

Dataducto para construir Emuladores Estadísticos con datos de NARCCAP

Johan Fonseca¹ Marcela Alfaro-Córdoba¹ Luis Barboza¹

¹Centro de Investigación en Matemática Pura y Aplicada, Universidad de Costa Rica

1 Introducción

El proyecto de investigación B7166 de la Universidad de Costa Rica tiene como objetivo principal presentar un marco estadístico para realizar una emulación de variables obtenidas por medio del Modelo Climático Regional (MCR) usando variables obtenidas por medio de Modelos de Circulación General (MCG) como covariables, como una alternativa al MCR tradicional que es computacionalmente intensivos. Como complemen-

to a ese proyecto, esta es una propuesta para el flujo de trabajo con datos de Earth System Grid (repositorio de datos atmosféricos) utilizando R, que puede ser utilizado por cualquier otra persona interesada en hacer ejercicios de emulación estadística con los datos de Programa de Evaluación Regional de Cambio Climático de América del Norte (NARCCAP). El flujo de trabajo incluye la extracción de datos del repositorio, el resumen, modificación y crea-

ción de subconjuntos de los datos, y la obtención de un formato útil para construir los modelos estadísticos, así como un procedimiento de exploración y prueba de calidad de los datos. El flujo de trabajo es construido utilizando el conjunto de paquetes de tidyverse, además de otros paquetes para la exploración y visualización.

2 Diagrama de trabajo

Obtención



Utilizando el paquete Rvest se extraen datos que son manipulados con Dplyr para la creación de una tabla con la información de los distintos archivos de NARCCAP que pueden ser descargados con el uso del paquete Curl.



```
urls <-  
"https://www.earthsystemgrid.org/dataset/  
narccap.crcm.output.cesm.html"  
wp <- read_html(urls)  
inks1 <- html_nodes(wp, ".paddingTop div a")
```

Depuración

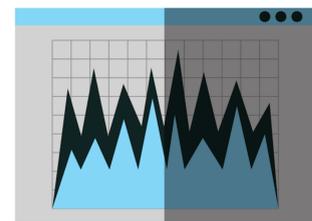


Los archivos descargados vienen dados con extensión .nc, por lo que se utiliza el paquete ncd4 para su lectura, stringr se utiliza para la lectura automática desde R del nombre de los archivos descargados. Luego el uso de Dplyr, Reshape2 para ordenar los datos en forma de tabla. Además el uso de Lubridate y Sp es en el proceso anterior para manipular las dimensiones temporales y espaciales que poseen los datos de NARCCAP.



```
Datos %>% mutate(Time=ymd_hms(as.character(Time)))  
%>%  
mutate(Year = year(Time),Month=month(Time)) %>%  
group_by(Year,Month,lon,lat) %>%  
summarise(Varm=mean(Var)) %>% ungroup()
```

Visualización

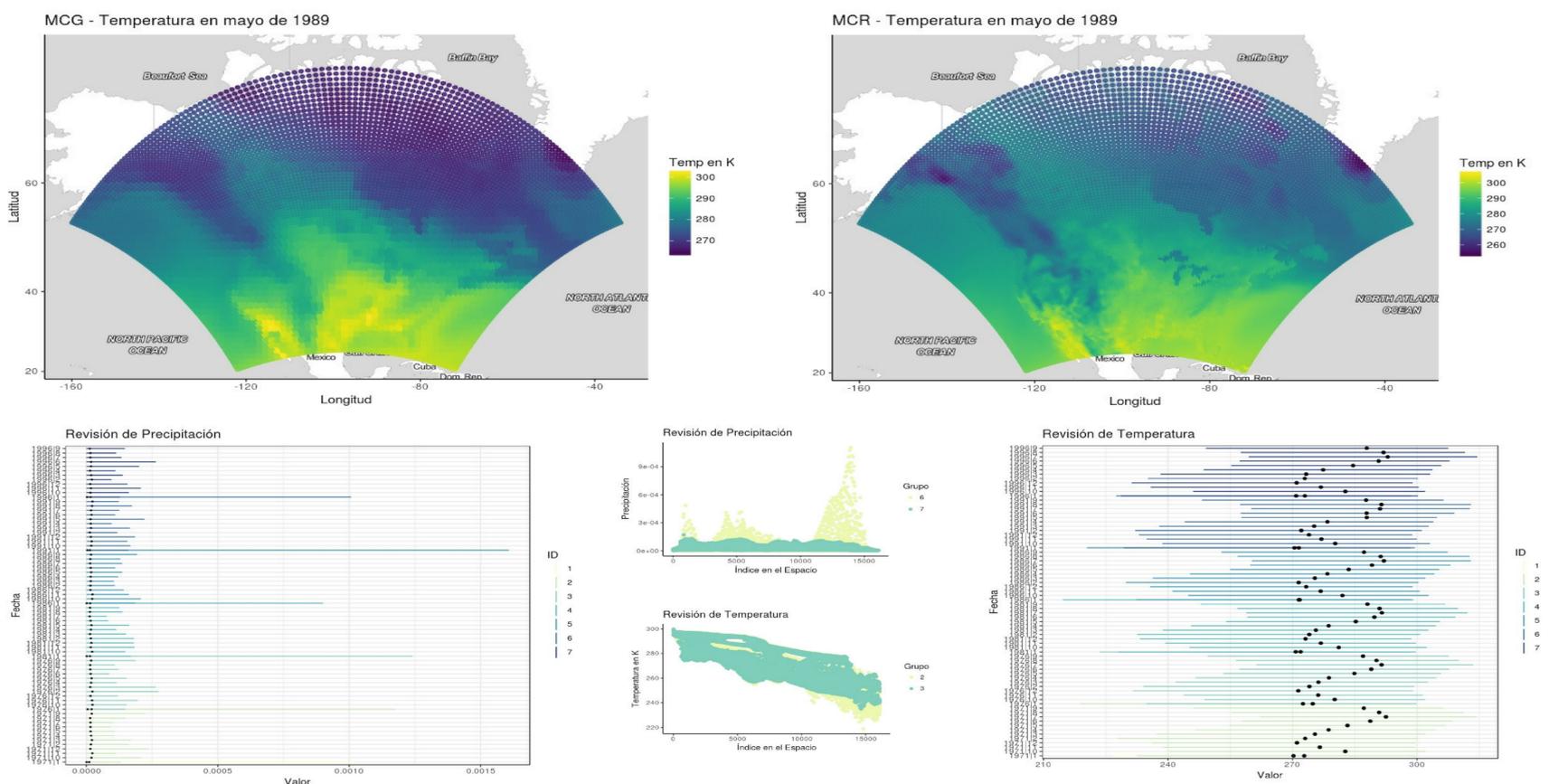


Una vez se tienen los datos en el formato deseado se realizan gráficos de control de calidad de los datos con ggplot2. Siendo uno de los problemas encontrados ubicaciones espacio-temporales repetidas con outliers.



```
ggmap(b) + geom_point(data =  
datos[datos$Year==YY&datos$Month==MM,],  
aes(lon-360,lat,color=TSm),alpha=0.9) +  
scale_color_viridis(name="Temp en K") +  
title=paste("Temperatura en ",mes[MM]," de",YY))  
theme_classic()
```

3 Visualización



4 Tareas pendientes

* Buscar un método de obtención de los datos del modelo global desde R.

* Crear un paquete con el formato de ROpenSci con este material.