

VALIDACIÓN DE DATOS

Introducción

Una vez recibidos los datos de las estaciones y antes de incorporarlos a las bases de datos, para su almacenamiento y publicación, y buscando minimizar el número de datos erróneos, se procede a realizar un proceso de validación basado en la Norma UNE 500540:2004 “Redes de estaciones meteorológicas automáticas: Directrices para la validación de registros meteorológicos procedentes de redes de estaciones automáticas.”

Rangos de Validación

Nivel 1: Validación de Rango

Es el primer nivel de validación que se realiza donde se comprueba el rango de los valores meteorológicos introducidos en la base de datos. Por rango se entiende el límite superior e inferior entre los que debe de estar el valor de un dato para ser considerado como válido.

Tabla de rangos de los datos horarios.

DATOS HORARIOS	Rango Mínimo	Rango Máximo	Unidades
Temperatura Media	-21	45	°C
Humedad Media	5	100	%
Velocidad viento	0	75	m/s
Dirección Viento	0	360	°
Radiación	0	1350	w/m2
Precipitación	0	150	mm

Tabla de rangos de los datos diarios.

DATOS DIARIOS	Rango Mínimo	Rango Máximo	Unidades
Temperatura Media	-21	45	°C
Temperatura Máxima	-21	45	°C
Hora-Minuto Temperatura Máxima	0	2400	
Temperatura Mínima	-21	45	°C
Hora-Minuto Temperatura Mínima	0	2400	
Humedad Media	5	100	%
Humedad Máxima	5	100	%
Hora-MinutoHumedad Máxima	0	2400	
Humedad Mínima	5	100	%
Hora-MinutoHumedad Mínima	0	2400	
Velocidad Viento	0	75	m/s
Dirección Viento	0	360	°

DATOS DIARIOS	Rango Mínimo	Rango Máximo	Unidades
Velocidad Viento Máxima	0	75	m/s
Hora-Minuto Velocidad Viento Máxima	0	2400	
Dirección Viento para Velocidad Máxima	0	360	°
Radiación	0	35	MJ/m ²
Precipitación	0	150	mm

Cualquier valor que esté fuera de su rango correspondiente se considera que es erróneo y se etiqueta como tal.

Los datos meteorológicos, pese a que sus valores estén dentro de rango, pueden presentar incoherencias con respecto a los valores de otros datos meteorológicos o ser sospechosos por el valor meteorológico que exponen. Por lo tanto los valores que han superado el primer nivel son sometidos a niveles superiores de validación.

Nivel 2: Validación de la Consistencia Interna

En este nivel de validación se tienen en cuenta las relaciones meteorológicas que existen entre las observaciones efectuadas en la misma estación. Valores medidos al mismo tiempo y en el mismo lugar no pueden ser inconsistentes entre ellos. (UNE 500540:2004)

Este nivel de validación sólo se realiza en los datos diarios. Las comprobaciones que se realizan en esta validación son las siguientes:

- El valor de Velocidad del Viento Máxima no puede ser inferior al valor de Velocidad del Viento.
- El valor de Velocidad del Viento no puede ser igual a cero.
- El valor de Dirección del Viento no puede ser igual a cero.
- El valor de Humedad Media no puede ser inferior a un 50% cuando el valor del dato de Precipitación es superior a 0,2.
- El valor de Temperatura Mínima no puede ser igual o superior al de Temperatura Máxima ni al de Temperatura Media. Del mismo modo, el valor del Temperatura Media no puede ser igual o superior al de Temperatura Máxima.
- El valor de Humedad Mínima no puede ser igual o superior al de Humedad Máxima ni al de Humedad Media. Del mismo modo, el valor de Humedad Media no puede ser igual o superior al de Humedad Máxima.
- El valor de Radiación no puede ser igual a cero.
- La diferencia entre el valor del dato de la Hora registrada para la Temperatura Mínima y el valor del dato de la Hora registrada para la Temperatura Máxima no puede ser inferior a una hora.
- La diferencia entre el valor del dato de la Hora registrada de la Humedad Mínima y el valor del dato de la Hora registrada de la Humedad Máxima no puede ser inferior a una hora.
- La suma de los porcentajes de los valores de los datos de Calmas y No Calmas no pueden superar el valor 1, que representa el 100%.

- La suma de los porcentajes de los valores de los datos de los rangos de temperatura siguientes: -40° a -30°, de -30° a -20°, de -20° a -10°, de -10° a 0°, de 0° a 10°, de 10° a 20°, de 20° a 30°, de 30° a 40°, de 40° a 50° y de 50° a 60° no pueden superar el valor 1, que representa el 100%.

Nivel 3: Validación de la Consistencia Temporal

En este nivel se compara el cambio entre observaciones consecutivas. Si la diferencia excede un valor permitido, distinto para cada parámetro, el dato será etiquetado como “sospechoso”. Si alguno de los dos datos usados en la comparación no existe, en este paso no se etiquetará el dato, a no ser que no se vaya a aplicar ninguna otra validación, entonces se etiquetara como “sospechoso”. (UNE 500540:2004)

Este nivel de validación evalúa la coherencia del valor de un dato con respecto a los valores de los datos precedentes y los datos posteriores. Este tipo de validación sólo afecta a los datos Horarios.

Por defecto, son cinco el número de datos anteriores y posteriores que utilizará la aplicación para validar su consistencia temporal. Este parámetro es configurable en la base de datos. Si se encontrase algún nulo entre esos datos, la validación no sería posible.

Para realizar la validación se halla la media de los 5 registros precedentes y de los 5 registros posteriores. Del mismo modo, aparte de la media, también se calcula la desviación típica de estos datos.

Si el valor de un dato horario rebasa cualquiera de las medias calculadas +/- 3 veces (este valor es también configurable) la desviación típica de los datos implicados en esas medias, se considerará un dato sospechoso y, por tanto, inválido en cuanto a su consistencia temporal.

Cada vez que un dato es validado como erróneo en su consistencia temporal, también se comprueba si dicho dato es también inválido en las estaciones correlacionadas. Si es así, es muy probable que la anomalía no sea tal. Por ello, ante esta circunstancia, el dato sería validado correctamente.

Las variables meteorológicas cuyos datos son validados en su consistencia temporal son: la Temperatura Media, la Humedad Media y la Radiación.

Nivel 4: Validación Espacial de los datos meteorológicos

Los valores de un mismo parámetro medidos al mismo tiempo en estaciones cercanas no pueden diferir demasiado unos de otros. (UNE 500540:2004)

La validación espacial coteja los datos de una determinada estación con respecto a las estaciones correlacionadas, aquellas que tienen un comportamiento climático parecido en cuanto a los datos adquiridos y que se encuentran situadas espacialmente a una distancia inferior a 50 Km. de la estación estudiada.

Esta validación sólo se realiza en los registros Diarios.

Para realizar la validación de una variable climática determinada, se toman las estaciones de mayor correlación, las que superan un coeficiente de correlación establecido, hasta un máximo de 5 estaciones, de las cuales se calcula la media y la desviación típica de los datos de la variable estudiada. Con estos dos parámetros estadísticos se establece una regla de validación aplicable al dato estudiado.

Esta regla de validación consiste en establecer unos límites, superados los cuales se considera al dato de la variable estudiada como sospechoso, y por tanto, inválido en cuanto a la validación de la consistencia temporal.

La regla de validación establecida para el contraste de un dato consiste en establecer un entorno cerrado, que debe de cumplirse para que el dato sea validado espacialmente:

$$D_i \in [\bar{x} \pm 3 \cdot \sigma]$$

$D_i \equiv$ Dato a validar

$\bar{x} \equiv$ Media de las estaciones correlacionadas

$\sigma \equiv$ Desviación típica de las estaciones correlacionadas

(el múltiplo de σ puede ser configurable)

Índices que marcan el grado de validación

Código de Validación	Descripción	Dato Horario	Dato Diario
A	Validación de Rango Correcta	X	X
B	Validación de Consistencia Interna Correcta	X	
B	Validación de Consistencia Temporal Correcta		X
C	Validación Espacial y de Rango Correcta		X
D	Validación Espacial y Consistencia Interna Correcta		X
M	Dato Modificado	X	X
R	Dato Fuera de Rango	X	X
S	Sospechoso por Inconsistencia Espacial		X
T	Sospechoso por Inconsistencia Temporal	X	
I	Sospechoso por Inconsistencia Interna		X

Dentro del código de validación I, datos sospechosos por Inconsistencia Interna, existe la siguiente subdivisión:

Código de Validación	Descripción
I1	Relación incorrecta entre medias, máximas y mínimas
I2	Valor de Velocidad, Dirección de viento constante de 0
I3	Humedad máxima <50 % y presencia de precipitación
I4	Temperatura máxima y mínima registrada en un intervalo inferior a 1 hora
I5	Humedad máxima y mínima registrada en un intervalo inferior a 1 hora