

Estimación de Precipitación Utilizando Sensores Remotos y Simulaciones Numéricas Mediante Redes Neuronales

Ezequiel Geslin¹, Juan Ruiz^{1,2}, Sergio Gonzalez^{1,2,3}, Luciano Vidal³, Pablo Negri⁴

¹ Departamento de Ciencias de la Atmósfera y los Océanos (DCAO-FCEN-UBA), Buenos Aires, Argentina

² Centro de Investigaciones del Mar y la Atmósfera (CIMA-CONICET/UBA), Buenos Aires, Argentina

³ Servicio Meteorológico Nacional (SMN), Buenos Aires, Argentina

⁴ Instituto de Ciencias de la Computación (ICC - CONICET/UBA), Buenos Aires, Argentina
ezegeslin@gmail.com, jruiz@cima.fcen.uba.ar, sergio.gonzalez@cima.fcen.uba.ar, lvidal@smn.gob.ar, pnegri@gmail.com

Resumen. La precipitación es uno de los componentes más importantes del balance hidrológico, y su cuantificación es esencial para el modelado preciso de dicho balance. Debido a la escasez de datos observacionales in situ, se recurre a sensores remotos para ampliar la cobertura espacial y temporal de las condiciones atmosféricas, y a modelos estadísticos o basados en aprendizaje automático que permiten vincular las variables sensadas con la precipitación. En este trabajo, se ha desarrollado un modelo para la estimación cuantitativa de precipitación utilizando redes neuronales profundas. La red neuronal se alimenta con datos de temperatura de brillo en el rango del infrarrojo proporcionados por el GOES-16, y se propuso también incluir pronósticos numéricos del agua precipitable provistos por el GFS. El modelo propuesto se entrena contra las tasas de precipitación instantáneas estimadas a partir del radar a bordo del satélite GPM. Los resultados al utilizar únicamente la temperatura de brillo fueron satisfactorios; sin embargo, al incorporar las simulaciones, las métricas en el conjunto de prueba mejoraron y se logró una mejor representación de los máximos de precipitación.

Palabras clave: Estimación de Precipitación, Sensores Remotos, U-Net, Pronóstico numérico.

Precipitation Estimation Using Remote Sensing and Numerical Simulation Through Neural Networks

Abstract. Precipitation is one of the most critical components of the hydrological balance, and its accurate quantification is essential for reliable modeling. Due to the limited availability of in-situ observational data, remote sensing is used to enhance the spatial and temporal coverage of atmospheric conditions. Additionally, statistical and machine learning models are employed to link remotely sensed variables with precipitation. In this work, a deep neural network model was developed for quantitative precipitation estimation. The network is fed with infrared brightness temperature data from the GOES-16 satellite, and numerical forecasts of precipitable water from the GFS were also considered as input. The proposed model is trained against instantaneous precipitation rates estimated from the GPM satellite's onboard radar. Results using only brightness temperature as input were satisfactory; however, the inclusion of the numerical forecasts improved performance on the test dataset and allowed for a better representation of maximum precipitation..

Key Words: Precipitation Estimation, Remote Sensing, U-Net, Numerical Forecast.

Introducción

La precipitación (PP) es el principal componente del ciclo hidrológico y su monitoreo es fundamental para determinar la humedad del suelo a través de modelos de balance hídrico y, entre otras cosas, determinar posibles rendimientos de cultivos. Sin embargo, la densidad de estaciones pluviométricas disponibles es insuficiente para poder representar la variabilidad espacial y temporal de la PP, sumado a la falta de estaciones en las regiones rurales y oceánicas.

Para complementar la información provista por las mediciones en superficie, la PP puede estimarse a partir de información provista por los satélites meteorológicos. Uno de los satélites más relevantes en este contexto es el Global Precipitation Measurement (GPM), un satélite de órbita baja que cuenta con el Dual-frequency Polarization Radar (DPR), un radar de doble frecuencia que ofrece estimaciones de PP de alta calidad. No obstante, dicho satélite presenta limitaciones en su resolución temporal en comparación con los satélites geoestacionarios. El Geostationary Operational Environmental Satellite-16 (GOES-16) ofrece una cobertura temporal y espacial mucho más amplia. Sin embargo, la estimación de precipitación producida a partir de las radianzas en el rango del infrarrojo, medidas por dicho satélite, es de menor calidad en comparación con la proporcionada por el DPR.

Los modelos numéricos de predicción del tiempo (NWP) proporcionan simulaciones atmosféricas de alta calidad, basadas en las leyes de fluidos geofísicos. Estos modelos pueden estimar la distribución espacio-temporal de diferentes cantidades físicas que son relevantes para la estimación de precipitación (como la humedad disponible en la atmósfera, la circulación en diferentes niveles, entre otros) [1,2,3,4].

En este trabajo se propone entrenar una red neuronal de tipo U-Net para estimar la tasa de precipitación en superficie a partir de las mediciones de temperatura de brillo en el infrarrojo obtenidas por el satélite GOES-16 que puedan ser implementadas en tiempo real para la estimación de la precipitación con una frecuencia temporal de 10 minutos. Por otro lado, este trabajo se propone medir el valor agregado de incorporar como variables de entrada del modelo datos de simulaciones numéricas globales de agua precipitable que brinden información de aspectos que no puedan ser cuantificados a partir de la información satelital.

Datos

Para el desarrollo de las redes se utilizaron las tasas instantáneas de precipitación estimadas por el radar de doble frecuencia (Dual-frequency Polarization Radar, DPR)[5] a bordo del satélite de órbita baja Global Precipitation Measurement (GPM). Los datos del GPM fueron segmentados en imágenes de 48 x 48 píxeles (aproximadamente 240 x 240 km).

Como variable de entrada se emplearon las temperaturas de brillo (TB) en el rango del infrarrojo, obtenidas a partir del satélite GOES-16 [6] con una resolución espacial de 2 kilómetros y una frecuencia temporal de 10 minutos. En

todos los casos se seleccionó la imagen del GOES más cercana a cada una de las imágenes obtenidas por el satélite GPM.

Por otro lado, se utilizaron como entrada del modelo datos de simulaciones numéricas provistas por el Global Forecasting System (GFS). En todos los casos se utilizó el campo simulado cuyo horario fuera el más cercano a la imagen satelital del GOES para garantizar la mayor consistencia posible entre los datos de entrada al modelo. La variable utilizada fue el agua precipitable, que consiste en la integral vertical del vapor de agua en cada punto de la malla del modelo.

Tanto los datos del GOES, como del modelo, fueron interpolados a la misma retícula de los datos del GPM para facilitar el proceso de entrenamiento e inferencia.

El dominio espacial se centró en Sudamérica y sus océanos adyacentes. Se seleccionaron para el entrenamiento y validación aquellas imágenes del satélite GPM en las que al menos el 15% de los valores de precipitación superaran los 0.1 mm/h, y además presentarían una correlación inferior a -0.1 entre la PP estimada por el GPM y la temperatura de brillo del GOES. La muestra total cuenta con cerca de 10.000 imágenes entre enero de 2020 y junio de 2021, de las cuales el 80% se destinó al entrenamiento y el 20% a la validación. Para la etapa de testeo, se utilizaron los datos de los meses de enero y julio de 2021, con el fin de analizar el desempeño del modelo tanto en condiciones estivales como invernales.

Metodología

Dado el carácter espacialmente estructurado de los datos de entrada, se optó por el uso de una red neuronal convolucional de tipo U-Net [7], ampliamente utilizada en el estado del arte de las estimaciones de precipitación [8,9]. Esta arquitectura permite combinar de manera eficiente información multiescala gracias al uso de conexiones de salto que transfieren directamente detalles espaciales desde el codificador hacia el decodificador. Para incluir los datos de las simulaciones numéricas, los mismos son primero procesados por una red de tipo encoder y posteriormente concatenados a los datos satelitales en el bottleneck de la red U-Net a partir del cual ambos conjuntos de datos son procesados en conjunto a través del decodificador.

El desarrollo del modelo se realizó empleando la librería PyTorch 2.3.1 ejecutando el entrenamiento en unidades de procesamiento gráfico (GPU, NVIDIA A40) lo cual permitió reducir significativamente los tiempos de cómputo. Para el entrenamiento, se utilizó como función de pérdida el error cuadrático medio (MSE), y como optimizador se empleó ADAM, con un batch size de 64 imágenes y una tasa de aprendizaje inicial de 0.001. Se definió un máximo de 1000 épocas de entrenamiento, y se utilizó una técnica de reducción de la tasa de aprendizaje, que la redujo en un orden de magnitud cuando no se observaron mejoras en la función de pérdida de validación durante 20 épocas consecutivas. En caso de no registrarse mejoras tras dos reducciones sucesivas de la tasa, se aplicó un criterio de detención anticipada (early stopping) para reducir el riesgo de overfitting. Como función de activación se utilizó Leaky ReLU, con

una pendiente negativa de 0.01.

Resultados y conclusiones

Los modelos desarrollados fueron evaluados utilizando el conjunto de testing y calculando el RMSE, el BIAS y el CORR entre las predicciones generadas por las redes y las tasas de precipitación observadas por el DPR. Cada configuración fue entrenada con 5 inicializaciones independientes con el objetivo de obtener una mayor robustez en las métricas reportadas. Los resultados obtenidos, habiendo promediado las 5 inicializaciones, se resumen en la Tabla 1.

Tabla 1. Métricas correspondientes al conjunto de testeo

INPUTS	RMSE [mm.h ⁻¹]	BIAS [mm.h ⁻¹]	CORR
Temperatura de brillo [K]	1.53	0.11	0.21
Temperatura de brillo[K] y agua precipitable [mm]	1.51	0.07	0.23

Si bien el modelo entrenado únicamente con TB tiene resultados satisfactorios, que indican una capacidad adecuada para estimar la tasa de precipitación en superficie, al compararlo con la red entrenada con la inclusión del agua precipitable, esta última produce un error cuadrático medio y un bias menor al mismo tiempo que reporta una mayor correlación con las imágenes del DPR, sugiriendo que la combinación con simulaciones numéricas mejora el rendimiento del modelo.

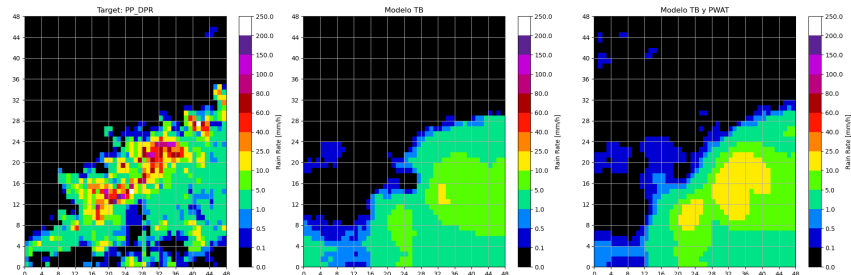


Fig. 1. Ejemplo de un caso correspondiente al conjunto de testeo imagen DPR (izquierda), predicción del modelo entrenado con TB (centro) y predicción del modelo entrenado con TB y agua precipitable (PWAT) (derecha).

En la Figura 1, se presenta una imagen del sensor DPR correspondiente al conjunto de testeo junto con las predicciones generadas por los modelos. En ella, se puede observar cómo la inclusión del agua precipitable le permitió al modelo representar valores más extremos de precipitación con una mayor precisión espacial.

Como líneas futuras de investigación, se propone seguir explorando distintas formas de integrar la variable simulada en la arquitectura del modelo. Además, se evaluará el impacto de redes más profundas en el desempeño del modelo. También se contempla incorporar nuevas variables provenientes del modelo numérico, como la tasa de precipitación, energía potencial convectiva disponible (CAPE), energía inhibitoria de la convección (CINE), viento y altura geopotencial en distintos niveles, así como variables de localización espacio-temporal.

Bibliografía

- [1] Li, X., Yang, Y., Mi, J., Bi, X., Zhao, Y., Huang, Z., Liu, C., Zong, L., & Li, W. (2021). Leveraging machine learning for quantitative precipitation estimation from Fengyun-4 geostationary observations and ground meteorological measurements. *Atmospheric Measurement Techniques*, 14, 7007–7023. <https://doi.org/10.5194/amt-14-7007-2021>
- [2] Upadhyaya, S. A., Kirstetter, P.-E., Kuligowski, R. J., & Searls, M. (2022). Towards improved precipitation estimation with the GOES-16 advanced baseline imager: Algorithm and evaluation. *Quarterly Journal of the Royal Meteorological Society*, 148(739), 2484–2500. <https://doi.org/10.1002/qj.4368>
- [3] Lyu, Y., & Yong, B. (2024). A novel double machine learning strategy for producing high-precision multi-source merging precipitation estimates over the Tibetan Plateau. *Water Resources Research*. <https://doi.org/10.1029/2023WR035643>
- [4] Wang, F., Tian, D., & Carroll, M. (2023). Customized deep learning for precipitation bias correction and downscaling. *Geoscientific Model Development*, 16, 535–556. <https://doi.org/10.5194/gmd-16-535-2023>
- [5] Iguchi, T., Seto, S., Meneghini, R., Yoshida, N., Awaka, J., Le, M., Chandrasekar, V., Brodzik, S., Kubota, T., & Takahashi, N. (2021). *GPM/DPR Level-2 Algorithm Theoretical Basis Document* (Revised Dec 2021). NASA.
- [6] NOAA/NASA GOES-R Program. (2019). *GOES-R Series Data Book* (Revision A). <https://www.goes-r.gov/downloads/resources/documents/GOES-RSeriesDataBook.pdf>
- [7] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 234–241). Springer. https://doi.org/10.1007/978-3-319-24574-4_28
- [8] Mosaffa, H., Ciabatta, L., Filippucci, P., Sadeghi, M., & Brocca, L. (2024). Hr-Precipnet: A machine learning framework for 1-km high-resolution satellite precipitation estimation. *SSRN*. <https://doi.org/10.2139/ssrn.4795786>
- [9] Sadeghi, M., Nguyen, P., Hsu, K., & Sorooshian, S. (2020). Improving near real-time precipitation estimation using a U-Net convolutional neural network and geographical information. *Environmental Modelling & Software*, 134, 104856. <https://doi.org/10.1016/j.envsoft.2020.104856>