	0.00e+00
	$p > 0,001$	2.70e-102	0.00e+00	0.00e+00	5.71e-186	0.00e+00
Buenaventura	D’Agostino K^2	676.678	556 024.250	15 865.992	14 950.433	58 812.295
	p -valor	1.15e-147	0.00e+00	0.00e+00	0.00e+00	0.00e+00
	$p > 0,01$	1.15e-147	0.00e+00	0.00e+00	0.00e+00	0.00e+00
	$p > 0,001$	1.15e-147	0.00e+00	0.00e+00	0.00e+00	0.00e+00
Tumaco	D’Agostino K^2	1 236.739	520 088.356	900.495	3 210.358	59 184.113
	p -valor	2.79e-269	0.00e+00	2.88e-196	0.00e+00	0.00e+00
	$p > 0,01$	2.79e-269	0.00e+00	2.88e-196	0.00e+00	0.00e+00
	$p > 0,001$	2.79e-269	0.00e+00	2.88e-196	0.00e+00	0.00e+00

Para complementar el análisis gráfico de las distribuciones, se aplicó el *test* de normalidad de D’Agostino–Pearson a cada una de las series de tiempo por variable y por localidad. Esta prueba fue seleccionada debido a su idoneidad para tamaños de muestra grandes, ya que combina medidas de asimetría y curtosis para evaluar desviaciones de la normalidad. A diferencia de otros *tests* como el de *Shapiro–Wilk*, que es más sensible en muestras pequeñas, el *test* de D’Agostino ofrece mayor robustez estadística cuando se dispone de un número elevado de observaciones [86],[87].

Los resultados obtenidos, presentados en la Tabla 3.5, muestran que todas las variables analizadas (presión, precipitación acumulada, componentes zonal (U) y meridional del viento (V), y nivel del mar) presentan desviaciones significativas respecto a una distribución normal en las tres localidades evaluadas (Ciudad Mutis, Buenaventura y Tumaco). En todos los casos, los valores del estadístico

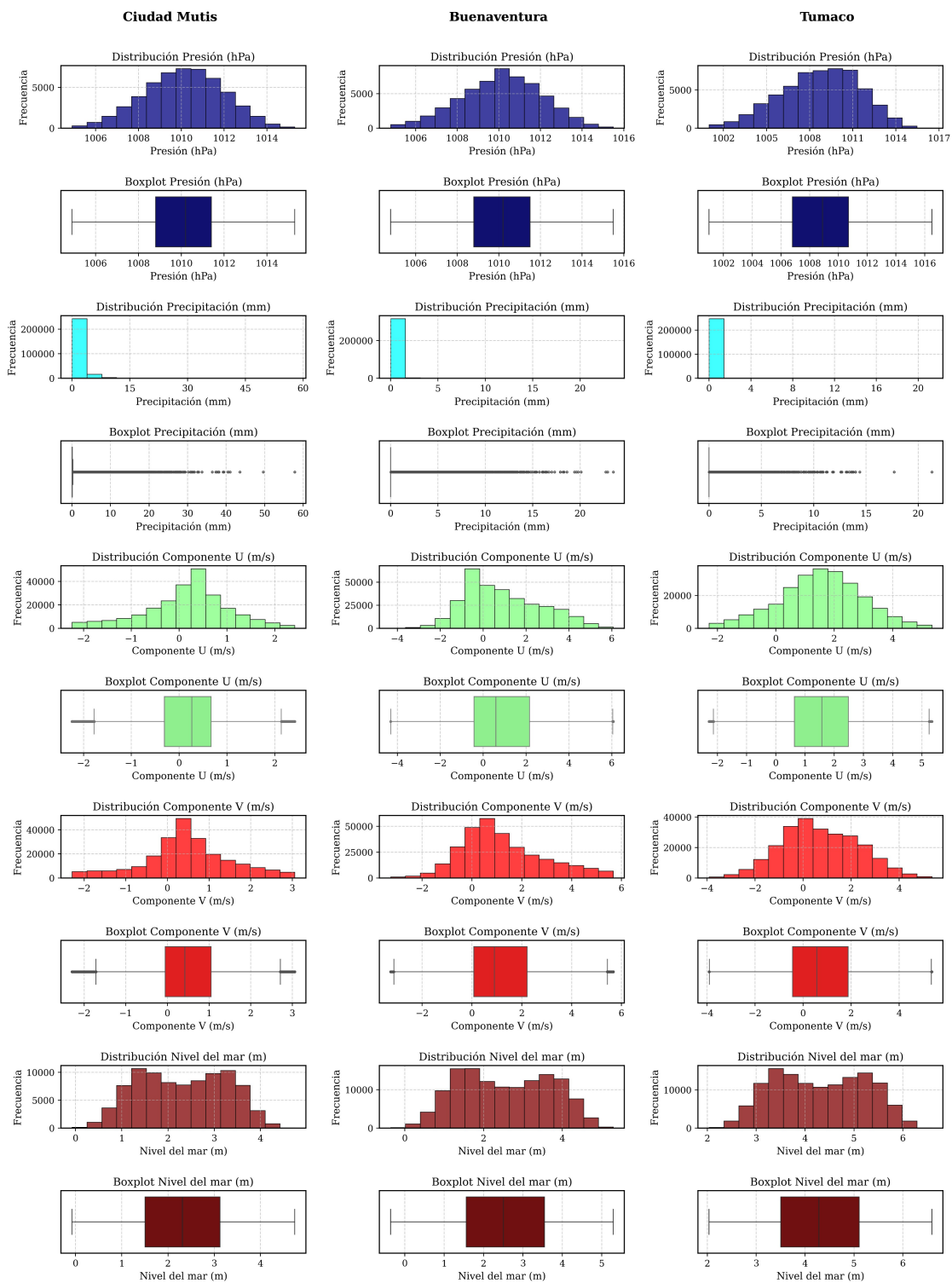


Figura 3.13: Distribución de frecuencias de las variables océano-atmosféricas medidas en puntos fijos.

K^2 fueron considerablemente altos y los p – valores resultaron mucho menores que los umbrales convencionales de significancia (0.01 y 0.001), lo que indica un rechazo contundente de la hipótesis nula de normalidad al nivel de significancia del 1 % y del 0.1 %.

La presión mostró un comportamiento más cercano a una distribución simétrica en comparación con las otras variables, aunque el *test* de D’Agostino también indicó una desviación significativa respecto a la normalidad en todas las localidades analizadas. Los valores del estadístico K^2 oscilaron entre 467.7 en Ciudad Mutis y 1 236.7 en Tumaco, con p – valores extremadamente pequeños (menores a 10^{-100}).

La precipitación acumulada fue la variable que mostró la mayor desviación, con valores de K^2 superiores a 500 000 en Buenaventura y Tumaco, lo que es consistente con su comportamiento típicamente sesgado y discontinuo. Por su parte, las componentes del viento (U y V) también presentaron distribuciones no normales, especialmente en Buenaventura, donde se obtuvieron valores de K^2 cercanos a 15 000. En cuanto al nivel del mar, dada la naturaleza de su variable (bimodal), esta también se aleja de una distribución normal teórica, pese a su aparente regularidad.

3.4.2. Exploración estadística de los datos *LiDAR*

3.4.2.1. Distribución de Frecuencias de las alturas ortométricas.

Ciudad Mutis

El análisis *LiDAR* del casco urbano de Bahía Solano reveló la mayor variabilidad altitudinal entre las áreas urbanas examinadas, con un rango que se extendió desde un mínimo de -1.28 m hasta un máximo considerable de 276.59 m. La altura media para esta área se situó en 27.39 m, mientras que la mediana fue de 13.03 m, con un rango intercuartílico de 23.7 m. Estos valores indicaron una notable dispersión de alturas en torno al promedio y un pronunciado sesgo en la distribución. De hecho, la distribución de frecuencias de estas alturas (cota Z), capturadas mediante *LiDAR* sobre el casco urbano de Bahía Solano (Figura 3.15 A) y presentada en la Figura 3.14, evidenció una clara asimetría hacia las elevaciones superiores. Esta asimetría se reflejó en la distribución de los datos: aproximadamente el 50 % de las alturas se concentró entre -1.28 m y 12 m, el siguiente 25 % se ubicó entre 12 m y 27 m, dejando el 25 % restante distribuido entre 27 m y el valor máximo de 276.59 m.

La Figura 3.15 B presenta la frecuencia espacial de los puntos *LiDAR* por cada celda de la grilla utilizada en el análisis de los datos. Se identificaron celdas con frecuencias de hasta 1 200 puntos en ciertas zonas de la escena analizada. No obstante, la mayor parte del área presentó una baja densidad de puntos, con la excepción de las áreas más despejadas, como las calles del casco urbano y las zonas de playa, donde se registró una mayor frecuencia de datos.

Buenaventura

Las métricas de altura derivadas del procesamiento *LiDAR* en el área urbana de Buenaventura revelaron el rango altitudinal más restringido de las tres zonas analizadas. La elevación máxima registrada fue de 37.81 m, correspondiente a una geomorfología más elevada, en contraste con el mínimo de -2.71 m, propio de áreas de baja mar. Con una media de 3.21 m y una mediana de 1.37 m, el rango intercuartílico de 5.50 m evidenció una menor dispersión altimétrica que en Bahía

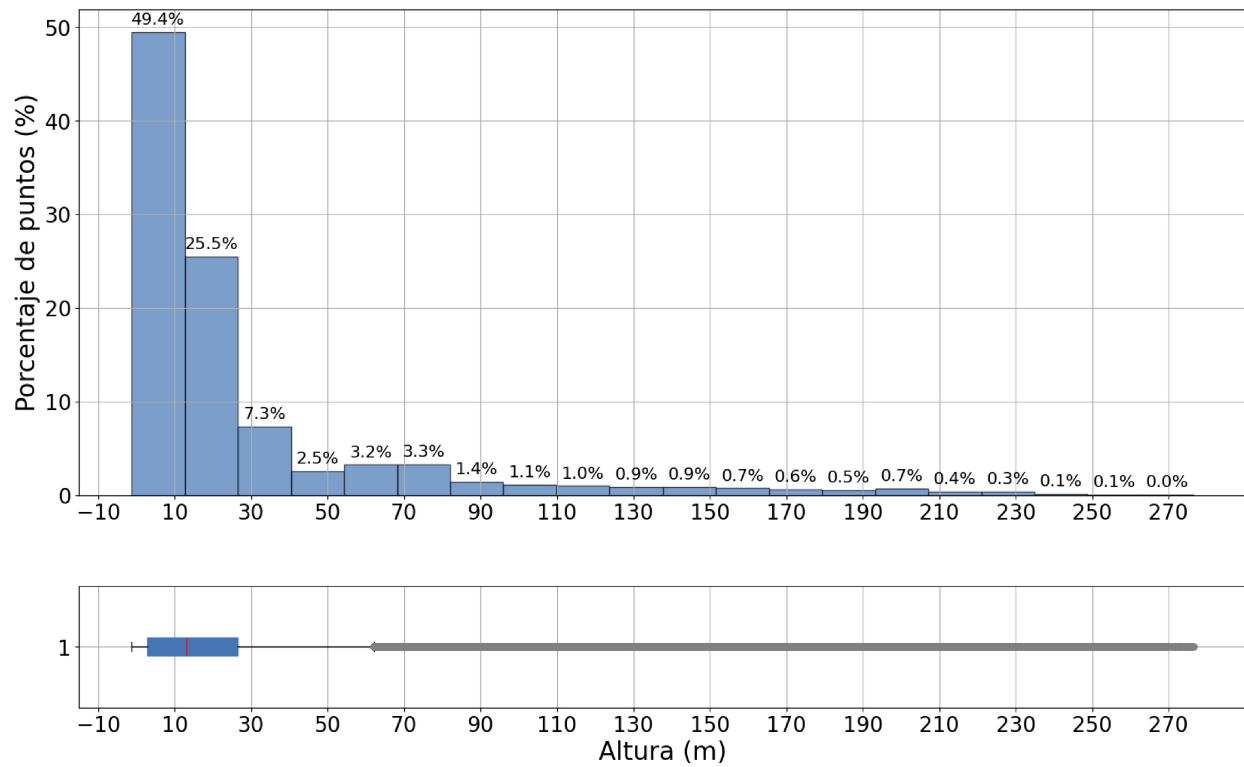


Figura 3.14: Histograma de las alturas ortométricas del casco urbano de Ciudad Mutis.

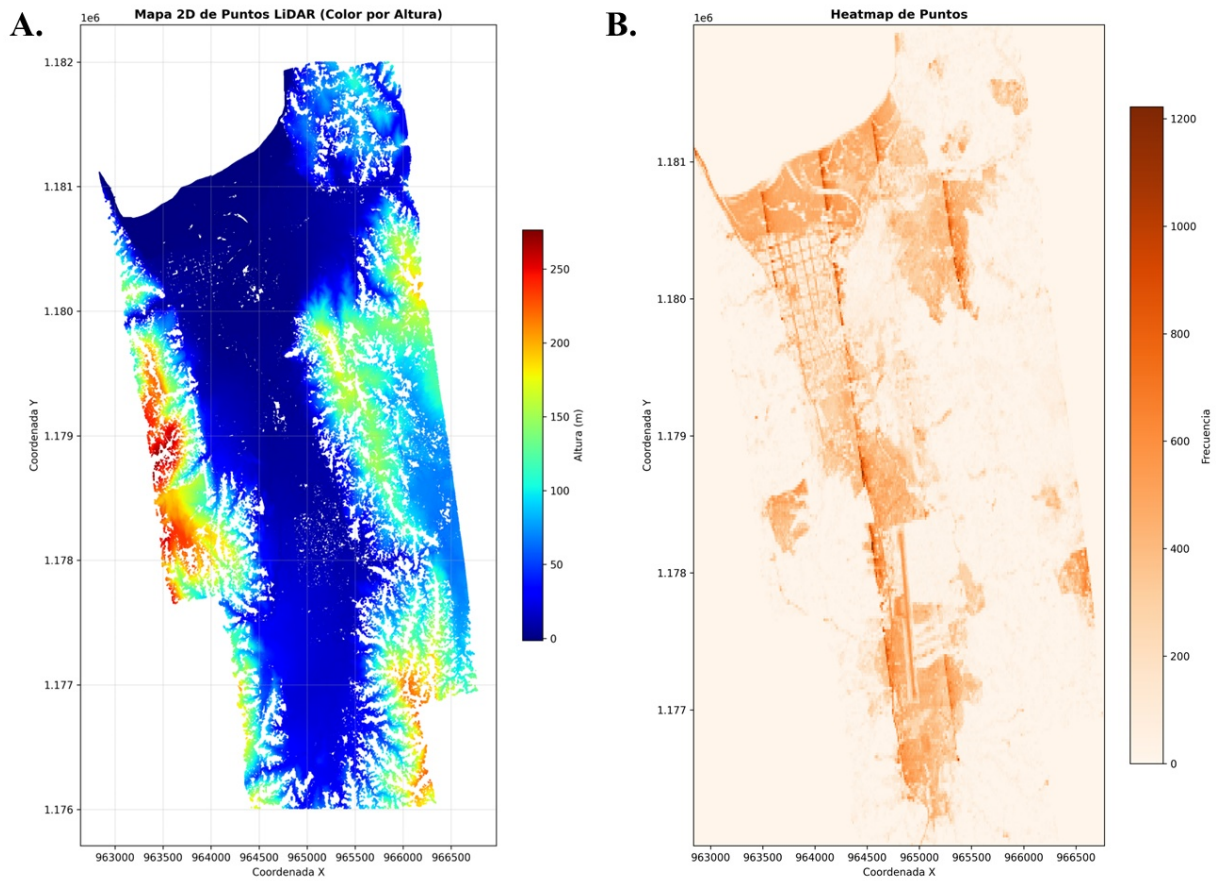


Figura 3.15: A. Alturas ortométricas del casco urbano de Ciudad Mutis. B. Frecuencias espaciales de los retornos de los datos *LiDAR* del casco urbano de Ciudad Mutis.

Solano, aunque persistió un sesgo hacia elevaciones superiores. El histograma de frecuencias (Figura 3.16) ilustró esta distribución, donde aproximadamente el 75 % de los datos se concentró entre los -2.71 m y los 4.8 m, y el 25 % restante entre los 4.8 y los 38 m. El análisis cuantitativo confirmó una asimetría positiva pronunciada (coeficiente de sesgo de 1.8), indicativa de la predominancia de valores bajos con la presencia de valores atípicos en las cotas más altas. Esta configuración sugiere la presencia de extensas áreas planas interrumpidas por estructuras verticales dispersas dentro del tejido urbano estudiado (Figura 3.17 A).

Se observó una mayor concentración de puntos en áreas despejadas como calles y zonas de bajamar (Figura 3.17 B), sugiriendo una menor complejidad superficial, mientras que otras áreas exhiben una menor densidad, posiblemente asociada a la presencia de vegetación o estructuras que influyeron en la cantidad de retornos (Figura 3.17 B).

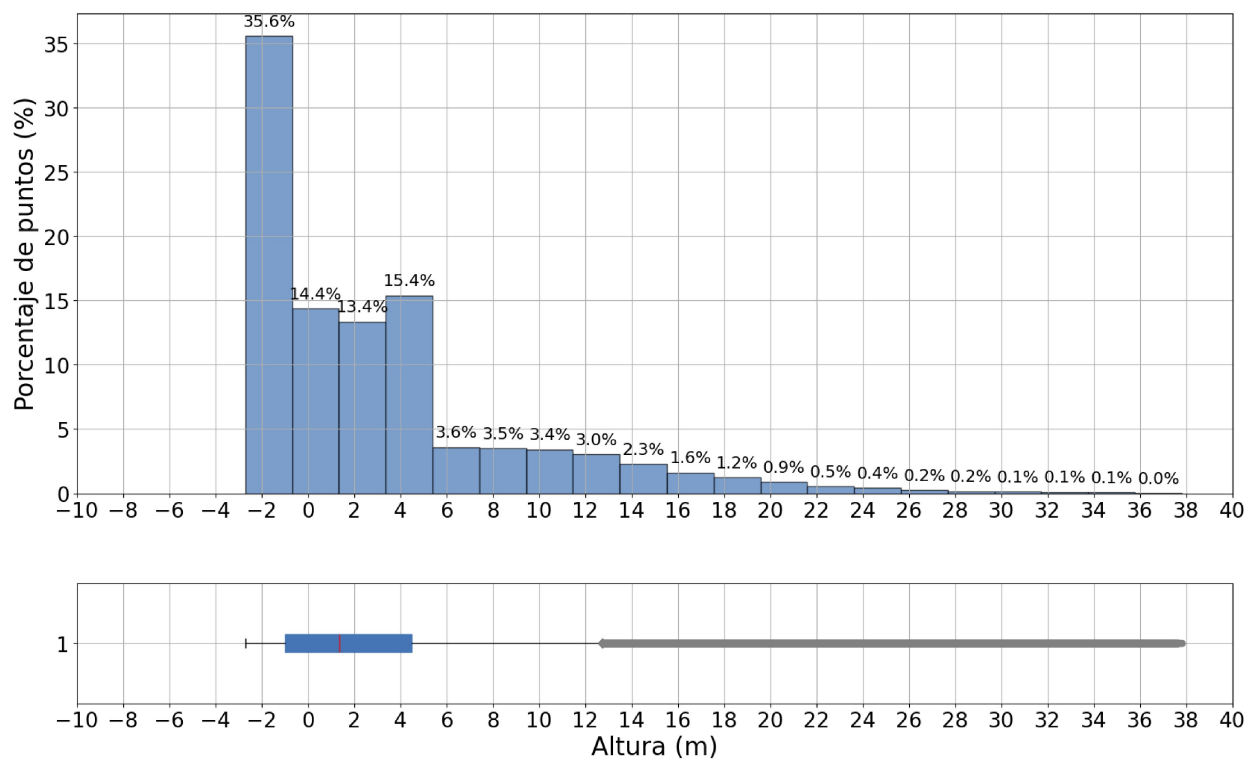


Figura 3.16: Histograma de las alturas ortométricas del casco urbano de Buenaventura.

Tumaco

Las métricas de altura derivadas del análisis *LiDAR* sobre el casco urbano de Tumaco revelaron un rango altitudinal que se extiende desde una altura mínima de -8.48 m hasta una máxima de 80.60 m (Figura 3.18). La altura media registrada en esta zona fue de 2.70 m, con una mediana de 2.50 m, lo cual indica una distribución relativamente homogénea de las elevaciones en torno a este

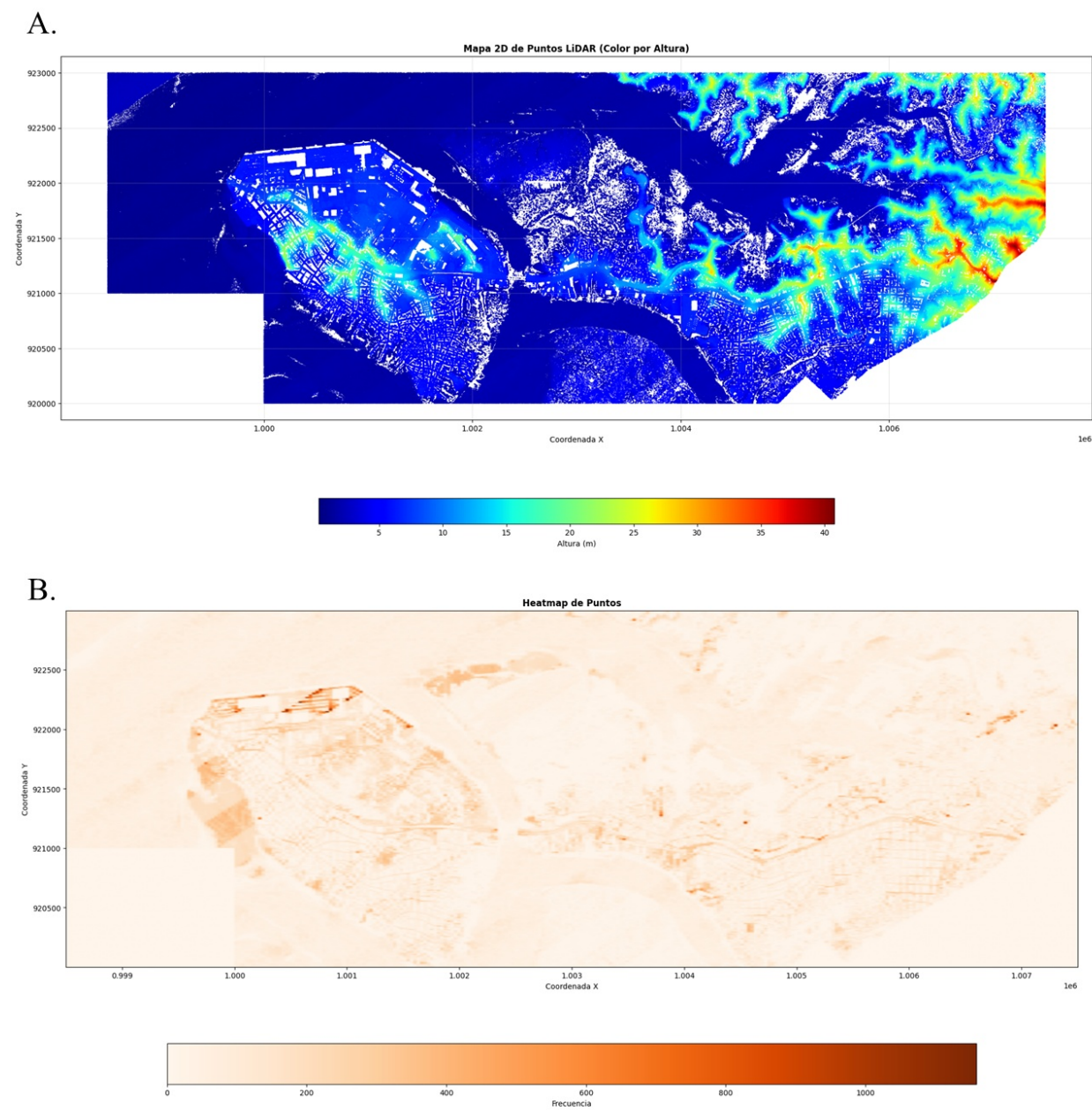


Figura 3.17: A. Alturas ortométricas del casco urbano de Buenaventura. B. Frecuencias espaciales de los retornos de los datos *LiDAR* del casco urbano de Buenaventura.

valor promedio. Más del 90 % de los datos están en elevaciones entre los 1.2 y los 5 m, un 6 % entre los 5 y los 10 m y el resto del 2 % entre los 8 y los 80 m.

El valor máximo de 80.60 m corresponde a una formación en forma de colina en la isla más externa de la ciudad de Tumaco. No obstante, su presencia como valor extremo dentro de una distribución de baja variabilidad (reflejada en la desviación estándar) sugiere que se trata de un dato atípico respecto al grueso de los retornos registrados.

La presencia de valores negativos en la altura mínima sugiere que algunos retornos del sensor *LiDAR* fueron capturados por debajo del nivel del mar o del *datum* de referencia vertical utilizado en el procesamiento. Este tipo de registros suele estar asociado a errores sistemáticos o artefactos provocados por reflexiones en superficies altamente reflectantes, como cuerpos de agua o estructuras metálicas. En zonas urbanas cercanas al litoral, como Tumaco, es común que estos errores se presenten debido a la interacción del haz láser con superficies acuáticas poco estables o con estructuras flotantes.

Esta interpretación se refuerza al observar la distribución estadística de las alturas mostrada en la Figura 3.19, donde el histograma revela una fuerte concentración de valores en el rango de 0 a 10 metros. A su vez, el diagrama de caja permite identificar la presencia de numerosos valores atípicos (*outliers*) por encima del cuarto cuartil, lo cual es consistente con una estructura urbana de baja altura, intercalada con algunas edificaciones de mayor elevación. Estos valores atípicos, aunque no representativos del conjunto, son importantes para caracterizar los elementos más altos del entorno construido. En conjunto, los resultados permiten caracterizar la morfología urbana de Tumaco como predominantemente plana, con escasas elevaciones significativas y un relieve sin grandes pendientes.

3.4.2.2. Análisis de la estructura espacial de la elevación.

Para analizar si existe una estructura espacial en los datos de elevación, es necesario aplicar pruebas de autocorrelación espacial, las cuales permiten identificar patrones de dependencia entre observaciones geográficamente cercanas. En este contexto, el correlograma se convierte en una herramienta exploratoria fundamental, ya que permite cuantificar y visualizar cómo los valores de una variable, como la altitud, se correlacionan en función de la distancia. Esta técnica es especialmente útil en la construcción y validación de modelos digitales del terreno (MDT), donde la continuidad espacial de las elevaciones es un supuesto clave para la interpolación y el modelado geoespacial. Según Malakar *et. al.* [88], la autocorrelación espacial positiva a distancias cortas es una característica común en datos topográficos, y su análisis mediante correlogramas permite determinar la escala de variación espacial, lo cual es esencial para seleccionar el tamaño de celda adecuado en un MDT [88]. Además, investigaciones recientes han demostrado que la incorporación de estructuras de autocorrelación espacial mejora significativamente la precisión de modelos de elevación generados mediante técnicas como *kriging*, regresión geoespacial y correlación digital de imágenes (DIC), esta última ampliamente utilizada para medir deformaciones y variaciones topográficas con alta precisión [88]. En particular, el uso de correlación digital de imágenes en modelos tridimensionales ha permitido capturar con mayor fidelidad las variaciones del terreno, lo cual es crucial en aplicaciones como monitoreo de deslizamientos, erosión y planificación territorial [88].

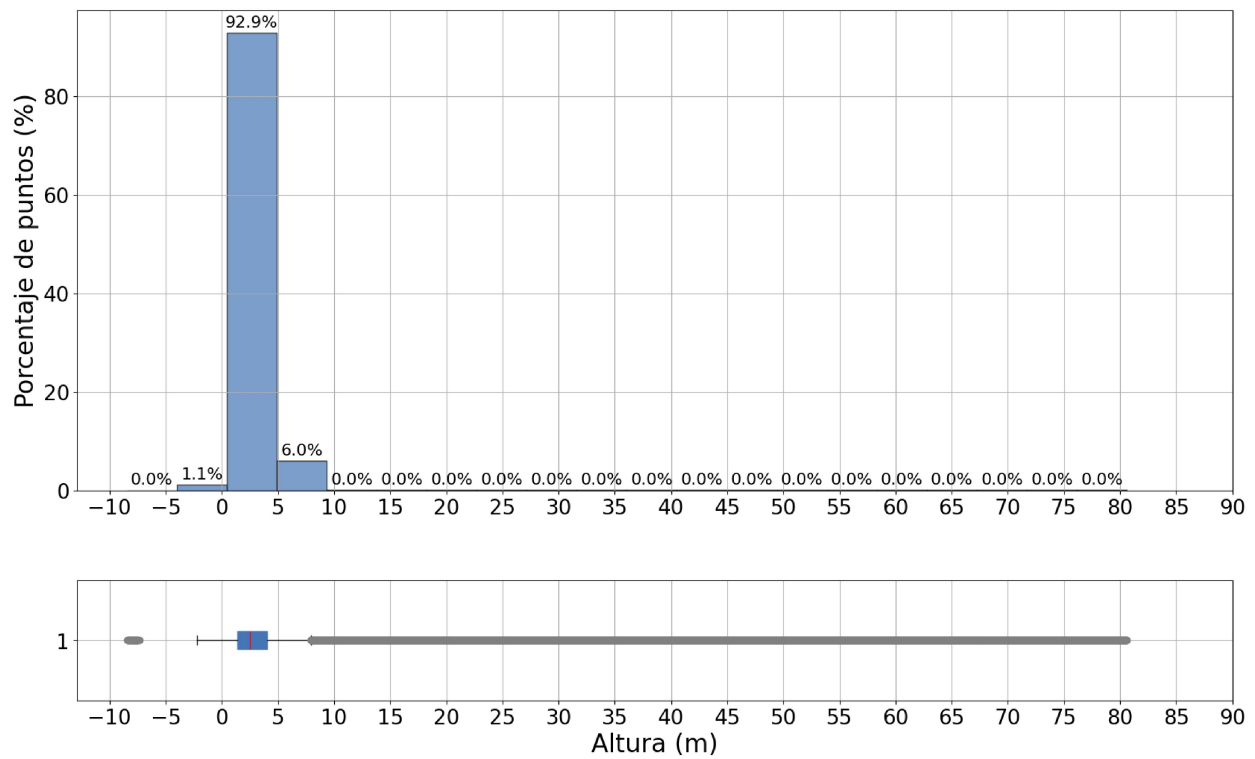


Figura 3.18: Histograma de las alturas ortométricas del casco urbano de Tumaco.

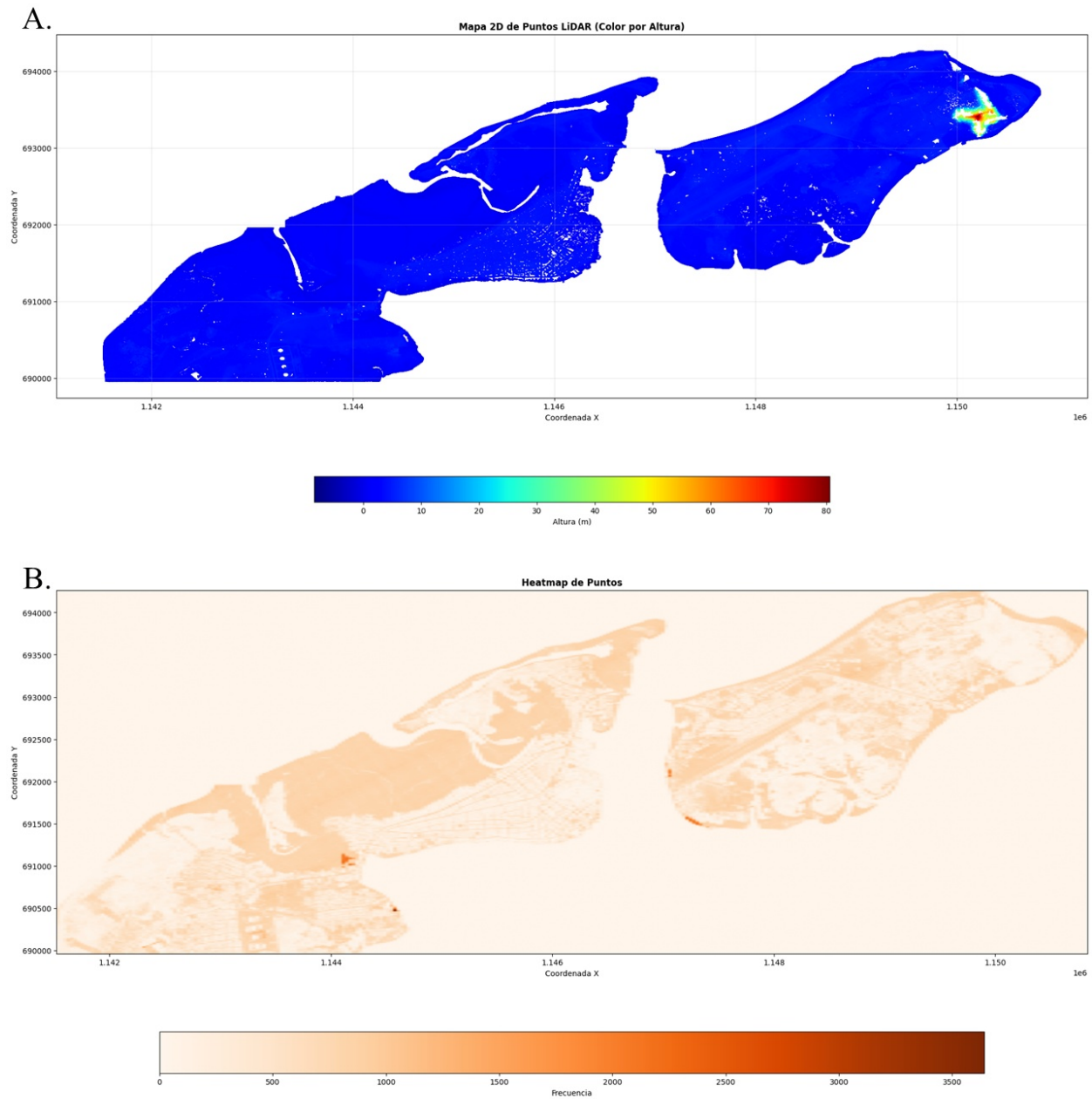


Figura 3.19: A. Alturas ortométricas del casco urbano de Tumaco. B. Frecuencias espaciales de los retornos de los datos *LiDAR* del casco urbano de Tumaco.

Para analizar la estructura espacial de los datos en cada casco urbano, se construyó un correlograma utilizando la librería `scipy.spatial` para calcular las distancias entre todos los pares de puntos. A partir de estas distancias, se agruparon las observaciones en diferentes clases de distancia (*lags*) y se calculó el coeficiente de correlación entre los valores de la variable de interés para cada clase. Este procedimiento permite identificar si existe dependencia espacial —es decir, si los valores cercanos tienden a parecerse más que los distantes.

Con el fin de evaluar la significancia estadística de las correlaciones observadas, se empleó un enfoque de simulación basado en un modelo nulo sin estructura espacial (aleatoriedad completa). Se generaron 100 permutaciones aleatorias de los valores observados, manteniendo fijas las coordenadas espaciales, lo cual rompe cualquier posible patrón espacial. Para cada permutación se calculó un correlograma, obteniendo así una distribución de correlaciones esperadas bajo la hipótesis nula. Con estas simulaciones se construyó un intervalo de confianza (del 95 %) que representa la zona de no significancia. Si las correlaciones observadas superan estos límites, se interpreta como evidencia de una estructura espacial significativa.

Ciudad Mutis

El correlograma de las elevaciones en Ciudad Mutis revela una marcada autocorrelación espacial positiva a distancias cortas (menores a 100 metros), con coeficientes de correlación elevados ($\hat{\rho} \geq 0.75$), lo que indica que las elevaciones cercanas tienden a ser similares entre sí (Figura 3.20). A medida que la distancia entre puntos se incrementa, dicha correlación disminuye progresivamente —alcanzando valores por debajo de 0.75 alrededor de los 250 metros— y se aproxima a cero en torno a los 1 000 metros, lo que sugiere una pérdida gradual de la dependencia espacial.

Notablemente, a partir de distancias superiores a los 1 300 metros, el patrón se invierte ligeramente, presentando valores de correlación negativa ($\hat{\rho} \leq -0.1$), lo cual podría interpretarse como una tendencia hacia una distribución más uniforme de las elevaciones en escalas más amplias. Este comportamiento es típico en fenómenos geográficos donde las condiciones topográficas exhiben continuidad espacial y estructuras organizadas, y refuerza la hipótesis de que las elevaciones en la zona no son aleatorias, sino que responden a un patrón espacial bien definido (Figura 3.15).

Buenaventura

El correlograma de las elevaciones del casco urbano de Buenaventura revela una estructura espacial compleja, caracterizada por una alternancia entre autocorrelación positiva y negativa a lo largo del gradiente de distancias (Figura 3.21). En las distancias más cortas, se observa una autocorrelación espacial positiva moderada a fuerte, con coeficientes de correlación estimados que alcanzan valores cercanos a $\hat{\rho} = 0.75$, lo que indica que las elevaciones cercanas tienden a ser similares entre sí. Este patrón es consistente con la continuidad espacial típica de variables topográficas.

A medida que la distancia entre puntos aumenta, la correlación disminuye progresivamente y se aproxima a cero, lo que sugiere una pérdida gradual de dependencia espacial. Sin embargo, en ciertos tramos del correlograma, especialmente a distancias intermedias y largas, se evidencian valores negativos de $\hat{\rho}$, alcanzando valores de $\hat{\rho} = -0.75$, lo cual indica una posible inversión en la relación espacial: puntos separados por esas distancias tienden a presentar valores contrastantes.

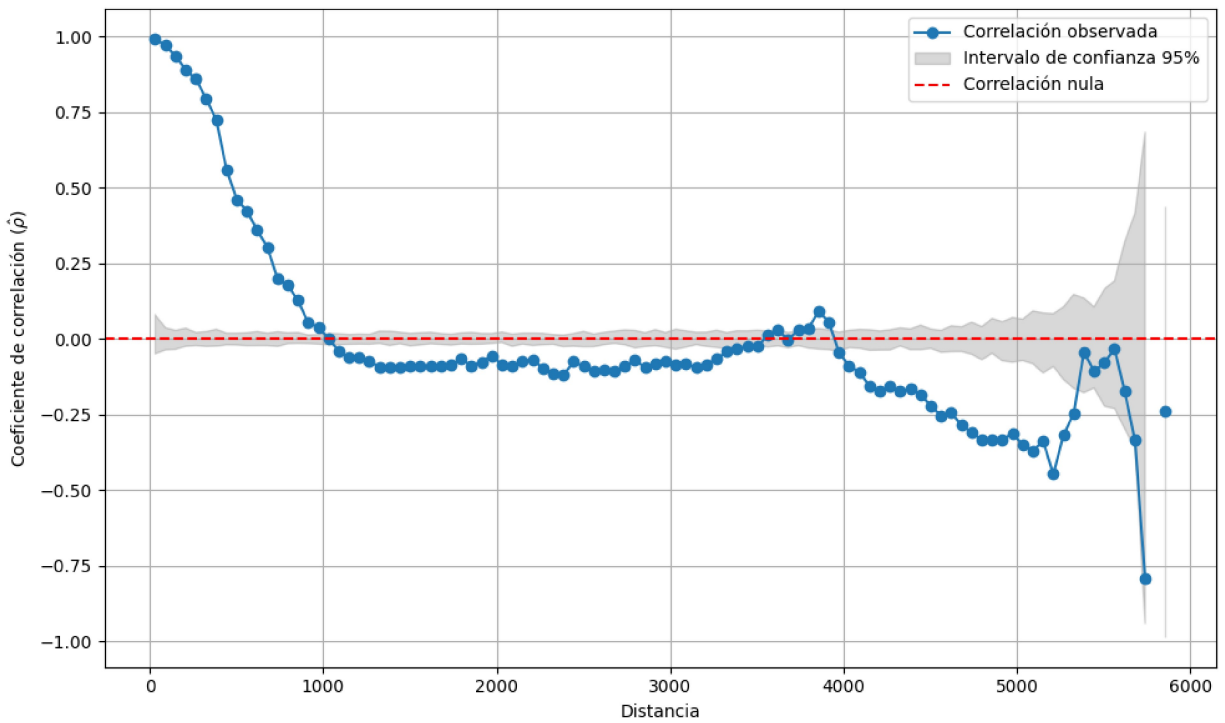


Figura 3.20: Correlograma de las alturas ortométricas del casco urbano de Ciudad Mutis.

Este comportamiento refleja la presencia de estructuras topográficas alternantes, como crestas y depresiones, o patrones periódicos en el relieve en la zona de Buenaventura (Figura 3.17).

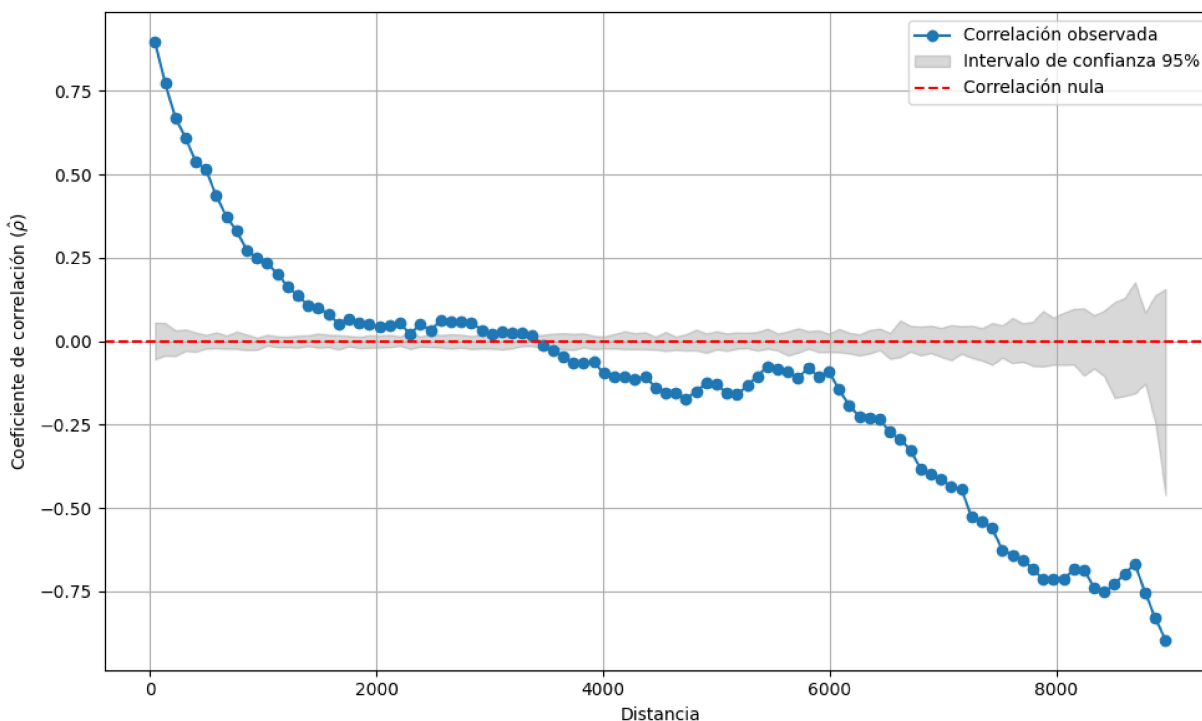


Figura 3.21: Correlograma de las alturas ortométricas del casco urbano de Buenaventura.

Tumaco

El correlograma obtenido para las elevaciones de Tumaco evidencia una estructura espacial significativa, con una marcada variabilidad en los coeficientes de correlación espacial ($\hat{\rho}$) a lo largo del gradiente de distancias (Figura 3.22). En las distancias más cortas, se observa una autocorrelación positiva moderada, con valores de $\hat{\rho}$ cercanos a 0.5, lo que indica que las elevaciones próximas tienden a ser similares entre sí. Este patrón es característico de variables geográficas que presentan continuidad espacial.

A medida que la distancia aumenta, la correlación disminuye progresivamente, alcanzando valores cercanos a cero y posteriormente negativos. En particular, se observan tramos con autocorrelación negativa significativa, con valores de $\hat{\rho}$ que descienden hasta -1.0, lo cual sugiere una alternancia sistemática en los valores de elevación a escalas más amplias. Este comportamiento puede estar asociado a la presencia de estructuras topográficas repetitivas o patrones ondulatorios en el relieve.

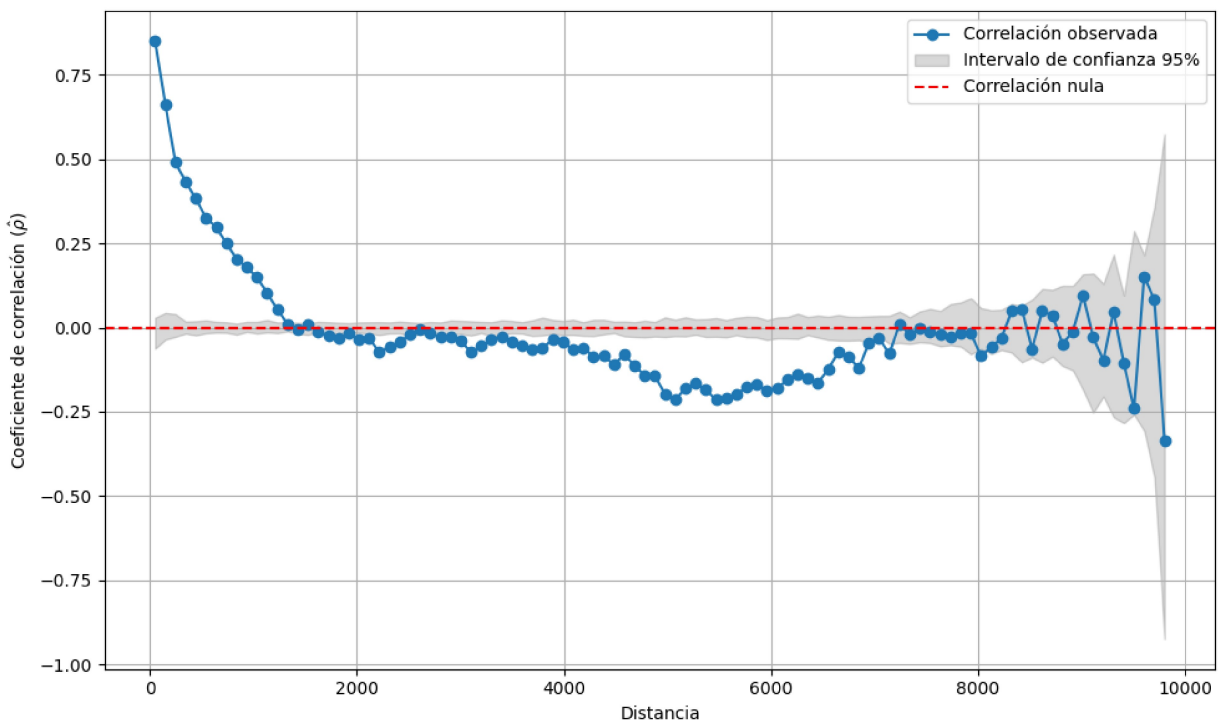


Figura 3.22: Correlograma de las alturas ortométricas del casco urbano de Tumaco.

3.5. Transformación de Datos

3.5.1. Series de tiempo oceanográficas y atmosféricas

3.5.1.1. Transformación de series de tiempo medidas en puntos fijos de observación

Con el objetivo de estandarizar la resolución temporal de este conjunto de series de tiempo y facilitar su análisis, se procedió a la agregación temporal de todas las variables a una resolución mensual. Este proceso consistió en transformar datos originalmente disponibles en frecuencias horarias, diarias y minutas (como las observaciones *in situ* de precipitación y nivel del mar) a valores mensuales, mediante el cálculo de promedios o acumulados, según la naturaleza de cada variable (Figura 3.24).

La elección de la escala mensual responde a múltiples consideraciones técnicas. En primer lugar, el fenómeno ENOS se manifiesta típicamente en escalas de tiempo interanuales y sus fluctuaciones son comúnmente caracterizadas a través de índices climáticos mensuales basados en las anomalías de la temperatura superficial del mar (TSM) en distintas regiones del Pacífico ecuatorial (Figura 3.23). Entre estos destacan:

- **Niño 1+2:** 0°–10°S, 90°W–80°W (costa de Sudamérica).
- **Niño 3:** 5°N–5°S, 150°W–90°W.
- **Niño 3.4:** 5°N–5°S, 170°W–120°W.
- **Niño 4:** 5°N–5°S, 160°E–150°W.

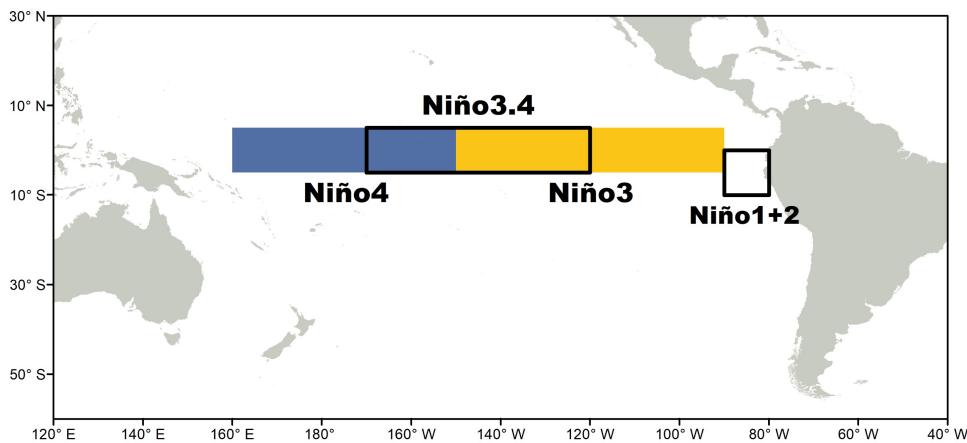


Figura 3.23: Regiones Niño ubicadas en el Pacífico Ecuatorial Tropical. Fuente: CIIFEN

Estos índices se calculan como anomalías mensuales de la TSM respecto a un período base climatológico (usualmente 30 años), y permiten identificar tanto la fase como la intensidad de

eventos El Niño y La Niña [4]. Por tanto, trabajar con series mensuales es coherente con la escala temporal dominante del ENOS y necesario para garantizar la comparabilidad con estas métricas globales.

Además, la mayoría de los productos de reanálisis y datos satelitales utilizados en este estudio, como ERA5 y Copernicus, están también disponibles a escala mensual. Esta homogeneización de la resolución temporal permite una integración adecuada de las fuentes y evita inconsistencias analíticas derivadas de diferentes granularidades temporales.

Es importante destacar que la agregación temporal puede influir en las propiedades estadísticas de las series de tiempo, como la varianza y la autocorrelación. Sin embargo, estudios previos han demostrado que, para análisis climáticos y meteorológicos, la resolución mensual proporciona un equilibrio adecuado entre la captura de variabilidad significativa y la reducción del ruido de alta frecuencia [89], [90].

3.5.1.2. Transformación de series de tiempo con cobertura espacio-temporal (fuentes externas)

Validación de fuentes externas

Una vez obtenidas las series *in situ* con resolución mensual para cada estación de medición ubicadas en los centros poblados de interés, se extrajeron las series de tiempo correspondientes al punto geográfico más cercano a la ubicación de cada estación dentro de los archivos multidimensionales provenientes de fuentes globales (satelitales y de reanálisis). Las coordenadas empleadas fueron:

- **Buenaventura:** lat = 3.85, lon = -77.15
- **Bahía Solano:** lat = 6.20, lon = -77.48
- **Tumaco:** lat = 1.82, lon = -78.73

Este procedimiento permitió emparejar espacialmente los datos *in situ* con las series derivadas de productos globales, facilitando su comparación temporal y la posterior evaluación de su consistencia (Figura 3.25).

Para evaluar la correspondencia entre las observaciones *in situ* y las fuentes de datos externas (satelitales y reanálisis), se calcularon dos métricas estadísticas: la raíz del error cuadrático medio (RMSE) y el coeficiente de correlación de Spearman ($\hat{\rho}$), este último empleado dada la distribución no normal de las variables. Los resultados se presentan en la Tabla 3.6.

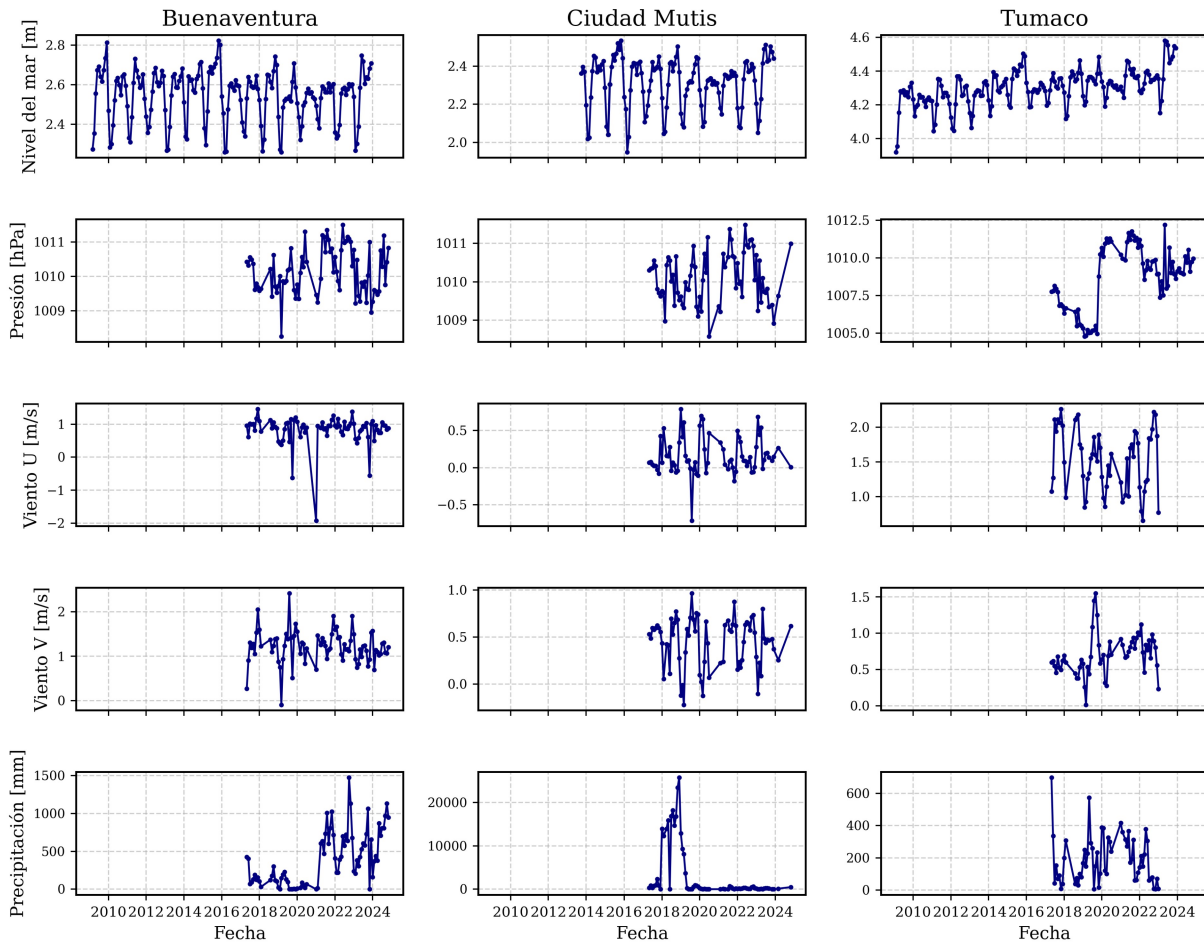


Figura 3.24: Promedios y acumulados mensuales de series de tiempo de variables océano - atmosféricas en las estaciones de Ciudad Mutis, Buenaventura y Tumaco (2009–2024).

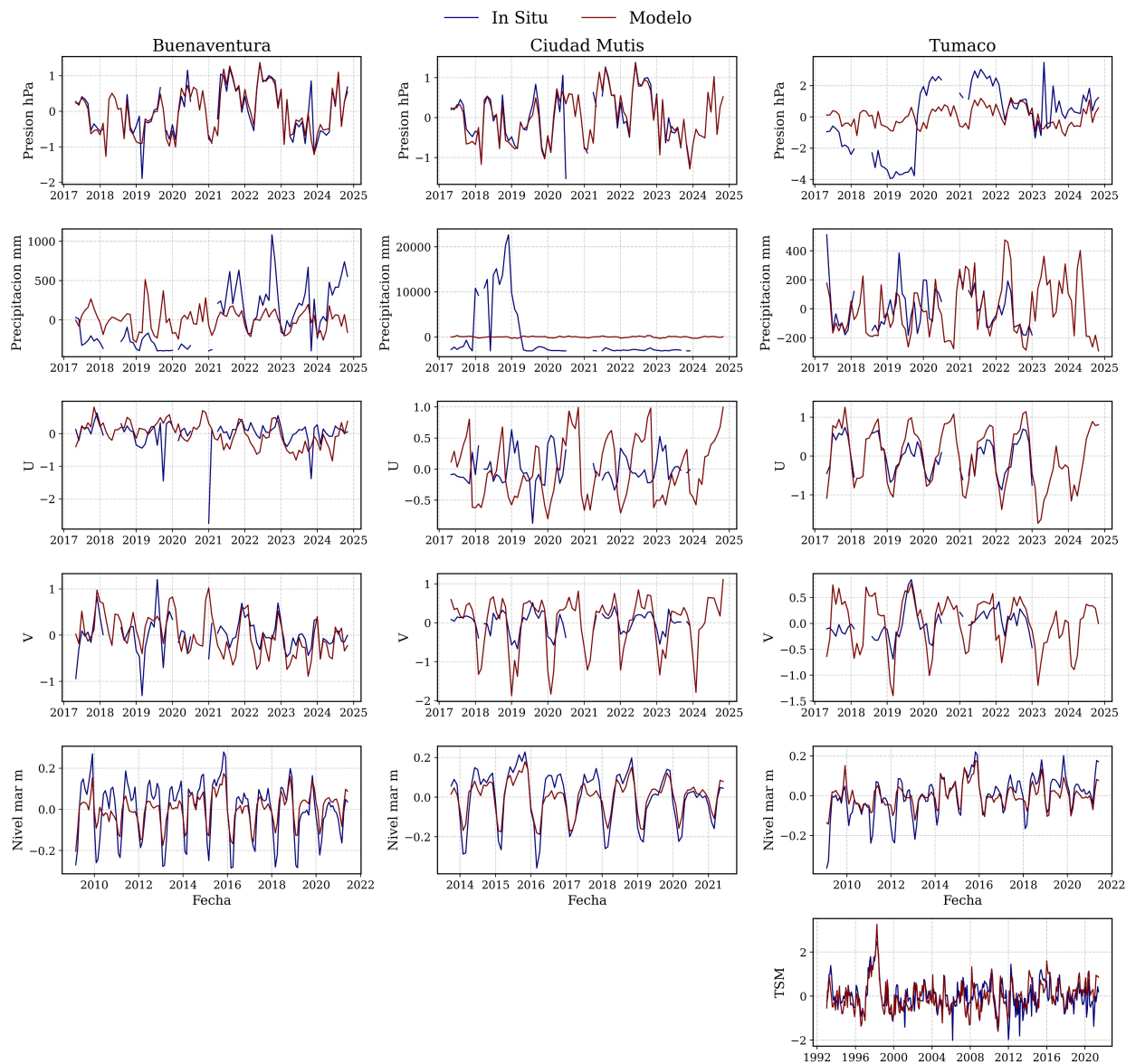


Figura 3.25: Comparación de las series mensuales *in situ* y de fuentes externas por estación y variable (2017–2024).

Tabla 3.6: Métricas estadísticas para la validación de fuentes externas (satelitales/reanálisis) por localidad y variable.

Variable/Estación	Buenaventura		Ciudad Mutis		Tumaco	
	RMSE	$\hat{\rho}_s$	RMSE	$\hat{\rho}_s$	RMSE	$\hat{\rho}_s$
Nivel del mar (m)	0.08	0.76	0.06	0.86	0.06	0.86
Precipitación Acumulada (mm)	338.08	0.36	6 235.25	-0.03	145.55	0.67
Presión (hPa)	0.28	0.90	0.31	0.90	1.88	0.45
TSM	—	—	—	—	0.51	0.63
Componente U	0.54	0.45	0.61	-0.55	0.32	0.93
Componente V	0.38	0.63	0.48	0.77	0.42	0.41

Los mejores niveles de concordancia se evidenciaron en la variable nivel del mar, con valores de $\hat{\rho}$ superiores a 0.75 en todas las estaciones y errores RMSE inferiores a 0.08 m, lo cual respalda la confiabilidad de los productos satelitales para esta variable. La presión media a nivel del mar también mostró una alta correlación en Buenaventura ($\hat{\rho} = 0.90$) y Ciudad Mutis ($\hat{\rho} = 0.90$), aunque en Tumaco la relación fue más limitada ($\hat{\rho} = 0.45$), posiblemente debido a interferencias locales o limitaciones de los sensores.

En contraste, la precipitación presentó una variabilidad espacial considerable. En Ciudad Mutis se observó una correlación prácticamente nula ($\hat{\rho} = -0.03$) y un RMSE superior a 6 000 mm, lo cual podría no deberse exclusivamente a limitaciones de los productos satelitales, sino a posibles fallas o deterioros en el sensor *in situ*, como saturación ante eventos extremos o problemas de mantenimiento y calibración. Estas condiciones comprometen la representatividad de la serie observada. Por el contrario, en Tumaco se obtuvo una correlación moderada ($\hat{\rho} = 0.67$), lo que sugiere un comportamiento más confiable del pluviómetro y una mejor correspondencia entre ambas fuentes de datos.

Respecto al viento, ambas componentes presentaron comportamientos dispares. La componente zonal (U) obtuvo una correlación alta en Tumaco ($\hat{\rho} = 0.93$), en contraste con valores negativos en Ciudad Mutis y moderados en Buenaventura, lo que muestra una mayor coherencia con los datos registrados en la estación de Tumaco. Por su parte, la componente meridional (V) presentó correlaciones de moderadas a bajas, con $\hat{\rho}$ entre 0.41 y 0.77, lo que evidencia que los modelos globales aún tienen dificultades para capturar adecuadamente la dirección y magnitud del viento en regiones costeras con alta complejidad topográfica. Por último, la temperatura superficial del mar (TSM), disponible únicamente en Tumaco, mostró una correlación moderada con los datos satelitales ($\hat{\rho} = 0.63$) y un error RMSE aceptable, indicando una representación razonable del campo térmico oceánico superficial en esta zona.

De forma complementaria, teniendo en cuenta la baja calidad de los datos de precipitación de las series de tiempo *in-situ* para validar la base de datos MSWEP, se realizó un análisis de validación cualitativa, comparando los datos mensuales del conjunto MSWEP y las climatologías históricas generadas por el IDEAM para los meses de marzo, junio y septiembre. La Figura 3.26 muestra

la distribución espacial de la precipitación acumulada para cada uno de estos meses, según ambas fuentes.

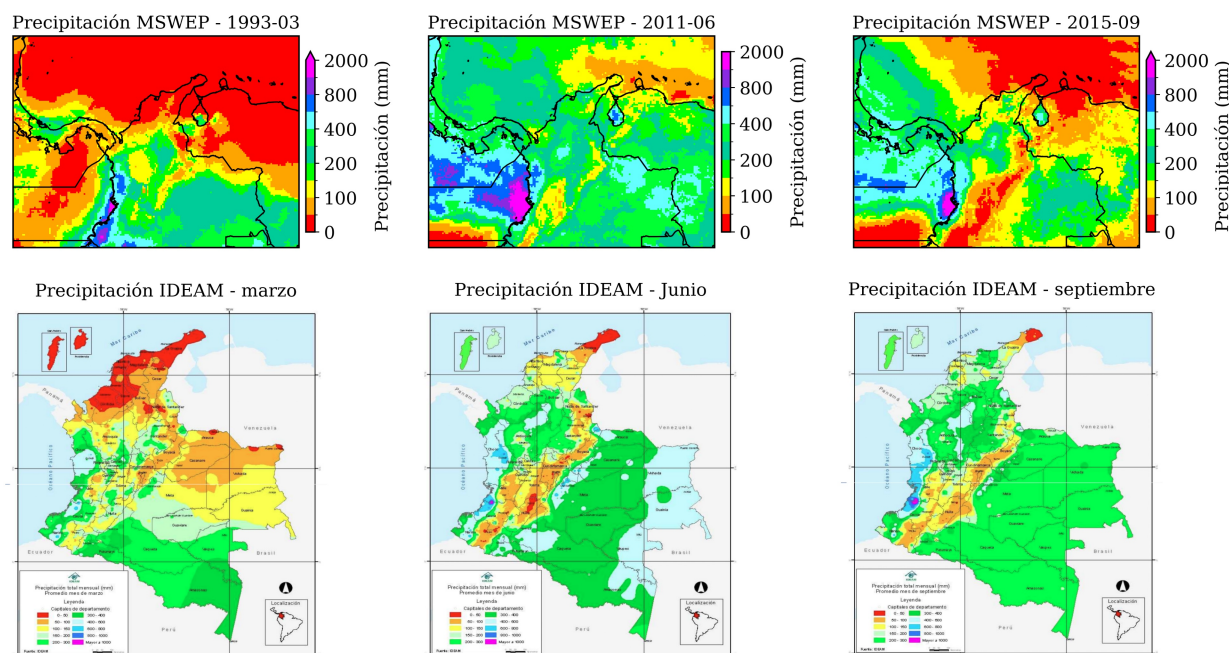


Figura 3.26: Comparación cualitativa de la precipitación mensual para los meses de marzo, junio y septiembre entre la base de datos MSWEP (arriba) y las climatologías del IDEAM (abajo).

En términos generales, se observa que el producto MSWEP logra reproducir de forma cualitativa los principales gradientes regionales de precipitación, especialmente la franja húmeda que se extiende desde el piedemonte amazónico hacia el litoral Pacífico y la región del Darién. Esta coherencia es particularmente evidente en marzo y septiembre, donde tanto MSWEP como IDEAM reflejan máximos de precipitación sobre el Pacífico colombiano (más de 800 mm) y valores bajos hacia la región Caribe y la Orinoquia.

Sin embargo, también se identifican algunas discrepancias. En junio, por ejemplo, MSWEP tiende a sobreestimar los valores sobre la vertiente oriental de los Andes, donde la topografía compleja introduce grandes incertidumbres en las estimaciones satelitales.

Este análisis permitió concluir que, si bien el conjunto de datos MSWEP presenta limitaciones en la representación puntual de la precipitación, particularmente en regiones montañosas o de baja densidad de observación, muestra una capacidad notable para capturar los patrones espaciales predominantes en el territorio colombiano. Esta representación es especialmente adecuada en la franja costera del Pacífico, donde se reproducen de manera coherente los máximos de precipitación observados en los registros climáticos del IDEAM. Esta consistencia espacial, aunque no exenta de errores en zonas específicas, respalda el uso de MSWEP como fuente válida para el análisis de variabilidad

climática regional y como insumo para el alcance de esta investigación.

Con base en las relaciones estadísticas obtenidas entre los datos medidos *in situ* y las fuentes externas (satelitales o de reanálisis), se concluyó que, para la mayoría de las variables analizadas, los niveles de concordancia observados, en términos de correlación de Spearman y errores tipo RMSE, son aceptables desde el punto de vista operativo. Estas evidencias respaldan la utilización de los datos provenientes de fuentes externas como insumos válidos para el desarrollo de modelos predictivos. En consecuencia, se seleccionaron dichas series externas para la implementación de los algoritmos de aprendizaje profundo, considerando además su mayor extensión temporal y cobertura espacial, lo cual permitió robustecer los análisis.

Estandarización de series multidimensionales para análisis climático

Para el procesamiento de datos multidimensionales provenientes de fuentes externas, se empleó el software *Climate Data Operators (CDO)*. *CDO* es una colección de operadores de línea de comandos diseñada para manipular y analizar datos climáticos y de modelos de predicción numérica del tiempo (NWP). Este software admite múltiples formatos de datos, incluidos GRIB 1/2 y NetCDF 3/4, y ofrece más de 600 operadores que permiten realizar funciones estadísticas, aritméticas, selección de datos, submuestreo e interpolación espacial [91]. Dentro de los procedimientos de transformación se destacan los siguientes:

Interpolación: Con el fin de estandarizar la resolución espacial de los diferentes conjuntos de datos provenientes de fuentes globales, se aplicó una interpolación bilineal a una rejilla común de $0.083^\circ \times 0.083^\circ$. Este procedimiento se realizó utilizando la función `remapbil` del software *CDO* [91], la cual implementa un esquema de interpolación bilineal que calcula el valor de una celda objetivo como una combinación lineal ponderada de los cuatro puntos más cercanos del campo original.

Matemáticamente, la estimación del valor interpolado $\hat{X}(x, y)$ en un punto (x, y) dentro de una celda se expresa como:

$$\hat{X}(x, y) = \sum_{j=1}^4 w_j X_j \quad (3.1)$$

donde X_j representa los valores de las cuatro celdas vecinas en el campo de origen y w_j son los coeficientes de ponderación que dependen de la distancia relativa entre el punto objetivo y las celdas vecinas.

Climatología: Se calculó la climatología mensual para el período base 1993–2020 utilizando el operador `ymonmean` del software *CDO* [91]. Este operador permite obtener la media mensual a lo largo de los años, proporcionando así una referencia climática por cada mes del año. Este procedimiento se define matemáticamente como:

$$\bar{X}_m = \frac{1}{N} \sum_{i=1}^N X_{m,i} \quad (3.2)$$

donde \bar{X}_m es la climatología del mes m , $X_{m,i}$ es el valor observado en el mes m del año i , y N es el número total de años del período base.

Anomalías: Las anomalías mensuales se obtuvieron restando la climatología mensual a los valores mensuales observados correspondientes, mediante el operador `ymonsub` de *CDO* [91]. Definida como:

$$A_{m,i} = X_{m,i} - \bar{X}_m \quad (3.3)$$

donde $A_{m,i}$ representa la anomalía del mes m en el año i , $X_{m,i}$ es el valor mensual observado y \bar{X}_m la climatología mensual previamente estimada.

Suavizado mediante media móvil de 12 puntos: Para atenuar el posible ruido inducido por fluctuaciones estacionales y resaltar las señales de variabilidad interanual asociadas al fenómeno ENOS, se aplicó un suavizado mediante una media móvil centrada de 12 meses, utilizando el operador `tinselmean` de *CDO* [91]. Esta técnica permite reducir la variabilidad de alta frecuencia y facilitar la detección de patrones persistentes de mayor escala. La ecuación utilizada corresponde a una media móvil centrada simétrica:

$$S_t = \frac{1}{2k+1} \sum_{j=-k}^k A_{t+j} \quad (3.4)$$

donde S_t representa el valor suavizado en el tiempo t , A_{t+j} son las anomalías en los tiempos adyacentes, y $k = 6$ para una ventana de 12 meses. Esta transformación es especialmente útil en estudios climáticos, ya que el ENOS se manifiesta principalmente en escalas interanuales, y su señal puede verse enmascarada por componentes estacionales de corta duración.

Estandarización utilizando Z-score: Finalmente, las anomalías suavizadas fueron transformadas mediante una estandarización basada en el Z-score, con el fin de unificar la escala de las variables y permitir su comparación entre dominios y variables distintas. La ecuación utilizada fue:

$$Z_{m,i} = \frac{A_{m,i} - \mu_m}{\sigma_m} \quad (3.5)$$

donde $Z_{m,i}$ es el valor estandarizado, $A_{m,i}$ es la anomalía, μ_m es la media de las anomalías del mes m y σ_m es la desviación estándar correspondiente. Esta transformación convierte la serie en una escala adimensional centrada en cero con desviación estándar unitaria, facilitando el análisis comparativo multivariado.

La Figura 3.27 representa la integración de todos los procedimientos de transformación aplicados a las series de tiempo espacio-temporales: cálculo de climatología, anomalías, suavizado y estandarización. Como ejemplo ilustrativo, se muestran las anomalías correspondientes a enero del año 2000, donde se aprecia la distribución espacial de las desviaciones respecto a la climatología mensual de referencia, así como la identificación de regiones con valores anómalos significativos para cada una de las variables en la Bahía de Panamá.

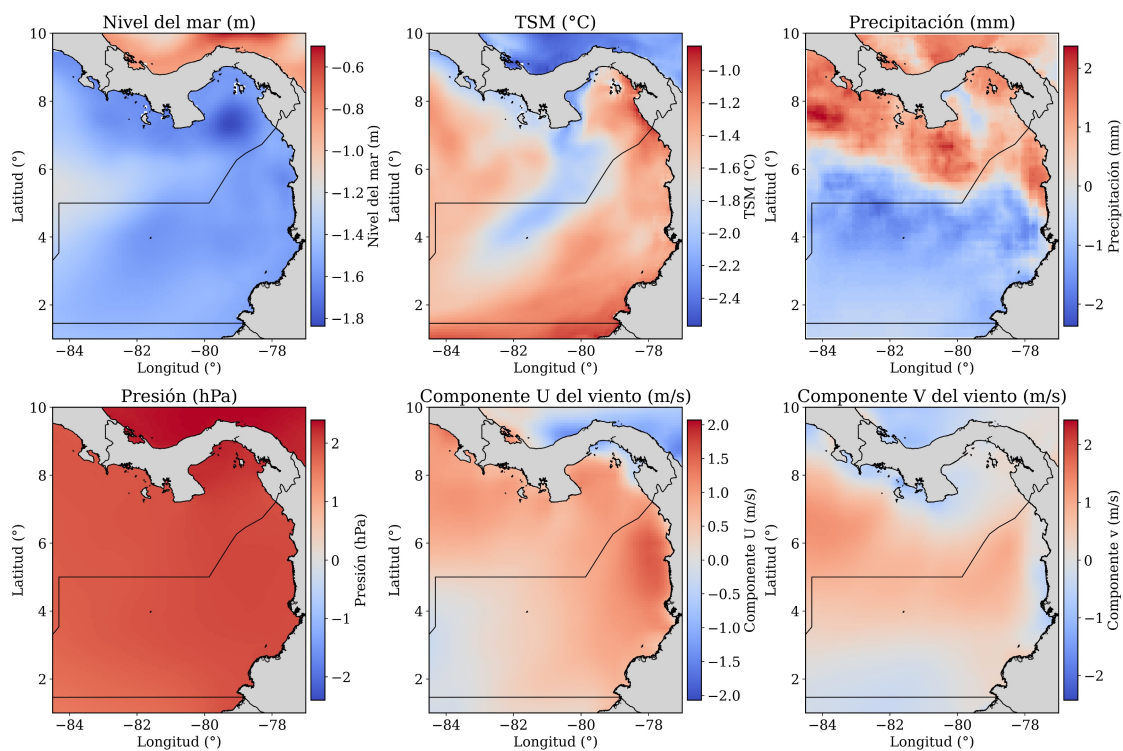


Figura 3.27: Anomalías de enero del 2000 de variables océano atmosféricas, suavizadas con una media móvil de 12 puntos y estandarizadas.

3.5.2. Conversión de alturas *LiDAR* a sistemas de referencia vertical: *MSL* y *LAT*

Los datos *LiDAR* obtenidos para esta investigación fueron capturados en el sistema de referencia geodésico MAGNA-SIRGAS-BOGOTA-OESTE (Ciudad Mutis y Buenaventura) y MAGNA-SIRGAS-BOGOTA-OESTE-OESTE (Tumaco), lo cual implica que las alturas asociadas a cada retorno corresponden a alturas elipsoidales (h), es decir, distancias verticales medidas desde la superficie del elipsoide de referencia global. Sin embargo, para propósitos de análisis geográfico, modelación del terreno e interpretación en contextos costeros, es necesario convertir estas alturas a sistemas de referencia vertical más significativos desde el punto de vista físico y operativo, como el nivel medio del mar (*Mean Sea Level - MSL*) y el nivel más bajo de marea astronómica (*Lowest Astronomical Tide - LAT*).

Para convertir las alturas elipsoidales a alturas ortométricas referidas al *MSL*, se utilizó la ondulación geoidal (N) obtenida a partir del modelo geoidal de Colombia GEOCOL 2004, el cual proporciona la separación entre el elipsoide MAGNA-SIRGAS y el geoide a nivel de Suramérica con una resolución de aproximadamente 1 minuto de arco. La relación utilizada fue:

$$H = h - N \quad (3.6)$$

donde H es la altura ortométrica (referida al *MSL*), h es la altura elipsoidal proveniente del sensor *LiDAR*, y N es la ondulación geoidal en el punto geográfico correspondiente. Para la implementación computacional, se empleó un *script* en Python, lo cual permitió transformar de forma automatizada los millones de puntos *LiDAR* registrados, asegurando precisión submétrica en la conversión.

Posteriormente, con el fin de representar las alturas en un sistema relevante para aplicaciones costeras y náuticas, se realizó una conversión adicional de las alturas ortométricas a alturas referidas al (*LAT*), *datum* comúnmente utilizado en batimetría y cartas náuticas. Esta transformación implicó restar a las alturas ortométricas el valor de la diferencia local entre el *MSL* y el *LAT*, información obtenida a partir de los registros mareográficos oficiales suministrados por la DIMAR de Colombia para los puertos de Ciudad Mutis, Tumaco y Buenaventura. La expresión matemática empleada fue:

$$Z_{LAT} = H - (MSL - LAT) \quad (3.7)$$

Donde Z_{LAT} representa la altura con respecto al nivel *LAT*, H es la altura ortométrica, y $(MSL - LAT)$ es la diferencia entre ambos niveles de referencia. Esta corrección permitió que los modelos de elevación obtenidos sean compatibles con el *datum* náutico, facilitando su integración con datos batimétricos y mejorando la interpretación de escenarios de inundación, vulnerabilidad costera y planificación territorial.

Modelado

4.1. Contexto

El modelado de datos constituyó una etapa central en esta investigación, traduciendo los datos ambientales en herramientas predictivas y de representación espacial. Esta fase se enfocó en la estimación de la respuesta del nivel del mar a los forzantes climáticos, así como en la generación de superficies de elevación costera para el análisis de exposición.

En este contexto, la metodología se centró en el desarrollo de modelos de aprendizaje profundo para la predicción del nivel del mar, particularmente utilizando una red neuronal de tipo recurrente con memoria a largo plazo (*Long Short-Term Memory, LSTM*). Este modelo fue seleccionado por su reconocida capacidad para aprender patrones temporales complejos en secuencias de datos, lo cual es indispensable para comprender la dinámica del nivel del mar influenciada por variables oceánicas y atmosféricas asociadas al fenómeno El Niño-Oscilación del Sur (ENOS).

De forma complementaria, se generaron Modelos Digitales de Elevación (MDE) a partir de datos *LiDAR* de alta resolución, empleando técnicas de interpolación geoestadística como el *kriging* y técnicas con algoritmos deterministas. Estas técnicas permitieron representar de forma precisa la morfología costera, elemento clave para integrar las predicciones del nivel del mar en contextos espaciales y realizar análisis de exposición y vulnerabilidad.

Las secciones siguientes describen, en primer lugar, la arquitectura, el entrenamiento y la evaluación de la red neuronal *LSTM*, junto con los criterios de selección de hiperparámetros, los procedimientos de validación implementados y los resultados obtenidos en términos de rendimiento predictivo. Adicionalmente, se presentan los procesos de construcción de Modelos Digitales de Elevación (MDE) a partir de datos *LiDAR*, incluyendo el ajuste de semivariogramas experimentales, la implementación de técnicas de interpolación geoestadística como el *kriging* y la evaluación de métodos determinísticos tales como la interpolación lineal, cúbica y por vecinos más cercanos. Se detallan los parámetros empleados, las limitaciones computacionales enfrentadas y los resultados de cada técnica en términos de fidelidad topográfica y coherencia espacial. Esta etapa de modelado representó la integración de los datos recopilados previamente, aportando no solo herramientas predictivas para el nivel del mar, sino también insumos esenciales para caracterizar la morfología costera y evaluar la vulnerabilidad frente a escenarios de sobre elevación marina en el Pacífico colombiano.

4.2. Modelado de series de tiempo

4.2.1. Selección de variables

4.2.1.1. Análisis exploratorio de asociaciones oceano-atmosféricas

Dado que los resultados del *test* de normalidad (Tabla 3.5) indicaron que las series de tiempo analizadas no siguen distribuciones normales, se optó por emplear el coeficiente de correlación de Spearman para explorar las asociaciones entre las variables oceánicas y atmosféricas consideradas en el estudio.

La Figura 4.1 muestra la distribución conjunta de las variables estandarizadas junto con sus respectivos coeficientes de correlación de Spearman ($\hat{\rho}$). Aunque no todas las relaciones son estrictamente lineales, se identificaron asociaciones claramente crecientes o decrecientes entre varios pares de variables. Se destaca, en particular, la correlación positiva alta entre el nivel del mar y la temperatura superficial del mar (TSM), con un coeficiente de $\hat{\rho} = 0.86$. Esta relación es coherente con los procesos de expansión térmica del agua marina, en los cuales el calentamiento de las capas superficiales del océano incrementa su volumen, elevando el nivel del mar, un mecanismo ampliamente documentado como una de las principales causas del aumento del nivel del mar en el contexto del cambio climático [92], [93].

Asimismo, se observaron correlaciones negativas moderadas entre el nivel del mar y la presión ($\hat{\rho} = -0.46$), así como con la componente zonal del viento (U; $\hat{\rho} = -0.56$). Estos patrones reflejan mecanismos de forzamiento atmosférico que modulan la distribución regional del nivel del mar, tales como el efecto barométrico inverso y el transporte inducido por el viento a través del estrés superficial [94].

Por otro lado, la precipitación evidenció una correlación muy baja con las demás variables ($|\hat{\rho}| < 0.10$), lo cual se atribuye a su elevada variabilidad espacio-temporal. También se destaca una correlación positiva relativamente alta entre las componentes zonal y meridional del viento ($\hat{\rho} = 0.73$), lo que indica cierta coherencia en la dirección predominante de los vientos en la región.

Este análisis de correlación bivariada sirvió como punto de partida para la selección preliminar de variables predictoras. Las variables TSM, presión media al nivel del mar y viento zonal mostraron tanto consistencia estadística como relevancia física, lo que respalda su inclusión en los modelos predictivos del nivel del mar.

No obstante, es importante advertir que la presencia de correlación entre variables predictoras, especialmente si supera el 90 %, puede generar redundancia o problemas de multicolinealidad, lo cual podría limitar la capacidad de generalización del modelo y favorecer el sobreajuste. Aunque en este caso ninguna relación entre variables superó dicho umbral, se implementó un análisis complementario (análisis multivariado) para manejar de forma controlada estas interdependencias.

4.2.1.2. Análisis de correlación cruzada con índices ENOS

Dado que el análisis exploratorio evidenció una fuerte asociación entre la TSM y el nivel del mar, se consideró pertinente incorporar los índices ENOS como posibles predictores en el proceso de modelado. Esta decisión se fundamenta en que dichos índices representan anomalías térmicas del

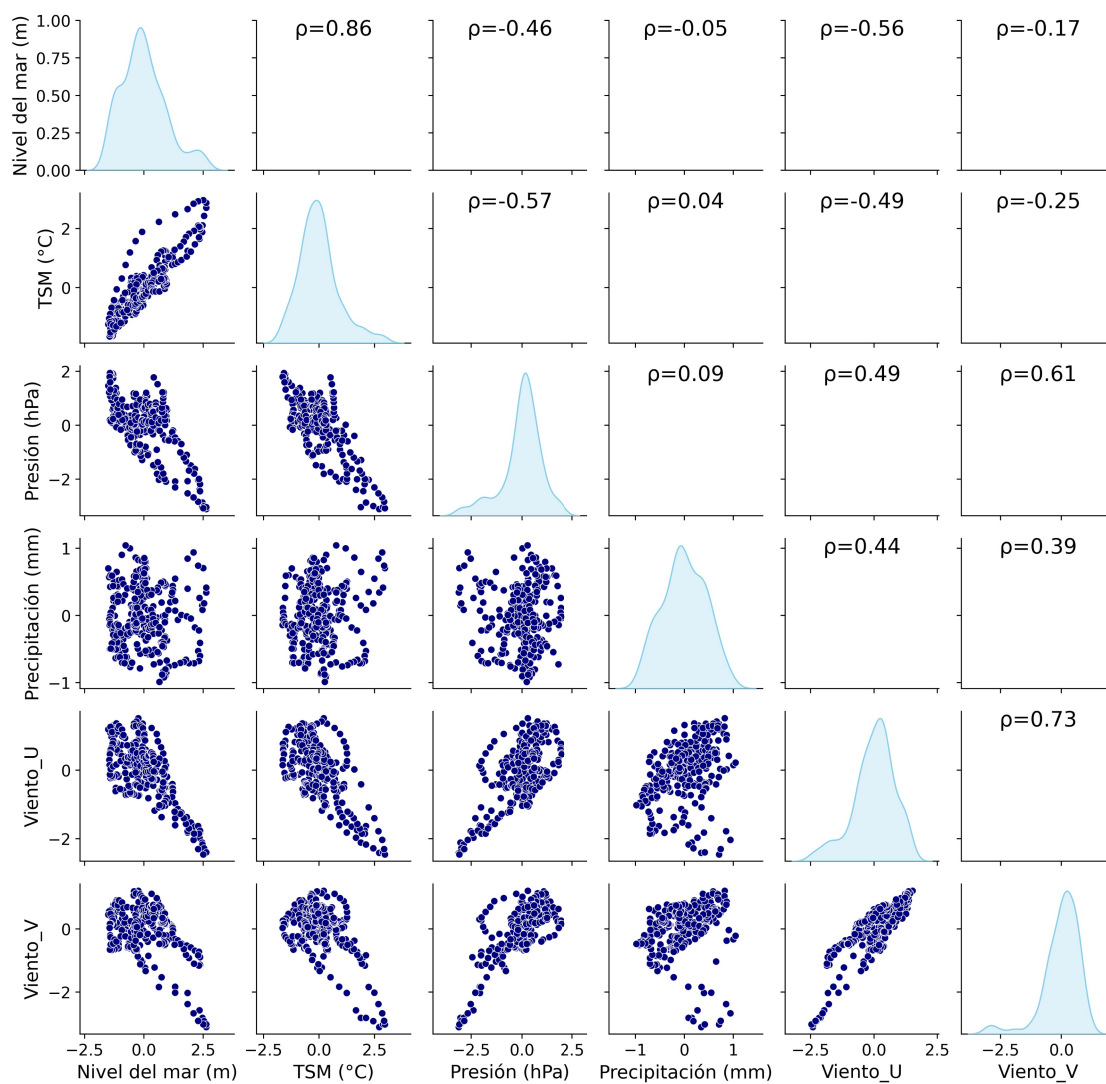


Figura 4.1: Análisis de correlación de Spearman entre variables oceánicas y atmosféricas estandarizadas. Los valores indican el coeficiente $\hat{\rho}$ y se visualiza la distribución conjunta de cada par de variables.

agua en el Pacífico ecuatorial (figura 3.23), capturando las diferentes modalidades de génesis del fenómeno ENOS [40], [95]. En particular:

- El índice **Niño 1+2** refleja eventos de tipo costero, con anomalías térmicas frente a las costas de Sudamérica.
- Los índices **Niño 3** y **Niño 3.4** representan condiciones del Pacífico central-oriental, caracterizando eventos canónicos con fuerte acoplamiento océano-atmósfera.
- El índice **Niño 4** identifica anomalías en el Pacífico central, típicas de eventos Modoki.

La inclusión de estos índices permitió incorporar explícitamente la señal climática interanual del ENOS en el conjunto de variables predictoras, reconociendo su influencia indirecta sobre el nivel del mar mediante procesos remotos de acoplamiento océano-atmósfera. Para asegurar una representación coherente y libre de ruido estacional, los índices fueron suavizados mediante una media móvil centrada de 12 meses y estandarizados posteriormente utilizando la técnica del Z-score, en coherencia con el resto de las variables empleadas.

Posteriormente, se realizó un análisis de correlación cruzada entre las series de tiempo de los índices ENOS y el nivel del mar mensual observado en la región del Pacífico colombiano, con el objetivo de identificar el desfase temporal (*lag*) óptimo en el que se alcanzaba la máxima asociación. Esta metodología permitió evaluar la respuesta retardada del nivel del mar frente a las anomalías térmicas en el Pacífico ecuatorial, aspecto fundamental para mejorar el diseño del modelo predictivo.

Los resultados de este análisis evidenciaron una señal significativa en toda la región costera del Pacífico colombiano, con coeficientes de Spearman superiores a 0.8 en la mayoría de los puntos (Figura 4.2). Esto demuestra la fuerte influencia de las condiciones térmicas del Pacífico ecuatorial sobre la dinámica del nivel del mar en la zona de estudio.

Particularmente, los índices Niño 1+2 y Niño 3 mostraron mayor coherencia espacial con el nivel del mar en el Pacífico colombiano, tanto en magnitud como en cobertura. Para el Niño 1+2, la correlación alcanzó valores máximos con un desfase de 1 a 2 meses en el sector costero suroccidental, y entre 2 y 4 meses en la zona centro y norte. Esta respuesta es coherente con la proximidad geográfica de la región Niño 1+2 al litoral colombiano y la ocurrencia de eventos tipo Niño Costero.

En contraste, el índice Niño 3 presentó una señal espacialmente más homogénea y altamente correlacionada con el nivel del mar, aunque con un desfase predominante de 2 meses en toda la cuenca, salvo en la zona norte, donde se observaron desfases entre 2 y 3 meses para la correlación máxima. Este comportamiento sugiere un efecto retardado típico de eventos El Niño canónicos, pero con repercusiones más rápidas sobre las costas colombianas que las generadas en la región Niño 1+2.

El índice Niño 3.4 también mostró correlaciones significativas, aunque algo menores que las del Niño 3, y con desfases similares (1 a 3 meses). Por su parte, el índice Niño 4 evidenció correlaciones significativas en todo el dominio, pero con desfases más variables (de -1 a 6 meses). Esta dispersión podría deberse a que el Niño 4, asociado a eventos tipo Modoki, ejerce una influencia más débil y difusa sobre el nivel del mar en el Pacífico colombiano.

En conjunto, estos hallazgos permitieron entender la conexión entre la variabilidad oceánica remota y la dinámica regional del nivel del mar. No obstante, dado que estos índices representan

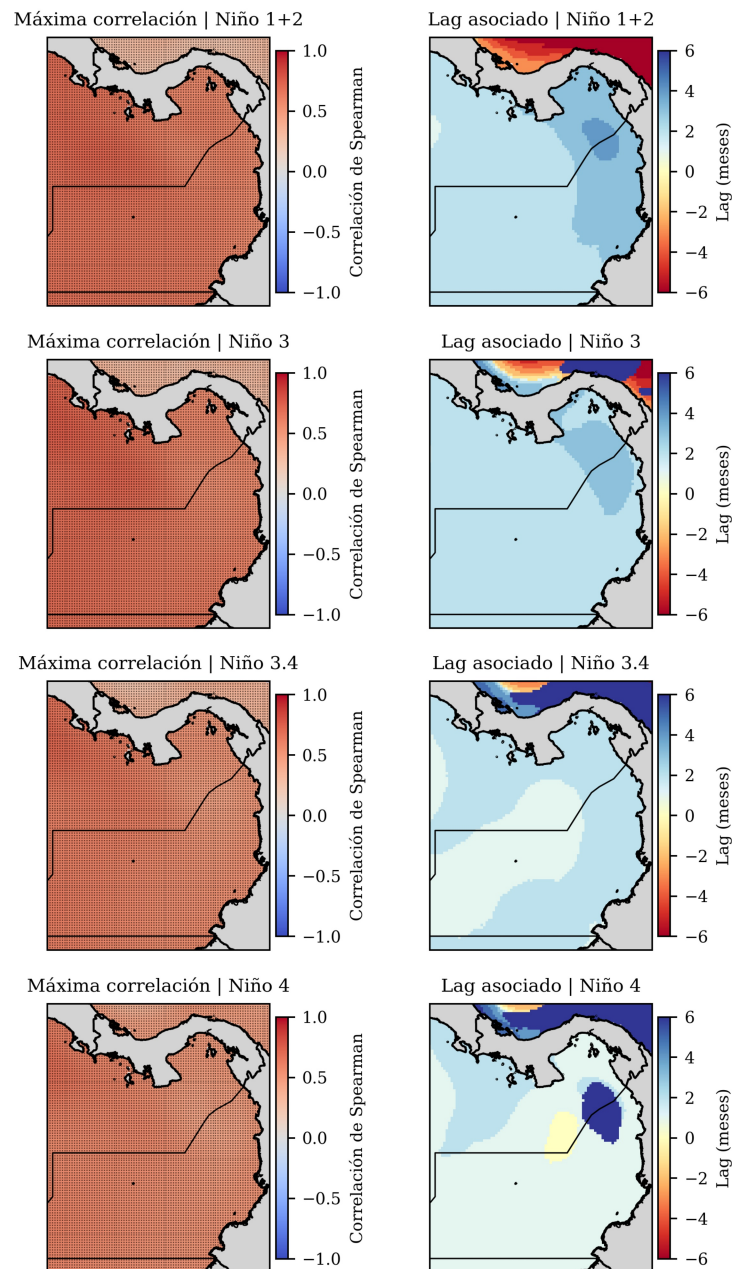


Figura 4.2: Correlación cruzada de Spearman entre los índices ENOS (Niño 1+2, Niño 3, Niño 3.4, Niño 4) y el nivel del mar. A la izquierda se muestra el valor máximo estimado de correlación para cada punto (los puntos negros indican significancia al 5%), y a la derecha el lag (en meses) en el que esta ocurre.

promedios regionales y sintetizan la señal climática en un único valor, se exploró una estrategia alternativa basada en la combinación de todas las variables predictoras mediante Análisis de Correlación Canónica Regularizada (rCCA), como se detalla en la siguiente subsección.

4.2.1.3. Análisis de patrones de asociación multivariados mediante correlación canónica

Para identificar patrones de asociación multivariada entre el nivel del mar y las variables climáticas potencialmente predictoras, se evaluó en primer lugar la validez de la hipótesis de normalidad multivariante, condición fundamental para la aplicación del Análisis de Correlación Canónica clásica (CCA). Aunque análisis previos habían evidenciado que las series individuales mostraban comportamientos no normales (a partir de la prueba univariada D'Agostino-Pearson), era necesario verificar si las variables, consideradas en conjunto, cumplían con la suposición de normalidad multivariante.

Con este propósito, se aplicó el *test* de Mardia (1970), ampliamente utilizado en estadística multivariante[96]. Los resultados mostraron valores elevados de asimetría multivariante en la mayor parte del dominio espacial, lo cual fue visualizado en el mapa de la Figura 4.3. En dicho mapa, se observó un patrón espacialmente coherente de asimetría, con valores superiores a 5 en extensas zonas del Pacífico colombiano, particularmente frente a Buenaventura y Tumaco, indicando sesgos significativos en la distribución conjunta de las variables. Estos resultados confirmaron que el supuesto de normalidad multivariante no se cumplía, lo cual podía afectar la estabilidad y validez del análisis CCA clásico.

Ante la evidencia de no normalidad multivariante, se optó por implementar un Análisis de Correlación Canónica Regularizada (rCCA). Esta técnica extiende la CCA clásica, introduciendo términos de penalización que permiten estabilizar la solución en contextos de alta colinealidad, no normalidad y conjuntos de datos con un número elevado de predictores en relación con el tamaño de muestra[97]

Matemáticamente, el rCCA busca vectores de pesos \mathbf{a} y \mathbf{b} que definan las variables canónicas S y P como:

$$S = \mathbf{a}^\top \mathbf{X}, \quad P = \mathbf{b}^\top \mathbf{Y} \quad (4.1)$$

$$\mathbf{a}^\top (\Sigma_{XX} + \lambda_X \mathbf{I}) \mathbf{a} = 1, \quad \mathbf{b}^\top (\Sigma_{YY} + \lambda_Y \mathbf{I}) \mathbf{b} = 1 \quad (4.2)$$

donde Σ_{XX} y Σ_{YY} son las matrices de covarianza de X y Y , respectivamente; λ_X y λ_Y son parámetros de regularización (penalización ridge), e \mathbf{I} es la matriz identidad.

\mathbf{X} representó el conjunto de variables predictoras (TSM, presión, componentes zonal y meridional del viento, precipitación, e índices ENOS) y \mathbf{Y} correspondió exclusivamente a la serie temporal del nivel del mar. A diferencia de la CCA clásica, la rCCA incorpora una penalización tipo $L2$ (o ridge), que consiste en sumar un múltiplo de la matriz identidad a las matrices de covarianza Σ_{XX} y Σ_{YY} . Este procedimiento estabiliza el cálculo de las soluciones canónicas, evitando problemas de colinealidad y sobreajuste. La magnitud de esta regularización está gobernada por el parámetro λ , cuyo aumento produce una mayor reducción de la varianza de los estimadores, aunque a costa de

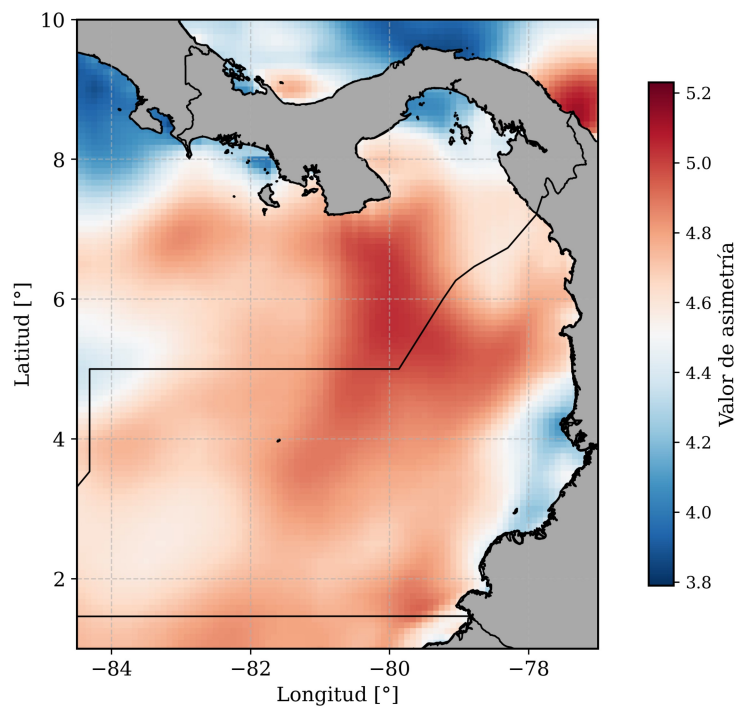


Figura 4.3: Mapa espacial de la asimetría multivariante calculada mediante el *test* de Mardia (1970) para el conjunto predictor. Los valores altos (en rojo) indican sesgos positivos y mayor desviación de la normalidad multivariante.

introducir cierto sesgo. Valores pequeños de λ permiten mayor flexibilidad, mientras que valores grandes inducen soluciones más suavizadas.

La implementación se llevó a cabo empleando el paquete `mixOmics` (versión 6.20) de *R*, utilizando la función `rcc`. Para la selección del parámetro de penalización óptimo, se exploraron tres valores: $\lambda = 0.1, 0.01$ y 0.001 . Se adoptó como criterio de selección la maximización de la correlación canónica entre las combinaciones lineales de \mathbf{X} y \mathbf{Y} . Como se muestra en la Figura 4.4, el valor $\lambda = 0.001$ permitió alcanzar las correlaciones más altas, superiores a 0.85 en amplias zonas de la cuenca del Pacífico colombiano. En términos espaciales, se observó que en casi toda la franja costera las correlaciones superaron el umbral de 0.80, reflejando una fuerte asociación multivariada entre las variables predictoras y la respuesta. No obstante, se identificaron áreas específicas donde la correlación disminuyó, particularmente en el extremo norte cercano a la frontera con Panamá y en el sur hacia la costa ecuatoriana. Este comportamiento podría vincularse a dinámicas locales más complejas, como la influencia del Chorro de Panamá en el sector septentrional, que podría modular las interacciones océano-atmósfera y alterar la coherencia espacial de los patrones canónicos en esas regiones.

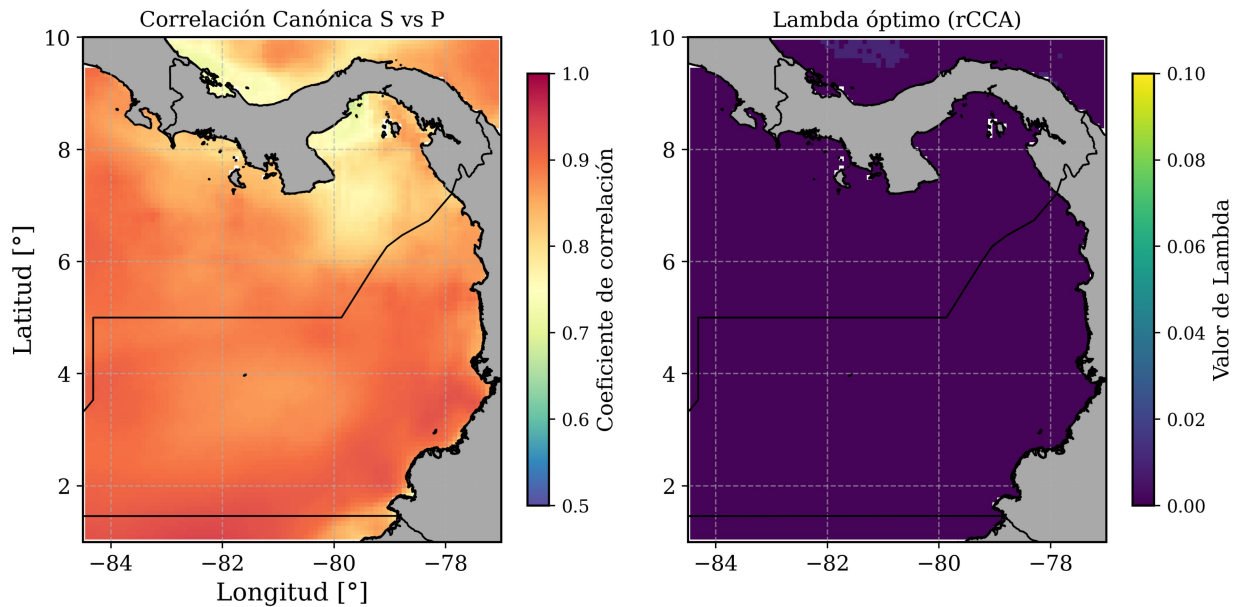


Figura 4.4: Izquierda: mapa espacial del coeficiente de correlación canónica entre el primer componente canónico predictor S y la respuesta P . Derecha: valores óptimos del parámetro λ seleccionados para cada punto del dominio geográfico mediante análisis *rCCA*.

Tras la obtención del primer componente regularizado S , se procedió a analizar la contribución relativa de cada predictor mediante los pesos canónicos. La Figura 4.5 presenta los mapas espaciales de dichos pesos, revelando patrones espaciales heterogéneos y diferenciados entre variables. Los pesos

canónicos permiten interpretar cómo cambios en cada predictor influyen en el valor del componente S . A diferencia de la CCA clásica, el rCCA tiende a “encoger” los pesos hacia cero, reduciendo la influencia de variables menos informativas y destacando aquellas con mayor peso canónico.

En el caso de la TSM, se observaron pesos marcadamente positivos en la mayor parte del dominio, alcanzando valores entre aproximadamente $+0.6$ y $+1.4$, especialmente al sur y en el oeste de la cuenca. Esto indicó que anomalías cálidas de TSM contribuyeron a incrementar el valor del primer componente canónico S , reflejando una relación directa entre calentamientos superficiales y la variabilidad del nivel del mar.

Para la presión media al nivel del mar, predominó un patrón negativo en la mayor parte del dominio, con pesos que alcanzaron valores de hasta -1.0 . Esto sugirió que incrementos en la presión tendieron a asociarse con disminuciones en el componente S , lo que resultó coherente con el efecto barométrico inverso, según el cual presiones más bajas elevan el nivel del mar.

La precipitación presentó en general pesos moderados positivos en el centro del dominio, con valores entre $+0.2$ y $+0.5$, mientras que en sectores costeros del norte y sur aparecieron pequeños parches de pesos negativos. Esto indicó un aporte positivo, aunque no dominante, de las anomalías de precipitación sobre el componente predictor S .

La componente zonal del viento (U) mostró pesos negativos significativos (hasta -1.5) en amplias zonas al centro y sureste del dominio, mientras que en el resto predominaron pesos cercanos a cero o levemente positivos (noroeste). Esto sugirió que vientos del oeste (valores positivos de U) tendieron a reducir el componente predictor S , posiblemente asociados con procesos de divergencia costera, es decir, desplazamientos de las aguas superficiales hacia el norte o mar afuera, lo que puede inducir descensos locales en el nivel del mar en sectores costeros centro y sur del Pacífico colombiano.

Por su parte, la componente meridional (V) presentó pesos predominantemente positivos (entre $+0.2$ y $+0.8$) en casi todo el dominio, exceptuando valores negativos moderados, especialmente en la zona oeste. Esto reflejó que vientos dirigidos hacia el norte tendieron a aumentar el componente predictor S , posiblemente asociados a cambios en la circulación atmosférica regional.

En cuanto a los índices ENOS, se identificaron patrones bien definidos, aunque con hallazgos particulares en el presente estudio:

- El **índice Niño 1+2** presentó pesos negativos moderados (alrededor de -0.3 a -0.6) en la mayor parte del dominio, salvo sectores neutros en la franja norte. Esto indicó que anomalías cálidas costeras, típicas del evento Niño Costero, tendieron en promedio a asociarse con descensos en S . Este comportamiento puede estar asociado a procesos locales y es coherente con la menor influencia directa que este índice ejerce sobre la costa colombiana, dado que sus anomalías térmicas se concentran principalmente al sur de Ecuador y Perú.
- El **índice Niño 3** mostró pesos positivos elevados (superiores a $+0.8$) en casi todo el dominio, con máximos frente a la costa centro y sur. Esto reflejó su fuerte asociación con el aumento de S , consistente con el impacto del Niño Canónico en Colombia, vinculado a la propagación de ondas Kelvin que elevan el nivel del mar en la región.
- El **índice Niño 3.4** se caracterizó por pesos negativos notables (entre -0.6 y -1.8), especialmente en zonas centro costeras y norte en el golfo de Panamá, lo cual indicó que anomalías

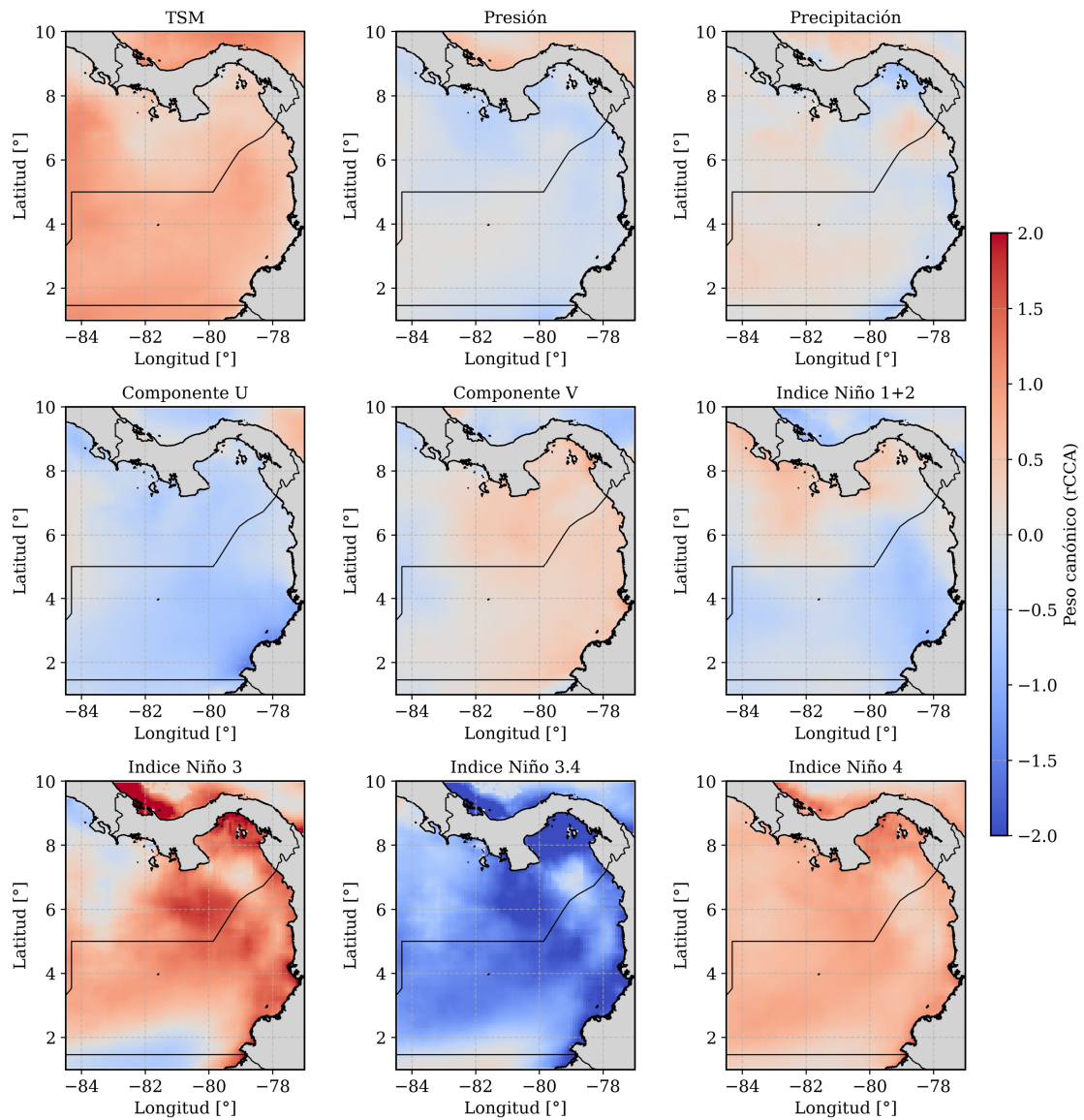


Figura 4.5: Mapas espaciales de los pesos canónicos obtenidos mediante $rCCA$ para cada variable predictora considerada.

cálidas en el Pacífico central-oriental tendieron a asociarse con reducciones en S . Cabe resaltar que, en términos físicos, Niño 3 y Niño 3.4 suelen presentar efectos similares y positivos sobre el nivel del mar durante eventos canónicos. Sin embargo, en este análisis rCCA emergió un patrón inverso entre ambos, lo que constituye un hallazgo particular y sugiere posibles diferencias regionales en la fase o en la respuesta del océano ante calentamientos en el Pacífico central. Este resultado refleja la capacidad del método regularizado para revelar matices específicos en la estructura de las relaciones multivariadas.

- Finalmente, el **índice Niño 4** presentó un patrón positivo dominante (entre +0.4 y +0.9) sobre la mayoría del dominio, indicando que anomalías cálidas en el Pacífico occidental-central contribuyeron a incrementar S . Este comportamiento es coherente con la influencia más difusa y persistente que ejerce el Niño Modoki, aunque con menor impacto global que el Niño Canónico.

En términos generales, los pesos canónicos revelaron la importancia relativa y el sentido de la contribución de cada variable al patrón canónico conjunto. La relación entre las variables predictoras y el nivel del mar se explicó en función de la combinación lineal total que optimizó la correlación entre ambos conjuntos de variables, y no por el efecto individual de una sola variable. Este análisis espacial de los pesos canónicos proporcionó información determinante para comprender las regiones y variables que ejercieron una mayor influencia en la variabilidad del nivel del mar y para fundamentar la selección de predictores en los modelos de aprendizaje profundo desarrollados posteriormente.

4.2.1.4. Interpretación local de los pesos canónicos

Para profundizar el análisis, se examinó la magnitud y el signo de los pesos canónicos del primer componente rCCA en las tres localidades de interés: Ciudad Mutis, Buenaventura y Tumaco. La Figura 4.6 muestra estos valores mediante diagramas de barras, lo cual permitió comparar la contribución relativa de cada variable en cada punto costero.

En **Ciudad Mutis**, los pesos positivos más altos correspondieron al índice Niño 3 (+1.22), el índice Niño 4 (+0.67) y, en menor medida, a la TSM (+0.46). La contribución positiva de los índices y TSM determinó que, dentro de la combinación lineal que definió el primer componente canónico, anomalías cálidas en el Pacífico central-oriental y colombiano, tendieron a reforzar el valor de dicho componente. Por el contrario, el índice Niño 3.4 (-1.35) y presión (-0.27) mostraron pesos negativos, lo que indicó que sus anomalías positivas ejercieron un efecto reductor sobre el primer componente canónico en Ciudad Mutis. Las componentes del viento tuvieron pesos de moderados a bajos y de signo opuesto: U (-0.16) y V (+0.45), siendo más relevante la componente meridional, posiblemente asociada a la dinámica del chorro de Panamá.

En **Buenaventura**, los índices Niño 3.4 y 3 fueron las variables dominantes (-2.93 y 2.16, respectivamente), destacando la fuerte conexión entre estos índices y la variabilidad regional del nivel del mar. La TSM local presentó un peso prácticamente nulo (0.11). Las variables viento zonal (U) (-0.82), precipitación (-0.23), presión (-0.16) y Niño 1+2 (-0.20) mostraron pesos negativos. En contraste, los índices Niño 4 (+1.42) y el viento meridional (V) (+0.97) mantuvieron pesos positivos,

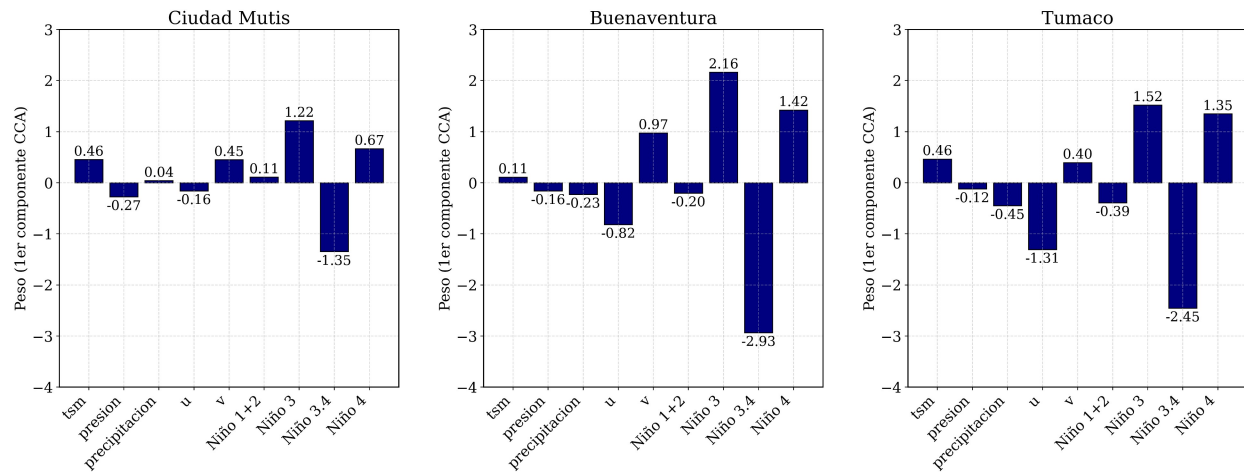


Figura 4.6: Contribución relativa de las variables predictoras al primer componente canónico del rCCA en las estaciones de Ciudad Mutis, Buenaventura y Tumaco.

reflejando que sus anomalías positivas tendieron a reforzar el componente canónico en esta región. Este patrón enfatizó la relevancia de diferenciar las distintas modalidades de ENOS, dado que no todos los eventos generan impactos iguales sobre el nivel del mar.

En **Tumaco**, el índice Niño 3.4 (-2.45) se mantuvo como el predictor de mayor peso negativo, lo que subrayó su influencia en la dinámica del nivel del mar en esta localidad. A diferencia de las otras poblaciones, Tumaco presentó un peso negativo mayor en el índice Niño 1+2 (-0.39). Esto sugiere que anomalías térmicas positivas en el sector costero oriental del Pacífico ecuatorial contribuyeron a reducir el componente canónico predictor en Tumaco. La TSM local mostró un peso moderadamente positivo (+0.46), mientras que las componentes del viento exhibieron pesos de signo opuesto ($U = -1.31$, $V = +0.40$), resaltando la posible influencia de la componente zonal, vinculada a la recurvatura de los vientos alisios del sureste en esta región. Finalmente, los índices Niño 3 (+1.52) y Niño 4 (+1.35) tuvieron pesos positivos, indicando que sus anomalías cálidas tendieron a aumentar el primer componente canónico en Tumaco.

En términos generales, estos resultados confirmaron que los pesos canónicos reflejaron el grado y el sentido en que cada variable contribuyó a la combinación lineal que maximizó la correlación con el nivel del mar. No obstante, su interpretación física debe entenderse en el marco multivariado del análisis, ya que el efecto de cada variable depende de su interacción con el resto de predictores presentes en el componente canónico.

En síntesis, los análisis exploratorios realizados permitieron no solo caracterizar las asociaciones bivariadas entre las variables oceánicas y atmosféricas, sino también revelar patrones multivariados robustos que vinculan la dinámica del nivel del mar con los distintos índices del fenómeno ENOS en el Pacífico colombiano. Estos resultados fueron fundamentales para la posterior selección de variables predictoras y para orientar el diseño y la parametrización de los modelos implementados, los cuales

se describen en las siguientes secciones. Esto garantiza que los modelos no solo estén sustentados estadísticamente, sino que se mantengan coherentes con la dinámica física subyacente al sistema océano-atmósfera.

4.2.2. Selección de algoritmos

Diversos estudios en el campo de la oceanografía y la climatología han implementado modelos como *Random Forest*, *XGBoost*, Support Vector Regression (SVR) o ARIMA para predecir el nivel del mar y analizar la influencia del fenómeno ENOS [98], [99]. Sin embargo, estos enfoques, aunque eficaces en escenarios específicos, se centran principalmente en capturar relaciones estáticas o lineales y presentan limitaciones para modelar las complejas dependencias temporales y no lineales características del sistema océano-atmósfera.

Considerando los hallazgos del análisis exploratorio desarrollado en esta investigación, donde se evidenciaron patrones de asociación multivariada, rezagos temporales y la influencia de señales climáticas de gran escala como el ENOS, se optó por utilizar redes neuronales del tipo *LSTM*. Estas redes resultaron especialmente adecuadas para abordar el objetivo de identificar y modelar relaciones dinámicas entre las anomalías de variables océano-atmosféricas y el nivel del mar en el Pacífico colombiano.

Las *LSTM* constituyen una extensión de las redes neuronales recurrentes (RNN), diseñadas para mitigar los problemas de desvanecimiento y explosión del gradiente que afectan al aprendizaje en secuencias largas. Incorporan estructuras internas denominadas *celdas de memoria*, que contienen puertas de entrada, olvido y salida. Dichas puertas controlan el flujo de información, permitiendo a la red mantener o descartar datos relevantes a lo largo del tiempo, lo que es esencial para capturar fenómenos con memoria larga, como los procesos modulados por el ENOS [100].

La dinámica interna de una celda *LSTM* puede expresarse mediante las siguientes ecuaciones:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (\text{puerta de olvido}) \quad (4.3)$$

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (\text{puerta de entrada}) \quad (4.4)$$

$$\tilde{C}_t = \tanh(W_C[h_{t-1}, x_t] + b_C) \quad (\text{estado candidato}) \quad (4.5)$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \quad (\text{estado de la celda}) \quad (4.6)$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (\text{puerta de salida}) \quad (4.7)$$

$$h_t = o_t \odot \tanh(C_t) \quad (\text{estado oculto}) \quad (4.8)$$

Donde x_t es el vector de entrada en el instante t ; h_t es el estado oculto; C_t es el estado interno de la celda de memoria; y f_t , i_t , o_t corresponden, respectivamente, a las puertas de olvido, entrada y salida. Las funciones σ (sigmoide) y \tanh (tangente hiperbólica) se utilizan como funciones de activación. Los parámetros W y b representan los pesos y sesgos aprendidos durante el entrenamiento [100].

En esta investigación, el modelo *LSTM* fue estructurado para procesar entradas multivariadas conformadas por series de tiempo mensuales de anomalías de TSM, presión media a nivel del mar,

componentes zonal y meridional del viento y precipitación acumulada. Estas variables fueron seleccionadas tanto por su relevancia física como por los resultados obtenidos en los análisis de correlación y *rCCA* previamente descritos. Además, fueron estandarizadas con el fin de estabilizar el proceso de entrenamiento y optimizar la convergencia del modelo.

A diferencia de los algoritmos de árboles o modelos lineales, las *LSTM* son capaces de representar explícitamente la evolución temporal del sistema, integrando de manera natural los efectos rezagados y las interacciones acumulativas que caracterizan los procesos océano-atmosféricos. Por ello, se seleccionó la arquitectura *LSTM* como modelo final para esta investigación, dada su idoneidad para capturar la dinámica compleja del nivel del mar bajo la influencia de forzantes climáticos de gran escala, contribuyendo así al cumplimiento del objetivo de modelar y pronosticar los impactos asociados al ENOS en el Pacífico colombiano.

4.2.3. Parametrización de algoritmos

La parametrización del modelo *LSTM* se diseñó cuidadosamente para optimizar su desempeño ante la complejidad inherente de las series multivariadas, no lineales y con dependencias temporales de largo alcance, características del sistema océano-atmósfera. Este proceso fue clave para garantizar que el modelo lograra un equilibrio entre capacidad predictiva y generalización, evitando el riesgo de sobreajuste.

4.2.3.1. Definición de ventana y horizonte

La selección del horizonte de predicción se fundamentó en el análisis de correlación cruzada entre los índices ENOS y el nivel del mar, el cual evidenció máximas correlaciones con desfases positivos de hasta seis meses. Por ello, se definió un horizonte de pronóstico (*horizon*) de seis pasos mensuales. La longitud de la ventana de entrada (*window_size*) se fijó en 48 meses, considerando que el fenómeno ENOS presenta ciclos interanuales que suelen variar entre 2 y 7 años, como se mencionó en capítulos anteriores. Escoger cuatro años como memoria histórica definió un compromiso razonable para capturar la persistencia del fenómeno sin introducir un exceso de complejidad en el modelo.

4.2.3.2. Tamaño y partición de datos

Las series de tiempo mensuales utilizadas abarcaron el período comprendido entre junio de 1993 y junio de 2020, cubriendo 27 años, es decir, un total de 325 observaciones en cada variable. Para cada modelo, los datos se dividieron en conjuntos de entrenamiento y prueba en una proporción del 80 % y 20 %, respectivamente. Esta partición se realizó mediante la función `train_test_split`, estableciendo la opción `shuffle=False` para asegurar que no se alterara el orden cronológico de las series. Aunque las muestras fueron segmentadas en bloques consecutivos, se preservó la estructura de las ventanas temporales definidas previamente (inputs y outputs), garantizando que la secuencia de datos dentro de cada muestra respetara la lógica temporal de cada variable. De este modo, se mantuvo la coherencia propia de un análisis de series de tiempo, en el cual las predicciones se basan en información disponible en el pasado.

4.2.3.3. Estrategia de validación

Para la evaluación del desempeño y la selección de hiperparámetros, se implementa una validación cruzada bloqueada de tipo *K-Fold*, adecuada para series de tiempo. En este enfoque, el conjunto de entrenamiento se divide en cinco bloques cronológicos consecutivos ($K=5$). En cada iteración, se utiliza uno de estos bloques como conjunto de validación, mientras que los bloques previos se emplean para el entrenamiento. Esta configuración permite obtener estimaciones robustas de las métricas de desempeño sin reducir en exceso el tamaño de los conjuntos de validación, lo cual resulta relevante dado que las series han sido segmentadas previamente en ventanas temporales (inputs y outputs).

4.2.3.4. Arquitectura y búsqueda de hiperparámetros

La arquitectura básica consistió en dos capas *LSTM* secuenciales, seguidas de una capa densa para producir el horizonte de salida. Se exploraron distintas combinaciones de neuronas en las dos capas *LSTM* (n_1, n_2) utilizando una búsqueda sistemática sobre el espacio de hiperparámetros. Para cada combinación candidata se entrenó el modelo y se evaluó su rendimiento mediante la métrica RMSE promedio sobre los *folders* de validación.

4.2.3.5. Regularización y optimización

Se integraron técnicas específicas para evitar sobreajuste:

- **Dropout:** Se fijó en 0.2 tanto para las conexiones estándar como para las conexiones recurrentes. Esta técnica introduce aleatoriamente el “apagado” de un subconjunto de neuronas en cada iteración de entrenamiento, lo que ayuda a prevenir la dependencia excesiva de ciertos patrones en los datos.
- **Penalización L2:** Se incorporó una penalización L2 con parámetro $\lambda = 0.001$ en las capas *LSTM*. Este término penaliza grandes valores en los pesos de la red, favoreciendo soluciones con menor complejidad y evitando el sobreajuste en datasets ruidosos o con alta dimensionalidad, como en este caso.
- **Early Stopping:** Se utilizó la técnica de *early stopping* con una paciencia de 20 épocas, deteniendo el entrenamiento si no se observaba mejora en el conjunto de validación. Esta estrategia permitió evitar iteraciones innecesarias y proteger al modelo de sobreajustes, asegurando eficiencia computacional.

4.2.3.6. Configuración general

La función de activación elegida para las capas *LSTM* fue **tanh**, en razón de su capacidad para manejar información tanto positiva como negativa, lo que resulta esencial en series de tiempo oceánicas que incluyen anomalías. La capa densa final utilizó activación lineal, adecuada para problemas

de regresión. El optimizador empleado fue *Adam*, por su robustez y eficiencia en la convergencia, especialmente útil en escenarios con gradientes dispersos o series largas.

Todas las variables predictoras previamente estandarizadas fueron reescaladas a su dominio original, junto con las predicciones generadas para facilitar su interpretación en términos absolutos (metros).

4.2.3.7. Dimensión espacial

El modelo *LSTM* se aplicó sobre un dominio espacial de 109×91 puntos (latitud-longitud), correspondiente al área de la Bahía de Panamá, incluyendo el Pacífico colombiano, definido en los conjuntos de datos globales. El entrenamiento se realizó de manera puntual, es decir, para cada nodo de la malla, empleando los mismos principios de parametrización.

Con base en los análisis exploratorios previos, incluyendo la correlación de Spearman, la correlación cruzada y el Análisis de Correlación Canónica Regularizada, se definieron tres escenarios para implementar las redes *LSTM*. Todos compartieron la misma configuración de hiperparámetros y los mismos criterios de validación y selección de arquitectura, con excepción de las variables de entrada y el número de neuronas en las capas ocultas, el cual varió entre modelos y poblaciones. Esta estrategia permitió evaluar la sensibilidad del rendimiento del modelo ante cambios en la arquitectura y optimizar el equilibrio entre capacidad de ajuste y generalización.

La selección final de la arquitectura por ubicación se fundamentó en la minimización del error cuadrático medio (RMSE) promedio obtenido en la validación cruzada, asegurando un desempeño robusto y consistente en las predicciones.

La Tabla 4.1 resume los hiperparámetros empleados de manera común a los tres modelos, así como la arquitectura específica (número de neuronas en las dos capas *LSTM*) que resultó seleccionada para cada ubicación costera en cada uno de los modelos evaluados.

Esta parametrización aseguró que el modelo *LSTM* se ajustara a las particularidades dinámicas y climáticas de cada localidad, capturando de manera eficaz los efectos acumulativos y retardados de los forzamientos atmosféricos y oceánicos sobre el nivel del mar en el Pacífico colombiano.

4.2.4. Implementación de los modelos *LSTM*

La implementación de los modelos *LSTM* se desarrolló en el lenguaje Python, utilizando las bibliotecas *TensorFlow*, *scikit-learn* y *Keras*, las cuales proporcionaron herramientas robustas para la definición arquitectónica, regularización y entrenamiento de redes neuronales profundas. Esta infraestructura permitió configurar distintas variantes de modelos, en consonancia con los objetivos del presente trabajo.

En este estudio se exploraron tres enfoques de modelado, derivados de los análisis previos de correlación:

- **Modelo 1.** Implementación de una red *LSTM* en la cual la única variable predictora correspondió al primer componente canónico resultante del análisis de correlación canónica regularizado (*rCCA*), calculado utilizando simultáneamente las nueve variables predictoras (las cinco

Tabla 4.1: Resumen de la parametrización del modelo *LSTM* y configuraciones de arquitectura por estación y modelo.

Parámetro	Valor o configuración		
Función de activación	tangente hiperbólica (\tanh)		
Optimización	Adam		
Función de pérdida	Error cuadrático medio (MSE)		
Regularización	Dropout = 0.2, L2 ($\lambda = 0.001$)		
Criterio de parada	EarlyStopping con paciencia de 20 épocas		
Horizonte de predicción	6 pasos mensuales		
Ventana de entrada	48 pasos mensuales		
Validación cruzada	K-Fold ($k = 5$)		
Tamaño dominio espacial	109×91		
Arquitecturas seleccionadas por población y modelo			
Estación	Modelo 1	Modelo 2	Modelo 3
Ciudad Mutis	(30, 30)	(30, 20)	(30, 30)
Buenaventura	(30, 30)	(30, 30)	(30, 30)
Tumaco	(30, 30)	(30, 30)	(30, 30)

variables ambientales locales más los cuatro índices ENOS). La variable objetivo fue el nivel del mar.

- Modelo 2.** Implementación de una red *LSTM* basada en el primer componente canónico del *rCCA*, pero esta vez calculado únicamente a partir de las variables seleccionadas como dominantes según sus pesos canónicos absolutos mayores a un umbral de 0.3 en cada punto geográfico. Este enfoque redujo la dimensionalidad de entrada, priorizando únicamente aquellas variables con mayor contribución al patrón canónico identificado. La variable objetivo al igual que el caso anterior fue el nivel del mar.
- Modelo 3.** Implementación de una red *LSTM* multivariada que empleó como variables predictoras las series de tiempo de anomalías estandarizadas de las cinco variables ambientales locales (*TSM*, presión, componentes zonal y meridional del viento, y precipitación acumulada). El objetivo continuó siendo el nivel del mar.

Para los tres modelos se mantuvieron consistentes los criterios de preprocesamiento y los hiperparámetros de entrenamiento, detallados en el apartado anterior.

Para verificar preliminarmente la potencialidad de los modelos, se realizó un primer entrenamiento sin aplicar validación cruzada, con el propósito de verificar la capacidad de ajuste de cada arquitectura bajo condiciones ideales. Los resultados obtenidos se presentan en las Figuras 4.7, 4.8 y 4.9 para cada uno de los modelos considerados.

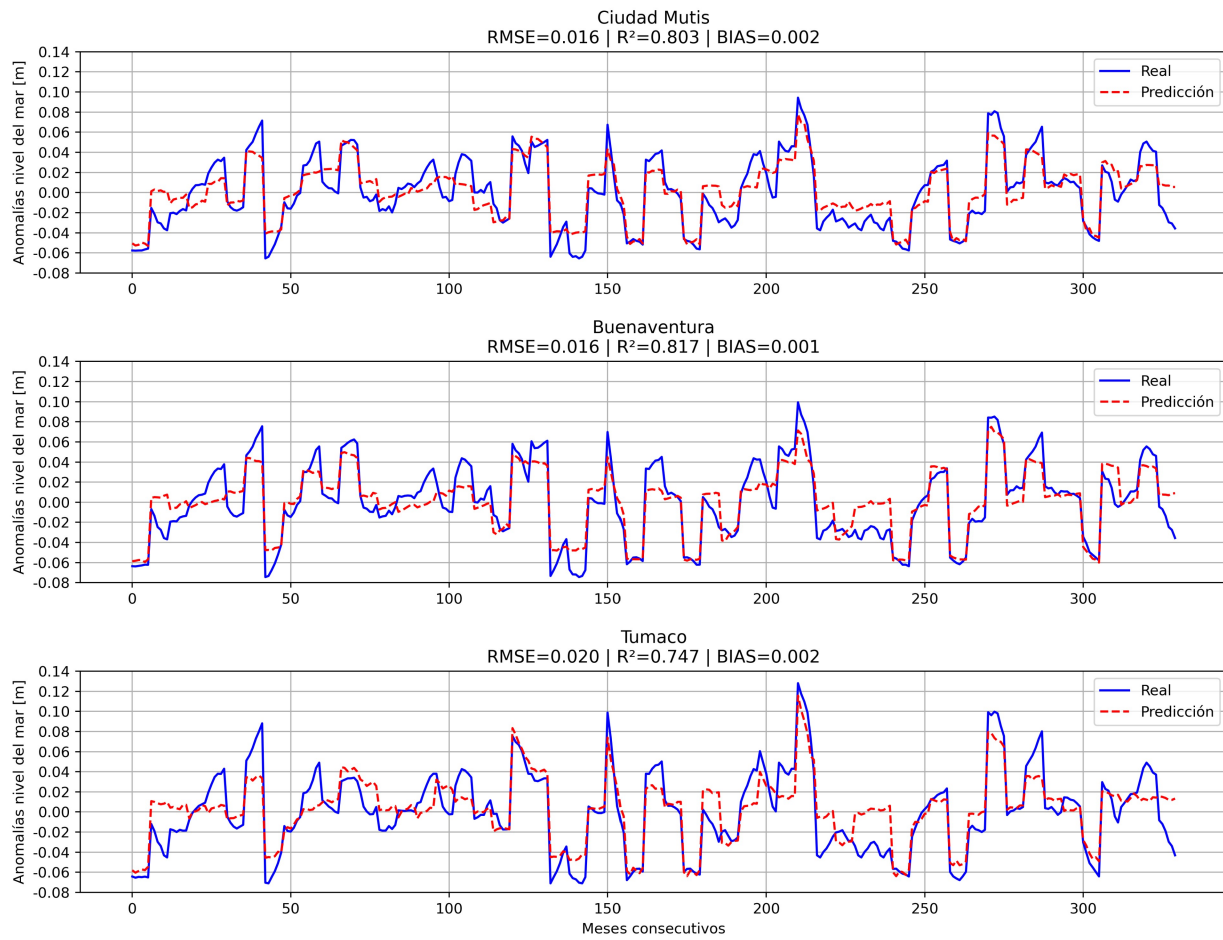


Figura 4.7: Resultados preliminares del **Modelo 1**. Predicción del nivel del mar utilizando como predictor el primer componente canónico del rCCA calculado con las nueve variables (cinco ambientales y cuatro índices ENOS).

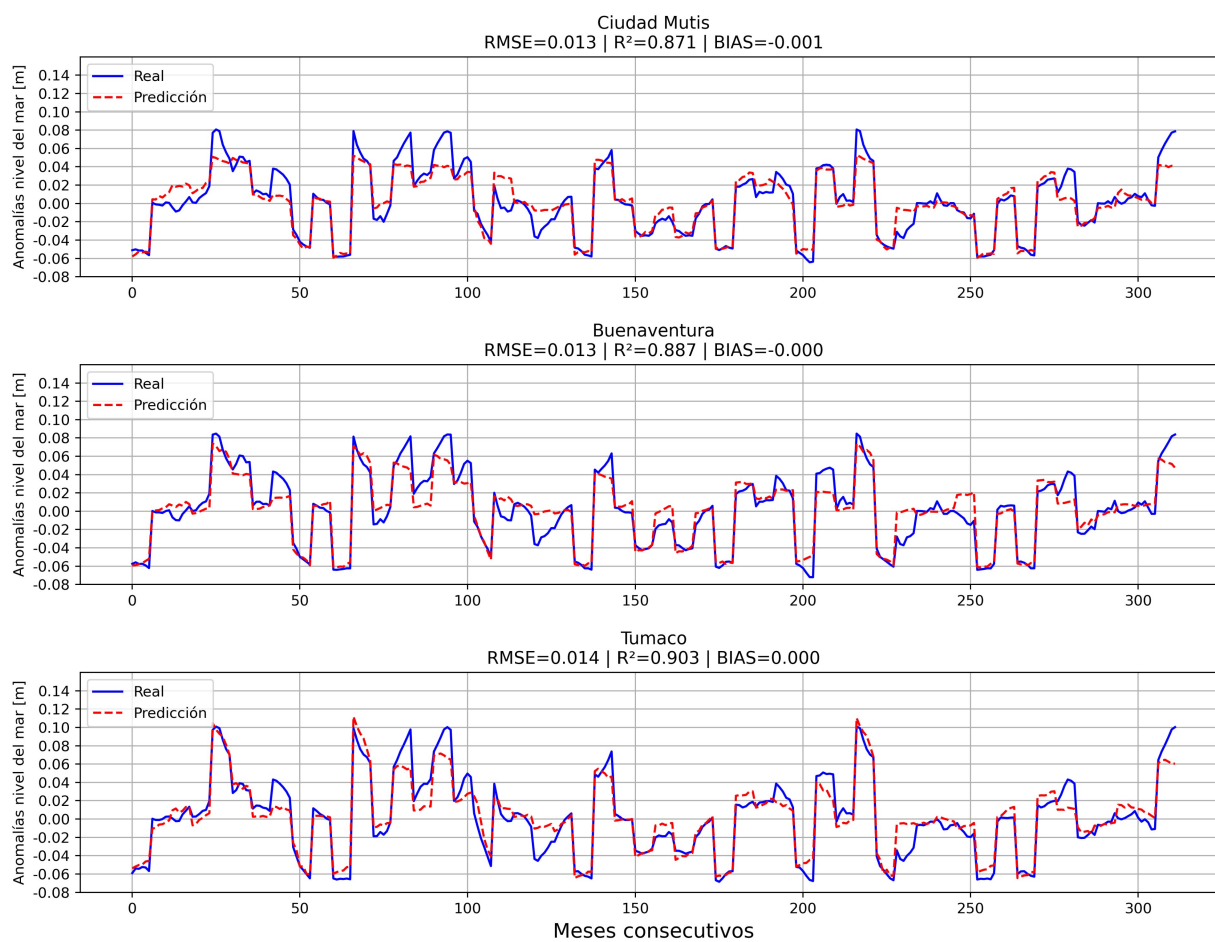


Figura 4.8: Resultados preliminares del **Modelo 2**. Predicción del nivel del mar utilizando como predictor el primer componente canónico del rCCA calculado únicamente con las variables seleccionadas como dominantes en cada localidad.

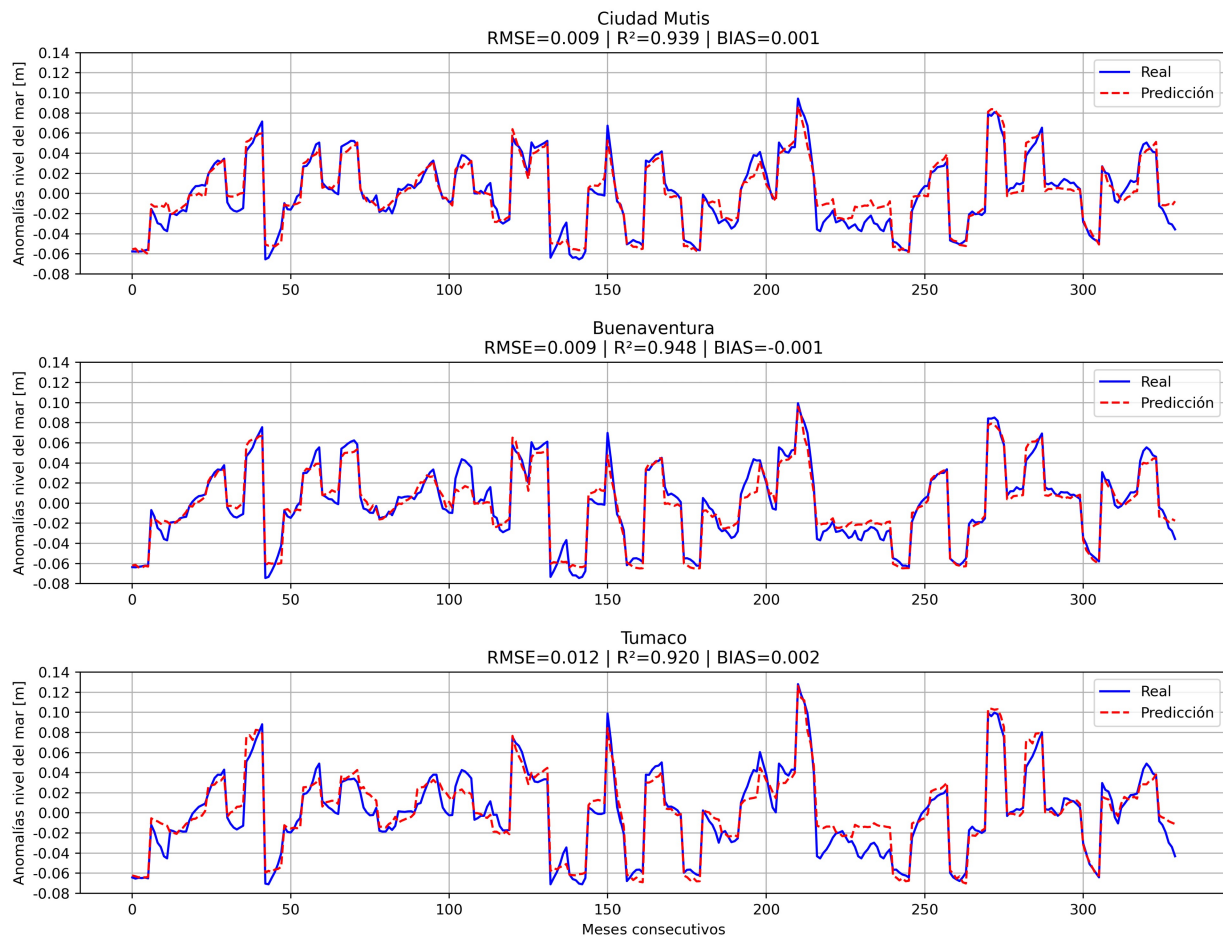


Figura 4.9: Resultados preliminares del **Modelo 3**. Predicción del nivel del mar utilizando las series de tiempo de las anomalías estandarizadas de las cinco variables ambientales como predictores.

Los resultados evidenciaron diferencias sustanciales en el ajuste entre los tres modelos. El *Modelo 1*, que utilizó el componente canónico global de las nueve variables, presentó un desempeño moderadamente inferior, con R^2 oscilando alrededor de 0.75 a 0.82, destacando la sensibilidad del ajuste a la representación sintética que realiza el rCCA sobre todas las variables (Figura 4.7).

Por su parte, el *Modelo 2*, al integrar únicamente las variables dominantes definidas localmente, logró una mejoría en el ajuste, obteniendo R^2 aproximado de 0.88 en Ciudad Mutis, 0.90 en Tumaco, y alrededor de 0.89 en Buenaventura. Esta variante demostró que una reducción cuidadosa de la dimensionalidad, basada en los pesos canónicos, puede mantener la calidad predictiva al tiempo que simplifica la complejidad del modelo (Figura 4.8).

Finalmente, el *Modelo 3*, basado en todas las variables ambientales, mostró un ajuste notable en las tres poblaciones, con valores de R^2 superiores a 0.92 en Buenaventura y Ciudad Mutis y ligeras diferencias en Tumaco. (Figura 4.9).

Estos resultados en primera instancia, confirman la idoneidad de las redes *LSTM* para capturar la variabilidad temporal del nivel del mar en el Pacífico colombiano, incluso bajo distintos esquemas de modelado. No obstante, para una evaluación más robusta y generalizable, se presenta en el siguiente capítulo (Evaluación) el análisis mediante validación cruzada y el estudio comparativo de métricas estadísticas, con el propósito de identificar el modelo óptimo por estación y por configuración arquitectónica.

4.3. Modelos Digitales de Elevación

4.3.1. Selección de algoritmos

4.3.1.1. Análisis probabilísticos

Una vez realizadas las pruebas de dependencia espacial, se identificó, en todas las áreas analizadas, una estructura positiva de autocorrelación espacial. Este hallazgo resulta fundamental, ya que indica que los valores de elevación en ubicaciones cercanas tienden a ser más similares entre sí que aquellos situados a mayor distancia. Esta dependencia espacial constituye el fundamento teórico para la aplicación de técnicas avanzadas de interpolación geoestadística, como el *kriging*.

El *kriging* es un conjunto de técnicas de interpolación consideradas óptimas desde el punto de vista geoestadístico, en tanto que generan estimaciones insesgadas con varianza mínima bajo determinados supuestos [101], [102]. A diferencia de los métodos determinísticos, como la distancia inversa ponderada o la interpolación lineal por triangulación, el *kriging* no solo permite predecir valores en ubicaciones no muestreadas [103], sino que también proporciona una estimación explícita de la incertidumbre asociada a dichas predicciones [101], [102]. Esta característica se basa en el uso del semivariograma, una función que cuantifica la autocorrelación espacial de la variable de interés, en este caso, la elevación, como función de la distancia y, en ciertos casos, de la dirección entre pares de observaciones.

El marco teórico del *kriging* requiere asumir algún tipo de estacionariedad espacial. En el *kriging* ordinario, se presupone estacionariedad de segundo orden, lo que implica una media constante y una covarianza que depende únicamente de la distancia entre puntos [101], [102]. Aunque no es

estrictamente necesario que los datos sigan una distribución normal, dicha condición suele considerarse deseable para realizar inferencias estadísticas adicionales, como la construcción de intervalos de confianza. La efectividad del *kriging* depende, en gran medida, de la existencia de una estructura espacial continua en los datos, la cual debe ser representada adecuadamente mediante la selección y el ajuste de un modelo de semivariograma apropiado [101], [102].

Existen diversos modelos matemáticos para representar el semivariograma, cada uno con propiedades específicas que pueden ajustarse mejor a distintos patrones de comportamiento espacial [103], [104]. La elección del modelo adecuado es indispensable, ya que influye directamente en la calidad de las predicciones y en la estimación de la incertidumbre. A continuación, se presentan los modelos seleccionados para este trabajo interpretados a partir de las funciones que ofrece PyKriging [105]:

Modelo Exponencial es adecuado cuando la correlación decae rápidamente pero nunca desaparece del todo, lo cual es útil en fenómenos con continuidad pero sin diferenciabilidad [103], [104], [105].

Dado un *lag* h (distancia entre dos puntos), la semivarianza $\gamma(h)$ se expresa como:

$$\gamma(h) = n + c \left(1 - e^{-\frac{h}{a/3}}\right) \quad (4.9)$$

donde:

- h : valores de la distancia a los cuáles se calcula el variograma
- n : efecto pepita (*nugget*)
- c : contribución estructural (*sill - nugget*)
- a : parámetro de alcance (*range*)

Modelo Esférico tiene un alcance definido, a partir del cual los puntos dejan de estar correlacionados [103], [104], [105].

$$\gamma(h) = \begin{cases} n + c \left[\frac{3h}{2a} - \frac{h^3}{2a^3} \right], & \text{si } h \leq a \\ n + c, & \text{si } h > a \end{cases} \quad (4.10)$$

donde:

- h : valores de la distancia a los cuáles se calcula el variograma
- n : efecto pepita (*nugget*)
- c : contribución estructural (*sill - nugget*)
- a : parámetro de alcance (*range*).

Modelo de Potencia (*Power*) es una función de semivarianza que representa una relación de tipo potencial entre la distancia y la variabilidad espacial. A diferencia de otros modelos que alcanzan un umbral (*sill*), el modelo de Potencia no presenta un alcance definido, lo que lo hace adecuado para fenómenos donde la variabilidad continúa aumentando con la distancia sin estabilizarse [103], [104], [105].

$$\gamma(h) = n + s \cdot h^a \quad (4.11)$$

donde:

- h : valores de la distancia a los cuáles se calcula el variograma
- n : efecto pepita (*nugget*)
- s : coeficiente de escala (contribución estructural)
- a : exponente de potencia, con $0 < a < 2$

Este modelo es útil en contextos donde la variabilidad espacial no presenta un límite claro, como en ciertos procesos naturales de larga escala o en datos con estructuras fractales.

4.3.1.2. Análisis determinísticos

Los modelos determinísticos de interpolación se basan en funciones matemáticas explícitas para estimar los valores en ubicaciones no muestreadas [106], sin considerar la variabilidad espacial ni la estructura de dependencia estadística entre los datos [107], [106], [108]. Entre estos métodos existen: el modelo lineal, el modelo cúbico (o interpolación cúbica), el método de vecinos más cercanos, la red de triángulos irregulares y algunos otros [103], [104]. Estos métodos son rápidos y fáciles de implementar, aunque son menos precisos en contextos donde la estructura espacial de los datos es importante [103], [104]. Los modelos seleccionados para este trabajo fueron los siguientes:

Modelo lineal asume que la variable cambia linealmente entre los puntos muestreados. En 2D, la interpolación se realiza usualmente dentro de triángulos (por ejemplo, usando la triangulación de Delaunay) [103], [104].

$$y = y_1 + \frac{(x - x_1)(y_2 - y_1)}{x_2 - x_1} \quad (4.12)$$

donde:

- x_1 e y_1 son las primeras coordenadas
- x_2 e y_2 son las segundas coordenadas
- x es el punto en el que se hace la interpolación
- y es el valor interpolado

Modelo cúbico utiliza funciones polinomiales de tercer grado para estimar valores intermedios. Proporciona una superficie continua y suave, incluyendo derivadas continuas, por lo que es útil en modelos donde se requiere una transición suave (por ejemplo, en variables meteorológicas o topográficas) [103], [104].

$$P(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3 \quad (4.13)$$

donde:

- x_i es el punto base de interpolación
- a, b, c, d coeficientes calculados a partir de los puntos vecinos y sus derivadas
- $P(x)$ es el valor interpolado

Modelo vecinos más cercanos (*Nearest Neighbor*) asigna a cada punto no muestreado el valor del punto de datos más cercano. Es una interpolación por regiones constantes, muy simple y rápida, pero genera superficies con saltos abruptos. No considera la distancia ni la continuidad espacial [103], [104].

$$f(\mathbf{x}) = f(\mathbf{x}_i) \quad \text{donde} \quad \mathbf{x}_i = \arg \min_{\mathbf{x}_j \in D} \|\mathbf{x} - \mathbf{x}_j\| \quad (4.14)$$

donde:

- x punto donde se quiere interpolar
- D conjunto de puntos muestreados
- x_i punto más cercano a x

4.3.2. Parametrización de los modelos

En esta sección, se detalla el proceso de configuración y la selección de parámetros para los modelos de interpolación geoestadística y determinística implementados para la generación del Modelo Digital de Elevación (MDE). La correcta parametrización fue fundamental para asegurar la calidad y precisión del modelo resultante.

4.3.2.1. Modelos Geoestadísticos (*Kriging*)

Para la implementación de las técnicas de *kriging* se utilizó la librería `PyKrige` [105] de `Python` [109], la cual resultó determinante para el análisis geoestadístico de los datos espaciales. Esta herramienta permitió construir semivariogramas experimentales a partir de los datos de elevación y facilitó la exploración visual y cuantitativa de la estructura de dependencia espacial. El proceso incluyó una evaluación cuidadosa de la forma del semivariograma experimental, lo que permitió seleccionar y ajustar el modelo teórico más adecuado (como esférico, exponencial ó potencia) para

representar de manera precisa la variabilidad espacial. `PyKrige` [105] proporciona una variedad de funciones para calcular, ajustar y visualizar estos modelos, así como métricas de validación que ayudan a optimizar los parámetros del variograma (alcance [*range*], umbral [*sill*] y efecto pepita [*nugget*]).

Cálculo del Semivariograma Experimental

Inicialmente, se calculó el semivariograma experimental a partir de los datos de elevación y sus coordenadas espaciales utilizando la función `UniversalKriging` de `PyKrige`. Para ejecutar el análisis del semivariograma se creó una función en `Python` para controlar los parámetros de entrada del comando `UniversalKriging`. En la tabla 4.2 se enumeran los parámetros de control de la función creada. El estimador de semivarianza utilizado fue el de *Matheron* [110], el cual es simple de implementar y no es tan costoso computacionalmente [105]. La fórmula que utiliza este estimador es la siguiente:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} (Z(x_i) - Z(x_i + h))^2 \quad (4.15)$$

donde:

- $\gamma(h)$: es la semivarianza estimada para una distancia h .
- $N(h)$: número de pares de puntos separados por una distancia aproximadamente igual a h .
- $Z(x_i)$: valor de la variable en la ubicación x_i .
- $Z(x_i + h)$: valor en la ubicación separada por h .

Pruebas de estacionariedad espacial

La estacionariedad espacial se refiere a la propiedad de un campo aleatorio donde las estadísticas conjuntas de dos puntos cualesquiera en el campo dependen únicamente del vector de separación (tanto en magnitud como en dirección) entre los puntos [111], y no de su ubicación absoluta en el espacio. En otras palabras, las características estadísticas (como la media y la varianza) de la variable que se está modelando son las mismas en cualquier lugar del área de estudio, y la correlación entre dos puntos depende solo de la distancia y dirección que los separa [111]. Para probar la estacionariedad en los datos de elevación analizados se realizaron dos aproximaciones; en la primera se aplicó una prueba de *Kruskal-Wallis* basada en cuatro máscaras espaciales y se compararon las medianas de estas máscaras para ver si hay diferencias espaciales en cada una de ellas. Esta prueba se implementó con la librería `Scipy` de `Python` [112].

La segunda aproximación consistió en el análisis de Regresión ponderada geográficamente (*Geographically Weighted Regression - GWR*) la cual es una técnica de análisis espacial que permite modelar relaciones entre variables que varían a lo largo de un área geográfica [113]. Esta regresión se utiliza principalmente para analizar datos que exhiben no estacionariedad [113], lo que significa que las relaciones entre las variables no son uniformes en el espacio. Esto puede ocurrir debido a

factores contextuales, diferencias culturales, condiciones ambientales u otros procesos espaciales que varían geográficamente [113]. La fórmula para el cálculo de esta regresión es la siguiente:

$$y_i = \beta_{i0} + \sum_{k=1}^m \beta_{ik} x_{ik} + \varepsilon_i \quad (4.16)$$

donde:

- y_i es la variable dependiente en la ubicación i .
- x_{ik} es la k -ésima variable independiente en la ubicación i .
- m es el número total de variables independientes.
- β_{i0} es el intercepto (ordenada al origen) en la ubicación i .
- β_{ik} es el coeficiente de regresión local para la k -ésima variable independiente en la ubicación i .
- ε_i es el término de error aleatorio en la ubicación i .

Pruebas de anisotropía

Para evaluar la presencia de anisotropía en los datos espaciales, se calcularon semivariogramas direccionales en los ángulos 0° , 45° , 90° y 135° . Cada uno de estos semivariogramas se ajustó a uno de los ángulos, lo que permitió comparar la estructura espacial según diferentes orientaciones. La similitud o divergencia entre estas curvas proporcionó una base para determinar si el fenómeno estudiado presenta dependencia direccional significativa.

Ajuste del Modelo Teórico de Semivariograma

Una vez obtenido el semivariograma experimental, el siguiente paso fue ajustar un modelo teórico que lo aproximara. La librería de Python `Pykrige` ofrece varios modelos predefinidos [105], incluyendo 'spherical', 'exponential', 'gaussian', 'linear', 'power' y 'Hole-Effect'. Para cada área de estudio, se probaron diferentes modelos teóricos, y el ajuste de sus parámetros se realizó mediante la función `variogram model parameters` del objeto `UniversalKriging` [105]. Los parámetros a ajustar para cada modelo son:

- **Nugget (efecto pepita):** Representa la varianza a distancia cero y puede atribuirse a errores de medición o variabilidad a escalas menores que la distancia de muestreo más pequeña.
- **Sill (meseta):** Es la semivarianza máxima a la que tiende el variograma a distancias grandes, representando la varianza total de los datos. La contribución estructural es la diferencia entre la meseta y el efecto pepita.
- **Range (alcance):** Es la distancia a la cual la autocorrelación espacial se vuelve insignificante.

Tabla 4.2: Descripción de los hiperparámetros para la creación de los semivariogramas y los modelos de *Kriging* utilizando PyKrige [105].

Parámetro	Descripción	Explicación de su uso
<code>y</code>	vector de coordenadas de latitud	Necesaria para el cálculo de las distancias
<code>x</code>	vector de coordenadas de longitud	Necesaria para el cálculo de las distancias
<code>z</code>	Valores de la variable de elevación.	Se pasa directamente como argumento a la función. Puede ser muestreado si hay más de 10,000 datos.
<code>variogram_model</code>	Modelo teórico del variograma. Define la forma de la curva de semivarianza.	Se utilizó una lista para automatizar la selección de los diferentes modelos
<code>nlags</code>	Número de intervalos (bins) de distancia para calcular la semivarianza.	Se definió el número de lags o intervalos de distancia (<code>nlags</code>) en 50 para obtener una representación detallada de la variabilidad espacial sin comprometer la robustez de la estimación en cada <i>lag</i>
<code>drift_terms</code>	Términos de deriva que se utilizan en kriging universal	Estos términos modelan una tendencia sistemática en los datos espaciales, es decir, una parte del valor que cambia suavemente en el espacio y no es puramente aleatoria.
<code>exponente</code>	Término para ajustar el modelo <code>power</code>	Este parámetro permite modificar el exponente con que se ajusta el modelo de potencia

- **Factor de escalamiento (exponente):** Es la potencia de ajuste del modelo de potencia.

El proceso de ajuste se realizó mediante la minimización de la suma de los errores cuadráticos entre el semivariograma experimental y el modelo teórico.

4.3.2.2. Interpolación con *Kriging*

El *kriging* es una técnica de interpolación geoestadística basada en teoría de procesos estocásticos. Su objetivo principal es estimar el valor de una variable espacial $Z(\mathbf{s}_0)$ en una ubicación no muestreada \mathbf{s}_0 , utilizando una combinación lineal ponderada de valores observados en ubicaciones cercanas [102]. Esta técnica busca minimizar el error cuadrático medio de predicción, bajo el supuesto de que existe una dependencia espacial entre las observaciones, modelada a través de una función de semivarianza $\gamma(h)$

Existen diversas variantes del *kriging* según los supuestos sobre la media del proceso. El ***kriging* Ordinario** asume que la media es constante pero desconocida dentro del vecindario local de interpolación [101]. Bajo este enfoque, se estiman los pesos λ_i que minimizan la varianza de predicción, imponiendo la restricción de que la suma de los pesos sea igual a uno ($\sum_{i=1}^n \lambda_i = 1$), garantizando así una estimación insesgada.

$$\hat{Z}(\mathbf{s}_0) = \sum_{i=1}^n \lambda_i Z(\mathbf{s}_i), \quad \text{con} \quad \sum_{i=1}^n \lambda_i = 1 \quad (4.17)$$

Por otro lado, el ***kriging* Universal** extiende este enfoque al considerar que la media no es constante, sino que varía sistemáticamente en el espacio [102]. Se modela como una combinación lineal de funciones conocidas de la ubicación, es decir, $m(\mathbf{s}) = \sum_{k=1}^p \beta_k f_k(\mathbf{s})$, donde f_k son funciones determinísticas (por ejemplo, términos lineales o cuadráticos), y β_k son coeficientes desconocidos [102]. Esta variante es adecuada cuando se sospecha la presencia de una deriva o tendencia espacial, como puede ocurrir en superficies topográficas o gradientes ambientales.

$$\hat{Z}(\mathbf{s}_0) = \sum_{i=1}^n \lambda_i Z(\mathbf{s}_i) + \sum_{k=1}^p \mu_k f_k(\mathbf{s}_0) \quad (4.18)$$

En ambos casos, el *kriging* proporciona no solo una estimación puntual del valor desconocido, sino también una medida de la incertidumbre asociada, lo que lo convierte en una muy buena herramienta para el modelado espacial y la toma de decisiones en diferentes disciplinas.

4.3.2.3. Modelos Determinísticos

La implementación de los modelos determinísticos de interpolación (Lineal, Cúbico y Vecino Más Cercano) se llevó a cabo mediante la función `interpolate` de la biblioteca `SciPy` [112]. Esta herramienta permite interpolar datos dispersos multidimensionales sobre una malla regular, aplicando directamente varios métodos determinísticos a puntos irregulares.

El modelo lineal interpoló a través de *simplices*, creando planos lineales entre los elementos geométricos formados por los puntos vecinos, ofreciendo rapidez y transiciones suaves sin oscilaciones

[112]. El modelo cúbico utilizó funciones cúbicas suavizadas, requiriendo mayor densidad de puntos pero generando superficies continuas y suaves, ideal para representar topografía o gradientes ambientales [112]. El modelo de Vecinos Más Cercanos asignó el valor del punto más cercano a cada punto de la malla, siendo rápido y sin parámetros complejos, aunque con posibles discontinuidades abruptas [112].

4.3.3. Implementación de los modelos

La generación de Modelos Digitales de Elevación (MDE) a partir de datos *LiDAR* demandó considerables recursos computacionales debido al inherente volumen masivo de puntos característicos de estas capturas aéreas de alta resolución. La densidad de la información *LiDAR*, si bien es decisiva para la obtención de modelos detallados, impuso una carga computacional significativa. Los conjuntos de datos procesados comprendieron 7 471 712 puntos distribuidos en 17 escenas para Ciudad Mutis, 4 988 597 puntos en 17 escenas para Buenaventura y un total de 15 657 467 puntos abarcando 20 escenas para Tumaco.

Durante la implementación de los modelos probabilísticos, específicamente en la etapa de cálculo y ajuste del semivariograma experimental a los modelos teóricos propuestos, se encontraron limitaciones de *hardware*. El proceso resultó ser prohibitivo en términos de memoria RAM disponible, impidiendo la computación para la totalidad de los conjuntos de datos. Adicionalmente, la ausencia de funcionalidades de paralelización o procesamiento mediante Unidad de Procesamiento Gráfico (GPU) en la librería *PyKriging* para estas operaciones intensivas limitó las estrategias de optimización. Ante esta restricción computacional, se optó por implementar los modelos probabilísticos utilizando un subconjunto representativo de datos, consistente en los 1 000 puntos con los que el sistema pudo completar el cálculo del semivariograma y su ajuste; para los modelos determinísticos sí se logró realizar las corridas con todo el conjunto de datos de cada casco urbano, utilizando parámetros basados en *Compute Unified Device Architecture - CUDA* implementados a través de *PyCUDA* [114].

4.3.3.1. Ciudad Mutis

Modelos Geoestadísticos (*Kriging*)

El semivariograma para Ciudad Mutis muestra una tendencia clara de aumento de la semivarianza con la distancia (Figura 4.10). Inicialmente, para distancias cortas, la semivarianza es baja, lo que indica que los puntos de elevación cercanos entre sí presentan valores bastante similares, reflejando una alta autocorrelación espacial local. A medida que la distancia entre los puntos se incrementa, la semivarianza también aumenta de manera constante y relativamente lineal hasta una distancia de aproximadamente unos 1 500 metros; luego hay variaciones de esta semivarianza entre los 1 500 y los 2 000 metros. Esta tendencia sugiere que la similitud en la elevación disminuye progresivamente con la distancia, lo que implica que la autocorrelación espacial se debilita a medida que los puntos están más alejados; sin embargo, después de los 1 500 metros, la semivarianza cae y presenta un patrón oscilatorio (Figura 4.10).

La Figura 4.11 muestra que no hay un efecto direccional en la relación entre el espacio y la elevación en el casco urbano de Ciudad Mutis. La prueba de *Kruskal-Wallis* para la estacionariedad,

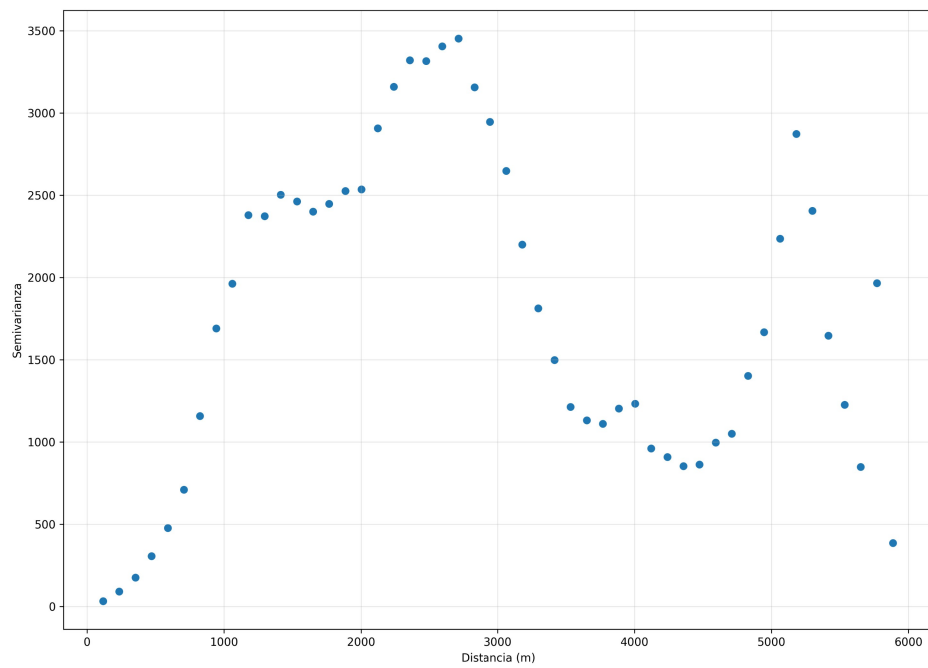


Figura 4.10: Semivariograma de las elevaciones del casco urbano de de Ciudad Mutis.

mostró que entre las cuatro máscaras espaciales construidas para este análisis, existen diferencias significativas basadas en las medianas, por lo que no hay una estacionariedad de la elevación en este casco urbano ($H = 558$, p – valor = 0.000, Mediana por máscara: [26.41, 0.79, 21.7, 19.27], Media global: 30.26).

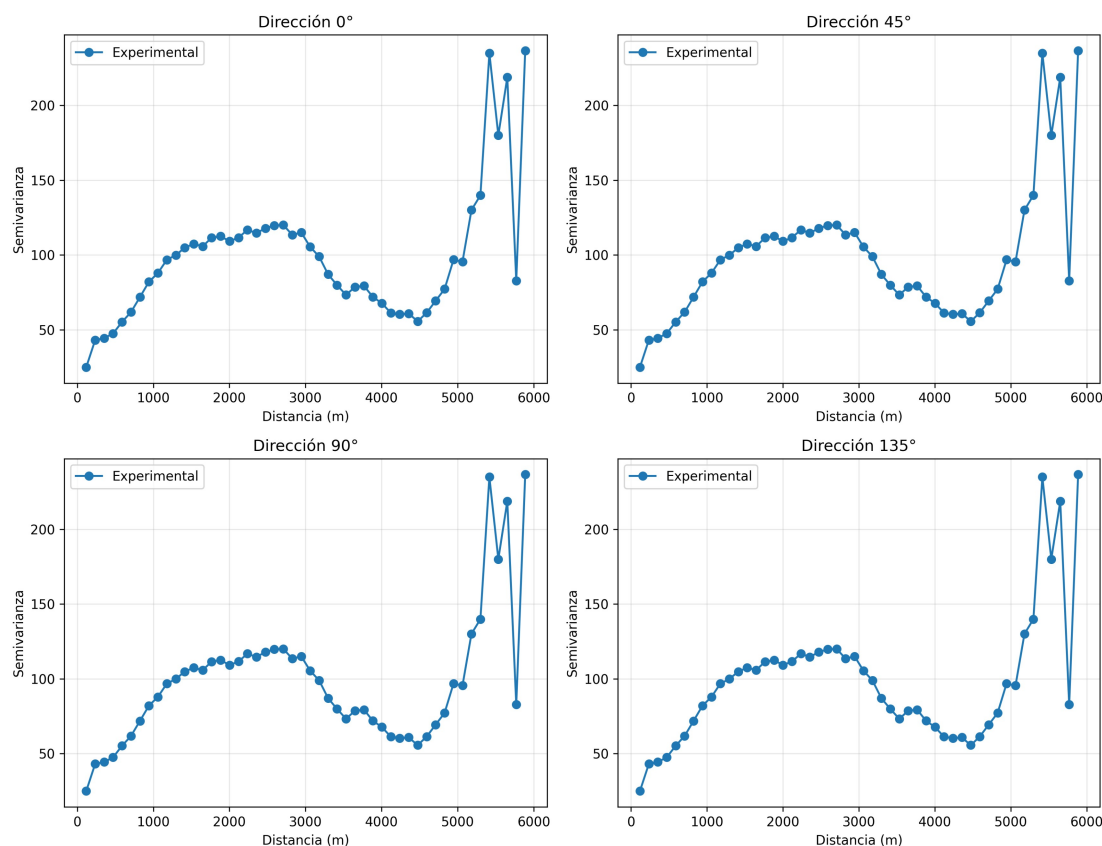


Figura 4.11: Semivariogramas con efectos direccionales en ángulos de 0°, 45°, 90° y 135°, de las elevaciones del casco urbano de de Ciudad Mutis, usando los 10 mil datos seleccionados.

El Análisis de *GWR* (Figura 4.12) mostró que no hay evidencia de estacionariedad en toda el área de estudio para la elevación; por lo tanto, se debe considerar esto en el ajuste del modelo de *Kriging*. Para considerar este supuesto de no estacionariedad con las elevaciones de Ciudad Mutis, se consideró el uso de *Kriging* Universal implementado con la librería *PyKrige*.

La Figura 4.13 exhibe el ajuste de los modelos teóricos de semivariograma (Exponencial, Esférico y Potencia) al semivariograma experimental derivado de los datos *LiDAR* de elevación para Ciudad Mutis. Los puntos azules representan la varianza observada en función de la distancia (*lag*), mientras que la línea negra ilustra el modelo teórico ajustado. La calidad de este ajuste es un factor

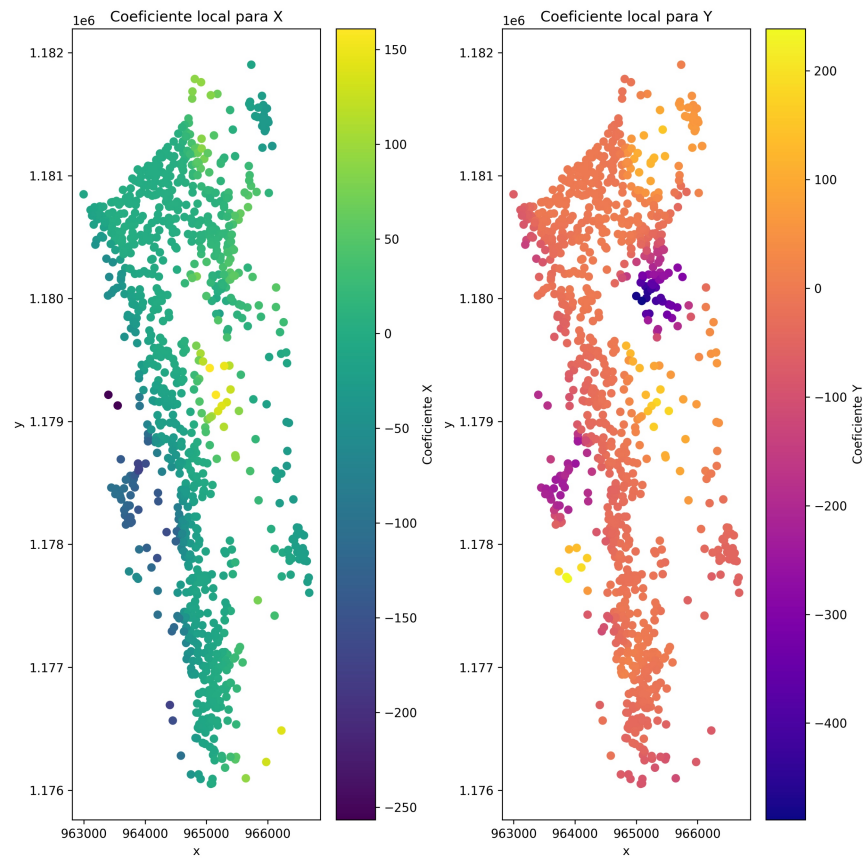


Figura 4.12: Análisis de Regresión Ponderada Geográficamente para la elevación de Ciudad Mutis.

importante, ya que el modelo seleccionado determinará la precisión de las predicciones obtenidas mediante la técnica de *Kriging*.

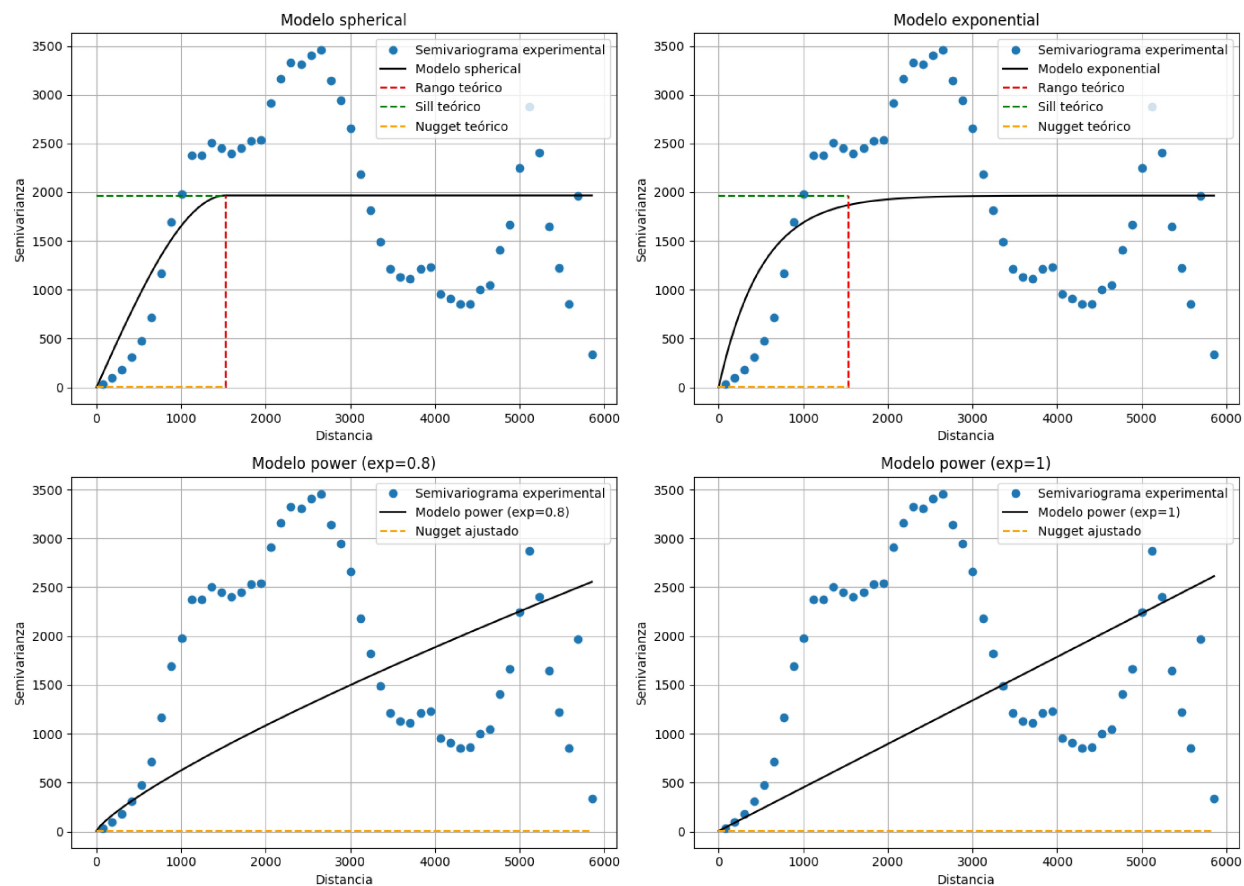


Figura 4.13: Comparación del ajuste de modelos teóricos de semivariograma (Exponencial, Esférico, Potencia) al semivariograma experimental de elevaciones en el casco urbano de Ciudad Mutis.

Los ajustes de los modelos teóricos al semivariograma experimental permitieron evaluar el comportamiento espacial de los datos de elevación. El modelo exponencial reprodujo adecuadamente la pendiente inicial del semivariograma, capturando la rápida ganancia de semivarianza a distancias cortas y estabilizándose cerca del *sill*, aunque no logró representar las oscilaciones observadas a mayores distancias. De manera similar, el modelo esférico ofreció un ajuste comparable, con una transición más suave hacia el *sill* y una representación aceptable de la estructura espacial hasta el rango estimado, pero sin capturar la variabilidad presente más allá de los 3 000 metros. El modelo *power* no representó ni la pendiente inicial ni las oscilaciones posteriores observadas en el semivariograma con ninguno de los dos exponentes analizados, mostrando una discrepancia significativa

Tabla 4.3: Valores de los parámetros de los variogramas teóricos ajustados al variograma experimental para las elevaciones de Ciudad Mutis.

Modelo	Sill	Range	Nugget
<i>Spherical</i>	1967	1531	2.0×10^{-5}
<i>Exponential</i>	1965	1531	1.0×10^{-5}
<i>Power</i> exp = 0.8	-	-	5
<i>Power</i> exp = 1	-	-	2

con respecto a la variabilidad real de los datos. Esta falta de ajuste sugiere que el modelo no es adecuado para describir la estructura espacial de los datos de elevación en este caso.

La Figura 4.14 presentó los mapas de elevación interpolados resultantes de la aplicación de *Kriging* utilizando cada uno de los cuatro modelos de semivariograma ajustados para Ciudad Mutis. Se observaron diferencias visuales significativas entre los mapas generados por cada modelo. Los modelos Exponencial y Esférico produjeron superficies de elevación continuas con patrones similares. El modelo *power* con sus exponentes mostró patrones similares con un rango de elevaciones dentro del rango encontrado en los datos.

Modelos determinísticos

La Figura 4.15 compara los Modelos Digitales de Elevación (MDE) para Ciudad Mutis generados mediante métodos determinísticos lineal (A), cúbico (B) y vecino más cercano (C). El modelo lineal ofreció una representación continua y detallada con un rango de elevación esperado, capturando patrones topográficos coherentes. En marcado contraste, el modelo cúbico generó valores de elevación anómalos, que oscilan entre -2 000 y 2 000 metros, lo cual es irreal para la zona y sugiere un sesgo hacia la sobreestimación en áreas con menos datos. El modelo del vecino más cercano, aunque con un rango de elevación más plausible, presentó discontinuidades visuales debido a su método de interpolación discreto. En general, el modelo lineal demostró ser el más adecuado para representar el terreno de Bahía Solano entre los métodos determinísticos evaluados.

4.3.3.2. Buenaventura

Modelos Geoestadísticos (*Kriging*)

El semivariograma experimental obtenido para la ciudad de Buenaventura evidencia una clara estructura de autocorrelación espacial positiva, manifestada en el incremento sistemático de la semivarianza con el aumento de la distancia entre puntos de muestreo (Figura 4.16). Este comportamiento es característico de procesos geoestadísticos en los que las observaciones espacialmente cercanas tienden a presentar valores similares, mientras que dicha similitud disminuye progresivamente conforme aumenta la separación.

Durante las primeras distancias (<6 000 metros), la semivarianza exhibe un crecimiento moderado y progresivo, lo que sugiere una dependencia espacial de baja a media escala. No obstante, a partir de los 7 000 metros —y con mayor énfasis entre los 8 000 y 9 000 metros— se registra un

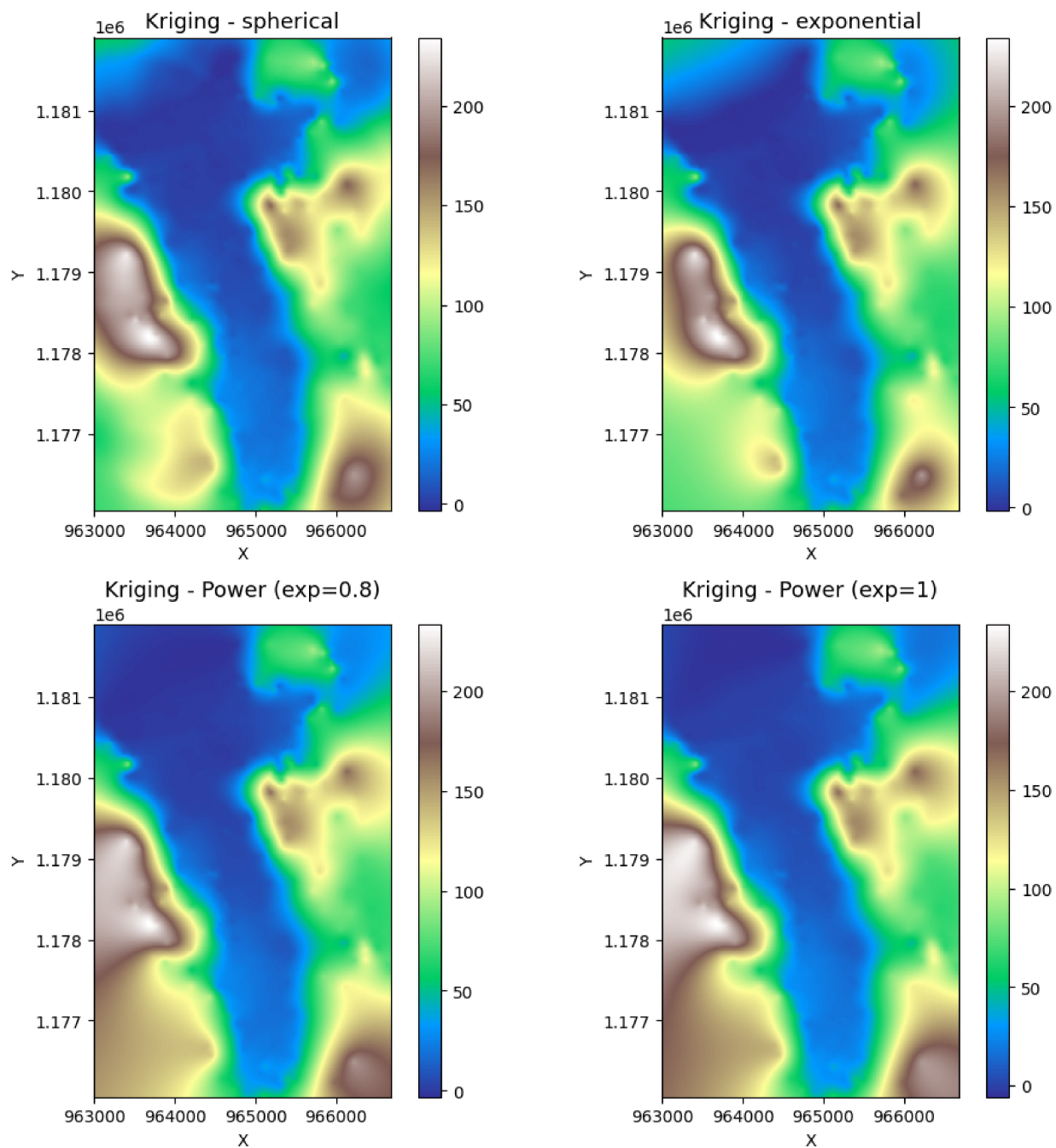


Figura 4.14: Interpolación de los modelos Exponencial, Esférico y Potencia usando la técnica de *kriging* obtenidos de las elevaciones del casco urbano de Ciudad Mutis.

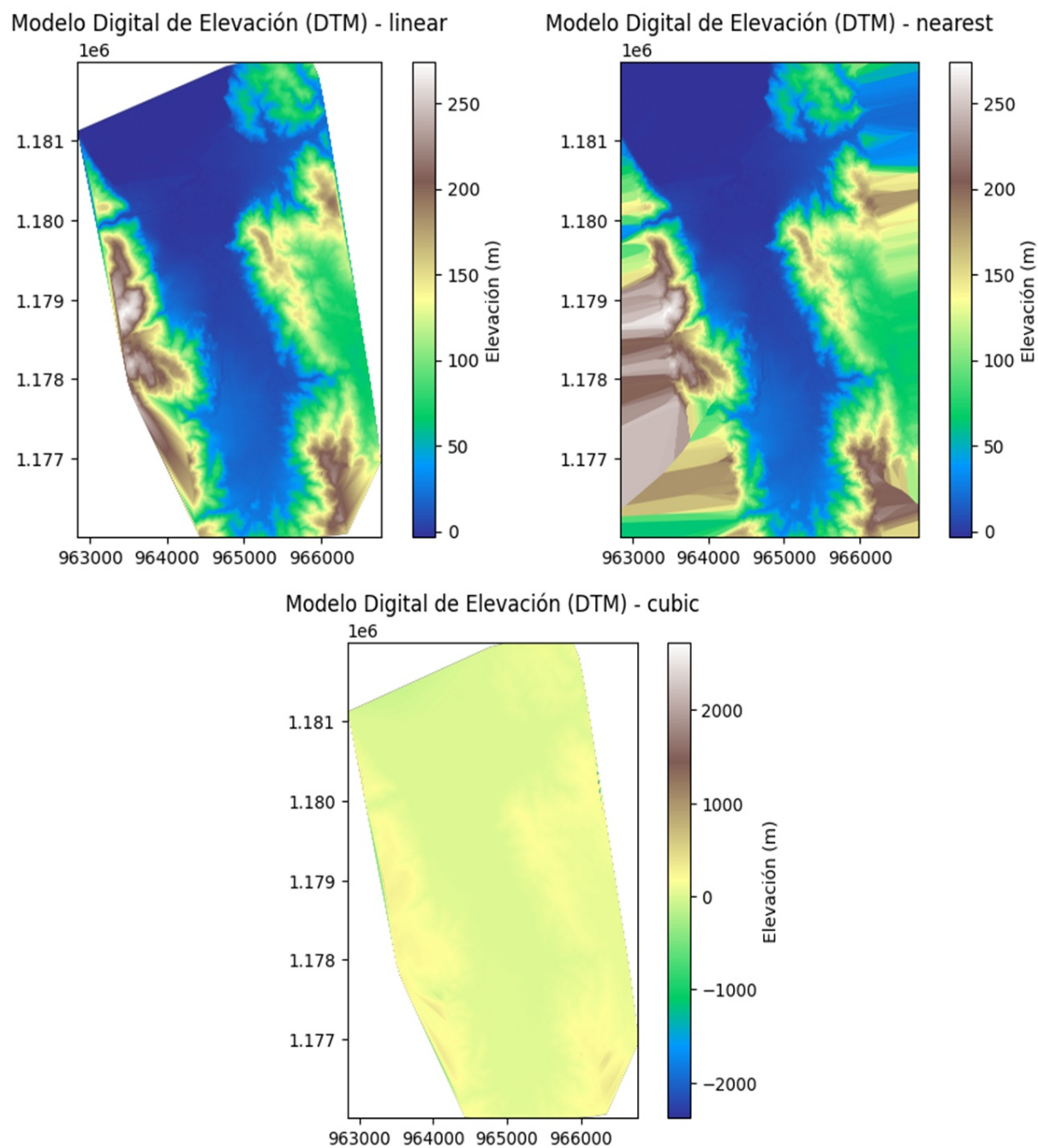


Figura 4.15: Modelos de elevación del casco urbano de Ciudad Mutis obtenidos por métodos determinísticos

incremento abrupto en la semivarianza. Este patrón sugiere la existencia de transiciones espaciales pronunciadas o posibles discontinuidades estructurales en el fenómeno analizado, potencialmente relacionadas con variaciones abruptas en la topografía, cambios de cobertura del suelo o condiciones ambientales heterogéneas.

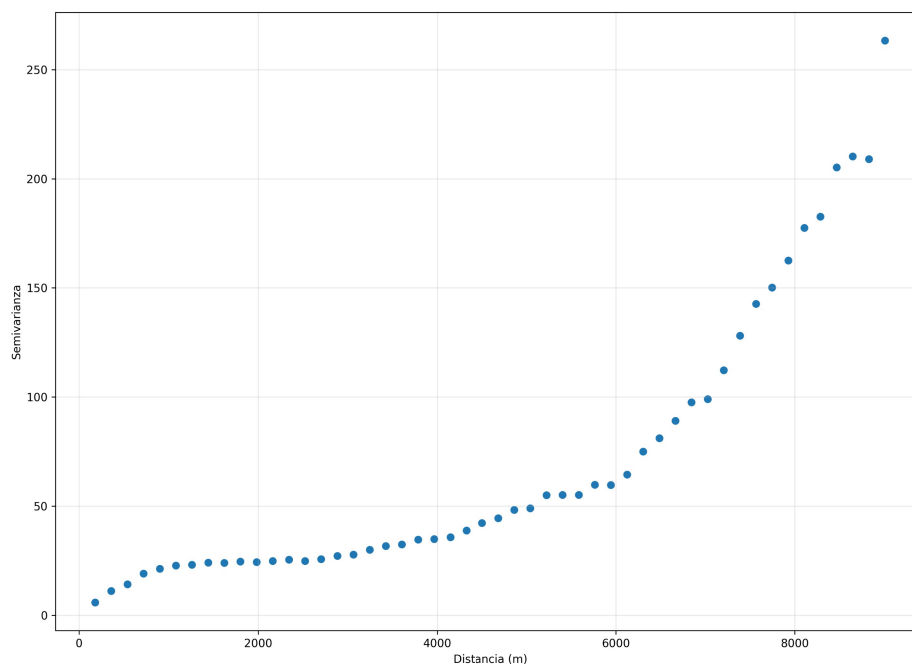


Figura 4.16: Semivariograma de las elevaciones del casco urbano de de Buenaventura.

La Figura 4.17 permite observar que las elevaciones en Buenaventura presentan una marcada tendencia espacial, evidenciada tanto por el crecimiento sostenido del semivariograma original como por la existencia de un gradiente altitudinal en el mapa de alturas. Una vez removida dicha tendencia, los residuos muestran una estructura espacial más débil y estacionaria, con un semivariograma considerablemente más estable y una distribución espacial de valores mucho más homogénea. Este resultado valida la necesidad de emplear técnicas como el *kriging* Universal, capaces de modelar datos que no cumplen con el supuesto de estacionariedad.

El análisis mediante *GWR* (Figura 4.18) corroboró la no estacionariedad de la elevación en toda el área de estudio, destacando la necesidad de incorporar esta condición en el ajuste del modelo geoestadístico de *kriging*.

La Figura 4.19 muestra que no se evidencian patrones de anisotropía en la relación espacio-elevación en el casco urbano de Buenaventura. No obstante, la prueba de *Kruskal-Wallis* para evaluar la estacionariedad espacial reveló diferencias significativas entre las medianas de cuatro

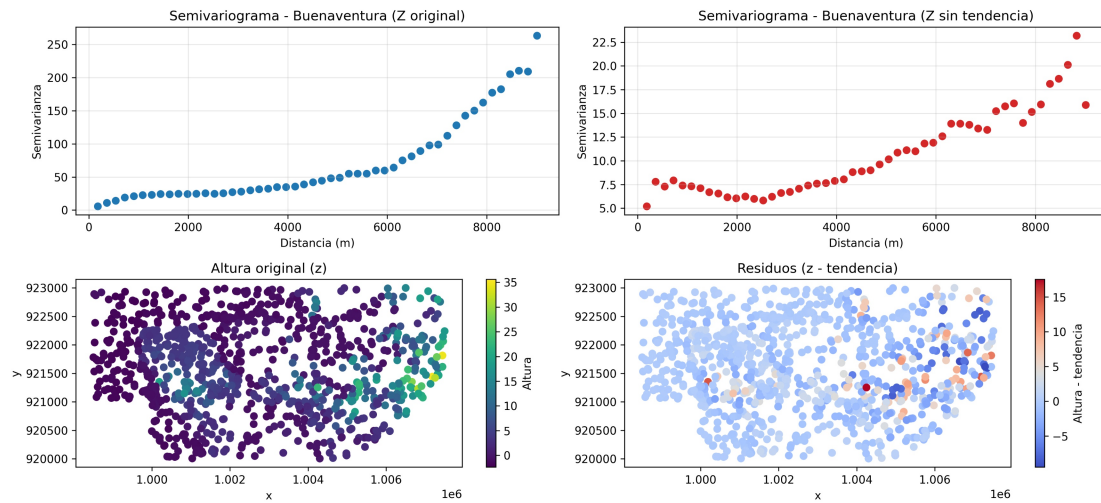


Figura 4.17: Semivariograma con la elevación original (h) y sin tendencia espacial no-estacionaria para el casco urbano de Buenaventura.

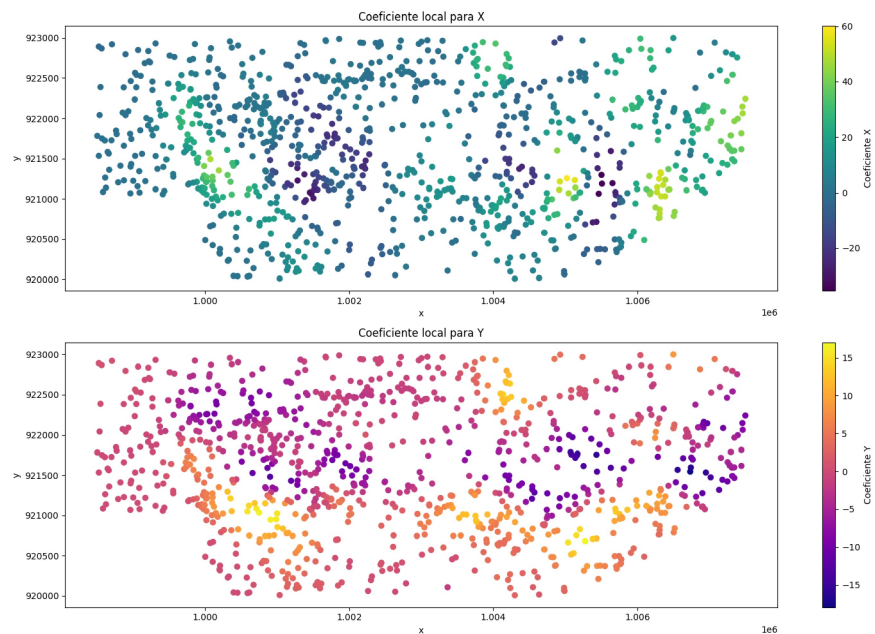


Figura 4.18: Análisis de Regresión Ponderada Geográficamente para la elevación de Buenaventura.

regiones espaciales definidas, lo cual indica la ausencia de estacionariedad en la variable elevación ($H = 185$, p -valor = 0.000, Medianas por máscara: [-0.44, -0.78, 3.64, -0.34], Media global: 3.28).

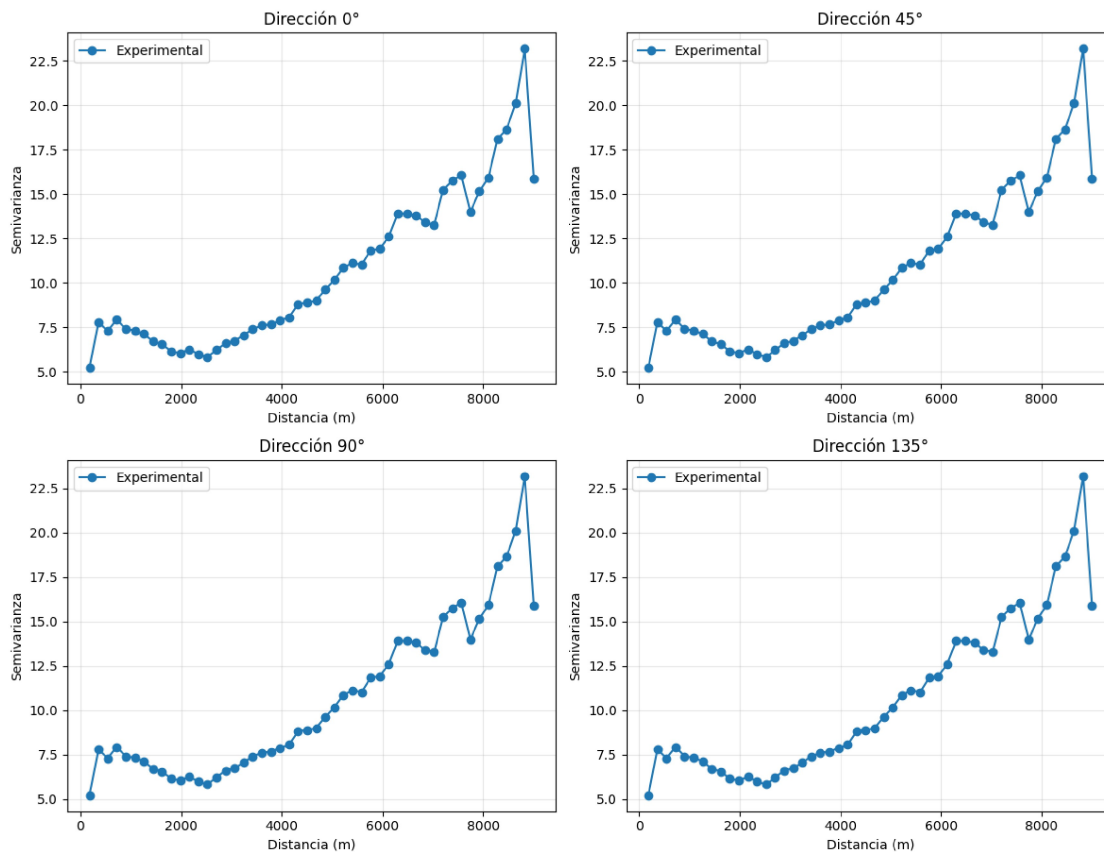


Figura 4.19: Semivariogramas con efectos direccionales en ángulos de 0°, 45°, 90° y 135°, de las elevaciones del casco urbano de de Buenaventura, usando los 10 mil datos seleccionados.

La Figura 4.20 presenta los ajustes de cuatro modelos teóricos de semivariograma (Exponencial, Esférico y Potencia) sobre el semivariograma experimental construido a partir de los datos de Buenaventura. En general, todos los modelos presentan limitaciones notorias para capturar adecuadamente la estructura espacial observada en los datos. Si bien los modelos exponencial y esférico ofrecen un ajuste razonable en las primeras distancias, es decir, en la zona de corto alcance o efecto local, tienden a subestimar significativamente la variabilidad en los tramos más largos, especialmente más allá de los 6000 metros.

Por su parte, el modelo de tipo *power*, evaluado con diferentes valores de exponente, mostró un comportamiento más flexible ante la estructura de dependencia espacial. En particular, el modelo *power* con un exponente de 1.3 logró el mejor ajuste visual y cuantitativo al semivariograma experimental, capturando de forma más precisa la tendencia creciente de la semivarianza con la distancia, sin imponer un umbral o *sill* bien definido como lo hacen los modelos anteriores.

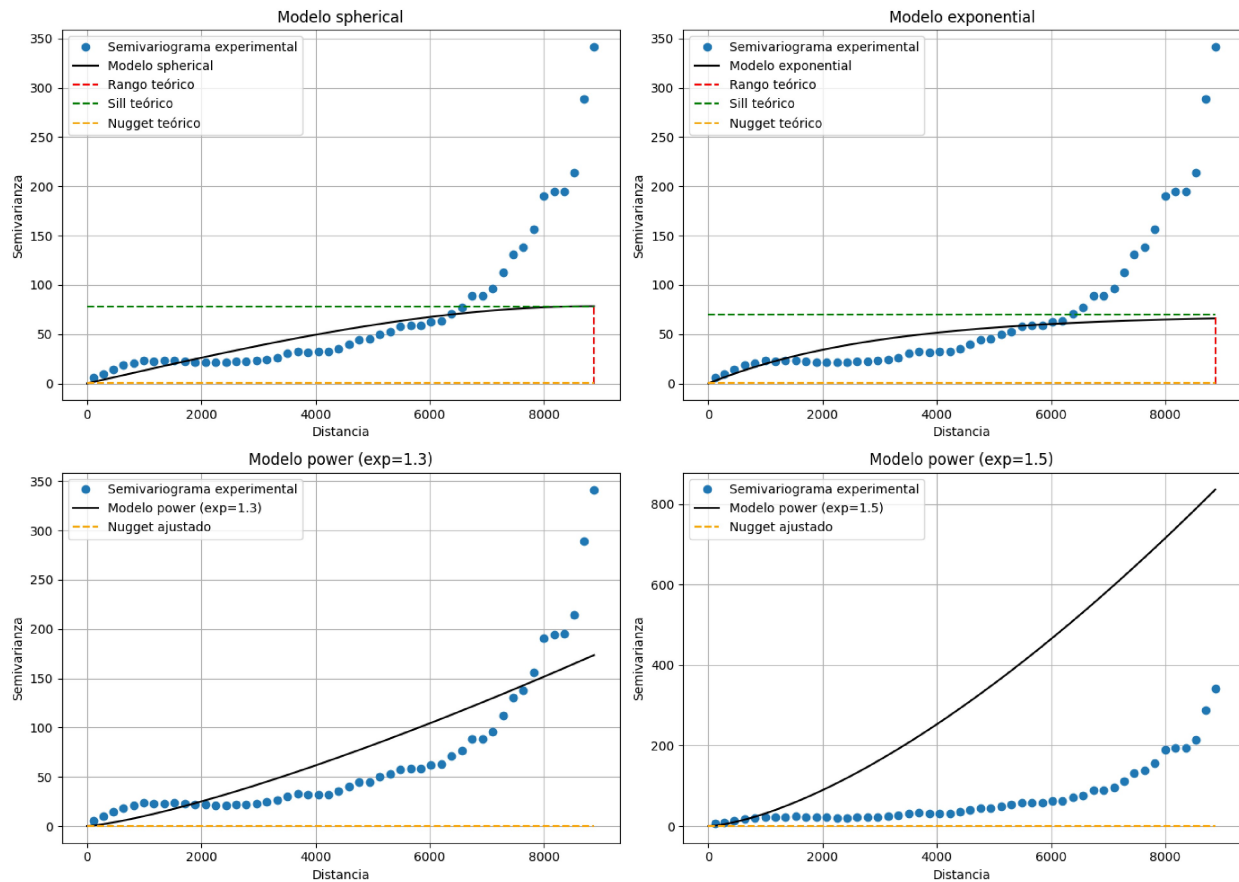


Figura 4.20: Comparación del ajuste de modelos teóricos de semivariograma (Exponencial, Esférico y Potencia) al semivariograma experimental de elevaciones en el casco urbano de Buenaventura.

Tabla 4.4: Valores de los parámetros de los variogramas teóricos ajustados al variograma experimental para las elevaciones de Buenaventura.

Modelo	Sill	Range	Nugget
<i>Spherical</i>	78	8876	1.0×10^{-10}
<i>Exponential</i>	69.4	8876	1.0×10^{-10}
<i>Power exp = 1.3</i>	-	-	1.0×10^{-25}
<i>Power exp = 1.5</i>	-	-	1.0×10^{-25}

La Figura 4.21 ilustra los mapas de interpolación generados mediante *kriging* Universal usando los diferentes modelos de semivariograma. Los modelos esférico, exponencial y lineal produjeron resultados visualmente similares y consistentes con la estructura topográfica observada, mostrando variaciones altitudinales bien definidas y transiciones espaciales suaves. Estos modelos lograron capturar las complejidades locales del relieve, particularmente en la zona oriental, donde se observan los mayores valores de elevación.

En contraste, el modelo *power* produjo un mapa de interpolación sustancialmente más uniforme y suavizado, incapaz de reflejar la variabilidad local presente en los datos originales. Esta pérdida de detalle se atribuye al mal ajuste del modelo *power* al semivariograma experimental, lo que conllevó a una representación espacial simplificada. En conjunto, los resultados refuerzan la importancia de una correcta elección del modelo de semivariograma para garantizar la fidelidad de la interpolación por *kriging*, siendo los modelos esférico, exponencial y lineal claramente superiores al modelo *power* en este caso.

Modelos determinísticos

La Figura 4.22 compara los Modelos Digitales de Elevación (MDE) para Buenaventura obtenidos mediante métodos determinísticos lineal, cúbico y vecino más cercano. El modelo lineal presenta una representación continua y detallada del terreno, con un rango de elevación esperado de -5 a 35 metros. En contraste, el modelo cúbico genera un rango de elevación anómalo, desde -300 hasta 100 metros, con una superficie suavizada y carente de detalle topográfico realista. El modelo del vecino más cercano, aunque mantiene el rango de elevación adecuado, muestra una representación discreta con bordes definidos.

El modelo lineal se destacó entonces, por proporcionar una representación visualmente coherente y realista de la elevación en Buenaventura, dentro del rango esperado. El modelo cúbico produce resultados no plausibles, mientras que el modelo del vecino más cercano sacrifica la continuidad por una representación basada en los puntos de datos más cercanos. Por lo tanto, el modelo lineal parece ser el método determinístico más adecuado para generar un MDE representativo del área de estudio.

4.3.3.3. Tumaco

Modelos Geoestadísticos (Kriging)

El semivariograma experimental para Tumaco muestra una estructura de autocorrelación espacial positiva en las primeras distancias, evidenciada por el incremento progresivo de la semivarianza hasta aproximadamente los 7 000–7 500 metros. Este comportamiento sugiere que los valores de elevación entre puntos cercanos son similares, lo que es característico de una dependencia espacial (Figura 4.23).

A partir de los 8 000 metros, el semivariograma comienza a mostrar un comportamiento más errático, con un aumento abrupto de la semivarianza que alcanza un máximo cercano a los 9 500 metros, seguido de una caída notable. Esta oscilación en los valores sugiere que a mayores distancias se pierde la dependencia espacial o incluso se podrían estar reflejando algo de ruido (Figura 4.23).

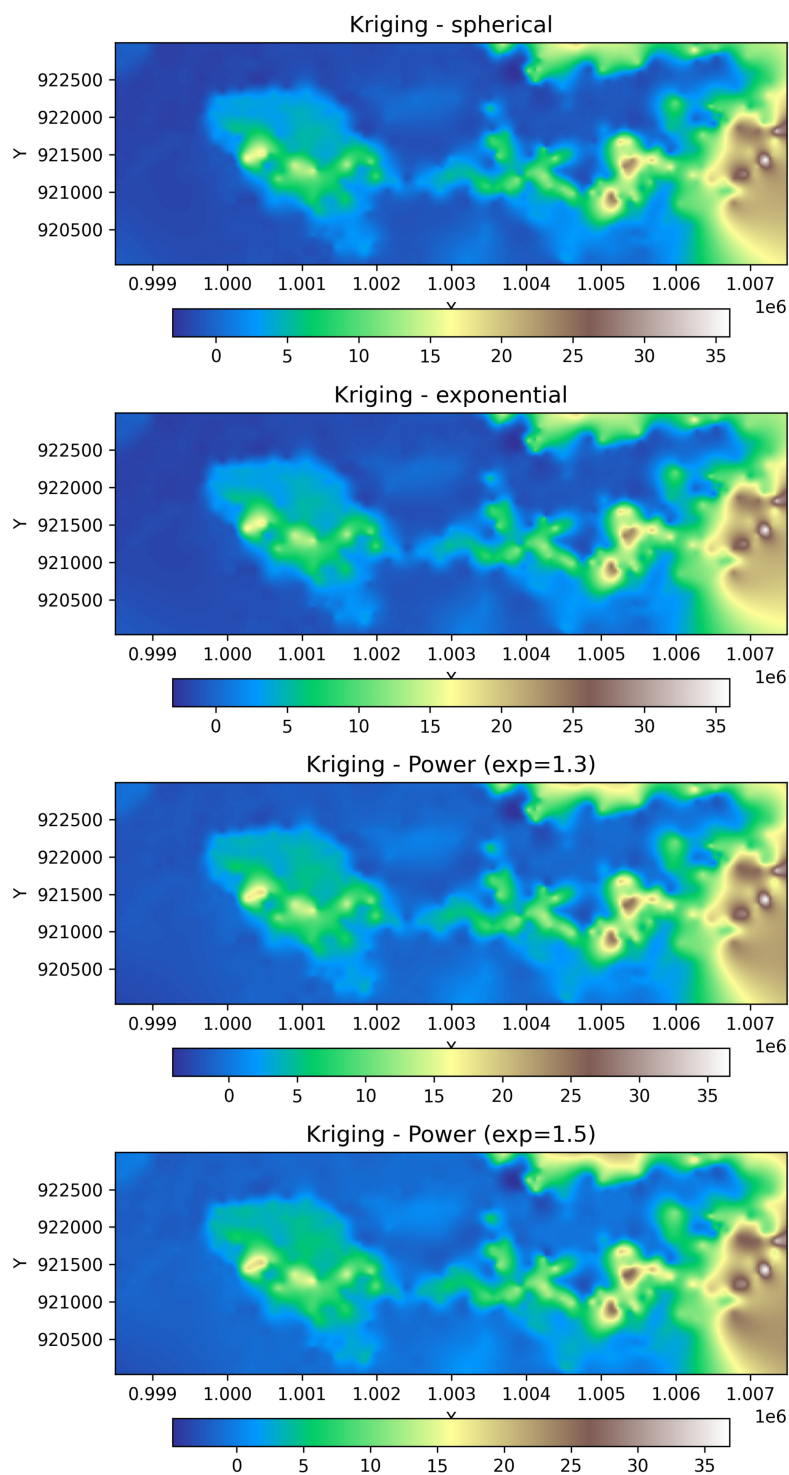


Figura 4.21: Interpolación de los modelos Exponencial, Esférico y Potencia usando la técnica de *kriging* obtenidos de las elevaciones del casco urbano de Buenaventura.

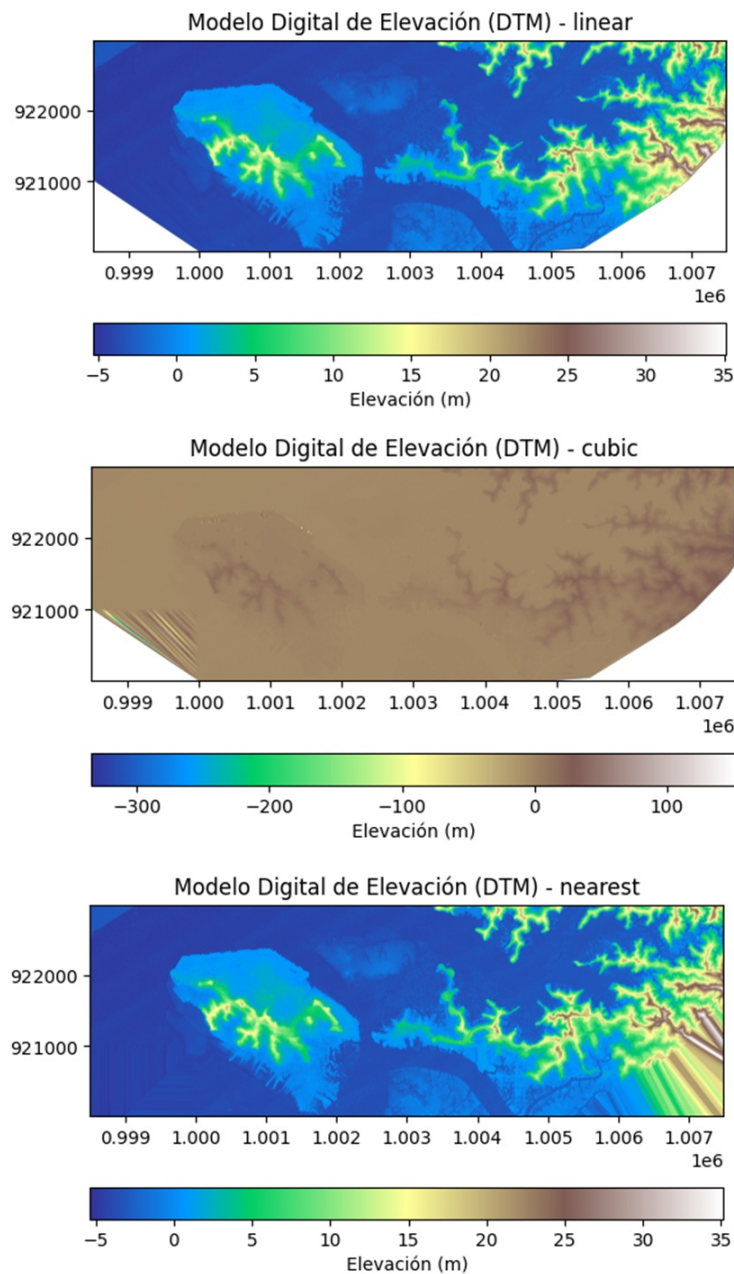


Figura 4.22: Modelos de elevación del casco urbano de Buenaventura obtenidos por métodos determinísticos.

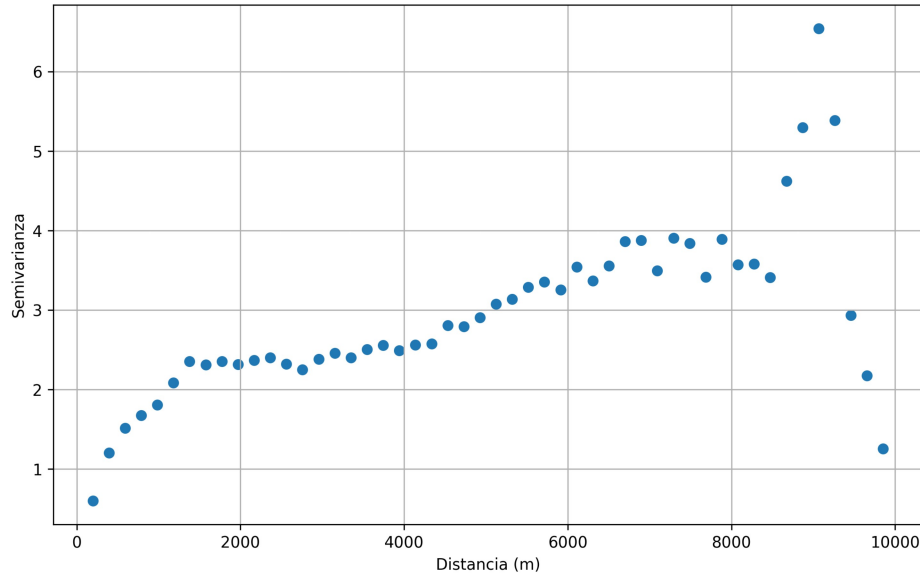


Figura 4.23: Semivariograma de las elevaciones del casco urbano de de Tumaco.

El análisis espacial de la elevación en Tumaco mostró inicialmente una fuerte tendencia espacial, reflejada en el crecimiento progresivo del semivariograma experimental. Tras eliminar esta tendencia, los residuos presentaron una estructura más estacionaria y homogénea, como se evidenció en el semivariograma reducido y en el mapa espacial de los residuos (Figura 4.24). Sin embargo, a pesar de esta corrección, persiste una clara estructura espacial en las diferencias de elevación, especialmente a partir de los 7 000 metros, lo que indica que las variaciones locales en el relieve no fueron completamente explicadas por la tendencia global y continúan influyendo en la dependencia espacial del terreno.

La Figura 4.25 muestra los coeficientes locales estimados mediante el modelo de regresión ponderada geográficamente (*GWR*) para las variables espaciales *longitud* (x) y *latitud* (y) en relación con la elevación en Tumaco. En ambos mapas se observa una marcada variabilidad espacial en los coeficientes, lo que indica que la relación entre las coordenadas espaciales y la altitud no es constante a lo largo del territorio. En el caso del coeficiente para *longitud* (x), se evidencian zonas con valores positivos elevados y otras con valores negativos, lo que refleja direcciones de pendiente opuestas según la ubicación. De manera similar, el mapa del coeficiente para *latitud* (y) muestra un patrón heterogéneo, con regiones donde el gradiente altitudinal es ascendente y otras descendente. Esta variabilidad en los coeficientes confirma la presencia de no estacionariedad espacial en la elevación, es decir, que la relación espacial no puede ser descrita adecuadamente por un modelo global uniforme.

La figura presenta los semivariogramas direccionales para el área de Tumaco 4.26, calculados en cuatro direcciones principales (0° , 45° , 90° y 135°), con el fin de evaluar la posible presencia de

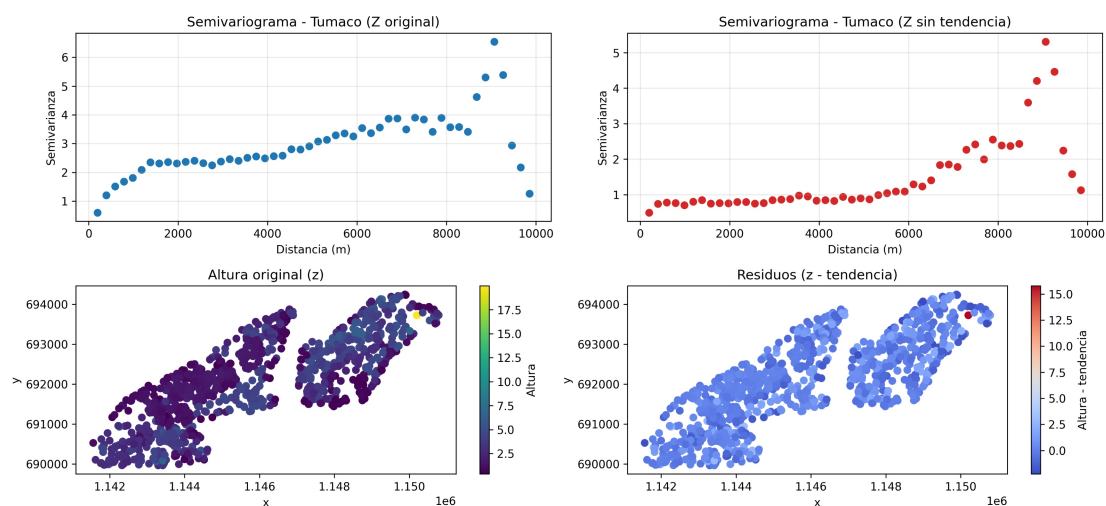


Figura 4.24: Semivariograma con la elevación original (h) y sin tendencia espacial no-estacionaria para el casco urbano de Tumaco.

anisotropía en la estructura espacial de la elevación. En todos los casos, las curvas del semivariograma muestran un patrón muy similar en términos de forma y magnitud, lo que indica que la semivarianza no varía significativamente con la dirección del análisis. Esta homogeneidad entre direcciones sugiere que no existe anisotropía espacial en los datos de elevación, es decir, la dependencia espacial es isotrópica.

El análisis de los modelos ajustados a los semivariogramas experimentales evidenció diferencias en la capacidad de representación de la estructura espacial de los datos topográficos de Tumaco (Figura 4.27). Los modelos exponencial y esférico ofrecieron un ajuste razonable en los tramos iniciales del semivariograma experimental, especialmente en presencia de una meseta o umbral en la semivarianza. El modelo esférico identificó un rango definido de aproximadamente 9 230 metros (Tabla 4.5), lo que muestra que en la extensión del casco urbano de Tumaco la estructura espacial no disminuye con la distancia, sino que se mantiene aumentando.

El modelo exponencial también presentó un ajuste similar al modelo esférico, representando bien la variación de la semivarianza con la distancia; sin embargo, mostró un mejor ajuste en cortas distancias. Por su parte, el modelo *power* (potencial) ofreció el mejor ajuste global, especialmente con exponentes bajos (0.05 y 0.1) (Tabla 4.5 lo que se refleja en la forma suavemente creciente del modelo ajustado sobre todo el dominio.

Los resultados de la interpolación de elevaciones en Tumaco mediante los distintos modelos de *kriging* revelaron diferencias significativas en la capacidad de cada uno para representar de manera adecuada el relieve local (Figura 4.28). El modelo *power* con un exponente de 0.1 generó una superficie más coherente con la topografía real de la zona, reflejando patrones altitudinales que

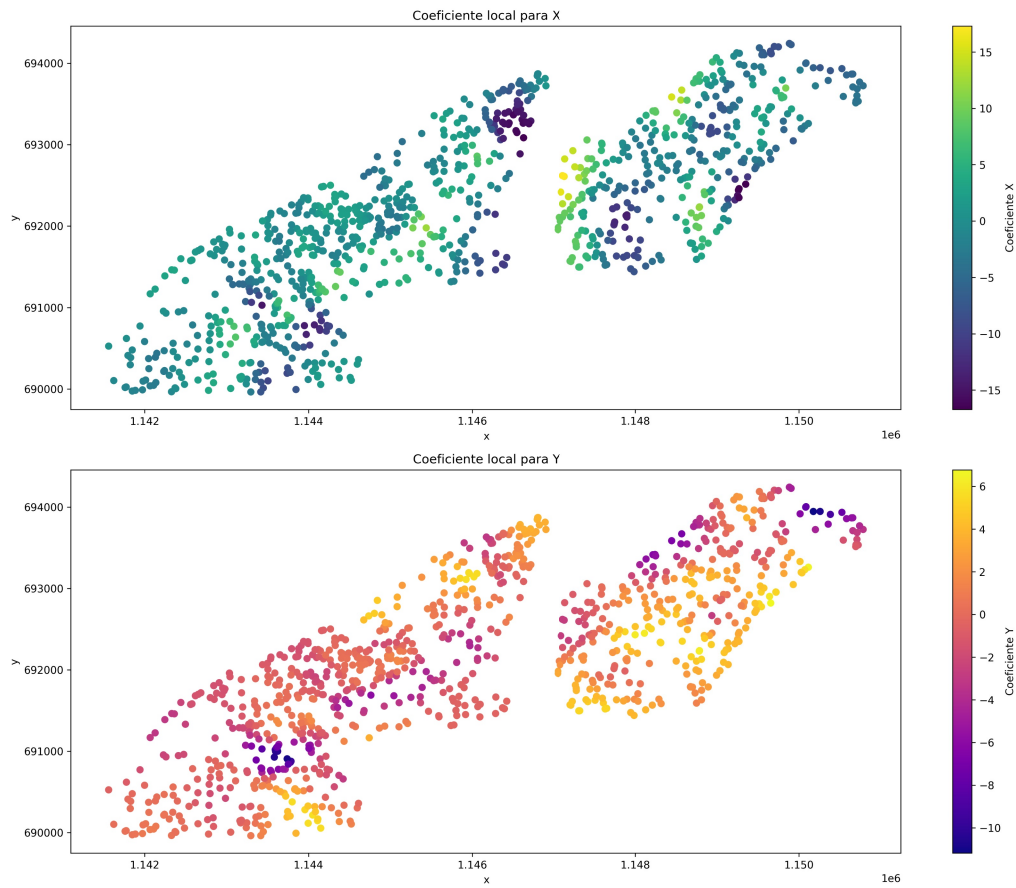


Figura 4.25: Análisis de Regresión Ponderada Geográficamente para la elevación de Tumaco.

Tabla 4.5: Valores de los parámetros de los variogramas teóricos ajustados al variograma experimental para las elevaciones de Tumaco.

Modelo	Sill	Range	Nugget
<i>Spherical</i>	2.7	9230	1.20
<i>Exponential</i>	3.16	9762	0.78
<i>Power exp = 0.05</i>	-	-	1.0×10^{-30}
<i>Power exp = 0.1</i>	-	-	1.0×10^{-30}

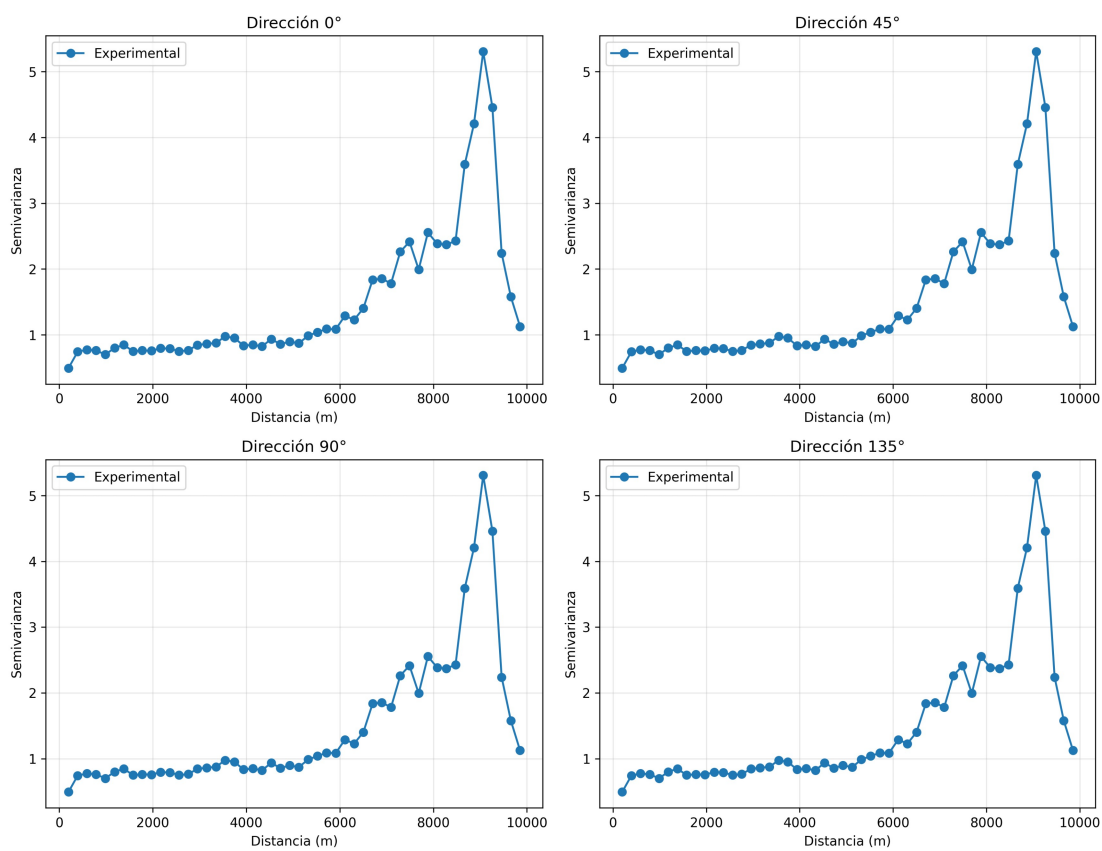


Figura 4.26: Semivariogramas con efectos direccionales en ángulos de 0° , 45° , 90° y 135° , de las elevaciones del casco urbano de de Tumaco, usando los 10 mil datos seleccionados.

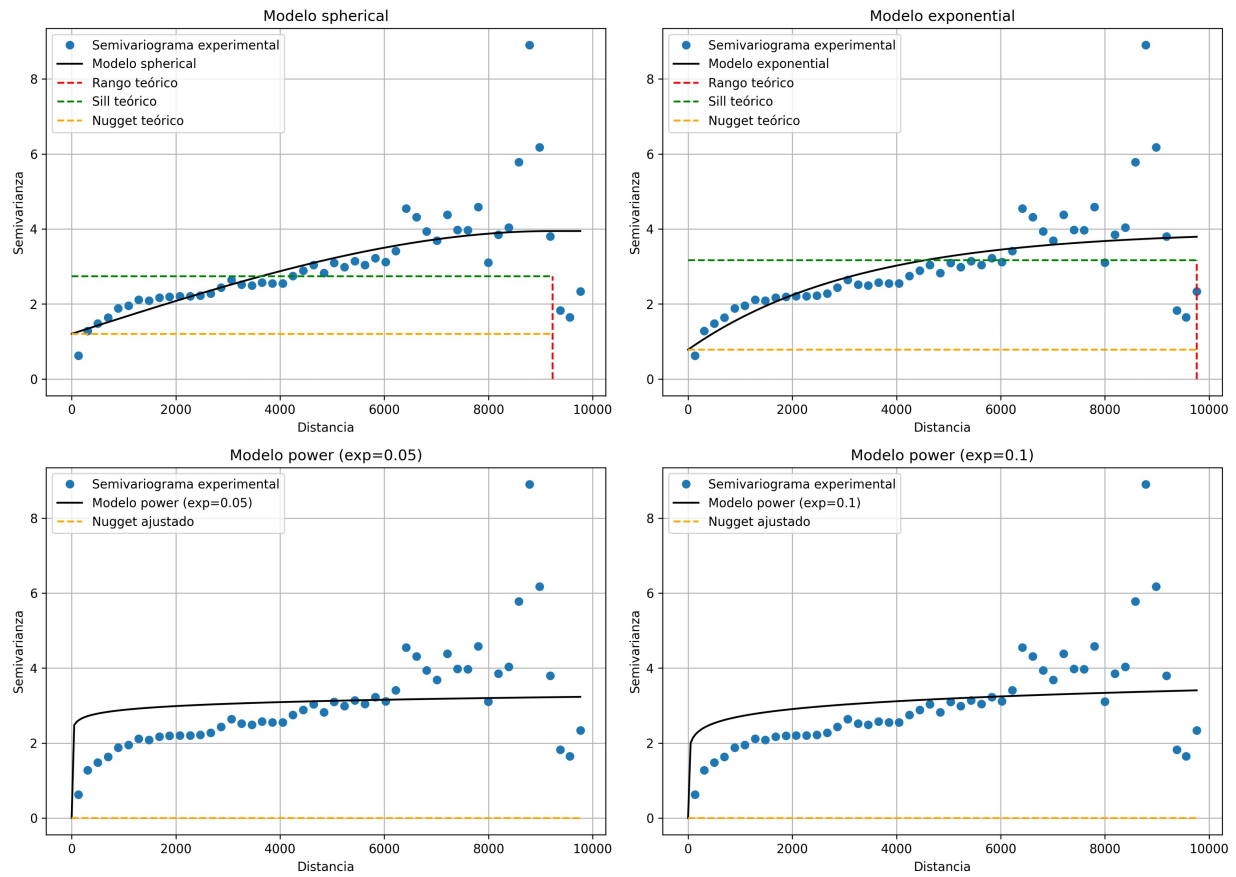


Figura 4.27: Comparación del ajuste de modelos teóricos de semivariograma (Exponencial, Esférico y Potencia) al semivariograma experimental de elevaciones en el casco urbano de Tumaco.

se ajustan mejor a lo esperado desde el punto de vista morfológico. No obstante, se observó una discrepancia entre los valores interpolados y las elevaciones reales, lo que podría atribuirse a la representatividad limitada de la muestra de datos utilizada. En contraste, los modelos esférico y exponencial mostraron una menor capacidad para capturar la variabilidad espacial del terreno, ya que las superficies generadas subestimaron sistemáticamente los valores de elevación, especialmente en las zonas más altas. Este comportamiento pone en evidencia la superioridad relativa del modelo de potencia en este contexto específico, al ofrecer una representación más precisa y realista del patrón altitudinal en el área urbana de Tumaco.

Modelos determinísticos

La Figura 4.29 presenta los Modelos Digitales de Elevación (MDE) para el casco urbano de Tumaco, generados mediante los métodos determinísticos. El modelo lineal (Figura 4.29) exhibe una representación continua de la elevación, con un rango que se extiende aproximadamente desde los -10 hasta los 70 metros, mostrando patrones topográficos suaves y definidos. En contraste, el modelo cúbico (Figura 4.29) revela una representación anómala, con un rango de elevación irreal que va desde -500 hasta 100 metros, presentando una superficie homogénea y poco detallada, lo que sugiere una inadecuada interpolación. El modelo del vecino más cercano (Figura 4.29) ofrece una representación discreta, con bloques de elevación definidos por el punto de datos más cercano, y un rango de elevación más acorde con las expectativas, similar al modelo lineal. En resumen, el modelo lineal parece proporcionar la representación más coherente y dentro del rango esperado de elevaciones para el área de estudio de Tumaco.

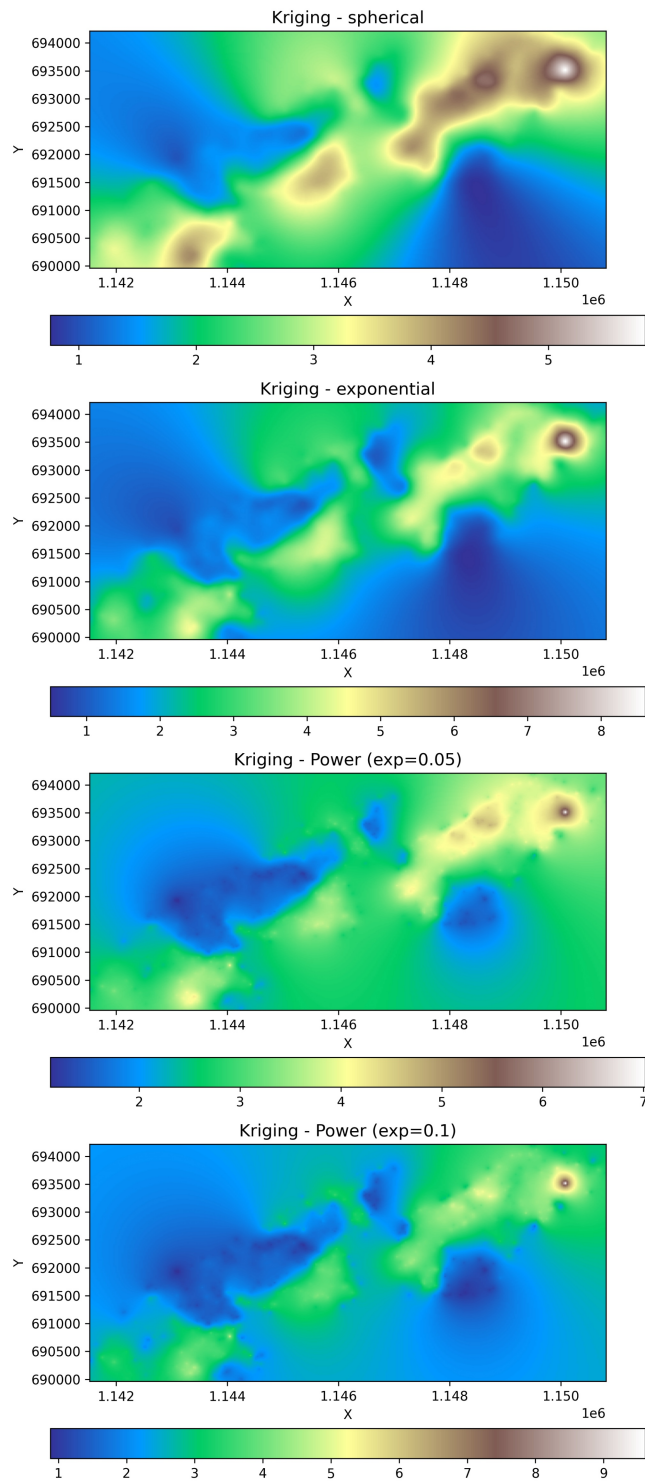


Figura 4.28: Interpolación de los modelos Exponencial, Esférico y Potencia usando la técnica de *kriging* obtenidos de las elevaciones del casco urbano de Tumaco.

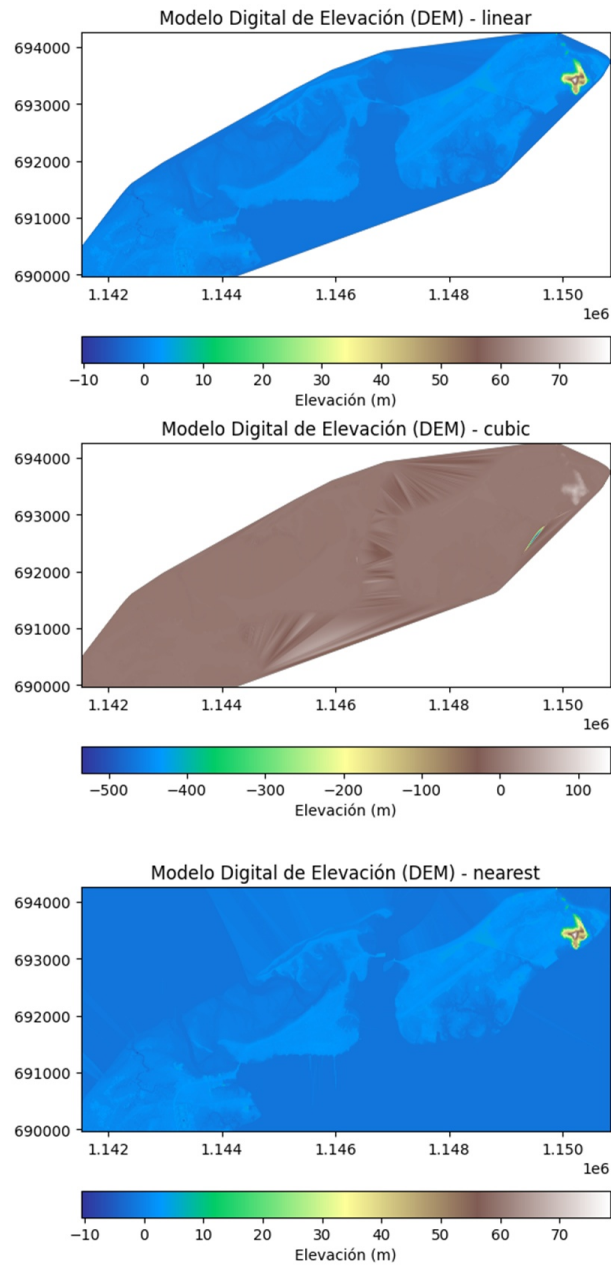


Figura 4.29: Modelos de elevación del casco urbano de Tumaco obtenidos por métodos determinísticos.

Evaluación

5.1. Contexto

La validación de modelos representa una etapa fundamental en todo proceso de modelado predictivo, ya que permite cuantificar el grado de ajuste entre las predicciones generadas por los algoritmos y los valores observados. En este capítulo se presentan los resultados de la evaluación del desempeño de los modelos LSTM implementados para predecir el nivel del mar, así como de los modelos generados para representar la elevación costera mediante técnicas de interpolación espacial.

Para el modelado del nivel del mar, se entrenaron las tres variantes de redes neuronales tipo *Long Short-Term Memory* (LSTM) previamente definidas, utilizando un esquema de validación cruzada bloqueada de tipo *K-Fold* específico por estación. El desempeño de cada enfoque se evaluó a partir de métricas estadísticas tales como el coeficiente de determinación (R^2), el error cuadrático medio (RMSE) y el sesgo (Bias). Sobre la base de estos resultados, se seleccionó para cada localidad el modelo que presentó el mejor equilibrio entre precisión y capacidad de generalización, con el propósito de emplearlo en la generación de escenarios de inundación que se detallarán en el capítulo siguiente.

Adicionalmente, la segunda sección de este capítulo se centra en la evaluación de los Modelos Digitales de Elevación (MDE) generados a partir de datos *LiDAR* para los cascos urbanos de Ciudad Mutis, Buenaventura y Tumaco, implementados tanto para enfoques probabilísticos (kriging) como métodos determinísticos (interpolación lineal, cúbica y vecino más cercano), con el objetivo de representar la topografía costera a partir de muestras puntuales de alta resolución. La calidad de los modelos se evaluó mediante inspección visual y validación cruzada, considerando el ajuste de semivariogramas experimentales, la continuidad espacial y la coherencia altimétrica del terreno.

Los resultados permitieron establecer comparaciones entre métodos, identificar patrones de error espacial y ofrecer recomendaciones sobre la idoneidad de cada enfoque en función del tipo de dato, la región y los objetivos del análisis.

5.2. Métricas de evaluación de modelos

5.2.1. Validación de los modelos de series de tiempo

En este apartado se presentan los resultados de la evaluación de desempeño de los modelos *Long Short-Term Memory* (LSTM) implementados para la predicción de anomalías mensuales del nivel del mar en el Pacífico colombiano. Tal como se detalló en el capítulo anterior, se entrenaron tres configuraciones de redes LSTM, diferenciadas por las variables empleadas como predictoras.

La validación de los modelos se llevó a cabo de manera independiente para cada centro poblado, Ciudad Mutis, Buenaventura y Tumaco, mediante un esquema de validación cruzada bloqueada de tipo *K-Fold*. Se aseguró la preservación de la secuencialidad inherente a las series de tiempo gracias al uso de ventanas temporales, tal como se describió previamente. Para cuantificar el ajuste y la precisión de cada modelo, se calcularon métricas como el error cuadrático medio (RMSE), el coeficiente de determinación (R^2) y el sesgo (Bias).

A continuación, se exponen los resultados obtenidos para cada configuración de red y su desempeño comparativo.

5.2.1.1. Modelo 1: LSTM entrenado con el primer componente canónico completo

Este modelo se entrenó empleando como único predictor el primer componente canónico (*rCCA1*) obtenido del análisis de correlación canónica regularizada (*rCCA*), calculado a partir de las nueve variables climáticas y el nivel del mar como la variable objetivo en todos los casos.

La Figura 5.1 muestra los resultados obtenidos por este modelo en las tres localidades analizadas. Los valores de R^2 se situaron en un rango entre 0.74 y 0.82, mientras que los valores de RMSE fluctuaron entre 0.017 m y 0.021 m. La mejor capacidad explicativa se alcanzó en Buenaventura ($R^2 = 0.82$), mientras que en Tumaco y Ciudad Mutis se evidenció el menor ajuste con $R^2 = 0.74$ y $R^2 = 0.78$, respectivamente. En todas las ubicaciones geográficas, el sesgo (Bias) se mantuvo próximo a cero, lo cual indica ausencia de errores sistemáticos y una predicción globalmente balanceada.

En cuanto a las curvas de pérdida (`loss` y `val_loss`), se observó un descenso pronunciado del error durante las primeras 20 épocas en las tres ubicaciones, reflejando que el modelo logró capturar patrones relevantes en las series temporales. Sin embargo, las trayectorias posteriores muestran matices diferenciados por localidad.

En Ciudad Mutis, las curvas de entrenamiento y validación se mantuvieron relativamente cercanas y estables a partir de la época 20, lo cual indicó un aprendizaje consistente y bajo riesgo de sobreajuste. El error convergió hacia valores bajos sin oscilaciones pronunciadas, sugiriendo que las características locales fueron bien representadas por el modelo.

En Buenaventura, aunque las curvas también descendieron rápidamente, se observan picos notables entre las épocas 20 y 40 en la curva de validación. Estos repuntes indican posibles dificultades del modelo para generalizar en ciertos segmentos temporales, probablemente debido a mayor variabilidad o ruido en la serie de nivel del mar en esta localidad. No obstante, tras la época 40, la curva de validación retomó una trayectoria decreciente y se estabilizó, lo que indica que el entrenamiento logró adaptarse y evitar un sobreajuste crítico.

Por su parte, en Tumaco, las curvas de pérdida mostraron una disminución progresiva pero mantuvieron una ligera separación constante entre entrenamiento y validación durante casi todo el proceso. Aunque esta brecha es moderada, sugiere que existe cierta complejidad en la señal local que el modelo no logra capturar completamente, lo cual se ve reflejado en el valor más bajo de R^2 obtenido en esta estación. Sin embargo, no se observaron incrementos abruptos en el error de validación, lo que confirma que, pese a esas diferencias, el modelo mantuvo estabilidad durante el aprendizaje en todas las localidades.

Estos resultados evidenciaron que el uso exclusivo del primer componente canónico permite capturar tendencias generales del nivel del mar, aunque con limitaciones para reproducir detalles locales específicos en algunas zonas.

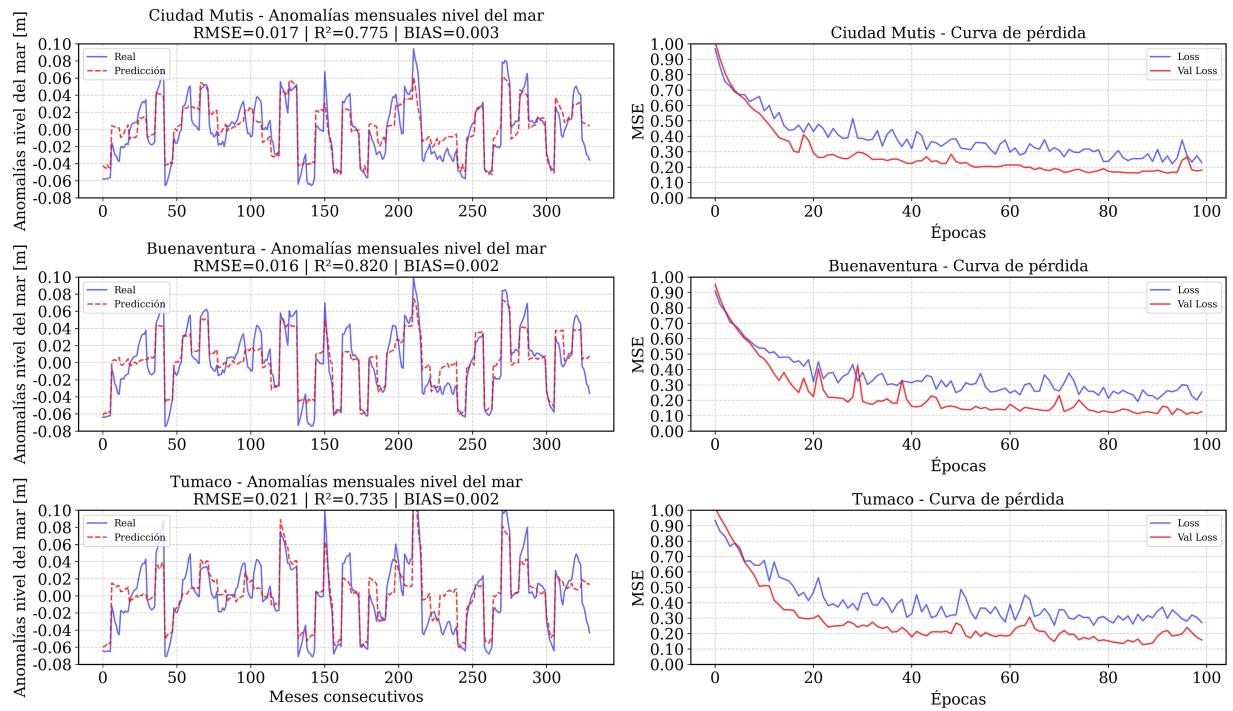


Figura 5.1: Resultados del Modelo 1 (LSTM entrenado con el primer componente canónico completo). A la izquierda, anomalías mensuales predichas y observadas; a la derecha, curvas de pérdida por población.

5.2.1.2. Modelo 2: LSTM entrenado con el primer componente canónico de variables seleccionadas

En el segundo enfoque, se seleccionaron únicamente aquellas variables cuyo peso absoluto en el primer componente canónico del análisis de correlación canónica (rCCA) superó el umbral de 0.3. Con estas variables dominantes se reconstruyó el componente canónico local, que se utilizó como único predictor de entrada para el modelo LSTM. Esta estrategia buscó reducir la dimensionalidad, limitando la entrada a las variables con mayor poder explicativo, y evitar la inclusión de predictores con escasa relevancia local.

La Figura 5.2 presenta los resultados obtenidos por este modelo para cada localidad. Los valores del coeficiente de determinación R^2 oscilaron entre 0.88 y 0.89, mientras que los RMSE se situaron

entre 0.012 m y 0.014 m. Tumaco alcanzó el mejor desempeño, con $R^2 = 0.89$ y RMSE de 0.014 m, lo que indica una capacidad relativamente alta para reproducir las oscilaciones del nivel del mar a partir de las variables más relevantes. En contraste, Ciudad Mutis presentó la menor capacidad explicativa, con $R^2 = 0.88$ y un RMSE de 0.013 m, mostrando que la selección de variables dominantes podría no haber capturado completamente la variabilidad local en esta zona. En todas las localidades, el sesgo (Bias) permaneció cercano a cero, lo cual refleja la ausencia de errores sistemáticos en las predicciones.

En cuanto a las curvas de pérdida (`loss` y `val_loss`) del Modelo 2, se observó un descenso pronunciado del error durante las primeras 20 épocas en las tres estaciones, indicando que el modelo logró capturar patrones relevantes de las series temporales de forma eficiente. Posteriormente, las trayectorias de las curvas se estabilizaron, mostrando convergencia hacia valores bajos sin incrementos abruptos en el error de validación, lo que evidencia un entrenamiento robusto y bajo riesgo de sobreajuste.

En Ciudad Mutis, las curvas de entrenamiento y validación se mantuvieron cercanas entre sí a partir de la época 20, reflejando un ajuste consistente y estable, en línea con el buen desempeño observado en las métricas (RMSE = 0.012 m, $R^2 = 0.883$). En Buenaventura, se evidenció un comportamiento similar, aunque la curva de validación se situó ligeramente por debajo de la de entrenamiento a lo largo de las épocas, lo cual puede sugerir un ligero efecto de regularización, pero sin señales de sobreajuste. Las métricas alcanzadas fueron muy satisfactorias (RMSE = 0.013 m, $R^2 = 0.888$).

Por su parte, en Tumaco, aunque el descenso inicial fue igualmente marcado, se mantuvo una pequeña separación entre las curvas de entrenamiento y validación durante casi todo el proceso. No obstante, esta brecha fue menor que en el Modelo 1 y el error convergió a valores estables, logrando un buen ajuste final (RMSE = 0.014 m, $R^2 = 0.890$).

En las curvas de pérdida del Modelo 2, se reflejó un comportamiento más equilibrado y estable que en el Modelo 1, confirmando la efectividad de emplear únicamente las variables de mayor peso del primer componente canónico en la construcción del predictor.

5.2.1.3. Modelo 3: LSTM entrenado con series de tiempo multivariadas

Finalmente, el Modelo 3 que fue entrenado directamente sobre las series de tiempo de las anomalías estandarizadas de las cinco variables ambientales. En este enfoque, no se incluyeron los índices ENOS como predictores directos.

La Figura 5.3 presenta los resultados obtenidos para este modelo en cada población. El desempeño del Modelo 3 fue claramente superior en términos globales, con valores del coeficiente de determinación R^2 que superaron el 0.92 en todas las ubicaciones. En particular, Buenaventura alcanzó el mayor ajuste, con $R^2 = 0.96$ y un RMSE de apenas 0.008 m. Ciudad Mutis obtuvo $R^2 = 0.94$ con un RMSE de 0.009 m, mientras que Tumaco alcanzó $R^2 = 0.93$ y un RMSE de 0.011 m. El sesgo (Bias) fue prácticamente nulo en todas las localidades, lo que refleja la ausencia de errores sistemáticos en las predicciones.

En cuanto a las curvas de pérdida (`loss` y `val_loss`) del Modelo 3, se observó un descenso rápido y sostenido del error durante las primeras 20 a 30 épocas en las tres estaciones analizadas, lo

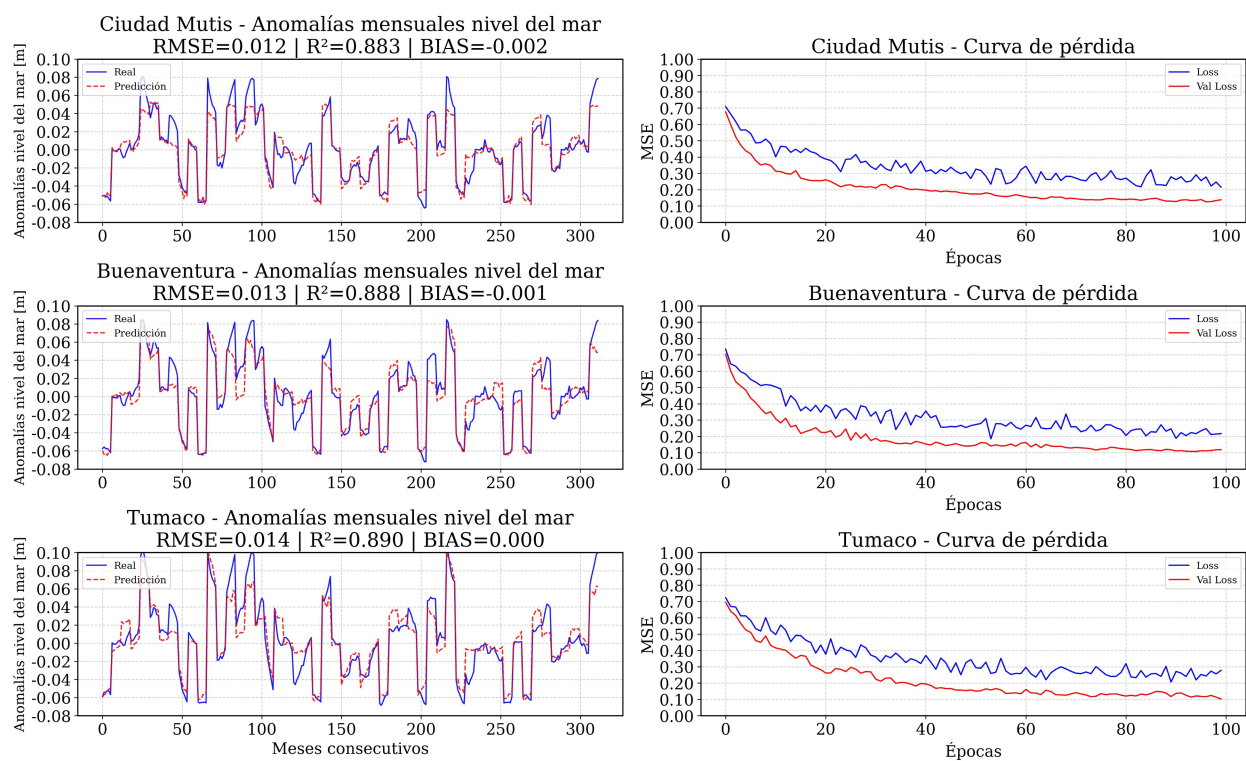


Figura 5.2: Resultados del Modelo 2 (LSTM entrenado con el primer componente canónico de variables seleccionadas). A la izquierda, anomalías mensuales predichas y observadas; a la derecha, curvas de pérdida por población.

cual refleja que la red neuronal fue capaz de aprender patrones robustos de la dinámica del nivel del mar a partir de las variables climáticas originales. Posteriormente, ambas curvas convergieron de manera suave hacia valores bajos y estables, sin evidencias de incrementos abruptos ni oscilaciones marcadas en el error de validación, lo que indica un entrenamiento altamente estable y libre de sobreajuste.

En Ciudad Mutis, las curvas de entrenamiento y validación se mantuvieron prácticamente superpuestas tras la fase inicial, señalando un ajuste muy sólido. En Bahía Buenaventura, se obtuvo el mejor desempeño global. Las curvas de pérdida presentaron la mayor estabilidad entre todas las estaciones, manteniendo `val_loss` consistentemente por debajo de la curva de entrenamiento. Finalmente, en Tumaco, las curvas de pérdida mostraron una caída inicial marcada y luego una fase de estabilidad. Si bien la señal predicha tiende a suavizar algunos picos extremos, el ajuste general fue notablemente bueno, sin indicios de sobreajuste ni errores sistemáticos.

En conjunto, las curvas de pérdida del Modelo 3 reflejan un comportamiento más estable, preciso y consistente que los modelos anteriores. Este excelente ajuste se tradujo en valores de sesgo prácticamente nulos en las tres localidades, lo que confirma la robustez de la arquitectura LSTM entrenada sobre las series originales de variables climáticas.

Comparativamente, el Modelo 3 ofreció los mejores resultados, tanto en términos de ajuste (R^2) como de precisión absoluta (RMSE). Aunque el Modelo 1 logró capturar las tendencias generales a partir del primer componente canónico, su capacidad para reproducir la variabilidad local fue más limitada. El Modelo 2 representó una mejora respecto al Modelo 1 al excluir variables poco informativas, pero sin alcanzar la precisión obtenida con las series originales multivariadas.

La Tabla 5.1 resume las métricas obtenidas para cada modelo y estación. En función de estos resultados, se seleccionó el Modelo 3 como el mejor candidato para generar los pronósticos y, posteriormente, la definición de los escenarios.

Tabla 5.1: Comparación de métricas de desempeño entre modelos LSTM por estación.

Modelo	Estación	RMSE (m)	Bias (m)	R^2
Modelo 1	Ciudad Mutis	0.017	0.003	0.775
	Buenaventura	0.016	0.002	0.820
	Tumaco	0.021	0.002	0.735
Modelo 2	Ciudad Mutis	0.012	-0.002	0.883
	Buenaventura	0.013	-0.001	0.888
	Tumaco	0.014	0.000	0.890
Modelo 3	Ciudad Mutis	0.009	0.001	0.936
	Buenaventura	0.008	-0.000	0.958
	Tumaco	0.011	0.000	0.927

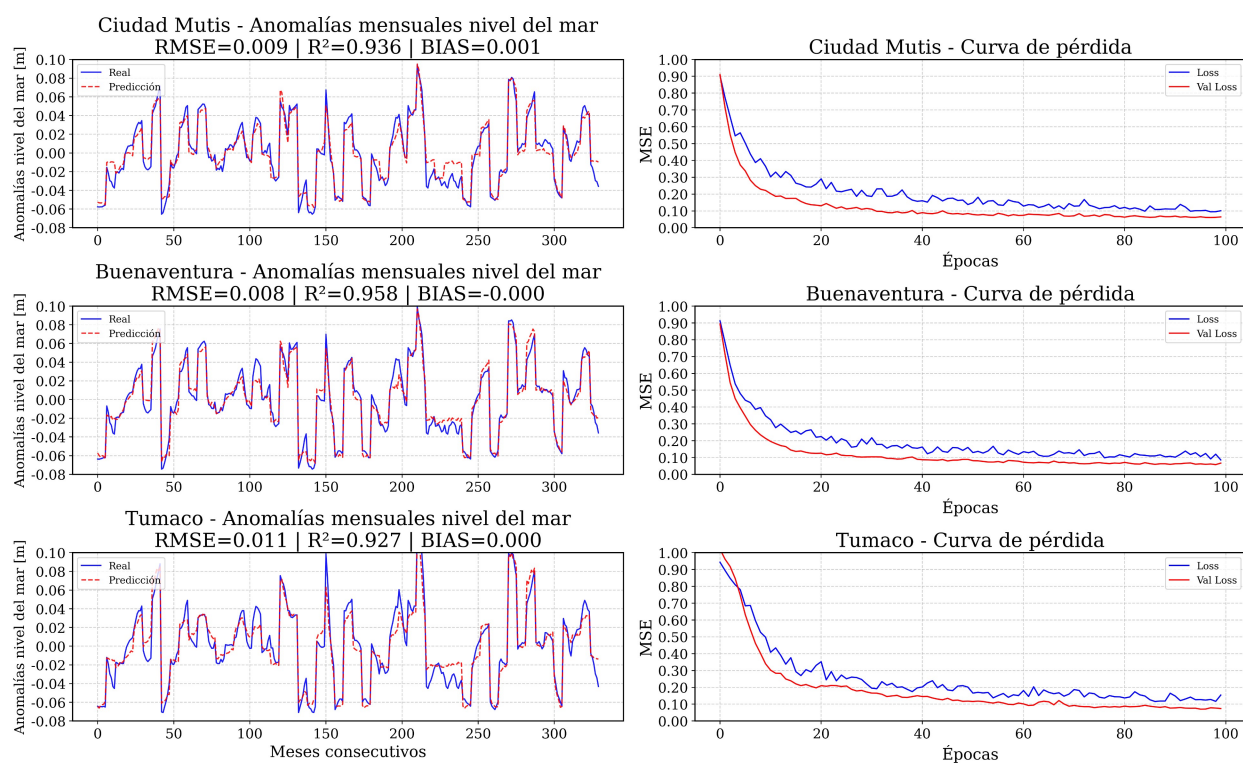


Figura 5.3: Resultados del Modelo 3 (LSTM entrenado con series de tiempo multivariadas). A la izquierda, anomalías mensuales predichas y observadas; a la derecha, curvas de pérdida por población.

5.2.2. Validación de los modelos digitales de terreno

Para evaluar el desempeño de los modelos determinísticos y probabilísticos aplicados en el análisis de elevaciones, se implementó un esquema de validación cruzada tipo k -fold con $k = 5$, lo que permitió una estimación robusta de la capacidad predictiva de cada método. En cada iteración, los datos se dividieron en cinco subconjuntos, utilizando cuatro de ellos para el entrenamiento del modelo y uno para la validación, rotando sucesivamente hasta cubrir todos los conjuntos. A partir de este procedimiento, se calcularon los estadísticos de evaluación correspondientes a cada partición, y se reportaron los valores promedio del coeficiente de determinación ($\overline{R^2}$), la raíz del error cuadrático medio (\overline{RMSE}) y el error absoluto medio (\overline{MAE}). Estos indicadores permitieron comparar de manera objetiva la precisión y ajuste de los diferentes enfoques de interpolación espacial.

5.2.2.1. Ciudad Mutis

Modelos probabilísticos

Los resultados obtenidos para los modelos probabilísticos aplicados en Ciudad Mutis (Tabla 5.2) muestran un rendimiento destacado de los modelos *exponential*, *spherical* y *power*, especialmente este último con exponentes ajustados. El modelo *power* con exponente 1 obtuvo el mejor desempeño global, con un \overline{RMSE} de 8.94 y un \overline{MAE} de 3.38, además del coeficiente de determinación promedio más alto ($\overline{R^2} = 0.96$).

Tabla 5.2: Resultados de los Modelos Probabilísticos para Ciudad Mutis.

Modelo	\overline{RMSE}	\overline{MAE}	$\overline{R^2}$
<i>spherical</i>	9.65	3.43	0.95
<i>exponential</i>	9.07	3.37	0.95
<i>power</i> exp = 0.08	9.27	3.51	0.96
<i>power</i> exp = 1	8.94	3.38	0.96

Muy cerca se ubicó la variante *power* con exponente 0.08, que alcanzó un \overline{RMSE} de 9.27 y un \overline{MAE} de 3.51, también con un elevado $\overline{R^2} = 0.96$. Los modelos *exponential* y *spherical* también presentaron un buen ajuste, con valores similares de \overline{RMSE} (9.07 y 9.65) y \overline{MAE} (3.37 y 3.43), y un coeficiente de determinación ligeramente inferior ($\overline{R^2} = 0.95$), lo que indica una capacidad sólida para representar la variabilidad espacial de las elevaciones (Tabla 5.2).

La figura 5.4 muestra la distribución de los errores en la elevación que se estimó a partir de cada uno de los ajustes que se realizaron con el *kriging*. Se observó que los modelos esférico y exponencial presentaron rangos de error más amplios (5 a 45), lo cual sugiere una mayor incertidumbre en la estimación de la elevación en ciertas zonas del área de estudio; y un marcado sobreajuste espacial de los errores. En particular, estos modelos mostraron concentraciones de errores más altos en regiones periféricas, posiblemente asociadas a una menor densidad de puntos de control o a una mayor heterogeneidad del terreno. Por otro lado, los modelos de potencia, tanto con exponente 0.8 como con exponente 1, presentaron errores más contenidos (5 a 35), lo que indica una estimación

más homogénea y con menor variabilidad espacial y una predicción de la elevación más allá de los puntos. En estos casos, los errores se distribuyeron de manera más uniforme, sin concentraciones marcadas, lo que podría reflejar una mejor adaptación del modelo a la estructura espacial de los datos (Figura 5.4).

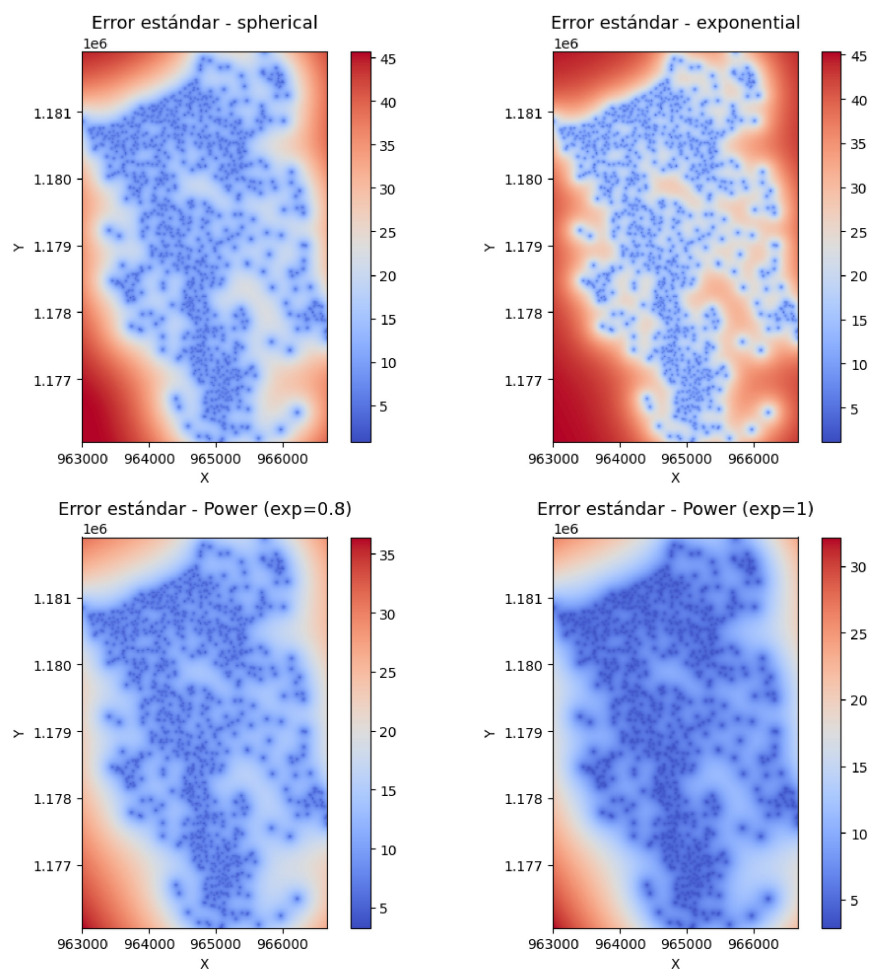


Figura 5.4: Distribución espacial de los errores promedio obtenidos por los modelos, exponenciales, esféricos y potencia del *kriging* para el casco urbano de Ciudad Mutis.

Modelos determinísticos

Los resultados obtenidos para los modelos determinísticos aplicados a la interpolación de elevaciones en Ciudad Mutis evidenciaron un rendimiento excepcional del modelo *lineal*, el cual superó ampliamente a las demás alternativas en todos los indicadores de evaluación. Este modelo registró la menor raíz del error cuadrático medio promedio ($\overline{\text{RMSE}} = 0.148$), con una desviación estándar muy baja de 0.002, lo que indica una alta precisión y estabilidad en sus predicciones a lo largo de los

diferentes subconjuntos de validación. Asimismo, presentó el menor error absoluto medio promedio ($\overline{MAE} = 0.059$), con una desviación estándar prácticamente nula (0.00008), reforzando su consistencia en la estimación de los valores reales. El coeficiente de determinación promedio ($\overline{R^2} = 0.99999$) fue el más alto entre los modelos evaluados, con una desviación estándar insignificante, lo cual refleja una capacidad casi perfecta para explicar la variabilidad espacial de las elevaciones.

En comparación, los modelos *cúbico* y de *vecino más cercano* también mostraron desempeños satisfactorios, aunque ligeramente inferiores. El modelo cúbico obtuvo un \overline{RMSE} de 0.166 y un \overline{MAE} de 0.061, mientras que el modelo de vecino más cercano alcanzó un \overline{RMSE} de 0.214 y un \overline{MAE} de 0.141. Ambos modelos mantuvieron valores altos de coeficiente de determinación, con $\overline{R^2}$ de 0.99998 y 0.99873 respectivamente, lo que indica un buen ajuste general, aunque con una ligera pérdida de precisión en comparación con el modelo lineal.

En términos generales, los tres enfoques determinísticos mostraron una excelente capacidad para reproducir la topografía de Ciudad Mutis. No obstante, el modelo lineal se destacó por su precisión superior, su consistencia entre los diferentes *folds* de validación y su casi perfecta representación de la variabilidad espacial de los datos (Tabla 5.3). La distribución homogénea de los errores entre los diferentes *folds* sugiere, además, una alta estabilidad y robustez en las predicciones de este modelo.

Tabla 5.3: Resultados de los Modelos Determinísticos para Ciudad Mutis.

Modelo	\overline{RMSE}	σ_{RMSE}	\overline{MAE}	σ_{MAE}	$\overline{R^2}$	σ_{R^2}
Linear	0.148	0.002	0.059	0.00008	0.99999	3.3×10^{-7}
Cubic	0.166	0.002	0.061	0.00004	0.99998	4.0×10^{-7}
Nearest	0.214	0.001	0.141	0.00017	0.99873	1.1×10^{-5}

5.2.2.2. Buenaventura

Modelos probabilísticos

Los resultados obtenidos para los modelos probabilísticos aplicados a la interpolación de elevaciones en Buenaventura (Tabla 5.4) evidencian un desempeño muy similar entre las distintas configuraciones evaluadas, con una ligera ventaja de los modelos *spherical* y *exponential*. Ambos modelos presentaron los errores más bajos, con una raíz del error cuadrático medio promedio (\overline{RMSE}) de 2.52 y un error absoluto medio promedio (\overline{MAE}) de 1.32. Además, alcanzaron un coeficiente de determinación promedio ($\overline{R^2}$) de 0.81, lo que indica una capacidad adecuada para capturar la variabilidad espacial de las elevaciones en la zona de estudio (Tabla 5.4).

Por su parte, los modelos *power* (Figura 5.6) mostraron un desempeño ligeramente inferior. La variante con exponente 1.3 obtuvo un \overline{RMSE} de 2.57, un \overline{MAE} de 1.33 y un $\overline{R^2}$ de 0.80, mientras que el modelo con exponente 1 presentó el mayor error entre los cuatro enfoques analizados, con un \overline{RMSE} de 2.61, un \overline{MAE} de 1.34 y el menor coeficiente de determinación ($\overline{R^2} = 0.79$) (Tabla 5.4).

La figura 5.6 presenta los mapas de error estándar resultantes de los diferentes modelos probabilísticos de kriging aplicados a la interpolación de elevaciones en Buenaventura. En general, los mapas exhiben una distribución espacial homogénea del error en la mayor parte del dominio, con

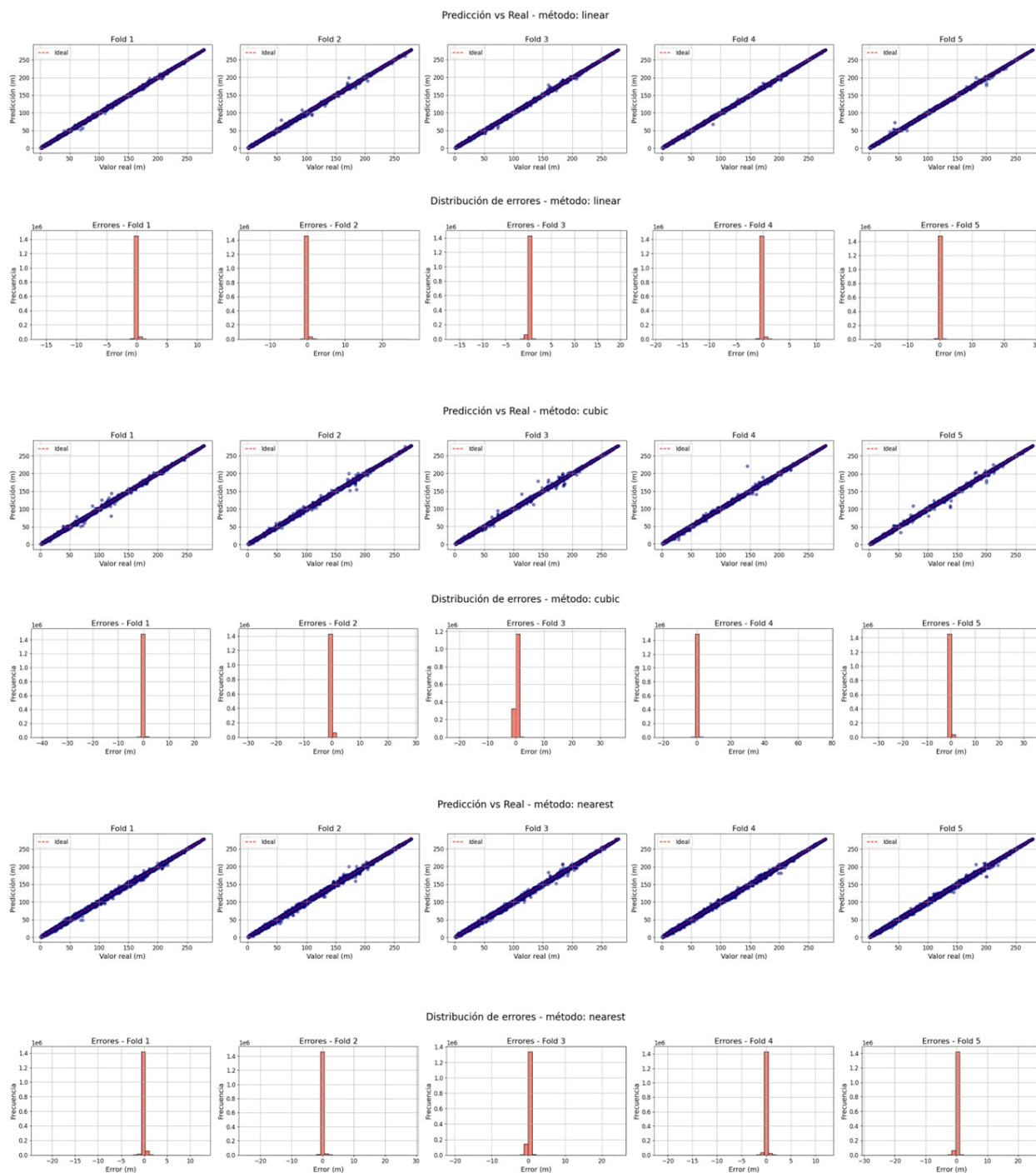


Figura 5.5: Desempeño de los modelos de interpolación lineal, cúbica y de vecino más cercano de la elevación del casco de Ciudad Mutis: gráficos de elevación predicha vs elevación real y Distribución de errores en validación cruzada (5 *Folds*).

Tabla 5.4: Resultados de los Modelos Probabilísticos para Buenaventura.

Modelo	$\overline{\text{RMSE}}$	$\overline{\text{MAE}}$	$\overline{R^2}$
<i>spherical</i>	2.52	1.32	0.81
<i>exponential</i>	2.52	1.32	0.81
<i>power</i> exp = 1.3	2.57	1.33	0.80
<i>power</i> exp = 1	2.61	1.34	0.79

valores bajos de error estándar en la zona central, donde se concentra la mayor densidad de puntos de observación. Por el contrario, los bordes del área —especialmente hacia el sureste— muestran un incremento notable del error estándar en todos los modelos, lo que es consistente con la menor densidad de datos en esas zonas y con el efecto de borde común en los procesos de interpolación.

El modelo *power* con exponente 1.3 muestra la distribución más uniforme y con menor magnitud general de error, mientras que el modelo *power* con exponente 1.5 alcanza los valores máximos de error estándar, particularmente hacia las esquinas del dominio. Los modelos *spherical* y *exponential* presentan patrones similares, con errores moderadamente bajos y concentraciones de incertidumbre en los extremos.

Modelos determinísticos

El análisis de los resultados de los modelos determinísticos aplicados a la interpolación de elevaciones en Buenaventura indicó un rendimiento consistentemente alto en los tres enfoques evaluados, aunque con diferencias pequeñas en cuanto a precisión y estabilidad. El modelo *lineal* se destacó como el de mejor desempeño general, al registrar la raíz del error cuadrático medio promedio ($\overline{\text{RMSE}}$) más baja, con un valor de 0.169 y una desviación estándar de apenas 0.001, lo que sugiere una excelente precisión y una alta estabilidad entre los distintos subconjuntos de validación. Del mismo modo, presentó el menor error absoluto medio ($\overline{\text{MAE}} = 0.115$), acompañado de una desviación estándar muy baja (0.00014), lo cual refuerza su fiabilidad en las predicciones puntuales. Además, el coeficiente de determinación promedio ($\overline{R^2}$) fue de 0.99920, con una desviación estándar de 8.2×10^{-6} , indicando una capacidad sobresaliente para explicar la variabilidad espacial de las elevaciones con un ajuste prácticamente perfecto.

En comparación, los modelos *cúbico* y de *vecino más cercano* también ofrecieron resultados muy satisfactorios, aunque con un desempeño ligeramente inferior al modelo lineal. El modelo cúbico presentó un $\overline{\text{RMSE}}$ de 0.225 y un $\overline{\text{MAE}}$ de 0.139, con una desviación estándar de 0.006 y 0.00021 respectivamente, lo que sugiere una mayor variabilidad en el error entre folds. Por su parte, el modelo de vecino más cercano logró un $\overline{\text{RMSE}}$ de 0.214 y un $\overline{\text{MAE}}$ de 0.141, con desviaciones estándar de 0.001 y 0.00017, mostrando un comportamiento más estable que el cúbico pero ligeramente menos preciso. Aunque ambos modelos mantuvieron valores elevados de coeficiente de determinación promedio ($\overline{R^2} = 0.99859$ para el cúbico y 0.99873 para el vecino más cercano), estos fueron marginalmente inferiores al del modelo lineal.

La figura 5.7 muestra los resultados del proceso de validación cruzada (k-fold con k=5) aplicado a

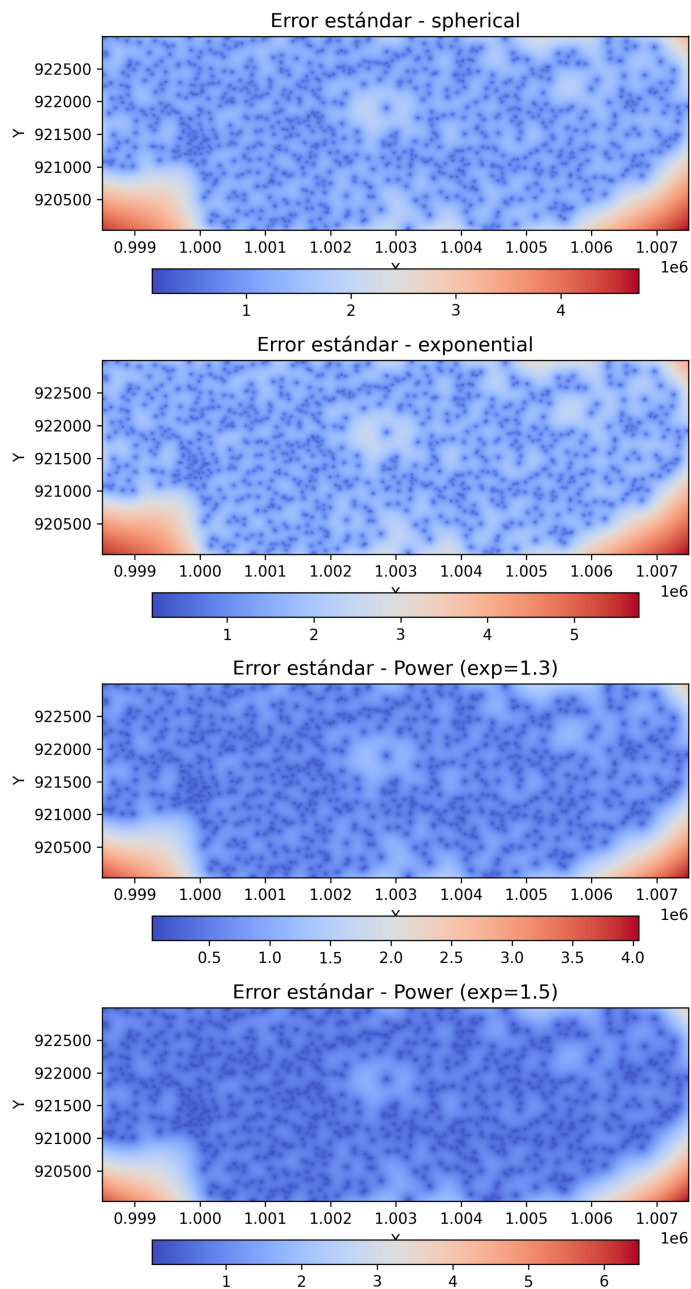


Figura 5.6: Distribución espacial de los errores promedio obtenidos por los modelos, exponenciales, esféricos, lineal y potencia del *kriging* para el casco urbano de Buenaventura.

Tabla 5.5: Resultados de los Modelos Determinísticos para Buenaventura.

Modelo	\overline{RMSE}	σ_{RMSE}	\overline{MAE}	σ_{MAE}	$\overline{R^2}$	σ_{R^2}
Linear	0.169	0.001	0.115	0.00014	0.99920	8.2×10^{-6}
Cubic	0.225	0.006	0.139	0.00021	0.99859	7.8×10^{-5}
Nearest	0.214	0.001	0.141	0.00017	0.99873	1.1×10^{-5}

los modelos determinísticos para la interpolación de elevaciones en Buenaventura. Para cada modelo —*linear*, *cubic* y *nearest* se presentan dos tipos de visualizaciones por fold: la comparación entre valores predichos y reales, y la distribución de los errores de predicción.

En las gráficas de predicción vs. valor real (filas impares), se observa un fuerte alineamiento de los puntos a lo largo de la diagonal en los tres modelos, lo que evidencia un buen ajuste general en todos los folds. Este comportamiento es especialmente notable en el modelo lineal, donde la nube de puntos es más compacta y alineada, reflejando menor dispersión y mayor precisión en las estimaciones. El modelo cúbico presenta ligeras desviaciones y algunos valores atípicos, mientras que el modelo de vecino más cercano muestra un patrón similar, aunque con mayor concentración de puntos alrededor de la línea de identidad.

Las gráficas de distribución de errores (filas pares) complementan esta evaluación mostrando que, en todos los modelos, los errores se distribuyen de manera estrecha alrededor del cero, con sesgos mínimos y colas poco pronunciadas. Nuevamente, el modelo lineal se destaca por una mayor simetría y menor amplitud de errores, lo que refuerza su estabilidad y consistencia entre los folds. Los modelos cúbico y de vecino más cercano, aunque también presentan distribuciones centradas, muestran colas ligeramente más extendidas y una mayor variabilidad en la dispersión de los errores.

5.2.2.3. Tumaco

Modelos probabilísticos

El análisis de los resultados obtenidos para los modelos probabilísticos aplicados a la interpolación de elevaciones en Tumaco (Tabla 5.6) evidenció un rendimiento más moderado en comparación con los resultados obtenidos en los cascos urbanos de Ciudad Mutis y Buenaventura. El modelo *exponential* se destacó como el de mejor desempeño dentro del conjunto evaluado, alcanzando la menor raíz del error cuadrático medio promedio ($\overline{RMSE} = 1.00$) y el menor error absoluto medio ($\overline{MAE} = 0.62$), acompañado de un coeficiente de determinación aceptable ($\overline{R^2} = 0.58$), lo cual sugiere una capacidad razonable para representar la variabilidad espacial de las elevaciones en esta zona del litoral.

El modelo *power*, con exponentes bajos (0.05 y 0.1), arrojó resultados aceptables, especialmente la variante con exponente 0.1, que logró un \overline{RMSE} de 1.07, un \overline{MAE} de 0.69 y un $\overline{R^2}$ de 0.52, acercándose al rendimiento del modelo *exponential*. En contraste, el modelo *power* con exponente 0.05 mostró un ajuste menos preciso, con un \overline{RMSE} de 1.14, un \overline{MAE} de 0.77 y un coeficiente de determinación de 0.50.

Por su parte, el modelo *spherical* obtuvo resultados similares a los del modelo *power* con expo-

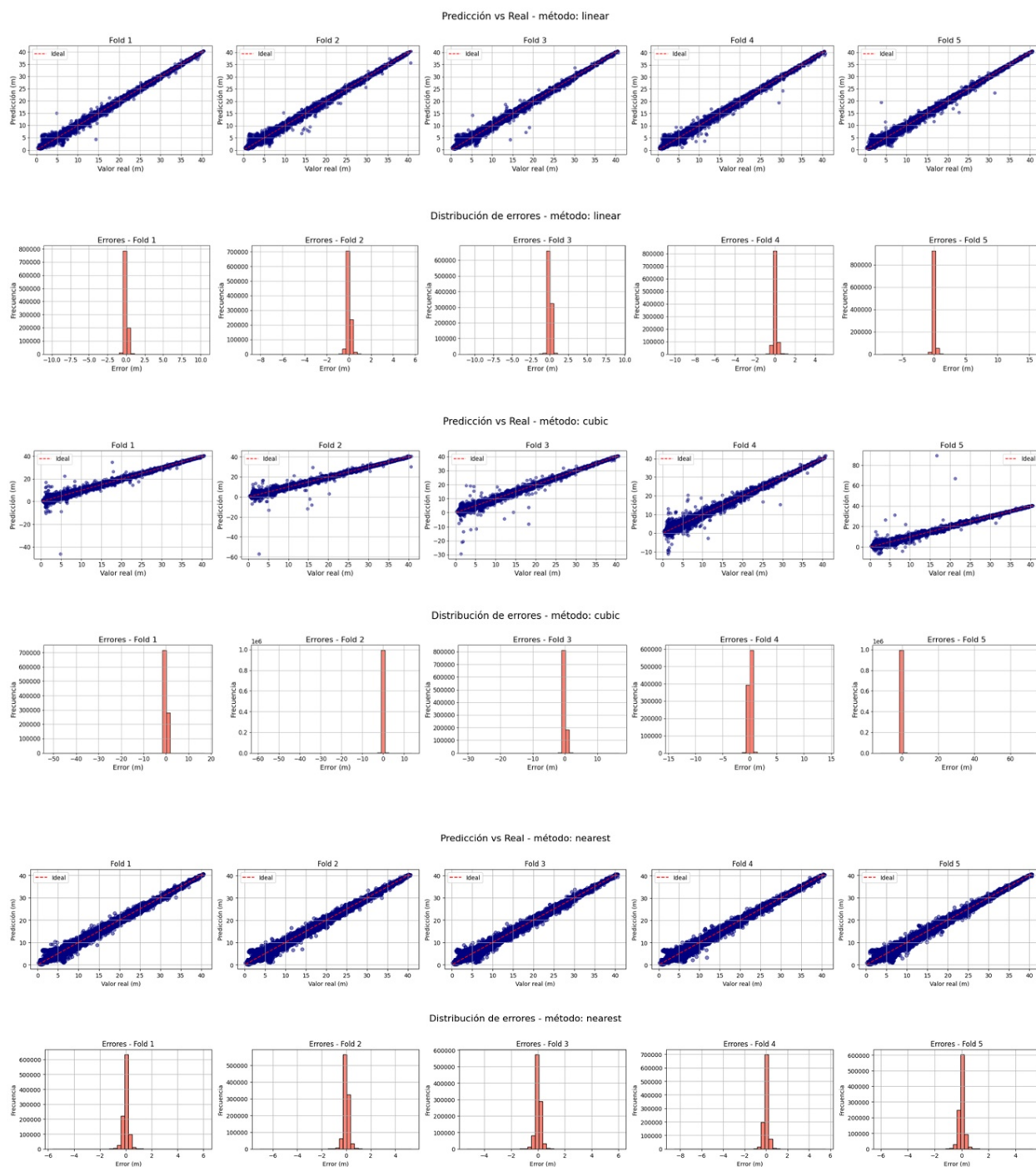


Figura 5.7: Desempeño de los modelos de interpolación lineal, cúbica y de vecino más cercano de la elevación del casco urbano de Buenaventura: gráficos de elevación predicha vs elevación real y Distribución de errores en validación cruzada (5 *Folds*).

nente 0.1, con un \overline{RMSE} de 1.10, un \overline{MAE} de 0.71 y un $\overline{R^2}$ también de 0.50, lo que indica un ajuste limitado a la estructura espacial de los datos. Finalmente, el modelo *linear* fue el que presentó el peor desempeño general, con los valores de error más altos ($\overline{RMSE} = 1.18$, $\overline{MAE} = 0.83$) y el menor coeficiente de determinación ($\overline{R^2} = 0.44$), reflejando una capacidad reducida para capturar la complejidad del terreno en el área de Tumaco.

Tabla 5.6: Resultados de los Modelos Probabilísticos para Tumaco.

Modelo	\overline{RMSE}	\overline{MAE}	$\overline{R^2}$
<i>spherical</i>	1.1	0.71	0.50
<i>exponential</i>	1.00	0.62	0.58
<i>power</i> exp = 0.05	1.14	0.77	0.5
<i>power</i> exp = 0.1	1.07	0.69	0.52

La figura 5.8 muestra la distribución espacial del error estándar estimado mediante los modelos probabilísticos de kriging aplicados a la interpolación de elevaciones en el área de Tumaco. Se presentan los resultados para cuatro modelos: *spherical*, *exponential* y dos variantes del modelo *power*, con exponentes de 0.05 y 0.1.

En general, se observa un patrón claro de incremento del error estándar en las zonas periféricas del dominio, particularmente hacia las esquinas y los bordes norte y suroeste. Esto sugiere una menor densidad de puntos de observación en estas regiones, lo que genera mayor incertidumbre en la estimación de elevaciones. En contraste, en el área central —donde la concentración de datos es mayor— todos los modelos presentan errores estándar más bajos, con superficies de interpolación más confiables.

El modelo *exponential* es el que muestra la menor magnitud global del error, con valores más uniformes y una transición más suave entre zonas de alta y baja incertidumbre, lo cual coincide con sus buenos indicadores estadísticos presentados en la Tabla 5.6. El modelo *spherical* presenta un patrón similar pero con errores ligeramente más altos y una estructura más segmentada, lo que sugiere un ajuste más local y menos generalizado.

Por otro lado, los modelos *power* con exponentes bajos presentan errores estándar considerablemente mayores en todo el dominio. En particular, el modelo con exponente 0.05 exhibe una superficie de error más amplia y homogéneamente elevada, indicando una mayor sensibilidad al parámetro de potencia utilizado y un ajuste menos preciso en zonas con menor soporte de datos. La variante con exponente 0.1 muestra una mejora ligera, aunque sigue reflejando mayores niveles de incertidumbre que los modelos *exponential* y *spherical*.

Modelos determinísticos

El análisis de los resultados de los modelos determinísticos aplicados en la interpolación de elevaciones para Tumaco reveló un desempeño muy satisfactorio, sobre todo del modelo *linear*. Este modelo presentó la menor raíz del error cuadrático medio promedio ($\overline{RMSE} = 0.061$), acompañada de una desviación estándar muy reducida ($\sigma_{\overline{RMSE}} = 0.001$), lo cual refleja una notable precisión y

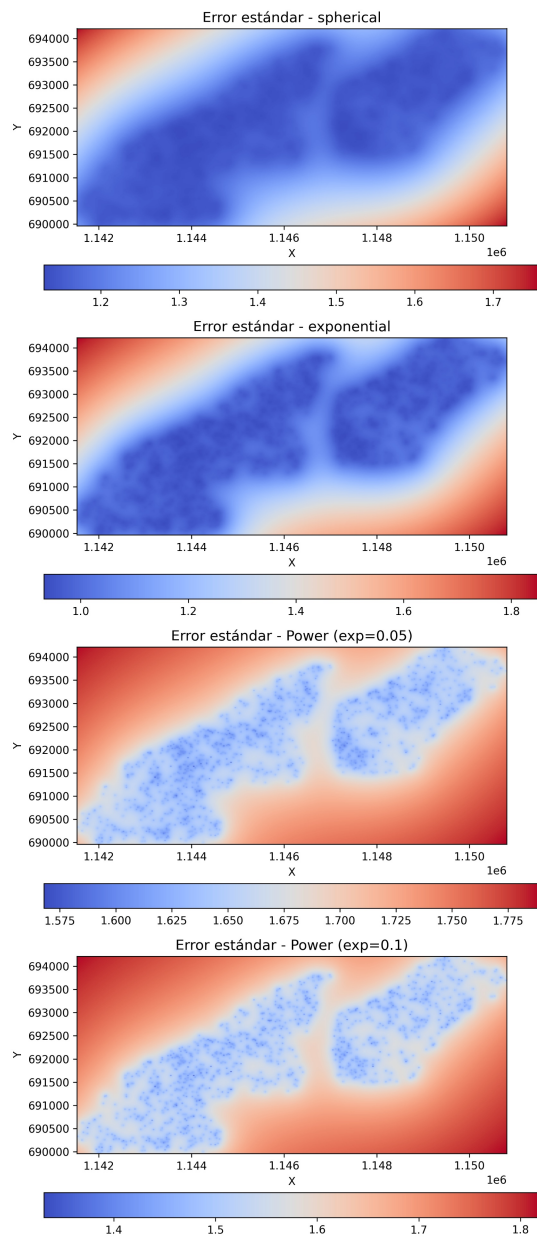


Figura 5.8: Distribución espacial de los errores promedio obtenidos por los modelos, exponenciales, esféricos, lineal y potencia del *kriging* para el casco urbano de Tumaco.

consistencia en sus estimaciones. Asimismo, el error absoluto medio promedio ($\overline{\text{MAE}}$) fue el más bajo entre los modelos, con un valor de 0.030 y una desviación estándar mínima de $4,0 \times 10^{-5}$. En cuanto a la capacidad explicativa del modelo, el coeficiente de determinación promedio ($\overline{R^2}$) alcanzó un valor sobresaliente de 0.99905, con una desviación estándar de apenas 1.3×10^{-5} , lo que indica una explicación casi total de la variabilidad espacial de los datos de elevación.

Los modelos *cubic* y *nearest* también mostraron un desempeño destacable, aunque ligeramente inferior al modelo lineal. El modelo cúbico obtuvo un $\overline{\text{RMSE}}$ de 0.073 y un $\overline{\text{MAE}}$ de 0.032, mientras que el modelo de vecino más cercano registró un $\overline{\text{RMSE}}$ de 0.079 y un $\overline{\text{MAE}}$ de 0.038. Ambos modelos conservaron valores elevados de coeficiente de determinación promedio (0.99861 para el cúbico y 0.99842 para el vecino más cercano), aunque sus desviaciones estándar fueron ligeramente mayores, particularmente en el modelo cúbico, cuya σ_{R^2} fue de 4.1×10^{-4} .

Tabla 5.7: Resultados de los Modelos Determinísticos para Tumaco.

Modelo	$\overline{\text{RMSE}}$	σ_{RMSE}	$\overline{\text{MAE}}$	σ_{MAE}	$\overline{R^2}$	σ_{R^2}
Linear	0.061	0.001	0.030	4.0×10^{-5}	0.99905	1.3×10^{-5}
Cubic	0.073	0.010	0.032	6.3×10^{-5}	0.99861	4.1×10^{-4}
Nearest	0.079	0.001	0.038	3.4×10^{-5}	0.99842	2.2×10^{-5}

La Figura 5.9 presenta los resultados del proceso de validación cruzada (*5-fold cross-validation*) aplicado a los modelos determinísticos evaluados para la interpolación de elevaciones en el área de Tumaco: *linear*, *cubic* y *nearest*.

En los diagramas de dispersión “Predicción vs Real”, se observa que los puntos se agrupan de forma consistente en torno a la línea de identidad (línea roja), lo cual indica una alta precisión en las estimaciones de los tres métodos. Esta tendencia es especialmente marcada en el modelo *linear*, donde la alineación es casi perfecta en todos los *folds*.

Las distribuciones de error, representadas como histogramas, revelan que los errores están centrados en torno a cero, con muy baja dispersión y sin evidencia de sesgos sistemáticos. En el caso del modelo *linear*, los histogramas son notablemente estrechos y concentrados, lo que confirma su estabilidad y baja variabilidad entre los *folds*. Para los modelos *cubic* y *nearest*, aunque también presentan buenas distribuciones centradas, se observa una ligera mayor dispersión, coherente con sus errores promedio levemente superiores reportados en la Tabla 5.7.

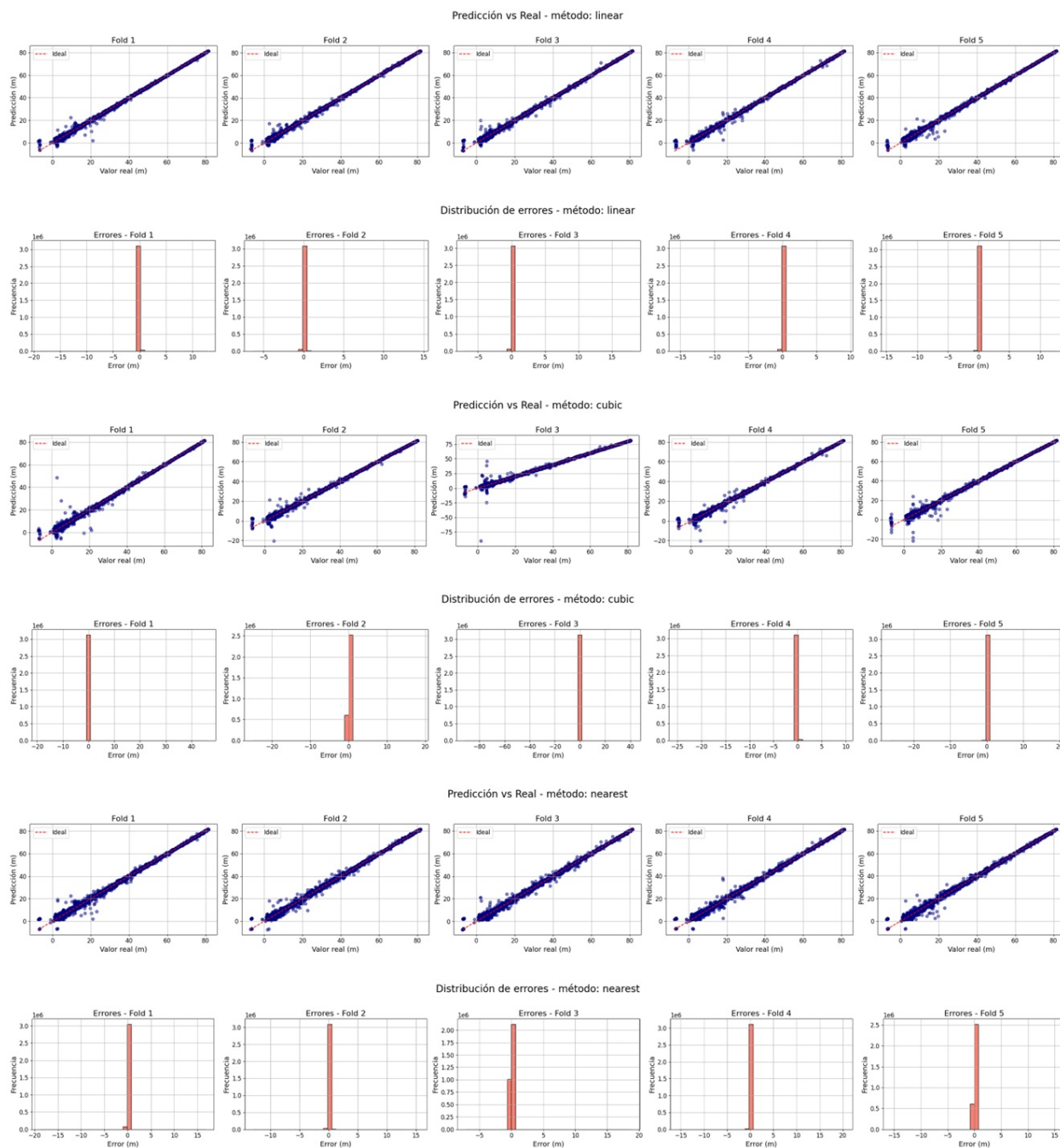


Figura 5.9: Desempeño de los modelos de interpolación lineal, cúbica y de vecino más cercano de la elevación del casco urbano de Tumaco: gráficos de elevación predicha vs elevación real y Distribución de errores en validación cruzada (5 *Folds*).

Resultados y Discusión

6.1. Contexto

Este capítulo presenta los principales hallazgos obtenidos a partir de la implementación del modelo *LSTM* y su aplicación a escenarios predictivos del nivel del mar inducido por el fenómeno El Niño-Oscilación del Sur (ENOS) en el litoral Pacífico colombiano. El análisis incluyó la estimación de cotas de inundación, las cuales se definieron como la suma de tres componentes: el *run-up* por oleaje, la cota de marea máxima y la sobreelevación generada por el ENOS. Esta integración permitió representar el nivel máximo que podría alcanzar el mar en condiciones extremas, con el objetivo de identificar zonas susceptibles a inundaciones costeras bajo diferentes manifestaciones del ENOS: Canónico, Modoki y Costero.

Se construyeron escenarios de elevación máxima a partir de estos tres forzantes oceanográficos y atmosféricos, considerando su efecto acumulado en el espacio y el tiempo. Para ello, se combinaron predicciones derivadas del modelo *LSTM* (que simula la respuesta del nivel del mar a condiciones climáticas multivariadas), observaciones históricas de nivel del mar (mareográficas) referidas al plano LAT y estimaciones de *run-up* obtenidas a partir de parámetros significativos del oleaje.

La integración de estos componentes se basó en datos históricos y de predicción, sin requerir coincidencia temporal exacta, dado que el propósito fue generar escenarios máximos de exposición costera. Los resultados se presentan mediante series de tiempo, mapas de inundación potencial y tablas comparativas por estación, los cuales se analizan en este capítulo.

6.2. Escenarios predictivos sobre la elevación del nivel del mar y su relación con el ENOS

En el contexto de esta investigación, la cota de inundación se define como la máxima elevación que puede alcanzar el nivel del mar en la línea de costa, resultado de la combinación de diversos forzantes oceánicos y atmosféricos. Este parámetro, de naturaleza aditiva, resulta fundamental para la evaluación de la exposición costera frente a escenarios extremos de ascenso del mar, y constituye un insumo técnico para la planificación territorial, la gestión del riesgo y el diseño de infraestructuras resilientes frente a la variabilidad climática [115], [116].

La metodología adoptada considera tres componentes principales:

- **Sobreelevación asociada al fenómeno ENOS:** representa el incremento del nivel medio del mar inducido por las condiciones oceánico-atmosféricas interanuales vinculadas al fenómeno El Niño-Oscilación del Sur. Su estimación se basó en la implementación de modelos de

aprendizaje profundo (red neuronal *LSTM*), entrenados para capturar la señal ENOS a partir de predictores climáticos multivariados (ver Sección 4.2.2).

- **Cota de marea observada:** corresponde a la altura máxima de la marea y condiciones meteorológicas normales, referida al plano vertical LAT (*Lowest Astronomical Tide*). Este valor se calculó a partir de registros históricos de nivel del mar en las estaciones mareográficas disponibles en las poblaciones de interés (ver Sección 3.2).
- **Run-up por oleaje:** expresa el ascenso del agua sobre la playa debido al impacto del oleaje. Se estimó mediante una fórmula empírica validada internacionalmente, que integra parámetros como la altura significativa del oleaje, el periodo medio espectral y la pendiente de la playa [117].

A continuación se describen los resultados obtenidos para estimar cada uno de estos componentes y su integración en escenarios predictivos diferenciados para los centros poblados de Ciudad Mutis, Buenaventura y Tumaco, considerando distintos tipos de eventos El Niño (Canónico, Costero y Modoki).

6.2.0.1. Determinación de la cota de marea observada

Dado que los registros del nivel del mar estaban referidos a distintos planos verticales (por ejemplo, al Nivel Medio del Mar o al cero instrumental), fue necesario nivelar todas las series a un plano de referencia común. Para ello, se adoptó el plano *LAT* (*Lowest Astronomical Tide*), ampliamente utilizado en aplicaciones hidrográficas, cartográficas y de planificación costera [118].

Una vez homogeneizadas, se construyeron las curvas de distribución acumulada de frecuencia para cada estación, y se seleccionó el percentil 90 (P90) como valor representativo de la cota máxima de marea. Este umbral fue adoptado con base en criterios de gestión del riesgo, dado que el P90 permite capturar los eventos de marea más altos sin ser distorsionado por errores instrumentales o valores atípicos. Su uso ha sido documentado en estudios previos sobre exposición costera [115], [119], [120].

Tabla 6.1: Cota de marea estimada a partir del percentil 90 % de la distribución acumulada en cada estación.

Estación	Cota de marea P90 (m)
Buenaventura	4.022
Ciudad Mutis	3.585
Tumaco	3.377

Los valores de cota de marea estimados evidenciaron una marcada variabilidad espacial entre las tres localidades analizadas. Buenaventura presentó la cota más elevada (4.02 m), seguida por Ciudad Mutis (3.59 m) y Tumaco (3.38 m). Esta diferencia puede explicarse por factores locales como la configuración geomorfológica de cada bahía, la batimetría costera y los efectos de resonancia

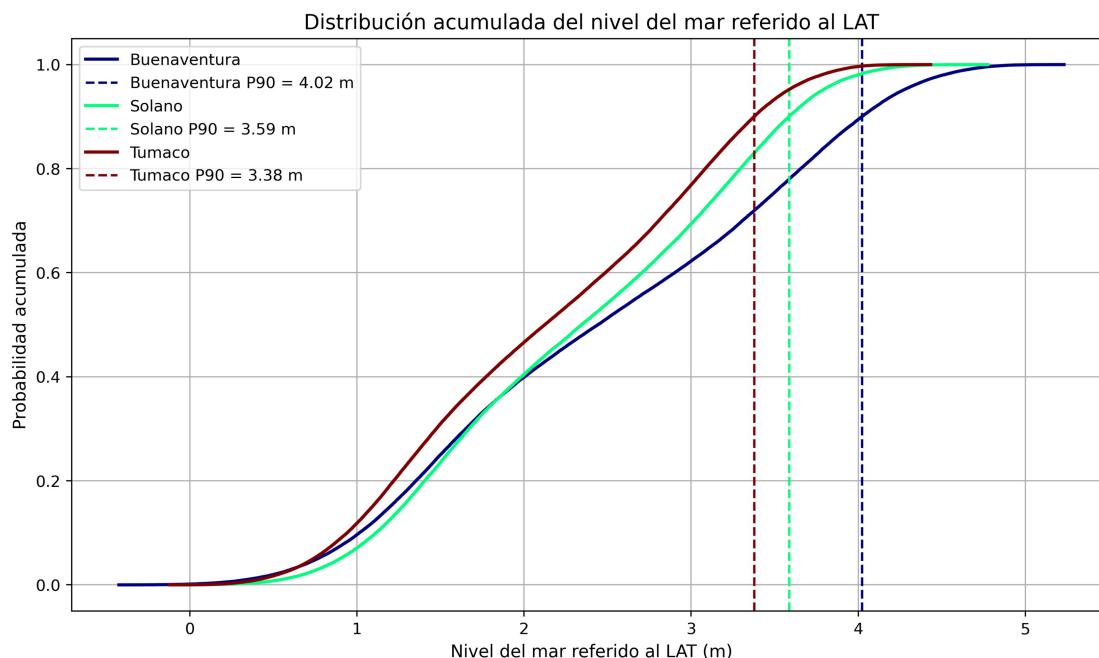


Figura 6.1: Distribución acumulada del nivel del mar referido al *LAT*. Se muestra el percentil 90 (*P90*) de cada población, utilizado como cota de marea máxima para escenarios de inundación.

interna, los cuales influyen significativamente en la respuesta hidrodinámica ante el forzante mareal [121], [122].

En el caso de Buenaventura, los valores obtenidos se encuentran en el rango de estimaciones previas. Por ejemplo, el estudio hidrodinámico de Otero et al. (2005) reportó alturas de marea cercanas a los 4.7 m para el 95 % de no excedencia (los valores estimados en esta investigación obedecen al 90 % de no excedencia), en una ubicación próxima al mareógrafo del puerto [123]. Este resultado es coherente con la configuración semiabierta de la bahía, donde se intensifican procesos de resonancia debidos a fenómenos de difracción y refracción, particularmente frente a la Isla de Buenaventura, que actúa como un dique natural amplificando las oscilaciones de largo período [124]. Esto refuerza la necesidad de considerar dicha amplificación en evaluaciones de riesgo costero.

Por su parte, la cota estimada en Tumaco (3.38 m) resultó inferior a los valores reportados por Gutiérrez y Puentes (2020), en el estero Aguaclara en Tumaco, quienes identificaron un nivel máximo de 3.93 m con una probabilidad de no excedencia del 99.9 % [125]. Esta diferencia puede atribuirse al hecho de que el valor aquí estimado se refiere al percentil 90.

En síntesis, los valores aquí obtenidos proporcionan una base coherente y conservadora para el cálculo de escenarios de inundación, manteniendo compatibilidad con estudios previos en el Pacífico colombiano.

6.2.0.2. Determinación del *Run-up*

El cálculo del *run-up* significativo se llevó a cabo utilizando la expresión empírica propuesta por Stockdon et al. [117], ampliamente aceptada en estudios de dinámica costera por su capacidad para integrar parámetros morfodinámicos y condiciones oceánicas de mar abierto. Esta formulación permitió estimar la altura del ascenso del agua en la zona de *swash*, excluyendo el sobrepaso. La ecuación empleada fue la siguiente:

$$R_{2\%} = 1.1 \left[0.35\beta_f(H_0L_0)^{1/2} + \frac{(H_0L_0(0.563\beta_f^2 + 0.004))^{1/2}}{2} \right] \quad (6.1)$$

donde:

- H_0 : altura significativa del oleaje en aguas profundas (m), obtenida de la variable VHM0 del producto Copernicus;
- L_0 : longitud de onda en aguas profundas (m), estimada como $L_0 = \frac{gT^2}{2\pi}$, siendo T el periodo medio espectral (VTM10) y $g = 9.81 \text{ m/s}^2$;
- β_f : pendiente de la playa en la zona de *swash*;
- $R_{2\%}$: altura excedida únicamente por el 2% de los eventos de *run-up*.

El cálculo se aplicó a cada punto espacial del dominio cubierto por los datos de oleaje. No obstante, dadas las limitaciones en la resolución espacial de estos productos cerca de la línea costera, se seleccionó el valor correspondiente al punto más cercano a la estación mareográfica. Este valor constituye una aproximación razonable a las condiciones locales, aunque no representa con total fidelidad la batimetría y morfología específicas del frente costero.

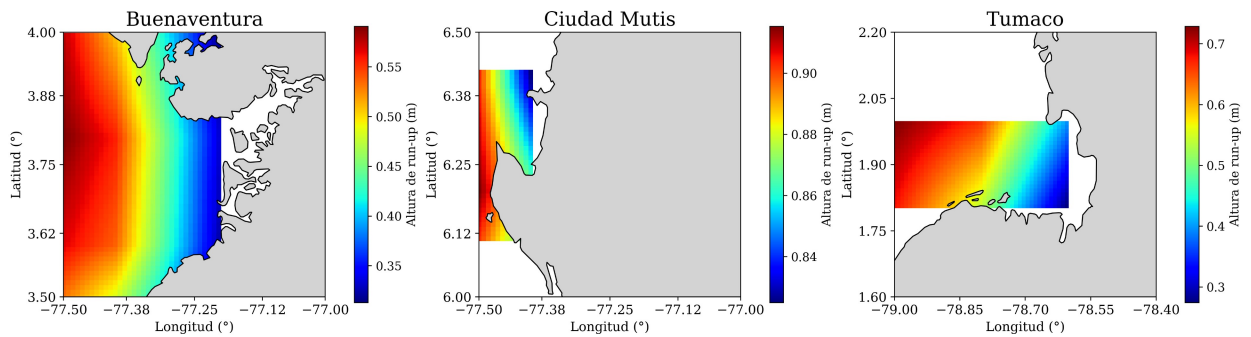


Figura 6.2: Altura del run-up por oleaje del percentil 98 para Ciudad Mutis, Buenaventura y Tumaco.

Los valores estimados del *run-up* para cada localidad fueron:

Tabla 6.2: Altura del *run-up* significativo (percentil 98) estimada para cada localidad costera.

Estación	<i>Run-up</i> (m)
Ciudad Mutis	0.87
Buenaventura	0.33
Tumaco	0.45

Estos resultados muestran una marcada variabilidad espacial, reflejando diferencias en la exposición al oleaje, la pendiente de la playa y la configuración geomorfológica. La mayor altura registrada en Ciudad Mutis concuerda con lo reportado en la compilación oceanográfica del CCCP [126], que señala un mayor régimen energético de oleaje en el norte del litoral Pacífico colombiano, debido a la incidencia predominante de trenes de onda del suroeste y la intensificación de los vientos alisios durante el segundo semestre del año.

Sin embargo, la ubicación de Ciudad Mutis dentro de una bahía parcialmente protegida modula esta energía, como lo evidencian estudios sobre la resonancia de onda larga y refracción local [124]. Esta contradicción aparente entre alta energía en mar abierto y amortiguamiento costero resalta la necesidad de integrar información batimétrica detallada y efectos morfodinámicos específicos en futuras estimaciones.

En contraste, los valores más bajos en Buenaventura (0.33 m) y Tumaco (0.45 m) podrían indicar una mayor protección del oleaje debido a la configuración de sus bahías, que tienden a disipar parte de la energía del oleaje incidente, o a pendientes de playa más suaves.

Es fundamental reconocer que la validación directa de estos valores de *run-up* con estudios locales de campo es limitada en el contexto del Pacífico colombiano. La mayoría de las investigaciones en la región se han centrado en la evaluación de la amenaza por tsunami o en la modelación de inundaciones costeras a gran escala, donde se estiman cotas de inundación máximas que, aunque conceptualmente relacionadas con el *run-up*, provienen de eventos de mayor magnitud y procesos hidrodinámicos distintos ([127], [120]). Estos estudios, sin embargo, sí resaltan la importancia de la topografía de la playa y la batimetría en la extensión de la inundación, un principio que sustenta la metodología de Stockdon et al. (2006).

Es importante señalar que, aunque los valores estimados mediante esta metodología empírica han sido utilizados ampliamente [128], su precisión depende de contar con parámetros locales como la pendiente β_f . En ausencia de mediciones de campo, se recurrió a valores estándar ($\beta_f = 0,05$ pendiente de playa intermedia), lo cual introduce una fuente de incertidumbre no despreciable, especialmente en zonas con playas de poca inclinación como Ciudad Mutis.

Finalmente, a pesar de las limitaciones mencionadas, los valores de *run-up* obtenidos representan una herramienta útil para integrar con otros componentes (marea y sobreelevación ENOS) en la construcción de escenarios de inundación. Su incorporación permite un análisis más holístico de la exposición costera, apoyando estrategias de mitigación y adaptación en contextos de cambio climático y aumento del nivel del mar [117].

6.2.0.3. Determinación incremento de nivel del mar inducido por ENOS

Para estimar el incremento del nivel del mar asociado a eventos del fenómeno ENOS, se emplearon los modelos *LSTM*, previamente entrenados. Estos modelos fueron alimentados con las series correspondientes a eventos específicos de El Niño y La Niña, segmentados por sus modalidades Canónica, Modoki y costera (Tablas 6.3, 6.4), dentro de la ventana de tiempo de los datos disponibles, de acuerdo con la clasificación propuesta por el Instituto Geofísico del Perú (IGP) [95],[129] y estudios previos relevantes sobre la diversidad de manifestaciones del ENOS [40].

Tabla 6.3: Clasificación de eventos ENOS en sus modalidades Canónica y Modoki [40].

El Niño Canónico				La Niña Canónica			
Dic-Feb	Mar-May	Jun-Ago	Sep-Nov	Dic-Feb	Mar-May	Jun-Ago	Sep-Nov
1973	1983	1972	1972	1971	1970	1970	1970
1983	1987	1983	1982	1974	1971	1971	1971
1987	1992	1987	1987	1976	1974	1973	1973
1992	1993	1997	1997	1989	1975	1975	1975
1998	1998	2009	2002	2000	1985	1988	1988
2010	2015	2015	2006	2008	1989	1999	1999
			2009		2007	2000	2007
			2015			2007	2010
El Niño Modoki				La Niña Modoki			
1978	1982	1982	1977	1974	1984	1984	1973
1987	1991	1987	1982	1989	1989	1988	1975
1991	1992	1991	1986	1999	1999	1989	1983
1992	1993	1993	1987	2000	2000	1998	1988
1993	1995	1994	1990	2006	2008	1999	1995
1995	1997	2002	1991	2008	2011	2000	1998
2003	2003	2004	1993	2009		2008	1999
2005	2015	2015	1994	2011		2010	2000
2010			2002	2012			2008
2015			2004				2010
			2006				2011
			2009				
			2015				

Para cada evento, se generaron pronósticos a seis meses del nivel del mar, empleando ventanas de entrada de 48 meses. Posteriormente, se aplicó un análisis de compuestos (*composite analysis*), calculando promedios y valores extremos (máximos para El Niño, mínimos para La Niña) sobre el conjunto de predicciones asociadas a cada tipo de evento. Esta técnica, ampliamente utilizada en climatología, permite identificar respuestas típicas del sistema frente a forzantes específicos, reduciendo la variabilidad no forzada y maximizando la señal [130, 131].

Las Figuras 6.3 , 6.4 y 6.5 presentan los resultados del análisis de compuestos para las loca-

Tabla 6.4: Clasificación de eventos ENOS Costeros [95]

El Niño Costero						La Niña Costera					
Año inicial	Mes inicial	Año final	Mes final	Duración	Magnitud	Año inicial	Mes inicial	Año final	Mes final	Duración	Magnitud
1951	5	1951	11	7	Moderado	1950	2	1950	12	11	Fuerte
1953	3	1953	6	4	Débil	1952	6	1952	9	4	Débil
1957	3	1957	12	10	Fuerte	1954	1	1956	2	26	Fuerte
1965	3	1965	9	7	Moderado	1956	10	1956	12	3	Débil
1969	4	1969	7	4	Débil	1960	5	1960	7	3	Débil
1972	3	1973	1	11	Moderado	1961	6	1961	9	4	Débil
1976	5	1976	11	7	Moderado	1962	2	1962	7	6	Fuerte
1982	7	1983	11	17	Extraordinario	1964	2	1964	12	11	Fuerte
1986	12	1987	12	13	Moderado	1966	3	1966	8	6	Fuerte
1991	7	1992	6	12	Moderado	1967	7	1968	6	12	Fuerte
1993	3	1993	9	7	Moderado	1970	4	1971	11	20	Fuerte
1994	11	1995	1	3	Débil	1973	5	1974	2	10	Moderado
1997	3	1998	9	19	Extraordinario	1974	11	1975	1	3	Moderado
2002	3	2002	5	3	Débil	1975	6	1976	1	8	Fuerte
2002	10	2002	12	3	Débil	1978	6	1978	9	4	Débil
2006	8	2007	1	6	Débil	1985	3	1985	9	7	Moderado
2008	3	2008	9	7	Débil	1988	5	1988	11	7	Moderado
2009	5	2009	9	5	Débil	1996	4	1996	7	4	Moderado
2012	3	2012	7	5	Débil	2001	9	2001	11	3	Débil
2014	5	2014	11	7	Débil	2007	4	2007	12	9	Moderado
2015	4	2016	7	16	Fuerte	2010	8	2010	11	4	Moderado
2016	12	2017	5	6	Moderado	2013	4	2013	8	5	Fuerte
2018	11	2019	3	5	Débil	2017	11	2018	3	5	Débil
2023	2	2024	3	14	Fuerte	2021	11	2022	7	9	Débil
						2022	9	2022	11	3	Débil

lidades de Ciudad Mutis, Buenaventura y Tumaco, respectivamente. En cada gráfico se muestran las curvas de anomalías compuestas promedio (línea punteada) y las de máximos compuestos correspondientes al evento de mayor o menor sobreelevación (línea sólida), diferenciando entre las distintas modalidades de El Niño y La Niña.

Los resultados del presente análisis muestran que los eventos El Niño, en sus diferentes modalidades (Canónico, Costero y Modoki), inducen incrementos diferenciales en el nivel del mar a lo largo del litoral Pacífico colombiano. Las mayores sobreelevaciones fueron registradas durante eventos tipo Canónico y Costero, con anomalías máximas que alcanzaron 0.11 m en Tumaco y 0.10 m en Buenaventura durante el primer mes posterior al inicio del evento. En contraste, los eventos Modoki presentaron anomalías más moderadas, con incrementos en el orden de 0.07 a 0.09 m, aunque con una señal más persistente en el tiempo.

Por su parte, los eventos La Niña indujeron anomalías negativas en el nivel del mar, con reducciones aproximadas entre -0.01 y -0.06 m, siendo estas muy similares durante los eventos de tipo Modoki y Canónico. Estas disminuciones están asociadas a la intensificación de los alisios, el ascenso de agua fría en superficie (upwelling) y un patrón de presión atmosférica que favorece condiciones más estables y descendentes en la columna de agua.

Aunque el presente estudio se enfocó en variables océano-atmosféricas (nivel del mar, TSM, viento, presión atmosférica y precipitación), los patrones identificados presentan analogías con los resultados de Navarro-Monterroza et al. [40], quienes evaluaron la variabilidad de la precipitación en Colombia durante las distintas fases del ENOS. En su investigación, concluyeron que El Niño Canónico tiende a generar reducciones más intensas de la precipitación respecto al Modoki, mientras que La Niña Modoki puede inducir precipitaciones mayores que su contraparte Canónica,

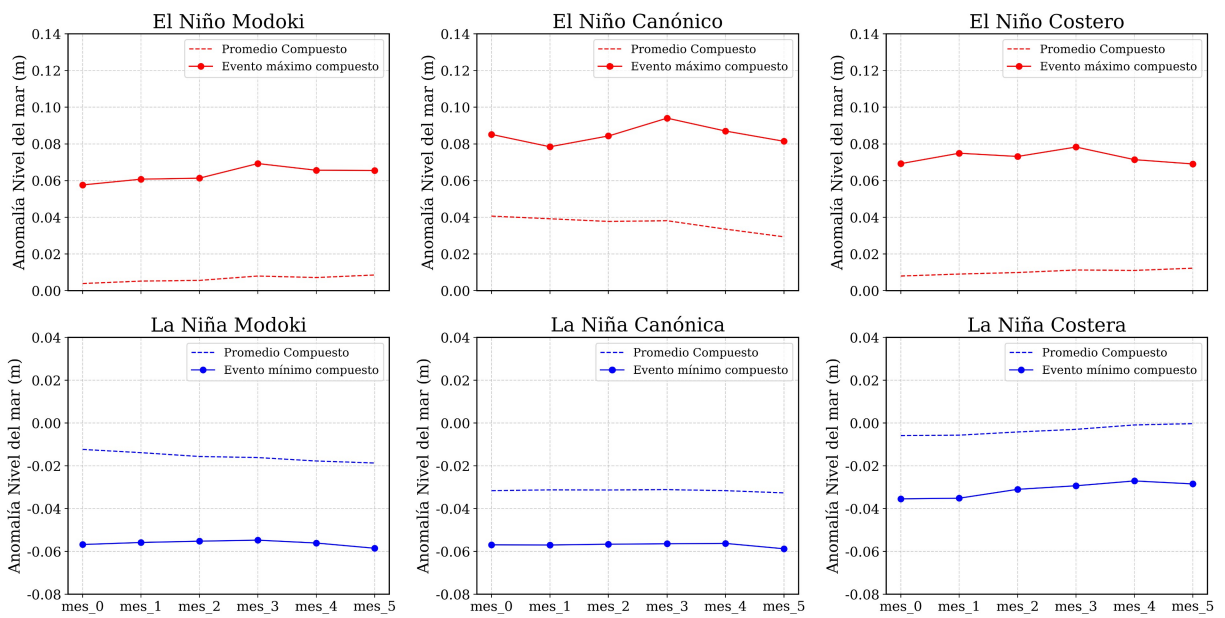


Figura 6.3: Pronóstico de anomalías del nivel del mar en Ciudad Mutis para cada tipo de evento ENOS.

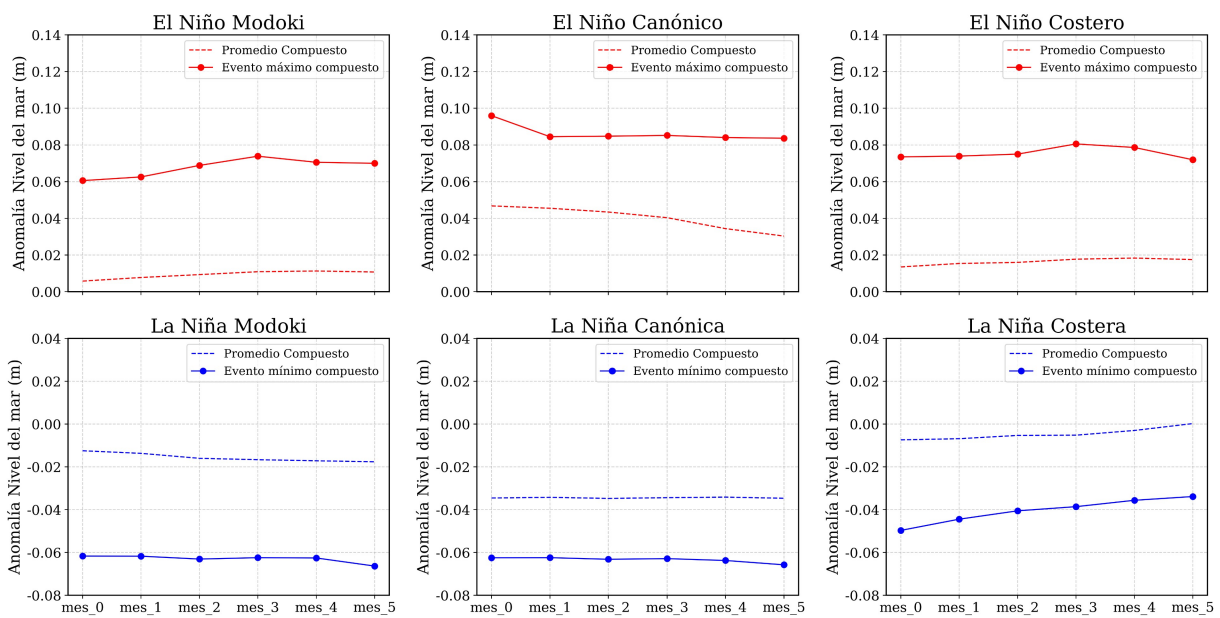


Figura 6.4: Pronóstico de anomalías del nivel del mar en Buenaventura para cada tipo de evento ENOS. Se muestran promedios compuestos y eventos extremos.

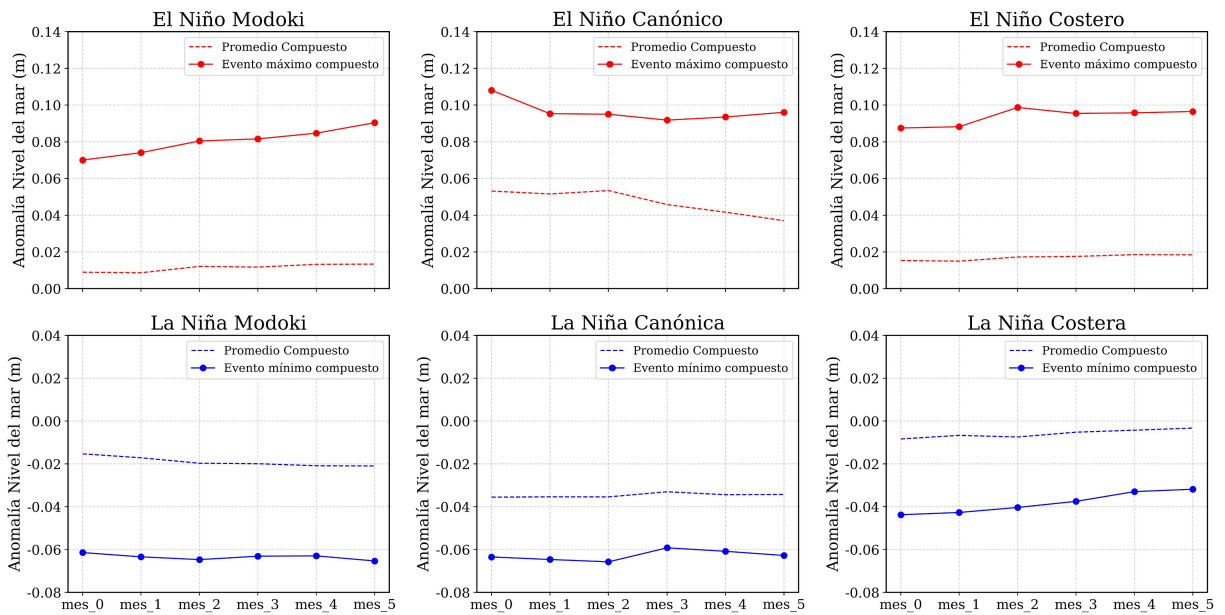


Figura 6.5: Pronóstico de anomalías del nivel del mar en Tumaco para cada tipo de evento ENOS.

particularmente durante el trimestre diciembre-enero-febrero (DJF).

A pesar de abordar variables distintas, ambos estudios coinciden en que la respuesta del sistema climático colombiano frente al ENOS está fuertemente condicionada por la modalidad del evento. Esta dependencia se refleja, por un lado, en los patrones espaciales y temporales de la precipitación (como se observa en Navarro-Monterroza et al. [40]) y, por otro, en las anomalías del nivel del mar estimadas en el presente trabajo. Además, ambos análisis destacan que dichos efectos pueden persistir hasta seis meses posteriores al inicio del evento, lo cual es coherente con el horizonte de predicción empleado en los modelos *LSTM* utilizados en este estudio.

Esta persistencia también fue confirmada mediante el análisis de correlación cruzada (Figura 4.2), donde se observaron correlaciones máximas en distintos desfases temporales, particularmente asociados a los índices Niño 3 y Niño 1+2. Estos índices representan las regiones oceánicas vinculadas con la génesis de los eventos El Niño Canónico y Costero, respectivamente, los cuales fueron precisamente los que mayores efectos indujeron en el ascenso del nivel del mar en las localidades costeras analizadas.

Tabla 6.5: Sobreelevación (o reducción) del nivel del mar estimada mediante análisis compuesto por tipo de evento ENOS y localidad. Se presentan los valores máximos (El Niño) y mínimos (La Niña) absolutos y los promedios durante los seis meses posteriores al inicio del evento.

Estación	El Niño Modoki (m)		El Niño Canónico (m)		El Niño Costero (m)	
	Máximo	Promedio	Máximo	Promedio	Máximo	Promedio
Ciudad Mutis	0.07	0.01	0.09	0.04	0.08	0.01
Bahía Buenaventura	0.07	0.01	0.10	0.04	0.08	0.02
Bahía Tumaco	0.09	0.01	0.11	0.05	0.10	0.02

Estación	La Niña Modoki (m)		La Niña Canónica (m)		La Niña Costera (m)	
	Máximo	Promedio	Máximo	Promedio	Máximo	Promedio
Ciudad Mutis	-0.05	-0.02	-0.06	-0.03	-0.03	0.00
Bahía Buenaventura	-0.06	-0.02	-0.06	-0.03	-0.03	0.00
Bahía Tumaco	-0.06	-0.02	-0.06	-0.03	-0.03	-0.01

La Tabla 6.5 muestra el resumen de los valores compuestos promedio absoluto, como los extremos absolutos calculados, los cuales constituyen el insumo faltante para la construcción de los escenarios de inundación, ya que permiten representar no sólo la respuesta media del océano a cada tipo de evento, sino también el comportamiento más severo registrado históricamente.

6.2.0.4. Definición de escenarios predictivos de incremento del nivel del mar para los diferentes eventos ENOS

Con el propósito de construir escenarios de incremento del nivel del mar orientados a la gestión del riesgo costero, se integraron de manera aditiva tres componentes clave: la cota de marea observada (P90), el *run-up* por oleaje y la sobreelevación del nivel del mar inducida por el fenómeno ENOS. Esta metodología no exige coincidencia temporal exacta entre las series, ya que el objetivo

principal es caracterizar condiciones potencialmente extremas que podrían dar lugar a eventos de inundación marina.

Dado el enfoque preventivo del análisis, centrado en la identificación de situaciones críticas para la infraestructura y las comunidades costeras, se decidió considerar únicamente los eventos El Niño (en sus modalidades Modoki, Canónica y costera), excluyendo los eventos La Niña. Esto se fundamenta en el hecho de que los eventos El Niño suelen asociarse a sobreelevaciones positivas del nivel del mar en el Pacífico colombiano, aumentando el riesgo de inundación en zonas bajas o urbanizadas.

En este contexto, los escenarios pronosticados para las distintas variantes del fenómeno ENOS en cada una de las localidades analizadas del litoral Pacífico colombiano se presentan en las Tablas 6.6, 6.7 y 6.8. En dichas tablas se reporta el cálculo de la cota total de inundación, obtenida a partir de la suma del *run-up* del oleaje, la cota de marea observada (P90) y la sobreelevación inducida por el evento ENOS correspondiente.

Esta información fue posteriormente empleada en la elaboración de los mapas de escenarios de inundación, los cuales permitieron identificar y delimitar las zonas más vulnerables dentro de cada núcleo urbano costero. Estos resultados obtenidos constituyen una herramienta técnica de gran utilidad para respaldar decisiones estratégicas relacionadas con el ordenamiento territorial, la planificación de infraestructuras portuarias y la gestión del riesgo frente al ascenso del nivel del mar asociado a la variabilidad climática.

Tabla 6.6: Cotas de inundación estimadas a partir del del *run-up* de oleaje, la marea observada (P90), y la sobreelevación generada por eventos El Niño Modoki, en cada estación.

Estación	Run-up (m)	Marea (m)	El Niño Modoki (m)	Cota inundación (m)
Ciudad Mutis	0.87	3.59	0.07	4.52
Buenaventura	0.33	4.02	0.07	4.42
Tumaco	0.45	3.38	0.09	3.92

Tabla 6.7: Cotas de inundación estimadas a partir del *run-up* de oleaje, la marea observada (P90) y la sobreelevación generada por eventos El Niño Canónico, en cada estación.

Estación	Run-up (m)	Marea (m)	El Niño Canónico (m)	Cota inundación (m)
Ciudad Mutis	0.87	3.59	0.09	4.55
Buenaventura	0.33	4.02	0.10	4.44
Tumaco	0.45	3.38	0.11	3.94

Tabla 6.8: Cotas de inundación estimadas a partir del run-up de oleaje, la marea observada (P90) y la sobreelevación generada por eventos El Niño Costero, en cada estación.

Estación	Run-up (m)	Marea (m)	El Niño Costero (m)	Cota inundación (m)
Ciudad Mutis	0.87	3.59	0.08	4.53
Buenaventura	0.33	4.02	0.08	4.43
Tumaco	0.45	3.38	0.10	3.93

6.2.1. Modelación espacial del impacto del nivel del mar sobre los principales poblados del pacífico colombiano

6.2.1.1. Ciudad Mutis

La Figura 6.6 presenta el escenario final de inundación modelado para Ciudad Mutis bajo condiciones de sobreelevación del nivel del mar inducidas por el fenómeno de El Niño. El mapa muestra la extensión espacial de la zona potencialmente inundada mediante una capa sombreada en azul, la cual representa las áreas que quedarían cubiertas por agua como resultado de la combinación de la marea alta observada, el *run-up* y la elevación anómala del nivel del mar asociada a los tres tipos de eventos El Niño: Canónico, Costero y Modoki. Estas tres variantes muestran diferencias estimadas de entre 2 y 3 cm en la altura total del nivel del mar (ver Tabla 6.7). El modelado se realizó integrando datos de topografía de alta resolución obtenidos mediante sensores *LiDAR* y utilizando un modelo de interpolación espacial basado en el modelo determinístico lineal.

El análisis espacial permite observar que el ingreso del agua hacia el área urbana ocurre principalmente a través del cauce del río que bordea Ciudad Mutis, generando un desbordamiento que afecta directamente uno de los barrios colindantes con el margen fluvial. Este comportamiento indica que, bajo condiciones extremas de sobreelevación oceánica, el río puede actuar como vía de propagación de la marea hacia el interior del casco urbano, superando los niveles normales de marea astronómica y provocando anegamientos localizados.

La Figura 6.7 muestra la versión vectorizada del escenario de inundación, permitiendo identificar de manera precisa las coberturas de uso del suelo afectadas. Se destaca un impacto casi total sobre los ecosistemas de manglar presentes en el área urbana de Ciudad Mutis, lo cuál nos muestra las consecuencias sobre este terreno si este ecosistema dejara de existir. Por otra parte, se estima que aproximadamente el 12% del área construida correspondiente a viviendas e infraestructura urbana se encuentra dentro de la zona de inundación. Aunque esta proporción puede parecer baja, reviste una alta importancia considerando la concentración poblacional y el valor socioeconómico de estas edificaciones, muchas de las cuales presentan condiciones estructurales precarias. Estos resultados refuerzan la necesidad de incorporar escenarios de elevación del nivel del mar en la planificación territorial y el diseño de medidas de adaptación a la variabilidad climática en zonas costeras vulnerables como Ciudad Mutis en Bahía Solano.

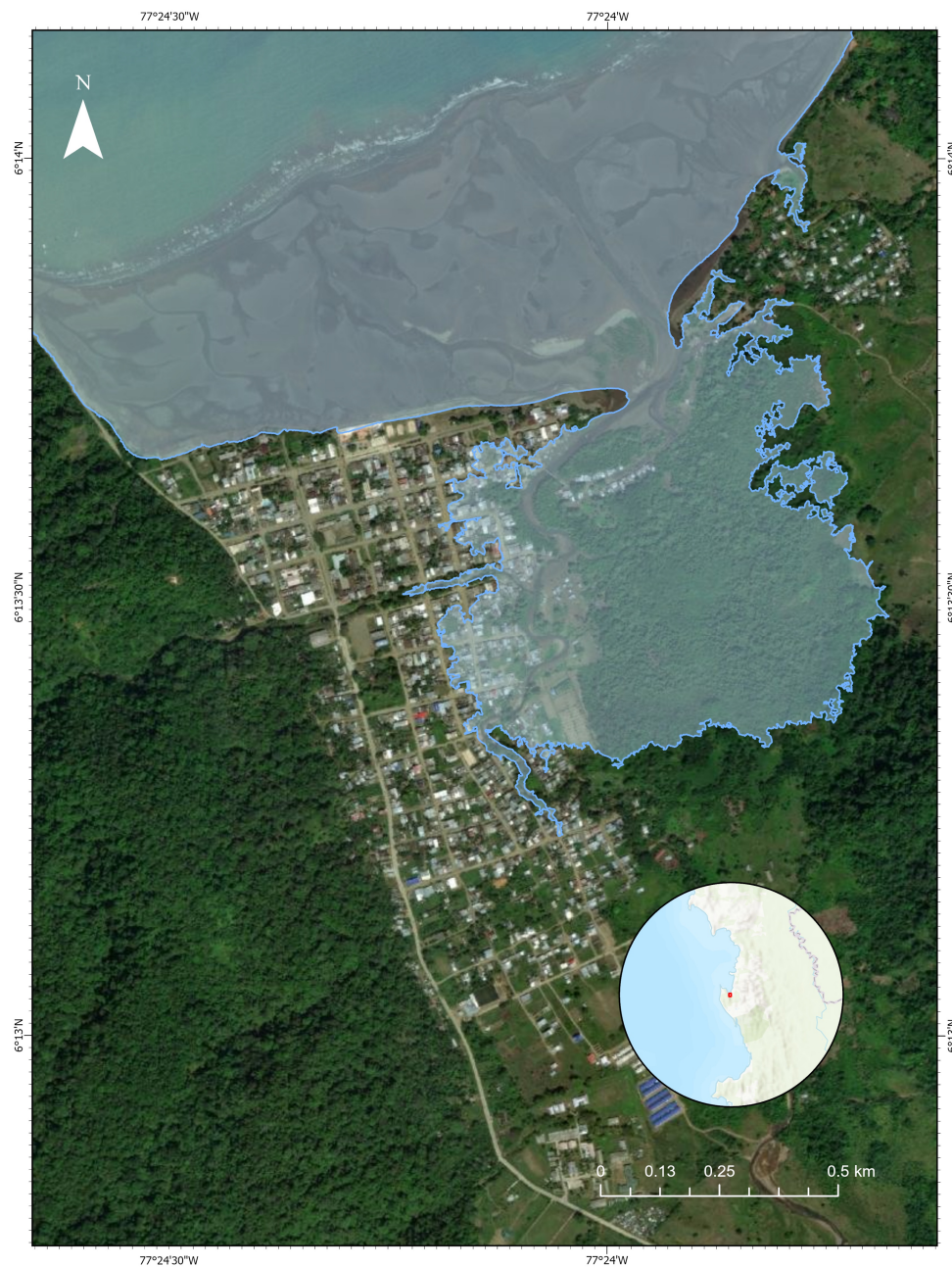


Figura 6.6: Mapa del escenario de inundación para Ciudad Mutis

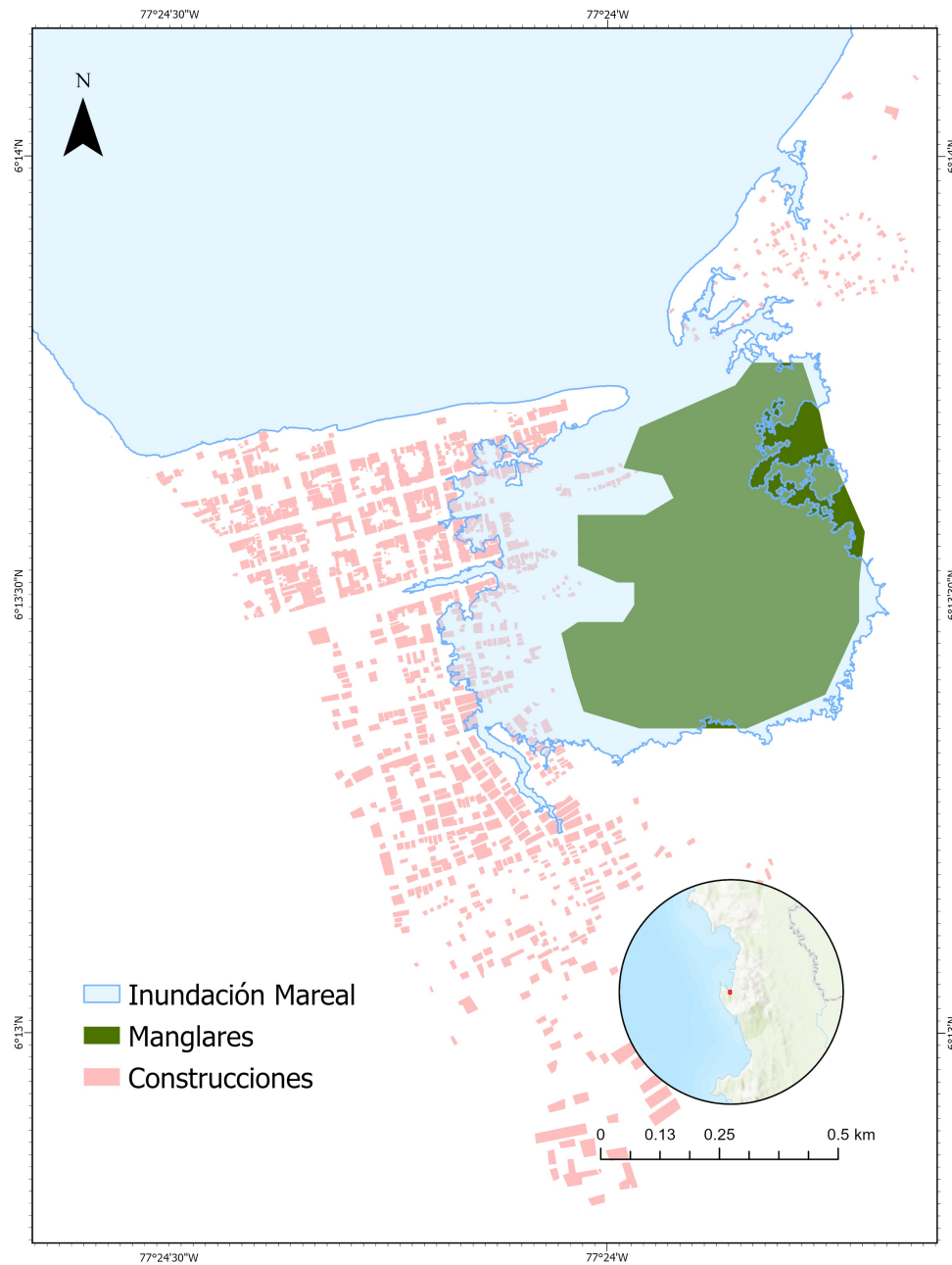


Figura 6.7: Mapa vectorizado del escenario de inundación para Ciudad Mutis, resaltando los ecosistemas de manglar, las construcciones de las viviendas y la porción inundada por la marea.

6.2.1.2. Buenaventura

La Figura 6.8 muestra el escenario de inundación modelado para la ciudad de Buenaventura, resultado de la interacción entre la marea astronómica, el *run-up* y la sobreelevación asociada a eventos extremos vinculados al fenómeno de El Niño. El análisis espacial revela que las zonas más afectadas se concentran principalmente en la Isla Cascajal, donde predominan asentamientos informales construidos sobre viviendas palafíticas. Estas estructuras, ubicadas en áreas de baja altitud y sin infraestructura adecuada de contención, presentan una alta vulnerabilidad frente a fluctuaciones significativas del nivel del mar. En el sector continental, las áreas más impactadas corresponden a la desembocadura del río Dagua, en donde se observa una amplia extensión de zonas urbanas costeras expuestas a inundaciones recurrentes. Los barrios construidos al borde del litoral, sobre rellenos o plataformas elevadas artificialmente, también muestran altos niveles de exposición ante escenarios de sobreelevación oceánica.

Por otro lado, la Figura 6.9 destaca el impacto potencial del escenario de inundación sobre los ecosistemas de manglar presentes en el entorno urbano de Buenaventura. El análisis espacial indica que más del 95 % de la cobertura de manglar en el casco urbano se vería afectada bajo este escenario, lo cual representa un riesgo significativo para la integridad y la protección costera natural, en caso de que este ecosistema sea intervenido. En cuanto al componente urbano construido, se estima que aproximadamente el 15 % de la superficie de cobertura urbana de Buenaventura sería directamente impactada por las aguas, comprometiendo infraestructuras habitacionales, redes de movilidad y servicios básicos.

6.2.1.3. Tumaco

La Figura 6.10 muestra el escenario de inundación modelado para el casco urbano de Tumaco, bajo condiciones extremas que combinan la marea alta con una sobreelevación proyectada del nivel del mar dada por la expansión térmica por el ENOS. El mapa presenta la extensión espacial de las áreas potencialmente inundadas como resultado de la interacción de estos factores, considerando la topografía local y la respuesta geomorfológica del terreno.

En la Figura 6.11, se visualiza el impacto espacial de este escenario de inundación sobre las principales coberturas del suelo, destacando una afectación significativa sobre los ecosistemas de manglar, los cuales resultan cubiertos en aproximadamente un 80 % de su extensión dentro del área urbana. Este resultado es especialmente relevante, dado el rol ecológico clave que estos ecosistemas desempeñan en la protección costera, la biodiversidad y el sustento de comunidades locales.

En cuanto al componente construido, la inundación afecta las tres islas que conforman la ciudad: El Morro, La Viciosa y la isla central de Tumaco. La mayor proporción del impacto urbano se concentra en esta última, que alberga la mayor densidad poblacional y una infraestructura más consolidada, pero también más vulnerable por su cercanía al nivel del mar y la presencia de viviendas palafíticas en condiciones precarias.

En conjunto, los resultados ponen de manifiesto la alta vulnerabilidad física y social de Tumaco frente a escenarios extremos de incremento del nivel del mar por la variabilidad climática. Las zonas afectadas combinan factores de exposición ambiental (baja altitud, proximidad al mar, ecosistemas

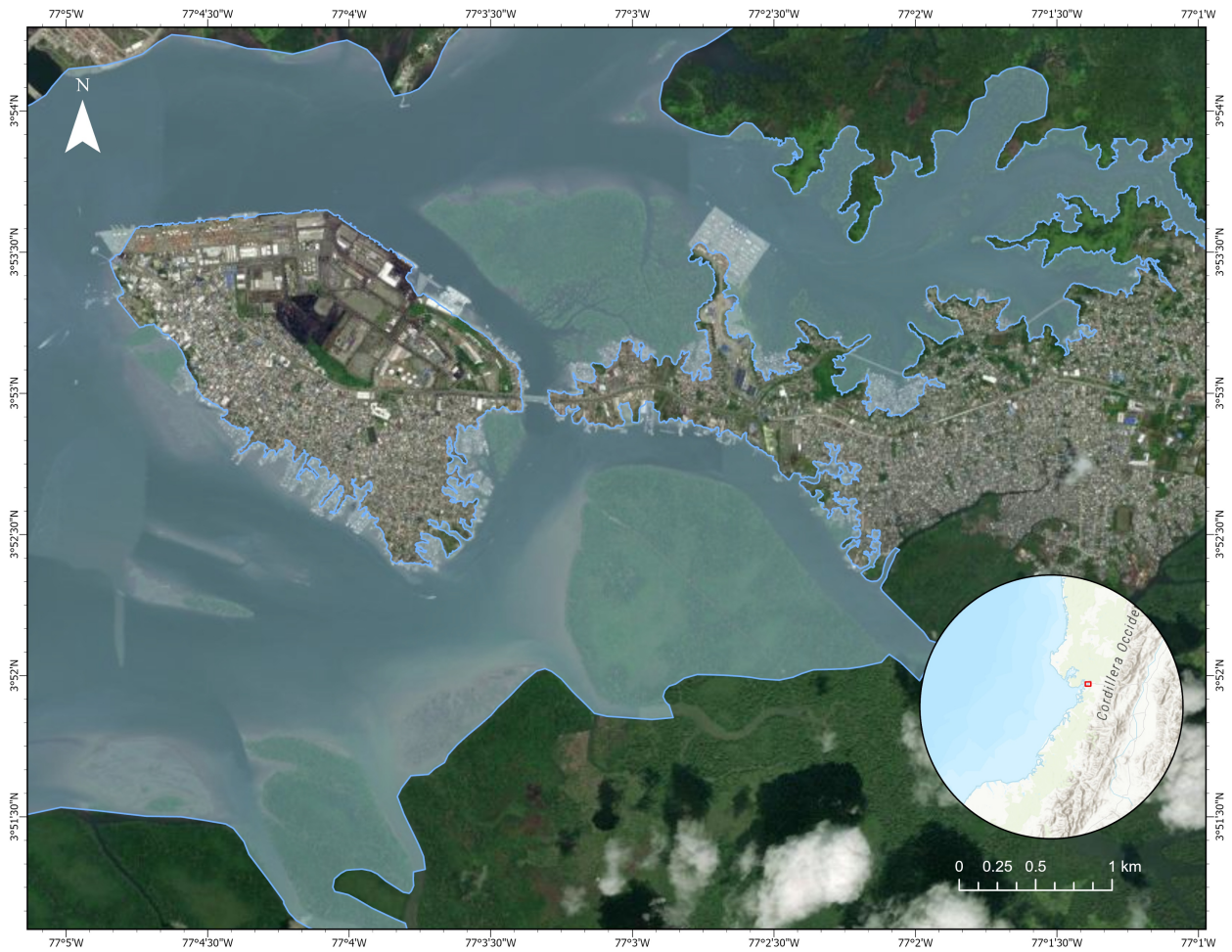


Figura 6.8: Mapa del escenario de inundación para Buenaventura.

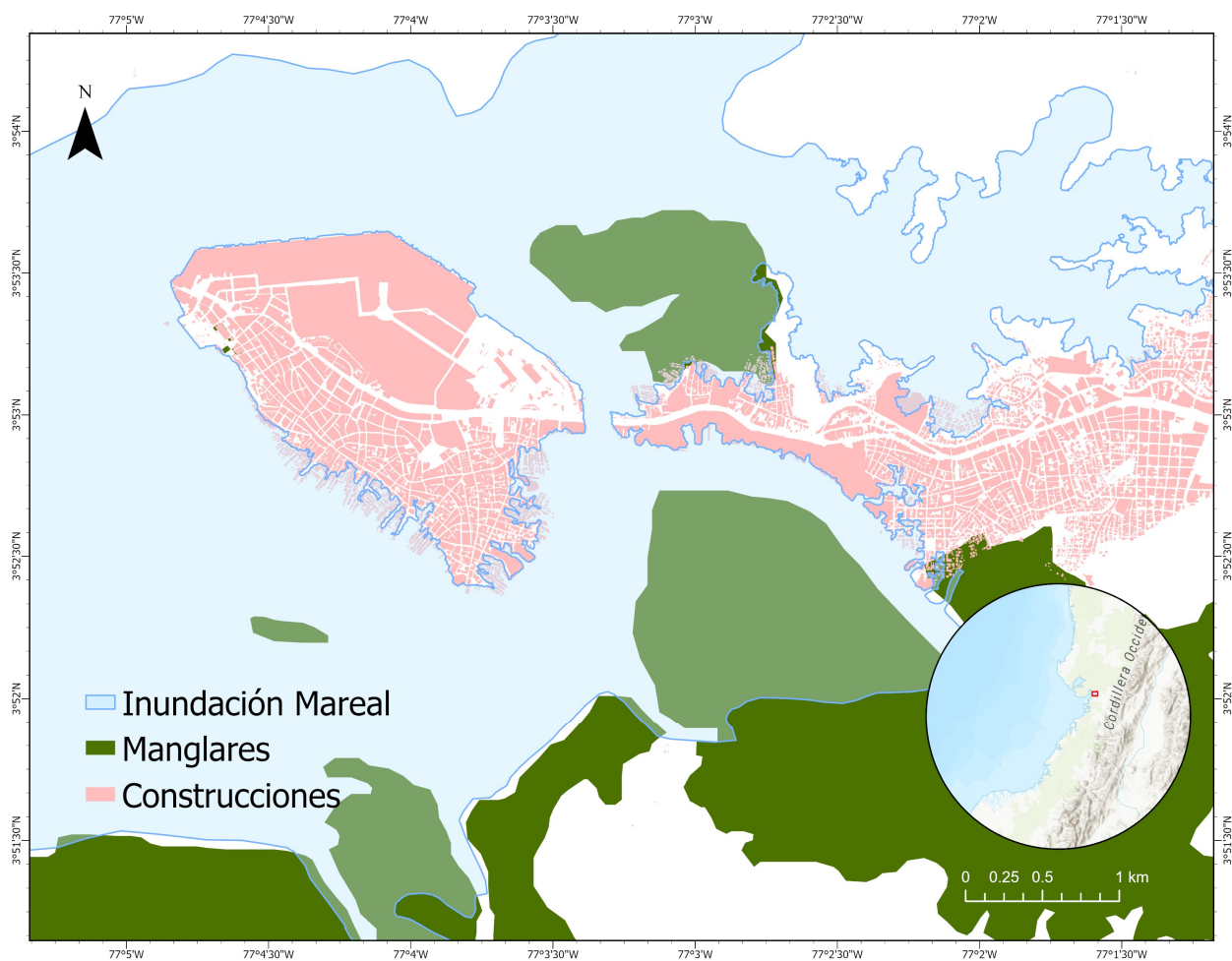


Figura 6.9: Mapa vectorizado del escenario de inundación para Buenaventura, resaltando los ecosistemas de manglar y la porción inundada por la marea.

sensibles) con factores socioeconómicos críticos (altas densidades poblacionales y condiciones habitacionales frágiles), lo que refuerza la urgencia de implementar estrategias de adaptación climática y ordenamiento territorial con enfoque de reducción de riesgo.

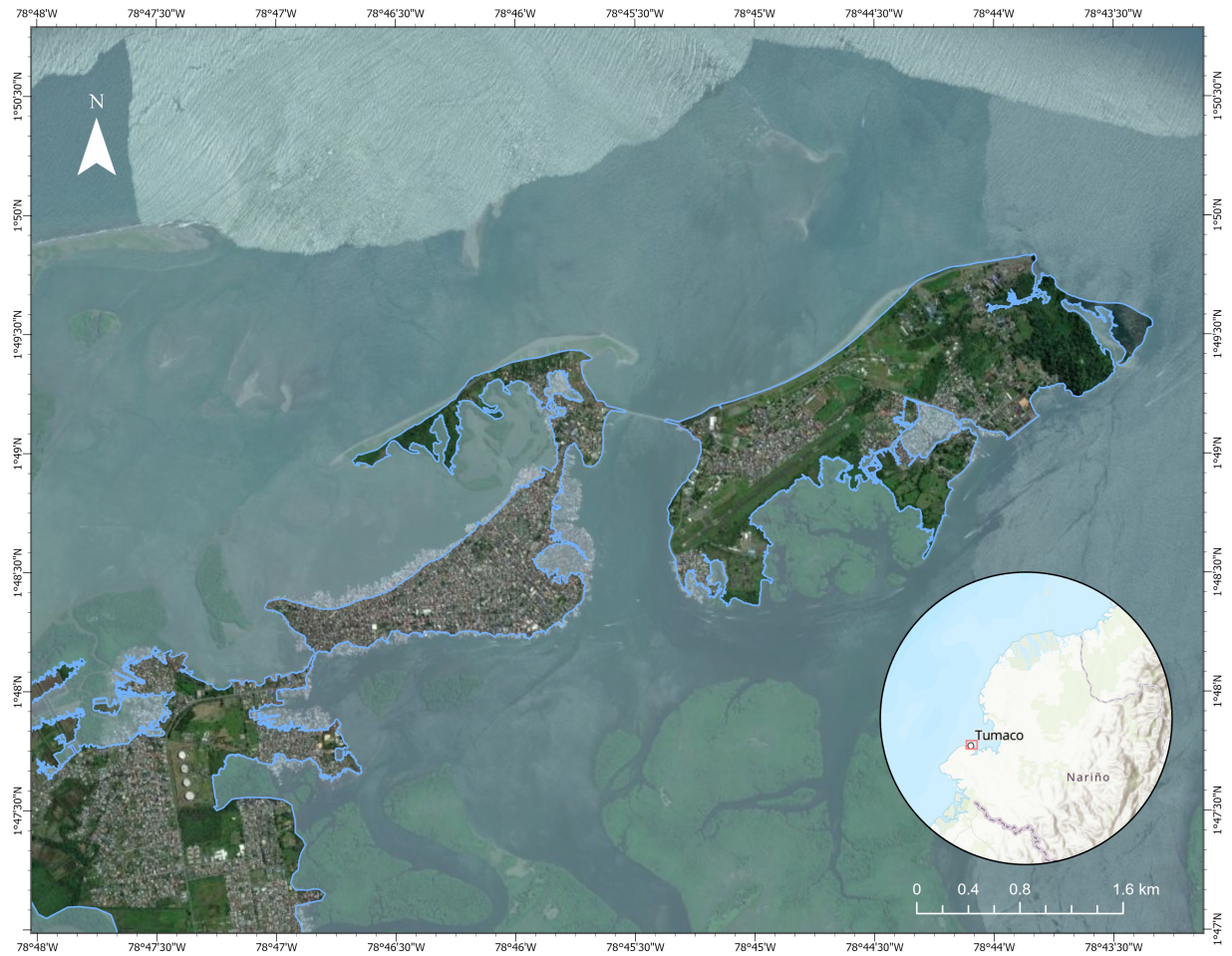


Figura 6.10: Mapa del escenario de inundación para Tumaco.

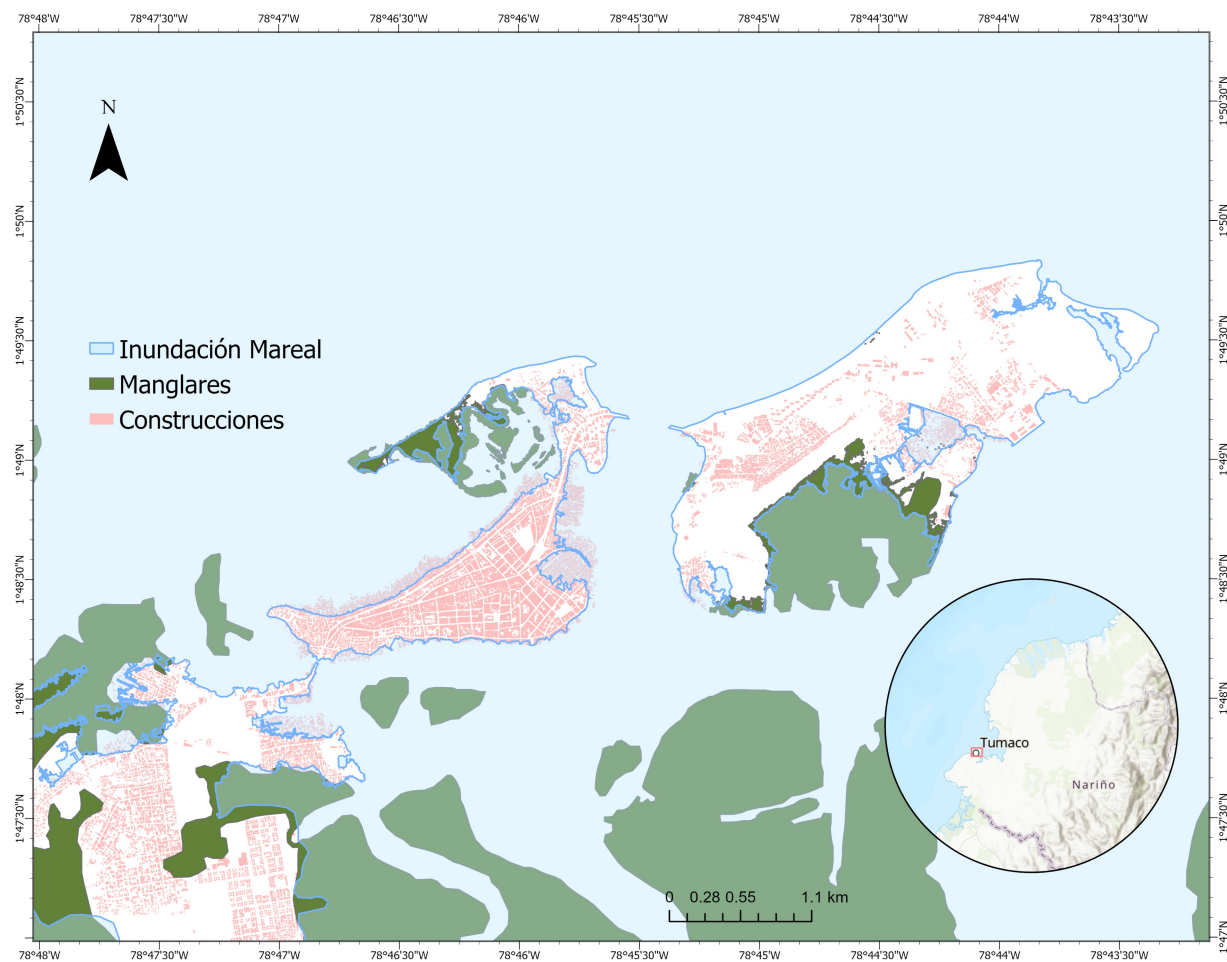


Figura 6.11: Mapa vectorizado del escenario de inundación para Tumaco, resaltando los ecosistemas de manglar, las construcciones de las viviendas y la porción inundada por la marea.

Conclusiones y Trabajos Futuros

7.1. Conclusiones

- Se identificaron patrones robustos de asociación entre las variables océano-atmosféricas (temperatura superficial del mar, nivel del mar, viento y presión atmosférica) y las distintas variantes del fenómeno El Niño (Canónico, Modoki y Costero), mediante el análisis de correlación canónica regularizada (*rCCA*) y modelos de series de tiempo multivariados. Los resultados mostraron que, aunque tanto la temperatura superficial del mar (TSM) costera como los índices climáticos de El Niño representan anomalías térmicas oceánicas, fueron los índices Niño, en particular el Niño 3.4 y el Niño 3, los que alcanzaron los mayores pesos canónicos, destacando su papel dominante como integradores de la señal climática interanual asociada al ENOS y su capacidad para explicar la variabilidad del nivel del mar en el Pacífico colombiano. En contraste, la TSM local presentó pesos canónicos más bajos y heterogéneos, lo cual se atribuyó a su carácter eminentemente local, sujeto a procesos subregionales como surgencias, corrientes costeras y forzamientos atmosféricos inmediatos, que introducen mayor ruido y reducen su estabilidad como predictor robusto. Por su parte, las componentes zonal (U) y meridional (V) del viento, así como la presión media al nivel del mar, exhibieron patrones de asociación diferenciados según la zona geográfica, reflejando la influencia de procesos locales, orográficos y de circulación regional, como el Chorro de Panamá o la advección ecuatorial. Estos hallazgos permiten concluir que la respuesta del nivel del mar frente al ENOS está condicionada tanto por la magnitud y el tipo de evento (Canónico, Modoki o Costero), como por la interacción local entre múltiples variables océano-atmosféricas, resaltando la relevancia de emplear enfoques multivariados como el *rCCA* para caracterizar y comprender esta compleja relación.
- A partir de los patrones de correlación identificados, se desarrollaron escenarios predictivos sobre la elevación del nivel del mar para los distintos tipos de El Niño (Canónico, Modoki y Costero), utilizando modelos de series de tiempo basados en redes neuronales tipo *Long Short-Term Memory (LSTM)*. Los resultados demostraron que el Modelo LSTM entrenado con series multivariadas originales de las variables ambientales (Modelo 3) superó significativamente en desempeño a los modelos basados en componentes canónicos (Modelos 1 y 2), alcanzando valores de R^2 superiores al 0.92 en todas las estaciones, con errores *RMSE* mínimos entre 0.008 m y 0.011 m. Estos modelos permitieron simular de manera precisa las anomalías mensuales del nivel del mar, capturando tanto su variabilidad interanual como las sobreelevaciones inducidas por eventos ENOS. Los análisis de compuestos evidenciaron que los eventos El Niño Canónico y Costero generan los mayores incrementos del nivel del mar en las localida-

des estudiadas, con anomalías máximas de hasta 0.11 m en Bahía Tumaco y 0.10 m en Bahía Buenaventura. En contraste, los eventos Modoki produjeron sobreelevaciones más moderadas, aunque más persistentes en el tiempo. Los eventos La Niña, por su parte, indujeron descensos del nivel del mar entre -0.01 m y -0.06 m. Estas conclusiones consolidan la evidencia de que el comportamiento del nivel del mar bajo diferentes escenarios ENOS es diferenciable tanto en magnitud como en persistencia, y que es factible predecir dicho comportamiento mediante modelos de series de tiempo multivariados, lo cual constituye un insumo crítico para la gestión del riesgo costero y la planificación territorial en el litoral Pacífico colombiano.

- Los modelos de interpolación aplicados a los datos *LiDAR* evidenciaron diferencias notables en cuanto a precisión, estabilidad y adecuación a las características geomorfológicas del litoral Pacífico colombiano. Entre los enfoques geoestadísticos evaluados, el *kriging* con modelos de semivariograma exponencial y esférico se destacó por ofrecer resultados aceptables en términos de ajuste y capacidad predictiva, particularmente en las localidades de Ciudad Mutis y Buenaventura, donde se alcanzaron coeficientes de determinación (R^2) cercanos a 0.99 y bajos errores (*RMSE* y *MAE*), reflejando una representación fiel de la variabilidad espacial del terreno.

Sin embargo, la aplicación de estos modelos en Tumaco reveló importantes limitaciones metodológicas. La distribución irregular de los puntos de muestreo y la escasa densidad en ciertas áreas generaron múltiples advertencias durante la ejecución del *kriging*, relacionadas con la falta de vecinos suficientes para realizar interpolaciones confiables. Este fenómeno afectó directamente la robustez de las estimaciones, reduciendo la precisión y generando incertidumbre espacial en los resultados. Asimismo, el ajuste de los semivariogramas experimentales presentó dificultades, especialmente en los tramos intermedios de distancia, donde la semivarianza no se estabilizó de forma clara, impidiendo una caracterización adecuada de la dependencia espacial. Estos resultados ponen en evidencia la sensibilidad del *kriging* a la densidad y distribución espacial de los datos.

Para abordar estas limitaciones y ampliar la aplicabilidad del *kriging* a conjuntos de datos de gran tamaño, como los provenientes de sensores *LiDAR* de alta resolución, es necesario explorar estrategias de escalamiento computacional. Una alternativa prometedora es la implementación de versiones paralelizadas o distribuidas del *kriging*, aprovechando tecnologías como *Dask* o bibliotecas especializadas en procesamiento en GPU, como *CuPy* o *RAPIDS cuSpatial*. Estas herramientas permiten dividir el dominio espacial en bloques (técnica conocida como *block kriging* o *moving window kriging*) y distribuir la carga de cálculo entre múltiples núcleos de CPU o unidades de procesamiento gráfico, lo cual facilita la interpolación eficiente en dominios extensos sin comprometer la calidad del ajuste.

Otra opción viable consiste en integrar modelos de reducción de dimensionalidad o técnicas de aprendizaje automático geoespacial como el *Gaussian Process Regression (GPR)* [132] con *kernels* adaptativos, que permiten una estimación más flexible de la variabilidad espacial sin requerir una grilla de puntos homogénea. De esta forma, se abre la posibilidad de implementar esquemas de interpolación híbridos, que combinen la robustez estadística del *kriging* con la

escalabilidad y capacidad de generalización de los modelos basados en inteligencia artificial.

- Por otro lado, los métodos determinísticos ofrecieron una alternativa notablemente más eficiente y precisa frente a los enfoques probabilísticos, especialmente en contextos que demandan tiempos de procesamiento reducidos y estabilidad en la generación de superficies continuas. Entre estos métodos, la interpolación lineal se destacó como la técnica más robusta y consistente. En los tres cascos urbanos analizados, Ciudad Mutis, Buenaventura y Tumaco, este método logró valores de coeficiente de determinación (R^2) superiores a 0.999, acompañados de errores mínimos de interpolación, con valores de raíz del error cuadrático medio ($RMSE$) por debajo de los 0.15 metros. Esta elevada precisión se mantuvo de forma consistente entre los distintos pliegues de la validación cruzada, evidenciando una alta capacidad de generalización y baja sensibilidad a la configuración del modelo.

A diferencia de otros métodos determinísticos como la interpolación cúbica o la basada en vecinos más cercanos, la interpolación lineal evitó problemas comunes asociados a cada uno de ellos: el sobreajuste característico del método cúbico, que tiende a introducir oscilaciones no deseadas en regiones con alta variabilidad, y las discontinuidades abruptas observadas en los métodos de vecino más cercano, que comprometen la coherencia espacial de las superficies generadas. En contraste, la interpolación lineal generó superficies topográficas suaves, continuas y coherentes con la morfología costera observada en campo, representando adecuadamente transiciones altimétricas en zonas complejas como bordes de manglares, barras litorales y depresiones inundables.

En este contexto, la interpolación lineal se consolidó como una herramienta altamente confiable, replicable y computacionalmente eficiente. Su desempeño fue óptimo incluso en áreas urbanas costeras complejas, caracterizadas por alta densidad de edificaciones, presencia de palafitos y ecosistemas costeros adyacentes, donde logró mantener coherencia espacial, continuidad en los gradientes altimétricos y una baja sensibilidad frente a la distribución irregular de los puntos *LiDAR*.

Además, su bajo costo computacional la hace ideal para ser implementada en procesos operativos de monitoreo y evaluación periódica de cambios costeros, así como en la actualización continua de inventarios de riesgo.

- Las aplicaciones prácticas de este estudio son múltiples. Los modelos desarrollados permiten estimar escenarios de inundación bajo diferentes condiciones del ENOS, proporcionando insumos valiosos para la planificación territorial y el diseño de estrategias de adaptación climática. No obstante, se identifican áreas de mejora para futuras investigaciones, como la incorporación de variables adicionales (corrientes oceánicas, datos de oleaje *in situ*) y el uso de técnicas de computación de alto rendimiento para optimizar el procesamiento de datos *LiDAR*. Asimismo, se recomienda explorar el acoplamiento de estos modelos con escenarios de cambio climático para evaluar impactos a más largo plazo. En conclusión, esta investigación evidencia el potencial de las técnicas de aprendizaje automático, particularmente las *LSTM*, combinadas con análisis geoespaciales avanzados, para fortalecer los sistemas de alerta temprana y la gestión

integral de riesgos costeros en el Pacífico colombiano.

7.2. Trabajos futuros

A partir de los hallazgos y limitaciones identificadas en el presente estudio, se proponen las siguientes líneas de investigación como continuidad y fortalecimiento del trabajo desarrollado:

- **Implementación de modelos de aprendizaje profundo con enfoque espacio-temporal:** Se propone incorporar arquitecturas como las redes neuronales convolucionales (*CNN*) o modelos híbridos *CNN-LSTM*, que permitan capturar simultáneamente la variabilidad temporal y espacial de las variables océano-atmosféricas, y con ello, mejorar la capacidad predictiva sobre el comportamiento del nivel del mar a escalas regionales y locales.
- **Desarrollo de esquemas híbridos entre modelación numérica y aprendizaje automático:** Una línea de trabajo relevante consiste en combinar resultados de modelos numéricos (e.g., *Delft3D*, *Mike 21*, *XBeach*) con técnicas de *machine learning* o *deep learning*, con el objetivo de propagar hacia la costa las condiciones de borde en mar abierto, permitiendo representar con mayor realismo la dinámica local de las variables predictoras.
- **Optimización de la interpolación espacial mediante técnicas geoestadísticas avanzadas:** se recomienda que futuros estudios consideren la implementación de técnicas de interpolación geoestadística en entornos de cómputo de alto rendimiento (por ejemplo, mediante *GPU* o procesamiento paralelo), lo cual permitiría superar las limitaciones actuales de memoria y escalar los análisis a dominios más extensos. Asimismo, la integración de datos batimétricos y la actualización periódica de los insumos *LiDAR* contribuirían a mejorar la precisión de los modelos y su aplicabilidad en contextos dinámicos. En conjunto, estas mejoras fortalecerían la capacidad predictiva de los sistemas de alerta temprana y la planificación adaptativa frente a la variabilidad climática en zonas costeras.
- **Actualización de insumos altimétricos a partir de datos *LiDAR* recientes:** Considerando los cambios morfológicos que pueden haberse generado en las zonas costeras en los últimos años, se plantea actualizar los datos altimétricos mediante nuevas campañas *LiDAR*, lo cual permitiría una mejor caracterización del relieve costero y, por ende, una mayor precisión en la estimación de cotas de inundación.

Estas líneas de trabajo permitirán avanzar hacia una caracterización más precisa de los riesgos asociados al ascenso del nivel del mar en el Pacífico colombiano, integrando tecnologías emergentes con un enfoque orientado a la toma de decisiones en el contexto de la gestión y planificación costera.

Bibliografía

- [1] C. García, *Oceanografía y Satélites*. Editorial Tébar, S. L., 2009. [Online]. Available: <https://books.google.com.co/books?id=G3XQcQAACAAJ>
- [2] C. Wang and P. C. Fiedler, “ENSO variability and the eastern tropical Pacific: A review,” *Progress in oceanography*, vol. 69, no. 2-4, pp. 239–266, 2006.
- [3] W. Cai, A. Santoso, M. Collins, B. Dewitte, C. Karamperidou, J. S. Kug, M. Lengaigne, M. J. McPhaden, M. F. Stuecker, A. S. Taschetto, A. Timmermann, L. Wu, S. W. Yeh, G. Wang, B. Ng, F. Jia, Y. Yang, J. Ying, X. T. Zheng, T. Bayr, J. R. Brown, A. Capotondi, K. M. Cobb, B. Gan, T. Geng, Y. G. Ham, F. F. Jin, H. S. Jo, X. Li, X. Lin, S. McGregor, J. H. Park, K. Stein, K. Yang, L. Zhang, and W. Zhong, “Changing El Niño–Southern Oscillation in a warming climate,” *Nature Reviews Earth and Environment*, vol. 2, pp. 628–644, 9 2021.
- [4] K. E. Trenberth, “The definition of El Niño,” *Bulletin of the American Meteorological Society*, vol. 78, no. 12, pp. 2771–2777, 1997.
- [5] M. J. McPhaden, S. E. Zebiak, and M. H. Glantz, “Enso as an integrating concept in earth science,” *science*, vol. 314, no. 5806, pp. 1740–1745, 2006.
- [6] S. Power, F. Delage, C. Chung, G. Kociuba, and K. Keay, “Robust twenty-first-century projections of el niño and related precipitation variability,” *Nature*, vol. 502, pp. 541–545, 2013.
- [7] Y. M. Yang, J. H. Park, S. I. An, B. Wang, and X. Luo, “Mean sea surface temperature changes influence ENSO-related precipitation changes in the mid-latitudes,” *Nature Communications*, vol. 12, 12 2021.
- [8] C. F. Ropelewski and M. S. Halpert, “Global and regional scale precipitation patterns associated with the El Niño/Southern Oscillation,” *Monthly Weather Review*, vol. 115, no. 8, pp. 1606–1626, 1987.
- [9] S. Córdoba-Machado, R. Palomino-Lemus, S. R. Gámiz-Fortis, Y. Castro-Díez, and M. J. Esteban-Parra, “Assessing the impact of El Niño Modoki on seasonal precipitation in Colombia,” *Global and Planetary Change*, vol. 124, pp. 41–61, 1 2015.
- [10] K. E. Trenberth and J. M. Caron, “The Southern Oscillation revisited: Sea level pressures, surface temperatures, and precipitation,” *Journal of Climate*, vol. 13, no. 24, pp. 4358–4365, 2000.
- [11] A. Cazenave and W. Llovel, “Contemporary sea level rise,” *Annual Review of Marine Science*, vol. 2, pp. 145–173, 1 2010.

- [12] B. D. Hamlington, R. R. Leben, R. S. Nerem, W. Han, and K.-Y. Kim, "The effect of the El Niño–Southern Oscillation on the global sea level budget," *Geophysical Research Letters*, vol. 41, no. 14, pp. 5123–5130, 2014.
- [13] N. J. M. Ardila, O. J. Rodríguez, and K. Robertson, "Amenazas naturales en el litoral Pacífico colombiano asociadas al ascenso del nivel del mar," *Cuadernos de Geografía: Revista Colombiana de Geografía*, no. 14, pp. 83–96, 2005.
- [14] M. A. Martínez, D. Rodríguez, and E. Pardo, "Vulnerabilidad y riesgo climático en la región del Pacífico colombiano," *Gestión y Ambiente*, vol. 20, no. 2, pp. 33–45, 2017.
- [15] IDEAM, "Atlas de vulnerabilidad al cambio Climático en Colombia," 2020, disponible en: <https://www.ideam.gov.co>.
- [16] Y. Zhao, X. Li, and R. Wang, "A hybrid LSTM model for sea level prediction based on climate indices," *Ocean Engineering*, vol. 248, p. 110777, 2022.
- [17] X. Wang, H. Liu, and Y. Zhang, "Sea level prediction using ensemble empirical mode decomposition and machine learning algorithms," *Journal of Atmospheric and Oceanic Technology*, vol. 37, no. 11, pp. 1949–1964, 2020.
- [18] A. Braakmann-Folgmann, R. Roscher, S. Wenzel, B. Uebbing, and J. Kusche, "Sea level anomaly prediction using recurrent neural networks," *CoRR*, vol. abs/1710.07099, 2017. [Online]. Available: <http://arxiv.org/abs/1710.07099>
- [19] K. Ashok, S. K. Behera, S. A. Rao, H. Weng, and T. Yamagata, "El Niño Modoki and its possible teleconnection," *Geophysical Research Letters*, vol. 34, no. L13705, 2007.
- [20] K. Takahashi and B. Dewitte, "Strong and moderate nonlinear El Niño events in the eastern Pacific," *Climate Dynamics*, vol. 36, no. 5, pp. 1031–1044, 2011.
- [21] X. Liu, "Airborne lidar for dem generation: some critical issues," *Progress in physical geography*, vol. 32, no. 1, pp. 31–49, 2008.
- [22] U. Wandinger, "Introduction to lidar," in *Lidar: range-resolved optical remote sensing of the atmosphere*. Springer, 2005, pp. 1–18.
- [23] J. E. Pabón Caicedo, José Daniel Montealegre Bocanegra, "Los fenómenos de el niño y de la niña, su efecto climático e impactos socioeconómicos," 2017. [Online]. Available: <https://repositorio.accefyn.org.co/handle/001/113>
- [24] E. Navarro-Monterroza, P. A. Arias, and S. C. Vieira, "El niño-oscilación del sur, fase modoki, y sus efectos en la variabilidad espacio-temporal de la precipitación en Colombia," *Revista de la Academia Colombiana de Ciencias Exactas, Físicas y Naturales*, vol. 43, no. 166, pp. 120–132, 2019.

- [25] P. C. Fiedler, “Environmental change in the eastern tropical pacific ocean: review of enso and decadal variability,” *Marine Ecology Progress Series*, vol. 244, pp. 265–283, 2002.
- [26] A. Devis Morales, “Análisis de las condiciones oceanográficas y meteorológicas de la bahía de tumaco y su relación con eventos de escala global.” *Boletín Científico CCCP*, vol. 9, pp. 1–21, 2002.
- [27] B. Wang, “Kelvin waves,” *Encyclopedia of atmospheric sciences*, vol. 1062, 2002.
- [28] J. D. Pabón-Caicedo and J. E. Montealegre-Bocanegra, *Los fenómenos de El Niño y de La Niña, su efecto climático e impactos socioeconómicos*. ACADEMIA COLOMBIANA DE CIENCIAS EXACTAS, FÍSICAS Y NATURALES COLECCIÓN JORGE ÁLVAREZ LLERAS No. 34, 2017, vol. 978-958-9205-93-8.
- [29] A. D. Morales, “Análisis de las condiciones oceanográficas y meteorológicas de la bahía de tumaco y su relación con eventos de escala global,” *Boletín Científico CCCP*, vol. 9, no. 9, pp. 1–21, 2003.
- [30] J.-M. Sayol, L. M. Vásquez, J. L. Valencia, J. R. Linero-Cueto, D. García-García, I. Vigo, and A. Orfila, “Extension and application of an observation-based local climate index aimed to anticipate the impact of el niño–southern oscillation events on colombia,” *International Journal of Climatology*, vol. 42, no. 11, pp. 5403–5429, 2022. [Online]. Available: <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.7540>
- [31] M. Bagheri, Z. Z. Ibrahim, I. D. Wolf, M. F. Akhir, W. I. A. W. Talaat, and B. Oryani, “Sea-level projections using a narx-nn model of tide gauge data for the coastal city of kuala terengganu in malaysia,” *Environmental Science and Pollution Research*, vol. 30, no. 34, pp. 81 839–81 857, 2023.
- [32] S. D. D. M. Deguenon, C. G. Hounmenou, R. Adade, O. Teka, I. I. Toko, D. W. Aheto, and B. Sinsin, “Simulation of the Impacts of Sea-Level Rise on Coastal Ecosystems in Benin Using a Combined Approach of Machine Learning and the Sea Level Affecting Marshes Model,” *Sustainability*, vol. 15, no. 22, p. 16001, 2023.
- [33] V. Spiridonov and M. Ćurić, *General Circulation of the Atmosphere*. Cham: Springer International Publishing, 2021, pp. 229–251. [Online]. Available: https://doi.org/10.1007/978-3-030-52655-9_15
- [34] J. Marshall and R. Plumb, *Atmosphere, Ocean and Climate Dynamics: An Introductory Text*, ser. International Geophysics. Elsevier Science, 2007. [Online]. Available: <https://books.google.com.co/books?id=KvJfvYBHiegC>
- [35] IPCC, *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.), Ed. Geneva, Switzerland: IPCC, 2014.

- [36] V. Spiridonov and M. Ćurić, *Energy and Radiation*. Cham: Springer International Publishing, 2021, pp. 35–58. [Online]. Available: https://doi.org/10.1007/978-3-030-52655-9_5
- [37] T. KE, “Changes in precipitation with climate change,” *Climate Research*, vol. 47, no. 1-2, pp. 123–138, 2011. [Online]. Available: <https://www.int-res.com/abstracts/cr/v47/n1-2/p123-138>
- [38] E. Navarro-Monterroza, P. A. Arias, and S. C. Vieira, “El niño-oscilación del sur, fase modoki, y sus efectos en la variabilidad espacio-temporal de la precipitación en colombia,” *Revista de la Academia Colombiana de Ciencias Exactas, Físicas y Naturales*, vol. 43, no. 166, p. 120–132, abr. 2019. [Online]. Available: <https://raccefyn.co/index.php/raccefyn/article/view/704>
- [39] I. J. Ramírez and F. Briones, “Understanding the el niño costero of 2017: The definition problem and challenges of climate forecasting and disaster responses,” *International Journal of Disaster Risk Science*, vol. 8, pp. 489–492, 12 2017.
- [40] E. Navarro-Monterroza, P. A. Arias, and S. C. Vieira, “El niño-oscilación del sur, fase modoki, y sus efectos en la variabilidad espacio-temporal de la precipitación en colombia,” *Revista de la Academia Colombiana de Ciencias Exactas, Físicas y Naturales*, vol. 43, no. 166, p. 120–132, Apr. 2019. [Online]. Available: <https://raccefyn.co/index.php/raccefyn/article/view/704>
- [41] A. Convenios and D. Protocolos, “Estatuto y reglamento de la cpps,” *Secretaría General. Guayaquil, Ecuador*, 2007.
- [42] *Manual de medición e interpretación del nivel del mar v. V: medidores de radar*, UNESCO, Paris, 2016. [Online]. Available: https://unesdoc.unesco.org/ark:/48223/pf0000246981_spa
- [43] W. Sheng and H. Li, *Tropical Atmosphere Ocean (TAO) Buoy*. Singapore: Springer Singapore, 2020, pp. 1–8. [Online]. Available: https://doi.org/10.1007/978-981-10-6963-5_80-1
- [44] T. Frederikse, F. Landerer, L. Caron, S. Adhikari, D. Parkes, V. W. Humphrey, S. Dangendorf, P. Hogarth, L. Zanna, L. Cheng, and Y. H. Wu, “The causes of sea-level rise since 1900,” *Nature*, vol. 584, pp. 393–397, 8 2020.
- [45] N. A. A. B. S. Bahari, A. N. Ahmed, K. L. Chong, V. Lai, Y. F. Huang, C. H. Koo, J. L. Ng, and A. El-Shafie, “Predicting Sea Level Rise Using Artificial Intelligence: A Review,” *Archives of Computational Methods in Engineering*, vol. 30, no. 7, pp. 4045–4062, 2023. [Online]. Available: <https://doi.org/10.1007/s11831-023-09934-9>
- [46] L.-L. Fu, “Topex/poseidon mission overview,” *Journal of Geophysical Research*, vol. 99, 1994.
- [47] B. Haines, W. Bertiger, S. Desai, D. Kuang, T. Munson, L. Young, and P. Willis, “Initial orbit determination results for jason-1: Towards a 1-cm orbit *,” 2002.
- [48] B. Tapley and C. Reigber, “Grace: a satellite-to-satellite tracking geopotential mapping mission,” in *Proceedings of Second Joint Meeting of the Int. Gravity Commission and the Int. Geoid Commission*, 1998, pp. 7–12.

- [49] K.-P. Bartholoma, R. Benz, D. Demuth, P. Dubock, B. Gardini, G. Graf, and G. Ratier, “Envisat-1-on its way to hardware.”
- [50] F. Mercier, L. Cerri, and J. P. Berthias, “Jason-2 doris phase measurement processing,” *Advances in Space Research*, vol. 45, pp. 1441–1454, 6 2010.
- [51] T. Ezer and W. B. Corlett, “Is sea level rise accelerating in the chesapeake bay? a demonstration of a novel new approach for analyzing sea level data,” *Geophysical Research Letters*, vol. 39, 10 2012.
- [52] J. Guo, Z. Hu, J. Wang, X. Chang, and G. Li, “Sea level changes of china seas and neighboring ocean based on satellite altimetry missions from 1993 to 2012,” *Journal of Coastal Research*, no. 73, pp. 17–21, 2015.
- [53] S. Dangendorf, D. Rybski, C. Mudersbach, A. Müller, E. Kaufmann, E. Zorita, and J. Jensen, “Evidence for long-term memory in sea level,” *Geophysical Research Letters*, vol. 41, pp. 5530–5537, 8 2014.
- [54] K. M. A. Hassan, M. A. Haque, and S. Ahmed, “Comparative Study of Forecasting Global Mean Sea Level Rising using Machine Learning,” *Proceedings of International Conference on Electronics, Communications and Information Technology, ICECIT 2021*, pp. 1–4, 2021.
- [55] W. Feng, M. Zhong, and H. Xu, “Sea level variations in the south china sea inferred from satellite gravity, altimetry, and oceanographic data,” *Science China Earth Sciences*, vol. 55, pp. 1696–1701, 2012.
- [56] Y. Cheng, H. P. Plag, B. D. Hamlington, Q. Xu, and Y. He, “Regional sea level variability in the bohai sea, yellow sea, and east china sea,” *Continental Shelf Research*, vol. 111, pp. 95–107, 12 2015.
- [57] T. Yan, K. Yu, L. Jiang, Y. Li, and N. Zhao, “Significant sea-level fluctuations in the western tropical pacific during the mid-holocene,” *Paleoceanography and Paleoclimatology*, vol. 39, 5 2024.
- [58] B. D. Hamlington, S. H. Cheon, P. R. Thompson, M. A. Merrifield, R. S. Nerem, R. R. Leben, and K. Y. Kim, “An ongoing shift in pacific ocean sea level,” *Journal of Geophysical Research: Oceans*, vol. 121, pp. 5084–5097, 7 2016.
- [59] G. Poveda and O. J. Mesa, “On the existence of lloró (the rainiest locality on earth): Enhanced ocean-land-atmosphere interaction by a low-level jet,” *Geophysical research letters*, vol. 27, no. 11, pp. 1675–1678, 2000.
- [60] S. López Isaza, J. Restrepo-Angel, J. Restrepo, C. Monroy, H. Mora, and E. Rodríguez-Rubio, “Nivel relativo del mar en la costa pacifica sur de colombia: variabilidad, tendencias e implicaciones en la dinámica deltaica,” *Boletín Geológico*, vol. 42, pp. 53–66, 01 2009.

- [61] E. Rodríguez-Rubio, "A multivariate climate index for the western coast of colombia," *Advances in Geosciences*, vol. 33, pp. 21–26, 04 2013.
- [62] V. Lai, A. N. Ahmed, M. A. Malek, H. Abdulmohsin Afan, R. K. Ibrahim, A. El-Shafie, and A. El-Shafie, "Modeling the nonlinearity of sea level oscillations in the malaysian coastal areas using machine learning algorithms," *Sustainability*, vol. 11, no. 17, p. 4643, 2019.
- [63] V. Nieves, C. Radin, and G. Camps-Valls, "Predicting regional coastal sea level changes with machine learning," *Scientific Reports*, vol. 11, no. 1, pp. 1–6, 2021. [Online]. Available: <https://doi.org/10.1038/s41598-021-87460-z>
- [64] N. A. Hazrin, K. L. Chong, Y. F. Huang, A. N. Ahmed, J. L. Ng, C. H. Koo, K. W. Tan, M. Sherif, and A. El-Shafie, "Predicting sea levels using ml algorithms in selected locations along coastal malaysia," *Heliyon*, vol. 9, no. 9, 2023.
- [65] T. O. Muslim, A. N. Ahmed, M. Malek, H. Abdulmohsin Afan, R. Khaleel Ibrahim, A. El-Shafie, M. Sapitang, M. Sherif, A. Sefelnasr, and A. El-Shafie, "Investigating the influence of meteorological parameters on the accuracy of sea-level prediction models in sabah, malaysia," *Sustainability*, vol. 12, no. 3, p. 1193, 2020.
- [66] A. F. Jamali, A. Mustapha, and S. A. Mostafa, "Prediction of sea level oscillations: Comparison of regression-based approach." *Engineering Letters*, vol. 29, no. 3, 2021.
- [67] A.-L. Balogun and N. Adebisi, "Sea level prediction using arima, svr and lstm neural network: assessing the impact of ensemble ocean-atmospheric processes on models' accuracy," *Geomatics, Natural Hazards and Risk*, vol. 12, no. 1, pp. 653–674, 2021.
- [68] V. Lai, A. N. Ahmed, M. A. Malek, H. A. Afan, R. K. Ibrahim, A. El-Shafie, and A. El-Shafie, "Modeling the Nonlinearity of Sea Level Oscillations in the Malaysian Coastal Areas Using Machine Learning Algorithms," *Sustainability (Switzerland)*, vol. 11, no. 17, 2019.
- [69] M. Bagheri, Z. Z. Ibrahim, S. Mansor, L. Abd Manaf, M. Akhir, W. Talaat, and I. D. Wolf, "Hazard assessment and modeling of erosion and sea level rise under global climate change conditions for coastal city management," *Natural Hazards Review*, vol. 24, no. 1, p. 04022038, 2023.
- [70] L. Kumar, M. S. Afzal, and M. M. Afzal, "Mapping shoreline change using machine learning: a case study from the eastern indian coast," *Acta Geophysica*, vol. 68, no. 4, pp. 1127–1143, 2020.
- [71] A. Alshouny, M. T. Elnabwy, M. R. Kaloop, A. Baik, and Y. Miky, "An integrated framework for improving sea level variation prediction based on the integration wavelet-artificial intelligence approaches," *Environmental Modelling & Software*, vol. 152, p. 105399, 2022.

- [72] J.-F. Legeais, M. Ablain, L. Zawadzki, H. Zuo, J. A. Johannessen, M. G. Scharffenberg, L. Fenoglio-Marc, M. J. Fernandes, O. B. Andersen, S. Rudenko *et al.*, “An improved and homogeneous altimeter sea level record from the esa climate change initiative,” *Earth System Science Data*, vol. 10, no. 1, pp. 281–301, 2018.
- [73] H. Hersbach, B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers *et al.*, “The era5 global reanalysis,” *Quarterly journal of the royal meteorological society*, vol. 146, no. 730, pp. 1999–2049, 2020.
- [74] H. E. Beck, E. F. Wood, M. Pan, C. K. Fisher, D. G. Miralles, A. I. Van Dijk, T. R. McVicar, and R. F. Adler, “Mswep v2 global 3-hourly 0.1 precipitation: methodology and quantitative assessment,” *Bulletin of the American Meteorological Society*, vol. 100, no. 3, pp. 473–500, 2019.
- [75] M. y. E. A. I. Instituto de Hidrología, *Atlas Climatológico de Colombia*. Bogotá, Colombia: IDEAM, 2017.
- [76] D. G. M. (DIMAR), *Estudio de la Meteorología Marina en el Pacífico Colombiano*. Bogotá, Colombia: Ministerio de Defensa Nacional, 2020.
- [77] J. W. Tukey, *Exploratory Data Analysis*. Reading, MA: Addison-Wesley, 1977.
- [78] V. Barnett and T. Lewis, *Outliers in Statistical Data*, 3rd ed. New York: Wiley, 1994.
- [79] UNESCO, *Manual de control de calidad de datos de nivel del mar*, París, 1993, iOC Manuals and Guides No. 14.
- [80] M. G. G. Foreman and R. F. Henry, *Manual for Tidal Heights Analysis and Prediction*, 2009. [Online]. Available: <https://www.dfo-mpo.gc.ca>
- [81] Copernicus Marine Service, “Global ocean gridded l4 sea surface heights and derived variables reprocessed,” <https://marine.copernicus.eu>, 2021, product ID: SEALEVEL_GLO_PHY_L4_REP_OBSERVATIONS_008_047.
- [82] G. DiBiarboure, M.-I. Pujol, M.-H. Rio, A. Pascual, and F. Mertz, “Handbook of ocean topography: From mean sea surface to dynamic heights,” CLS, AVISO+, Tech. Rep., 2011, aVISO Handbook Document Ref: SALP-MU-P-EA-15693-CLS.
- [83] S.-B. Scientific, *Data Processing Manual*, Bellevue, WA, 2014. [Online]. Available: <https://www.seabird.com>
- [84] Copernicus Marine Service, “Cmems product quality strategy document,” <https://marine.copernicus.eu>, 2022, accessed May 2025.
- [85] Environmental Systems Research Institute, “Arcgis pro.” [Online]. Available: <https://www.esri.com/en-us/arcgis/products/arcgis-pro/resources>

- [86] K. Le Boedec, “Sensitivity and specificity of normality tests and consequences on reference interval accuracy at small sample size: a computer-simulation study,” *Veterinary Clinical Pathology*, vol. 45, no. 4, pp. 648–656, 2016, pMID: 27556235. [Online]. Available: <https://doi.org/10.1111/vcp.12390>
- [87] C. Marín Navas, J. V. Delgado Bermejo, A. K. McLean, J. M. León Jurado, M. E. Camacho Vallejo, and F. J. Navas González, “Modeling climate change effects on genetic diversity of an endangered horse breed using canonical correlations,” *Animals*, vol. 14, no. 5, p. 659, 2024, pMCID: PMC10931216. [Online]. Available: <https://doi.org/10.3390/ani14050659>
- [88] K. D. Malakar, S. Roy, and M. Kumar, *Geographic Information Systems: The Basics and Beyond*. Cham: Springer Nature Switzerland, 2025, pp. 125–157. [Online]. Available: https://doi.org/10.1007/978-3-031-92017-2_5
- [89] K.-G. Karlsson, N. Håkansson, J. P. D. Mittaz, T. Hanschmann, and A. Devasthale, “Impact of avhrr channel 3b noise on climate data records: Filtering method applied to the cm saf clara-a2 data record,” *Remote Sensing*, vol. 9, no. 6, p. 568, 2017.
- [90] M. Tudaji, Y. Nan, and F. Tian, “Assessing the value of high-resolution rainfall and streamflow data for hydrological modeling: An analysis based on 63 catchments in southeast china,” *Hydrology and Earth System Sciences*, vol. 29, pp. 1919–1937, 2025.
- [91] U. Schulzweida, *CDO User Guide*, 2022, versión 2.1.0. [Online]. Available: <https://zenodo.org/records/7112925>
- [92] I. P. on Climate Change, “Climate change 2021: The physical science basis - chapter 9,” Intergovernmental Panel on Climate Change. Working Group I, Tech. Rep., 2021, contribution of Working Group I to the Sixth Assessment Report. [Online]. Available: <https://www.ipcc.ch/report/ar6/wg1/>
- [93] NASA, “Sea level: Vital signs of the planet,” <https://climate.nasa.gov/vital-signs/sea-level/>, 2023, consultado en mayo de 2025.
- [94] P. Fraile-Jurado, A. Muñoz, and J. García-Lafuente, “La influencia de la presión atmosférica y el viento en la variabilidad del nivel del mar en el mareógrafo de bonanza (cádiz),” *Revista de Climatología*, vol. 16, pp. 1–15, 2016. [Online]. Available: <https://www.researchgate.net/publication/305636676>
- [95] Instituto Geofísico del Perú, “Lista de eventos El Niño,” 2024, accedido: 3 de junio de 2025. [Online]. Available: http://met.igp.gob.pe/elniño/lista_eventos.html
- [96] K. V. Mardia, “Measures of multivariate skewness and kurtosis with applications,” *Biometrika*, vol. 57, no. 3, pp. 519–530, 1970.
- [97] I. González, S. Déjean, P. G. Martin, and A. Baccini, “Cca: An r package to extend canonical correlation analysis,” *Journal of Statistical Software*, vol. 23, no. 12, pp. 1–14, 2008.

- [98] E.-J. Lee, J.-Y. Chae, and J.-H. Park, “Reconstruction of Sea Level Data around the Korean Coast Using Artificial Neural Network Methods,” *Journal of Coastal Research*, vol. 95, no. sp1, pp. 1172 – 1176, 2020. [Online]. Available: <https://doi.org/10.2112/SI95-227.1>
- [99] A.-L. Balogun and N. Adebisi, “Sea level prediction using arima, svr and lstm neural network: assessing the impact of ensemble ocean-atmospheric processes on models’ accuracy,” *Geomatics, Natural Hazards and Risk*, vol. 12, no. 1, pp. 653–674, 2021. [Online]. Available: <https://doi.org/10.1080/19475705.2021.1887372>
- [100] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [101] H. Wackernagel and H. Wackernagel, “Ordinary kriging,” *Multivariate geostatistics: an introduction with applications*, pp. 79–88, 2003.
- [102] M. Oliver and R. Webster, “A tutorial guide to geostatistics: Computing and modelling variograms and kriging,” *Catena*, vol. 113, pp. 56–69, 2014.
- [103] J. Li and A. D. Heap, *A review of spatial interpolation methods for environmental scientists*. Geoscience Australia Canberra, 2008.
- [104] ———, “Spatial interpolation methods applied in the environmental sciences: A review,” *Environmental Modelling & Software*, vol. 53, pp. 173–189, 2014.
- [105] B. S. Murphy, “Pykrige: development of a kriging toolkit for python,” in *AGU fall meeting abstracts*, vol. 2014, 2014, pp. H51K–0753.
- [106] N. S.-N. Lam, “Spatial interpolation methods: a review,” *The American Cartographer*, vol. 10, no. 2, pp. 129–150, 1983.
- [107] L. Khazaz, H. J. Oulidi, S. El Moutaki, and A. Ghafri, “Comparing and evaluating probabilistic and deterministic spatial interpolation methods for groundwater level of haouz in morocco,” *Journal of Geographic Information System*, vol. 7, no. 6, pp. 631–642, 2015.
- [108] S. Shahmohammadi-Kalalagh and F. Taran, “Evaluation of the classical statistical, deterministic and geostatistical interpolation methods for estimating the groundwater level,” *International Journal of Energy and Water Resources*, vol. 5, no. 1, pp. 33–42, 2021.
- [109] G. Van Rossum and F. L. Drake Jr, *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam, 1995.
- [110] G. Matheron, “Principles of geostatistics,” *Economic geology*, vol. 58, no. 8, pp. 1246–1266, 1963.
- [111] F. Fouedjio, “Stationarity,” in *Encyclopedia of Mathematical Geosciences*. Springer, 2021, pp. 1–5.

- [112] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burrowski, P. Peterson, W. Weckesser, J. Bright *et al.*, “Scipy 1.0: fundamental algorithms for scientific computing in python,” *Nature methods*, vol. 17, no. 3, pp. 261–272, 2020.
- [113] A. S. Fotheringham, C. Brunsdon, and M. Charlton, “Geographically weighted regression,” *The Sage handbook of spatial analysis*, vol. 1, pp. 243–254, 2009.
- [114] A. Klöckner, N. Pinto, B. Catanzaro, Y. Lee, P. Ivanov, and A. Fasih, “Gpu scripting and code generation with pycuda,” 2013. [Online]. Available: <https://arxiv.org/abs/1304.5553>
- [115] M. Parry, O. Canziani, J. Palutikof, P. van der Linden, and C. Hanson, *Climate Change 2007: Impacts, Adaptation and Vulnerability*. Cambridge University Press, 01 2007, pp. 315–356.
- [116] I. de Hidráulica Ambiental de la Universidad de Cantabria (IHCantabria), “Documento temático cota de inundación: Atlas de inundación en el litoral peninsular español,” IHCantabria, Tech. Rep., 2015, accedido el 3 de junio de 2025. [Online]. Available: <https://smc.ihcantabria.es/SMC25/wp-content/uploads/2015/08/Documento-tematico-de-cota-de-inundacion.pdf>
- [117] H. Stockdon, R. Holman, P. Howd, and A. Sallenger, “Empirical parameterization of setup, swash, and runup,” *Coastal Engineering*, vol. 53, no. 7, pp. 573–588, 2006.
- [118] Organización Hidrográfica Internacional (OHI), *Manual sobre los datums hidrográficos y las referencias verticales*, Mónaco, 2000, publicación Técnica de la OHI, edición de 2000. [Online]. Available: https://iho.int/uploads/user/pubs/standards/s-44/S-44_Esp_5.0.1.pdf
- [119] L. M. V. L., J. D. I. S., and R. E. S. E., “Determinación de la cota máxima de inundación en la frontera colombo-ecuatoriana aplicando modelación numérica,” *Boletín Científico CIOH*, vol. 37, pp. 3–16, 2018. [Online]. Available: <https://ojs.dimar.mil.co/index.php/CIOH/article/view/446>
- [120] A. M. Guerrero and R. E. Sánchez Escobar, “Evaluación de la amenaza por tsunami en poblaciones del sur, centro y norte del litoral pacífico colombiano,” *Boletín Científico CIOH*, vol. 38, no. 2, pp. 15–40, 2019. [Online]. Available: <https://ojs.dimar.mil.co/index.php/CIOH/article/view/524>
- [121] M. Vogel, “Sea-level science: Understanding tides, surges, tsunamis and mean sea-level changes, by david pugh and philip woodworth,” *Contemporary Physics*, vol. 56, pp. 1–1, 01 2015.
- [122] G. Godin, *The Analysis of Tides*. University of Toronto Press, 1972. [Online]. Available: <https://books.google.com.co/books?id=3kbjAAAACAAJ>
- [123] L. J. O. Díaz, “Aplicación de un modelo hidrodinámico bidimensional para describir las corrientes y la propagación de la onda de marea en la bahía de Buenaventura,” *Boletín Científico CIOH*, vol. 12, pp. 9–21, 2005.

- [124] J. A. Orejarena Cuartas, “Determinación de la resonancia de onda larga en el pacífico colombiano y otras bahías naturales mediante modelación numérica,” *Revista de Ciencias*, vol. 23, no. 1, ene. 2019. [Online]. Available: https://revistaciencias.univalle.edu.co/index.php/revista_de_ciencias/article/view/8619
- [125] J. E. Gutiérrez Fuentes and M. G. Puentes Galindo, “Determinación del régimen medio y extremo del nivel del mar, en esteros y rios con dinámica mareal en la Bahía de Tumaco,” *Boletín Científico CIOH*, vol. 29, pp. 110–121, 2011.
- [126] Centro de Investigaciones Oceanográficas e Hidrográficas del Pacífico (CCCP), *Compilación Oceanográfica del Pacífico Colombiano. Volumen II*. Tumaco, Colombia: Dirección General Marítima - DIMAR, 2022.
- [127] R. Sánchez, M. Puentes, J. Gutiérrez, and G. Velasco, “Estimación de la amenaza por tsunami en el municipio de San Andrés de Tumaco, Pacífico colombiano, utilizando información LiDAR,” *Boletín Científico CIOH*, vol. 30, pp. 29–42, 2016.
- [128] A. Pistolesi, “Shore-parallel sediment transport by waves and tides in inner shelf waters,” Ph.D. dissertation, Massachusetts Institute of Technology, 01 2016.
- [129] Comité Multisectorial Encargado del Estudio Nacional del Fenómeno El Niño (ENFEN), “Nota técnica - informe del comité enfen sobre la situación del fenómeno el niño,” 2024, accedido: 3 de junio de 2025. [Online]. Available: http://met.igp.gob.pe/elnino/enfen/ICEN-Nota_Tecnica.pdf
- [130] Boschat, Simmonds, Purich, Cowan, and Bernardes, “On the use of composite analyses to form physical hypotheses: An example from heat wave – sst associations,” *Scientific Reports*, no. 6, 2016. [Online]. Available: <https://doi.org/10.1038/srep29599>
- [131] Laken, B.A. and Čalogović, J., “Composite analysis with monte carlo methods: an example with cosmic rays and clouds,” *J. Space Weather Space Clim.*, vol. 3, p. A29, 2013. [Online]. Available: <https://doi.org/10.1051/swsc/2013051>
- [132] E. Schulz, M. Speekenbrink, and A. Krause, “A tutorial on gaussian process regression: Modelling, exploring, and exploiting functions,” *Journal of mathematical psychology*, vol. 85, pp. 1–16, 2018.