# CalculadorHD

## Datos básicos

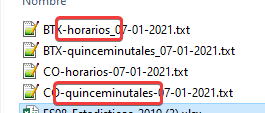
### Descripción

Este programa es el encargado de generar documentación en la cual se aplican cálculos a unos datos de entrada. Esos datos de entrada serán archivos csv. Serán válidos tanto los archivos de texto (txt) como los archivos csv (csv). Cualquier implementación/modificación del proyecto puede dar lugar a que esta documentación no se ajuste al 100%.

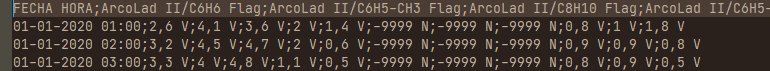
#### formato de entrada

Los archivos de entrada deberán cumplir por obligación las siguientes especificaciones:

* Los archivos deben estar nombrados con las palabras "quinceminutales" u "horarios" para que el programa pueda distinguirlos.
* Además, esas palabras deberán ir delimitadas por guiones medios.



* Los datos que hay dentro del archivo deben tener la primera columna que sea la "fecha hora", los datos separados por ";" y los decimales separados por ",". Recordar que para los octohorarios necesitamos 1 día extra de datos para realizar los cálculos.



### Datos técnicos

tecnologías utilizadas: Node + Electron

Documentación: En la carpeta, existen imágenes png. Contienen datos ocultos de los esquemas. Si se abre con Draw.io se puede recuperar ese esquema y modificarlo.

Icono: Debe tener como mínimo 256x256 de tamaño para que Electron lo acepte. 512x512 es aceptable.

## Configuraciones adicionales

* Se generará un dato inválido (-9999 "N") si los datos de entrada no cumplen un porcentaje mínimo del 75% de datos válidos.
* Se fusionarán los datos del mismo tipo según su contaminante. Es decir, por ejemplo, todos los archivos de entrada de BTX se fusionarán para dar lugar a 1 archivo BTX horario, otro diario y otro octohorario.
* Todos los contaminantes menos el CO y el BTX deberán ser redondeados a 0 decimales.
* CO y BTX deberán tener 1 decimal como máximo

## Algoritmos

### Horarios

Este documento se genera a partir de los quinceminutales. Los archivos que ya estén en "horarios" no necesitan pasar por este proceso.

Debe cumplir:

1. Debe tener como mínimo 75% (3 de los 4 valores) de los quinceminutales para calcular el dato horario. De lo contrario, el dato horario será inválido ( -9999 "N")
2. De esos datos, verificar que sean datos validos por lo menos el 75% (3 de ellos). De lo contrario, el dato horario será inválido ( -9999 "N")
3. Una vez el punto 1 y 2 comprobado, hacer la media aritmética de los valores (recordar que pueden ser 4 valores o 3 ).

### Diarios

Este documento se genera a partir de los documentos horarios. Es decir que, si lo calculamos partiendo de un documento que contenga quinceminutales, primero deberemos transformar esos datos a horarios.

El funcionamiento es igual que el de los horarios. Se deben tener un mínimo del 75% de los datos horarios validos para poder generar un dato diario válido. Es decir que de 24 horas, 18 como mínimo deberán ser validos.

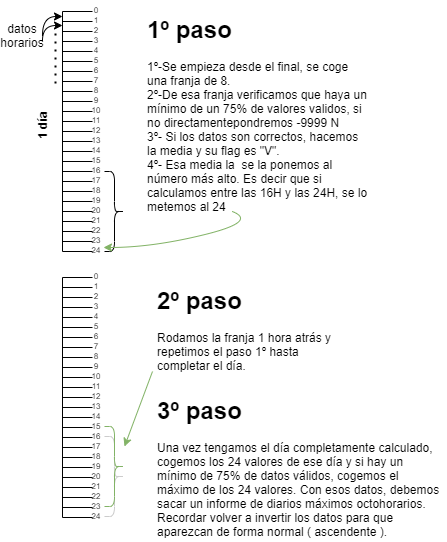
### Octohorarios máximos diarios.

Este documento se genera a partir de los datos horarios.

El octohorario solo se aplica a algunos contaminantes. A continuación, se puede ver la lista. Dicha lista puede cambiar con el tiempo. Lista: O3, CO

Debe cumplir:

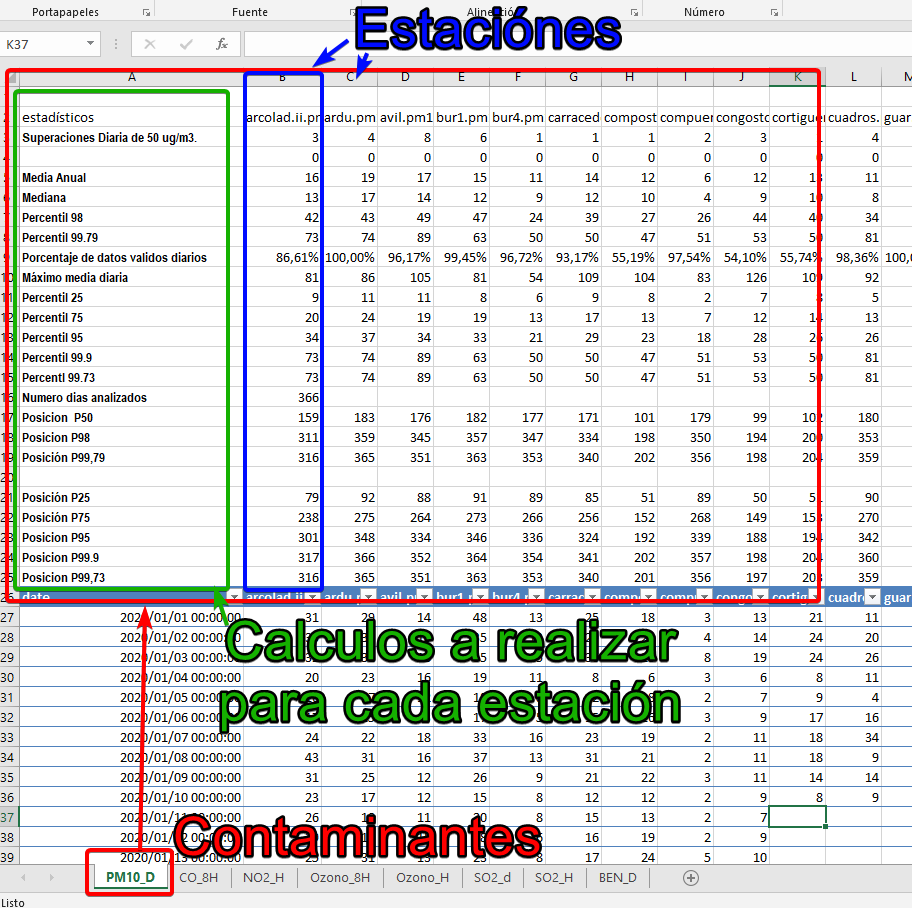
* 75% de los datos deben ser correctos para constar como válidos
* se debe tener 1 día extra anterior a los datos a calcular debido a que el octohorario se desplaza atrás 8 horas en cada cálculo.



## Estadísticos

### Datos importantes

* en el archivo que Luis me dio de estadísticos, solo falta el PM2.5\_D. Ese es igual en todo al PM10\_D salvo que las superaciones diarias son de 25 ug/m3
* Los archivos de entrada del "PM2.5" se llamarán "PM25". Pero se debe tratar como si fueran "2.5" y la salida deberá tener "2.5".
* Tanto los estadísticos como los archivos básicos de salida deberán ser el producto de la fusión de datos clasificados por contaminantes. Es decir que si tenemos un archivo "BTX-horarios" y otro "BTX-quinceminutales", se tendrá que generar un archivo BTX con los datos de ambos fusionados.
* Para saber si un valor es válido, tener en cuenta la Flag con prioridad al valor.
* Los cálculos se realizarán para datos horarios, 8horarios y diarios. Los 15minutales no se usan para esta fase.

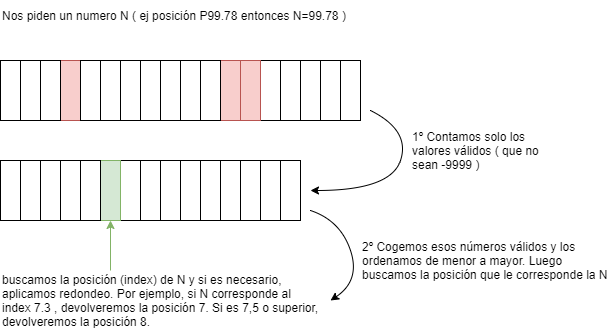


### Cálculos

#### Posición

Dado un número "N" buscaremos la posición al que le corresponde en el array después de haber ignorado los valores incorrectos y haber reordenado los valores de menor a mayor.

Es decir que sería: ( N \* array.length ) / 100



#### Percentil

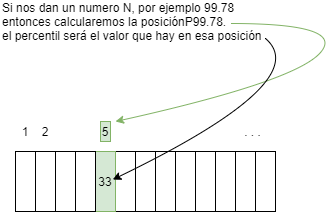
Se utiliza la posición para calcular el percentil. Es el valor que se encuentra en N. Tener en cuenta que los arrays en javascript empiezan a contar desde 0.

ejemplo de como se calcula la posición y el percentil en javascript:

let posicion = Math.round((perc \* dat1.length) / 100);

let percent = `${dat1[posicion-1]}`.replace('.', ',');

// se pone -1 porque en javascript se cuentan a partir de 0



#### Media anual

* En el cálculo de la media anual, no se tendrá en cuenta el 75% de valores válidos.
* La media anual nunca será -9999 N.
* Siempre se calcula usando valores válidos. Los inválidos no contabilizan ni para la suma, ni para la división.
* No usaremos el número de días que tiene el año. Sumaremos los datos validos y los dividiremos por la cantidad de valores usados en la suma.
* Si una estación tiene 0 datos válidos, eso generará una división inválida ( 0 /0 ) . En ese caso, pondremos "SD" que significa "sin datos" en vez del valor.
* los decimales seguirán las mismas reglas que los archivos horarios, diarios y octohorarios. Es decir, que habrá que ver el archivo de configuración.

#### Mediana

La mediana es lo mismo que decir percentil 50. Por lo tanto, deberemos calcular la posición 50.

#### Superación diaria

Las medidas son siempre en ug/m3 . Que supere la medida sin incluirla. Es decir que, si buscamos valores que superen 50um/m3, solo contaremos los valores incluidos entre el 51 en adelante. El 50 no se contabiliza. Solo datos válidos.

#### Porcentaje de datos validos

* Lo primero es verificar si el año que estamos calculando es bisiesto o no y por lo tanto cuantos días contiene. No podemos fiarnos de las filas que contiene el documento ya que puede ser incompleto. Una vez tengamos el número de días (si estamos haciendo los cálculos diarios) o de horas (si estamos haciendo los cálculos horarios), ese será el 100%. Ahora contamos el numero de días válidos (u horas) y sacamos el porcentaje.
* Tendrá 2 decimales. No sigue la misma regla de decimales que en la generación de ficheros horarios, diarios y octohorarios.

formula: (valores validos de la columna \* 100) / días u horas del año natural completo.

#### Máximo media diaria / horaria

Cuando trabajamos con datos diarios ( D) u horarios (H), ya tenemos los valores medios de cada día / hora. Simplemente debemos ver cuál es el máximo de esos días. Este cálculo no se realiza para otros

#### Número de horas/días analizados

* Este es el número de días o de horas que tiene el año que estamos calculando. Se utilizan las fechas (es decir, que debemos consultar los días/horas que tiene ese año con una librería tipo "moment.js"). No se deben usar los registros del archivo porque pueden estar mal o incompletos, pero eso se deben calcular manualmente.
* Si estamos en el contaminante PM10\_D, significa que estamos consultando los datos diarios, por lo tanto, se usará el "número de días analizados" y si el contaminante es PM10\_H, significará que estamos calculando datos horarios y por lo tanto usaremos el "número de horas analizadas".
* En el caso del octohorario, también calcularíamos el "número de días del año".