

HERRAMIENTA PARA EL POSPROCESAMIENTO DE LOS DATOS NUMÉRICOS DE MODELOS REGIONALES: SCAHpy

Fiorela Castellón¹ e Ivonne Montes¹

¹ Instituto Geofísico del Perú, Lima, Perú

Palabras clave: Software de posprocesamiento, WRF, Python, análisis, visualización

Citar como Castellón, F., & Montes, I. (2024). Herramienta para el posprocesamiento de los datos numéricos de modelos regionales: SCAHpy. *Boletín científico El Niño*, Instituto Geofísico del Perú, Vol. 11 n.º 03, pág 13 - 17.

¿Cuál es la **contribución** de esta investigación para la comunidad científica?

Los resultados de esta investigación demuestran que SCAHpy facilita el análisis de datos numéricos de modelos atmosféricos, oceánicos e hidrológicos, por lo que la comunidad científica y académica puede beneficiarse enormemente con su uso puesto que simplifica el tiempo que se invierte en construir los algoritmos para el análisis de datos y, en consecuencia, permite dedicarse con exclusividad a la investigación en sí misma. Además, SCAHpy es de código abierto, por lo que los investigadores peruanos y extranjeros pueden contribuir con el desarrollo continuo del *software* para ampliar sus capacidades y diagnósticos.

Resumen

El presente avance de investigación muestra, en su versión inicial, el conjunto de programas de cómputo denominado SCAHpy, paquete de código abierto desarrollado en Python que permite el análisis y visualización de datos provenientes de modelos numéricos. Específicamente, SCAHpy ha sido desarrollado como herramienta de posprocesamiento para la componente atmosférica, oceánica e hidrológica del modelo Sistema Tierra regional de pronóstico para el territorio peruano y el océano Pacífico implementado por el Instituto Geofísico del Perú (IGP RESM-COW), y contiene una serie de programas para la extracción e interpolación de datos de coordenadas sigma a cartesianas, visualización de series de tiempo, mapas

horizontales, secciones verticales, entre otros. Cabe señalar que, en su fase inicial, el análisis y visualización de datos numéricos solo se enfoca en las salidas de la componente atmosférica del IGP RESM-COW, la cual se basa en el modelo atmosférico WRF.

1. Introducción

El modelado numérico es una herramienta fundamental para compensar la falta de datos observacionales (p. ej., Montes et al., 2010), así como para llevar a cabo una serie de experimentos simulados sobre situaciones observadas (p. ej., Moya et al., 2020) e idealizadas que han ocurrido, podrían haber ocurrido o pueden ocurrir (p. ej., Ticse, 2018) en la naturaleza.

Esto debido a que, en el campo de la geofísica, los procesos en la naturaleza pueden ser expresados a través de ecuaciones diferenciales parciales, las cuales describen la evolución espacial y temporal del proceso, es decir, su dinámica. Por tanto, el modelado numérico es una representación matemática de las propiedades físicas de la Tierra y sus interacciones, las cuales se basan en las leyes fundamentales de la naturaleza (p. ej., conservación de energía, masa, *momentum*).

Si nos referimos a los modelos numéricos del clima, para resolver las ecuaciones se debe definir un espacio físico, que puede ser todo el planeta o una región específica, tal que el espacio para resolver las ecuaciones en una malla dividida en pequeñas celdas que pueden ser regulares o irregulares. Por tanto, los resultados de los modelos que representan la dinámica de uno de los compartimentos del Sistema Tierra (atmósfera, hidrósfera, criósfera, litósfera y biósfera) o las interacciones entre ellos, se obtienen como un volumen de datos definidos en el espacio y tiempo, es decir, las variables que simulan tienen entre 3 y 4 dimensiones. Como consecuencia, se generan cantidades masivas de datos que requieren un tiempo considerable para su análisis posterior (es decir, para el posprocesamiento). Cabe señalar que, para la coordenada vertical, no se usa generalmente la coordenada cartesiana, sino que se construyen nuevas coordenadas para ahorrar el costo de la experimentación real. Por ello, tener una nueva coordenada implica un nivel de complejidad mayor para el análisis de los datos numéricos.

En tal sentido, la comunidad científica, bajo la necesidad de simplificar el costo del posprocesamiento, ha desarrollado *software* o herramientas de análisis tanto de datos observacionales como numéricos, entre los cuales se encuentran OceanSpy (Almansi et al., 2019), Metpy (May et al., 2022), Argopy (Maze et al., 2020), Pyleoclim (Khider et al., 2022), entre otros. Inspirados en lo anterior, se ha desarrollado SCAHPy, un paquete de código abierto desarrollado en Python que permite el análisis y visualización de datos provenientes de modelos numéricos que presentan la estructura vertical en coordenada sigma como el modelo atmosférico WRF o el modelo oceánico CROCO. Específicamente, SCAHPy ha sido desarrollado como herramienta de posprocesamiento de los datos de salida de la componente atmosférica, oceánica e hidrológica del modelo Sistema Tierra regional de pronóstico para el

territorio peruano y el océano Pacífico implementado por el IGP (denominado IGP RESM-COW v1; Segura et al., 2023; Montes et al., 2023), y contiene una serie de programas para la extracción e interpolación de datos de coordenadas sigma a cartesianas, visualización de series de tiempo, mapas horizontales, secciones verticales, entre otros. En el presente artículo damos a conocer una descripción preliminar de SCAHPy (v0.1.0) y el avance que se tiene en la implementación en su versión inicial.

2. Metodología

El paquete SCAHPy, desarrollado en Python, aprovecha las capacidades de paralelización de datos espaciales para acelerar el procesamiento y la visualización de las salidas del modelo. Este *software* no solo permite un análisis más eficiente de los resultados del modelo, sino que también ofrece la flexibilidad necesaria para adaptarse a las salidas de otros modelos atmosféricos, como el WRF.

Esta herramienta de código abierto, distribuida bajo la licencia GPL v3.0, ha sido desarrollada incorporando principalmente las librerías *xarray*, *matplotlib* y *cartopy*. La documentación completa, que incluye los tutoriales de aplicación, se encuentra disponible en el siguiente enlace: <https://fiorelacl.github.io/SCAHPy/>. Además, el paquete está disponible en el repositorio de PyPI (Python Package Index), en el siguiente enlace: <https://pypi.org/project/scahpy>, desde donde puede ser descargado e instalado fácilmente.

La componente atmosférica del modelo acoplado IGP RESM-COW v1 está basada en el modelo atmosférico WRF. Este modelo tiene la característica de tener el sistema de grillas Arakawa tipo C y coordenadas verticales que siguen el terreno (sigma), por lo que los datos requieren un tratamiento previo a su visualización. Para trabajar con datos grillados en Python, se manejan tres conceptos: dimensión, coordenadas y variables. El primero es la etiqueta o nomenclatura, la segunda es una estructura de datos que almacena los valores que conforman cada etiqueta. Un archivo de salida del modelo WRF suele denominarse `wrfout_d0x_YYYY-MM-DD_HH:00:00`. Este archivo, al ser leído en Python para el caso del modelo acoplado, está conformado por 8 dimensiones, 7 coordenadas y 164 variables.

Archivos sin SCAHpy

Dimensión

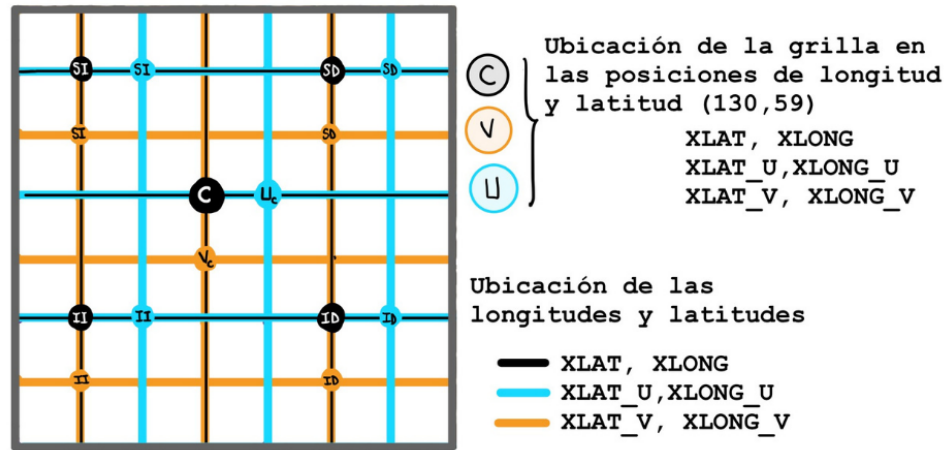
Time, south_north, west_east, bottom_top, bottom_top_stag, soil_layers_stag, west_east_stag, south_north_stag

Coordenadas

XLAT, XLONG, XTIME, XLAT_U, XLONG_U, XLAT_V, XLONG_V

Variables

164 Variables



Archivos con SCAHpy

Dimensión

time, lat, lon

Coordenadas

time, lat, lon

Variables

Variables seleccionadas

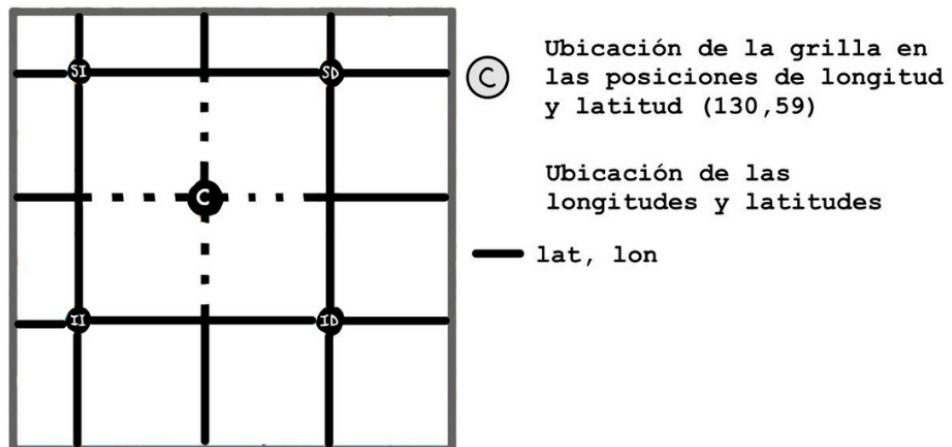


Figura 1. Lectura de archivos de la componente atmosférica del modelo IGP RESM-COW v1 sin el paquete SCAHpy (panel superior) y empleando el paquete SCAHpy (panel inferior).

En la Figura 1 se observa que, al leer el archivo sin SCAHpy, las dimensiones y las coordenadas no coinciden y, por ende, la indexación no funciona. Por ejemplo, la latitud está denominada como 'south_north'; no obstante, los valores de latitud están guardados en la coordenada XLAT para variables centradas y XLAT_U o XLAT_V para los campos de viento. El paquete SCAHpy, en su módulo de lectura de datos, realiza la homogeneización de coordenadas y dimensiones, tras lo cual se queda solo con los principales atributos, tales como latitud, longitud, tiempo y verticales de ser necesario.

Entre las funcionalidades más destacadas del paquete se encuentran:

1. Lectura de datos: El paquete ofrece dos funciones principales, ds_wrf_single y ds_wrf_multi, diseñadas para leer archivos individuales o múltiples, respectivamente. Ambas funciones leen únicamente las variables indicadas por el usuario; asimismo, homogenizan las coordenadas de las variables al pasar las coordenadas de los campos de velocidad de los bordes de la grilla hacia el centro (Figura 1).

- 2. Extracción de datos en puntos de estación:** La función extract_station_wrf permite extraer los valores de las variables de interés en ubicaciones de estaciones proporcionadas a través de archivos .csv o .shp.
- 3. Interpolación vertical:** La función vert_levs permite la interpolación de variables seleccionadas a los niveles de presión indicados por el usuario. En caso contrario, se interpolan a 13 valores de presión preestablecidos.
- 4. Obtención de climatologías:** La función dmy_var convierte los valores horarios en valores diarios, mensuales o anuales, permitiendo sintetizar variables mediante acumulación, promedio o mediana. Por otro lado, monthly_clim calcula la climatología mensual de las variables indicadas utilizando la media o la mediana.
- 5. Visualización:** La función map_pp_uv10_sst genera un gráfico 2D (Figura 2a) que representa las variables de precipitación, viento zonal y meridional a 10 m, así como la temperatura superficial del mar. Además, cross_section_yz produce un corte vertical que muestra la distribución de una variable con respecto a distintos niveles de presión (Figura 2b).

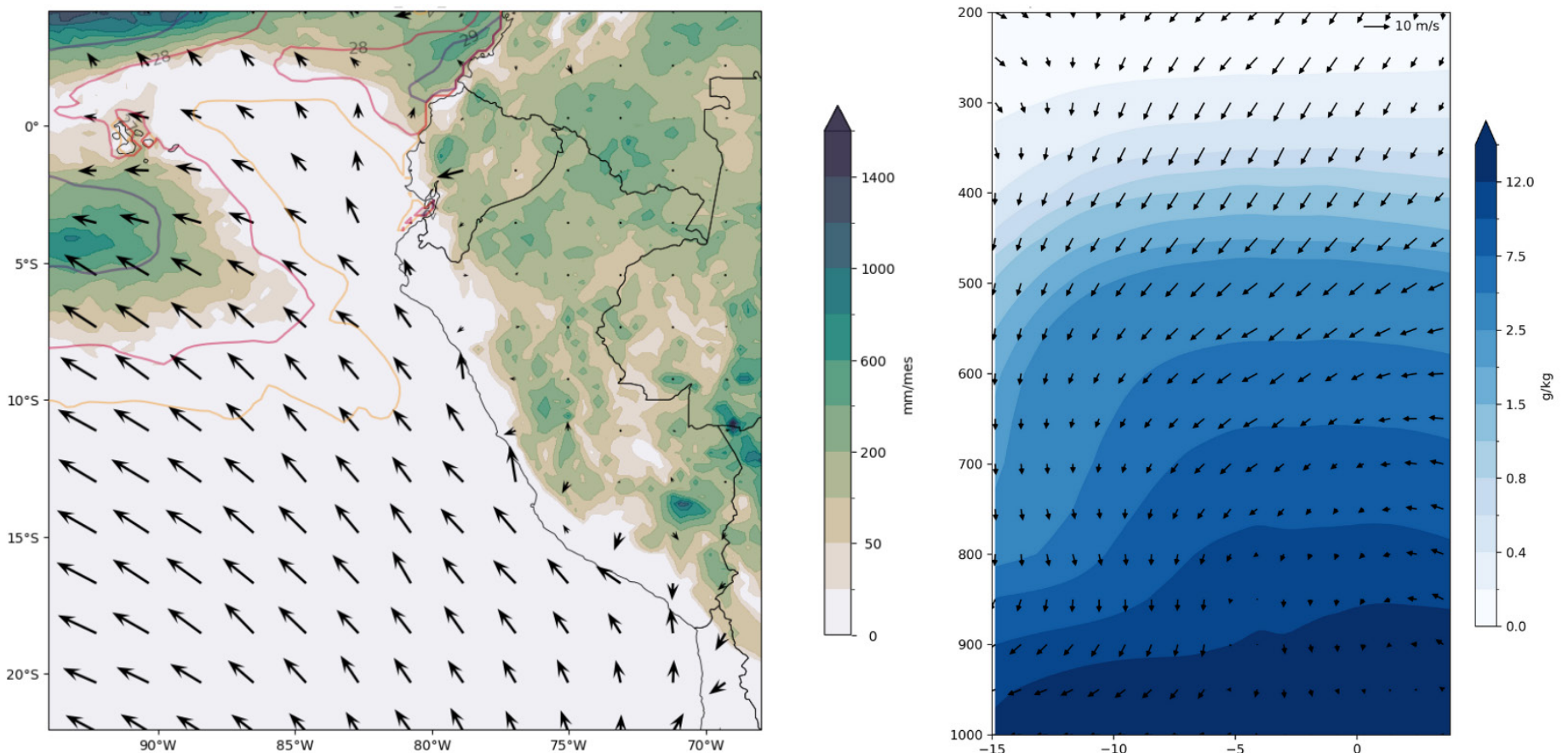


Figura 2. (Panel izquierdo) Mapa en superficie que muestra la precipitación acumulada mensual (izquierda), vientos (flechas), contornos de temperatura superficial del mar (línea amarilla, roja y morada corresponden a las isotermas de 26 °C, 27 °C y 28 °C, respectivamente) y (panel derecho) sección meridional de la atmósfera que muestra la distribución vertical de humedad específica y vientos de la climatología promedio de febrero-marzo-abril .

En cuanto a las visualizaciones, el paquete ofrece la capacidad de generar mapas con hasta tres variables en un mismo nivel, como es el caso de la precipitación, los vientos y los contornos de la temperatura superficial del mar (Figura 2a). Además, para variables distribuidas en distintos niveles verticales, SCAHpy posibilita la creación de gráficas tipo Hovmoller que ilustran el transporte vertical de la humedad específica, tal como se muestra en la Figura 2b.

3. Conclusiones preliminares

El paquete SCAHpy ofrece una solución integral para el procesamiento de archivos de salida del modelo WRF que permite proporcionar archivos netCDF con coordenadas limpias y formato de tiempo adecuado para una fácil lectura con programas como CDO. Además, facilita la interpolación vertical a los niveles especificados y genera gráficos de diagnósticos, incluyendo mapas y secciones verticales.

Para futuras versiones, se planifica la implementación de diagnósticos adicionales para la componente oceánica, así como la inclusión de funciones para la obtención de otros diagnósticos relevantes en

la componente atmosférica, como el transporte de humedad y el cálculo de la humedad relativa. También se espera que el paquete esté disponible en Conda así como en PyPI, lo que ampliará su accesibilidad y la facilidad de instalación para los usuarios.

Referencias

Almansi, M., Gelderloos, R., Haine, T. W. N., Saberi, A., & Siddiqui, A. H. (2019). OceanSpy: A Python package to facilitate ocean model data analysis and visualization. *Journal Of Open Source Software*, 4(39), 1506. <https://doi.org/10.21105/joss.01506>

Beuzen, T., & Timbers, T. (2022). *Python packages*. Chapman & Hall/CRC The Python Series.

Castillón, F., Berlin, S., y Montes, I. (2023). Validación de la componente atmosférica del sistema acoplado regional océano-atmósfera del Pacífico sudeste. *Boletín científico El Niño*, Instituto Geofísico del Perú, 10 (1), 9-12.

Geir Arne Hjelle, Real Python (2022, diciembre 1). *How to Publish an Open-Source Python Package to PyPI*. Recuperado el 05 Enero 2024 de <https://realpython.com/pypi-publish-python-package/> High-Performance computing market. (2023, marzo 30). Recuperado el 05 Enero 2024 de <https://www.futuremarketinsights.com/reports/high-performance-computing-market>

Khider, D., Emile-Geay, J., Zhu, F., James, A., Landers, J., Ratnakar, V., & Gil, Y. (2022). Pyleoclim: Paleoclimate Timeseries Analysis and Visualization With Python. *Paleoceanography And Paleoclimatology (Online)*, 37(10). <https://doi.org/10.1029/2022pa004509>

May, R. M., Goebbert, K. H., Thielen, J. E., Leeman, J. R., Camron, M. D., Bruick, Z. S., Bruning, E. C., Manser, R. P., Arms, S. C., & Marsh, P. T. (2022). MetPy: A Meteorological Python Library for Data Analysis and Visualization. *Bulletin Of The American Meteorological Society*, 103(10), E2273-E2284. <https://doi.org/10.1175/bams-d-21-0125.1>

Maze, G., & Balem, K. (2020). argopy: A Python library for Argo ocean data analysis. *Journal Of Open Source Software*, 5(53), 2425. <https://doi.org/10.21105/joss.02425>

Montes, I., Colas, F., Capet, X., & Schneider, W. (2010). On the pathways of the equatorial subsurface currents in the eastern equatorial Pacific and their contributions to the Peru-Chile Undercurrent. *Journal Of Geophysical Research*, 115(C9). <https://doi.org/10.1029/2009jc005710>

Montes, I., Mosquera, K., Luna, M., Gilt, H., Gazen, D., & Woodman, R. (2016). Sistema computacional de alto rendimiento para la simulación de fluidos geofísicos HPC-Linux-Clúster. *Boletín técnico: Generación de información y monitoreo del Fenómeno El Niño, Instituto Geofísico del Perú*, 3(3). <https://repositorio.igp.gob.pe/handle/20.500.12816/4663>

Moya-Álvarez, A. S., Martínez-Castro, D., Kumar, S., Rojas, J. L. F., Estevan, R., Saavedra, M., & Silva, Y. (2020). Statistical characterization of vertical meteorological profiles obtained with the WRF-ARW model on the central Andes of Peru and its relationship with the occurrence of precipitation on the region. *Atmospheric Research (Print)*, 239, 104915. <https://doi.org/10.1016/j.atmosres.2020.104915>

Segura, B. y Montes, I. (2021). Análisis de sensibilidad del sistema acoplado regional COW para el Pacífico sudeste. *Boletín científico El Niño, Instituto Geofísico del Perú*, 8 (2), 12-17.

Segura, B., Montes, I., Castellón, F., Manay R. & Takahashi, K. (2023). Implementación del componente acoplado océano-atmósfera del Modelo Regional del Sistema Tierra (RESM, por sus siglas en inglés) para el territorio peruano y el océano Pacífico oriental: periodo enero-julio 2023. *Boletín científico El Niño, Instituto Geofísico del Perú*, vol. 10 n.º 11, págs. 10-13.

Srinath, K.R. (2017) Python—The Fastest Growing Programming Language. *In-ternational Research Journal of Engineering and Technology (IRJET)*, 4, 354-357.

Ticse, K. E. (2018). Caracterización de la dinámica oceánica de la costa peruana bajo diferentes escenarios de esfuerzo de vientos (Tesis para obtener el título profesional de Licenciado en Física). Universidad Nacional Mayor de San Marcos, Lima, Perú.

Tucker, T., Giglio, D., Scanderbeg, M., & Shen, S. S. P. (2020). Argovis: A Web Application for Fast Delivery, Visualization, and Analysis of Argo Data. *Journal Of Atmospheric And Oceanic Technology*, 37(3), 401-416. <https://doi.org/10.1175/jtech-d-19-0041.1>