

# Generación de catálogo de clima

*Centro Internacional de Agricultura Trópical grupo de Big-Data AEPS Hugo Dorado*

*April 26, 2018*

## Introducción

El procesamiento de datos climáticos es una parte fundamental en el análisis de Agricultura Específica por Sitio, debido a que el clima desempeña un rol importante en la explicación de la variación del rendimiento. En la práctica normalmente en la colecta de información, tenemos sitios donde se cultiva, que son georreferenciados y al adicionar una fecha de siembra y cosecha obtenemos una unidad observacional que llamaremos ciclo de cultivo **crop cycle**, el cual en algunas ocasiones también lo hemos llamado por el nombre de evento de cultivo **cropping event**. Por otro lado, tenemos otra fuente de información que son las estaciones meteorológicas, estos datos pueden venir con información diversa respecto a fechas de disponibilidad de información, también no siempre contienen todas las variables climáticas, por ejemplo tiende mucho a faltar variables tales como la radiación solar y humedad relativa.

La vinculación de la información de los ciclos de cultivo con las estaciones meteorológicas se hace por medio de una secuencia de pasos que involucran evaluar disponibilidad de información, distancias espaciales y diferencias de altura. Para ello se requiere por un lado tener la información procesada y cruda de las estaciones meteorológicas y por otro lado un catálogo de estaciones meteorológicas que reporte disponibilidad de información y fechas de inicio y finalización.

Para ello, hemos desarrollado el script **Stations\_Catalog.R**, el cual permite organizar la información climática en una estructura que facilita su consulta y genera insumos que serán utilizados en la asignación de eventos de cosecha a estaciones meteorológicas. Para poder ejecutarlo efectivamente es necesario haber realizado el control de calidad y estimación de datos faltantes que se puede encontrar en el proceso descrito en Control de datos de clima. Este manual se compone por una sesión de preparación de datos donde se organiza la estructura de los datos de entrada, luego la extracción de altitud de los registros georreferenciados y posteriormente la generación y estructuración del manual de clima. A modo de ejemplo reproducible, se generaron aleatoriamente 230 ciclos de cultivo (georeferenciación y fechas de siembra) en el estado de Chiapas México y se procesó 32 estaciones meteorológicas dentro del mismo estado, provenientes INEGI.

## Preparación de datos

Los comandos deben ser ejecutados utilizando el proyecto **Vinculacion\_clima\_lotes\_comerciales.Rproj** que está contenido dentro de la carpeta. La ubicación de los archivos para lectura y escritura se hará utilizando el paquete **here**. En el proceso manipularemos información geoespacial para extraer los datos de altura, por tanto necesitaremos la librería **raster**. También se hará uso de funciones de estructuración de matrices para lo cual se utilizará la librería **reshape**.

```
rm(list=ls())

library(here)
library(raster)
library(reshape2)
```

### Datos de entrada

#### Estaciones meteorológicas

Para procesar los datos de las estaciones meteorológicas se requiere crear un archivo de csv, como se muestra continuación, lo ideal es conservar exactamente los mismos nombres de los encabezados, donde la primera columna **Station\_Name** debe indicar el nombre o código de la estación meteorológica y este deber ser único por estación, seguido por la latitud (**Latitude**) y longitud (**Longitude**) en escala de decimales, posteriormente la fecha de comienzo (**Ini\_Date**) y final (**End\_Date**) de datos de la estación, y por último cada variable (**Variable**) contenida dentro de la estación separada en cada fila.

```
# Reading datasets

stations <- read.csv(here::here('BASIC_FILES', 'estaciones_variables_fechas.csv'))
head(stations)
```

```
##      Station_Name Latitude Longitude Ini_Date   End_Date Variable
## 1 Chiapas5deMarzo 15.04469 -92.53328 3/1/2009 9/20/2015      P
## 2 Chiapas5deMarzo 15.04469 -92.53328 3/1/2009 9/20/2015     RH
## 3 Chiapas5deMarzo 15.04469 -92.53328 3/1/2009 9/20/2015     TM
## 4 Chiapas5deMarzo 15.04469 -92.53328 3/1/2009 9/20/2015     TX
## 5 ChiapasAgroipsa 17.58973 -91.94636 9/3/2013 12/13/2017      P
## 6 ChiapasAgroipsa 17.58973 -91.94636 9/3/2013 12/13/2017     RH
```

## Ciclos de cultivo

En cuanto a georreferenciación y fechas, el formato es similar al de estaciones meteorológicas, el identificador de cada cultivo debe ser nombrado por una variable llamada **FID**

```
crop_cycles <- read.csv(here::here('BASIC_FILES', 'cultivos_georeferenciados.csv'))
head(crop_cycles)
```

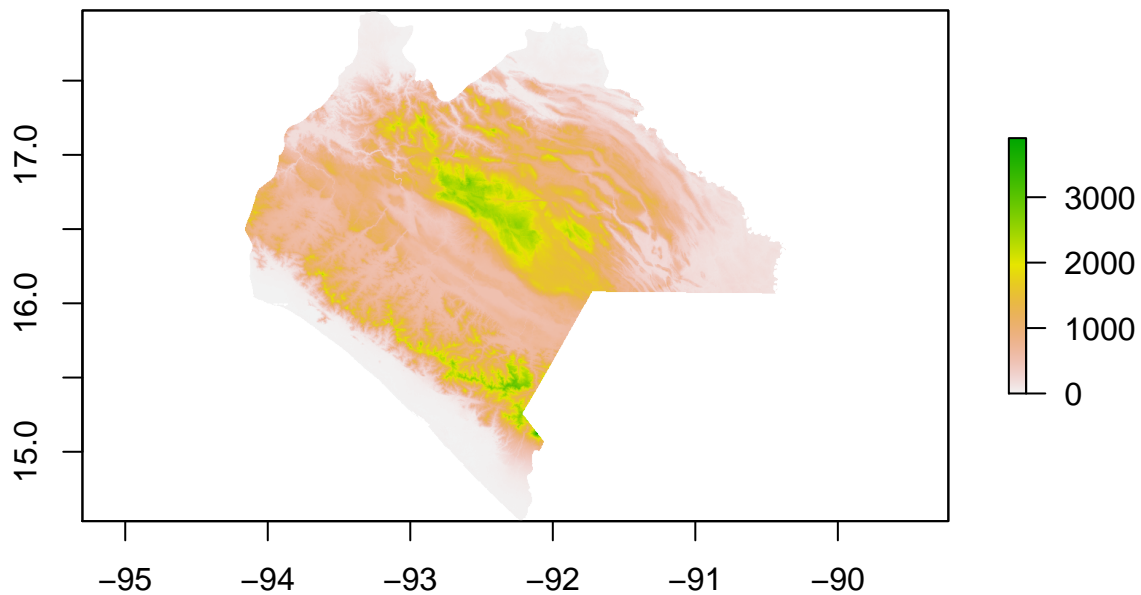
```
##   FID Latitude Longitude   Ini_Date   End_Date
## 1   0 15.16654 -92.81051 3/1/2009 6/30/2009
## 2   1 15.16966 -92.79103 9/3/2013 1/2/2014
## 3   2 15.36830 -92.78351 4/18/2011 8/17/2011
## 4   3 15.18909 -92.77942 3/1/2009 6/30/2009
## 5   4 15.27548 -92.62670 11/10/2007 3/10/2008
## 6   5 15.22505 -92.87760 7/1/2008 10/30/2008
```

Tanto para las estaciones meteorológicas como para ciclos de cultivo, se debe haber echo una revisión de que las fechas de inicio no sean mayores a las de finalización, que los ciclos de siembra tengan una duración razonable y que no hayan datos perdidos o faltantes, se recomienda no utilizar ni acentos ni letras que puedan generar inconvenientes con la codificación.

## Datos de altura

Como parte del procesamiento, es necesario estimar la altura de las estaciones y los lotes, dado que esto hace parte del criterio para evaluar que tan similar puede ser el clima de dos localidades, para ello es necesario contar con un archivo en formato de raster **chiapas\_srtm** de la localidad que se esté analizando, en este caso Chiapas.

```
elevation <- raster(here::here('BASIC_FILES', 'chiapas_srtm'))
plot(elevation)
```



### Ajustar formato de fecha

Posterior a la lectura de datos, es necesario especificar el formato de fecha a través de la función `as.Date` e indicando en el campo de fecha, cual es la ubicación del mes, día y año.

```
# Date formate

stations$Ini_Date    <- as.Date(stations$Ini_Date,'%m/%d/%Y')
stations$End_Date    <- as.Date(stations$End_Date,'%m/%d/%Y')
crop_cycles$Ini_Date <- as.Date(crop_cycles$Ini_Date,'%m/%d/%Y')
crop_cycles$End_Date <- as.Date(crop_cycles$End_Date,'%m/%d/%Y')
```

### Extraer altitud de las estaciones y los lotes

En este paso se realiza el cálculo del valor de altura de cada evento de cosecha y estación meteorológica utilizando los datos de coordenadas y el objeto raster que contiene el mapa de altura de Chiapas, cada altura es agregada a la base de datos,

```
# Elevation estimation

stations$Elevation <- extract(elevation ,stations[c('Longitude','Latitude')])
```

```
head(stations)
```

```
##      Station_Name Latitude Longitude   Ini_Date   End_Date Variable
## 1 Chiapas5deMarzo 15.04469 -92.53328 2009-03-01 2015-09-20      P
## 2 Chiapas5deMarzo 15.04469 -92.53328 2009-03-01 2015-09-20     RH
## 3 Chiapas5deMarzo 15.04469 -92.53328 2009-03-01 2015-09-20     TM
## 4 Chiapas5deMarzo 15.04469 -92.53328 2009-03-01 2015-09-20     TX
## 5 ChiapasAgroipsa 17.58973 -91.94636 2013-09-03 2017-12-13      P
## 6 ChiapasAgroipsa 17.58973 -91.94636 2013-09-03 2017-12-13     RH
##      Elevation
## 1           10
## 2           10
## 3           10
## 4           10
## 5           59
## 6           59
```

```
crop_cycles$Elevation <- extract(elevation ,crop_cycles[c('Longitude','Latitude')])
```

```
head(crop_cycles)
```

```
##      FID Latitude Longitude   Ini_Date   End_Date Elevation
## 1    0 15.16654 -92.81051 2009-03-01 2009-06-30         0
## 2    1 15.16966 -92.79103 2013-09-03 2014-01-02        16
## 3    2 15.36830 -92.78351 2011-04-18 2011-08-17        40
## 4    3 15.18909 -92.77942 2009-03-01 2009-06-30        10
## 5    4 15.27548 -92.62670 2007-11-10 2008-03-10       100
## 6    5 15.22505 -92.87760 2008-07-01 2008-10-30        22
```

Como parte del procesamiento, se descartan filas en las cuales la altura no ha sido posible calcularse y han quedado con valores de NA, debido a problemas con el raster, o que han quedado por fuera del área de cobertura del mismo.

```
# Remove stations or crop_cycles with missing values in elevation
```

```
stations <- stations[!is.na(stations$Elevation),]
```

```
crop_cycles <- crop_cycles[!is.na(crop_cycles$Elevation),]
```

## Generar catálogo de clima

Para generar el catálogo de clima, se estructuran los datos en una forma longitudinal de manera que la disponibilidad de cada variable se pueda observar en una sola fila en la base de datos.

```
# Generation of the variables catalog
```

```
station_groups_Date <- reshape(stations,v.names=c("Ini_Date","End_Date"),
                                idvar = c("Station_Name"),
                                timevar= "Variable",direction = "wide",sep = "_")
```

```
head(station_groups_Date)
```

```
##      Station_Name Latitude Longitude Elevation Ini_Date_P End_Date_P
## 1    Chiapas5deMarzo 15.04469 -92.53328         10 2009-03-01 2015-09-20
## 5    ChiapasAgroipsa 17.58973 -91.94636         59 2013-09-03 2017-12-13
```

```
## 10 ChiapasAlpujarras 15.07486 -92.16931      998      <NA>      <NA>
## 14   ChiapasArgelia 14.83969 -92.50369        4      <NA>      <NA>
## 17   ChiapasArgovia 15.12858 -92.29531     712 2007-11-10 2015-09-20
## 20   ChiapasAustralia 16.25031 -92.46119     601 2008-07-01 2013-03-05
##      Ini_Date_RH End_Date_RH Ini_Date_TM End_Date_TM Ini_Date_TX End_Date_TX
## 1    2009-03-01  2015-09-20  2009-03-01  2015-09-20  2009-03-01  2015-09-20
## 5    2013-09-03  2017-12-13  2013-09-03  2017-12-13  2013-09-03  2017-12-13
## 10   2011-04-18  2015-09-20  2011-04-18  2015-09-20  2011-04-18  2015-09-20
## 14      <NA>      <NA>  2009-03-01  2015-09-20  2009-03-01  2015-09-20
## 17      <NA>      <NA>  2007-11-10  2015-09-20  2007-11-10  2015-09-20
## 20      <NA>      <NA>  2008-07-01  2013-03-05  2008-07-01  2013-03-05
##      Ini_Date_SR End_Date_SR
## 1      <NA>      <NA>
## 5    2013-09-03  2017-12-13
## 10   2011-04-18  2015-09-20
## 14   2009-03-01  2015-09-20
## 17      <NA>      <NA>
## 20   2008-07-01  2013-03-05
```

## Disponibilidad de datos por cada variable

También agregamos un indicador binario siendo 1 disponibilidad de datos y 0 ausencia de datos de dicha estación (fila) en dicha variable (columna).

```
# Weather variables availability inspection

stations$availability <- 1

stations_availability <- dcast(stations, Station_Name ~ Variable, sum)

Stations_catalog <- merge(station_groups_Date, stations_availability, by='Station_Name', all=T, sort = F)

head(Stations_catalog)
```

```
##      Station_Name Latitude Longitude Elevation Ini_Date_P End_Date_P
## 1   Chiapas5deMarzo 15.04469 -92.53328        10 2009-03-01 2015-09-20
## 2   ChiapasAgroipsa 17.58973 -91.94636         59 2013-09-03 2017-12-13
## 3 ChiapasAlpujarras 15.07486 -92.16931        998      <NA>      <NA>
## 4   ChiapasArgelia 14.83969 -92.50369         4      <NA>      <NA>
## 5   ChiapasArgovia 15.12858 -92.29531        712 2007-11-10 2015-09-20
## 6 ChiapasAustralia 16.25031 -92.46119        601 2008-07-01 2013-03-05
##      Ini_Date_RH End_Date_RH Ini_Date_TM End_Date_TM Ini_Date_TX End_Date_TX
## 1    2009-03-01  2015-09-20  2009-03-01  2015-09-20  2009-03-01  2015-09-20
## 2    2013-09-03  2017-12-13  2013-09-03  2017-12-13  2013-09-03  2017-12-13
## 3    2011-04-18  2015-09-20  2011-04-18  2015-09-20  2011-04-18  2015-09-20
## 4      <NA>      <NA>  2009-03-01  2015-09-20  2009-03-01  2015-09-20
## 5      <NA>      <NA>  2007-11-10  2015-09-20  2007-11-10  2015-09-20
## 6      <NA>      <NA>  2008-07-01  2013-03-05  2008-07-01  2013-03-05
##      Ini_Date_SR End_Date_SR P RH SR TM TX
## 1      <NA>      <NA> 1 1 0 1 1
## 2    2013-09-03  2017-12-13 1 1 1 1 1
## 3    2011-04-18  2015-09-20 0 1 1 1 1
## 4    2009-03-01  2015-09-20 0 0 1 1 1
## 5      <NA>      <NA> 1 0 0 1 1
## 6    2008-07-01  2013-03-05 1 0 1 1 1
```

## Estructurar el orden de las variables

Agregamos una variable que indique la fecha inicial más temprana y la fecha final más tardía, finalmente redistribuimos las posiciones de las variables con el fin de tener una mejor organización. De la misma manera, en la base de datos de eventos de cosecha, reagrupamos el orden de variables.

```
# Sorting the variables
```

```
nams_catalog <- names(Stations_catalog)
```

```
Stations_catalog$Min_Date <- apply(Stations_catalog[grep('Ini',nams_catalog)],1,min,na.rm=T)
```

```
Stations_catalog$Max_Date <- apply(Stations_catalog[grep('End',nams_catalog)],1,min,na.rm=T)
```

```
head(Stations_catalog)
```

```
##      Station_Name Latitude Longitude Elevation Ini_Date_P End_Date_P
## 1 Chiapas5deMarzo 15.04469 -92.53328      10 2009-03-01 2015-09-20
## 2 ChiapasAgroipsa 17.58973 -91.94636      59 2013-09-03 2017-12-13
## 3 ChiapasAlpujarras 15.07486 -92.16931     998      <NA>      <NA>
## 4 ChiapasArgelia 14.83969 -92.50369        4      <NA>      <NA>
## 5 ChiapasArgovia 15.12858 -92.29531     712 2007-11-10 2015-09-20
## 6 ChiapasAustralia 16.25031 -92.46119     601 2008-07-01 2013-03-05
##      Ini_Date_RH End_Date_RH Ini_Date_TM End_Date_TM Ini_Date_TX End_Date_TX
## 1 2009-03-01 2015-09-20 2009-03-01 2015-09-20 2009-03-01 2015-09-20
## 2 2013-09-03 2017-12-13 2013-09-03 2017-12-13 2013-09-03 2017-12-13
## 3 2011-04-18 2015-09-20 2011-04-18 2015-09-20 2011-04-18 2015-09-20
## 4      <NA>      <NA> 2009-03-01 2015-09-20 2009-03-01 2015-09-20
## 5      <NA>      <NA> 2007-11-10 2015-09-20 2007-11-10 2015-09-20
## 6      <NA>      <NA> 2008-07-01 2013-03-05 2008-07-01 2013-03-05
##      Ini_Date_SR End_Date_SR P RH SR TM TX  Min_Date  Max_Date
## 1      <NA>      <NA> 1 1 0 1 1 2009-03-01 2015-09-20
## 2 2013-09-03 2017-12-13 1 1 1 1 1 2013-09-03 2017-12-13
## 3 2011-04-18 2015-09-20 0 1 1 1 1 2011-04-18 2015-09-20
## 4 2009-03-01 2015-09-20 0 0 1 1 1 2009-03-01 2015-09-20
## 5      <NA>      <NA> 1 0 0 1 1 2007-11-10 2015-09-20
## 6 2008-07-01 2013-03-05 1 0 1 1 1 2008-07-01 2013-03-05
```

```
Stations_catalog <- Stations_catalog[,c(1:4,15:19,5:14,20:21)]
```

```
head(Stations_catalog)
```

```
##      Station_Name Latitude Longitude Elevation P RH SR TM TX Ini_Date_P
## 1 Chiapas5deMarzo 15.04469 -92.53328      10 1 1 0 1 1 2009-03-01
## 2 ChiapasAgroipsa 17.58973 -91.94636      59 1 1 1 1 1 2013-09-03
## 3 ChiapasAlpujarras 15.07486 -92.16931     998 0 1 1 1 1      <NA>
## 4 ChiapasArgelia 14.83969 -92.50369        4 0 0 1 1 1      <NA>
## 5 ChiapasArgovia 15.12858 -92.29531     712 1 0 0 1 1 2007-11-10
## 6 ChiapasAustralia 16.25031 -92.46119     601 1 0 1 1 1 2008-07-01
##      End_Date_P Ini_Date_RH End_Date_RH Ini_Date_TM End_Date_TM Ini_Date_TX
## 1 2015-09-20 2009-03-01 2015-09-20 2009-03-01 2015-09-20 2009-03-01
## 2 2017-12-13 2013-09-03 2017-12-13 2013-09-03 2017-12-13 2013-09-03
## 3      <NA> 2011-04-18 2015-09-20 2011-04-18 2015-09-20 2011-04-18
## 4      <NA>      <NA>      <NA> 2009-03-01 2015-09-20 2009-03-01
## 5 2015-09-20      <NA>      <NA> 2007-11-10 2015-09-20 2007-11-10
## 6 2013-03-05      <NA>      <NA> 2008-07-01 2013-03-05 2008-07-01
```

```
##   End_Date_TX Ini_Date_SR End_Date_SR   Min_Date   Max_Date
## 1  2015-09-20      <NA>      <NA> 2009-03-01 2015-09-20
## 2  2017-12-13 2013-09-03 2017-12-13 2013-09-03 2017-12-13
## 3  2015-09-20 2011-04-18 2015-09-20 2011-04-18 2015-09-20
## 4  2015-09-20 2009-03-01 2015-09-20 2009-03-01 2015-09-20
## 5  2015-09-20      <NA>      <NA> 2007-11-10 2015-09-20
## 6  2013-03-05 2008-07-01 2013-03-05 2008-07-01 2013-03-05
```

```
crop_cycles <- crop_cycles[,c(1:3,6,4:5)]
```

```
head(crop_cycles)
```

```
##   FID Latitude Longitude Elevation   Ini_Date   End_Date
## 1   0 15.16654 -92.81051          0 2009-03-01 2009-06-30
## 2   1 15.16966 -92.79103         16 2013-09-03 2014-01-02
## 3   2 15.36830 -92.78351         40 2011-04-18 2011-08-17
## 4   3 15.18909 -92.77942         10 2009-03-01 2009-06-30
## 5   4 15.27548 -92.62670        100 2007-11-10 2008-03-10
## 6   5 15.22505 -92.87760         22 2008-07-01 2008-10-30
```

## Almacenar resultados

Finalmente almacenamos el catálogo de estaciones climáticas en un csv para una posterior consulta y también lo almacenamos junto al de ciclo de cosecha en un RData

```
# Save outputs files
```

```
write.csv(Stations_catalog,here::here("RESULTS","Stations_catalog.csv"),row.names=F)
```

```
save(Stations_catalog,crop_cycles,file = here::here("RESULTS","Catalogs.RData"))
```

(C) 2018, Grupo Big Data y Agricultura Específica por Sitio, Centro Internacional de Agricultura Trópic