**MANUAL DE USUARIO PARA REALIZAR PREDICCIONES CLIMÁTICAS DESDE R**

**run\_CFSV2**= Este script toma como variable predictora la Temperatura Superficial del Mar (TSM) del modelo CFSv2 en el formato CPT disponible en la librería de datos del IRI, como variable respuesta se toman las estaciones o datos de CHIRPS de la zona de interés **sin datos faltantes**, el script genera una sola predicción climática estacional correspondiente al último periodo de datos disponible en la TSM.

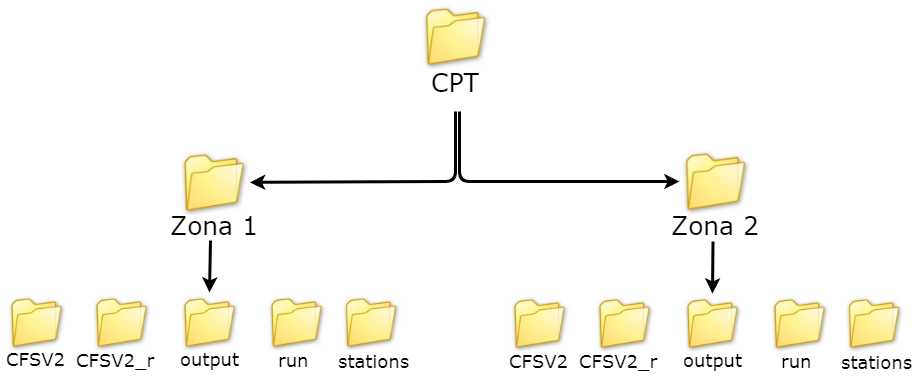


Figura 1. Esquema ejemplo de la Jerarquía de carpetas para correr el script.

La Figura 1 muestra la jerarquía de carpetas para realizar corridas masivas de CPT en versión batch para Windows desde el software estadístico R, en el esquema ejemplo solo se consideraron dos zonas, pero el script funciona con la cantidad de zonas que requiera el usuario, hay que tener en cuenta que la rapidez del proceso depende del número de zonas que se consideren, dentro de cada zona se encuentran las siguientes carpetas que deben contener los siguientes archivos:

**CFSV2:** Esta carpeta debe ser creada y además de esto se debe conservar exactamente el mismo nombre que muestra la Figura 1, dentro de esta carpeta se guardan los archivos de la TSM del modelo CFSv2 en formato CPT, los archivos deben de ser guardados conservando la siguiente estructura en sus nombres, por ejemplo se tienen los siguientes nombres de archivos **Apr\_Aug-Sep.tsv** y **Apr\_Aug-Sep-Oct.tsv**, las primeras tres letras en inglés corresponden al mes de las condiciones iniciales del modelo, las siguientes letras en ingles corresponden a los meses del periodo de interés separado por guion medio, en el primer nombre ejemplo el periodo de interés corresponde a un bimestre, mientras que el segundo nombre ejemplo corresponde a un trimestre, algo importante que se debe de tener en cuenta es que la primera letra de cada mes debe ser mayúscula tanto en las condiciones iniciales como en el periodo de interés, adicional a esto las condiciones iniciales deben estar separadas del periodo de interés por un guion bajo.

**CFSV2\_r:** Esta carpeta se crea automáticamente y dentro de ella se guardan automáticamente los archivos resultantes de la metodología de selección de área predictora, por cada archivo de la TSM se deben generar 10 archivos, **en esta carpeta no se debe ingresar ningún archivo.**

**output:** Esta carpeta se crea automáticamente y dentro de ella se guardan automáticamente las salidas de CPT, el script genera salidas por cada uno de los archivos incluidos en la carpeta **CFSv\_2\_r**, se definieron unas salidas por defecto en la corrida general del Batch, si se quiere considerar algunas salidas adicionales se debe modificar la función **run\_cpt** dentro del script, **en esta carpeta no se debe ingresar ningún archivo.**

**run:** Esta carpeta se crea automáticamente y dentro de ella se guardan automáticamente los archivos en .bat correspondientes a cada corrida realizada en CPT, esto se guarda con el objetivo de tener un control histórico de las corridas que se realizaron, para visualizar el contenido de los archivos se debe dar click derecho y luego dar en click en editar, si se realiza doble click en estos archivos se ejecutara la corrida de nuevo en la consola, **en esta carpeta no se debe ingresar ningún archivo.**

**stations:** Esta carpeta debe ser creada y además de esto se debe conservar exactamente el mismo nombre que muestra la Figura 1, dentro de esta carpeta se deben guardar los archivos de las estaciones de precipitación o los datos satelitales de precipitación , la forma de estos archivos debe ser igual a la que se considera en CPT y el formato debe ser en (.txt).

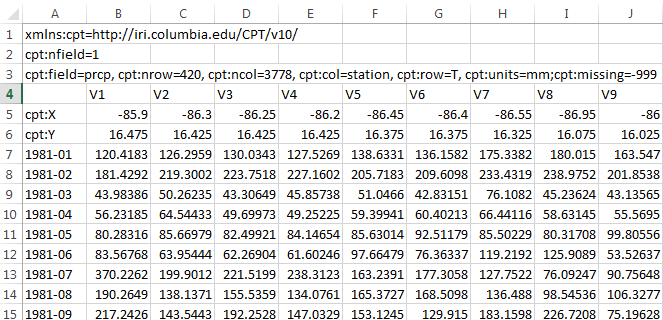


Figura 2. Ejemplo de la forma que debe tener el archivo de las estaciones de precipitación.

Los parámetros más importantes a tener en cuenta en este archivo (Figura 2) son el número de filas (nrow) y el número de columnas (ncol), el conteo de la filas se realiza sin tener en cuenta las coordenadas de las estaciones, y el resto de parámetros del encabezado no se modifican.

Adicional a estas carpetas el script genera dos archivos en formato csv, uno de estos archivos contiene un resumen de las métricas de todas las corridas, considerando la correlación de Pearson, 2AFC y el Goodness Index, el otro archivo contiene un resumen de las probabilidades de todas las corridas, finalmente el script guarda automáticamente la imagen con el área que seleccionó en todas las corridas.

Para hacer uso del script, antes de realizar la corrida se debe cambiar la variable **main\_dir** (línea 312) en el script por la ruta donde se encuentra tu carpeta principal llamada CPT (la ruta no debe contener la carpeta CPT), en R las ruta debe estar separada con slash (/), después de cambiar la ruta verifique que los archivos de entrada para CPT no estén abiertos, luego procede a correr el código en dos secciones la primera sección del código se debe correr desde la primera línea hasta la línea 374(*cat("\n Batch CPT realizado \n")*), aquí se debe esperar hasta que el proceso finalice todas las corridas de CPT en batch, es decir cuando se cierren todas las consolas de la cmd abiertas, luego se debe correr el código de la línea 374(*cat("\n Batch CPT realizado \n")*) hasta el final.

**RECOMENDACIONES PARA REALIZAR LA CORRIDA DE CPT EN BATCH DESDE R**

* Para poder hacer uso del script se debe haber generado la variable de entorno para poder llamar a CPT desde R. (para realizar este paso guiarse con las diapositivas respectivas).
* Recuerde que los datos de estaciones de precipitación o datos satelitales no pueden contener información faltante.
* Verifique que los nombres de las carpetas estén igual a la Figura 1.
* Los datos de la TSM deben estar descomprimidos.
* Verifique que los parámetros nrow y ncol en su archivo de estaciones o datos satelitales.
* Los nombres de las estaciones no deben tener más de 16 caracteres.
* Recuerde realizar la corrida del script en dos secciones, 1ra sección (línea 1 hasta línea 353) y la 2da sección (línea 355 hasta línea 377).
* Si llega a fallar el código en algún paso, lo mejor es que inicie de nuevo borrando todos los archivos que generó el script y el ambiente de variables de R.
* Verifique que los datos de las estaciones no contengan datos NA en las ultimas columnas de cada archivo, y si los tienen proceda a abrir el archivo desde Excel y seleccione todas las columnas en blanco y elimínelas, luego guarde el archivo de nuevo.

**DESCRIPCIÓN DE LAS FUNCIONES UTILIZADAS EN LOS SCRIPTS**

Dentro del script que se generó se encuentran unas funciones específicas para cada proceso, en la Tabla 1 se encuentra la descripción de cada una de estas funciones.

|  |  |  |  |
| --- | --- | --- | --- |
| Nombre de la  función | Descripción de la función  Tabla 1. Descripción de las funciones utilizadas en el script. | Entradas de la función | Salidas de la  función |
| transform\_raster | Esta función transforma un conjunto de datos en formato raster. | **X=** Datos que sé que quieren pasar a formato raster.  **Y=** Extensión de los datos. | Los datos en formato raster. |
| data\_raster | Esta función realiza la transformación de los datos en formato raster para todos los años de historia en los datos. | **dates**= Datos de la TSM en formato CPT | Un stack con la información en formato raster para todos los años. |
| quarterly\_data | Esta función genera el acumulado de la precipitación para el periodo de interés. (Acumula bimestres y trimestres) | **data=**Dataframe del conjunto de datos de precipitación que se quieren acumular.  **names=**Nombre del archivo de la TSM con el cual se va a realizar la corrida. | Un dataframe de los datos de precipitación acumulados. |
| pca\_x\_svd | Esta función realiza el análisis de componentes principales para un conjunto de datos utilizando la descomposición en valores singulares. | **X=**Matriz a la cual se le va a realizar el ACP.  **modos=** Número de componentes a obtener a partir del conjunto de datos.  **pon=** ponderación para cada uno de los pixeles, la ponderación se realiza con la latitud de cada pixel. | Un objeto de tipo lista que contiene las componentes principales y los vectores propios del conjunto de datos |
| pca\_x\_svd | Esta función realiza el análisis de componentes principales para un conjunto de datos utilizando la descomposición en valores singulares. | **X=**Matriz a la cual se le va a realizar el ACP.  **modos=** Número de componentes a obtener a partir del conjunto de datos. | Un objeto de tipo lista que contiene las componentes principales y los vectores propios del conjunto de datos |
| selection\_area | Esta función calcula la correlación promedio que tiene cada uno de los pixeles de la TSM con todos los posibles modos canónicos. | **X=** Matriz de datosde la TSM.  **Y=**Matriz de datos de la precipitación.  **pon=** ponderación para cada uno de los pixeles, la ponderación se realiza con la latitud de cada pixel. | Un vector que contiene la correlación promedio para cada uno de los pixeles de la TSM. |
| files\_x | Esta función construye los archivos de la TSM haciendo la selección de área predictora. | **raster=** Raster de la TSM.  **cor=** Correlación promedio de cada pixel.  **Na=** Nombre que va tener el archivo de salida.  **years=** Años disponibles en la TSM. | 10 archivo en formato CPT para TSM. |
| run\_cpt | Esta función realiza la corrida de CPT en batch. | **x=**  Ruta donde se encuentra el archivo de la TSM.  **y=**Ruta donde se encuentra el archivo de la estaciones.  **run=** Ruta donde se va a guardar la corrida histórica de CPT en formato .bat.  **output=** ruta donde se van a guardar las salidas de CPT. | Archivos de salida de CPT. |
| best\_GI | Esta función selecciona el mejor goodness index de las 10 corridas. | **x=** nombre del archivo de la TSM. | Un objeto con el nombre del archivo que obtuvo el mejor goodness index |
| metricas | Esta función genera un resumen de las métricas de cada corrida con CPT. | **x=** nombre del archivo de la TSM. | Un archivo resumen en formato csv de las métricas de las corridas de cada zona. |
| proba | Esta función genera un resumen de las probabilidades de cada corrida con CPT. | **x=** nombre del archivo de la TSM. | Un archivo resumen en formato csv de las probabilidades de las corridas de cada zona. |
| save\_areas | Esta función genera la gráfica con los pixeles seleccionados en cada corrida de CPT. | **ras=** datos de la TSM en formato raster.  **cor=** Correlación promedio de cada pixel.  **all\_name=**nombre del archivo de la TSM con el que se obtuvo la mejor corrida. | Un gráfico en formato .tiff de los pixeles que se seleccionaron en la mejor corrida |