

Corrección de sesgos en estimaciones de precipitación basadas en datos satelitales en infrarrojo

Sergio H. González¹, Luciano Vidal², Juan J. Ruiz³, Pablo A. Negri⁴ y Alejo A. Silvarrey Barruffa⁵

¹ Centro de Investigaciones del Mar y la Atmósfera (CIMA/CONICET-UBA), (DCAO,FCEyN/UBA), (IRL IFAECI/CNRS-IRD-CONICET-UBA), Buenos Aires, Argentina, sergio.gonzalez@cima.fcen.uba.ar

² Servicio Meteorológico Nacional de Argentina, Buenos Aires, Argentina, lvidal@smn.gob.ar

³ Centro de Investigaciones del Mar y la Atmósfera (CIMA/CONICET-UBA), (DCAO,FCEyN/UBA), (IRL IFAECI/CNRS-IRD-CONICET-UBA), Buenos Aires, Argentina, jruiz@cima.fcen.uba.ar

⁴ Instituto de Investigación en Ciencias de la Computación (ICC/CONICET-UBA), (DC,FCEyN/UBA), Buenos Aires, Argentina, pnegri@gmail.com

⁵ Universidad Católica del Uruguay, Punta del Este, Uruguay, alejo.silvarrey@ucu.edu.uy

Resumen. La precipitación es uno de los principales componentes del balance hidrológico y su cuantificación es fundamental para el modelado preciso de dicho balance. Debido a la escasez de datos observados in-situ, se recurre a sensores remotos para poder incrementar la cobertura espacial y temporal. Sin embargo, las estimaciones basadas en sensores remotos, presentan sesgos que afectan su calidad. En este trabajo, se propone el uso de un modelo de redes convolucionales profundas para corregir parcialmente los errores sesgos del algoritmo de estimación de precipitación del satélite GOES-16. El entrenamiento de la red se lleva a cabo utilizando datos del radar meteorológico a bordo del satélite GPM. El modelo basado en redes neuronales, se compara con una técnica ampliamente utilizada en corrección de sesgos basada en datos históricos y con estimaciones de precipitación proveniente del producto PDIR-Now. El desempeño de todos los productos es evaluado con observaciones de estaciones pluviométricas. Las estimaciones del modelo propuesto reducen la raíz cuadrada del error cuadrático medio en un 164% (9%) respecto de las estimaciones del GOES-16 (PDIR-Now) y en 5% de la corrección basadas en datos históricos.

Palabras clave: Estimación de precipitación, Satélite, Corrección de sesgo, U-Net.

1 Introducción

La precipitación (PP) es la principal componente del ciclo hidrológico que se encarga de la recarga de agua en el suelo desde la atmósfera. Su monitoreo es fundamental para determinar la humedad del suelo a través de modelos de balance hídrico y por ejemplo determinar posibles rendimientos de cultivos en base a dichas características, en conjunto con otras variables meteorológicas como la temperatura, la radiación solar y la humedad del aire. Sin embargo, la densidad de estaciones pluviométricas

disponibles no es adecuada para poder representar la variabilidad espacial y temporal de la PP.

Para complementar la información provista por datos in-situ, la PP puede ser estimada a partir de información provista por sensores remotos (estimación cuantitativa de precipitación o ECP). Esto permite monitorear la PP con mayor resolución espacial y temporal y con una mayor cobertura areal (e.g. zonas escasamente pobladas, océanos, etc).

Una problemática de las ECP de productos cuasiglobales como el IMERG[1], GsMAP [2], CMORPH [3], entre otros, es que las versiones finales de dichos productos que cuentan con tratamiento posterior para reducción del sesgo de las estimaciones, conllevan a periodos de latencia de los datos de hasta 3 meses, lo cual es un tiempo no apto para su uso operativo.

Los ajustes se realizan debido a que las ECP suelen presentar sesgos que afectan su calidad. Esto implica un deterioro en la iniciación de los modelos de balance hídrico. Por ello se han implementado diferentes metodologías con el objetivo de corregir parcialmente dichas estimaciones.

Una de las técnicas ampliamente utilizada es la conocida como PDF-Matching, la cual es una metodología sencilla basada en un ajuste respecto a datos históricos [4]. Otras metodologías de correcciones de sesgos son las basadas en técnicas de aprendizaje automático las cuales han demostrado resultados prometedores para su aplicación [5][6].

En este trabajo se propone utilizar un modelo de aprendizaje automático basado en una arquitectura U-Net [7] que corrija parcialmente los errores del algoritmo de ECP provenientes del satélite Geostationary Operational Environmental Satellites - 16 (GOES-16) [8]. La red es entrenada con ECP provenientes del sensor Dual-frequency Precipitation Radar (DPR) [9] a bordo del satélite Global Precipitation Measurement (GPM). Esta estimación tiene una calidad muy superior dado que está basada en datos de radar, aunque no es adecuado para su uso operativo debido a la baja resolución temporal y cobertura espacial.

Este modelo es comparado con la técnica PDF-Matching y con el PDIR-Now [10] que es una estimación global basada en redes neuronales profundas. El desempeño de las diferentes técnicas se evalúa con observaciones de estaciones pluviométricas (Est-Pluv).

2 Datos

Los datos de estimaciones de PP utilizados provienen de diferentes productos de diversos sensores remotos. Entre ellos el algoritmo de ECP del satélite GOES-16 (GOES-RRQPE), que es derivado a partir de datos en infrarrojo y posee una resolución temporal de 10 minutos y espacial de 2 km. El algoritmo de ECP del sensor Dual-frequency Precipitation Radar (DPR) del satélite Global Precipitation Measurement (GPM) que está basado en datos de radar, pero con un periodo de revisita de 1 o 2 días, y con un ancho de sensado de 245 km². El producto de ECP PDIR-Now está basado en mapas de características auto-organizados que es un tipo de red neuronal artificial de aprendizaje no supervisado, y luego una estimación por relación de ajuste entre la temperatura de brillo en rango del infrarrojo térmico y la PP.

También se cuenta con datos observados de PP de 167 estaciones pluviométricas cada 10 minutos sobre la región de Córdoba. Dichos datos son parte de la base de datos de estaciones automáticas integrada por el Servicio Meteorológico Nacional [11].

3 Metodología

3.A Construcción del Conjunto de Datos

El conjunto de datos está compuesto por los pares de datos de la estimación de PP en superficie del GOES-RRQPE y el RR-DPR. Los datos comprenden el periodo de Enero 2020 a Junio 2021 inclusive, principalmente sobre la región de Sudamérica ($90^{\circ}\text{O}-30^{\circ}\text{O}$ | $50^{\circ}\text{S}-10^{\circ}\text{N}$). El conjunto de datos para el entrenamiento del modelo consiste en sectores cuadrados de 240 km^2 tomados a lo largo de la pasada del RR-DPR y que se superponen en un 40% (esto último para incrementar el tamaño del conjunto de entrenamiento). Por otra parte se cuenta con estimaciones del GOES-RRQPE interpoladas a la retícula de los sectores del RR-DPR utilizando interpolación bi-lineal y el dato más cercano en tiempo.

Las imágenes utilizadas para el entrenamiento son seleccionadas mediante el siguiente criterio: que la imagen cuente con al menos un 20% de los píxeles con PP mayor a 0.1 mm^{-1} del RR-DPR. Mediante este criterio se obtuvieron un total 6105 muestras de entrenamiento.

3.B Métodos aplicados: U-Net y PDF-Matching

El modelo implementado es un modelo de redes neuronales convolucionales con una arquitectura del tipo U-Net (Fig. A.1 del apéndice). Para entrenar y evaluar el desempeño de la U-Net se utilizó un conjunto de datos de entrenamiento (90%) y validación (10%) de los datos de enero 2020 a junio 2021, que se excluye los datos de enero y junio de 2021, para el periodo de testeo.

La arquitectura de tipo U-Net tiene por propósito codificar la información de entrada a un espacio de mínima dimensión, la cual posteriormente es decodificada en conjunto con la información del mismo nivel de codificación conocidos como saltos de conexión. Para el entrenamiento de la red, se utilizó una función de costo de error cuadrático medio (MSE), y un optimizador ADAM con un batch size de 16 imágenes, un tasa de aprendizaje de $1e^{-3}$ y 150 épocas de entrenamiento

La técnica de PDF-Matching realiza una regresión por cuantiles. La base de datos del conjunto de entrenamiento se utiliza para determinar los cuantiles de las distribuciones del RR-DPR y GOES-RRQPE. La evaluación del desempeño de este método se realiza durante el periodo de testeo.

Para evaluar el desempeño de los métodos implementados, se utilizaron datos de PP observada en superficie de estaciones automáticas integradas por el SMN durante el periodo de testeo, y también con el producto PDIR-Now. Las métricas utilizadas son la raíz cuadrada del error cuadrático medio (RMSE), el sesgo (BIAS) y el coeficiente de correlación de Pearson (CORR).

4 Resultados y Conclusiones

En la tabla de la Fig. B.1 del apéndice se presentan el RMSE, BIAS y CORR, y las distribuciones de PP acumuladas diarias para los conjuntos evaluados. Se observa que el GOES-RRQPE es el que mayor RMSE y BIAS posee, pero así también la mayor correlación. La aplicación de la técnica de PDF-Matching conserva la correlación del GOES-RRQPE pero redujo notablemente el RMSE y el BIAS. La U-Net es el que menor RMSE posee. El valor de BIAS es mucho menor que para el GOES-RRQPE, pero a diferencia del PDF-Matching es negativo, indicando una subestimación de los acumulados de PP. Para el PDIR-Now, el desempeño es superior al del GOES-RRQPE en términos de RMSE y BIAS, pero no a la técnica de PDF-Matching y la U-Net.

A partir de las distribuciones de PP acumulada diaria, se observa que el GOES-RRQPE es el que genera mayores PP acumuladas diarias. Inversamente el modelo de U-Net, subestima la frecuencia de PP más altas registrados por los pluviómetros. La técnica PDF-Matching y el producto PDIR-Now poseen una distribución similar a la observada, con una subestimación de los valores más extremos observados por los pluviómetros. Esta subestimación de los valores más extremos puede obedecer en parte a la baja resolución espacial de los datos de RR-DPR utilizados en el entrenamiento del PDF-Matching y de la U-Net.

Por otra parte, la subestimación más marcada que se observa en la U-Net es probable que se deba a las características propias del MSE como función de costo y a la relativa poca información contenida en los datos de entrada de la red (basada en una combinación de temperaturas de brillo en el infrarrojo). Por otra parte, a medida que aumenta el valor de PP, disminuye su representación en la base de datos, con lo cual los valores más extremos suelen estar pobremente representados. Nuestra línea de trabajo actualmente busca formas alternativas de la función de costo que tengan en cuenta dicho desbalance en los datos, como funciones de costo pesadas, como así también la inclusión de variables adicionales como entrada a la red que contribuyan a la detección y determinación de las PP.

5 Apéndice

A Arquitectura U-Net

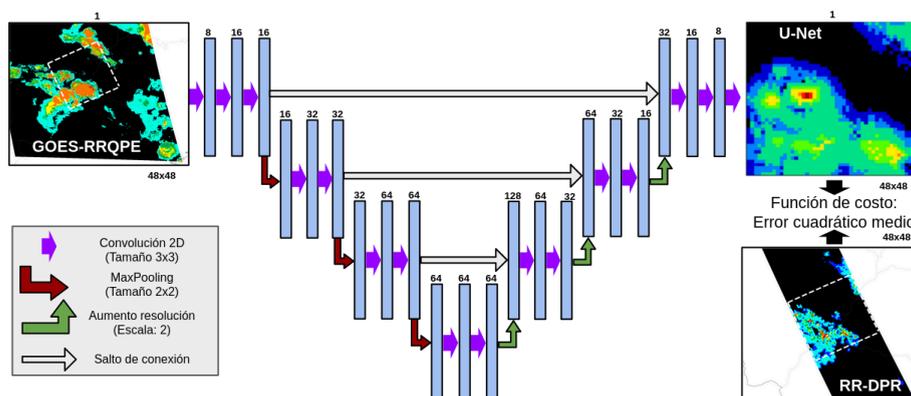


Fig A.1. Esquema de la arquitectura U-Net implementada. Las líneas blancas punteadas indican el sector correspondiente a la entrada en el GOES-RRQPE y al objetivo en el RR-DPR.

B Distribuciones y métricas de las ECP

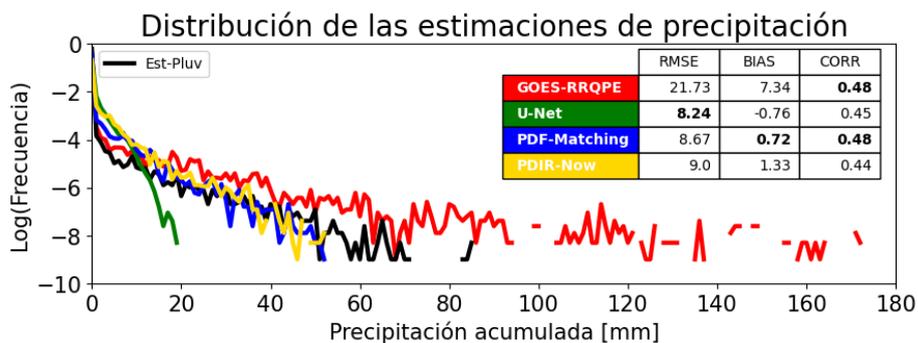


Fig. B.1. Distribuciones de PP acumulada diaria de GOES-RQPE (rojo), U-Net (verde), PDF-Matching (azul), PDIR-Now (amarillo) y Est-Pluv (negro). También se incluye las métricas evaluación obtenidas, raíz cuadrada del error cuadrático medio (RMSE), el sesgo (BIAS) y el coeficiente de correlación de Pearson (CORR) para los diferentes conjuntos respecto de las Est-Pluv en el periodo de testeo.

6 Bibliografía

- [1] G. Huffman et al., “Integrated Multi-satellite Retrievals for GPM (IMERG) Technical Documentation,” 2023. Available: https://gpm.nasa.gov/sites/default/files/2023-07/IMERG_TechnicalDocumentation_final_230713.pdf
- [2] T. Kubota et al., “Global Precipitation Map Using Satellite-Borne Microwave Radiometers by the GSMaP Project: Production and Validation,” in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 7, pp. 2259–2275, July 2007, doi: 10.1109/TGRS.2007.895337.
- [3] R. J. Joyce, J. E. Janowiak, P. A. Arkin, and P. Xie, “CMORPH: A Method that Produces Global Precipitation Estimates from Passive Microwave and Infrared Data at High Spatial and Temporal Resolution,” *Journal of Hydrometeorology*, vol. 5, no. 3, pp. 487–503, Jun. 2004, doi: [https://doi.org/10.1175/1525-7541\(2004\)005%3C0487:CAMTPG%3E2.0.CO;2](https://doi.org/10.1175/1525-7541(2004)005%3C0487:CAMTPG%3E2.0.CO;2)
- [4] P. Xie and A.-Y. Xiong, “A conceptual model for constructing high-resolution gauge-satellite merged precipitation analyses,” *Journal of Geophysical Research: Atmospheres*, vol. 116, no. D21, Nov. 2011, doi: <https://doi.org/10.1029/2011jd016118>.
- [5] H. Chen, L. Sun, R. Cifelli, and P. Xie, “Deep Learning for Bias Correction of Satellite Retrievals of Orographic Precipitation,” *IEEE transactions on geoscience and remote sensing*, vol. 60, pp. 1–11, Jan. 2022, doi: <https://doi.org/10.1109/tgrs.2021.3105438>.
- [6] S. Xiao, L. Zou, J. Xia, Z. Yang, and T. Yao, “Bias correction framework for satellite precipitation products using a rain/no rain discriminative model,” *Science of the total environment*, vol. 818, pp. 151679–151679, Apr. 2022, doi: <https://doi.org/10.1016/j.scitotenv.2021.151679>.
- [7] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation.” Available: <https://arxiv.org/pdf/1505.04597v1>
- [8] “GOES-R Advanced Baseline Imager (ABI) Algorithm Theoretical Basis Document For Rainfall Rate (QPE),” 2020. Accessed: Jun. 07, 2024. [Online]. Available: https://www.star.nesdis.noaa.gov/goestr/documents/ATBDs/Enterprise/ATBD_Enterprise_Rainfall_Rate_v3_2020-07-10.pdf
- [9] T. Iguchi et al., “GPM/DPR Level-2 Algorithm Theoretical Basis Document.” Accessed: Jun. 07, 2024. [Online]. Available: https://gpm.nasa.gov/sites/default/files/2022-06/ATBD_DPR_V07A.pdf
- [10] P. Nguyen et al., “PERSIANN Dynamic Infrared–Rain Rate Model (PDIR) for High-Resolution, Real-Time Satellite Precipitation Estimation,” *Bulletin of the American Meteorological Society*, vol. 101, no. 3, pp. E286–E302, Mar. 2020, doi: <https://doi.org/10.1175/bams-d-19-0118.1>.
- [11] G. M. Díaz, M. Vita, M. P. Hobouchian, L. J. Ferreira, y L. Giordano, “Expansión de la red del SMN empleando los datos de precipitación de las estaciones meteorológicas automáticas de terceros,” Apr. 2021. url: <https://repositorio.smn.gob.ar/handle/20.500.12160/1541>