# Protocolo para automatizar la descarga de datos climáticos desde la nube y generar indicadores biometeorológicos para el monitoreo epidemiológico de cultivos

# Protocol for automating the download of climatic data from the cloud and generating biometeorological indicators for the epidemiological monitoring of crops

Paccioretti P., Giannini-Kurina, F., Suarez F., Scavuzzo M., Balzarini. M.

Paccioretti P., (ORCID: 0000-0002-1267-6361), Universidad Nacional de Córdoba, Facultad de Ciencias Agropecuarias e Instituto de Altos Estudios Espaciales “Mario Gulich”, Comisión Nacional de Actividades Espaciales, Universidad Nacional de Córdoba, Falda del Cañete, Argentina. Giannini-Kurina, F. (ORCID: 0000-0001-6763-7792), Jornær Sektioner, Institut for Agroøkologi, Aarhus Universitet, Dinamarca. Suarez F., (ORCID: 0000-0001-6763-7792), Universidad Nacional de Córdoba, Facultad de Ciencias Agropecuarias y UFYMA CONICET, Córdoba, Argentina. Scavuzzo M., (ORCID: 0000-0003-0905-6361) Instituto de Altos Estudios Espaciales “Mario Gulich”, Comisión Nacional de Actividades Espaciales, Universidad Nacional de Córdoba, Falda del Cañete, Argentina. Balzarini, M. (ORCID: 0000-0002-4858-4637), Universidad Nacional de Córdoba, Facultad de Ciencias Agropecuarias y UFYMA CONICET, Córdoba, Argentina.

*Correspondencia a:* pablopaccioretti@agro.unc.edu.ar

## Resumen

Datos climáticos derivados de imágenes o productos satelitales disponibles en la nube tienen gran cobertura en espacio y tiempo, buena precisión y en general son de libre acceso. Sin embargo, la obtención y descarga de variables climáticas a diferentes escalas se encuentra limitada por la falta de procedimientos computacionales estandarizados. El objetivo de este estudio fue desarrollar un código computacional que facilite el manejo de imágenes satelitales para obtención de variables climáticas en un dominio espaciotemporal. El producto ERA5 del servicio de Cambio Climático *Copernicus* fue usado como fuente de datos climático. El protocolo incluye la descarga desde la plataforma *Google Earth Engine* con un código desarrollado en lenguaje R. Adiciona el preprocesamiento estadístico de los datos climáticos a escala quincenal y/o mensual. Combinando productos derivados de satélites con conocimiento agronómico sobre un cultivo, los datos climáticos pueden ser convertidos en variables biometeorológicas y usados para el monitoreo espaciotemporal de cultivos. El proceso desarrollado se validó superponiendo datos de variables biometeorológicas en cada sitio de un estudio epidemiológico sobre dos virus monitoreados por 15 años. El procedimiento puede ser aplicado a otros productos satelitales que generan datos espaciales. El script de R y datos utilizados se encuentra en https://github.com/PPaccioretti/descarga\_datos\_climaticos.

## Palabras claves:

Google Earth Engine – Software R – ERA5 – Computación en la nube.

## Summary

Climate variables data derived from satellite imagery or products available in the cloud have wide coverage in space and time, good accuracy, and are generally freely accessible. However, limited advances have been made to obtain and download climatic variables automatically from the cloud at different spatial and temporal scales. The objective of this study was to develop a computational code to facilitate handling satellite images in order to derive climatic variables for a given spatiotemporal domain. The climate data was obtained from the ERA5 product of the Copernicus Climate Change Service. The protocol includes data download from the Google Earth Engine platform with a code developed in R language. The developed protocol includes statistical preprocessing of climatic data at fortnightly and/or monthly scale. By combining satellite-derived products with agronomic knowledge about a crop, climate data can be converted into biometeorological variables and used for spatiotemporal monitoring of crops. The developed process was validated by joint data from biometeorological variables at each site of an epidemiological study on two viruses monitored for 15 years. However, it may be applied to other satellite products using spatial data. The R-script and sample data to run the proposed protocol are provided at https://github.com/PPaccioretti/descarga\_datos\_climaticos.

## Keywords

Google Earth Engine – R software – ERA5 – Cloud Computing

## Introducción

El clima es un factor usado para explicar la distribución y desarrollo de organismos vivos y su relación con los recursos naturales que los rodean (Yim y Kira, 1975). Cuando las variables meteorológicas se relacionan con estados del cultivo se obtienen indicadores o variables biometeorológicas (Gosling et al., 2014; Noce et al., 2020), por ejemplo, precipitaciones o temperaturas previo a la siembra, durante el llenado de grano o a cosecha.

Enmarcado dentro del Servicio de Cambio Climático Copernicus (*Copernicus Climate Change Service* - C3S) de la Agencia Espacial Europea (ESA), el Centro Europeo para Previsiones Meteorológicas a Medio Plazo (*European Centre for Medium-Range Weather Forecasts* - ECMWF) está produciendo grandes conjuntos de datos climáticos a escala global. Se combinan modelos climáticos y datos observacionales de todo el mundo. Este producto es la quinta generación de Reanálisis Europeo (*European ReAnalysis* - ERA5), sucesor de ERA-Interim (Muñoz-Sabater et al., 2021)(Muñoz-Sabater et al., 2021). Estos productos, ponen a disposición información para todo el mundo a escala temporal diaria y mensual desde el año 1979. Así, cada vez es más factible mapear y modelar las variaciones o cambios espaciotemporales de las condiciones meteorológicas imperantes antes, durante y después de los cultivos anuales.

Gracias a la disponibilidad de datos climáticos masivos en diversas plataformas de Internet, es posible obtener información climática variada para un sitio georreferenciado sobre un territorio de interés. Actualmente, se cuenta con diferentes herramientas informáticas para la obtención de productos satelitales almacenados en la nube. Entre ellas, Google Earth Engine (GEE) (Gorelick et al., 2017) es una plataforma ampliamente utilizada, la cual permite, a través de código de programación, la descarga y procesamiento de información proveniente de numerosas imágenes satelitales y datos geoespaciales desde la nube. Una ventaja de GEE es que permite el procesamiento en la nube *i.e.* realizados por medio de Internet fuera del ordenador del usuario. A través de GEE es posible preprocesar diferentes variables meteorológicas georreferenciadas a diferentes escalas, tanto espacial como temporal, para su posterior descarga. Los datos pueden obtenerse para un sitio o lote agrícola, un país o incluso a escala planetaria en diferentes períodos de tiempo.

La plataforma GEE permite la descarga de datos meteorológicos del producto ERA5 para cualquier sitio del planeta desde el año 1979. Los datos pueden descargarse por hora o agregados, tanto mensualmente o promedios mensuales por hora del día. Teniendo en cuenta el creciente interés por análisis espacialmente explícitos, la obtención de variables biometeorológicas puede beneficiarse del cálculo de los productos climáticos, datos de estaciones meteorológicas georreferenciadas y a través de la interpolación basada en regresión o el reanálisis climático (Hu et al., 2016). Sin embargo, se han realizado limitados avances para automatizar la obtención y descarga de variables meteorológicas a diferentes escalas espaciales y temporales de la nube para su posterior correlación con el estado de los cultivos, sometidos a diversas condiciones ambientales y la influencia de distintas enfermedades. El objetivo de este estudio es desarrollar un protocolo computacional que facilite la obtención, procesamiento, descarga y visualización de datos climáticos desde la nube para obtener variables biometeorológicas que permitan el monitoreo de un evento particular asociado al desarrollo de un cultivo y la presencia/ausencia de virus. El protocolo propuesto integra, de manera novedosa, nuevas herramientas utilizadas para el procesamiento de datos espaciales. El uso de las herramientas desarrolladas en el software R y GEE, facilita el accionar de futuros usuarios y sirve de base para otras implementaciones en el monitoreo de cultivos.

## Material y Métodos

### Datos

Utilizando la plataforma GEE a través del paquete *rgee* (Aybar et al., 2020) del software R (R Core Team, 2021) se descargaron datos cada 14 días de temperatura, precipitaciones y temperatura de punto de rocío para sitios con observaciones sobre la presencia de virus W*heat streak mosaic virus* (WSMV) y *High plains wheat mosaic virus* (HPWMoV) en trigo correspondiente a quince años. De esta manera la unidad observacional de la base de datos resultante corresponde a la combinación de sitio y año para los cuales se cuenta con información cada 14 días de cada variable biometeorológica en el periodo de interés. Las descargas se hicieron para 15 años con datos de campo comprendidos en el período 2002-2021 (excepto los años 2008, 2016, 2017, 2019 y 2020). En estos años se monitoreó la incidencia viral en 204 sitios de Argentina, acumulando 898 observaciones en el total de años muestreados, de cada muestra se determinó la presencia/ausencia del virus por técnica serológica DAS-ELISA con sueros comerciales específicos para cada virus (datos provistos por la Dra. Vanina Alemandri, UFYMA, INTA-CONICET). Ambos virus son transmitidos por un ácaro eriófido, *Aceria tosichella*, por lo que la presencia de la enfermedad puede estar asociada a las condiciones meteorológicas favorables para la diseminación de los ácaros vectores (Wosula et al., 2017). La región utilizada para visualizar la variabilidad espacio temporal de las variables climáticas está conformada por 11 provincias de Argentina (Figura 1): La Pampa, Buenos Aires, Córdoba, Santa Fe, Entre Ríos, Santiago del Estero, Corrientes, Salta, Chaco, Tucumán y Jujuy, donde se ha monitoreado la enfermedad. Los datos climáticos se obtuvieron del producto ERA5 Land (Muñoz-Sabater et al., 2021). Este producto contiene información climática de 50 variables (bandas) en una resolución espacial de 11132 × 11132 metros. En este trabajo se utilizaron datos de precipitaciones en metros, temperatura del aire y temperatura de rocío, ambos a dos metros de la superficie en grados Kelvin. Para una mejor interpretación de los valores, las precipitaciones se multiplicaron por 1000 para transformarlas a milímetros (mm) y las temperaturas se transformaron a grados centígrados (°C) restando 273,15 a cada valor. Para los muestreos realizados con relación al cultivo de trigo, la información biometeorológica descargada se obtuvo como el promedio o acumulación en caso de precipitaciones de periodos de 14 días durante los meses de abril, mayo y junio. La información espacial de los límites de las provincias muestreadas se obtuvo desde el *Global Administrative Unit Layers* (GAUL) distribuido por FAO (*Food and Agriculture Organization)* de las Naciones Unidas. En la Figura 1 se muestran los sitios de muestreo para los 15 años de estudio.

### Protocolo de descarga y análisis

Mediante la interfaz de programación de aplicaciones (*Application Programming Interface* - API), es posible utilizar y ejecutar código de GEE desde clientes que utilizan JavaScript o Python. En el software de programación R (R Core Team, 2021), se desarrolló el paquete *rgee* (Aybar et al., 2020) que permite el uso de la API de GEE de Python y la comunicación entre R y Python que es posible usando el paquete *reticulate* (Ushey et al., 2022). La secuencia de código, resultado de este trabajo, permite la obtención de límites provinciales desde la nube, unión de archivos espaciales entre ocurrencia de virus y provincia de muestreo, mapeo de datos espaciales, descarga, obtención de medidas descriptivas de las imágenes descargadas y visualización de datos climáticos temporales para cada sitio de muestreo y para la región en estudio. El script se encuentra disponible en https://github.com/PPaccioretti/descarga\_datos\_climaticos.

Además de permitir descargar una de las capas de las variables meteorológicas disponibles en ERA5 para un determinado dominio espacial y periodo de tiempo, el protocolo incluye herramientas de preprocesamiento para facilitar el manejo de datos tipo ráster utilizando el paquete *stars* (Pebesma, 2022) y la visualización de los datos. Es posible generar mapas de variabilidad espacial de las variables meteorológicas mediante el uso del paquete *tmap* (Tennekes, 2018) y *ggplot2* (Wickham, 2016). Concatenando la variabilidad espacial de una serie de imágenes temporales se describe la variabilidad espaciotemporal. Se calculan promedios o acumulaciones de valores de estas variables para distintos periodos o semanas relacionadas al ciclo del cultivo. Finalmente, para visualizar el impacto de las variables biometeorológicas seleccionadas se realizan gráficas usando el paquete *ggplot2* (Wickham, 2016) y de R para contrastar las medias de las variables climáticas entre sitios con presencia vs. sitios con ausencia del evento de monitoreo, en el caso de ilustración el evento monitoreado es la presencia del virus bajo estudio.

Finalmente, se analizó la presencia y ausencia de cada virus en relación con valores de las variables biometerológicas resumidas en periodos de 14 días para todo otoño (siete periodos). La relación fue estudiada estadísticamente mediante un modelo lineal generalizado para una variable binaria (presencia/ausencia del virus en el cultivo) y 21 variables regresoras, *i.e.* temperatura promedio, precipitación acumulada y punto de rocío promedio, para cada uno de los siete períodos en estudio. El modelo fue ajustado con enlace canónico utilizando la función *glm* del software R (R Core Team, 2021).

## Resultados y Discusión

### Visualización de variabilidad espacial de datos climáticos

En la Figura 2a se muestra la temperatura promedio para los meses de abril, mayo y junio en los años donde se realizaron muestreos. El promedio de temperatura para toda la región fue de 15.7°C, el valor máximo promedio observado en los años en estudio fue de 23.8°C, el valor mínimo promedio fue de -10°C observándose en la zona noroeste de la región. Las precipitaciones acumuladas promedio (Figura 3b) para la región en estudio, fueron de 190 mm para otoño, variando entre 5 mm y 954 mm. La temperatura de punto de rocío promedio fue de 8.9°C (Figura 3c). En el noroeste de la región se observaron las menores temperaturas de punto de rocío (-26.4°C), y la mayor altitud (Figura 3d).

### Visualización de variabilidad espacio temporal en datos climáticos

En la Figura 3 se visualizan las precipitaciones acumuladas de la región para los años en estudio. En el año 2009 se observaron las menores precipitaciones promedio (98 mm), mientras que en el año 2002 se observó la mayor precipitación promedio 278 mm). Las temperaturas promedio se encontraron entre 15°C y 17,1°C. El año con promedio de temperatura más bajo fue el 2007 y el año con mayor temperatura promedio fue el 2009.

### Relación datos bioclimáticos con presencia de enfermedad

La presencia/ausencia del virus HPWMoV en relación con valores de las variables biometeorológicas resumidas para períodos de 14 días, desde el comienzo del otoño (semana 0), se sustentó estadísticamente con el ajuste de un modelo lineal generalizado. Los resultados indicaron que una variable de mayor impacto es la precipitacion acumulada de las semanas 7 y 8 (relación negativa con p-valor <0,0001) y de las semanas 11-12 y 13-14 (relación positiva, p-valor = 0,0001 y p-valor <0,0001 respectivamente) (Figura 4). El impacto de las precipitaciones también se observó sobre la presencia del virus WSMV, que fue mayor cuando ocurrieron mayores precipitaciones durante la semana 5 y 6 (p-valor <0,0001) y menor con precipitaciones altas a finales del otoño (p-valor <0,0001) (Figura 5).

## Conclusión

El protocolo estadístico-computacional desarrollado, viabiliza la descarga de variables climáticas desde la nube en diferentes escalas espaciales y temporales, además, facilita el análisis exploratorio del impacto del clima sobre el evento de interés. Se implementó exitosamente para descargar datos de precipitación, temperatura y punto de rocío que fueron relacionados a la presencia/ausencia de virosis en trigo. El script desarrollado puede ser utilizado para el estudio de otros eventos de cultivo con distribución binaria (presencia/ausencia) que hayan sido georreferenciados.

## Agradecimientos

Las investigaciones que sustentan este estudio fueron posibles gracias al proyecto PUE UFYMA-CONICET. Además, se agradece a la Ing. Agr (Dra.) Vanina Alemandri por la base de datos utilizada para la ilustración.

## Bibliografía

Aybar, C., Wu, Q., Bautista, L., Yali, R., y Barja, A. (2020). rgee: An R package for interacting with Google Earth Engine. *Journal of Open Source Software*, *5*(51), 2272. https://doi.org/10.21105/joss.02272

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., y Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, *202*, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Gosling, S. N., Bryce, E. K., Dixon, P. G., Gabriel, K. M. A., Gosling, E. Y., Hanes, J. M., Hondula, D. M., Liang, L., Bustos Mac Lean, P. A., Muthers, S., Nascimento, S. T., Petralli, M., Vanos, J. K., y Wanka, E. R. (2014). A glossary for biometeorology. *International Journal of Biometeorology*, *58*(2), 277–308. https://doi.org/10.1007/s00484-013-0729-9

Hu, Z., Hu, Q., Zhang, C., Chen, X., y Li, Q. (2016). Evaluation of reanalysis, spatially interpolated and satellite remotely sensed precipitation data sets in central Asia. *Journal of Geophysical Research: Atmospheres*, *121*(10), 5648–5663. https://doi.org/10.1002/2016JD024781

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D. G., Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo, C., y Thépaut, J.-N. (2021). ERA5-Land: a state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data*, *13*(9), 4349–4383. https://doi.org/10.5194/essd-13-4349-2021

Noce, S., Caporaso, L., y Santini, M. (2020). A new global dataset of bioclimatic indicators. *Scientific Data*, *7*(1), 398. https://doi.org/10.1038/s41597-020-00726-5

Pebesma, E. (2022). *stars: Spatiotemporal Arrays, Raster and Vector Data Cubes*.

R Core Team. (2021). *R: A Language and Environment for Statistical Computing* (4.1.2). https://www.r-project.org/

Tennekes, M. (2018). tmap : Thematic Maps in R. *Journal of Statistical Software*, *84*(6). https://doi.org/10.18637/jss.v084.i06

Ushey, K., Allaire, J. J., y Tang, Y. (2022). *reticulate: Interface to “Python”*.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Wosula, E. N., Tatineni, S., Wegulo, S. N., y Hein, G. L. (2017). Effect of Temperature on Wheat Streak Mosaic Disease Development in Winter Wheat. *Plant Disease*, *101*(2), 324–330. https://doi.org/10.1094/PDIS-07-16-1053-RE

Yim, Y., y Kira, T. (1975). Distribution of forest vegetation and climate in the Korean peninsula: I. distribution of some indices of thermal climate. *Japanese Journal of Ecology*, *25*, 77–88. https://doi.org/https://doi.org/10.18960/seitai.25.2\_77

## Figuras

|  |
| --- |
|  |
| Figura 1: Sitios de muestreo para los virus W*heat streak mosaic virus* (WSMV) y *High plains wheat mosaic virus* (HPWMoV) en trigo. En color celeste se resaltan las provincias muestreadas según reportes de enfermedad. |

|  |
| --- |
|  |
| Figura 2: Mapa de variabilidad espacial de: a) temperatura, b) precipitaciones acumuladas y c) temperatura de punto de rocío, promedios calculados utilizando el período de abril, mayo y junio de la región en estudio en los años donde se realizó el muestreo, d) Elevación de la región en estudio. |

|  |
| --- |
|  |
| Figura 3: Variabilidad espaciotemporal de precipitaciones acumuladas de abril, mayo y junio en los años muestreados en la región de estudio. |

|  |
| --- |
| Imagen en blanco y negro  Descripción generada automáticamente con confianza baja |
| Figura 4: Relación entre presencia y ausencia del virus *High plains wheat mosaic virus* (HPWMoV) en cultivo de Trigo, en función del valor de punto de rocío promedio, temperatura promedio y precipitaciones acumuladas promedio estimadas de las variables biometeorológicas durante el otoño cada dos semanas. |

|  |
| --- |
| Gráfico, Gráfico de dispersión  Descripción generada automáticamente |
| Figura 5: Relación entre presencia y ausencia del virus *Wheat streak mosaic virus* (WSMV) en Trigo, en función del valor de punto de rocío promedio, temperatura promedio y precipitaciones acumuladas promedio de las variables biometeorológicas obtenidas cada 14 días desde el comienzo del otoño. |