

Validación de datos

2025



GOBIERNO
DE ESPAÑA

MINISTERIO
DE AGRICULTURA, PESCA
Y ALIMENTACIÓN



Sistema de
Información Agroclimática
para el Regadio

En la actualidad, la red dispone de tres modelos de sonda para medir estas variables, aunque el principio de medida Una vez recibidos los datos de las estaciones, antes de incorporarlos a la base de datos, para su almacenamiento y publicación, 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.”, buscando minimizar el número de datos erróneos,

Rangos de Validación

Siguiendo la norma, es necesario pasar cinco niveles de validación:

- Nivel 0: Validación de la estructura del registro y del instante de la medida
- Nivel 1: Validación de rango
- Nivel 2: Validación de la consistencia interna
- Nivel 3: Validación de la consistencia temporal
- Nivel 4: Validación espacial

Nivel 0: Validación de la estructura del registro y del instante de la medida

Comprobación de la estructura de los datos recibidos.

- Comprobación que el primer campo del registro contenga un tipo de registro válido (102, 202, 203, 199,500).
- Comprobación que el registro es completo, para ello debe de cumplir con un número mínimo de registros:
 - Mínimo Campos Registro Horario = 13;
 - Mínimo Campos Registro Diario = 23;
 - Mínimo Campos Registro Diario2 = 17;
- En caso de existir más campos, el registro no se descarta (para aquellas estaciones que tengan información de sensores adicional).
- Comprobación que los datos de cada campo sean compatibles con los tipos de datos que le corresponde (byte, short, decimal...).
- Comprobación que el IdProvincia de cada registro se corresponda con el del nombre del archivo.
- Comprobación que el IdEstacion de los registros se corresponda con el del nombre de archivo.
- Comprobación que el campo Año y el campo Día componen una fecha correcta y se guarda en el campo fecha del registro final de incorporación.
- Comprobación que el campo fecha no es anterior a una fecha anterior a 30 días. Los datos se incorporan al sistema, pero se genera un aviso
- Eliminación y notificación por Log de todos los registros con fecha superior a la fecha actual
- Eliminación y notificación por Log de todos los registros con fecha inferior a la fecha de instalación de la estación.
- Hora: debe encontrarse entre 30 o 2400, en caso contrario se descarta el dato horarios

Nivel 1: Validación de Rango

Comprobación del 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.

Rangos de temperatura:

Los cálculos de los rangos de temperatura se han determinado por provincia y por mes. Para el cálculo de estos rangos se ha tenido en cuenta las efemérides históricas registradas por la red SiAR en cada provincia y mes.

Rangos de humedad:

Rango mínimo de humedad: 3 % RH

Rango máximo de humedad: 100 % RH

Rangos de viento:

- Velocidad de viento mínima: 0 m/s
- Velocidad de viento máxima: 75 m/s
- Dirección de viento mínima: 0 °
- Dirección de viento máxima: 360 °

Radiación global:

- Radiación medio horaria: Determinación la hora de orto y ocaso por estación y establecimientos de rangos en función de ellos.
- Radiación diaria: Determinación de la radiación máxima en día despejado en función de su localización y el día del año.

Pluviometría:

- Precipitación mínima: horaria: 0 mm
- Precipitación máxima: horaria: 60 mm
- Precipitación mínima: diaria: 0 mm
- Precipitación máxima: diaria: 360 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: -40º a -30º, de -30º a -20º, de -20º a -10º, de -10º a 0º, de 0º a 10º, de 10º a 20º, de 20º a 30º, de 30º a 40º, de 40º a 50º y de 50º a 60º 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 etiquetará 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 ≡ Dato a validar

\bar{x} ≡ Media de las estaciones correlacionadas

σ ≡ 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