

## **CONTROL DE CALIDAD DE DATOS DE RADIACIÓN SOLAR GLOBAL OBTENIDOS POR ESTACIONES METEOROLÓGICAS AUTOMÁTICAS**

**Fernando Nollas<sup>1</sup>, Pablo F. Orte<sup>2,3</sup>, Julián Lell<sup>2</sup>, Rodrigo Alonso-Suarez<sup>4</sup>, German A. Salazar<sup>3,5</sup>**

<sup>1</sup>Servicio Meteorológico Nacional. SMN. Av. Dorrego 4019, CABA. CP: 1427. Argentina.

<sup>2</sup>Instituto de Investigaciones Científicas y Técnicas para la Defensa. CITEDEF, Argentina.

<sup>3</sup>Consejo Nacional de Investigaciones Científicas y Técnicas. CONICET. Argentina.

<sup>4</sup>Laboratorio de Energía Solar. LES. Uruguay.

<sup>5</sup>INENCO-CONICET, Avenida Bolivia 5150, Salta, A4408FVY, Argentina

Tel. 011-61575757 –e-mail: [fnollas@smn.gob.ar](mailto:fnollas@smn.gob.ar)

**RESUMEN:** En este trabajo se presenta una metodología preliminar para realizar el control de calidad de datos de irradiancia solar global horizontal proveniente de estaciones automáticas operadas por el Servicio Meteorológico Nacional y otras Instituciones en Argentina. La metodología se basa en la implementación de filtros automáticos en base a diferentes criterios, algunos de los cuales han sido ampliamente estudiados y validados en la literatura científica, mientras que otros han sido desarrollados en base a la detección de problemáticas observadas en las series de datos disponibles. La metodología presentada no asegura aplicabilidad universal a nivel científico dado que muchas de las estaciones no presentan un protocolo de mantenimiento estándar ni calibración frecuente. Sin embargo, y debido a la dificultad de conseguir datos de irradiancia solar medida en superficie, los filtros presentados en esta metodología permiten realizar un diagnóstico preliminar de la calidad de la información obtenida por las estaciones, poniendo a disposición del usuario información de importancia sobre la calidad de las observaciones analizadas para la toma de decisiones en cuanto a su utilización.

**Palabras clave:** Filtros, Calidad, Estaciones meteorológicas automáticas, Radiación solar.

### **INTRODUCCIÓN**

Las estaciones meteorológicas automáticas (EMAs) se han difundido considerablemente entre los usuarios de información meteorológica debido a la automatización que proporcionan en la adquisición de datos sin la necesidad de la intervención manual del personal, más que un mantenimiento instrumental básico. Las EMAs cuentan con software asociado que permite adaptar la frecuencia de adquisición de la medición a la necesidad del usuario.

La utilización de instrumentación automática donde no se realiza un mantenimiento y calibración adecuada o se realizan pruebas de diagnóstico rutinarias para evaluar el estado de las mediciones, suele generar grandes volúmenes de datos de baja confiabilidad.

El Servicio Meteorológico Nacional (SMN) cuenta con una red de EMAs distribuidas en el país en las cuales se obtienen mediciones de varios parámetros atmosféricos. Sin embargo, la suma de estaciones convencionales y EMAs pertenecientes al SMN no alcanza para cubrir todo el territorio argentino. Es por ello que, paralelamente, el SMN ha desarrollado convenios con distintas instituciones gubernamentales y privadas con la finalidad de recibir y concentrar datos de las EMAs que estos operan en el país. Esta posibilidad de contar con más información y extender el área de cobertura de las mediciones impacta de manera directa en los resultados de los modelos numéricos de pronóstico del tiempo y en la posibilidad de brindar productos de mayor calidad a los usuarios de los distintos servicios sectoriales del país.

Los parámetros meteorológicos para los cuales, típicamente, vienen preparadas las EMAs son: temperatura, humedad relativa, presión, intensidad y dirección del viento y precipitación. Sin embargo, algunas EMAs también incluyen sensores de irradiancia solar global horizontal (GHI) con distintos principios de

funcionamiento. La importancia de medir este parámetro radica en que la energía solar es el principal motor de procesos naturales, así como de distintas aplicaciones para su aprovechamiento utilizadas en diferentes disciplinas. Asimismo, conseguir datos confiables medidos en superficie es realmente un desafío debido a la dificultad que genera adquirir, mantener y calibrar los sensores que los obtienen. En Argentina existen actualmente muy pocas redes de medición de radiación solar (Orte et al., 2021) debido al alto costo y esfuerzo que ello requiere. Esto pone de manifiesto la importancia de los datos de GHI obtenidos por las distintas EMAs distribuidas en el país y sus posibles usos.

Al realizarse mediciones automáticas, puede ocurrir que no se efectúe el mantenimiento adecuado a los sensores y esto representa un inconveniente principalmente al analizar la información de radiación solar, dado que estos datos son los más proclives a incorporar errores (Moradi, 2009). El error en la medición de GHI en superficie se asocia principalmente con las incertezas y errores en los sensores que la miden (Younes et al., 2005), pero además se identifican errores involucrados con la operación de los mismos (Shi et al., 2008; Younes et al., 2005) como pueden ser una mala configuración del datalogger, nivelación incorrecta del sensor, sombras indeseadas o reflexiones causadas por construcciones cercanas, gotas de agua o nieve sobre la cúpula del sensor, entre otras.

Debido a las falsas conclusiones a las que se podría arribar utilizando series de datos con mediciones erróneas, se vuelve imprescindible contar con una metodología de control de calidad (QC) de los datos a través de la implementación de filtros, los cuales deben estar basados en criterios adecuados a este tipo de datos. Debe tenerse en cuenta que, de acuerdo a la frecuencia de adquisición de los datos y la ubicación en la que se encuentre una EMA, se pueden presentar diferentes situaciones que pueden hacer variar los datos de GHI por efectos transitorios como nubosidad (Gueymard, 2017) u otros efectos meteorológicos (Yabra et al., 2021). Por lo tanto, el desarrollo de una metodología unificada para las distintas bases de datos provenientes de EMA con diferentes configuraciones en cuanto a la frecuencia de adquisición y en un país con la extensión y la heterogeneidad de climas y relieves como Argentina es un desafío en sí mismo.

La mayoría de los filtros establecidos en la literatura requieren para su aplicación las tres componentes de radiación solar. Estas son: irradiancia solar directa normal (DNI, por sus siglas en inglés), irradiancia solar difusa horizontal (DHI, por sus siglas en inglés) y GHI.

Uno de los QC más usados actualmente para datos de radiación solar es el de la Baseline Surface Radiation Network (BSRN) (Long y Dutton, 2002; Long y Shi, 2008). El mismo está basado en 4 criterios: valores físicamente posibles, valores extremadamente raros, comparación entre las componentes de radiación (DNI, GHI y DHI) y detector de tracker-off. Cuando se cuenta con las 3 componentes de radiación solar, un control de consistencia puede ser llevado a cabo (Long y Shi, 2008; Journée y Bertrand, 2011; Pashiardis y Kalogirou, 2016; Perez-Astudillo et al., 2018). Este test tiene en cuenta que las tres componentes de la radiación solar se relacionan según la ecuación 1 y por lo tanto establece un valor de aceptación a la diferencia que existe entre la GHI medida directamente contra la estimada utilizando la DNI y DHI.

$$GHI = DNI \cdot \cos(SZA) + DHI \quad (1)$$

donde SZA corresponde al ángulo solar cenital.

Cuando se cuenta con mediciones simultáneas de GHI y DHI, es posible aplicar criterios basados en relaciones  $K_t$ -K (Rojas et al., 2018; Younes et al., 2005; Maxwell et al., 1993; Ruiz-Arias et al., 2010; Nollas et al., 2022) donde  $K_t$  representa el índice de claridad y K la fracción difusa. Estos parámetros se obtienen según las ecuaciones 2 y 3 donde  $I_0$  representa la irradiancia extraterrestre en un plano horizontal en el tope de la atmósfera corregida por la distancia Tierra-Sol variable.

$$K_t = \frac{GHI}{I_0 \cos(SZA)} \quad (2)$$

$$K = \frac{DHI}{GHI} \quad (3)$$

Algunos autores utilizaron modelos de cielo claro para establecer valores límite (Geuder et al 2015, Geiger et al.2002, Raichijk, 2012; Younes et al., 2005) así como también se han establecido criterios que utilizan la cantidad de horas de sol (heliofanía) para detectar valores sospechosos o erróneos (Moradi, 2009; Journée y Bertrand, 2011).

Entre los trabajos en los que se desarrollaron controles de calidad para analizar específicamente datos registrados con una frecuencia sub-horaria se pueden citar Geuder et al. (2015), Perez-Astudillo et al. (2018) y Journée y Bertrand (2011) así como las metodologías utilizadas por BSRN y propuestas por Long y Shi (2008). Asimismo, debe tenerse cuidado a la hora de trabajar con datos minutales o con menor frecuencia porque a esa resolución temporal se presentan efectos transitorios como aumentos o disminuciones repentinas de la irradiancia solar (Gueymard, 2017). Un punto en común en la mayoría de las metodologías de QC desarrolladas para datos sub-horarios es que requieren conocer dos o más componentes de la radiación solar y son insuficientes para asegurar la calidad de los datos de GHI y DHI en ausencia de la DNI. Asimismo, son muy pocos los trabajos en los que se establecieron filtros para valores GHI capaces de identificar valores erróneos independientemente de algún otro parámetro.

El objetivo de este trabajo es proponer una metodología preliminar y automática de control de calidad de datos de GHI con el objetivo de hacer un análisis inicial de la calidad de la información obtenida por EMAs dentro del país. Cabe aclarar que esta metodología fue desarrollada con la motivación de disponer de una serie de filtros que, al implementarse de manera automática, permitan facilitar el control de calidad de series de datos extensas. Si bien podría reducir la necesidad de realizar análisis visuales en busca de anomalías, siempre es recomendable esta práctica por problemas no contemplados en la presente metodología.

## MEDICIONES

Los sensores de radiación solar que poseen las EMAs analizadas en el presente trabajo son variados. Algunas poseen piranómetros termoeléctricos como el Hukseflux SP02 y Kipp y Zonen CMP6, mientras que otras utilizan sensores con principio de funcionamiento fotovoltaico como los de la marca LICOR.

Por otro lado, considerando que muchas de las EMAs pertenecen a redes de terceros, debe tenerse en cuenta que cada institución tiene la potestad de configurar su *datalogger* de la manera que considera adecuada. En el caso particular de las EMAs analizadas en el presente trabajo, los *dataloggers* fueron configurados por las distintas Instituciones para que adquirieran datos con intervalos temporales de 5,10,15,30 y 60 minutos. El intervalo temporal más frecuente en la base de datos de este trabajo fue de 10-min. Si bien es esperable que los valores nocturnos de GHI sean iguales a 0 o incluso menores por el principio de funcionamiento térmico de los piranómetros (Sanchez et al., 2016), debe tenerse en cuenta que las distintas instituciones no presentan el mismo adquisidor de datos, por lo que es posible que en la configuración del *datalogger* se genere un valor de *offset* de manera que, en horas nocturnas, se presenten valores positivos constantes o levemente variables pero con la particularidad de ser pequeños (por debajo de 2 W/m<sup>2</sup>).

Debido a que el SMN no tiene injerencia sobre las redes de terceros, en algunas circunstancias no se tiene conocimiento sobre el mantenimiento del instrumental o su calibración. Teniendo esto en cuenta, es importante aclarar que los filtros propuestos en el presente trabajo no aseguran una calidad óptima (a nivel científico) de los datos que los pasen, sino que representan un primer análisis de calidad de datos provenientes de sensores de los que se desconoce su mantenimiento. La utilidad de este tipo de datos se deja a criterio de los usuarios para aplicaciones en las que se permita cierto margen de tolerancia en cuanto a las incertidumbres asociadas. Esto implica que los usuarios que hagan uso de esta información deben estar al tanto de las limitaciones que estos datos poseen.

## FILTROS APLICABLES A MEDICIONES DE MÁXIMA FRECUENCIA

En esta sección se presentan cuatro filtros que pueden ser aplicados a las mediciones para la máxima frecuencia de adquisición definida de cada sitio de GHI, entendiendo estos como los disponibles para la máxima frecuencia de adquisición definida de cada sitio que puede ser 5-min, 10-min,15-min, etc. Los datos que cumplen con todos los criterios tienen asociado el *flag* 0. Dado que las EMAs obtienen

únicamente valores de GHI, los criterios utilizados para establecer los filtros aplicados en este trabajo consideran únicamente valores de esta variable.

**Filtro 1:** El objetivo principal de este filtro es establecer los máximos valores esperables en cada sitio de medición. Para ello se proponen diferentes criterios en función de la frecuencia de muestreo, teniendo en cuenta la posibilidad de fenómenos transitorios a medida que la frecuencia de muestreo aumenta. Asimismo, también se hace una diferenciación entre valores sospechosos y valores inválidos, considerando que estos últimos deben directamente ser descartados por el usuario de los datos debido a que su uso seguramente conlleve a resultados poco confiables. Uno de los criterios más utilizados para establecer los valores máximos alcanzables en la medición de GHI está dado por el uso de un modelo que permita obtener estimaciones de irradiancia de cielo claro ( $GHI_{cs}$ ), condición que asume la ausencia de nubes. En este trabajo se optó por utilizar el modelo de cielo claro detallado en Rigollier et al. (2000), el cual fue utilizado para obtener el Atlas de Radiación Solar de Europa (ESRA). La elección de este modelo empírico se basa en su sencillez, dado que depende de un solo parámetro, la turbidez de Linke establecida para una aeromasa de 2 (TL2), relacionada con la transparencia atmosférica. Si bien varios valores estándar han sido propuestos (Younes et al., 2005; Journée y Bertrand, 2011; Geiger et al., 2002; Ruiz-Arias et al., 2010; Moradi, 2009; Rajchick, 2012), en este trabajo se utilizó un valor fijo de 1.5 para el TL2 para todos los sitios analizados, como se propone en Abal et al. (2017). La ventaja de utilizar un valor fijo radica en la posibilidad de poder utilizar este filtro en tiempo real dado que no se necesitan otros parámetros de entrada procedentes de otras fuentes. Una desventaja es la baja representatividad geográfica y temporal de un valor fijo.

Se establece así que, si una medición de GHI presenta un valor superior a  $GHI_{cs}$  para ese mismo momento, el dato se considera “sospechoso” y se marca con el *flag* 1, siendo válido para valores obtenidos con cualquier frecuencia temporal y todas las latitudes.

Por otro lado, para establecer el valor límite a partir del cual una medición debe ser marcada como inválida, la consideración es distinta en función del tiempo de adquisición de los datos. Cuando el tiempo de adquisición es menor o igual a 10 minutos se hace uso de los valores extremadamente raros establecidos en los controles de calidad que aplica la BSRN (Long y Dutton, 2002; Long y Shi, 2008); esto es debido a que se contempla la posibilidad de efectos transitorios como *cloud enhancement events* (CEE). Si una medición de GHI presenta un valor superior a los valores extremadamente raros para ese mismo momento, entonces dicha línea se marca como inválida con el *flag* 2.

Para el caso de tiempos de adquisición mayores a 10 minutos se consideraron inválidos los valores que fueron mayores a la irradiancia solar horizontal extraterrestre promediada para el mismo periodo de adquisición.

**Filtro 2:** El filtro 2, establecido de manera empírica por Journée y Bertrand (2011), impone que los valores mínimos esperables de GHI en condiciones de cielo muy cubierto en función de SZA y a través del uso del índice de claridad. Si bien el criterio fue introducido para valores 10-min y 30-min, se encontró que es capaz de identificar valores sospechosamente bajos para las frecuencias consideradas en el presente trabajo, lo cual puede ser producto de impurezas sobre el domo de los sensores (naturales o antropogénicas), obstáculos en el horizonte de los sensores, malfuncionamiento del sensor, entre otros. Este filtro tiene la consideración de que, en sitios antárticos, por ejemplo, se presentan condiciones meteorológicas muy cambiantes en donde se pueden presentar muchos días con niebla y a su vez niveles de radiación muy bajos, por lo que para latitudes menores a  $-60^\circ$ , este filtro se reduce a que  $GHI > 0$  cuando  $SZA < 90^\circ$ . Los valores que no pasan este filtro se marcan como sospechosos con el *flag* 1.

**Filtro 3:** El filtro 3 tiene por acción determinar valores que cambien drásticamente y fuera de lo esperado entre registros consecutivos. Fue planteado por Journée y Bertrand (2011) para valores 10-minutales y 30-minutales y se basa en diferencias del índice de claridad en vez de valores absolutos de GHI. Este filtro detecta valores tal que la diferencia entre el índice de claridad para un determinado instante y el siguiente es mayor a 0.75 siempre y cuando SZA sea menor a  $88^\circ$ . Los valores que no cumplen la condición de este filtro se consideran sospechosos y son marcados con el *flag* 1.

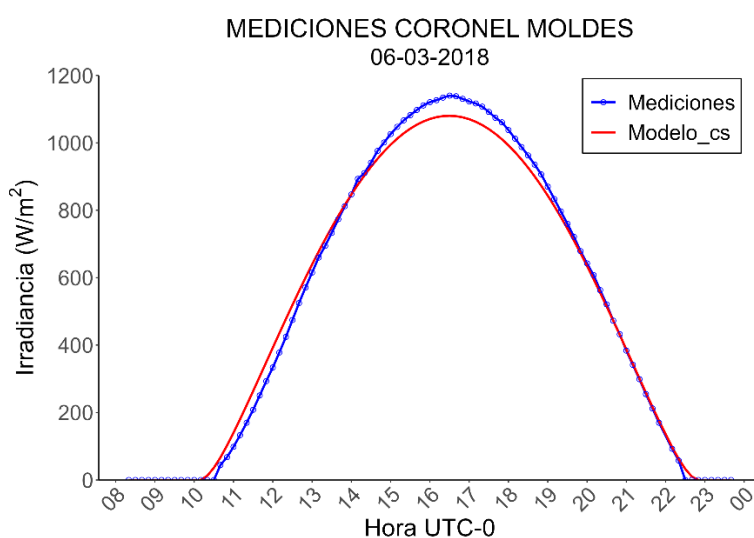
**Filtro 4:** El filtro 4 se enfoca en días en los que se presenta un valor truncado (fijo) a partir de un determinado valor de GHI. Lo más probable es que esto se deba a un error producido durante la configuración del *datalogger*, pero, independientemente del motivo, no deberían considerarse datos provenientes de una base de datos en la cual haya valores truncados. El criterio que se tomó para el desarrollo de este filtro establece que un valor de GHI debe ser, al menos, diferente en un 0.1% respecto al valor anterior cuando  $SZA < 90^\circ$ . A la hora de utilizar este filtro se deben tener en cuenta dos aspectos importantes: en primer lugar, los decimales con los que se adquieren los valores, ya que si se adquiere el dato sin decimales esto podría provocar que más datos sean filtrados. En segundo lugar, debe tenerse en cuenta la sensibilidad del tipo de instrumento con el que se está trabajando, debido a que si el instrumento no es lo suficientemente sensible entonces variaciones en la irradiancia no serán detectadas por el sensor y esto provocará que esos datos sean filtrados. Los valores detectados por este filtro se consideran sospechosos y son marcados con el *flag* 1.

## FILTROS DIARIOS

Se presentan filtros que requieren conocer todos los datos correspondientes a un determinado día, por lo que solo pueden ser aplicados en tiempo diferido (por ejemplo, en forma automática al final del día).

**Filtro 5:** En este caso no es un filtro en sí mismo, sino un criterio de detección de datos posiblemente erróneos basado en la cantidad de datos que fueron detectados por los filtros aplicados a las mediciones de máxima frecuencia. Este filtro surge de observar varios días en los que se detectó una considerable cantidad de valores sospechosos mediante los filtros 1 a 4 (con uno de ellos o varios combinados). El ejemplo más claro en el que se observó esto fue en datos donde las mediciones son más altas que los obtenidos por el modelo de cielo claro en cierto rango temporal, como se puede ver en la Figura 1.

En ella se puede ver que una porción de los datos medidos queda por debajo del modelo de cielo claro y no son marcados por ninguno de los filtros 1 a 4. Sin embargo, más del 56% de los datos diurnos del día en cuestión, para el sitio considerado, son superiores al modelo de cielo claro y por lo tanto marcados con el *flag* 1. Teniendo la posibilidad de analizar el comportamiento diario de los datos, se concluye que la confiabilidad de los datos no detectados por los filtros 1 a 4 dentro de un día en el que más del 40% de los datos sí fueron detectados por estos filtros es muy baja. Por ello, el filtro 5 establece que todos los datos que se encuentran en un día con más del 40% de los datos diurnos marcados con el *flag* 1 o 2 sean marcados con el *flag* 3, aun los que habían sido inicialmente marcados con el *flag* 1 o 2, de manera tal de que el usuario final decida si los utilizará o no.

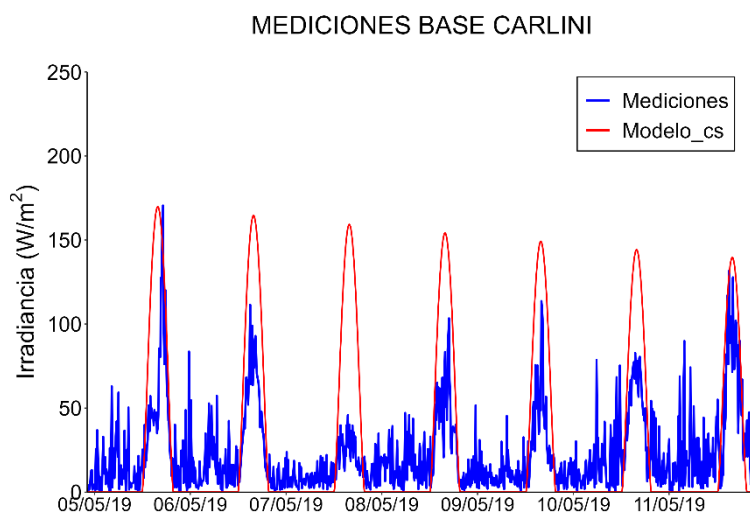


**Figura 1.** Mediciones obtenidas por la EMA instalada en Coronel Moldes y estimaciones del modelo de cielo claro.

**Filtro 6:** este filtro establece criterios para marcar con el *flag* 4 todos los valores correspondientes a un determinado día en base a tres parámetros que se pueden obtener una vez finalizado el mismo. Estos parámetros son la desviación estándar de los valores diurnos, la desviación estándar de los valores nocturnos y *offset* nocturno. El filtro 6 se subdivide en tres sub-filtros (6.1, 6.2 y 6.3 en la Tabla 1) de acuerdo al parámetro a utilizar y, como se hace uso de valores diarios, todos los valores de un determinado día se marcan con el *flag* 4 cuando no superan uno o más de estos sub-filtros.

**Sub-filtro 6.1:** Durante el análisis de los datos se encontraron estaciones que, durante un periodo de tiempo, presentaban valores con muy poca variación diurna, la cual no era esperable para su ubicación geográfica ni época del año. El sub-filtro 6.1 tiene como finalidad identificar los días en los que los valores diurnos de GHI presentan una muy baja variabilidad, lo cual se puede deber principalmente a un valor truncado a partir de cierto valor, valores que cambian muy poco durante el día, mucha suciedad en la cúpula del sensor, entre otras. También puede ocurrir que, por desconfiguración del datalogger, un día determinado haya perdido sus valores diurnos y solo tenga valores nocturnos. Teniendo en cuenta este tipo de situaciones, el sub-filtro 6.1 hace uso de la desviación estándar diaria de los datos diurnos y se establece que los valores del día son inválidos y marcados con el *flag* 4 cuando la desviación estándar es menor a  $10 \text{ W/m}^2$ . El sub-filtro 6.1 es aplicable únicamente para sitios en continente, ya que en sitios antárticos se pueden dar días con condiciones meteorológicas donde haya muy poco cambio en GHI debido a niebla o incluso si las fechas están cercanas a la noche polar. Por lo tanto, se debe aplicar para sitios en donde la latitud sea inferior a  $60^\circ\text{S}$ , siendo válido para todas las frecuencias de adquisición del dato.

**Sub-filtro 6.2:** Si bien se puede esperar que sensores termoelectricos presenten leves fluctuaciones en los valores nocturnos de GHI, no es deseable que haya una variación importante en los mismos y menos aún que se presenten valores positivos. Durante el análisis e inspección visual de los datos de GHI se encontraron días en los que la gran variación de los valores nocturnos genera sospechas sobre los valores diurnos, aun cuando los últimos pasaron los filtros anteriores. En la Figura 2 se puede observar el comportamiento de los datos de GHI para la base antártica Carlini en donde se concluye que los datos corresponden, mayormente, a ruido compatible con diversas causas (sensor mal conectado, agua en los bornes del datalogger, empalmes mal hechos, etc.) con la complicación de que, a pesar de ser visiblemente ruido, los valores diurnos se encuentran entre límites posibles por lo que se les asignaría el *flag* 0 si no se realiza algún tipo de control diario. Se encontró que estos tipos de errores tienen asociados valores nocturnos positivos muy variables y altos ( $>10 \text{ W/m}^2$ ).

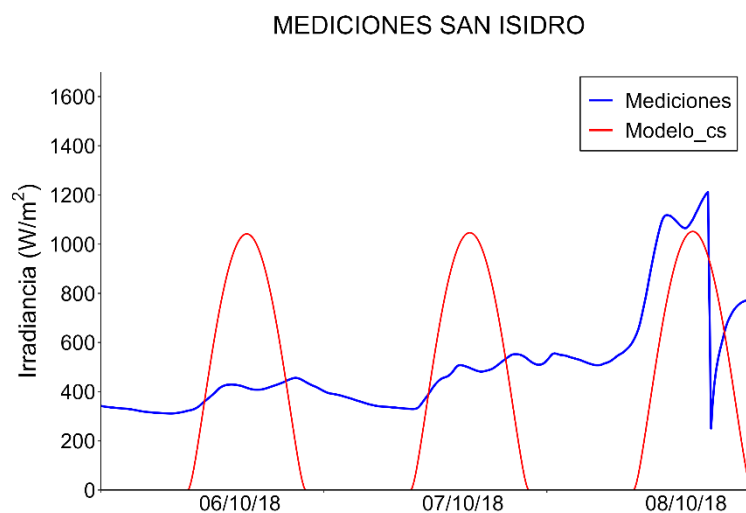


**Figura 2.** Mediciones obtenidas por la EMA instalada en base Carlini y estimaciones del modelo de cielo claro durante mayo de 2019. Las mediciones nocturnas corresponden a ruido que vuelve sospechosas las mediciones diurnas.

Luego de realizar un análisis de los valores nocturnos de GHI obtenidos por piranómetros termoelectrónicos calibrados y mantenidos, se encontró que los valores nocturnos típicamente mostraban valores de desviación estándar menores a  $2 \text{ W/m}^2$ , por lo que este fue el valor que se adoptó como límite para la desviación estándar nocturna obtenido por las EMAs. Por lo tanto, si un determinado día presenta un valor de desviación estándar con valores nocturnos mayor a  $2 \text{ W/m}^2$ , todos los valores de ese día son marcados con el *flag* 4. Cabe aclarar que, para obtener la desviación estándar de los valores nocturnos, se utilizaron todos los valores del día antes del amanecer y luego de la puesta de sol tales que  $\text{SZA} > 100^\circ$ , para evitar posibles efectos atmosféricos que incrementen los valores de GHI. El subfiltro 6.2 es aplicable para todo sitio independientemente de la latitud y del tiempo de adquisición de los datos.

**Sub-filtro 6.3:** Como ya se ha mencionado anteriormente, las instituciones no tienen todas el mismo tipo de datalogger, y cada una tiene la potestad para configurarlo como lo considere más conveniente; por ello es posible que, debido a diferencias en la electrónica y/o en la configuración del *datalogger*, además del comportamiento mismo del sensor, se pueda generar un valor de *offset* tanto negativo como positivo. Este *offset* se puede obtener de manera sencilla durante las horas nocturnas, debido a la ausencia de radiación solar, como un promedio de los valores de GHI tales que  $\text{SZA} > 100^\circ$ . Como se explicó en el subfiltro 6.2, el comportamiento de los valores nocturnos es un buen parámetro para poder detectar valores diurnos sospechosos. Es por esto que el subfiltro 6.3 se centra en los valores de *offset* nocturno con la finalidad de determinar un rango de aceptabilidad para este parámetro a través de la implementación de valores límite tanto superiores como inferiores. En la Figura 3 se pueden observar mediciones con un *offset* nocturno tan alto que, automáticamente, invalida las mediciones diurnas a pesar de que las mismas se encuentran dentro de valores adecuados, aunque sin ninguna apariencia de comportamiento típico de esta medida.

Para obtener el valor inferior se utiliza el enfoque propuesto por Long y Dutton (2002) y Long y Shi (2008), quienes establecieron que los valores nocturnos físicamente posibles pueden alcanzar valores de hasta  $-4 \text{ W/m}^2$ . Por otro lado, a través de la comparación con modelo de cielo claro, se encontraron sitios con datos de GHI con un *offset* nocturno positivo, pero con valores diurnos en su mayor parte dentro de rangos esperables. En este trabajo se encontró que un valor de *offset* nocturno de hasta  $2 \text{ W/m}^2$  sigue siendo adecuado teniendo en cuenta las diferencias en la tecnología de los dataloggers, su configuración y su calibración. Este subfiltro 6.3 es igualmente aplicable para todo sitio independientemente de la latitud y del tiempo de adquisición de los datos.



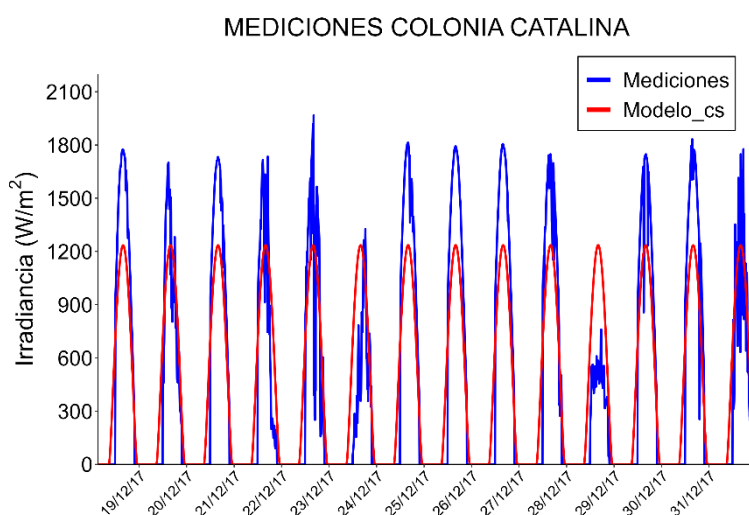
**Figura 3.** Mediciones obtenidas por la EMA instalada en San Isidro y estimaciones del modelo de cielo claro durante octubre de 2018. Las mediciones nocturnas presentan valores que vuelve sospechosas las mediciones diurnas.

**Filtro 7:** Así como trabajar con datos sospechosos de ser erróneos puede conducir a conclusiones erróneas, trabajar con datos incompletos también puede ser un problema a la hora del análisis. Por ello, el filtro 7 se

implementa en la presente metodología con la finalidad de detectar días que no cuenten con, al menos, el 80% de los datos diurnos. Para ello, en este filtro se calcula la cantidad teórica de datos diurnos que debería tener cada día en cada sitio en función de sus coordenadas, fecha y la frecuencia de muestreo. Este valor se compara con la cantidad de datos diurnos disponibles con el criterio de que si la cantidad de datos diurnos disponible es menor al 80% de los teóricos entonces todos los valores de dicho día se marcan con el *flag* 5.

## FILTRO MENSUAL

**Filtro 8:** Cuando se dispone de todos los valores de GHI correspondientes a un determinado mes luego de haber aplicado los filtros 1 al 7, es posible identificar días que, si bien superaron dichos filtros, deben ser marcados como sospechosos por estar rodeados de días que presentaron los *flags* diarios 6 y 7. Ejemplos de este tipo de comportamiento se pueden ver en las Figura 4. En esta figura se puede ver que los valores de GHI para el sitio en cuestión presentan valores muy altos, por encima incluso de la irradiancia extraterrestre, por lo que casi todos los días fueron marcados con el *flag* 5. El análisis se centra en los días 23-12 y 28-12 dado que, por la presencia de nubes, la mayor parte del día presenta valores de GHI que pasan los filtros 1 a 7. El hecho de que tantos días de un determinado mes hayan presentado un comportamiento que los invalida conlleva que esos días, donde el comportamiento parece adecuado, sea considerado también sospechoso. Es por esto que el filtro 8 de la presente metodología establece que si 10 o más días de un determinado mes han sido marcados con los *flags* 3 y/o 4 entonces los días de ese mes que no presenten esos *flags* quedarán marcados con el *flag* 6 y el usuario de los datos debe decidir si considerará esos datos en su análisis o no.



**Figura 4.** Mediciones en las que se observan que la mayor parte de las mismas se encuentran por encima del modelo de cielo claro en días despejados a excepción de días con marcada nubosidad.

Un resumen de todos los filtros presentados en la presente metodología, su correspondiente *flag* (también conocido como bandera o indicador), latitudes y frecuencia temporal en las que se aplican se muestra en la Tabla 1. Los datos que cumplen con todos los criterios tienen asociado el *flag* 0.

## JERARQUÍA DE LOS FILTROS

La metodología de QC implementada en este trabajo presenta 7 *flags*, de acuerdo al tipo de filtro utilizado. Los *flags* van desde 0 hasta 6, donde el *flag* 0 se reserva para todos los datos que pasaron los filtros. El hecho de plantear diferentes *flags* de acuerdo al tipo de filtro utilizado permite establecer una jerarquía entre los mismos, teniendo en cuenta la severidad o capacidad de detección de cada filtro, así como la invalidez de datos que cada uno de ellos presenta. Debe tenerse presente que el orden en el que los filtros se presentan, del 1 al 8, responde a la periodicidad con la que se pueden aplicar teniendo en cuenta que hay algunos que pueden aplicarse en tiempo real, otros a día vencido y otros a mes vencido. A pesar de esto, la



severidad asociada a cada uno de los flags no sigue el mismo orden que los filtros. En este trabajo se consideró que los *flags* 3 y 4 tienen la mayor jerarquía debido a que los mismos se asocian con filtros diarios, que son capaces de identificar comportamientos anómalos considerando la variabilidad diurna y nocturna. Por otro lado, los *flags* 2 y 1 siguen en el orden descendiente, respectivamente, debido a que están asociados con valores que no superan alguno de los criterios instantáneos por lo que se consideran inválidos, en el caso del *flag* 2, o sospechosos en el caso del *flag* 1. Luego continúa el *flag* 6, el cual establece la sospecha sobre los valores de todo un día, pero analizando el comportamiento de todo el mes, por lo que el criterio aplicado en el mismo es indirecto. Por último, el *flag* 5 es el que menor jerarquía o severidad presenta teniendo en cuenta que no representa un criterio para detectar datos sospechosos de ser erróneos, sino que marca los datos de acuerdo a la cantidad de datos diurnos disponibles, independientemente de su calidad.

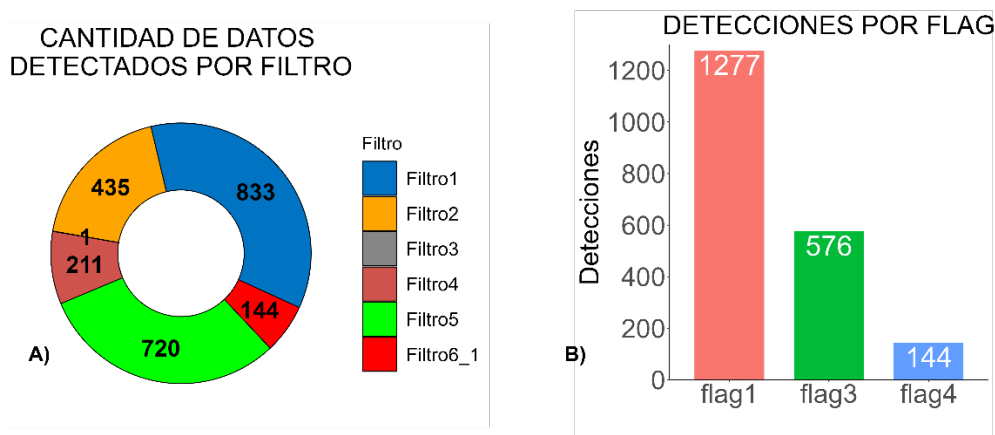
Tabla 1. Descripción de criterios para cada filtro, flag asociado, latitud para la que se aplica y tiempo de adquisición del dato para el cual se aplica.  $S$  representa la corrección en la distancia Sol-Tierra y  $E_{cs}$  la constante solar establecida en  $1361.1 \text{ W/m}^2$  (Gueymard, 2018).

Filtro	Criterio	Flag asociado	Latitud de aplicación	Tiempo de adquisición	Dato al que se aplica
1	$GHI > GHI_{cs}$	1	Todas	Todos	De máxima frecuencia
	$GHI > 1.2 * S * E_{sc} * (\cos(SZA))^{1.2} + 50 \text{ W/m}^2$	2	Todas	$\leq 10 \text{ min}$	De máxima frecuencia
	$GHI > S * E_{sc} * \cos(SZA)$	2	Todas	$> 10 \text{ min}$	De máxima frecuencia
2	$K_t < (80 - SZA) * 10^{-4}$ si $SZA \leq 80^\circ$ , $GHI \leq 0$ y $SZA > 80^\circ$	1	Latitud $< 60^\circ S$	Todos	De máxima frecuencia
3	$ K_t(t) - K_t(t-1)  > 0.75$ si $SZA < 88^\circ$	1	Todas	Todos	De máxima frecuencia
4	$(GHI(t) - GHI(t-1)) / GHI(t-1) < \pm 0.1\%$	1	Latitud $< 60^\circ S$	Todos	De máxima frecuencia
5	40% del día con el flag 1 o 2	3	Todas	Todos	Análisis diario
6.1	desviación estándar diurna de $GHI < 10 \text{ W/m}^2$	4	Latitud $< 60^\circ S$	Todos	Análisis diario
6.2	desviación estándar nocturna de $GHI > 2 \text{ W/m}^2$	4	Todas	Todos	Análisis diario
6.3	$offset \text{ nocturno} < -4 \text{ W/m}^2$	4	Todas	Todos	Análisis diario
	$offset \text{ nocturno} > 2 \text{ W/m}^2$				
7	Cantidad de datos diurnos $< 80\%$ de cantidad teórica de datos diurnos	5	Todas	Todos	Análisis diario
8	Mas de 10 días del mes marcados con el <i>flag</i> 3 o 4	6	Todas	Todos	Análisis mensual

## RESULTADOS

Para analizar una metodología de QC debe tenerse en cuenta que no es correcta la idea de que mientras más datos marque como sospechosos entonces más eficiente es y, por lo tanto, mejor que otra. Debe haber un equilibrio entre la rigurosidad y flexibilidad de los filtros en base a criterios relacionados con la GHI para evitar filtrar datos correctamente medidos de datos que deben, al menos, marcarse como sospechosos. Por lo tanto, expresar los resultados de esta metodología en base a la cantidad de datos que son detectados y filtrados por cada sitio no se considera la mejor manera de demostrar su utilidad ni eficiencia a la hora de determinar calidad. Esto es debido a que cada sitio de medición tiene sus características y, durante el análisis, se encontraron sitios cuyos valores fueron filtrados en un 10% y sitios con un gran nivel de abandono cuyos datos fueron marcados por la metodología en un 100%. El resultado puede evaluarse mejor observando los datos que pasaron todos los filtros, en los cuales no se observaron valores incorrectos. De

cualquier manera, y a modo de referencia, esta metodología se aplicó a datos confiables obtenidos por el SMN en la ciudad de Buenos Aires durante el año 2018. Para que los datos sean comparables con los obtenidos por las EMAs, los datos minutales fueron promediados para obtener un dato cada 10 minutos. Luego de realizar el promedio se obtuvo un total de 52417 datos 10-min considerando datos diurnos y nocturnos, de los cuales 2344 fueron marcados con algún flag. Luego del análisis realizado a estos datos, se puede ver que solo hubo valores detectados en los filtros 1 a 6.1 con una mayor cantidad de valores detectados por el filtro 1 y, en sintonía con los criterios de esta metodología, en segundo lugar, se encuentra el filtro 5. A su vez, la cantidad de valores detectados por el filtro 6.1 muestra que un día entero posee una desviación estándar muy baja. Por otro lado, también se obtuvo que, luego de aplicada la metodología a los datos de Buenos Aires, el resultado fue que 1277 valores se marcaron con el flag 1 mientras que 576 fueron marcados, de manera independiente, con el flag 3. Finalmente, la misma cantidad de valores detectados por el filtro 6.1 fue, también, marcada con el flag 4.



**Figura 5.** A) Cantidad de valores detectados por cada filtro al aplicar la metodología de QC a datos medidos en Buenos Aires durante el 2018. B) Cantidad de valores asociados con los flags luego de aplicada la jerarquía de los mismos.

El análisis realizado con datos confiables obtenidos en Buenos Aires no es representativo de lo que debería obtenerse en otros sitios y debe aclararse que el porcentaje de datos detectados por los filtros en este sitio no representa un valor mínimo que deba esperarse en general.

## CONCLUSIONES Y MEJORAS

Se presenta una metodología de QC novedosa para el filtrado de datos de irradiancia solar global obtenidos con estaciones meteorológicas automáticas. Esta metodología se desarrolló en base a los datos disponibles obtenidos, tanto por redes del SMN como de otras instituciones, y contempla posibles efectos transitorios y geográficos, implementación automática y la posibilidad de realizar los controles en tiempo real y/o diferido. Esto último es una de las principales ventajas ya que el análisis en base a valores diarios o mensuales permite identificar problemas que no se evidencian al trabajar con los datos adquiridos con una determinada frecuencia (5-min a 60-min). Más allá de las bondades de la metodología presentada en este trabajo, deben tenerse en cuenta que no es capaz de detectar algunos problemas conocidos asociados a los datos de GHI. Un ejemplo de esto es la desnivelación que se puede producir en los sensores y que requiere la intervención de una persona para asegurar el mantenimiento continuo. Esta metodología tampoco detecta si un sensor determinado está obteniendo valores de GHI menores a los esperables debido a la tierra o polvo que pueda tener la cúpula externa de un sensor o incluso debido a un factor de calibración inadecuado que genere la obtención de valores levemente inferiores al modelo de cielo claro. Debido a la dificultad que conlleva detectar disminuciones sistemáticas para un determinado SZA, la metodología tampoco es capaz de detectar posibles bloqueos de la radiación solar que produzcan sombras sistemáticas sobre los sensores. Una de las limitaciones se encuentra en el filtro 1 debido a que el uso de un valor fijo de TL2 en el modelo de cielo claro puede no ser aplicable a todas las escalas temporales, especialmente las más pequeñas, o tener

espacio para ser más ajustado dependiendo del sitio o época del año. Una mejora que se podría implementar sería la utilización de algún modelo más complejo que presente un mejor desempeño, lo cual se consigue utilizando mayor cantidad de parámetros de entrada.

Otro aspecto que podría incluirse en busca de una mejora en la metodología sería la inclusión de incertidumbres en las mediciones, aunque esto se torna muy complicado cuando no se conoce ni siquiera el régimen de calibraciones de varios de los sensores ni tampoco la incertidumbre en sus respectivas constantes de calibración

## REFERENCIAS

- Abal, G., Aicardi, D., Alonso-Suárez, R., Laguarda, A., 2017. Performance of empirical models for diffuse fraction in Uruguay. *Solar Energy*, 141, 166-181. doi: 10.1016/j.solener.2016.11.030
- Geiger, M., Diabaté, L., Ménard, L., y Wald, L. 2002. A web service for controlling the quality of measurements of global solar irradiation. *Solar Energy*, 73(6), 475-480. doi:10.1016/s0038-092x(02)00121-4
- Geuder, N., Wolfertstetter, F., Wilbert, S., Schüler, D., Affolter, R., Kraas, B., ... Espinar, B. 2015. Screening and Flagging of Solar Irradiation and Ancillary Meteorological Data. *Energy Procedia*, 69, 1989-1998. doi:10.1016/j.egypro.2015.03.205
- Gueymard, C. A. 2017. Cloud and albedo enhancement impacts on solar irradiance using high-frequency measurements from thermopile and photodiode radiometers. Part 1: Impacts on global horizontal irradiance. *Solar Energy*, 153, 755-765. doi:10.1016/j.solener.2017.05.004
- Gueymard, C. A. 2018. A reevaluation of the solar constant based on a 42-year total solar irradiance time series and a reconciliation of spaceborne observations. *Solar Energy*, 168, 2-9. doi:10.1016/j.solener.2018.04.001
- Journée, M., y Bertrand, C. 2011. Quality control of solar radiation data within the RMIB solar measurements network. *Solar Energy*, 85(1), 72-86. doi:10.1016/j.solener.2010.10.021
- Long, C.N., Dutton, E.G., 2002. BSRN Global Network recommended QC tests, V2.0. Available online at [https://epic.awi.de/id/eprint/30083/1/BSRN\\_recommended\\_QC\\_tests\\_V2.pdf](https://epic.awi.de/id/eprint/30083/1/BSRN_recommended_QC_tests_V2.pdf) (último acceso 08/08/2023).
- Long, C.N., Shi, Y., 2008. An automated quality assessment and control algorithm for surface radiation measurements. *Open Atmos. Sci. J.* 2, 23-27. <http://dx.doi.org/10.2174/1874282300802010023>
- Maxwell, E., Wilcox, S., y Rymes, M. 1993. Users manual for seri qc software, assessing the quality of solar radiation data. Solar Energy Research Institute, Golden, CO. Available online at <<https://www.nrel.gov/docs/legosti/old/5608.pdf>> (último acceso 08/08/2023).
- Moradi, I. 2009. Quality control of global solar radiation using sunshine duration hours. *Energy*, 34(1), 1-6. doi:10.1016/j.energy.2008.09.006
- Nollas, F. M., Salazar, G. A., y Gueymard, C. A. (2022). Quality control procedure for 1-minute pyranometric measurements of global and shadowband-based diffuse solar irradiance. *Renewable Energy*. <https://doi.org/10.1016/j.renene.2022.11.056>
- Orte, F., Wolfram, E., Luccini, E., D'Elia, R., Lusi, A., Pallotta, J., y otros (2021). Saver-Net UV-total solar irradiance monitoring network in Argentina. *Revista Meteorológica, Argentina*, 47(2), 150002-10. doi: 10.24215/1850468Xe016. Disponible en: <https://repositorio.uca.edu.ar/handle/123456789/15292>
- Pashiardis, S.; Kalogirou, S.A. 2016. Quality control of solar shortwave and terrestrial longwave radiation for surface radiation measurements at two sites in Cyprus. *Renewable Energy*, 96(), 1015-1033. doi:10.1016/j.renene.2016.04.001
- Perez-Astudillo, D., Bachour, D., y Martin-Pomares, L. 2018. Improved quality control protocols on solar radiation measurements. *Solar Energy*, 169, 425-433. doi:10.1016/j.solener.2018.05.028
- Raichijk, C. (2012). Control de calidad de mediciones de radiación solar. *Avances en Energías Renovables y Medio Ambiente*. Vol. 16, 17-22. ISSN 0329-5184
- Rigollier, C., Bauer, O., y Wald, L. 2000. On the clear sky model of the ESRA — European Solar Radiation Atlas with respect to the heliosat method. *Solar Energy*, 68(1), 33-48. doi:10.1016/s0038-092x(99)00055-9
- Rojas, R. G., Alvarado, N., Boland, J., Escobar, R., y Castillejo-Cuberos, A. 2018. Diffuse fraction estimation using the BRL model and relationship of predictors under Chilean, Costa Rican and Australian climatic conditions. *Renewable Energy*. doi:10.1016/j.renene.2018.09.079

- Ruiz-Arias, J. A., Alsamamra, H., Tovar-Pescador, J., y Pozo-Vázquez, D. 2010. Proposal of a regressive model for the hourly diffuse solar radiation under all sky conditions. *Energy Conversion and Management*, 51(5), 881–893. doi:10.1016/j.enconman.2009.11.024
- Sanchez, G., M. L. Cancillo, and A. Serrano 2016, An intercomparison of the thermal offset for different pyranometers, *J. Geophys. Res. Atmos.*, 121, 7901–7912, doi:10.1002/2016JD024815.
- Shi G. Y., Hayasaka T, Ohmura A, Chen ZH, Wang B, Zha JQ, y otros. 2008 Data quality assessment and the long-term trend of ground solar radiation in China. *J Appl Meteorol Climatol* 2008;47:1006e16.
- Yabra, M. S., R. de Elia, L. Vidal, M. Nicolini, R. Vasques Ferro, C. Ribero, L. Chiaparrí, E. Fernández, C. Campetella, O. Bonfili, M. Ceballos, G. Barrera, N. Troche, V. López, M. Schizzano, N. Bentancor, L. Berengua y M. Steven, 2021: Las nieblas en los aeropuertos argentinos: revisión de literatura y perspectiva de los pronosticadores. *Nota Técnica SMN 2021-89*
- Younes, S., Claywell, R., y Muneer, T. 2005. Quality control of solar radiation data: Present status and proposed new approaches. *Energy*, 30(9), 1533–1549. doi:10.1016/j.energy.2004.04.031

## QUALITY CONTROL OF GLOBAL SOLAR RADIATION DATA OBTAINED BY AUTOMATIC WEATHER STATIONS

**ABSTRACT:** In this work, a preliminary methodology is presented to perform a quality control of horizontal global solar irradiance data from automatic stations operated by both the National Meteorological Service and other Institutions in Argentina. The methodology is based on the implementation of filters that have different criteria, some of which have been widely studied and validated in the scientific literature, while others have been developed based on the detection of different problems observed in the available data series. The methodology presented is preliminary and does not ensure universal applicability at a scientific level, since many of the stations do not have a standard maintenance protocol or frequent calibration. However, and due to the difficulty of obtaining solar irradiance data measured on the surface, the filters presented in this methodology allow a preliminary diagnosis of the quality of the information obtained by the stations, making important information available to the user about the quality of the databases analyzed for decision-making regarding their use.

**KEYWORDS:** filters, quality, automatic weather stations, solar radiation.