

FORTHCOMING 62J12-10-23-01**ESTIMACIÓN ESPACIAL DE LA LLUVIA DIARIA EN UNA REGIÓN DE CUBA USANDO MODELOS LINEALES GENERALIZADOS Y COVARIABLES TOPOGRÁFICAS**

Abel Centella-Artola^{1*}, Roberto Serrano-Notivoli^{2**}, Cecilia Fonseca-Rivera^{3*}, Arnoldo Bezanilla-Morlot^{4*}, Maibys Sierra-Lorenzo^{5*}

* Instituto de Meteorología, Loma de Casa Blanca, Regla, La Habana 11700, Cuba

** Departamento de Geografía y Ordenación del Territorio, Instituto Universitario de Ciencias Ambientales, Universidad de Zaragoza, Zaragoza, 50009, España

ABSTRACT

This paper presents the results of several simulations performed to make spatial estimates of daily rainfall in a region of Cuba at a spatial resolution of 0.00097° (~1 km). The methodology used is based on generalised linear models (GLM), using different combinations of pairs of covariates associated with terrain characteristics in addition to elevation, latitude, and longitude. It was determined that among the eight topographic covariates initially calculated the terrain roughness index, the cost distance, and the terrain orientation, were the best to be included in the models. The results have shown that in simulations where the models included a couple of terrain covariates, the precision of the estimate was better than when longitude, latitude, and elevation were just considered. It is also shown that by increasing the spatial resolution to 1 km, the results are more precise than those achieved in previous studies using 3 km. This work is a valuable contribution to research activities related to climate variability and change in Cuba.

KEYWORDS: Cuba, rainfall, Generalised Linear Models, R software

MSC: 62J12

RESUMEN

Se presentan los resultados de varias simulaciones realizadas, para hacer estimaciones espaciales de la lluvia diaria en una zona del extremo occidental de Cuba, a una resolución espacial de 0.00097° (~1 km). La metodología se basa en el empleo de modelos lineales generalizados con diferentes combinaciones de pares de covariables asociadas con las características del terreno, además de la altura, la latitud y la longitud. Se determinó que, de las ocho covariables topográficas inicialmente calculadas, el índice de rugosidad del terreno, la distancia de costo y la orientación del terreno, fueron las mejores. Los resultados obtenidos muestran que, en todas las simulaciones donde el modelo incluyó un par de covariables del terreno, la precisión de la estimación fue mejor que cuando sólo se utilizó la longitud, latitud y altura. También se demuestra que, al incrementar la resolución espacial a 1 km, los resultados son mejores que los que se obtienen a 3 km. Este trabajo es una contribución valiosa al proceso de actualización del mapa isoyético de Cuba, así como a las investigaciones relacionadas con la variabilidad y el cambio climático en Cuba.

PALABRAS CLAVES: Cuba, Precipitación, Modelos Lineales Generalizados, R software

1. INTRODUCCIÓN

La lluvia es el elemento climático que tiene mayor variación espacial y temporal. Por eso, la estimación espacial de su ocurrencia y de su cantidad en aquellas áreas donde no hay mediciones directas, es un gran reto científico. Se conoce que las mediciones de precipitación utilizando pluviómetros ubicados en tierra continúan siendo las más precisas, pero a la misma vez hay serias limitaciones para tomar las muestras en lugares apartados, de difícil acceso y con topografía compleja. Existen otras técnicas, como el uso de la información de los satélites, que permiten obtener estimados continuos en el espacio a muy alta resolución, pero dichas estimaciones aun no son tan precisas como las mediciones directas con los pluviómetros sobre tierra.

La lluvia tiene dos características particulares que influyen en su modelación. Por un lado, es una variable intermitente (ocurre o no ocurre) y cuando ocurre, adquiere magnitudes que pueden variar sustancialmente tanto en tiempo como en espacio. De hecho, la literatura internacional reporta una amplia variedad de métodos para hacer

estimaciones espaciales de esta variable utilizando métodos no geoestadísticos, geoestadísticos y mezclas de ambos tipos ([1], [2], [3]). Por otro lado, las series de lluvia diaria presentan una elevada frecuencia de valores cero, pues como se conoce, en la mayoría de los lugares, la cantidad de días que llueve es mucho menor a la que no llueve. Por esta razón, los métodos de estimación que se utilicen deben considerar esta característica particular.

A pesar de los retos anteriores, existe la constante necesidad de crear conjuntos de datos con estimados precisos en forma de rejilla regular, con el objetivo de satisfacer la demanda de múltiples aplicaciones científicas y prácticas. Ejemplos de esas necesidades son la modelación de procesos hidrológicos, la evaluación de los recursos hídricos de una región, o el estudio de las variaciones y cambios en el clima. En todos los casos se requiere de información espacial continua, en lugar de la que se tiene con estaciones de medición irregularmente distribuidas.

Recientemente, Centella et. al [4] desarrollaron el primer conjunto de datos de lluvia diaria para Cuba en forma de rejilla regular con resolución espacial de 3 km, para el período 1961-2008. Los autores emplearon modelos lineales generalizados y utilizaron covariables como la longitud, la latitud y la altura del terreno. Con este método se realizó el control de calidad de las series de las estaciones, el rellenado de los datos faltantes y finalmente se estimaron los valores hacia una rejilla regular con la resolución mencionada. En ese trabajo se utilizó el paquete *reddPrec* de R software [5], el cual se sustenta en la metodología desarrollada por Serrano-Notivoli, et.al [6]. Los resultados fueron satisfactorios, pues al evaluar la habilidad del esquema de predicción para los días con lluvia y sin lluvia, se obtuvo que los porcentajes de Verdadero Cero, Falso Cero, Verdadero Positivo y Falso Positivo fueron de 92.43%, 7.57%, 70.97% y 29.03%, respectivamente. Así mismo, la capacidad de estimación de las magnitudes de lluvia (valores mayores que cero) fue muy buena, aunque en algunas regiones de Cuba, donde la densidad de pluviómetros era baja, la capacidad predictiva resultó menor aunque con valores de correlación de Pearson cercanos a 0.6.

Si bien los resultados alcanzados fueron muy satisfactorios y útiles, los mismos pueden ser mejorados en la precisión de los estimados, además de actualizados hasta fechas más recientes. De esta forma se puede hacer una contribución importante a la actualización del Mapa Isoyético de Cuba y al desarrollo de otras investigaciones relacionadas con las variaciones del clima y el pronóstico estacional de la lluvia. Se conoce que, a la luz de las variaciones y cambios observados en el clima de Cuba, la actualización del Mapa Isoyético facilitará una mejor evaluación de los recursos hídricos nacionales.

A partir de los antecedentes mencionados, surge la siguiente pregunta ¿Se puede obtener una rejilla de lluvia diaria más precisa y con mayor resolución espacial utilizando la misma metodología, pero empleando covariables adicionales? En esencia, esta es la hipótesis del presente trabajo.

Tomando en consideración lo anterior, el objetivo de este trabajo es “Evaluar las posibilidades de mejorar la precisión de la estimación de la rejilla de valores de lluvia diaria en Cuba, incrementando la resolución espacial a 1 km, e incluyendo la información de factores topográficos adicionales a la altura”. La evaluación se realizará en forma de estudio de caso, considerando solamente una porción del territorio cubano y una muestra aleatoria de días seleccionada entre enero de 1961 y diciembre de 2008.

A continuación, se describe la metodología utilizada para alcanzar el objetivo propuesto. Seguidamente se presentan los resultados y luego se realiza una discusión de los mismos. Finalmente se presentan las conclusiones, incluyendo las recomendaciones que resulten pertinentes para investigaciones futuras.

2. MATERIALES Y MÉTODOS

El abordaje metodológico incluye las siguientes tareas: i) Descripción del conjunto de datos diarios de lluvia empleado en el estudio; ii) Determinación, cálculo y análisis de los índices o covariables relacionados con la topografía en el área de estudio y; iii) Modelación de la lluvia diaria con diferentes combinaciones de covariables para la estimación de los valores por puntos de rejilla. Estas tareas se describen a continuación:

2.1 Descripción del conjunto de datos diarios de lluvia

El conjunto de datos que se utiliza en este trabajo es un subconjunto del empleado por Centella et. al [4] y se corresponde con las series de lluvia diaria de 48 estaciones pluviométricas localizadas en el extremo occidental de Cuba, para el período de enero de 1961 a diciembre de 2008 (Figura 1). Esta zona tiene una baja densidad de pluviómetros con información en comparación con el resto del país, por eso se considera que el reto de hacer las estimaciones espaciales resulta más interesante. Por otro lado, el área identificada para el estudio ocupa el cuarto

puesto dentro de las zonas de topografía más compleja de Cuba, aunque ciertamente Cuba no es un país que se destaque por la complejidad en su relieve.

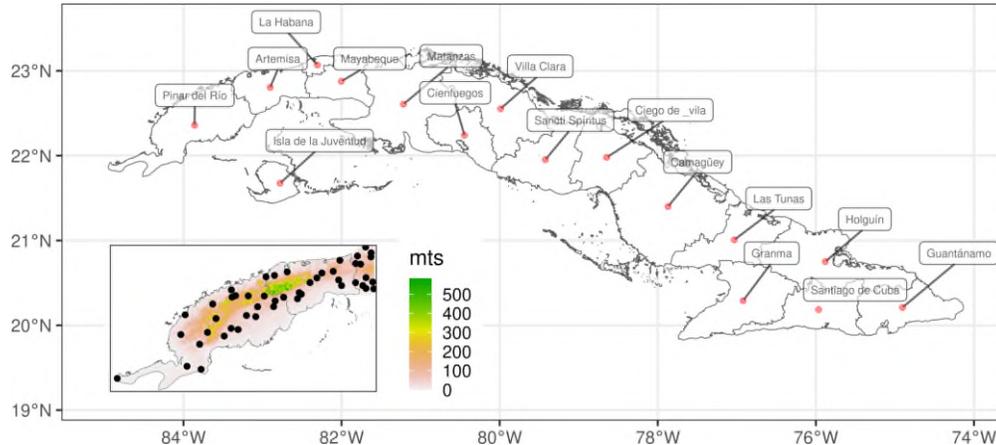


Figura 1: Mapa de Cuba y sus provincias. La figura insertada muestra el área de estudio con su relieve (altura en metros) y la localización de los pluviómetros

El resumen estadístico de la muestra utilizada indica que los valores medios están entre 2.7 y 4.3 mm/día, con desviaciones estándar cercanas a 12 mm/día. Las magnitudes máximas de lluvia registradas en la región de estudio varían notablemente entre las estaciones (214.0 mm - 675.7 mm), evidenciando las diferencias en el comportamiento espacial de los eventos lluviosos, así como las distintas características de las localidades donde se registraron esos valores extremos. La información faltante en las series de datos oscila entre el 6 % y el 10 %, como resultado del proceso de control de calidad a que fueron sometidas las series, pues Centella et al [4] removieron los datos que eran sospechosos.

La Figura 2 muestra los patrones de variación anual de los totales mensuales para cada pluviómetro y los valores promedios diarios de todos los pluviómetros. Se puede observar el carácter estacional de la variable, con la existencia de 2 períodos bien definidos. El período de mayo-octubre se denomina período lluvioso [7] y es donde se registran los mayores acumulados, mientras el que va de noviembre a abril, se le define como período poco lluvioso [7]. Se destaca además que en los meses de julio y agosto se produce una reducción relativa de los valores, haciendo que el régimen de lluvias en la región tenga un carácter bimodal. También se puede distinguir que las variaciones espaciales de los valores medios mensuales en el período lluvioso son mayores, mostrando asimetría positiva (Figura 2a).

La Figura 2b refleja un comportamiento similar, pero en este caso los valores representan la media de todos los pluviómetros por día. Se puede ver cómo cambia la distribución de la lluvia diaria entre los meses de un período y otro e incluso, entre los meses de un mismo período. Nótese que aun cuando abril y noviembre forman parte del período poco lluvioso, los valores se distribuyen ligeramente distintos a diciembre-marzo. Por esta y otras razones a estos meses se les cataloga como meses de transición. Así mismo, la forma de la distribución de los valores en el mes de julio, es más cercana a la de los meses del período poco lluvioso, aunque los acumulados diarios son mayores. Las grandes colas a la derecha, con carácter multimodal, principalmente en los meses de mayo a octubre, reflejan la influencia de los diferentes procesos que originan los mayores valores de lluvia diaria. Entre esos procesos están la actividad convectiva, las ondas del este y los ciclones tropicales.

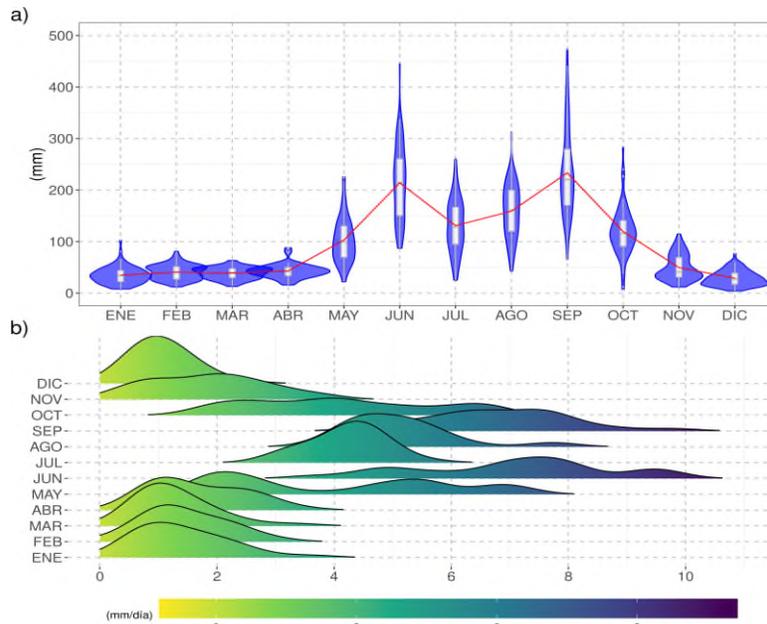


Figura 2: Régimen de variación anual de la lluvia en la región de estudio según las magnitudes registradas por los pluviómetros. a) Patrón de valores totales mensuales para cada pluviómetro y b) Patrón de los valores diarios promedio de todos los pluviómetros

2.2 Cálculo de los índices relacionados con la topografía

Se conoce que diferentes rasgos del relieve (altura, exposición, pendiente, rugosidad, etc.) influyen en mayor o menor medida sobre la distribución espacial de las variables climáticas. Por esta razón, a partir de los datos de altura del Modelo de Elevación Digital (DEM) de Cuba con resolución de 1x1 km, se decidió calcular distintos indicadores topográficos o covariables que pueden describir las características del terreno. Las covariables obtenidas son las siguientes:

- Índice de rugosidad del terreno (TRI): Media de las diferencias absolutas entre una celda y sus 8 vecinas.
- Rugosidad (roughness): Diferencia entre los valores máximo y mínimo de una celda y sus 8 vecinas
- Pendiente (slope): Valor de la pendiente de la celda
- Aspecto (aspect): Orientación de la pendiente de la celda (grados en el sentido del reloj)
- Distancia de costo (cost_dist): distancia de cada punto a la costa, pero asumiendo el costo de atravesar el relieve. El costo está dado por una superficie de fricción, la cual se multiplica por la distancia para obtener el valor final.
- Distancia a la costa (DIST2COAST): Distancia de cada punto a la línea de costa.

Todas las variables anteriores fueron guardadas en ficheros raster independientes con formato GeoTIFF, los cuales son utilizados posteriormente en la etapa de modelación de la lluvia diaria.

2.3 Modelación espacial de la lluvia diaria

Como se mencionó anteriormente, en este trabajo se emplea la misma metodología propuesta por Serrano-Notivoli et al [6], pero incorporando nuevos predictores e incrementando la resolución espacial de las estimaciones. Este método está implementado en el paquete *reddPrec* (cran.r-project.org/web/packages/reddPrec) y con el mismo se pueden hacer 3 tareas: 1) Control de calidad; 2) Rellenado de series y; 3) Estimación de la lluvia hacia otros puntos

del espacio (p.ej. Una rejilla espacial regular). Los estimados de la lluvia (llamados “valores de referencia” o RV) se obtienen para cada día de forma independiente, empleando la información de las 10 estaciones más cercanas (NNS). En el caso del control de calidad, los estimados se comparan con los registrados en cada ubicación, y en función de varios criterios (ver documentación del paquete *reddPrec*), se decide clasificar el dato observado como sospechoso o no. En el rellenado de las series, donde se asume que se emplean series validadas, los RV sirven para sustituir los valores faltantes. En este paso los RV siempre son calculados sin la participación de la estación candidata, por eso se llega a tener series de pares de valores observados y estimados, que permiten realizar una validación cruzada del tipo LOOCV (leave one out cross-validation). Finalmente, para la estimación de nuevos puntos en el espacio (rejillas), el cálculo de los RV es similar y funciona mejor si la información de entrada ya fue validada y/o rellenada previamente. En todos los casos, los resultados del método son mejores (peores) cuando la densidad de las estaciones es alta (baja).

En síntesis, el valor de RV se determina a partir de dos estimados: i) una predicción binomial (BP) de la probabilidad de ocurrencia de un día húmedo; y ii) una estimación de la magnitud (MP) de precipitación. BP emplea los valores de las 10 NNS codificados como una variable binomial (día de lluvia o día sin lluvia) para calcular la probabilidad de ocurrencia de lluvia en el día i y localidad l ($Bp_{i,l}$):

$$BP_{i,l} = \beta_{0_{i,l}} + \beta_{1_{i,l}}alt_l + \beta_{2_{i,l}}lon_l + \beta_{3_{i,l}}lat_l + \varepsilon_{1_{i,l}} \quad (1)$$

Donde: $\beta_{n_{0_{i,l}}}$ son los coeficientes de regresión; alt_l , lat_l y lon_l , son la altura, latitud y longitud de las NN y $\varepsilon_{1_{i,l}}$ es el error. El modelo es implementado en R utilizando la función *glm()* con la familia binomial.

En el caso de MP también se emplea la función *glm()* pero con la familia quasi-binomial, utilizando los valores de lluvia (Pcp) previamente son transformados con la fórmula 2 antes de ajustar el modelo (3).

$$Pcp_{i,l} = \frac{NN_i - (min_i - (Q2_i - Q1_i))}{(max_i + (Q3_i - Q1_i)) - (min_i - (Q2_i - Q1_i))} \quad (2)$$

Donde: $Pcp_{i,l}$ son los valores transformados para el día i y la localidad l ; NN_i son los valores de lluvia en las estaciones NNS; min , max , $Q1$, $Q2$ y $Q3$ son los valores mínimo, máximo, y los cuartiles 1, 2 y 3 de los valores de las NN en el día i .

$$MP_{i,l} = \beta_{4_{i,l}} + \beta_{5_{i,l}}alt_l + \beta_{6_{i,l}}lon_l + \beta_{7_{i,l}}lat_l + \varepsilon_{2_{i,l}} \quad (3)$$

Donde: $\beta_{n_{4_{i,l}}}$ son los coeficientes de regresión; alt_l , lat_l y lon_l , son la altura, latitud y longitud de las NN y $\varepsilon_{2_{i,l}}$ es el error.

El valor final de RV es determinado por la combinación de MP y BP, considerando un umbral de BP mayor o igual que 0.5 para determinar un día húmedo. Así, si $BP \geq 0.5$, RV es igual a MP, de lo contrario RV es cero.

De acuerdo con los autores, en el desarrollo del método de los RV se utilizó la regresión lineal múltiple asintótica adaptativa [6] tomando en cuenta que otros métodos generaban sobrestimaciones importantes. De hecho, una característica especial de este método es que para cada día y localidad, siempre se ajusta un modelo particular e independiente. Es decir, no se considera que las relaciones entre un pluviómetro y sus vecinos se mantienen constantes a lo largo del tiempo. Esta característica tiene la ventaja de que siempre se puede utilizar toda la información disponible en un día determinado, sin tomar en consideración el tamaño de las series de observaciones o que las observaciones válidas de las estaciones se correspondan exactamente en el tiempo. Por otro lado, como en la escala de tiempo diaria la distribución espacial de la lluvia es muy irregular y cambia de forma drástica de un día a otro, el uso de modelos adaptados a cada día y lugar puede ofrecer mejores resultados en la estimación de la lluvia local.

2.4 Determinación de la habilidad del método de los RV utilizando predictores adicionales

Esta etapa del trabajo se organizó con el fin de responder las dos preguntas siguientes: 1) ¿Se puede mejorar la precisión del método utilizado por Centella et. al [4], si se incorporan nuevos predictores asociados a la topografía? y 2) ¿Cuál sería la combinación de covariables que mejores resultados produce? Con este propósito se realizan 4 simulaciones para estimar los valores de lluvia diaria en los puntos de una rejilla regular de 1x1 km. Una simulación se realizó con los predictores originales (latitud, longitud y altura), mientras que en las tres restantes se adicionó un par de covariables (indicadores topográficos) cada vez. La cantidad de simulaciones a realizar quedó determinada por los resultados del análisis de las covariables .

Las simulaciones se realizaron utilizando una muestra aleatoria de 5 259 días, logrando que la misma incluyera una diversidad de casos similares a los de la muestra total (p. ej. desde días con valores altos de lluvia en la mayor parte de los pluviómetros, hasta días en que no se registraron precipitaciones en toda la zona de estudio). También debe aclararse que cuando se utilizó solamente la latitud, longitud y altura, el modelo de regresión logística multivariado incluyó la información de 10 NNS. Sin embargo, en los otros casos, al incluirse las otras 2 covariables, se emplearon 15 NNS para evitar la pérdida de grados de libertad. Desde luego que esto se hace a riesgo de que se incluya el efecto de estaciones más lejanas y que la estimación sea entonces más regional que local, aunque se tenga una alta resolución espacial de 1 km.

Como se mencionó antes, en la metodología de los RV no se obtiene un único modelo para la estimación espacial de la lluvia, sino que para cada día y punto se ajustan modelos diferentes. De esta forma, la manera de evaluar las estimaciones de cada simulación se realiza mediante la comparación los estimados con las observaciones, determinando así su habilidad o capacidad predictiva. Para esto se tomaron las rejillas con las estimaciones de cada simulación y se extrajeron los valores de las celdas correspondientes a las coordenadas de los pluviómetros. Los valores estimados extraídos, fueron entonces comparados con las observaciones de cada pluviómetro en los días de la muestra aleatoriamente seleccionada. Se construyeron diagramas de dispersión para evaluar la habilidad general de cada simulación, así como tablas de contingencia para reflejar la precisión de la predicción de los días con y sin lluvia (mayores que cero y menores o iguales a cero, respectivamente).

Para hacer una evaluación más completa, también se calcularon diferentes métricas generales como el sesgo (BIAS), el error medio absoluto (MAE), el error medio cuadrático (MSE), la raíz del error medio cuadrático (RMSE), el coeficiente de correlación de Pearson (r), la razón entre los valores medios estimados y observados (b), la razón entre los coeficientes de variación de los valores estimados y observados (v) y el índice de eficiencia modificado de Kling-Gupta (KGE) [8]. El KGE es un indicador adimensional que se emplea con mucha frecuencia en el campo de la hidrología y la climatología, para realizar evaluaciones de modelos o hacer comparaciones de bases de datos de variables relacionadas con el ciclo hidrológico. La formulación del mismo es como sigue:

$$KGE = 1 - \sqrt{(r - 1)^2 + (b - 1)^2 + (v - 1)^2} \quad (4)$$

$$r = \frac{\sum_i^n (O_i - \bar{O})(S_i - \bar{S})}{\sqrt{\sum_i^n (O_i - \bar{O})^2} \sqrt{\sum_i^n (S_i - \bar{S})^2}} \quad (5)$$

$$b = \frac{\bar{S}}{\bar{O}} \quad (6)$$

$$v = \frac{CV_S}{CV_O} \quad (7)$$

Donde: n es el número de observaciones, O_i y S_i son los datos observados y estimados, respectivamente en la observación i ; \bar{O} , \bar{S} , CV_O y CV_S son los valores medios y del coeficiente de variación de las observaciones y los estimados, respectivamente. Según Kling et al [8] con el uso del coeficiente de variación en lugar de la desviación estándar, se asegura que los elementos representados por el sesgo y la variabilidad no estén correlacionados de manera cruzada.

Por su propia formulación se evidencia que el valor óptimo del índice KGE y de sus distintas componentes (r , b y v) es la unidad. En esencia, r mide la correlación lineal entre las observaciones y los estimados, b refleja la tendencia de los valores estimados a ser mayores o menores que las observaciones, mientras que v refleja algo similar para el caso de la variabilidad.

2.5 Evaluación de la precisión de las estimaciones con el incremento de la resolución espacial

Con el objetivo de conocer si el incremento de la resolución tiene un impacto positivo en los resultados de la estimación, se extrajeron los valores puntos de rejilla de la base de datos de 3 km obtenida por Centella et. al [4] que se corresponden con las coordenadas de los 48 pluviómetros. La comparación de los valores de las estaciones con los de sus correspondientes puntos de rejilla, permite tener una idea de la calidad de las estimaciones a la resolución de 3 km sobre el área geográfica utilizada en el presente trabajo. Entonces, si se contrastan estos resultados con los de la simulación 4, se podrá conocer la influencia del incremento de la resolución.

3. RESULTADOS Y DISCUSIÓN

3.1 Análisis de las covariables relacionados con la topografía

Al observar la Figura 3 se puede tener una idea sobre el comportamiento espacial de los indicadores relacionados con las características del terreno, incluyendo la altura (heigh), que no fue calculada sino extraída directamente del DEM. Se puede apreciar que la estructura espacial de varios indicadores resulta bastante similar (p.ej. elevación, TRI, pendiente, rugosidad) sugiriendo que su inclusión en el modelo generaría problemas de multicolinealidad. Para evaluar objetivamente las relaciones existentes entre las covariables, se extrajeron los valores de las mismas en las celdas donde se ubican los pluviómetros y se realizó una comparación multivariada (Figura 4). El balance de los resultados anteriores permite indicar que:

- Entre las covariables heigh, TRI, pendiente y rugosidad existe una relación positiva y fuerte, toda vez que las mismas adquieren sus mayores y menores valores en las zonas altas y bajas, respectivamente. Nótese también que las estructuras espaciales de esas variables son muy similares (Fig 3 a, b, d y f);
- Las relaciones de cost_dist y DIST2COST con las otras variables son pobres, aunque la correlación entre ellas es alta y estadísticamente significativa. Esa relación es lógica si se comprende que cost-dist refleja el “costo de atravesar el relieve” el cual, en el caso particular de esta región, es mayor en las zonas alejadas de la costa, donde precisamente se localizan las mayores alturas y el relieve más complejo.

Este análisis sirve de base para apoyar las 4 simulaciones realizadas, combinando las covariables longitud, latitud y altura con TRI, cost-dist y aspect, tal y como se muestra en la Tabla 1.

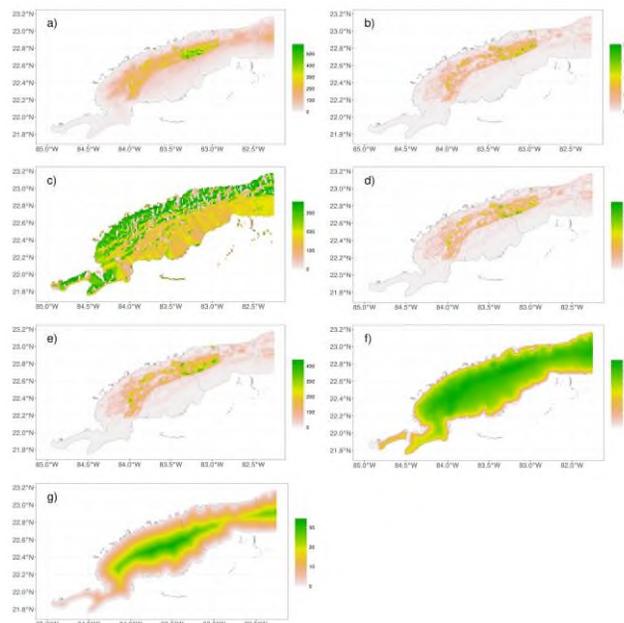


Figura 3. Variación espacial de las características del terreno en la zona de estudio. a) altura (heigh) (mts), b) Índice de rugosidad del terreno (TRI), c) Aspecto (aspect), d) Pendiente (slope), e) Rugosidad (roughness), f) Distancia de costo (costdist), g) Distancia a la costa (DIST2COAST)(km) . Por simplicidad se utiliza la misma escala de colores en todos los mapas.

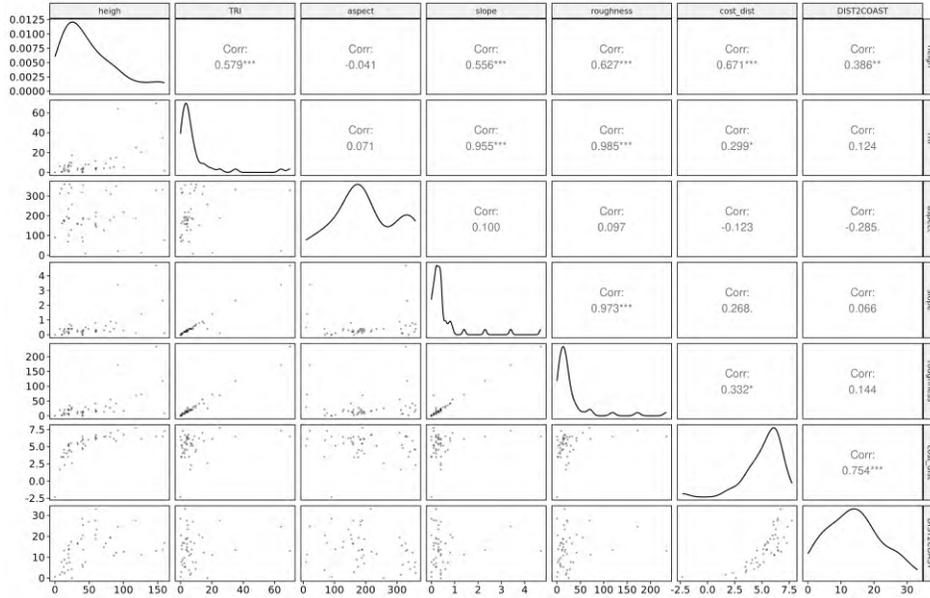


Figura 4: Comparación multivariada de los datos de las covariables relacionadas con las características del terreno. El triángulo superior muestra los coeficientes de correlación de Pearson con la significación estadística (*, ** y ***, se refiere a los niveles de significación 0.01, 0.05 y 0.1, respectivamente), la diagonal refleja la distribución de cada variable y el triángulo inferior muestra los diagramas de dispersión correspondientes.

Tabla 1: Simulaciones, covariables y cantidad de estaciones vecinas (NNS)

	COVAR1	COVAR2	COVAR3	COVAR4	COVAR5	NNS
Simulación 1	Latitud	Longitud	Altura	TRI	Cost-dist	15
Simulación 2	Latitud	Longitud	Altura	aspect	TRI	15
Simulación 3	Latitud	Longitud	Altura	aspect	Cost-dist	15
Simulación 4	Latitud	Longitud	Altura			10

3.2 Evaluación de las simulaciones con covariables adicionales

De acuerdo con los resultados mostrados en los diagramas de dispersión de la Figura 5, se puede indicar que en todas las simulaciones donde se combinaron las variables adicionales, los resultados fueron mejores que cuando se consideró solamente la posición y la altura. Como regla, los valores estimados (RV) tienden a ser menores que las magnitudes observadas.

En la Tabla 3 se puede notar que la precisión en la estimación de días secos y húmedos es bastante buena en todas las simulaciones con covariables adicionales. Las principales diferencias respecto a la simulación 4 se observan en la reducción de FP y el consecuente incremento de los VP. Si bien en algunos casos las simulaciones 1-3 presentan resultados muy cercanos, los valores de la 3 son ligeramente mejores.

En la Figura 6 se distinguen mejor las ligeras diferencias entre las tres primeras simulaciones respecto a la 4. En esencia, esas simulaciones logran una estimación más precisa de los días secos, principalmente en los meses más lluviosos, y al mismo tiempo tienen mucha mayor precisión en la predicción de los días húmedos, durante los meses más secos. Por otro lado, en el mes de julio los valores de las diferencias entre las tres primeras simulaciones y la cuarta son más parecidas a los observados en los meses del período seco. Nótese que aun cuando el mes de julio forma parte de los que integran la temporada lluviosa en Cuba, es el de menores acumulados de lluvia y la distribución de los valores de lluvia diaria en el mismo se asemeja más a la de los meses del período poco lluvioso (Figura 2b). En esencia, los resultados anteriores sugieren que la incorporación de detalles relacionados con la topografía tiene una influencia positiva en incrementar la precisión de la estimación de los días secos y húmedos.

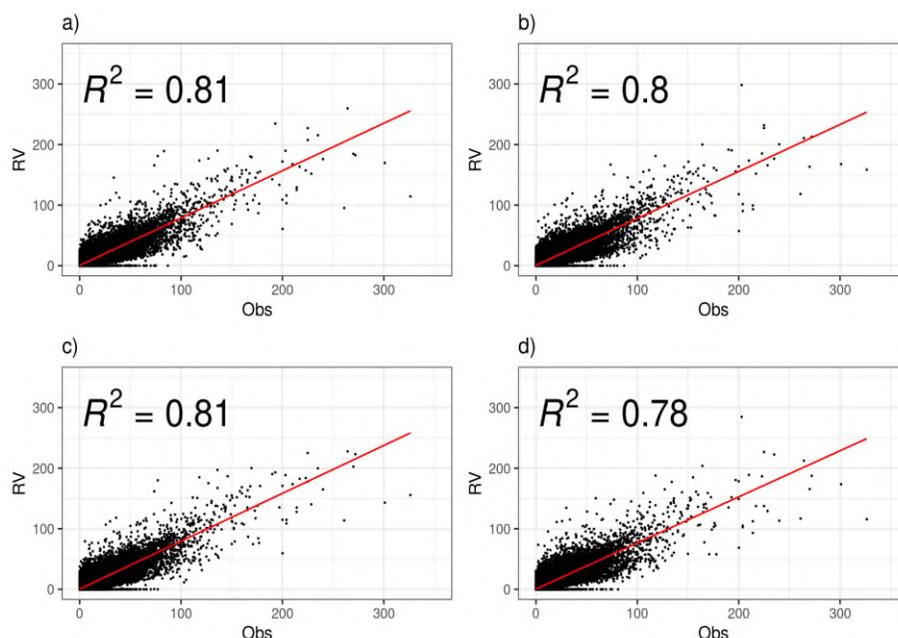


Figura 5: Diagramas de dispersión de los valores observados y estimados para todas las estaciones. Los paneles a), b), c) y d) representan las simulaciones 1, 2, 3 y 4 (Tabla 2).

Tabla 3: Evaluación de la precisión en la estimación de días secos y húmedos (ceros y unos). Las etiquetas VC, FC, VP y FP, significan Verdadero Cero, Falso Cero, Verdadero Positivo y Falso Positivo. Los números asociados a las etiquetas reflejan la simulación (Tabla 2)

MESES	VC1	FC1	VP1	FP1	VC2	FC2	VP2	FP2	VC3	FC3	VP3	FP3	VC4	FC4	VP4	FP4
TODOS	96.52	3.48	92.65	7.35	96.47	3.53	92.64	7.36	96.62	3.38	92.95	7.05	96.05	3.95	91.21	8.79
ENE	98.26	1.74	90.70	9.3	98.26	1.74	90.51	9.49	98.27	1.73	90.28	9.72	98.10	1.90	88.08	11.92
FEB	97.82	2.18	91.70	8.3	97.70	2.30	91.25	8.75	97.83	2.17	91.59	8.41	97.54	2.46	89.30	10.70
MAR	98.74	1.26	91.06	8.94	98.66	1.34	91.58	8.42	98.75	1.25	91.77	8.23	98.33	1.67	88.72	11.28
ABR	98.73	1.27	95.43	4.57	98.67	1.33	94.82	5.18	98.84	1.16	95.34	4.66	98.53	1.47	93.72	6.28
MAY	96.37	3.63	94.16	5.84	96.24	3.76	94.17	5.83	96.49	3.51	94.41	5.59	96.00	4.00	92.80	7.20
JUN	93.06	6.94	93.56	6.44	93.18	6.82	93.85	6.15	93.27	6.73	93.76	6.24	92.51	7.49	92.85	7.15
JUL	93.95	6.05	91.62	8.38	93.82	6.18	91.59	8.41	94.12	5.88	92.32	7.68	92.75	7.25	89.35	10.65
AGO	93.14	6.86	92.04	7.96	93.38	6.62	92.02	7.98	93.63	6.37	92.46	7.54	92.87	7.13	91.17	8.83
SEP	92.16	7.84	93.24	6.76	92.04	7.96	93.28	6.72	92.41	7.59	93.59	6.41	91.16	8.84	92.21	7.79
OCT	95.33	4.67	92.83	7.17	95.03	4.97	92.70	7.30	95.06	4.94	93.00	7.00	94.35	5.65	91.41	8.59
NOV	97.52	2.48	91.23	8.77	97.42	2.58	90.78	9.22	97.53	2.47	91.28	8.72	97.00	3.00	89.00	11.00
DIC	98.13	1.87	91.44	8.56	98.17	1.83	91.91	8.09	98.36	1.64	92.32	7.68	97.93	2.07	90.15	9.85

En la Tabla 4 aparecen los resultados de las métricas que permiten evaluar la estimación de la magnitud de la precipitación, las cuales fueron calculadas filtrando los valores observados mayores que cero. Nuevamente, todos los indicadores reflejan una mejor habilidad para las simulaciones 1, 2 y 3, mientras que la 3 vuelve a mostrar, muy ligeramente, los números más favorables.

Como se ha podido mostrar hasta aquí, las simulaciones con covariables adicionales ofrecen mejores resultados. Entre ellas, la simulación 3, y luego la 1, aventajan a la 2. Sin embargo, es difícil decidir sobre la conveniencia final de emplear la variante 3 sobre la 1, si se toma en consideración que:

1. La región de estudio no posee un relieve de características complejas, pues en la misma predominan las alturas por debajo de 200 m y sólo hay áreas muy limitadas con elevaciones superiores a 300 o 400 m.
2. La densidad de estaciones utilizada en el estudio es baja y el 90% de ellas se localiza en alturas inferiores a los 100 m. Sólo 5 pluviómetros están en alturas entre 140 y 150 m.
3. Al calcular las métricas presentadas en la Tabla 4, pero utilizando solamente los pluviómetros localizados en alturas sobre los 100 m (Tabla 5) se puede identificar que las simulaciones 2 y 1 pasan a ser las de mejores resultados. Esto puede indicar que, con el incremento de la altura, TRI se vuelve más importante en el incremento de la precisión. Sin embargo, como aquí sólo se tiene una cantidad muy pequeña de estaciones, resulta difícil poder llegar a una conclusión robusta al respecto.

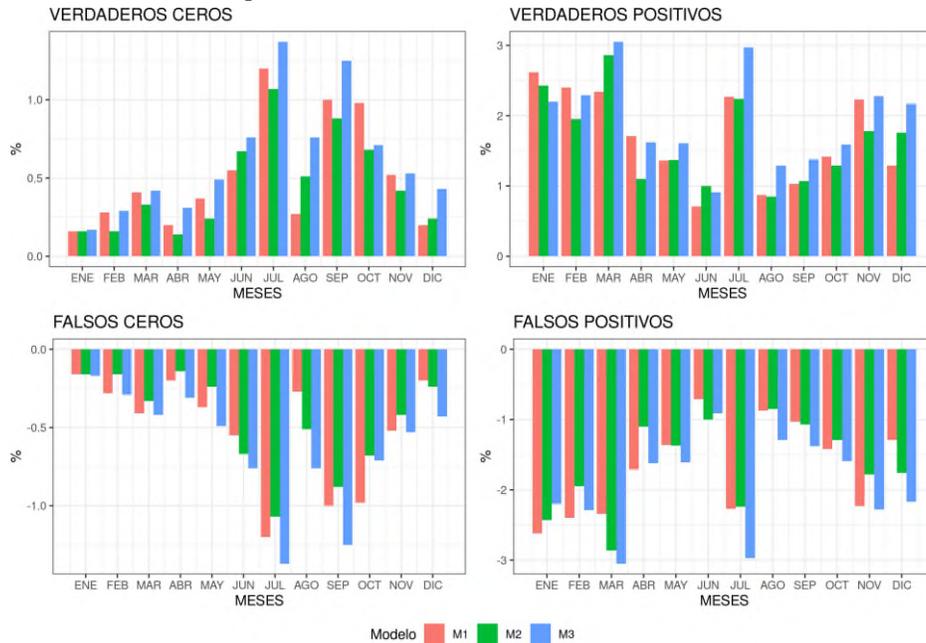


Figura 6: Diferencias de los porcentajes de VC, FC, VP y FP entre las simulaciones 1, 2 y 3 (M1, M2, y M3, respectivamente) vs la simulación 4.

Tabla 4: Valores de BIAS, MAE, MSE, RMSE, r , b , v y KGE calculados para cada simulación, considerando los días donde la lluvia observada fue superior a cero.

	BIAS	MAE	MSE	RMSE	r	b	v	KGE
Simulación 1	-1.6599	5.6403	97.3430	9.8663	0.8549	0.8836	0.9810	0.8130
Simulación 2	-1.8274	5.7012	101.0254	10.0511	0.8497	0.8719	0.9880	0.8021
Simulación 3	-1.5962	5.5490	93.6703	9.6783	0.8606	0.8881	0.9806	0.8202
Simulación 4	-1.8444	6.0041	108.6268	10.4224	0.8370	0.8707	0.9767	0.7906

Tabla 5: Valores de BIAS, MAE, MSE, RMSE, r , b , v y KGE calculados para cada simulación, considerando los días donde la lluvia observada fue superior a cero y las estaciones localizadas en alturas de 100 m o más.

	BIAS	MAE	MSE	RMSE	r	b	v	KGE
Simulacion1	-0.4166	3.6439	44.8066	6.6938	0.9321	0.9713	0.9443	0.9076
Simulacion2	-0.5856	3.3993	40.8518	6.3915	0.9385	0.9597	0.9715	0.9212

Simulacion3	-1.2156	4.7571	68.2106	8.2590	0.8968	0.9164	0.9499	0.8581
Simulacion4	-2.4284	6.4584	123.5841	11.1168	0.8085	0.8330	0.9613	0.7430

De acuerdo con lo anterior, la decisión de la inclusión de una u otra covariable, puede requerir de investigaciones futuras sobre este tema. Por ejemplo, hacer nuevas simulaciones en la misma región, pero con un número mayor de estaciones pluviométricas localizadas en rangos de alturas mayores, o desarrollar una investigación similar a la actual en una región de relieve más complejo y con mayor densidad de estaciones, pero utilizando solamente las simulaciones 1-3. También podrían investigarse otras ideas, como combinar ambas estimaciones ponderando los resultados de las mismas en función de las variaciones en la complejidad del terreno.

Finalmente, con el objetivo de apreciar y comprender mejor las diferencias espaciales entre las simulaciones realizadas, se construyeron los mapas donde se muestra la distribución del valor medio de la simulación 4 y las diferencias de las simulaciones 1, 2 y 3 con 4 (Figura 7). Se puede apreciar que las estructuras espaciales de 1 y 3 son muy parecidas, con zonas de valores negativos más intensos que en la simulación 2. Esto quiere decir que la distribución representada por 2 resulta más cercana a la 4, excepto en la región centro-sur del área de estudio, donde se manifiesta una sobrestimación importante en la simulación 2. Los resultados menos favorables de la simulación 4 quedan aún más claros al ver que los estimados de lluvia llegan a ser superiores a 7 mm/día, en las regiones más elevadas. Esa cantidad representa alrededor de 2 500 mm de precipitación acumulada en el año y es superior, en más de 600 mm, que los estimados de las simulaciones 1 y 3. De acuerdo con los mapas de la media histórica de lluvia anual [7] y [9], los máximos valores que se observan en esta región están en el rango de 1800 a 2000 mm, por lo que es una sobrestimación muy notable, del orden de 500 – 700 mm. Sin importar cuán significativa puede ser esa diferencia desde el punto de vista estadístico, se puede afirmar que el tamaño del efecto es grande pues su magnitud es mayor que el promedio de lluvia que se acumula en cualquiera de los meses más lluviosos del año (Figura 2a).

3.3 Evaluación del incremento de la resolución

La Figura 8a muestra la comparación de los valores observados en las 48 estaciones con los estimados obtenidos por Centella, et. al [4], tomando en cuenta el conjunto de estaciones utilizado en este trabajo. Se puede apreciar que los resultados alcanzados con la resolución de 3 km tienen menor habilidad que los obtenidos para 1 km con la simulación 4, con un valor de R^2 mucho menor. Así mismo, la distribución espacial de los estimados de lluvia en 3 km (Figura 8b) refleja estimaciones aún más elevadas que las obtenidas con la simulación 4. Esta comparación permite indicar que los resultados que se obtienen a mayor resolución resultan superiores, e invita a pensar que resultaría posible realizar estimaciones a resoluciones más altas que un kilómetro, mucho más si se incluyen las covariables del terreno empleadas en las simulaciones 1 ó 3.

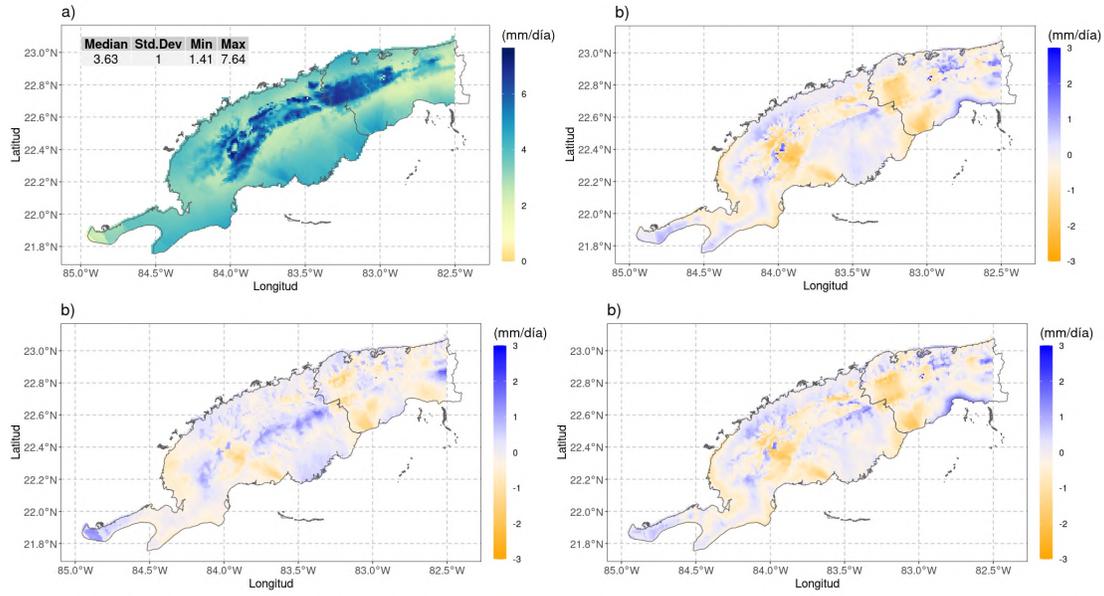


Figura 7: Distribución espacial del promedio de lluvia (mm/día) para la muestra utilizada en la simulación 4 (panel a) y las diferencias de las simulaciones 1, 2 y 3 con la simulación 4 (paneles b, c y d)

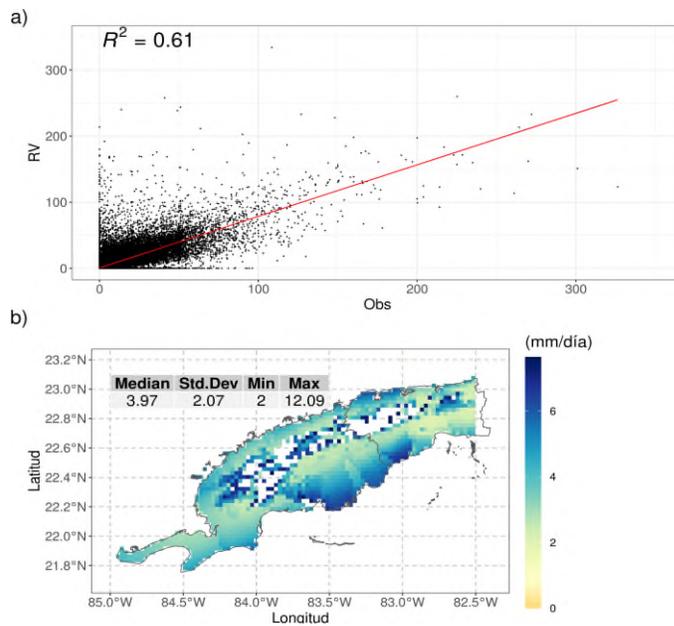


Figura 8: Resultados de las estimaciones según Centella et al. [4]. a) Diagrama de dispersión entre estimados y observados y, b) Distribución espacial del promedio de lluvia (mm/día). Las áreas en blanco en el mapa reflejan magnitudes de lluvia superiores a los valores máximos de la simulación 4.

5. CONCLUSIONES

En este documento se ha descrito el trabajo desarrollado para evaluar las posibilidades de mejorar la precisión de las estimaciones espaciales de la lluvia diaria a una resolución espacial de 1 km, incluyendo la información de distintas covariables relacionadas con la topografía del terreno. Luego de realizar diferentes etapas de trabajo que incluyeron

la preparación y análisis de las covariables, la realización y comparación de diferentes simulaciones con combinaciones de distintos predictores, se alcanzaron resultados que, en esencia, permiten concluir lo siguiente:

- 1 La inclusión de las covariables relacionadas con la topografía, incrementa la habilidad de la estimación de días secos y húmedos, así como la precisión de la magnitud de la lluvia.
- 2 Las simulaciones con mejores resultados fueron las que incluyeron las combinaciones de las covariables adicionales aspect y cost_dist (simulación 3) y TRI con cost_dist (simulación 1). Aunque la simulación 3 presentó resultados ligeramente mejores que la 1, esta última parece funcionar mejor en áreas donde el relieve resulta más complejo.
- 3 La decisión final sobre la inclusión de las covariables de las simulaciones 1 ó 3 puede requerir de investigaciones futuras sobre áreas de topografía más compleja y con mayor densidad de estaciones. Esto permitiría definir el uso de una covariable u otra o identificar estrategias de ponderar su empleo en función de la complejidad del terreno.
- 4 Se comprobó que al incrementar la resolución espacial de la rejilla hacia la que se estiman los valores, los resultados fueron más precisos. Este hecho invita a pensar que los resultados de este método pueden ser mejores con una mayor resolución espacial, aunque el efecto deberá ser evaluado con mayor robustez empleando una muestra de estaciones independientes a las que participan en el proceso de estimación.
- 5 Si bien la inclusión de covariables y el incremento de la resolución son elementos que adicionan complejidad e incrementan el costo computacional del método, no se considera que esto pueda ser un factor limitante a la hora de su implementación. Las razones para indicar esto son las siguientes:
 - a) Está claro que el método de estimación que se utiliza requiere de los recursos computacionales suficientes para resolver las estimaciones en rejillas de muy alta resolución. En estos casos se recomienda disponer de suficientes procesadores y memoria RAM para distribuir apropiadamente el trabajo y almacenar el volumen de datos. Sin embargo, se considera que el incremento de la precisión de los estimados día a día es un beneficio invaluable para las investigaciones y aplicaciones climáticas.
 - b) El mayor costo del método se produce una sola vez cuando se trata de desarrollar una base de datos para un período temporal largo (muchos años). Sin embargo, el costo de realizarlo para un día no es grande, aunque se reconoce nuevamente que el método demanda recursos de procesamiento y memoria.

RECEIVED: OCTOBER, 2023.
REVISED: FEBRUARY, 2024.

REFERENCIAS

- [1] CENTELLA-ARTOLA, A., BEZANILLA-MORLOT, A., SERRANO-NOTIVOLI, R., VÁZQUEZ-MONTENEGRO, R., SIERRA-LORENZO, M., and CHANG-DOMINGUEZ, D. (2023). A new long term gridded daily precipitation dataset at high-resolution for Cuba (CubaPrec1). **Data in brief**, 48, 109294. <https://doi.org/10.1016/j.dib.2023.109294>
- [2] HAYLOCK, M.R., HOFSTRA, N., KLEIN TANK, A.M.G., KLOK, E.J., JONES, P.D., NEW, M. (2008): A European daily high-resolution gridded data set of surface temperature and precipitation for 1950-2006. **J. Geophys. Res. Atmos.** 113 (20) <http://dx.doi.org/10.1029/2008JD010201> art. no. D20119.
- [3] KLING, H., FUCHS, M. and Paulin, M. (2012): Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios, **J. Hydrol.**, 424–425, 264–277, doi:10.1016/j.jhydrol.2012.01.011, 2012.
- [4] LECHA, L., PAZ, L.R., y LAPINEL, B. (1994): **El clima de Cuba**. Editorial ACC, La Habana, Cuba.
- [5] SERRANO-NOTIVOLI, R., De LUIS, M., and BEGUERÍA, S.(2017a): An R package for daily precipitation climate series re- construction, **Environ. Model. Softw.**, 89, 190–195, <https://doi.org/10.1016/j.envsoft.2016.11.005>.
- [7] SERRANO-NOTIVOLI, R., De LUIS, M., SAZ, M. A. and BEGUERÍA, S.(2017b): Spatially-based reconstruction of daily precipitation instrumental data series, **Clim. Res.**, 73, 167–186, <https://doi.org/10.3354/cr01476>
- [8] SERRANO-NOTIVOLI, R., TEJEDOR, E. (2021): From rain to data: A review of the creation of monthly and daily station-based gridded precipitation datasets. **WIREs Water**, 8 (6), <https://doi.org/10.1002/wat2.1555>.
- [9] TRUSOV, I.I., IZQUIERDO, A., DÍAZ, L. R. (1983): **Características espaciales y temporales de las precipitaciones atmosféricas en Cuba**. Editorial ACC, La Habana, Cuba.

[10] YANG X, XIE X, LIU DL, JI F, WANG L (2015): Spatial interpolation of daily rainfall data for local climate impact assessment over greater Sydney region. **Advances in Meteorology** 2015:1–12.
<https://doi.org/10.1155/2015/563629>