Requerimientos de clima 08/01/2016

Para trabajar con la información climática de una zona, esta deberá cumplir con algunos requisitos, con el fin de estandarizar la información que ingresa al proceso de control de calidad y llenado de datos faltantes.

Todo se almacenará en archivos de texto (\*.txt). Una sola estación puede registrar varias variables al tiempo, así que se tendrá un archivo por variable de cada estación y se identificará de la siguiente manera:

ID\_VAR.txt 🡪 donde ID es el identificador o nombre de la estación sin espacios, y VAR la variable que contendrá el archivo. Las variables contempladas en el análisis son RAIN, RHUM, TMAX, TMIN, ESOL\* y SBRI, que son precipitación (mm), humedad relativa (%), temperatura máxima (⁰C), temperatura mínima (⁰C), energía solar\* y brillo solar (horas).

*\*Para esta variable el formato del nombre del archivo es un poco diferente: ID\_UNIT\_VAR.txt, donde UNIT representa la unidad de medida en que se registra ESOL, que podrá ser WAM2, CCM2 o MJM2, indicando watts por metro cuadrado, calorías por centímetro cuadrado o mega julios por metro cuadrado.*

|  |
| --- |
| *Archivos de texto con el formato necesario para procesarlo* |
|  |

Es importante destacar que es posible que la información venga tanto diaria como horaria, por lo que será necesario entonces separar los archivos según su tipo de registros (una carpeta para los archivos diarios y otra para los horarios). Si existe el caso que una misma estación contenga ambos formatos de registro (tanto horario como diario), es preferible utilizar la información horaria si ambos archivos contienen las mismas fechas, de lo contrario, se podrán usar los dos, ya que el código contempla una fusión en estos casos para aprovechar ambas fuentes.

Al interior de los archivos la forma en que deben estar organizados los datos es la siguiente. Se recomienda respetar al menos el orden que aquí se expone.

La primera línea contendrá los encabezados o rótulos: Date Hour Value

|  |  |
| --- | --- |
|  | La fecha debe tener el formato yyyymmdd, representando el año con sus cuatro cifras, el mes y el día con dos cifras. La hora HH:MM con AM o PM. Para el valor la única petición es que los decimales estén representados por puntos y si no existe registro para una fecha y hora determinada usar NA para dato ausente. |

Esta organización se puede hacer fácilmente desde Excel. En el caso de la fecha se elige formato de celda, personalizar (*custom*) y se escribe “yyyymmdd”. En el caso de los valores, si se tienen representados los datos ausentes de forma especial (por ejemplo -9999 es común) se selecciona la columna entera y se utiliza la función de reemplazar dicha forma a la convenida (NA)

Si la información es diaria sólo se tendrá la columna de fecha (Date) y valor (Value) siguiendo los mismos lineamientos nombrados anteriormente

De todas las estaciones que se tenga información de SBRI, será necesario crear un archivo Est\_Lat.csv (con este nombre puntual), que se puede crear desde Excel, y contendrá:

|  |  |  |
| --- | --- | --- |
| 🡪 |  | Donde ID deberá ser el mismo manejado anteriormente y Lat, la latitud de la estacion en cordenada decimal. |

Por último, son necesarios dos archivos mas en formato \*.csv que contendrán los valores posibles máximos y mínimos por cada variable en la zona de estudio. Se requiere un archivo para el control horario y otro para el control diario, pues los valores posibles son diferentes en estos casos. Se aconseja que para el archivo horario se manejen valores extremos posibles según el historial del pais, y que en el archivo diario si sean mas acotados a la realidad de la zona. Por ejemplo en temperaturas para Colombia se podria dejar en Val\_REF\_QCHour de 0 a 50 grados y en humedad relativa los valores máximos posibles que son de 0 a 100 %, aunque se aclara que estos filtros son muy amplios pero la idea es hacer un filtro de valores que evidentemente estan muy alejados de la realidad.

|  |  |
| --- | --- |
|  |  |

***\*\*\*Es importante que se conserven estos formatos con sus respectivos encabezados, conservando incluso las letras mayúsculas o minúsculas que sean usadas.\*\*\****

**IMPORTANTE**

Es necesario contar con la mayor cantidad de información climática posible, pues cuando hay información faltante, la metodología se “apoyará” en la información disponible de otras estaciones para realizar la imputación de los datos ausentes. Por otro lado, también se requiere que ésta sea lo más cercana posible al punto de interés, pues variables como la precipitación y brillo solar son muy sensibles (cambian mucho) en cortas distancias. Un mínimo requerido de información sería contar con 3 estaciones con todas sus variables y que cada una de estas tenga a lo sumo, 30% o 40% de NA’s para el periodo a trabajar (tener en cuenta que entre más datos falten, mayor será la información simulada que al final se tendrá y esto no es muy conveniente). Si no se cuenta con información horaria, basta con tener la información diaria, pero no mensual.