

# Validación de estimativos de irradiancia global horizontal sobre la superficie uruguaya a escala diaria

Iñaki Sarazola 5.066.471-1

isarazola@fing.edu.uy

Tutor: **Agustin Laguarda**

2022



---

Firma del estudiante:

A handwritten signature in black ink, consisting of several sharp, angular strokes.

# Índice

<b>1. Introducción</b>	<b>2</b>
<b>2. Datos de Irradiancia solar en tierra</b>	<b>3</b>
<b>3. Modelos Validados</b>	<b>4</b>
3.1. LCIM - LES . . . . .	4
3.2. Heliosat-4 - CAMS . . . . .	4
3.3. POWER - NASA . . . . .	4
3.4. GL1.2 - INPE/CPTEC . . . . .	4
3.5. NSRDB - NREL . . . . .	5
3.6. MERRA2 - NASA . . . . .	5
<b>4. Metodología</b>	<b>6</b>
4.1. Control de calidad de los datos de referencia . . . . .	6
4.2. Procesamiento de los datos en tierra. . . . .	6
4.3. Control de calidad de modelos. . . . .	7
4.4. Métricas de desempeño de la validación. . . . .	8
<b>5. Resultados</b>	<b>10</b>
5.1. Resultados globales por sitio . . . . .	12
5.2. Resultado Según Nubosidad . . . . .	13
5.3. Resultados por estación del año. . . . .	15
<b>6. Conclusiones</b>	<b>16</b>
<b>7. Agradecimientos</b>	<b>17</b>
<b>8. Anexo</b>	<b>20</b>
8.1. Anexo: Resultados según Nubosidad . . . . .	20
8.2. Anexo: Distribución de nubosidad modelado por los estimativos . . . . .	20

# 1. Introducción

Conocer el comportamiento del recurso solar con precisión es de vital importancia para el desarrollo, dimensionamiento y valoración de riesgos en proyectos de aprovechamiento de energía de tal naturaleza; si bien en la actualidad el sector energético es el principal usuario de esta información, el material no deja de ser de suma importancia para aplicaciones agropecuarias de producción vegetal, investigación médica entre otras diversas áreas.

En ausencia de mediciones en tierra de calidad y larga duración, es usual utilizar información derivada de observaciones satelitales meteorológicas o modelaciones de la atmósfera, cuyas limitaciones en cobertura espacial y extensión temporal de registro es menor que las mediciones; sin embargo, el uso de estas estimaciones frecuentemente implica un compromiso en la incertidumbre de los datos utilizados y es este el factor principal que afecta la viabilidad financiera del emprendimientos o la fiabilidad de la investigación.

Existen varias bases de datos de irradiación solar generadas por modelos de distinta complejidad que utilizan información satelital de diversas fuentes para describir los parámetros determinantes del comportamiento de la región como lo es la nubosidad. Además de las características de cada modelo, el desempeño de los datos que proporcionan es dispar en el globo, dependiendo su certeza en las particularidades climáticas de cada región, esto hace que para conocer su incertidumbre sean necesarias evaluaciones locales.

El presente estudio analiza el desempeño de seis bases de datos de datos de irradiación solar diaria acumulada en plano horizontal ( $GHI$  [ $\frac{kWh}{m^2}$ ]) que son comúnmente utilizadas y de acceso público, tal validación se lleva a cabo para la superficie uruguaya a nivel diario para el periodo 2018-2021 siendo 7 las estaciones de medida que respaldan el recurso en el sitio.

## 2. Datos de Irradiancia solar en tierra

Como medidas de referencia se utilizaron datos provenientes de 7 estaciones de la Red de Medida de Irradiancia Solar (RMCIS: <http://les.edu.uy/rmcis/>) del LES (Laboratorio de Energía Solar), cuya disposición espacial se presenta en la figura 2.1.

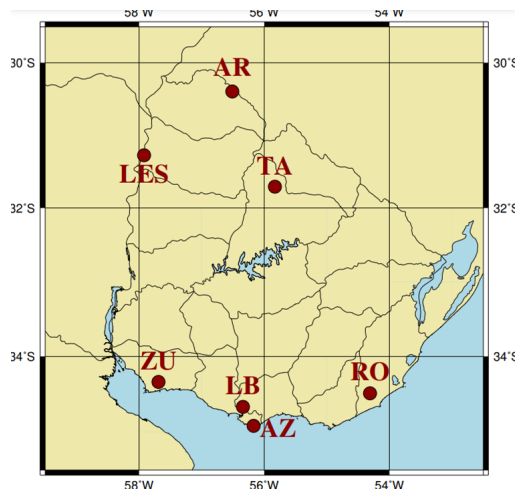


Figura 2.1: ubicación geográfica de las estaciones de medida.

Las presentes siglas del mapa 2.1 refieren a las estaciones presentadas en el cuadro 1.

Departamento	Local	Código	Latitud [°]	Longitud[°]	Altura[m]
Artigas	Aeropuerto Artigas	AR	-30.3984	-56.5117	136
Salto	Salto Grande	LES	-31.2827	-57.9181	56
Tacuarembó	La Magnolia	TA	-31.7387	-55.9792	142
Colonia	La Estanzuela	ZU	-34.3380	-57.6904	70
Canelones	Las Brujas	LB	-34.6720	-56.3401	38
Montevideo	Punta Carretas	AZ	-34.9182	-56.1665	58
Rocha	Ruta 9 y Ruta 15	RC	-34.4893	-54.3203	20

Cuadro 1: Estaciones de medida utilizadas.

El LES administra varias estaciones de medida de irradiancia solar distribuidas en el territorio uruguayo. Los datos se registran a nivel minutilizando piranómetros clase A o B (norma ISO 9060:2018) que se mantienen calibrados cada dos años con trazabilidad técnica a la referencia radiométrica mundial mantenida por el PMOD-World Radiation Center (Suiza). Los instrumentos en LE reciben mantenimiento diario (limpieza de cúpulas y nivelación de instrumentos). Las estaciones RC y AZ tiene mantenimiento al menos mensual. Las demás estaciones reciben mantenimiento con frecuencia semanal.

Los datos históricos de irradiación solar global diaria sobre plano horizontal están disponibles para visualización abierta al público y descarga a través del Sistema de Gestión de Datos del LES [1].

Se utilizan en cada estación de medida los datos minutil entre el periodo 2018 - 2021, con estos, luego de aplicar controles de calidad, cuyo procedimiento se describe en la sección 4.1, se obtienen los acumulados diarios y con tales se realiza el análisis.

### 3. Modelos Validados

#### 3.1. LCIM - LES

El presente se basa en el concepto central de tomar como base un modelo de cielo claro y modificar sus estimativos en función de un índice de nubosidad (CIM significa Cloud Index Model) basado en información satelital.

$$I = I_{cc}F(\eta) \quad (3.1)$$

En particular, para modelar la irradiación global horizontal en condiciones de cielo completamente despejado, LCIM hace uso de el modelo ESRA [2] (ya validado para la region ) [3] que depende del factor de 'Turbidez de Linke' [4] llamado  $T_L$ , para una masa de aire igual a dos para representar el estado de la atmósfera en un único término. Es decir, incluye los efectos de dispersión y absorción de gases residuales y absorción de vapor de agua; en la presente versión se utiliza un ciclo estacional de  $T_L$  promediado de ciclos medios de TL en 10 estaciones distintas de la región [5]. Por otro lado, para el planteo de el factor de atenuación  $F(\eta)$ , se propone una dependencia lineal.

$$F(\eta) = a(1 - \eta) + b \quad (3.2)$$

Donde los parámetros a y b son ajustados al sitio de utilización. En su versión operativa, LCIM utiliza parámetros empíricos promediados espacialmente sobre diez sitios de la Pampa Húmeda. Esto no supone gran pérdida de precisión, dado que los parámetros presentan baja variabilidad espacial en la región [6]. Los datos se encuentran disponibles para descarga en la pagina del laboratorio de energía solar [7]

#### 3.2. Heliosat-4 - CAMS

El método Heliosat-4 [8] ya fue previamente validado para la región a nivel horario [9] y a nivel diario [10], [3]. Estima la radiación de onda corta descendente recibida a nivel del suelo en todos los condiciones del cielo. Es un modelo totalmente físico que utiliza una aproximación rápida, pero aún así precisa del modelo de transferencia radiativa y, por lo tanto es, es muy adecuado para el uso de información satelital geoestacionaria.

Está compuesto por dos modelos basados utilizados en forma de ábacos, también llamados tablas de consulta (o *LUT* por sus siglas en ingles 'look up table'): el modelo McClear para el cálculo de irradiancia en condiciones sin nubes y el modelo McCloud calcula la extinción de la irradiancia debido a las nubes. Ambos se han realizado utilizando el modelo de transferencia radiativa libRadtran.

Las entradas principales al modelo se refieren a información sobre las propiedades de la atmósfera como vapor de agua, propiedades de los aerosoles y contenido de ozono; esta información es provista por el Copernicus Atmosphere Monitoring Service (CAMS) con resolución tri-horaria.

Las propiedades de las nubes se derivan de imágenes de los satélites Meteosat Segunda Generación (MSG) en su resolución temporal de 15 min usando un esquema APOLLO (AVHRR Processing Scheme Over Clouds, Land and Ocean) adaptado. Se disponen de datos a escala diaria entre 2018 y 2021, estos pueden están disponibles a diferentes escalas en el portal [11].

#### 3.3. POWER - NASA

La plataforma POWER [12] plantea disponer de un amplio abanico de predicciones de interés para el área de las energías renovables, entre ellas, presenta un estimativo de radiación solar; para el periodo pertinente en el presente análisis, POWER utiliza datos provenientes de 'NASA's CERES SYN1deg Edition 4.1', el proyecto CERES aprovecha las capacidades de generación de imágenes GEO de ultima generación para recuperación de nubosidad; tambien se utiliza información de insolación solar superficial por parte del Proyecto de intercambio de agua y energía global (GEWEX) de la NASA (Surface Radiation Budget (SRB)). Los datos modelados son de acceso libre, disponibles a escala diara en el portal [13]

#### 3.4. GL1.2 - INPE/CPTEC

El modelo GL1.2 [14] es un modelo de radiación solar en superficie a nivel diario utilizado y desarrollado por el INPE/CPTEC (Instituto nacional de pesquisas espaciais /centro de previsao de

estudios climáticos) de Brasil. El modelo plantea una solución física y simple al balance radiativo del sistema tierra atmósfera. Fue diseñado mediante simplificaciones consistentes con la estructura que describe la transferencia de radiación solar de onda corta en el sistema Tierra-atmósfera. Se divide el espectro radiativo en intervalo de longitudes de onda entre visible, ultravioleta e infrarrojo, y analizando la atmósfera en: troposfera (*altura < 17km*) y estratosfera(*altura > 17km*). Aplicando distintas hipótesis para cada una de las secciones en cuestión, utilizando información de índole satélite proveniente del satélite GOES y haciendo uso de parametrizaciones y balances radiativos se obtienen los correspondientes estimativos.

### 3.5. NSRDB - NREL

Por sus siglas en ingles, Net Surface Radiation Data Base [15]- [16], utiliza el modelo PSM v3 -Physical Solar Model -, un modelo de dos pasos 'Two-step model' que pretende comprender la física completa que afecta la transmisión de la radiación solar entre el tope de la atmósfera y la superficie terrestre; El modelo utiliza información de nubes proveniente de los satélites geoestacionarios del entorno operativo (GOES) que cubren América del Norte y del Sur; variables como la profundidad óptica del aerosol o el vapor de agua se extraen de la base de datos MERRA2 y se utilizan metodologías de reducción de escala para hacer coincidir los datos MERRA2 de resolución más baja con la cuadrícula NSRDB de 4 km. Considerando el amplio abanico de información consultada por el modelo, este requiere un método computacional de alto rendimiento mediante el cual se resuelven las correspondientes ecuaciones de transferencia radiativa mediante el algoritmo FARMS (Fast All-sky Radiation Model for Solar applications) [17]. Los sets de datos utilizados son de publico acceso y se encuentran disponibles en la plataforma del NSRDB [18]

### 3.6. MERRA2 - NASA

Por sus siglas en ingles 'Modern-Era Retrospective Analysis for Research and Applications'[19] es un método de reanálisis ; proporciona un análisis global dinámicamente coherente de la características atmosféricas globales al combinar el modelado geofísico dinámico de fluidos de la atmósfera y mediciones satelitales, en tierra, de aeronaves y sondas meteorológicas.

El modelo incluye importantes procesos físicos como la transferencia radiativa y la convección. Debido a que los datos de reanálisis se obtienen mediante simulación numérica, no pueden reemplazar por completo los datos observados para describir el verdadero estado tridimensional de la atmósfera. Los errores en los productos de reanálisis de radiación ocurren debido a la heterogeneidad de las diferentes fuentes de datos y los diferentes esquemas de asimilación de datos. Por lo tanto, comprender la incertidumbre y el sesgo de los datos de reanálisis es un requisito previo para el uso racional de los datos de reanálisis.

Las bases de reanálisis también, al no depender de imágenes satélites tienen la ventaja de presentar *datasets* ininterrumpidos, es decir sin 'huecos'; además, otro gran atributo que presentan los modelos de esta naturaleza es la posibilidad de cubrir grandes períodos temporales ([1980 - Presente ] para MERRA2) y tienen gran cobertura espacial en todo el globo (en el caso de MERRA2 se tiene una resolución  $0,5^{\circ} \times 0,625^{\circ}$ )

Los datos modelados se encuentran disponibles en [20].

## 4. Metodología

Para llevar a cabo el estudio fue necesario calcular los acumulados diarios de las medidas en tierra (utilizadas como referencia) y de los modelos que proveen información intradiaria. Para el caso de los datos en tierra, eso incluye un exhaustivo control de calidad que se describe a continuación, para los datos modelados el acumulado implica también un conjunto de decisiones a cerca del enmascarado propuesto por la fuente. Para llevar a cabo el análisis se utilizó el lenguaje de programación Python.

### 4.1. Control de calidad de los datos de referencia.

Como primera instancia de control de calidad de datos en tierra se realizó un análisis visual sobre los datos crudos, de forma de descartar anomalías evidentes o errores en la etiqueta temporal; en la figura 4.1 se presenta un día descartado por posible mal funcionamiento del equipo de medición.

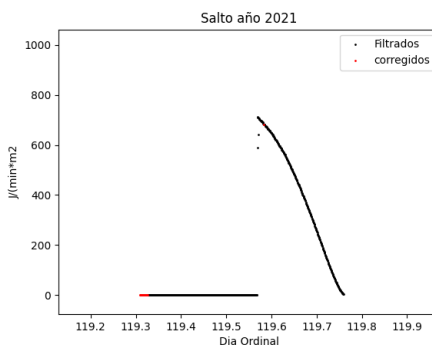


Figura 4.1: Día descartado por análisis visual

A partir de la serie temporal bien etiquetada se procede con el filtrado de calidad de datos; en donde, además de despreciar el intervalo nocturno que no es de interés, se aplican filtros BSRN para datos extremadamente raros ([21])

$$-2 \frac{W}{m^2} < GHI < 1,2G_s \cos^2 \theta_z + 50 \frac{W}{m^2}, \quad (4.1)$$

donde  $\theta_z$  es el ángulo cenital y  $G_s = 1361 \frac{W}{m^2}$  es la llamada constante solar [22]; este criterio, supone un filtrado básico, eliminando únicamente valores físicamente posibles pero altamente improbables. Para el estudio en cuestión, no se consideran valores negativos, por lo tanto se añade un filtro que limita GHI a valores únicamente positivos, es decir  $GHI > 0$ . En la tabla 2 se presentan la cantidad de datos descartados en cada estación, como datos disponibles, se interpretan los datos diurnos de medida.

Datos	Artigas	Salto	Canelones	Colonia	Rocha	Tacuarembó	Montevideo
Disponibles (min)	1012550	1047507	1048631	1047792	1043687	1045943	1035811
Descartados (min)	42254	24091	2869	3661	16273	7152	22061
Descartados (%)	4.17 %	2.29 %	0.27 %	0.35 %	1.56 %	0.68 %	2.13 %

Cuadro 2: Datos disponibles para el estudio.

### 4.2. Procesamiento de los datos en tierra.

Considerando que el propósito del control de datos explicado en la sección 4.1 es obtener un acumulado diario representativo de irradiación diaria en la correspondiente ubicación, en algunos casos los 'baches' generados por el filtrado deben ser rellenados. Esta acción se lleva a cabo cuando los datos faltantes son pocos de forma que el acumulado diario sea representativo. En este estudio se recurre al método de interpolación para la corrección de datos; tal interpolación se lleva a cabo en la magnitud  $k_t$  -índice de claridad minutil-, para excluir el efecto de la posición relativa del sol. Es importante destacar que interpolar datos en un día con información faltante puede sesgar el resultado del acumulado

diario, por lo que se aplica únicamente en casos específicos como se describe a continuación. Ya que la variable de interés es diaria se observa cada día como grupo de datos y se pondera la validez de tal día. Se analiza la cantidad de correcciones en el día; en una primera instancia, se descartan todos los días con un intervalo interpolado ininterrumpido mayor a treinta minutos, como se ejemplifica en la figura 4.2a, cuanto mas largo es el intervalo interpolado, más incertidumbre se genera en la interpolación, en la figura 4.2b se muestra un día descartado por tener mayor cantidad de datos faltantes que el umbral propuesto.

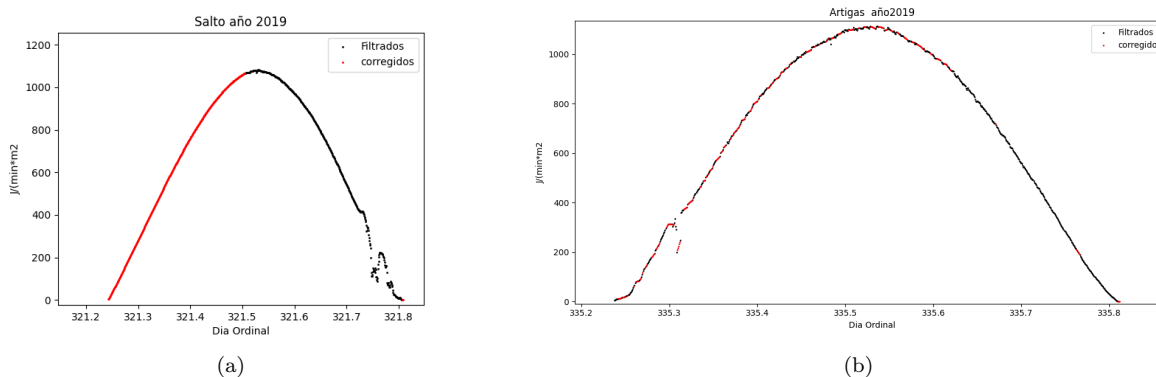


Figura 4.2: Ejemplos de días descartados por datos faltantes.

Por otro lado, se analiza como afecta el disponer de un alto numero de datos interpolados pero no necesariamente contiguos; tal situación se ejemplifica en 4.2b , en el presente caso, se presume que puede existir alguna falla en el equipamiento de medición o almacenamiento de datos, de manera que se toma como un posible error y se descarta el día; para determinar cual es la cantidad de datos interpolados crítica, se presenta el diagrama 4.3.

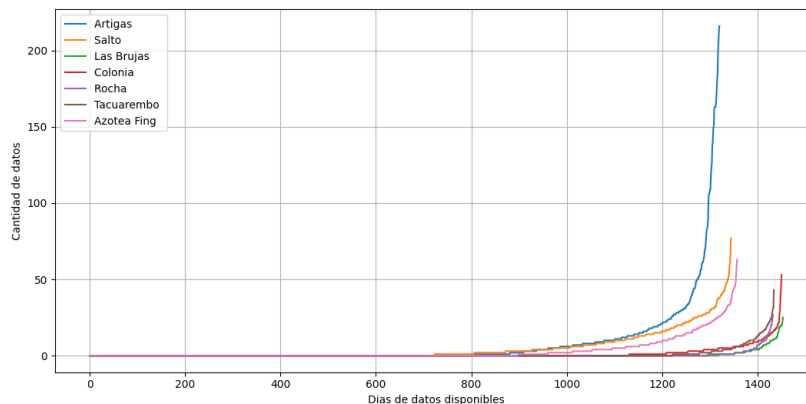


Figura 4.3: Cantidad de datos interpolados (descartando los mencionados en 4.2a)

Observando 4.3 se estima que un valor discriminante para este segundo error es también treinta minutos, por lo tanto se descartan todos los días que dispongan de mas de 30 datos interpolados. Finalmente se obtiene la tabla 3 que representa la cantidad total de datos interpolados utilizados en cada estación.

### 4.3. Control de calidad de modelos.

Los modelos analizados proveen los datos en distintas escalas (horaria, diezminutal, minutal, etc...), los que vienen en escala diaria pasan directamente a la etapa de análisis, y los que disponen de una



Datos	Artigas	Salto	Canelones	Colonia	Rocha	Tacuarembó	Montevideo
Interpolados (min)	3843	4850	753	1576	596	1277	3451
Interpolados (%)	0.20 %	0.26 %	0.04 %	0.08 %	0.03 %	0.06 %	0.18 %
Disponibles (días)	1240	1300	1454	1447	1433	1432	1333

Cuadro 3: Datos interpolados utilizados en el estudio y acumulados diarios obtenidos.

escala temporal inferior se compila su acumulado diario; en tal situación, de manera de no perjudicar el desempeño del estimativo, se descarta cualquier día que incluya alguna alerta proveniente de la fuente. Por otro lado, la mayoría de los estimativos se basan en información satelital, la cual no siempre esta disponible, es por esto que se generan faltantes en el *set* de datos, tales se corrigen con el criterio anterior de no tener un intervalo faltante mayor a treinta minutos y se recurre al interpolado. Con este enmascarado presente, se observa que no fue utilizado ningún dato interpolado, lo cual se justifica por el hecho natural de que si se carece de información satelital, el fallo generalmente es prolongado (de mas de media hora), por lo que se descartara tal día automáticamente.

#### 4.4. Métricas de desempeño de la validación.

Para evaluar el desempeño de los modelos se utilizan una serie de indicadores que caracterizan el desvío en las estimaciones. La evaluación se realiza comparando la estimación del modelo ( $\hat{y}$ ) contra la verdad medida en tierra ( $y^{ref}$ ). Se han propuesto varios indicadores para cuantificar el desempeño de modelos en el contexto del recurso solar [23], la mayoría de los cuáles son estadísticos de uso común. De las opciones disponibles, se reportará el desvío promedio (MBD), el desvío absoluto promedio (MAD) y el desvío cuadrático medio (RMSD), definidos según:

$$MBD = \frac{1}{N} \sum_{i=1}^{i=N} (\hat{y}_i - y_i^{ref}), \quad (4.2)$$

$$MAD = \frac{1}{N} \sum_{i=1}^{i=N} |\hat{y}_i - y_i^{ref}|, \quad (4.3)$$

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^{i=N} (\hat{y}_i - y_i^{ref})^2}, \quad (4.4)$$

donde  $N$  es la cantidad de muestras en los datasets  $\hat{y}$  e  $y^{ref}$ . La definición del MBD es tal que un valor positivo indica sobrestimación del modelo frente a las medidas y un valor negativo indica subestimación. Los indicadores relativos de estas cantidades, rMBD, rMAD y rRMSD, se expresan como porcentaje de la media de las medidas ( $\bar{y}^{ref}$ ). Los indicadores anteriores miden el sesgo y la dispersión del desvío en términos individuales (muestra a muestra). Otra categoría de indicadores busca medir la diferencia estadística de ambos conjuntos de datos ( $\hat{y}$  e  $y^{ref}$ ) comparando sus distribuciones de probabilidad. Dado un set de datos  $y_A$  de tamaño  $N_A$ , la distribución de probabilidad acumulada ( $CDF_A$ ) se define como:

$$CDF_A(y) = Prob(y \leq y_A) = \frac{1}{N_A} \sum_{l=1}^{l=N:A} \Theta(y - y_A(l)) \quad (4.5)$$

donde  $\Theta(y)$  es la función escalón de Heavyside. En base a la CDF de  $\hat{y}$  ( $CDF_{\hat{y}}$ ) y de  $y^{ref}$  ( $CDF_{ref}$ ), se define su diferencia absoluta como,

$$D(y) = |CDF_{\hat{y}}(y) - CDF_{ref}(y)| \quad (4.6)$$

$D(y)$  es una función positiva cuyo máximo indica el valor de  $y$  para el cual hay mayor diferencia estadística entre el estimado y la medida. Este máximo ( $D^{max}$ ) se utiliza en el test no paramétrico de Kolmogorov-Smirnov (KS) para determinar si dos conjuntos de datos son estadísticamente similares [24].  $D^{max}$  se compara con un umbral  $V_C$  que depende de la cantidad de muestras  $N$  y el nivel de

confianza  $\alpha$ . Para datasets con  $N > 35$  y un nivel de confianza del 1 %, el umbral se puede aproximar por  $V_C \approx \frac{1.63}{\sqrt{N}}$  [25]. La integral de  $D(y)$  se denomina KSI (Kolmogorov-Smirnov Integral), de la forma:

$$KSI = \int_{\delta_y} D(y)dy, \quad (4.7)$$

KSI es un indicador que mide la distancia entre las dos distribuciones. También en este estudio se incluirá el coeficiente de correlación llamándolo *corr*, este cuantifica la intensidad de la relación lineal entre las dos variables en un análisis de correlación.

Otros autores reportan la desviación estándar (SD) . La desviación estándar se relaciona con los indicadores MBD y RMSD como  $SD = \sqrt{RMSD^2 - MBD^2}$  y se expresa en forma porcentual (rSD) relativo a la media  $\bar{y}^{ref}$ .

El indicador MBD mide el sesgo sistemático que un modelo puede introducir en la estimación de largo plazo, y debe ser muy pequeño para que un modelo sea adecuado para caracterizar el recurso. Los modelos ajustados localmente generalmente tienen sesgo nulo ( $MBD \approx 0$ ), por lo que el indicador más utilizado es el RMSD, que además coincide con el SD para estos casos. La información contenida en el MAD es similar a la del RMSD y en este trabajo se reportará por completitud, puesto que permite comparar con trabajos que reporten este indicador. El KSI mide la diferencia estadística entre los datos, a diferencia del RMSD que mide la dispersión del desvío. Usualmente un modelo de bajo RMSD tiene también bajo KSI, pero formalmente miden características distintas del desempeño del estimado.

En este análisis se hará uso fundamental de los indicadores relativos a la media de las medidas, ya que facilita la visualización de los efectos en un trabajo de comparación, sin embargo, se incluirá la media de los datos para poderse recuperar el valor absoluto si el lector así lo desea.

## 5. Resultados

Una vez los datos ya procesados y filtrados para cada una de las estaciones se procede con el análisis y comparación de los datos modelados con los validos. Para evidenciar la asimilación de los estimativos con los datos en tierra, se presenta el gráfico 5.1 (se realizo de igual forma para el resto de las localidades de medición) , donde, al contraponer medida estimada a medida real se espera que los datos se aproximen a una recta  $y = x$ , siendo un caso perfecto; de forma adicional, se utiliza un código de colores para diferenciar las etapas estacionales del año.

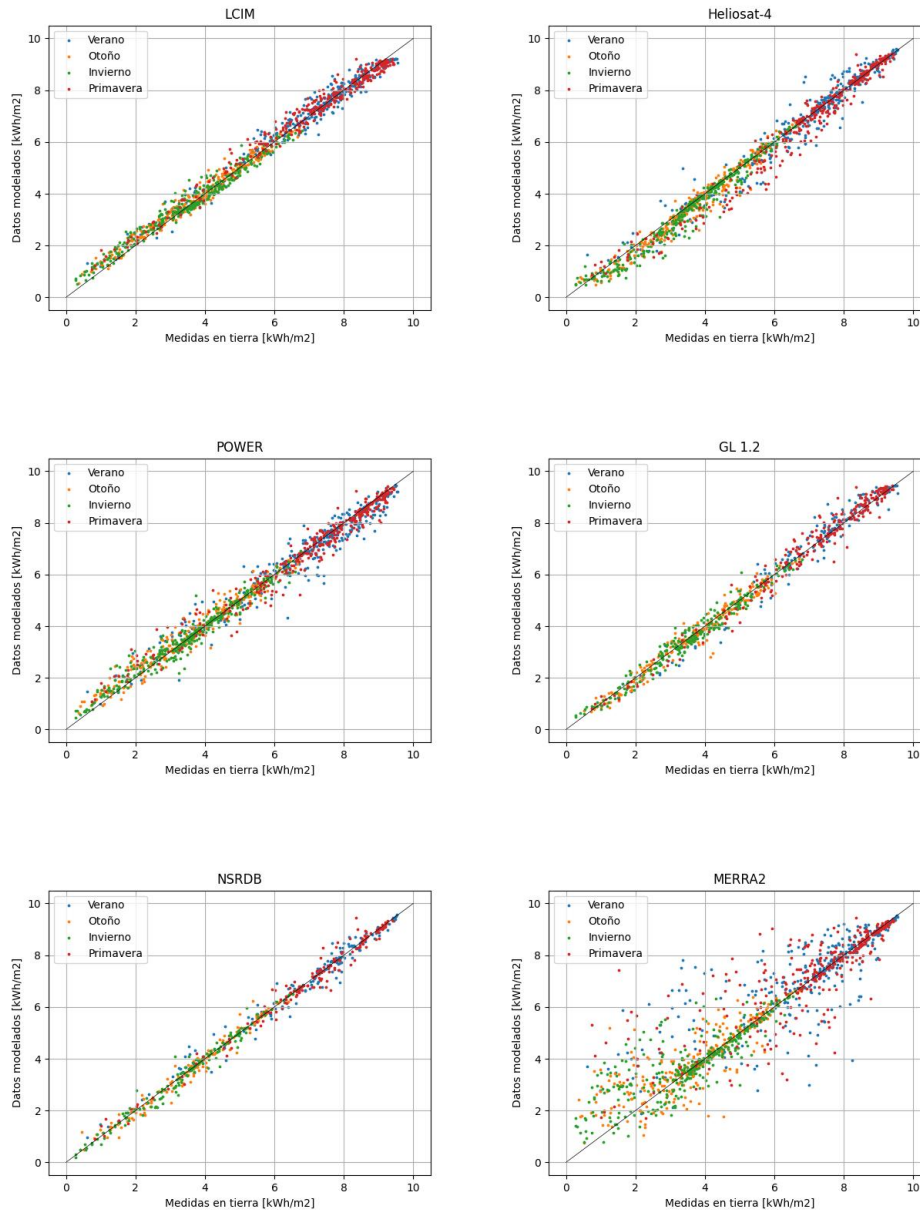


Figura 5.1: Medidas vs Modelo, estación de salto -LES-

A simple vista, se observa una buen comportamiento de los estimativos, también es de destacar que se obtiene los primeros indicios de la diferencia de la naturaleza entre MERRA 2 y el resto. De igual forma y a modo de una inspección visual preliminar, se presenta la serie temporal total de todos los modelos en la gráfica 5.2 tomando como ejemplo la estación de Estanzuela.

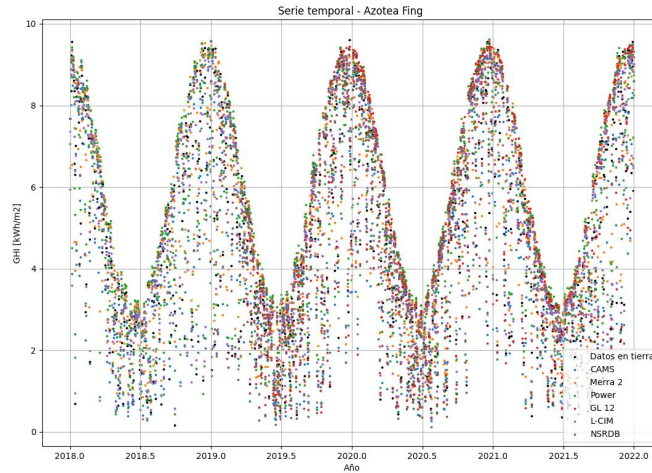


Figura 5.2: Serie temporal - Estanzuela, Colonia.

Debido a la gran acumulación de puntos, la figura 5.2 es compleja de interpretar, sin embargo, el propósito de la misma es corroborar que todos los modelos siguen el comportamiento esperado en parámetros generales y eliminar *disclaimers*, lo cual

## 5.1. Resultados globales por sitio

En la presente sección se presentaran los datos de desempeño global correspondiente a cada uno de los métodos de estimación.

Por motivos ajenos al estudio, los datos estimados por GL1.2 y NSRDB solo pudieron ser obtenidos para los períodos [04/2019-2021] y [2019-2020], respectivamente; de manera de conservar la mayor cantidad de datos posibles y así darle una mayor validez al estudio el análisis se realizara en paralelo de forma individual para estos dos estimativos mencionados y el resto de los modelos se comparan en la misma cantidad de datos, es decir en el rango temporal [2018-2021]. Para resaltar este comentario se agrega la referencia \* en el cuadro 4.

modelo	métrica	AR	LE	TA	ZU	LB	AZ	RC	Promedio
<b>LCIM</b>	rMBD (%)	2.1	1.6	2.1	-0.2	1.2	0.5	1.7	<b>1.3</b>
	rRMSD (%)	5.9	5.5	7.0	5.7	5.6	6.0	6.2	<b>6.0</b>
	rMAD (%)	4.6	4.3	5.3	4.7	4.4	4.8	4.9	<b>4.7</b>
	Corr.	0.994	0.994	0.992	0.994	0.995	0.994	0.994	<b>0.994</b>
	KSI [ $\frac{Wh}{m^2}$ ]	122.3	108.4	114.3	104.1	90.9	102.5	91.5	<b>104.9</b>
<b>NSRDB*</b>	rMBD (%)	2.6	0.1	2.7	0.3	2.3	-1.5	2.6	<b>1.3</b>
	rRMSD (%)	7.9	4.9	9.1	6.3	7.4	6.9	8.8	<b>7.3</b>
	rMAD (%)	4.5	3.3	5.7	4.4	5.3	4.6	6.1	<b>4.8</b>
	Corr.	0.989	0.994	0.986	0.993	0.992	0.991	0.989	<b>0.990</b>
	KSI [ $\frac{Wh}{m^2}$ ]	131.3	51.5	134.2	75.7	122.7	81.6	128.8	<b>103.7</b>
<b>GL1.2*</b>	rMBD (%)	0.5	0.1	-0.1	0.3	-0.9	0.3	-1.8	<b>-0.2</b>
	rRMSD (%)	5.8	5.9	6.9	13.5	6.9	7.1	8.6	<b>7.8</b>
	rMAD (%)	4.3	4.3	5.1	9.0	5.0	5.3	6.2	<b>5.6</b>
	Corr.	0.994	0.993	0.992	0.969	0.992	0.991	0.989	<b>0.988</b>
	KSI [ $\frac{Wh}{m^2}$ ]	80.9	81.5	127.7	157.7	133.3	89.5	144.9	<b>116.5</b>
<b>Power</b>	rMBD (%)	0.4	0.3	1.1	1.3	5.3	4.1	3.6	<b>2.3</b>
	rRMSD (%)	6.5	7.0	7.5	7.3	9.8	11.6	9.8	<b>8.5</b>
	rMAD (%)	4.8	4.8	5.5	5.2	7.1	8.0	7.0	<b>6.1</b>
	Corr.	0.992	0.989	0.990	0.990	0.988	0.976	0.985	<b>0.987</b>
	KSI [ $\frac{Wh}{m^2}$ ]	101.4	111.1	112.1	89.4	250.7	208.9	163.8	<b>148.2</b>
<b>Heliosat-4</b>	rMBD (%)	-0.8	-2.3	0.2	-2.5	-1.6	-5.0	-1.3	<b>-1.9</b>
	rRMSD (%)	8.3	7.3	7.6	7.9	8.6	10.1	8.6	<b>8.3</b>
	rMAD (%)	5.9	4.9	5.6	5.6	6.0	7.0	6.1	<b>5.9</b>
	Corr.	0.987	0.990	0.991	0.990	0.988	0.985	0.988	<b>0.988</b>
	KSI [ $\frac{Wh}{m^2}$ ]	137.1	148.0	123.9	152.2	133.3	251.1	115.8	<b>151.6</b>
<b>MERRA2</b>	rMBD (%)	6.3	4.7	10.3	6.3	8.6	6.1	11.0	<b>7.6</b>
	rRMSD (%)	20.5	19.6	24.6	20.5	21.9	19.7	23.9	<b>21.5</b>
	rMAD (%)	13.3	12.0	16.3	13.3	14.6	12.8	16.4	<b>14.1</b>
	Corr.	0.917	0.907	0.903	0.924	0.925	0.927	0.921	<b>0.918</b>
	KSI [ $\frac{Wh}{m^2}$ ]	318.7	247.6	495.9	307.9	407.1	302.7	496.5	<b>368.1</b>
<b># datos</b>		1220	1279	1409	1422	1430	1311	1410	<b>1354</b>
<b>media de las medidas [<math>\frac{kWh}{m^2}</math>]</b>		5.0	5.3	4.8	4.9	4.7	4.9	4.5	<b>4.9</b>
<b># datos* GL1.2  NSRDB</b>		800 520	870 644	959 721	971 724	976 728	898 660	958 706	<b>919 672</b>
<b>media GL1.2  NSRDB [<math>\frac{kWh}{m^2}</math>]</b>		5.0 5.0	5.2 5.2	4.8 4.8	4.8 4.9	4.7 4.7	4.8 5.0	4.5 4.5	<b>4.8 4.9</b>

Cuadro 4: Desempeño a nivel diario de modelos de GHI por sitio.

Como se analiza en el cuadro 4, como primer resultado, se respalda numéricamente el postulado de la diferencia en desempeño del modelo MERRA2 frente al resto; por otro lado, buscando la precisión, LCIM presenta mejores indicadores, tanto en desvío cuadrático medio como en consistencia espacial, seguido a esto se encuentran los modelos GL1.2 y NSRDB que como ya se comentó, no se puede afirmar la superioridad de los modelos frente al resto, pero lo que si es cierto es que se obtiene una tendencia a mejor desempeño por parte de los modelos que utilizan información satélite proveniente del satélite GOES (estos tres últimos mencionados) lo cual es coherente ya que el tal satélite tiene mejor vista sobre la región como se menciona en [9].

## 5.2. Resultado Según Nubosidad

Para desglosar de forma mas profunda el comportamiento de los estimativos; a continuación se plantea un análisis utilizando como parámetro fundamental la variable  $K_t$  índice de claridad el cual se define de la siguiente forma:

$$K_t = \frac{GHI_{medido}}{GHI_{TOA}} \quad (5.1)$$

Siendo  $GHI_{TOA}$  la irradiancia global horizontal en el tope de la atmósfera ('*Top of Atmosphere*'), en este caso, considerando que se trabaja con medidas de naturaleza diaria, la llamada variable es el índice de claridad diario. En este análisis, se busca analizar el comportamiento de los modelos en base a la nubosidad existente, por lo tanto los resultados numéricos presentados serán promedios espaciales de las estaciones.

Como primer análisis, se busca analizar la capacidad predictiva de cada uno de los estimativos; con este propósito se presentan los gráficos de la figura 5.3 donde se compara la función densidad de  $K_t$  predicha por el modelo contra la real, se presenta tal resultado en la estación de LB.

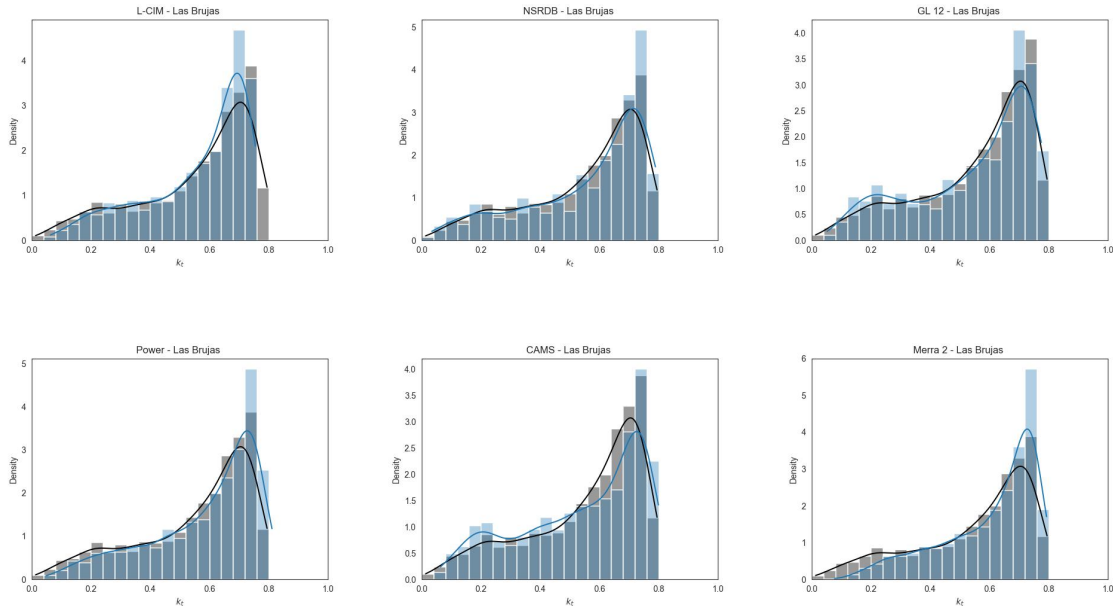


Figura 5.3: Distribución de nubosidad de los modelos - LB - en azul modelo y en gris real, la curva negra representa la tendencia de nubosidad, en el anexo se presentan las restantes estaciones.

Un aspecto importante a destacar de la figura 5.3 es que ninguno de los modelos resalta drásticamente sobre el resto, se observa un comportamiento relativamente uniforme a lo largo de los gráficos; sin embargo, se puede analizar la sensibilidad de la predicción de nubosidad de cada modelo. Hay modelos que tienen una tendencia a observar un mayor número de días claros que el real, como NSRDB, GL1.2 o CAMS; otros que ven menos días completamente nublados como MERRA2; por último, si se analiza el comportamiento de LCIM se observa una buena correlación entre lo predicho y lo real en días mediana y altamente nublados, pero por otro lado se presenta una subestimación de días completamente despejados, lo cual se traduce en una sobreestimación de días levemente nublados; confirmando las observaciones ya observadas en estudios anteriores [9].

Con el propósito de respaldar las observaciones con métricas de desempeño, se presentan las figuras 5.4b, 5.4a y 5.5 donde cada una de las barras indica la media espacial de la métrica de desempeño correspondiente en un intervalo de  $K_t$  de amplitud 0,1.

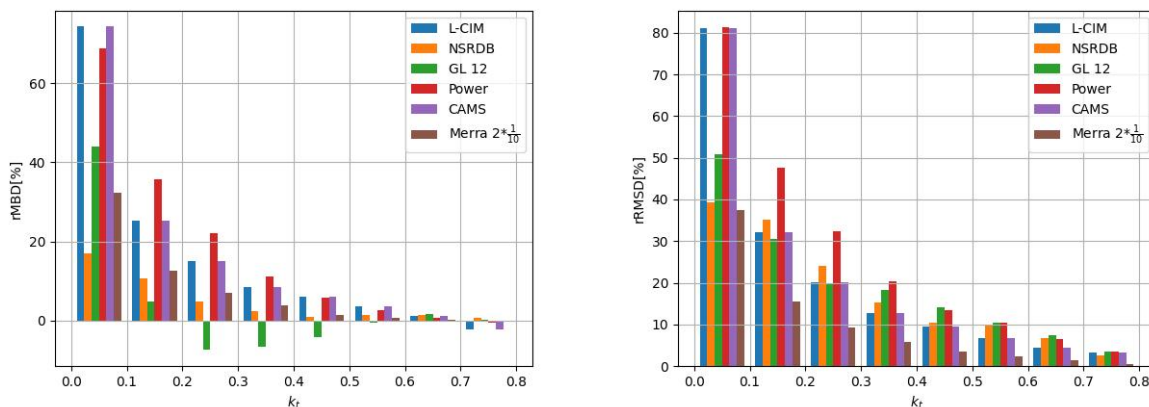


Figura 5.4: Desempeño promedio de los estimativos según índice de claridad diario  $K_t$

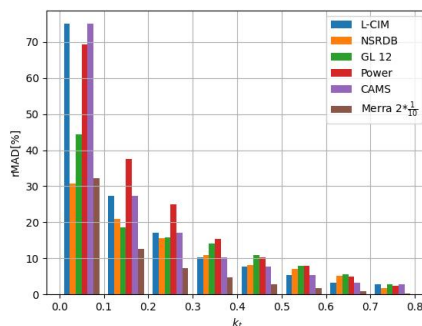


Figura 5.5: rMAD promediado espacialmente según índice de claridad diario  $K_t$

**Nota:** El desempeño de el estimador MERRA2 fue dividido entre 10 para obtener una mejor visualización del fenómeno, es decir, si la barra de MERRA2 tiene una altura de 35 % esto indica un desempeño del 350 %

Al analizar las figuras 5.4b, 5.4a y 5.5 se comienza a destacar un claro efecto de la nubosidad sobre la precisión de los estimativos. En general, se observa en 5.4b como los modelos amplifican su error relativo al aumentar su nubosidad; particularmente, en la figura 5.4a se analiza la naturaleza del error de cada uno de los modelos, en condiciones de cielo claro se observa como los modelos LCIM y CAMS subestiman la medida mientras que el resto estiman por exceso en tal condición de nubosidad, también, son estos dos los modelos que mas sobreestiman la medida en días completamente nublados ( $K_t$  muy bajos).

También, al momento de realizar este análisis es necesario observar el efecto la distribución de datos en cuanto a la nubosidad, analizando 5.3 y teniendo en cuenta la disposición de los datos en función de la nubosidad, se observa como la media de los datos se centra mas en días de poca nubosidad, osea se observa una curva creciente hasta  $K_t \approx 0,7$ , teniendo esto en cuenta es claro que el desempeño de los datos estimados en condiciones de baja nubosidad, es decir alto  $K_t$  serán los determinantes para el desempeño global del modelo.

Es de destacar también, el hecho de que los indicadores graficados son relativos, por lo que el causante de que el error baje para alto  $K_t$  puede deberse a dividir por una media mayor; pero en adición, los errores absolutos también decrecen para altos  $K_t$ , esto se ve reflejado en las gráficas en su versión no relativa, disponibles en el anexo 8.1.

### 5.3. Resultados por estación del año.

Con el fin de profundizar en el análisis, se estudia el desempeño estacional de cada uno de los modelos, cuyo resultado se presenta en las figuras 5.6 y 5.7.

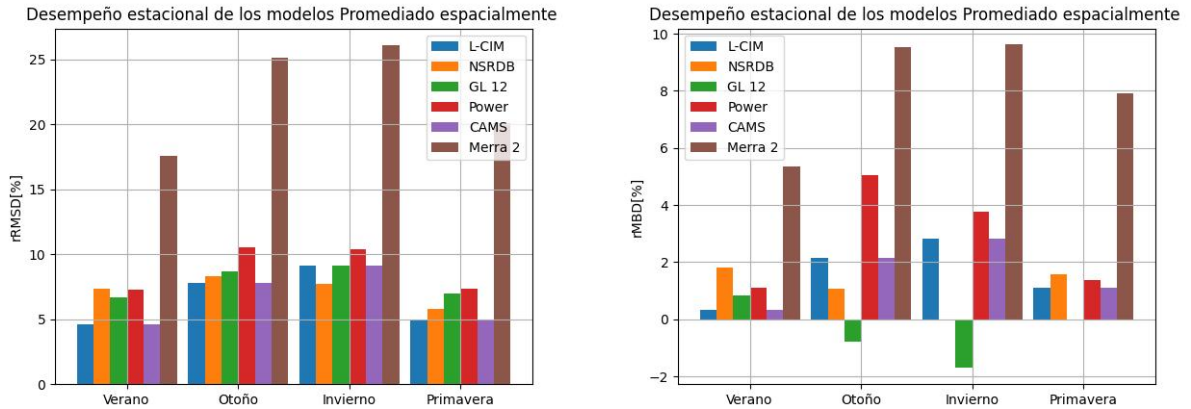


Figura 5.6: Desempeño promedio de los estimativos según estación del año.

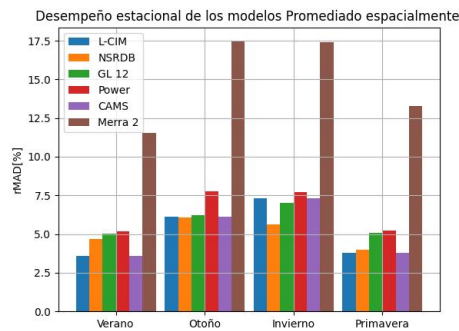


Figura 5.7: rMAD promediado espacialmente según estación del año

Como se observa en la imagen 5.6a la diferencia estacional no es alevosa, sin embargo observando el comportamiento de cada modelo (por color) se analiza una notoria diferencia ( $\approx 1\% - 4\%$ ) entre el desempeño de las estaciones verano - primavera y otoño - invierno siendo las primeras, épocas con menos nubosidad y mas energía incidente. Algo similar se observa en la imagen 5.6b, donde el sesgo, en general es mayor en otoño e invierno. Curiosamente, algunos estimativos son mas sensibles a este cambio estacional que otros; se observa como el GL1.2 por ejemplo presenta un sesgo considerablemente peor en invierno que en verano y algo similar se observa para POWER, CAMS o MERRA2 en el desvío cuadrático medio; NSRDB por otro lado no parece verse muy afectado por la estación del año.

También, es de destacar el hecho de que las gráficas analizadas corresponden a desempeños relativos, es decir, normalizados por la media de las medidas en tierra en la región temporal correspondiente por lo tanto un aumento en el error relativo no necesariamente representa un error absoluto mayor, sino que, en este caso puede verse causado por un descenso en la media de los datos; esta ultima observación indica como los modelos son poco sensibles a la magnitud de la medida, su error se comporta mas cercano a una incertidumbre que un error porcentual de medida.



## 6. Conclusiones

En el presente estudio se realizó un análisis sobre el desempeño de distintos estimativos de irradiancia a nivel diario para 7 sitios distribuidos en Uruguay. Los estimativos utilizados son de libre acceso, con excepción de GL1.2 que fue obtenido por comunicación directa con J.C. Ceballos de INPE/CPTEC.

Como resultado principal se destaca la alta precisión y consistencia espacial del modelo LCIM con un rRMSD y rMBD promedio de 6 % y 1 %, luego se posicionan modelos de buen desempeño como NSRDB y GL1.2 los cuales presentan desvíos del orden de presentado por LCIM y desvió cuadrático medio al rededor de 7.5 %, luego se sitúan Nasa-Power y Heliosat-4 con rMBD de 2 % y -2 % además de un rRMSD promedio rondando la cifra de 8.4 %, por ultimo MERRA2 presenta un rMBD de 7.5 % y un desvió cuadrático medio de 20 %. De este resultado se concluye que los modelos basados en el satélite GOES (con gran visibilidad sobre la zona) tienden a tener un mejor desempeño sobre otros como CAMS, basado en el satélite Europeo MSG; por otro lado, también se resalta importancia de la información satelital en el modelado del fenómeno en cuestión, esto se ve reflejado tanto en el desempeño de LCIM, diseñado específicamente para la pampa húmeda y entrenado con datos de la región, como también se nota la ausencia de tal información en el desempeño del modelo MERRA2.

Teniendo presente la variabilidad espacial del recurso, durante el estudio no se encontraron mayores efectos de la posición terrestre con el desempeño de los estimativos; al considerar el desempeño estacional, se observa una mayor precisión relativa de los modelos en épocas de mayor radiación (verano - otoño).

Finalmente, como resultado general de los modelos, se obtiene una correlación negativa entre el nivel de nubosidad y el desempeño de los mismos; se entiende que el efecto de la presencia de nubes es difícil de modelar con precisión por los modelos, aumentando su incertidumbre en esas condiciones.

## 7. Agradecimientos

Primeramente quiero agradecer a PEDECIBA por hacer posible esta pasantía, no únicamente por lo significativa y formativa que fue esta instancia para mi persona, sino que también por la inversión al conocimiento científico que esta oportunidad simboliza para cualquier estudiante que forme parte.

Quiero también, profundamente agradecer a los integrantes del Laboratorio de Energía Solar por la disposición a compartir información pertinente al estudio y mas allá, contribuir a mi formación tanto académica como personal orientándome en en diversas situaciones.

Gracias principalmente a mi tutor de este estudio y orientador Agustín Laguarda, quien sin prejuicios sobre mi escasa formación en el área se dispuso a presentar numerosas explicaciones y soluciones, además de en múltiples oportunidades brindar instancias de formación para mi persona.

## Referencias

- [1] Sistema de gestión de datos (sgd) del laboratorio de energía solar. <http://les.edu.uy/datos/sgd>. Accessed: 2022.
- [2] Christelle Rigollier, Olivier Bauer, and Lucien Wald. On the clear sky model of the esra—european solar radiation atlas—with respect to the heliosat method. *Solar energy*, 68(1):33–48, 2000.
- [3] A Laguarda, P Iturbide, X Orsi, MJ Denegri, S Luza, BL Burgos, V Stern, and R Alonso-Suárez. Validación de modelos satelitales heliosat-4 y cim-esra para la estimación de irradiancia solar en la pampa húmeda. *Energías Renovables y Medio Ambiente*, 48:1–9, 2021.
- [4] R Perez, Pierre Ineichen, R Seals, and A Zelenka. Making full use of the clearness index for parameterizing hourly insolation conditions. *Solar Energy*, 45(2):111–114, 1990.
- [5] Agustín Laguarda and Gonzalo Abal. Índice de turbidez de linke a partir de irradiación solar global en el uruguay. 2016.
- [6] Agustín Laguarda. Modelado de la irradiancia solar sobre la superficie terrestre: modelos físicos e híbridos utilizando información satelital sobre la pampa húmeda. 2021.
- [7] Estimaciones satelitales diarias de irradiación global -les-. <http://les.edu.uy/online/ghisat/>. Accessed: 2022.
- [8] Zhipeng Qu, Armel Oumbe, Philippe Blanc, Bella Espinar, Gerhard Gesell, BENOÏT GSCHWIND, Lars Klüser, Mireille Lefèvre, Laurent Saboret, Marion Schroedter-Homscheidt, et al. Fast radiative transfer parameterisation for assessing the surface solar irradiance: The heliosat-4 method. *Meteorologische Zeitschrift*, 26(1):33–57, 2017.
- [9] A Laguarda, G Giacosa, R Alonso-Suárez, and G Abal. Performance of the site-adapted cams database and locally adjusted cloud index models for estimating global solar horizontal irradiation over the pampa húmeda. *Solar Energy*, 199:295–307, 2020.
- [10] Joaquín Gonzalez, Vívian Teixeira-Branco, and Rodrigo Alonso-Suárez. Evaluation of the heliosat-4 and flashflux models for solar global daily irradiation estimate in uruguay. In *Proceedings of the ISES Solar World Congress 2019, Santiago de Chile, Chile, 4-7 de Noviembre, 2019*. ISES, 2019.
- [11] Copernicus atmosphere monitoring service (cams) radiation service. <https://www.soda-pro.com/web-services/radiation/cams-radiation-service>. Accessed: 2022.
- [12] William S Chandler, James M Hoell, David Westberg, Taiping Zhang, and Paul W Stackhouse Jr. Nasa prediction of worldwide energy resource high resolution meteorology data for sustainable building design. Technical report, 2013.
- [13] Power | data access viewer - prediction of worldwide energy resource. <https://power.larc.nasa.gov/data-access-viewer/>. Accessed: 2022.
- [14] Juan Carlos Ceballos, Marcus Jorge Bottino, and Jaidete Monteiro De Souza. A simplified physical model for assessing solar radiation over brazil using goes 8 visible imagery. *Journal of Geophysical Research: Atmospheres*, 109(D2), 2004.
- [15] Manajit Sengupta, Yu Xie, Anthony Lopez, Aron Habte, Galen Maclaurin, and James Shelby. The national solar radiation data base (nsrdb). *Renewable and sustainable energy reviews*, 89:51–60, 2018.
- [16] Manajit Sengupta, Aron Habte, Yu Xie, Anthony Lopez, and Christian A Gueymard. The national solar radiation data base (nsrdb) for csp applications. In *AIP Conference Proceedings*, volume 2126, page 190015. AIP Publishing LLC, 2019.
- [17] Yu Xie, Manajit Sengupta, and Jimy Dudhia. A fast all-sky radiation model for solar applications (farms): Algorithm and performance evaluation. *Solar Energy*, 135:435–445, 2016.

- [18] National solar radiation database -nsrdb-. <https://nsrdb.nrel.gov/data-viewer/download/intro>. Accessed: 2022.
- [19] Clara S Draper, Rolf H Reichle, and Randal D Koster. Assessment of merra-2 land surface energy flux estimates. *Journal of Climate*, 31(2):671–691, 2018.
- [20] Giovanni database - the bridge between data and science. <https://giovanni.gsfc.nasa.gov/giovanni/>. Accessed: 2022.
- [21] Atsumu Ohmura, Ellsworth G Dutton, Bruce Forgan, Claus Fröhlich, Hans Gilgen, Herman Hegner, Alain Heimo, Gert König-Langlo, Bruce McArthur, Guido Müller, et al. Baseline surface radiation network (bsrn/wcrp): New precision radiometry for climate research. *Bulletin of the American Meteorological Society*, 79(10):2115–2136, 1998.
- [22] Greg Kopp and Judith L Lean. A new, lower value of total solar irradiance: Evidence and climate significance. *Geophysical Research Letters*, 38(1), 2011.
- [23] Jie Zhang, Anthony Florita, Bri-Mathias Hodge, Siyuan Lu, Hendrik F Hamann, Venkat Banunayanan, and Anna M Brockway. A suite of metrics for assessing the performance of solar power forecasting. *Solar Energy*, 111:157–175, 2015.
- [24] Canavos George C. *Applied Probability and Statistical Methods*. Little, Brown and Company, first edition.
- [25] Frank J Massey Jr. The kolmogorov-smirnov test for goodness of fit. *Journal of the American statistical Association*, 46(253):68–78, 1951.

## 8. Anexo

### 8.1. Anexo: Resultados según Nubosidad

A modo de consulta se presentan las gráficas 5.4b, 5.4a y 5.5 en versión absoluta.

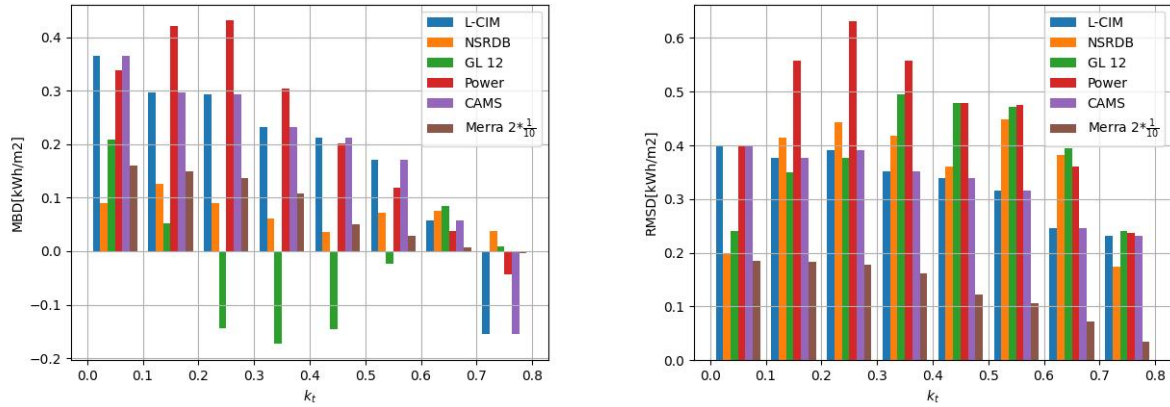


Figura 8.1: Desempeño absoluto promedio de los estimativos según índice de claridad diario  $K_t$

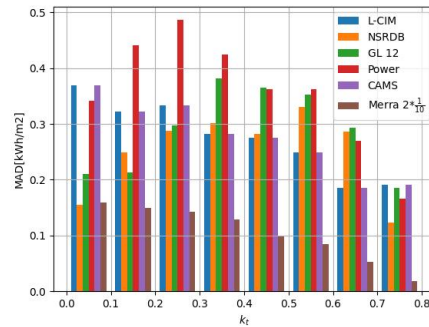
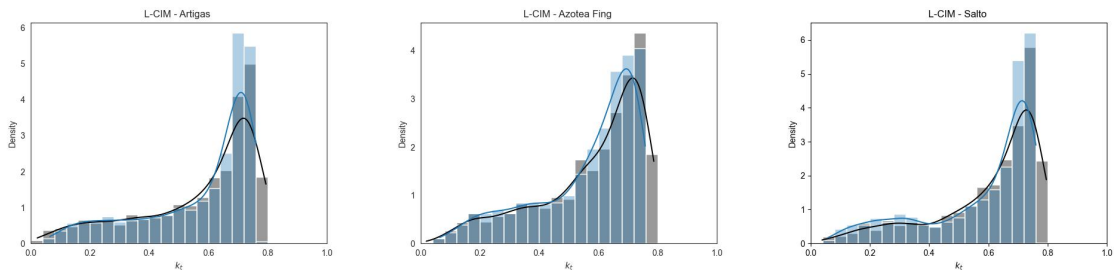


Figura 8.2: MAD promediado espacialmente según índice de claridad diario  $k_t$

### 8.2. Anexo: Distribución de nubosidad modelado por los estimativos

En la presente sección se presenta la distribución de nubosidad modelado por cada uno de los modelos, agrupados por estimativo.



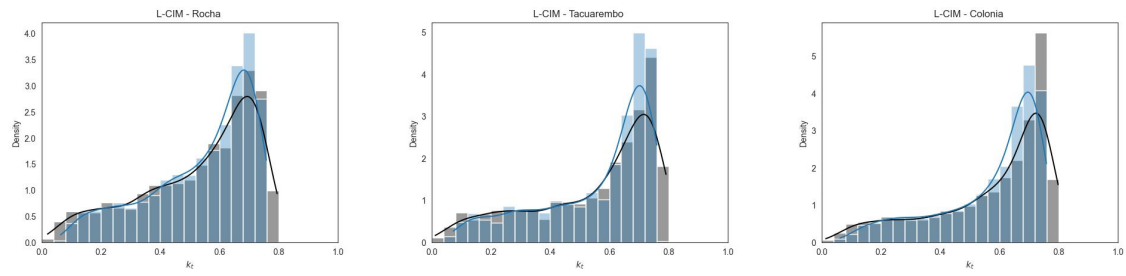


Figura 8.3: Distribución de nubosidad de LCIM .

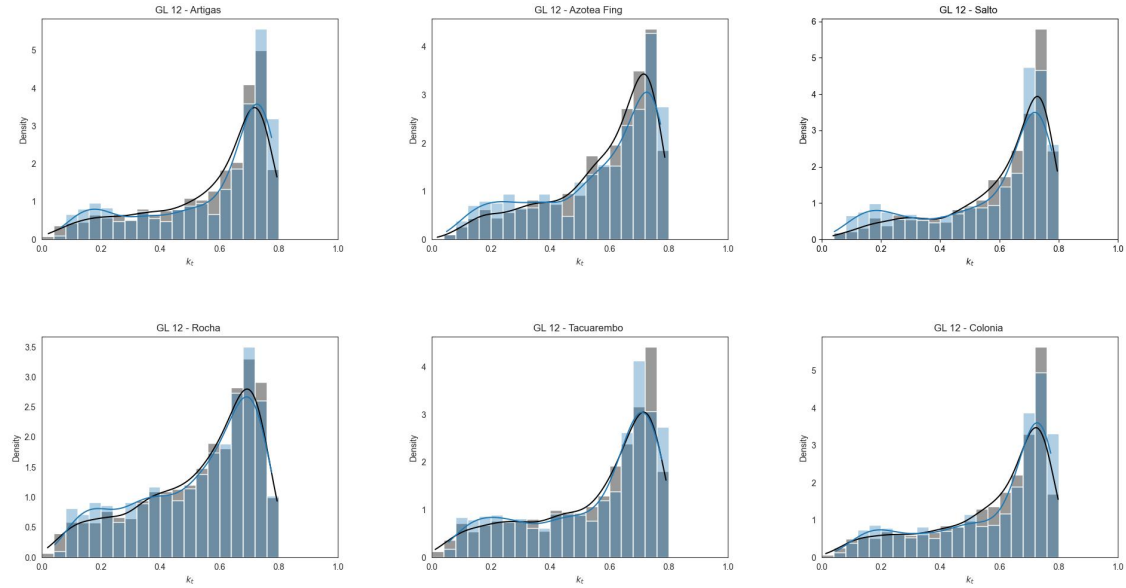


Figura 8.4: Distribución de nubosidad de GL1.2 .

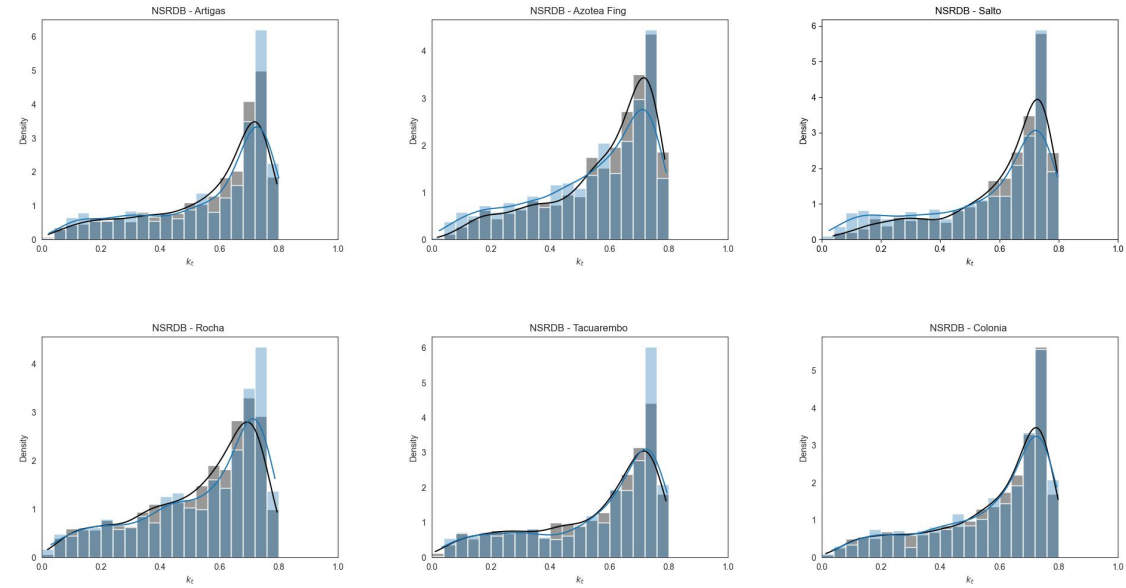


Figura 8.5: Distribución de nubosidad de NSRDB.

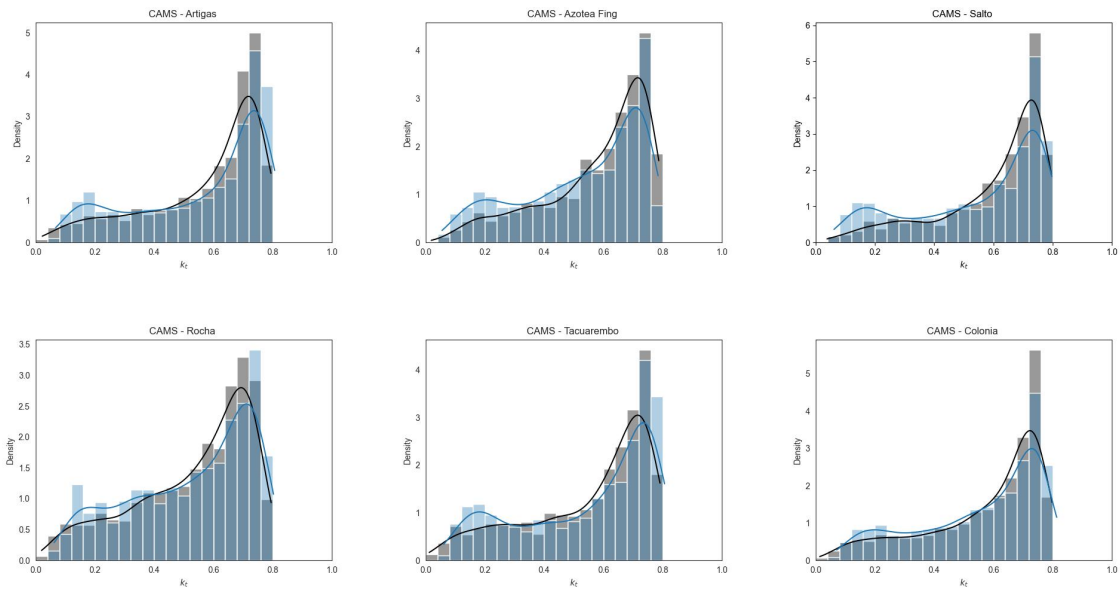


Figura 8.6: Distribución de nubosidad de Heliosat-4

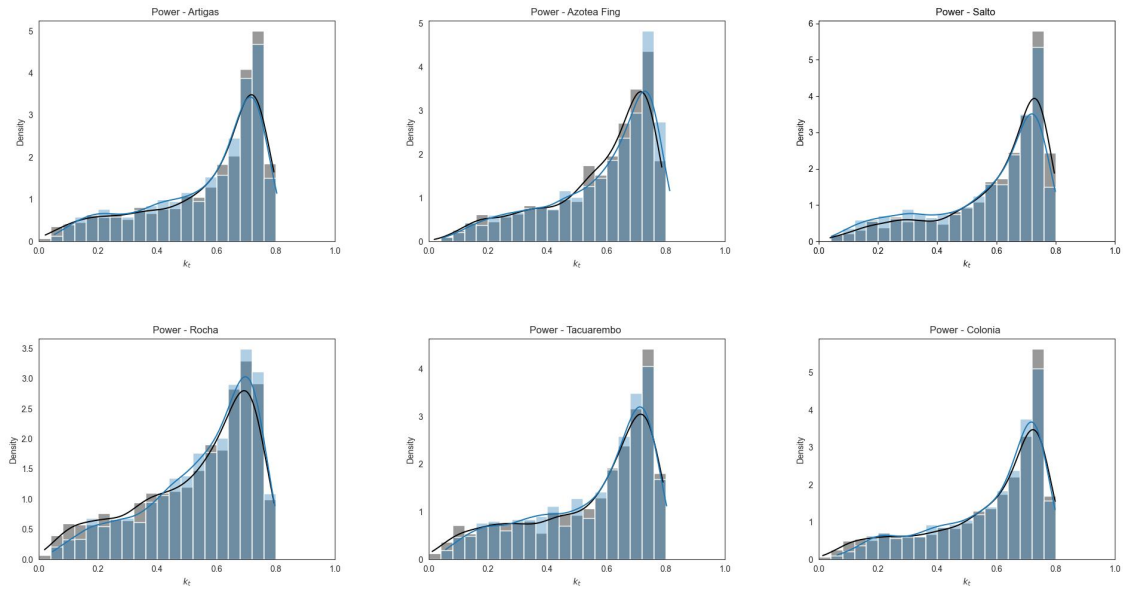
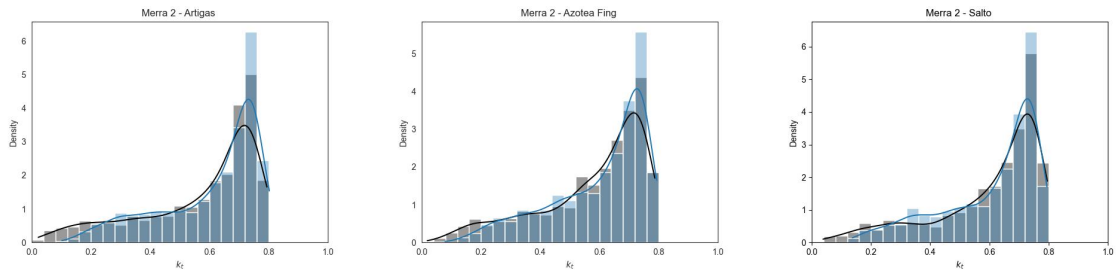


Figura 8.7: Distribución de nubosidad NASA-Power.



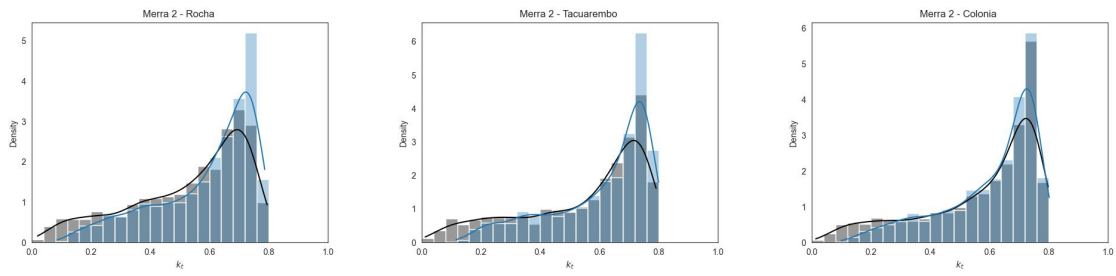


Figura 8.8: Distribución de nubosidad de MERRA2.