

Asignación de clima a ciclos de cultivo

Centro Internacional de Agricultura Trópica grupo de Big-Data AEPS Hugo Dorado

April 26, 2018

Introducción

La asignación de variables climáticas a ciclos de cultivo utilizando estaciones meteorológicas, se realiza por medio de un proceso que consiste inicialmente en definir la coincidencia entre sitios de cultivo y ubicación de las estaciones, dicha coincidencia se define por medio de diferencias entre altura y proximidad espacial, las cuales se evalúan por medio de restricciones que pueden variar por variable (por ejemplo la distancia aplicada para temperatura puede ser menos rigurosa que la distancia en precipitación). Un paso posterior a evaluar las coincidencias es el de calcular indicadores climáticos por cada una de las variables de interés dentro del ciclo, para lo cual se utilizan promedios, desviaciones estándar, acumulados y frecuencias.

El script `Distances_Process.R`, permite realizar el procesamiento de estaciones meteorológicas en escala diaria y ciclos de cultivo, con el propósito de vincularlos y generar indicadores climáticos que puedan ser de utilidad en el futuro para asociar información de suelo, manejo agronómico y rendimiento. Para que el proceso pueda ser ejecutado efectivamente se debe realizar una buena preparación de los datos utilizados, especialmente los de las estaciones meteorológicas.

El documento se divide en una parte de preparación de datos donde se detalla la estructura de los archivos iniciales y los filtros de información relevante, seguido de la vinculación de ciclos de cultivo a estaciones meteorológicas que permite aplicar las restricciones de distancia y altura, luego una sesión para identificar por variable la estación más próxima a cada ciclo de cultivo, posteriormente se generan indicadores climáticos, y finalmente se almacenan los datos procesados. El propósito es que en el futuro, otros investigadores, técnicos, académicos entre otros puedan replicar estos pasos para analizar el efecto del clima en sitios cercanos a estaciones meteorológicas.

Preparación de datos

Los paquetes necesarios, son `here` para leer y escribir archivos con la ubicación asignada para el proyecto, `data.table`, que es un paquete que permite manipular bases de datos de manera más eficiente, `plyr` que es un paquete de procesamiento y resumen de bases de datos, `Imap` que es una herramienta de mapeo interactivo y en este caso se usa para calcular las distancias entre puntos georeferenciados y `pgriness` que es un paquete de análisis espacial y minería de datos

```
rm(list = ls())

library(here)
library(data.table)
library(plyr)
library(Imap)
library(pgriness)
```

Archivos iniciales

El primer archivo que debemos cargar por medio de la función `source` es `Merge_Station_Funs.R`, dado muchas de las funciones que utilizamos posteriormente están programadas allí, lo ideal es que esté de la carpeta `SCRIPTS`

```
source(here::here('SCRIPTS', 'Merge_Stations_Funs.R'))
```

El siguiente archivo a cargar es el catálogo de estaciones climáticas, el cual debe contener a su vez una base de datos de ciclos de cultivo con la altura incluida, `Catalogs.RData` es un archivo que se genera en el proceso descrito en Generación de catálogo de clima, que está dentro de esta misma carpeta.

```
load(here::here('RESULTS', 'Catalogs.RData'))
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
```

```
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
```

`list_Station_Unprocess.RData` y `list_Stations_Process.RData` son archivos que provienen del proceso de limpieza y control de procesamiento de datos, `list_Station_Unprocess.RData` es una que lista contiene en cada componente un `data.frame` con fecha y medición de cada variable por separado, dicha medición está en un formato sin procesar, es decir puede contener NA's. El nombre de cada componente de la lista está compuesto por el identificador de la estación meteorológica y separado por raya piso (`_`) el nombre de la variable, (P,RH,SR,TM,TX).

```
load(here::here('BASIC_FILES', 'list_Station_Unprocess.RData'))
lapply(list_Stations_Unprocess[1:3], head)
```

```
## $ChiapasYunnuenRanchoAlegre_TX
##      Date var
## 1 2006-01-01 NA
## 2 2006-01-02 NA
## 3 2006-01-03 NA
```

```
## 4 2006-01-04 NA
## 5 2006-01-05 NA
## 6 2006-01-06 NA
##
## $ChiapasYunnuenRanchoAlegre_TM
##      Date var
## 1 2006-01-01 NA
## 2 2006-01-02 NA
## 3 2006-01-03 NA
## 4 2006-01-04 NA
## 5 2006-01-05 NA
## 6 2006-01-06 NA
##
## $ChiapasYunnuenRanchoAlegre_SR
##      Date      var
## 1 2007-11-08 113.19
## 2 2007-11-09 535.33
## 3 2007-11-10 491.75
## 4 2007-11-11 529.47
## 5 2007-11-12 550.99
## 6 2007-11-13 456.63
```

Por otro lado `list_Stations_Process.RData`, tiene los archivos ya procesados de las estaciones meteorológicas, siendo en este caso, listas que contienen `data.frames` con columnas de información limpias, y las variables trabajadas ya consolidadas en una tabla. El nombre corresponde con el identificador de la estación meteorológica.

```
load(here::here('BASIC_FILES', 'list_Stations_Process.RData'))
lapply(list_Stations_Process[1:3], head)
```

```
## $Chiapas5deMarzo
##      Date P    RH    SR    TM    TX
## 1 2009-03-01 1 78.81 406.05 17.1 32.2
## 2 2009-03-02 0 59.62 565.66 16.2 33.9
## 3 2009-03-03 0 68.06 557.36 15.5 33.1
## 4 2009-03-04 0 68.93 552.14 15.8 32.2
## 5 2009-03-05 0 57.01 547.86 15.0 33.7
## 6 2009-03-06 0 65.85 548.87 14.5 34.5
##
## $ChiapasAgroipisa
##      Date    P    RH    SR    TM    TX
## 1 2013-09-03 0.4 76.36 603.66 22.81849 32.97380
## 2 2013-09-04 0.0 86.14 492.61 22.15026 34.05704
## 3 2013-09-05 20.5 89.40 478.49 23.13098 33.98811
## 4 2013-09-06 0.7 90.14 390.46 24.30158 34.77380
## 5 2013-09-07 8.1 85.76 560.86 24.23131 35.02532
## 6 2013-09-08 41.8 89.16 428.35 23.32564 34.31834
##
## $ChiapasAlpujarras
##      Date    P    RH    SR    TM    TX
## 1 2011-04-18 1.2 88.41 221.49 19.7 27.2
## 2 2011-04-19 3.4 89.77 231.87 18.3 24.1
## 3 2011-04-20 52.2 91.05 287.99 17.5 25.3
## 4 2011-04-21 93.8 90.66 274.39 17.8 24.3
## 5 2011-04-22 43.4 92.78 194.94 15.9 23.5
```

```
## 6 2011-04-23 57.2 75.46 480.77 17.5 25.9
```

Filtrar por información útil de acuerdo a los periodos de interés y disponibilidad de datos

Descartar estaciones cuya información esté por fuera de las fechas de los ciclos de cultivo

```
# Filter stations per availability

earliest_date_crop_cycle <- min(crop_cycles$Ini_Date)
latest_data_crop_cycle <- max(crop_cycles$End_Date)

rule1 <- earliest_date_crop_cycle > Stations_catalog$Max_Date
rule2 <- latest_data_crop_cycle < Stations_catalog$Min_Date

Stations_target <- Stations_catalog[!(rule1|rule2),]
```

Descartar ciclos de cultivo cuya información esté por fuera de las fechas de las estaciones

```
# Filter station per cropping cycles

crop_cycles <- crop_cycles[min(Stations_target$Min_Date) < crop_cycles$Ini_Date,]
```

Vinculación de ciclos de cultivo a estaciones meteorológicas

La función `merge_Stations` permite coincidir ciclos de cultivo y estaciones meteorológicas de acuerdo a restricciones de distancia conservando unicamente las relaciones que cumplan los criterios mínimos, en este caso por defecto tenemos que para temperatura se admite una distancia máxima de 30.000 metros, 15.000 metros para precipitación y 150 metros de diferencia en elevación.

```
# Merge the station and cropping cycles according distances rules that are specified in the function,  
# the time period between station - cropping cycles is validated as well, station and crop_cycles which  
# don't
```

```
merge_station <- merge_Stations(Stations_target, crop_cycles, crop_cycle_Name= 'FID',  
                                Dtemp = 30000, Drain = 15000,  
                                Drhum = 30000, Drads = 30000,  
                                DifElv = 150)
```

```
lapply(merge_station, head)
```

```
## $usefullStation  
##      Station ID_crop_cycle      dist crop_cycle_Latitude  
## 1 ChiapasTezuitlan         0 16091.07         15.16654  
## 2 ChiapasRanchoSanJosé     0 24650.23         15.16654  
## 3 ChiapasRanchoSanJosé     1 22654.96         15.16966  
## 4 ChiapasTezuitlan         2 25641.18         15.36830  
## 5 ChiapasTezuitlan         3 12929.15         15.18909  
## 6 ChiapasRanchoSanJosé     3 21984.61         15.18909  
##      crop_cycle_Longitude crop_cycle_Elevation crop_cycle_Ini_Date  
## 1          -92.81051         0         2009-03-01  
## 2          -92.81051         0         2009-03-01  
## 3          -92.79103        16         2013-09-03  
## 4          -92.78351        40         2011-04-18
```

```

## 5          -92.77942          10          2009-03-01
## 6          -92.77942          10          2009-03-01
##   crop_cycle_End_Date Stat_Latitude Stat_Longitude Stat_Elevation Stat_P
## 1          2009-06-30          15.16953          -92.66081          16          1
## 2          2009-06-30          15.12906          -92.58442          14          1
## 3          2014-01-02          15.12906          -92.58442          14          1
## 4          2011-08-17          15.16953          -92.66081          16          1
## 5          2009-06-30          15.16953          -92.66081          16          1
## 6          2009-06-30          15.12906          -92.58442          14          1
##   Stat_RH Stat_SR Stat_TM Stat_TX Stat_Ini_Date_P Stat_End_Date_P
## 1          1          1          1          1          2009-03-01          2012-10-09
## 2          1          1          1          1          2008-05-09          2015-09-20
## 3          1          1          1          1          2008-05-09          2015-09-20
## 4          1          1          1          1          2009-03-01          2012-10-09
## 5          1          1          1          1          2009-03-01          2012-10-09
## 6          1          1          1          1          2008-05-09          2015-09-20
##   Stat_Ini_Date_RH Stat_End_Date_RH Stat_Ini_Date_TM Stat_End_Date_TM
## 1          2009-03-01          2012-10-09          2009-03-01          2012-10-09
## 2          2008-05-09          2015-09-20          2008-05-09          2015-09-20
## 3          2008-05-09          2015-09-20          2008-05-09          2015-09-20
## 4          2009-03-01          2012-10-09          2009-03-01          2012-10-09
## 5          2009-03-01          2012-10-09          2009-03-01          2012-10-09
## 6          2008-05-09          2015-09-20          2008-05-09          2015-09-20
##   Stat_Ini_Date_TX Stat_End_Date_TX Stat_Ini_Date_SR Stat_End_Date_SR
## 1          2009-03-01          2012-10-09          2009-03-01          2012-10-09
## 2          2008-05-09          2015-09-20          2008-05-09          2015-09-20
## 3          2008-05-09          2015-09-20          2008-05-09          2015-09-20
## 4          2009-03-01          2012-10-09          2009-03-01          2012-10-09
## 5          2009-03-01          2012-10-09          2009-03-01          2012-10-09
## 6          2008-05-09          2015-09-20          2008-05-09          2015-09-20
##   Stat_Min_Date Stat_Max_Date difElevation UsefullData
## 1          2009-03-01          2012-10-09          16 TX,TM,SR,RH
## 2          2008-05-09          2015-09-20          14 TX,TM,SR,RH
## 3          2008-05-09          2015-09-20           2 TX,TM,SR,RH
## 4          2009-03-01          2012-10-09          24 TX,TM,SR,RH
## 5          2009-03-01          2012-10-09           6 TX,TM,P,SR,RH
## 6          2008-05-09          2015-09-20           4 TX,TM,SR,RH
##
## $StationsList
##           Station Stat_Elevation Total_crop_cycles_coverage
## 1 Chiapas5deMarzo           10                26
## 2 ChiapasAgroipsa           59                 6
## 3 ChiapasAlpujarras          998                18
## 4 ChiapasArgelia             4                 18
## 5 ChiapasArgovia            712                17
## 6 ChiapasAustralia          601                 5

```

Existe la posibilidad que posterior al paso anterior, aún algunos eventos de cultivo no puedan ser cubiertos por las estaciones que están disponibles para la información climática, esto debido a que no cumplen los criterios establecidos, es decir están muy lejos o tienen una diferencia de altura muy grande, por tal razón estos eventos de cultivo son removidos de la base de datos de `crop_cycles`.

```
# Extract the usefullStation list
```

```
usefullStation <- merg_Station$usefullStation
```

```
# Selecting only fileds which have avaiable stations
```

```
crop_cycles <- crop_cycles[crop_cycles$FID %in% usefullStation$ID_crop_cycle,]
```

Identificar por variable, la estación que mejor represente la información climática.

La función `crop_cycle_vars_station_assignment`, está diseñada para definir cuál es la estación que mejor representa a cada variable que va a ser asignada a un ciclo de cultivo, para ello elegirá la estación con menor distancia, menor diferencia de altura y menor cantidad de datos perdidos en dicho periodo.

```
# Assigning each variable for each station
```

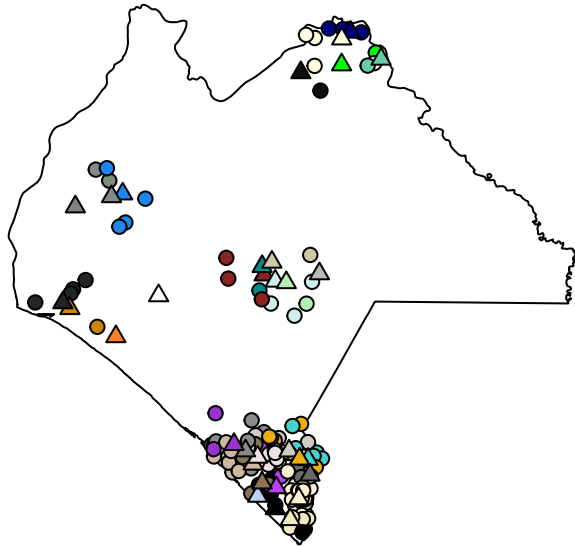
```
crop_cycle_vars_station_assignment <- assing_var_station(crop_cycles,usefullStation)
```

```
head(crop_cycle_vars_station_assignment)
```

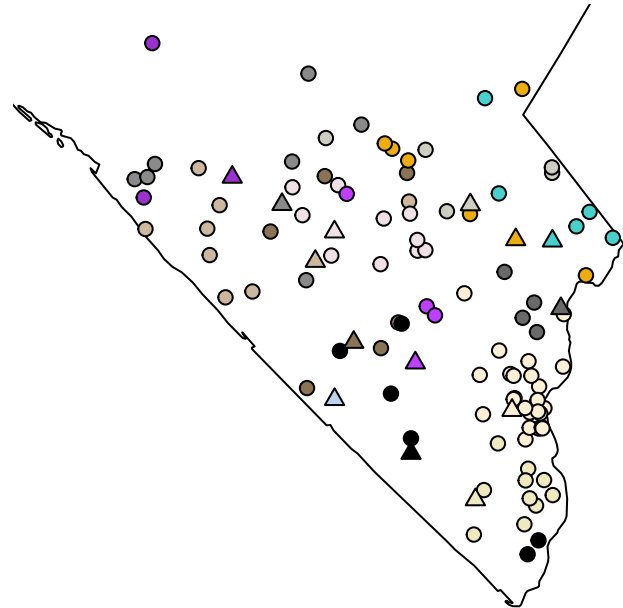
##	crop_cycle	Variable	Station	Ini_Date	End_Date
## 1	0	SR	ChiapasTezuitlan	2009-03-01	2009-06-30
## 2	0	P	ChiapasTezuitlan	2009-03-01	2009-06-30
## 3	0	RH	ChiapasTezuitlan	2009-03-01	2009-06-30
## 4	0	TX	ChiapasRanchoSanJosé	2009-03-01	2009-06-30
## 5	0	TM	ChiapasRanchoSanJosé	2009-03-01	2009-06-30
## 6	1	SR	ChiapasRanchoSanJosé	2013-09-03	2014-01-02

La figura a continuación muestra la distribución de estaciones (triángulos) y ciclos de cultivo (círculos), de acuerdo a la asignación obtenida en temperatura máxima representada por el mismo color.

Puntos vs estaciones (Tmax)



Puntos vs estaciones (Tmax) Chiapas sur



Por último se aplica la función `weather_assignment`, que permite traer los valores reales de las variables de las estaciones meteorológicas y agruparlas en un `data.frame` o matriz donde cada fila representa un día del ciclo. El objeto final es una lista que contiene la historia climática a nivel diario de cada ciclo de cultivo.

Assigning daily weather information for each event variable

```
Final_weather_assign <- weather_assignment(crop_cycle_vars_station_assignment, list_Stations_Process)
```

```
lapply(Final_weather_assign[1:3], head)
```

```
## $'0'
```

```
##      date      SR P    RH    TX    TM
## 1 2009-03-01    0.00 0 71.42 25.6 19.0
## 2 2009-03-02 581.80 0 59.02 34.9 15.7
## 3 2009-03-03 540.56 0 70.62 33.8 14.8
## 4 2009-03-04 545.25 0 72.96 32.4 14.3
## 5 2009-03-05 540.59 0 60.32 33.7 14.7
## 6 2009-03-06 531.51 0 70.03 34.3 13.9
##
```

```
## $'1'
```

```
##      date      SR    P    RH    TX    TM
## 1 2013-09-03 198.5278 131.4 91.39035 32.50599 18.630246
## 2 2013-09-04 424.0631  83.8 65.32661 32.18667  8.836046
## 3 2013-09-05 477.8690 141.6 62.06322 29.62901 19.787441
## 4 2013-09-06 417.2395 108.0 58.17717 32.60210  8.402100
## 5 2013-09-07 437.2993  83.8 75.54668 31.82549 18.739802
## 6 2013-09-08 401.3285 108.0 63.23239 32.31289 22.303101
```

```
##
## $'2'
##      date      SR   P    RH   TX   TM
## 1 2011-04-18 327.22 0.0 80.64 32.8 23.3
## 2 2011-04-19 291.50 0.0 79.32 32.7 22.2
## 3 2011-04-20 387.24 0.0 76.67 32.9 22.9
## 4 2011-04-21 319.98 0.0 81.65 31.5 21.5
## 5 2011-04-22 491.10 0.2 76.32 33.8 22.5
## 6 2011-04-23 494.18 0.0 69.69 35.2 20.5
```

El resultado que se obtiene en este paso sirve como insumo para realizar clustering de eventos de clima e indicadores climáticos, los cuales hacen parte de otros análisis o pasos previos de análisis de Agricultura Específica por Sitio (AEPS).

Generar indicadores climáticos por cada uno de los ciclos de acuerdo a la asignación.

Después de obtener las listas procedemos a calcular indicadores climáticos que sirvan como variables en análisis posteriores, para ello utilizamos la función `weather_indicators`, no obstante para agregar más indicadores y ver con mayor profundidad como trabaja esta función se sugiere explorar el repositorio de indicadores climáticos.

```
# Generating weather indicators
```

```
Final_indicators <- weather_indicators(Final_weather_assign)
head(Final_indicators)
```

```
##      crop_cycle  TX_Avg  TM_Avg  T_Avg  DR_Avg      SR  TX_34_Fq
## 0              0 32.89204 20.40739 26.64971 12.484651 49559.56 0.4344262
## 1              1 32.93010 21.14670 27.03840 11.783407 48158.53 0.1885246
## 2              2 33.43034 21.88626 27.65830 11.544079 45298.89 0.3852459
## 3              3 32.89204 20.40739 26.64971 12.484651 49559.56 0.4344262
## 11             11 29.14326 19.85129 24.49727  9.291972 45682.51 0.0000000
## 12             12 29.08355 18.69974 23.89164 10.383809 43276.45 0.0000000
##      P_Ac  P_10_Fq  TM_15_Fq  RH_Avg  RH_sd
## 0    725.0 0.1639344 0.04098361 77.19346 9.358854
## 1   2860.8 0.3196721 0.01639344 85.00486 7.617765
## 2   1221.4 0.2786885 0.00000000 84.16486 5.952203
## 3    725.0 0.1639344 0.04098361 77.19346 9.358854
## 11  3348.8 0.5163934 0.00000000 86.94393 3.908390
## 12  2549.8 0.5327869 0.00000000 87.62785 4.525919
```

Almacenar resultados

Finalmente almacenamos los resultados del procesamiento, todos los archivos son relevantes para consultas posteriores o seguir procesos de análisis AEPS.

```
# Files to save
```

```
save( Final_weather_assign , file=here::here("RESULTS","Final_weather_assign.RData"))
```

```
write.csv(crop_cycle_vars_station_assignment,here::here("RESULTS","crop_cycle_vars_station_assignment.csv"))
```

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