

Actividad 6

Análisis de datos con Pandas y Matplotlib.

Aaron Sebastian Castillo Espinoza

Departamento de Física de la Univerdsidad de Sonora

Grupo 3, Semestre 2019-1 — Física Computacional



“El saber de mis hijos
hará mi grandeza”

Introducción

Durante esta actividad hicimos un analisis de datos para determinar las horas de frío que hubo en el pasado invierno en la ciudad de Hermosillo, Sonora. Para esta actividad, se cuenta con los datos medidos por la estación ubicada en un cultivo de Vid, en el Km. 41 de la carretera Hermosillo a Bahía Kino.

El requerimiento de frío en especies frutales caducifolias, requisito también conocido como acumulación de frío, es un factor decisivo en la adaptación de estas especies a su ambiente. De este requerimiento depende la ruptura de la dormición de un amplio espectro de árboles y arbustos frutales de uso comercial, tales como las especies frutales de pepita (manzano, peral, membrillero), las de hueso o carozo (duraznero o melocotonero, ciruelo japonés, cerezo dulce, guindo, olivo, etc), las especies productoras de frutos secos (almendro, avellano, nogal, castaño, pecán, pistachero), los arbustos de hoja caduca (arándanos, frambueso, moras, zarzamora, grosellero), y las especies de hoja caduca trepadoras (vid, actinidia). Todas ellas tienen que estar expuestas a un período de bajas temperaturas durante el letargo invernal para una adecuada ruptura de la dormición e inicio de la nueva estación de crecimiento.

Esta actividad esta dividida en 2 partes, la primera fue utilizando el conocido modelo De Utah desarrollado por Arlo Richardson y colegas en 1974 e implementado por la Universidad de California. Esta parte también fue calificada como parte la primera evaluación del curso. En la segunda parte se utilizo otro método desarrollado por el Instituto de Investigaciones Forestales, Agrícolas y Pecuarias (INIFAP) en este casi implementado en el Campo Experimental Costa de Hermosillo (INIFAP-CECH) por Jose Grajeda Grajeda y colaboradores.

El propósito final de esta actividad es realizar una comparación entre ambos métodos

Modelo De Utah

Para empezar con este modelo lo que hice fue leer la lista de datos proporcionada por el profesor , de las cuales solo son necesarias 2 columnas de las 32 que tiene, la tituladas `Timestamp` y `AirTC_Avg`, que tienen registrados la hora cada 10 minutos y la temperatura correspondiente a esta hora y guardarla en un `DataFrame`. Para que estos métodos funcionen es necesario que las temperaturas estén acumulados por día. Para esto fue necesario primero separar los datos de la fecha por año, mes, día y hora. Posteriormente se filtran los datos importantes para el análisis, los cuales empiezan a partir del 1 de Noviembre.

Ahora es necesario obtener la temperatura promedio por hora, así como la temperatura máxima y mínima correspondiente a cada mes. Para obtener estos datos utilice la función `GroupBy` de pandas, la cual suma todos los datos numéricos que corresponden, en este caso a un mismo año, mes, día y hora, una vez que tiene la suma de estos datos le pido que me saque el promedio de los mismos que corresponderán a la temperatura promedio por hora. Una vez teniendo estos datos repito básicamente el mismo proceso, solo que ahora al la función `GroupBy` le pido que agrupe los valores de la temperatura promedio obtenida anteriormente por año, mes y día para así checar cual de estas es la máxima y la mínima y asignar esta a cada día. Cabe resaltar que estas no son necesarias para el análisis mediante el modelo de Utah, el profesor nos pidió que las obtuviéramos como una actividad extra.

Aquí vemos la gráfica del comportamiento de las temperaturas máximas y mínimas:

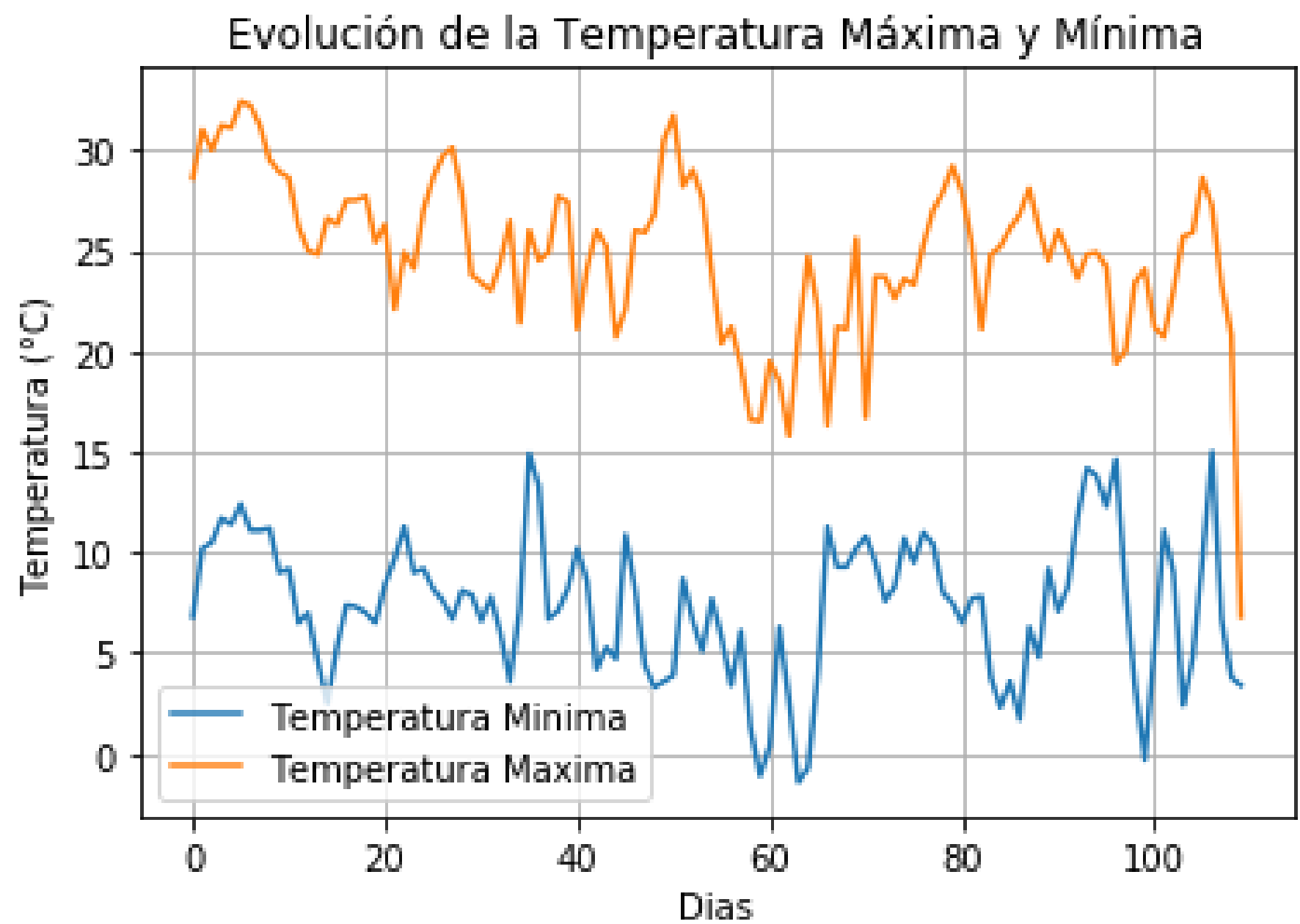


Figura 1: Evolución de las temperaturas máximas y mínimas por día

Una vez hecho creé un loop con un arreglo vacío llamado UF (Unidades Frías) que, checando todo el `DataFrame` y basándome en la tabla mostrada, dependiendo en la temperatura promedio el el valor que se le asigna a cada hora.

Una vez teniendo estos datos por hora vuelvo a hacer uso de la función `GroupBy` para sumar todos los valores de UF's de cada día en una nueva columna del `DataFrame` llamada `UF_24` donde se muestran las Unidades frías por día necesarias para la utilización de este modelo. Por ultimo creo un nuevo `DataFrame` en el cual guardo todos los datos obtenidos anteriormente deshaciéndome de los valores repetidos de `UF_24` por hora quedándome solo con los valores por día.

Por ultimo utilizo la funcion `CumSum` de Numpy para realizar una suma acumulada de los datos de `UF_24` para posteriormente graficarlo y ver su comportamiento

| Tabla 1: Relación de eficacia para la salida de la dormición, según el «modelo de Utah». | |
|--|---|
| Temperatura (°C) | UF correspondientes a 1 hora transcurrida a un dado rango térmico |
| < 1,4 | 0 |
| 1,5 a 2,4 | 0,5 |
| 2,5 a 9,1 | 1 |
| 9,2 a 12,4 | 0,5 |
| 12,5 a 15,9 | 0 |
| 16,0 a 18,0 | -0,5 |
| > 18 | -1 |

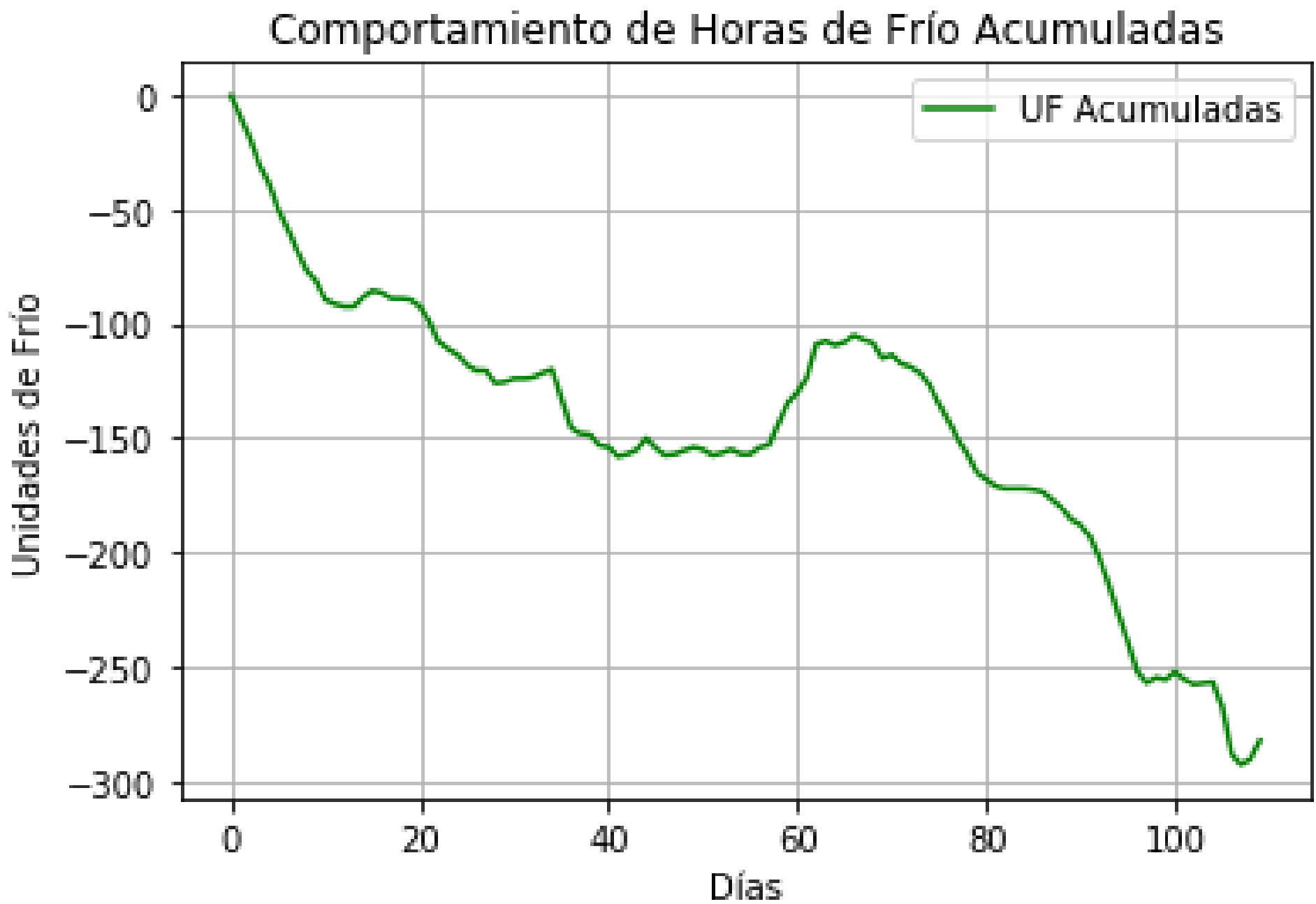


Figura 2: Evolución de las unidades de frio acumuladas

Modelo INIFAP-CECH

Para la realización de este modelo, el procedimiento es básicamente el mismo que el anterior, lo que cambia son los criterios a considerar al momento de calcular las horas frías.

Partiendo del mismo `DataFrame` creado para el metodo anterior donde tengo los datos de la temperatura promedio por hora, lo que hago es una copia de este para poder trabajar con el y no modificar el del metodo anterior. Con este nuevo `DataFrame` creo un loop con dos arreglos vacíos llamados `HC` (Horas Calientes) y `HF` (Horas Frías). En el loop checo las temperaturas promedio que se encuentran entre 0°C y 10°C, aquellas que se encuentren en ese rango le asigno un 1 al arreglo `HF` y un 0 a `HC`, en caso de que la temperatura se mayor a 25°C se le asigna un 0 a `HF` y un 1 a `HC`, por ultimo, en caso de que la temperatura promedio por hora no se encuentre en ninguno de los rangos dados, se le asigna un 0 a ambos arreglos. Una vez que se termina ese loop, agrego los arreglos al `DataFrame`.

Una vez hecho esto, con la función `GroupBy` para sumar los valores de `HC` y `HF` por día y los llamo `HC24` y `HF24` respectivamente. Ya que tengo estos datos, lo que sigue es hacer una resta entre las `HF24` menos las `HC24` como lo indica el método y a esto se le conoce como Horas de Frío Efectivas (HFE).

Por ultimo creé un nuevo `DataFrame` en el cual tengo los valores de la suma acumulada de `UF_24` y las `HFE` para poder hacer una comparación de ambas. Para esto solo copie una creado anteriormente para el método de Utah con la suma acumulada de UF's por día y le agregue las horas de frío efectivas obtenidas en este nuevo método.

La primer gráfica para comparar es la siguiente:

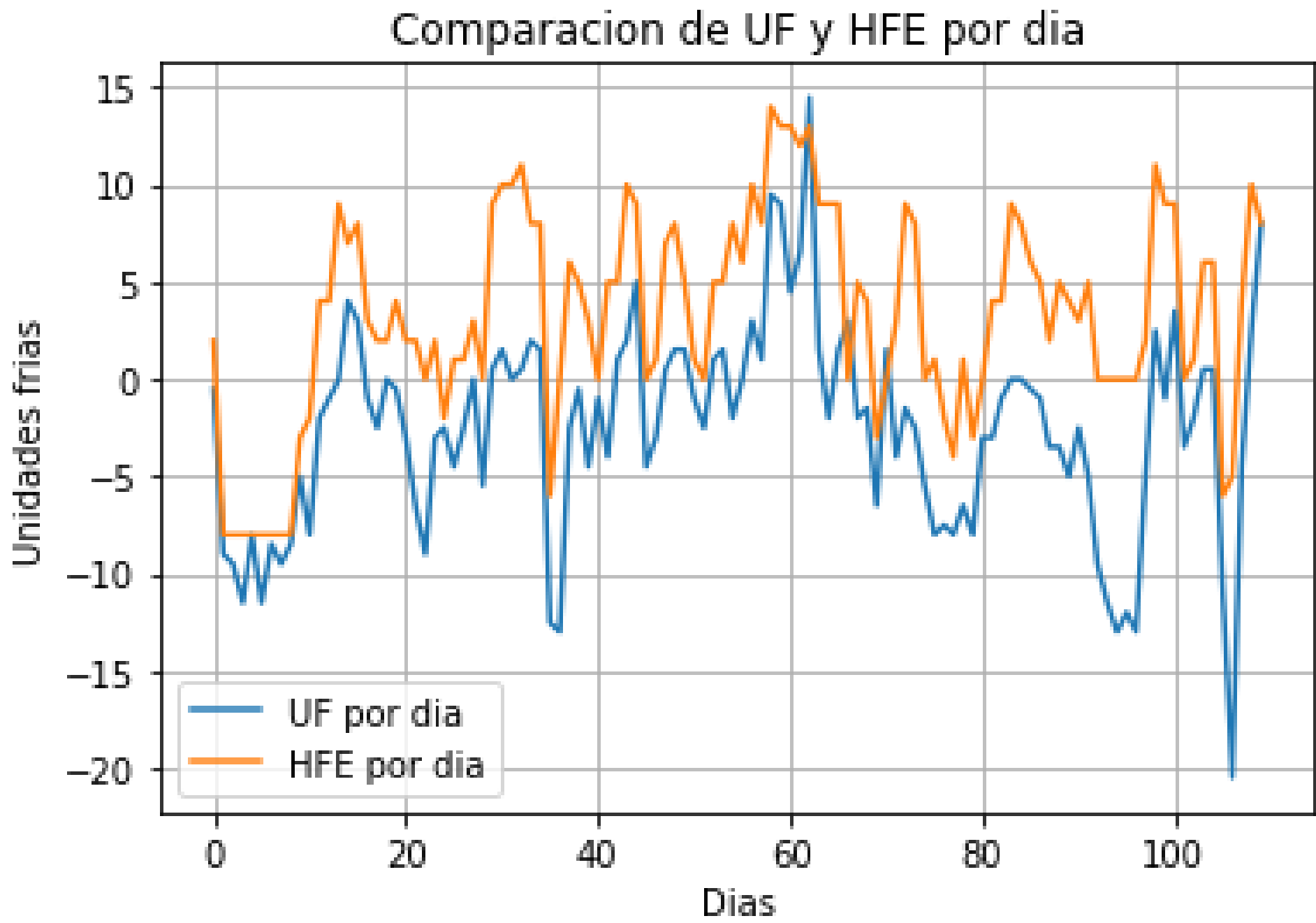


Figura 3: Unidades Frías y Horas de Frío Efectivas por Día

En esta vemos los datos obtenidos por ida, notamos que a pesar de las diferencias, son gráficos muy similares, lo que nos hace llegar a la conclusión de que ambos métodos son igual de buenos para analizar las horas de frío por día.

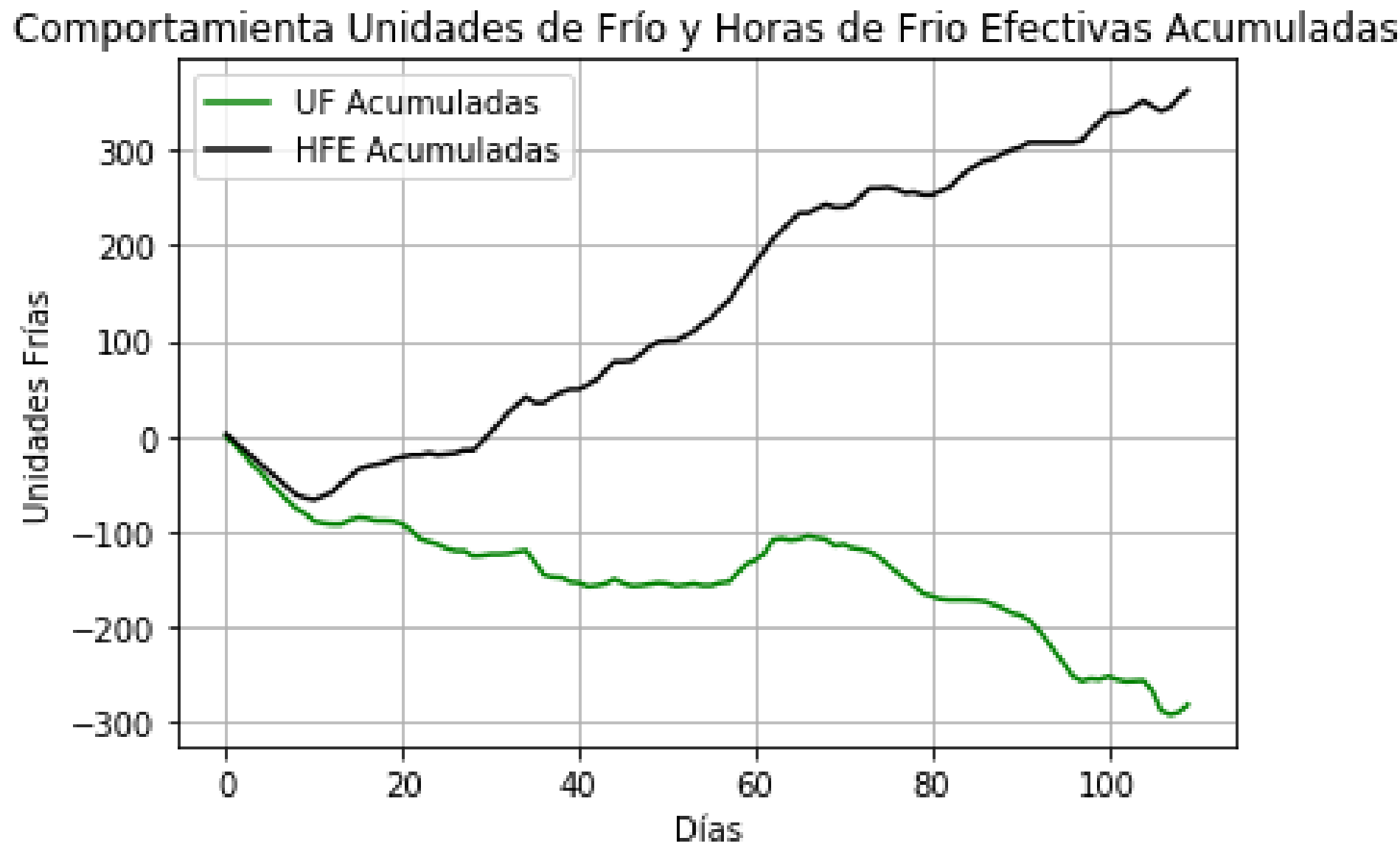


Figura 4: Horas de Frío Efectivas y Horas de Frío acumuladas

En esta ultima gráfica se nota como son las Unidades de Frío y las Horas de Frío Efectivas acumuladas, en esta se puede ver que tienen el mismo comportamiento, en esencia, solo que uno positivo y otro negativo, eso debido a la manera en que se realiza el análisis.

Conclusiones

Como vemos para hacer análisis de datos así como para la realización de códigos en python, no existe una sola manera de hacerlo y eso quedo en evidencia durante la evaluación y en la actividad 6 pues no todos hicimos el mismo código ni utilizamos las mismas funciones, sin embargo todos llegamos al mismo resultado, es decir obtuvimos las mismas gráficas. (4)

Referencias

1. Requerimiento de frío en especies frutales caducifolias”
https://es.wikipedia.org/wiki/Requerimiento_de_frio_en_especies_frutales_caducifolias#Modelo_de_Utah
2. Grajeda, J., Osorio, G. & Sabori R. Üso de Estaciones Meteorologicas Automatizadas en la Agricultura”, 2002.