

Tesis Doctoral
Ingeniería Energética

**Caracterización del recurso solar para el análisis
de la producción de centrales termosolares**

Dep. Ingeniería Energética
Escuela Técnica Superior de Ingeniería
Universidad de Sevilla

2017



Tesis Doctoral
Ingeniería Energética

Caracterización del recurso solar para el análisis de la producción de centrales termosolares

Autora:

Sara Moreno Tejera

Directores:

Dr. Manuel A. Silva Pérez

Dra. Lourdes Ramírez Santigosa

Dep. de Ingeniería Energética
Escuela Técnica Superior de Ingeniería
Universidad de Sevilla

Sevilla, 2017

Agradecimientos

Agradezco a mis directores Manuel Silva y Lourdes Ramírez el tiempo que me han dedicado, que me consta que es valioso. También quiero agradecerles el haber compartido conmigo durante estos cuatro años sus conocimientos y sus ideas, y por haber considerado siempre las mías. Especialmente, me gustaría agradecerle a Manuel su ayuda y su implicación en cada uno de los trabajos que me ha dirigido desde el Proyecto fin de Carrera hasta la presente Tesis, y siempre con el mejor de los tratos.

Le agradezco a Ale su paciencia y su comprensión. Me consta que esta Tesis tú también la has sufrido. Gracias por entenderlo y por animarme a pesar de todo.

A mi familia le agradezco su apoyo incondicional, y en especial a mi madre, sus infinitos favores multiplican mi tiempo.

No me puedo olvidar de mis amigos trajaneros, siempre dispuestos a alternar el deporte con las cervezas, el complemento perfecto para poder aclarar las ideas.

Y por supuesto, a mis compañeros del GTER los antiguos y los actuales. Especialmente a Nativi, Elena y Miguel, que han estado conmigo en estos últimos años. Investigar con gente entusiasta, inteligente y divertida hace que el trabajo sea casi un hobby.

Sara Moreno Tejera

Sevilla, 2017

Lista original de publicaciones

Esta Tesis doctoral se presenta por compendio de publicaciones de acuerdo con el artículo 9 de la normativa reguladora del régimen de tesis doctoral (Acuerdo 9.1/CG 19-4-12) en el marco del programa regulado por el RD 1393/2007.

La tesis la componen 3 artículos publicados (anexos A, B y C):

ANEXO A:

Moreno-Tejera, S., Ramírez-Santigosa, L., Silva-Pérez, M. a., 2015. **A proposed methodology for quick assessment of timestamp and quality control results of solar radiation data.** *Renew. Energy* 78, 531–537. doi:10.1016/j.renene.2015.01.031

ANEXO B:

Moreno-Tejera, S., Silva-Pérez, M.A., Lillo-Bravo, I., Ramírez-Santigosa, L., 2016. **Solar resource assessment in Seville, Spain. Statistical characterisation of solar radiation at different time resolutions.** *Sol. Energy* 132, 430–441. doi:10.1016/j.solener.2016.03.032

ANEXO C:

Moreno-Tejera, S., Silva-Pérez, M.A., Ramírez-Santigosa, L., Lillo-Bravo, I., 2017. **Classification of days according to DNI profiles using clustering techniques.** *Sol. Energy* 146, 319–333. doi:10.1016/j.solener.2017.02.031

Un cuarto artículo enviado a la revista *Solar Energy* se encuentra, en el momento de presentar este documento, en fase de revisión y se incluye como Anexo D:

ANEXO D:

Moreno-Tejera, S., Silva-Pérez, M.A., Ramírez-Santigosa, L., Lillo-Bravo, I., 2017. **The clear sky yield index: a new indicator for qualifying the solar resource for CSP plants.** (Enviado a *Solar Energy* el 07/06/2017, N^o manuscrito SE-D-17-01162)

Otros trabajos relacionados con la caracterización del recurso solar para las centrales termosolares desarrollados por la autora de esta Tesis:

- **Publicaciones en revistas indexadas:**

Larrañeta, M., Moreno-Tejera, S., Lillo-Bravo, I., Silva-Pérez, M.A., 2017. A methodology for the stochastic generation of hourly synthetic direct normal irradiation time series. *Theoretical and Applied Climatology*, 1-11. doi:10.1007/s00704-017-2188-4

Lillo, I., Pérez, E., Moreno, S., Silva, M., 2017. Process Heat Generation Potential from Solar Concentration Technologies in Latin America: The Case of Argentina. *Energies* 10, 383. doi:10.3390/en10030383

Fernández Peruchena, C.M., Ramírez, L., Silva-Pérez, M.A., Lara, V., Bermejo, D., Gastón, M., Moreno-Tejera, S., Pulgar, J., Liria, J., Macías, S., Gonzalez, R., Bernardos, A., Castillo, N., Bolinaga, B., Valenzuela, R.X., Zorzalejo, L.F., 2016. A statistical characterization of the long-term solar resource: Towards risk assessment for solar power projects. *Sol. Energy* 123, 29–39. doi:10.1016/j.solener.2015.10.051

Larrañeta, M., Moreno-Tejera, S., Silva-Pérez, M.A., Lillo-Bravo, I., 2015. An improved model for the synthetic generation of high temporal resolution direct normal irradiation time series. *Sol. Energy* 122, 517–528. doi:10.1016/j.solener.2015.09.030

Silva-Perez, M.A., Barea-Garcia, J.M., Larrañeta, M., Moreno-Tejera, S., Lillo-Bravo, I., 2014. Analysis of the distribution of measured and synthetic DNI databases and its effect on the expected production of a parabolic trough plant. *Energy Procedia* 49, 2512–2520. doi:10.1016/j.egypro.2014.03.266

Moreno-Tejera, S., Pérez-Aparicio, E., Barea-García, J.M., Lillo-Bravo, I., Silva-Pérez, M.A., 2014. Assessment of a global-to-direct empirical model for the long-term characterization of Direct Normal Insolation. *Energy Procedia* 49, 2377–2386. doi:10.1016/j.egypro.2014.03.252

Pagola, I., Gastón, M., Fernández-Peruchena, C., Moreno, S., Ramírez, L., 2010. New methodology of solar radiation evaluation using free access databases in specific locations. *Renew. Energy* 35, 2792–2798. doi:10.1016/j.renene.2010.04.034

- **Publicaciones en actas de congresos:**

Fernández-Peruchena, C.M., Lara-Faneho, V., Ramírez, L., Zorzalejo, L.F., Silva, M.,

Bermejo, D., Gastón, M., Moreno, S., Pulgar, J., Pavon, M., Macías, S., Valenzuela, R.X., 2017. A Methodology for Probabilistic Assessment of Solar Thermal Power Plants Yield. AIP Conf. Proc. 1850. doi:10.1063/1.4984514

Lara-Fanego, V., Pulgar, J., Fernández Peruchena, C.M., Gastón, M., Moreno, S., Ramírez, L., Valenzuela, R.X., Zarzalejo, L.F., Bermejo, D., Silva, M., Pavón, M., Bernardos, A., Macías, S., 2017. A novel procedure for generating solar irradiance TSYs. AIP Conf. Proc. 1850. doi:10.1063/1.4984523

Larrañeta, M., Moreno-Tejera, S., Lillo-Bravo, I., Silva-Pérez, M.A., 2017. Cloud transient characterization in different time steps. AIP Conf. Proc. 1850. doi:10.1063/1.4984524

Pavón, M., Fernández, C.M., Silva, M., Moreno, S., Guisado, M. V, Bernardos, A., 2017. Statistical Analysis of CSP Plants by Simulating Extensive Meteorological Series. AIP Conf. Proc. 1850. doi:10.1063/1.4984554

Pérez-Aparicio, E., Lillo-Bravo, I., Moreno-Tejera, S., Silva-Pérez, M., 2017. Economical and environmental analysis of thermal and photovoltaic solar energy as source of heat for industrial processes. AIP Conf. Proc. 1850. doi:10.1063/1.4984572

Peruchena, C.M.F., Ramírez, L., Silva, M., Lara, V., Bermejo, D., Gastón, M., Moreno, S., Pulgar, J., Liria, J., Macías, S., Gonzalez, R., Bernardos, A., Castillo, N., Bolinaga, B., Valenzuela, R.X., Zarzalejo, L., 2016. A methodology for calculating percentile values of annual direct normal solar irradiation series. AIP Conf. Proc. 1734. doi:10.1063/1.4949237

Moreno, S., Pérez, E., Larrañeta, M., Silva, M. a., 2011. Influence of data type in the development of TMY. SolarPaces Conf. Granada, Spain.

Moreno, S., Ramirez, L., Pérez, E., Bernardos, A., Silva, M., 2010. Long term estimation of DNI from solar global irradiation. A model to estimate DNI from monthly clearness index. SolarPaces Conf. Perpignan, France.

Barea, J.M., Moreno, S., Silva, M. a., 2010. Analysis of the influence of the monthly distribution of Direct Normal radiation in the production of Parabolic Trough plants using EOS. SolarPaces Conf. Perpignan, France.

Moreno, S., Silva, M., 2009. Comparison of methodologies to estimate direct normal irradiation from daily values of global horizontal irradiation. SolarPaces Conf. Berlin, Germany.

- **Otros trabajos:**

UNE 206013:2017, Centrales termosolares. Procedimiento de generación de años percentiles de radiación solar.

UNE 206011:2014, Centrales termosolares. Procedimiento de generación de Año Solar Representativo.

Resumen

En la presente Tesis se avanza en el estudio de las variables que deben tenerse en cuenta en la caracterización del recurso solar de un emplazamiento orientado a proyectos de centrales termosolares de concentración, y se presentan nuevas herramientas y procedimientos que facilitan su análisis.

El estudio se ha realizado empleando una extensa base de datos de medidas radiométricas registradas en Sevilla, España. Esta base de datos ha sido revisada y etiquetada, aplicando procedimientos de control de calidad comúnmente conocidos. Para interpretar el etiquetado, localizar los errores y tratar de identificar el origen de los mismos, se presenta una nueva herramienta visual, compuesta por dos tipos de gráficas. Esta herramienta se basa principalmente en la representación de las etiquetas resultantes de aplicar el control de calidad para cada variable abarcando cualquier periodo de tiempo en una misma gráfica, independientemente de la frecuencia de registro de los datos. Con esta herramienta se facilita la identificación de errores que no pueden asociarse a la incertidumbre de las medidas y que, por tanto, deben eliminarse. Posteriormente, se ha corregido y completado la base de datos aplicando un procedimiento que propone una solución específica para cada variable radiométrica en función del instante del día en el que se produce el error, el tiempo de persistencia del mismo y el tipo de día en el que tiene lugar (claro, nuboso, cubierto).

Una vez corregida y validada, esta extensa base de datos de valores registrados en su origen cada 5 s se analiza en diferentes escalas temporales. Teniendo en cuenta las escalas en las que la radiación solar se utiliza para evaluar la producción de las diferentes tecnologías solares, se caracterizan estadísticamente las series de valores instantáneos, diez-minutales, horarios, diarios, mensuales y anuales. El análisis de la base de datos de Sevilla en las menores escalas temporales muestra que las distribuciones estadísticas de los valores instantáneos y diez-minutales son semejantes entre sí, pero muy distintas de las de los valores horarios. Este resultado debe tenerse en cuenta a la hora de seleccionar la escala temporal adecuada en la simulación de centrales termosolares. El análisis de los valores diarios, mensuales y anuales identifica características del recurso útiles para el desarrollo de un proyecto

de una central termosolar.

Para completar la caracterización del recurso, se presenta una metodología para clasificar los días en función del estado del cielo a partir de las propiedades de la radiación que afectan a la operación y a la producción de las centrales termosolares. Esta metodología utiliza tres índices que representan tres propiedades de las curvas diarias de radiación directa normal (DNI): la variabilidad, la energía o radiación diaria acumulada y la distribución en el tiempo de esta energía. En la metodología, estas características son representadas a partir de índices. El número y las características de cada tipo de día se obtienen aplicando técnicas de *clustering* sin supervisión. Dos grupos diferentes de índices ($\{k_v, VI, F_m\}$ y $\{k_v, POP_D, F_m\}$) son evaluados mediante métodos de validación interna. Como resultado, el número óptimo de grupos obtenidos es 10 en ambos casos, aunque el grupo que incluye el índice VI obtiene una incertidumbre menor a la hora de encontrar el óptimo. Para validar los resultados obtenidos del *clustering*, se utilizan dos años de DNI clasificados por 5 expertos a partir de la visualización de sus curvas como la *clasificación de referencia*. De la comparativa entre las clasificaciones se concluye que el grupo VI tiene un mayor grado de coincidencia.

Finalmente, se caracteriza la representatividad de los tipos de días identificados desde el punto de vista de la producción de una central termosolar de captadores cilindroparábolicos (CCP) con y sin almacenamiento a partir de un nuevo índice, el índice de producción relativa de cielo despejado. Este índice adimensionaliza los efectos estacionales de la producción permitiendo evaluar las características del recurso solar desde la perspectiva de la producción y comparar el nivel de aprovechamiento del recurso entre centrales con diferente configuración. Previamente a la caracterización de los tipos de días, se analiza la relación entre los índices de clasificación y el índice de producción relativa de ambas centrales. La variabilidad y la distribución en el tiempo de las curvas de DNI tienen una alta influencia en las centrales de tecnología CCP sin almacenamiento. Esta influencia disminuye considerablemente cuando la central cuenta con un sistema de almacenamiento térmico. De la caracterización del índice de producción relativa de cielo despejado de los tipos de días se concluye que el número de tipos de días respecto a la producción de las centrales puede optimizarse. Para el emplazamiento analizado, los 10 tipos de días identificados en la metodología de clasificación a partir de las propiedades de las curvas de DNI se reducen a 7 tipos de días desde el punto de vista de la producción de una central CCP sin almacenamiento, y a 6 tipos para

una central con 7.5 h de almacenamiento térmico.

La relación entre los tipos de días y el índice de producción relativa al cielo despejado tiene una aplicación directa para la definición de la estrategia de operación de una central y la predicción de la energía generada a partir de la predicción cualitativa del tipo de día.

Abstract

This doctoral thesis presents advances in the study of the relevant characteristics of the solar resource from the standpoint of concentrated solar power (CSP) projects, as well as new tools and procedures to facilitate its analysis.

The study was carried out using an extensive database of radiometric measurements registered in Seville, Spain. This database has been checked and labeled using commonly known quality control procedures with the help of a new visual tool that facilitates the interpretation of the labels, the localization of the errors and the identification of their origin. This tool displays the data labels of each variable covering any period by means of a color-code in a single graph, regardless of the frequency of the data, helping to identify the gaps and other errors that can not be associated with the uncertainty of the measures. Then, the database has been corrected and filled by applying a new procedure that proposes a specific solution for each radiometric variable as a function of the time of the day in which the error occurs, the period of persistence of the error and the type of day according to the state of the sky (clear, cloudy, covered).

Once corrected and validated, this extensive database, recorded in its origin every 5 s, is analyzed on different time scales. Considering the time-steps in which solar radiation is used to evaluate the production of solar technologies, the instantaneous, ten-minute, hourly, daily, monthly and annual values are characterized statistically. The analysis of the Seville database on the smallest time-steps shows that the statistical distributions of the instantaneous and ten-minute values are similar, but very different from the distribution of the hourly values. This result should be considered when selecting the appropriate time scale to simulate the behaviour of solar thermal power plants. The analysis of the daily, monthly and annual values reveal solar resource characteristics relevant for the development of a CSP project.

To complete the characterization of the resource, a methodology to classify the days based on the solar radiation features that affect the operation and the production of the solar thermal power plants is presented. This methodology uses three indexes that represent three features of daily direct normal radiation (DNI) curves: energy,

variability and time distribution of the DNI. These characteristics are represented by daily dimensionless indexes. The number of clusters or day types and the characteristics of each day type are obtained by applying unsupervised clustering techniques. Two different sets of indexes ($\{k_b, VI, F_m\}$ and $\{k_b, POP_D, F_m\}$) are evaluated by internal validation methods. The optimal number of groups obtained is 10 with both sets, although the first set (VI set, which includes the VI index) shows a lower uncertainty finding the optimum. The visual classification of 2 years of daily DNI curves carried out by 5 experts is used as a reference to validate the results. The VI set shows a better coincidence with the reference classification.

Finally, the relation between the day types and the electricity yield of CSP parabolic trough plants (PTP) is studied by means of a new index, the clear sky yield index. This index allows the evaluation of the daily characteristics of the solar resource from the production perspective removing seasonal effects and is also helpful to compare the exploitation of the radiation of plants with different configurations for a specific period. Prior to the characterization of the day types, the relation between the classification indexes and the clear sky yield index is analyzed for two reference plants, one of them with a large Thermal Energy Storage (TES) system and the other with no TES. The variability and time distribution of DNI curves have a higher influence on the PTP without thermal storage. This influence decreases considerably when the plant has a thermal storage system. The characterization of the clear sky yield index for each day type shows that the number of day types with respect to the production of the plants can be optimized. For the analyzed location, the 10 day types identified with the classification methodology are reduced to 7 types of days from the point of view of the production of a PTP without storage, and to 6 types for a plant with 7.5 h of thermal storage.

The relation between day types and clear sky yield index has a direct application to the definition of operating strategies and forecasting of the energy generation based on the qualitative prediction of the type of day.

Índice

Agradecimientos	5
Lista original de publicaciones	7
Resumen	11
Abstract	15
Índice	17
Índice de Figuras	19
Índice de Tablas	25
Notación	27
1 Introducción	29
2. Objetivos	35
3. Estructura del documento	37
4. Metodología	39
4.1. <i>Base de datos</i>	39
4.1.1. Control de calidad de datos de radiación solar	40
4.1.2. Rellenado de huecos	42
4.2. <i>Caracterización estadística de la radiación solar en Sevilla</i>	44
4.2.1. Caracterización estadística de los valores instantáneos, horarios y 10-min.	45
4.2.2. Caracterización estadística de los valores diarios	46
4.2.3. Caracterización estadística de los valores mensuales y anuales	47
4.3. <i>Clasificación de días en función de las curvas de DNI</i>	47
4.3.1. Modelo de cielo despejado envolvente	48
4.3.2. Índices de clasificación	50
4.3.3. Procedimiento de clasificación	56
4.4. <i>Caracterización del tipo de día en función de la producción</i>	63
4.4.1. Definición del índice de producción relativa de cielo despejado.	63

4.4.2.	Datos de entrada	64
4.4.3.	Modelo de centrales y simulación	64
4.4.4.	Evaluación del índice de producción relativa de cielo despejado	66
4.4.5.	Caracterización del índice de producción relativa de cielo despejado de los tipos de días	67
5.	Resultados	69
5.1.	<i>Base de datos</i>	69
5.1.1.	Control de calidad	69
5.1.2.	Rellenado de huecos	70
5.2.	<i>Caracterización estadística de la radiación solar en Sevilla</i>	72
5.2.1.	Caracterización estadística de los valores instantáneos, horarios y diez-minutal.	72
1.1.1	Caracterización estadística de los valores diarios	77
5.2.2.	Caracterización estadística de los valores mensuales y anuales	78
5.3.	<i>Clasificación de días en función de las curvas de DNI</i>	80
5.3.1.	Número y tipos de días en función de los índices	80
5.3.2.	Validación de resultados	87
5.4.	<i>Caracterización de la producción en función del tipo de día</i>	92
5.4.1.	Evaluación del índice de producción relativa de cielo despejado diario	92
5.4.2.	El índice de producción relativa de cielo despejado frente a los índices de clasificación	94
1.1.2	Caracterización del índice de producción relativa de cielo despejado de cada tipo de día	98
6.	Discusión	107
7.	Conclusiones	113
	Referencias	117
	Anexo A	A
	Anexo B	B
	Anexo C	C
	Anexo D	D

Índice de Figuras

Figura 4-1. Representación de los resultados del control de calidad basado en los procedimientos de la BSRN.....	41
Figura 4-2.. A la izquierda, valores máximos de I_{b0}/ε_0 frente a la altura solar y el modelo de ASHRAE ajustado. A la derecha, modelo de ASHRAE envolvente máximo para los 365 días del año.	50
Figura 4-3. Representación de valores instantáneos del índice de transmitancia k_b , del índice de fracción de directa k_{b0} , de las medidas registradas de DNI y de los valores correspondientes del modelo envolvente máximo de cielo despejado de ASHRAE para un día en su mayoría despejado.....	52
Figura 4-4. Curvas teóricas de cielo despejado de DNI con valores instantáneos de k_b desde 0.1 hasta 1.....	54
Figura 4-5. Representación de los valores de DNI registrados durante 4 días en Sevilla durante el 2010 junto con la envolvente de cielo despejado y sus correspondientes valores de k_b , VI, POP_D y F_m	55
Figura 4-6. Representación de los valores del POP_D frente a los valores diarios del VI para todos los días del periodo 2000-12.....	57
Figura 4-7.. Representación de los dos grupos de índices calculados a partir de los 13 años de medidas de DNI utilizadas para clasificar los días en función del estado del cielo: k_b , VI y F_t (izquierda) y k_b , POP_D y F_t (derecha).....	59
Figura 4-8. . Ejemplos de las curvas correspondientes a los tipos de días definidos en la Tabla 2. Estas figuras son las utilizadas por los expertos como referencia para realizar la clasificación visual.....	62
Figura 4-9. DNI y potencia eléctrica bruta producida por la central CCP PTP7.5 el 21 y 22 de junio en condiciones de cielo despejado envolvente máximas.....	66
Figura 4-10. Descripción esquemática del procedimiento seguido para la evaluación del k_p diario: Primero se simula el comportamiento de un central CCP con y sin	

almacenamiento térmico durante 16 años a partir de las medidas de DNI diez-minutal de la base de datos del GTER y un año diez-minutal de valores de DNI en condiciones de cielo despejado (DNIcs). A continuación, se calculan, se analizan y se comparan los k_p diarios de ambas centrales durante dicho periodo. 67

Figura 4-11. Descripción esquemática del procedimiento seguido para la caracterización de los tipos de días desde el punto de vista de la producción de las centrales: Primero, se estudia la relación entre k_p y k_b , VI y F_m . Estos índices son los utilizados para representar las características de las curvas de la DNI (energía, variabilidad y distribución temporal) en la metodología de clasificación. A continuación, se caracterizan los valores del índice k_p de cada uno de los tipos de días resultantes de aplicar la metodología de clasificación de acuerdo a estas tres características de la DNI. 68

Figura 5-1. Resultados del control de calidad del año 2001 de la base de datos del GTER. La figura más a la izquierda representa los valores medidos de GHI frente a los calculados a partir de sus componentes. La figura central y la situada a su derecha representan las etiquetas de los procedimientos del control de calidad de la BSRN de las medidas de GNI y DNI. Las líneas del orto el ocaso y el medio día solar (en color azul) ayudan a analizar los resultados. 70

Figura 5-2. Función de densidad de probabilidad de los valores de k_t calculados a partir de la base de datos antes ("with gaps") y después ("filled") de aplicar el procedimiento de rellenado de huecos..... 71

Figura 5-3. Función de densidad de probabilidad de los valores de k_{t0} calculados a partir de la base de datos antes ("with gaps") y después ("filled") de aplicar el procedimiento de rellenado de huecos..... 71

Figura 5-4. Funciones de densidad de probabilidad de los valores instantáneos de k_t y k_{t0} para diferentes rangos de masa óptica relativa del aire. 73

Figura 5-5. Función de densidad de probabilidad de los índices k_t (izquierda) y k_{t0} (derecha) obtenidos a partir de los valores de GHI y DNI cada 5s, promediados cada 10 min y cada hora..... 75

Figura 5-6. Función de densidad de probabilidad de los valores 10 min y horarios de GHI y DNI. 76

Figura 5-7. Valores medios horarios de GHI (izquierda) y DNI (derecha) para cada

día del año.....	76
Figura 5-8. Función de densidad de probabilidad de k^4 y número de tipos de día promedios mensuales de acuerdo al estado del cielo (despejado, cubierto, parcialmente cubierto).....	78
Figura 5-9. Representación de los valores máximos, mínimos y medianos de SH y CH de las 1000 iteraciones en función al número de clúster, para los dos grupos de índices considerados.....	81
Figura 5-10. A la izquierda, representación de todos los posibles pares de rectas de ajuste de los valores máximos de SH y CH para los dos grupos de índices considerados para buscar el “knee point”, o número de clúster más óptimo al aplicar el L-method. A la derecha, se representa el valor de RMSE _c respecto al número de clústeres.....	82
Figura 5-11. Representación de los valores del POP _D (a la izquierda) y de VI (a la derecha) frente a k_b y F_t del periodo 2000-13. Los diferentes símbolos identifican los diferentes clústeres encontrados en cada caso.....	83
Figura 5-12. Curvas de DNI de los días correspondientes a los medoides de los clústeres del grupo de índices POP _D . Las etiquetas coinciden con las etiquetas de los clústeres de la Figura 22. Las figuras de la columna izquierda tienen distribución de energía concentrada principalmente por la mañana. Las de la columna central, energía homogénea a lo largo del día. Las de la columna derecha, energía concentrada principalmente por la tarde. La distribución de las figuras por filas está también relacionada con características similares de energía y variabilidad intradiaria de la DNI.....	85
Figura 5-13. Curvas de DNI de los días correspondientes a los medoides de los clústeres del grupo de índices VI. Las etiquetas coinciden con las etiquetas de los clústeres de la Figura 22. Las figuras de la columna izquierda tienen distribución de energía concentrada principalmente por la mañana. Las de la columna central, energía homogénea a lo largo del día. Las de la columna derecha, energía concentrada principalmente por la tarde. La distribución de las figuras por filas está también relacionada con características similares de energía y variabilidad intradiaria de la DNI.....	86
Figura 5-14. Porcentaje de días que han sido clasificados en el mismo tipo de día por 3 o más expertos.....	87

Figura 5-15. A la izquierda, se representa VI frente a k_b para el periodo 2000-12 y F_t igual a 1 (primera figura), 0.5 (segunda figura) y 0 (última figura). Las líneas de puntos definen los límites de cada clúster encontrado y cada medoide correspondiente es simbolizado con "X". A la derecha, se representan los valores de VI frente a k_b para F_t igual a 1 (primera figura), 0.5 (segunda figura) y 0 (última figura) de la serie de referencia elaborada a partir de la clasificación realizada por los expertos. En estas figuras, se incluyen las líneas que definen cada tipo de clúster.. 90

Figura 5-16. Representación de las curvas de DNI correspondientes a los medoides de los clústeres obtenidos finalmente como resultado de aplicar la metodología propuesta..... 91

Figura 5-17. Valores diarios de k_p y su correspondiente valor mediano para cada día del año (línea roja) de PTP0 (izquierda) y PTP7.5 (derecha) durante los 16 años de medidas considerados. 92

Figura 5-18. Representación de los valores de k_p de PTP7.5 frente a los de PTP0 durante el periodo 2000-15 (izquierda). Distribución de frecuencia de k_p para PTP0 (en azul) y para PTP7.5 (en rojo) durante el periodo 2000-15 (derecha)..... 93

Figura 5-19. Representación de los valores de k_p de PTP0 (izquierda) y de PTP7.5 (derecha) frente a los valores de k_b durante el periodo 2000-15..... 94

Figura 5-20. Distribución de frecuencia de la ratio k_p/k_b de PTP0 (azul) y de PTP7.5 (rojo) para el periodo 2000-2015..... 95

Figura 5-21. Representación de los valores de k_p de PTP0 (izquierda) y de PTP7.5 (derecha) frente a los valores normalizados de VI durante el periodo 2000-15..... 96

Figura 5-22. Representación de los valores de k_p de PTP0 (izquierda) y de PTP7.5 (derecha) frente a F_m de la serie de medidas de DNI y su serie simétrica. Las líneas de puntos verticales indican los valores límite de F_m seleccionados en la metodología de clasificación para definir los días de tarde ($F_m \leq 0.3$), los días homogéneos ($0.3 < F_m < 0.7$) y los días de mañana ($F_m \geq 0.7$)..... 97

Figura 5-23. Fila superior: valores de k_p de los días de tarde frente a los valores de los días de mañana para la serie medida y la serie simétrica de PTP0 (izquierda) y PTP7.5 (derecha); fila inferior: CDF de los valores de k_p de los días de tarde (línea violeta) y los días de mañana (línea verde) para ambas series de PTP0 (izquierda) y PTP7.5 (derecha). 98

Figura 5-24. Gráficas “Box and Whisker” de los valores k_p de cada tipo de día de PTP0 (izquierda) y PTP7.5 (derecha) para el periodo 2000-15. El percentil 25 y 75 definen las cajas de cada tipo de día y el 99.3% de los valores se encuentran entre los límites de los “whiskers”. Las líneas rojas horizontales situadas en cada caja, corresponden al valor mediano de cada serie. Las líneas de puntos verticales que se muestran en ambas figuras separan los días clasificados con energía alta, media, baja y nula. El porcentaje de días de la serie que pertenece a cada tipo de día se muestra bajo la etiqueta de cada uno de ellos..... 99

Figura 5-25. Histogramas de los valores de k_p para los tipos de días o clústeres a partir de la simulación de PTP0 (azul) y PTP7.5 (rojo) durante los 16 años de medidas. 102

Índice de Tablas

Tabla 4–1. Procedimiento de rellenado de huecos y corrección de datos erróneos.	43
Tabla 4–2. Tipos de días utilizados como referencia por los expertos para realizar la clasificación del periodo 2010-11. Los días han sido definidos de acuerdo a las características consideradas a través de los índices para aplicar el algoritmo de agrupación.	61
Tabla 4–3. Principales parámetros técnicos utilizados en SAM para modelar las centrales semejantes a Solnova1 (PTP0) y Andasol3 (PTP7.5).	65
Tabla 5–1. Valores mensuales y anuales medios diarios de GHI, DNI y k_p en Sevilla.	79
Tabla 5–2. Matriz de confusión de los resultados de la clasificación obtenida con el procedimiento basado en las técnicas de clustering frente a la clasificación de referencia obtenida a partir de la clasificación visual realizada por los expertos.	88
Tabla 5–3. Valores medios y medianos de k_p para cada tipo de día de los 16 años simulados de PTP7.5 y PTP0. En la última fila se muestran los porcentajes de días de cada clúster durante el periodo. Las líneas de puntos separan los tipos de días con energía alta (H), media (M), baja (L) y nula (N).	101
Tabla 5–4. Resultados del K-S test entre las distribuciones de los valores de k_p de cada tipo de día obtenido con la simulación de PTP0 durante el periodo 2000-15.	104
Tabla 5–5. Resultados del K-S test entre las distribuciones de los valores de k_p de cada tipo de día obtenido con la simulación de PTP7.5 durante el periodo 2000-15.	104
Tabla 5–6. Análisis cualitativo de los valores de k_p de cada clúster y central CCP. k_p toma valores insignificantes en los clúster con energía nula y baja con la excepción del clúster ML (k_p en rojo). El clúster BH muestra los mayores valores de k_p (k_p en verde) para ambas centrales.	105

Notación

$H^{d_{bn}}$	Irradiación directa normal diaria
$H^{d_{bn,m}}$	Irradiación directa normal registrada durante la primera mitad del día
$H^{d_{cs}}$	Irradiación directa normal diaria en condiciones de cielo despejado
I_{g0}	Irradiancia global horizontal
I_{bn}	Irradiancia directa normal
I_0	Irradiancia extraterrestre horizontal
I_n	Irradiancia extraterrestre normal
I_{cs}	Irradiancia directa normal en condiciones de cielo despejado
I_{sc}	Constante solar
ϵ_0	Factor de corrección de la distancia Tierra-Sol
θ_z	Ángulo cenital
α	Ángulo de altura solar
k_t	Índice de claridad
k_{t0}	Índice de fracción de directa
$k^{\Delta t_t}$	Índice de claridad para el intervalo de agregación Δt
$k^{\Delta t_{t0}}$	Índice de fracción de directa para el intervalo de agregación

	Δt
k_b	Índice de transmitancia
$k^{\Delta t}_b$	Índice de transmitancia para el intervalo de agregación Δt
VI	Índice de variabilidad
F_t	Fracción de tiempo
F_m	Fracción de mañana
F_a	Fracción de tarde
$k^{\Delta t}_p$	Fracción de producción de cielo despejado durante el periodo Δt
k^d_p	Fracción de producción de cielo despejado diaria
$P^{\Delta t}$	energía eléctrica producida por la central durante el periodo Δt
$P^{\Delta t}_{CS}$	energía eléctrica producida por la central durante el periodo Δt en condiciones de cielo despejado

1 INTRODUCCIÓN

Aunque la radiación solar es el recurso energético más abundante disponible para el ser humano, la dispersión espacial y su variabilidad condicionan y limitan su aprovechamiento. La producción de un pequeño sistema de generación de energía eléctrica a partir de energía solar puede pasar de producir a potencia máxima a dejar de producir en cuestión de segundos (Hoff and Perez, 2010). En los sistemas mayores, como las centrales solares, la influencia de la variabilidad depende de la tecnología. En una central fotovoltaica, cuanto mayor sea su extensión menos variable será su producción debido al paso de nubes (Hoff and Perez, 2010). Cuando hablamos de centrales termosolares las rampas en producción pueden ser del orden de minutos, dependiendo de la inercia térmica del sistema, que dependerá a su vez de la tecnología y del tamaño de la central (Hirsch et al., 2017).

La variabilidad intradiaria de la radiación afecta significativamente a la operación y producción de las centrales termosolares. Una generación intermitente y muy variable dificulta su integración en la red eléctrica. Por tanto, caracterizar y, en la medida de lo posible, predecir las propiedades del recurso que influyen en la producción resulta fundamental para la optimización de la producción de las centrales y asegurar la estabilidad de la red.

El desarrollo de un proyecto de energía solar pasa por diferentes fases desde su concepción hasta su operación y en todas ellas el recurso solar tiene un papel fundamental. Mientras que en las fases iniciales sólo son necesarias estimaciones aproximadas de los niveles anuales de radiación solar (global o directa, dependiendo de la tecnología), a medida que avanza el proyecto aumenta la necesidad de profundizar en la caracterización del recurso, considerando frecuencias de datos mayores y estimaciones a largo plazo que sólo pueden obtenerse a partir del análisis estadístico de una serie continua, validada y extensa de datos.

En una base de datos extensa de radiación es muy común encontrar huecos, registros incorrectos del tiempo o medidas erróneas (Kumar et al., 2013). Aplicar un control de calidad que identifique estos errores y etiquete los datos para posteriormente corregirlos o descartarlos resulta fundamental. La lectura y la

interpretación del etiquetado para aplicar un buen procedimiento de corrección y el rellenado de los posibles huecos en la serie original o en la que resulte tras los descartes no son tareas fáciles. En la estimación de valores mensuales y anuales a partir de una serie con huecos, el procedimiento más habitual que se utiliza consiste en rellenar los días con fallos con el valor diario promedio mensual de los días disponibles de la serie. Este procedimiento puede alterar las propiedades estadísticas de la serie original e incurrir en estimaciones inexactas, dependiendo del número de días que sea necesario rellenar, por lo que se han propuesto otros procedimientos más elaborados (Hoyer-Klick et al., 2009; Ogunsola and Song, 2014; Schwandt et al., 2014).

La caracterización estadística del recurso solar en un emplazamiento se debe llevar a cabo a partir de extensas series de medidas de radiación solar –entre 6 y 30 años, dependiendo del autor (Meyer, 2010; Stoffel et al., 2010)- con diferentes requerimientos en la resolución temporal dependiendo del uso específico de los datos. En función de la dinámica del sistema de energía solar bajo consideración, la resolución temporal de la radiación solar de los datos que se utilizan para simular podría tener un efecto significativo en los resultados del análisis de viabilidad (Hirsch and Schenk, 2010). En el caso de las centrales termosolares, donde la variable más relevante es la radiación directa normal (DNI), algunos autores (Cebecauer and Suri, 2015) sugieren que el uso de valores promediados cada 10 minutos es una elección adecuada para la simulación orientada a los análisis de viabilidad. Para un análisis detallado de la respuesta dinámica de la central ante procesos transitorios provocados por el paso de las nubes y ciertos procedimientos de operación, como arranques y paradas, se requieren datos instantáneos de alta resolución.

En los trabajos de Tovar-Pescador (Tovar-Pescador, 2008) y Fernández-Peruchena (Fernández-Peruchena and Bernardos, 2015) puede encontrarse una completa revisión sobre las características estadísticas de las componentes de la radiación solar en resoluciones mensual, diaria, horaria e instantánea. En los últimos años, el análisis del comportamiento de los valores instantáneos atrae especial atención en el contexto del modelado de las componentes de la radiación para la simulación de sistemas solares (Bright et al., 2015; Morf, 2013; Ngoko et al., 2014). Por otro lado, el análisis de los valores mensuales y anuales también es abordado por varios autores con la idea de cuantificar y cualificar el recurso solar en diferentes emplazamientos. Se pueden encontrar ejemplos de tales estudios en los que se analizan regiones con un prometedor potencial solar en (Bachour and Perez-Astudillo, 2014; Islam et al., 2010,

2009; Zawilska and Brooks, 2011), aunque muchos de estos estudios se fundamentan en bases de datos de periodos relativamente cortos (inferiores a 5 años).

Las metodologías de evaluación del recurso solar en las etapas preliminares de los proyectos de centrales termosolares generalmente se basan en valores anuales, mensuales y, en algunos casos, en valores diarios de DNI (Cebeacauer and Suri, 2015; Pagola et al., 2010; Wilcox and Marion, 2008). El principal objetivo de estos estudios es caracterizar estadísticamente el recurso solar de un emplazamiento en estas escalas temporales. Sin embargo, la producción de una central termosolar no es directamente proporcional a la energía radiante recibida, también depende de cuándo (periodo del día) y cómo (con alta o baja variabilidad) se recibe esa energía radiante. Un día con valores de irradiancia muy variables a lo largo del día, debido al paso de nubes, puede tener la misma irradiación que otro con valores de DNI muy estables durante una fracción del día y, sin embargo, la producción eléctrica de ambos días puede ser muy diferente, ya que la operación de la central será muy diferente también. De cara a la operación de una central termosolar, los transitorios de la radiación causado por el paso de las nubes juegan un papel importante. El instante en el que la energía radiante está disponible también lo es, como consecuencia de la dependencia del rendimiento óptico del sistema de captación con la posición del sol. Por lo tanto, una metodología que permita clasificar los días teniendo en cuenta la variabilidad y la distribución intradiaria de la DNI, además de la irradiación o energía diaria, puede ayudar a caracterizar mejor el recurso desde el punto de vista de las centrales termosolares. Ya que tanto la variabilidad como la distribución temporal afectan al funcionamiento de las centrales termosolares, esta clasificación puede ser útil para evaluar el comportamiento de la central o para definir la estrategia de operación a partir la predicción de dichas características (Jiménez-Pérez and Mora-López, 2016). Por ejemplo, Rohani et al. (2016) analizaron datos de operación de la central Andasol 3 (Guadix, España) considerando tres tipos de días (claro, nublado y parcialmente nublado) de diferentes estaciones del año para evaluar un modelo de simulación. La estrategia de operación del modelo fue adaptada a los datos de operación de la central real. Del análisis de los datos medidos, los autores concluyen que el parámetro más importante para decidir el modo de operación de la central cada día se basa en la predicción de la DNI disponible de los días próximos. Por tanto, implementan la estrategia de operación del sistema de almacenamiento de energía térmica (TES) en función de las condiciones de cielo de los dos días próximos.

En la bibliografía se presentan varias opciones para clasificar los días, según su irradiancia global horizontal (GHI), en función de su energía y variabilidad. La clasificación de los días en función a la acumulada de GHI diaria, lleva estudiándose desde hace décadas mediante el uso del índice de claridad k_t , Liu and Jordan introdujeron este concepto en 1960 (Liu and Jordan, 1960). Con el desarrollo de las tecnologías solares, surgen nuevos métodos más complejos que combinan este parámetro con otros que aportan información cualitativa sobre el tipo de día mejorando los resultados. Harrouni et al. (2005) clasifican días utilizando datos 10-min de radiación global horizontal (GHI) según su dimensión fractal y su índice de claridad diario. De igual forma, Rahim et al. (2004) y Baharuddin et al. (2010) los clasifican usando la fracción de nubes y la duración de horas de Sol. Kang y Tam (2013) evalúan la probabilidad de persistencia del índice de claridad instantáneo y clasifican los días con ayuda del índice de claridad diario. Otros autores plantean opciones alternativas. Soubdhan et al. (2009) identifican tipos de días mediante el análisis de índices de claridad instantáneos (1-s) a partir de distribuciones de Dirichlet combinadas que modelan las curvas de distribución de los k_t . En la bibliografía pueden encontrarse numerosas metodologías para la clasificación de tipos de días en función del estado del cielo, pero no existe un consenso claro en cuanto a los parámetros más adecuados ni en cuanto al número de tipos de días a identificar.

La clasificación de días, en su mayoría, se ha llevado a cabo a partir de parámetros calculados a partir de la componente global de la radiación, aunque existen otras opciones. Calbó et al. (2001) señalan que la componente difusa es un mejor factor discriminante de la variabilidad de la radiación solar que la componente global. De igual forma, la DNI parece ser también un mejor factor discriminante, ya que es más sensible que la global al paso de nubes y se ve menos afectada por los efectos de la geometría solar. Además, es la única componente aprovechable por las centrales termosolares. Pérez et al. (2011) utilizan el índice de transmitancia, calculado a partir de valores horarios de DNI, como entrada para la caracterización de la variabilidad intrahoraria de la GHI. M Gastón-Romeo et al. (2011) emplean técnicas de *clustering* para clasificar los días a partir de la morfología de las curvas de DNI. Esta metodología está orientada a establecer una clasificación de días útil para las centrales termosolares. El uso de técnicas basadas en morfología matemática consigue conservar el comportamiento dinámico de las curvas de DNI, el cual ayuda a caracterizar su variabilidad, aunque no considera la distribución temporal de la energía.

Los criterios de clasificación usados para clasificar días de acuerdo a las condiciones del cielo a partir de las características de la radiación deben ser adaptados a su aplicación. La clasificación de días en función de las condiciones del cielo con el objeto de analizar y mejorar la integración de las centrales fotovoltaicas en la red eléctrica, ha sido abordado por varios autores a partir de las características de la GHI (Kang and Tam, 2013b; Muselli et al., 2000; Soubdhan et al., 2009; Stein et al., 2012). La metodología propuesta por Gastón-Romeo et al. (2011) está orientada a establecer una clasificación de días útil para las centrales termosolares, aunque no estudia la relación existente entre los tipos de días obtenidos y la producción de una central termosolar. Para validar la aplicabilidad de los métodos de clasificación con la idea de mejorar la caracterización y la predicción del recurso desde el punto de vista de la producción de las centrales, esta comprobación resulta fundamental.

2. OBJETIVOS

El objetivo general del presente trabajo es mejorar los procedimientos de caracterización del recurso solar para el desarrollo de proyectos de centrales termosolares, con el objeto de optimizar la producción de las centrales utilizando esta información en las diferentes etapas de su vida útil, desde el diseño hasta la operación. Para este fin, se proponen a su vez tres objetivos concretos:

- Caracterizar estadísticamente el recurso solar en las diferentes escalas temporales que pueden influir en la producción de una central termosolar en un determinado emplazamiento. Esto nos ayudará a seleccionar el punto de diseño de la central, la escala temporal más adecuada para simular su comportamiento, a conocer la variabilidad del recurso, los valores máximos, los periodos más favorables y más desfavorables para producir, así como a caracterizar cuantitativamente el recurso.
- Proponer y aplicar un método de clasificación de días en función de las propiedades de las curvas diarias de DNI que influyen en el grado de aprovechamiento de la radiación solar por una central termosolar.
- Establecer una relación entre tipo de día y producción de energía eléctrica.

Esta información puede ayudar a conocer mejor las características del recurso disponible en la fase de diseño de la central, a caracterizar mejor el recurso a largo plazo para el estudio de viabilidad, a mejorar las metodologías de predicción (en la fase de post-procesamiento) y a planificar las estrategias de operación y mantenimiento de la central.

3. ESTRUCTURA DEL DOCUMENTO

La presente Tesis se estructura de acuerdo a las instrucciones establecidas en el artículo 9 de la normativa reguladora del régimen de tesis doctoral (Acuerdo 9.1/CG 19-4-12) dentro de la modalidad *Tesis por compendio de publicaciones*. El objeto de este documento es describir y unificar el trabajo realizado en las publicaciones, incluidas en los anexos, que conforman esta Tesis. Para este fin, el documento contempla los apartados: Introducción, Objetivos, Estructura de la Tesis, Metodología, Resultados, Discusión y Conclusiones.

Siguiendo esta estructura, en el apartado *Introducción* se presentan de manera resumida los procedimientos de caracterización del recurso solar para su aplicación en las tecnologías solares que se han utilizado hasta la actualidad, así como las principales motivaciones que nos llevan a plantear nuevas metodologías que mejoren la caracterización del recurso para optimizar la producción de las centrales termosolares. Este es el objetivo principal que se describe en el apartado *Objetivos*, donde además se plantean tres objetivos concretos relacionados con las dos metodologías de caracterización principales que se presentan en este trabajo. A continuación, el apartado metodología, se encuentra estructurado en cuatro sub-apartados:

El primer sub-apartado describe la base de datos que se va a utilizar para aplicar las metodologías de caracterización. En este apartado se presenta la herramienta visual para la interpretación e identificación de anomalías de la base de datos detallada en el artículo adjunto en el *Anexo A* (Moreno-Tejera et al., 2015), así como la metodología utilizada para corregir y rellenar los registros erróneos de la base de datos descrita en el artículo adjunto en el *Anexo B* (Moreno-Tejera et al., 2016).

El segundo sub-apartado describe la metodología que se propone para caracterizar el recurso solar en las diferentes escalas temporales desde el punto de vista de las centrales termosolares, detallada en el artículo adjunto en el *Anexo B* (Moreno-Tejera et al., 2016).

El tercer sub-apartado describe la metodología de clasificación de tipos de días en

función de las características de la DNI que pueden influir en la producción de las centrales termosolares, detallada en el artículo adjunto en el *Anexo C* (Moreno-Tejera et al., 2017).

En el cuarto sub-apartado se describe el procedimiento que se utiliza para caracterizar los tipos de días identificados al aplicar la metodología del tercer sub-apartado desde el punto de vista de la producción de dos centrales de captadores cilindroparábolicos con diferente configuración. En él se propone un nuevo índice adimensional que permite evaluar el grado de aprovechamiento del recurso solar de una central y comparar los resultados obtenidos entre centrales con diferente configuración para un mismo recurso. El procedimiento contenido en este apartado se encuentra detallado en el artículo adjunto en el *Anexo C* (pendiente de publicación).

La estructura de cuatro sub-apartados seguida en el apartado *Metodología* se repite para la presentación de los principales resultados obtenidos al aplicar las metodologías descritas a la base de datos de Sevilla en el apartado *Resultados*. En el apartado *Discusión*, se analizan los resultados para identificar las conclusiones más relevantes del trabajo, que se exponen en el apartado *Conclusiones*. En este mismo apartado, se proponen futuras líneas a desarrollar y trabajos actualmente en curso relacionados con los resultados obtenidos en la presente Tesis.

4. METODOLOGÍA

En este apartado se resume la metodología seguida en cada uno de los artículos que conforman la presente Tesis. Estos artículos se encuentran adjuntos en los *Anexos*, donde se describe con mayor detalle la metodología aquí resumida.

4.1. Base de datos

La estación radiométrica y meteorológica del GTER, lleva registrando medidas desde 1984. Desde su primera instalación hasta su estado actual, ha sido trasladada en una ocasión y ha ido evolucionando hasta convertirse en una estación completa de calidad.

Desde 1999, la estación cuenta con un seguidor automático Kipp & Zonen 2AP sobre el que se encuentra colocado un pirheliómetro Eppley NIP (Normal Incidence Pyrheliometer) y un piranómetro Kipp & Zonen CM21. Ambos sensores son considerados estándares secundarios según ISO. La incertidumbre diaria y horaria especificada por los fabricantes es de un $\pm 2\%$ en el caso de la GHI y de un $\pm 1\%$ en el caso de la DNI, asumiendo un adecuado mantenimiento de los sensores. El GTER lleva a cabo operaciones rutinarias de mantenimiento y calibración de los sensores de acuerdo a las recomendaciones de los fabricantes. La estación también cuenta con sensores radiométricos y meteorológicos para registrar medidas de otras variables: irradiancia difusa horizontal (DHI), irradiancia global sobre superficie inclinada (37°), temperatura y humedad ambiente y velocidad y dirección de viento. El sistema de adquisición de datos de la estación registra medidas cada 5 segundos. Esta frecuencia de registro de las medidas permite caracterizar la variabilidad de la radiación con gran detalle.

En este trabajo se han usado las medidas registradas en la estación durante el periodo 2000-2015. Los años anteriores a este periodo han sido descartados por contar con un mayor número de anomalías en el registro y menor fiabilidad en las medidas.

4.1.1. Control de calidad de datos de radiación solar

Para caracterizar el recurso solar en cualquier emplazamiento, se necesita un periodo de medidas tan extenso como sea posible. Registrar medidas de radiación solar correctamente y con la menor incertidumbre posible, no es una tarea fácil. Además de la propia incertidumbre de los dispositivos es habitual encontrar huecos, errores en el registro del tiempo e incluso en la propia medida en este tipo de bases de datos. El registro de medidas de radiación requiere de un seguimiento y control continuo, así como de operaciones de mantenimiento periódicas para evitar o minimizar el efecto de otras fuentes de error. Aun realizando estas prácticas, es necesario aplicar un control de calidad que identifique posibles errores. La identificación de errores en las medidas debe ir acompañada de un etiquetado que informe y facilite la generación de una serie completa y validada.

En el presente trabajo, se aplican los filtros propuestos por la Baseline Solar Radiation Network (BSRN) (König-Langlo et al., 2013; Long and Dutton, 2004) para localizar y etiquetar los datos. Los resultados de aplicar los procedimientos de control de calidad de la BSRN se visualizan y analizan mediante una herramienta visual (Moreno-Tejera et al., 2015) que facilita la lectura y la interpretación de los resultados del control de calidad. Esta herramienta permite identificar, registros del tiempo (*time-stamps*) incorrectos, huecos y datos que no cumplen con alguno de los tres límites establecidos por la BSRN en cada uno de sus procedimientos.

La herramienta visual se basa en dos tipos de gráficas. En la primera, se representan las etiquetas que resultan de aplicar los procedimientos del control de calidad a cada una de las variables durante un periodo de tiempo tan extenso como se quiera. En una misma gráfica se visualizan los resultados de todo el periodo proporcionando una idea global de la calidad de los datos. En la Figura 4-1 se muestra un ejemplo de este tipo de gráfica aplicado a la componente global durante un periodo anual.

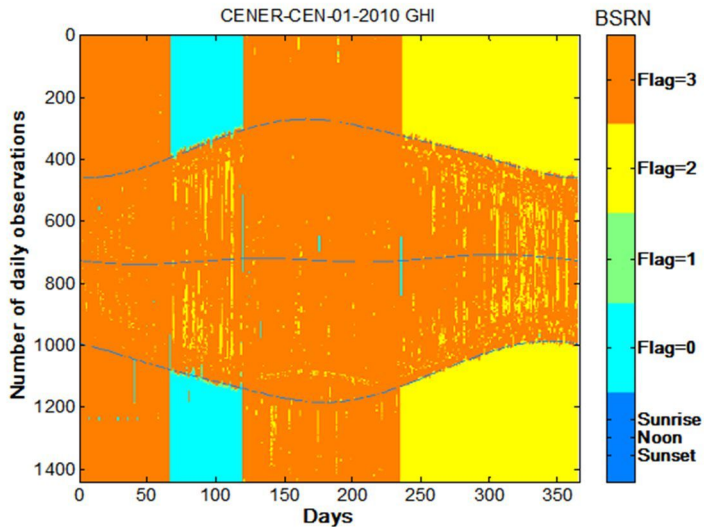


Figura 4-1. Representación de los resultados del control de calidad basado en los procedimientos de la BSRN.

Para cada día, se representan las etiquetas de cada registro con un código de colores en una línea vertical. Esta línea tiene tantos puntos como registros o números de observaciones se han registrado durante el día. Las líneas de salida, puesta de sol y medio día solar a lo largo del año se incluyen como referencia (en color azul). Estas líneas ayudan a identificar los errores en el registro del tiempo. Para la visualización de una extensa base de datos como es nuestro caso, se recomienda representar las etiquetas de un año completo como se muestra en el eje horizontal.

El segundo tipo de gráfica que se propone, sólo puede representarse si se han registrado las tres componentes de la radiación. Si este es el caso, se enfrentan gráficamente los valores medidos de GHI a los valores de GHI calculados a partir de sus dos componentes. En la misma gráfica, se sugiere representar como referencia dos líneas con una desviación de $\pm 8\%$ entre los valores medidos y calculados. Esta gráfica asiste en la identificación y cuantificación de algunos errores, tal y como se muestra en (Moreno-Tejera et al., 2015).

4.1.2. Rellenado de huecos

Los huecos y los datos incorrectos (incluyendo el registro del tiempo) son muy habituales en las bases de datos de radiación solar (Kumar et al., 2013). Sin embargo, en la bibliografía no es habitual encontrar una descripción detallada del procedimiento de relleno y corrección de estos huecos o datos incorrectos. La solución habitual es descartar estos datos en el análisis (Bachour and Perez-Astudillo, 2014; Islam et al., 2010; Zawilska and Brooks, 2011), aunque algunos autores proponen otras alternativas (Ogunsola and Song, 2014; Schwandt et al., 2014).

En este trabajo, la base de datos radiométricos se ha revisado, corregido y completado siguiendo un procedimiento desarrollado por el GTER (Moreno-Tejera et al., 2016), obteniendo como resultado 13 años (2000-2012)¹ de datos de DNI y GHI completos, validados y de alta calidad.

El procedimiento de corrección y relleno de huecos comienza por clasificar los días en función de su calidad y su nivel de nubosidad. Dicho procedimiento se detalla en la Tabla 4-1.

¹ El periodo 2013-2015 se ha revisado y completado siguiendo la misma metodología posteriormente a la publicación de Moreno-Tejera et al. (2016), por eso los resultados de la aplicación de la metodología a este periodo no se incluirá en el apartado 5.1.2.

Tabla 4–1. Procedimiento de rellenado de huecos y corrección de datos erróneos.

DÍAS CORREGIBLES	PROCEDIMIENTO DE RELLENADO
CASO 1. Medidas correctas de GHI y DNI con un hueco menor o igual a una hora cerca del orto/ocaso.	CASO 1A. Día nublado: interpolación lineal.
	CASO 1B. Día claro: modelo de cielo claro.
CASO 2. Medidas correctas y sin huecos de GHI. Medidas incorrectas o no existentes de DNI.	CASO 2. Modelo empírico de descomposición desarrollado a partir de las medidas correctas del GTER, si no hay medidas correctas de DHI disponibles.
CASO 3. Medidas correctas de GHI con un hueco menor o igual que una hora, cercano al orto/ocaso. Medidas incorrectas o no existentes de DNI.	CASO 3. Combinación de los casos 1 y 2.
CASO 4. Medidas correctas y sin huecos de DNI. Medidas incorrectas o no existentes de GHI.	CASO 4. Modelo empírico de descomposición desarrollado a partir de las medidas correctas del GTER, si no hay medidas correctas de DHI disponibles.
DÍAS DESCARTADOS	PROCEDIMIENTO DE RELLENADO
CASO 1. Archivo no creado	Los días descartados son sustituidos por otros días de la base de datos del GTER, cuya radiación acumulada diaria coincide con la registrada en la estación del IFAPA más cercana a la estación para esos días.
CASO 2. Medidas incorrectas de GHI y DNI	
CASO 3. Hueco superior a 1 h cerca del orto/ocaso en las medidas de GHI y de DNI.	
CASO 4. Hueco mayor de 30 min en las medidas de GHI y de DNI.	

Los días corregibles son aquellos en los que al menos una de las principales variables (GHI o DNI) se ha registrado correctamente o tiene un hueco en el registro inferior a una hora durante la primera hora de la mañana o la última de la tarde. Estos días se corrigen utilizando principalmente modelos de cielo despejados (Powell, 1984) y modelos empíricos de descomposición desarrollados a partir de medidas validadas

del GTER (Moreno, 2009). Estos modelos se basan en las reconocidas relaciones entre el índice de claridad y el índice de transmitancia de radiación directa (Iqbal, 1983; Skartveit and Olseth, 1992). Los días descartados son aquellos que no cumplen las condiciones descritas para ser considerados como días corregibles en la Tabla 4-1, para los que se necesita el uso de medidas de otras estaciones radiométricas como referencia. La base de datos utilizada en este caso procede de la Red de Estaciones agroclimáticas del Instituto de Agricultura y Pesca de Andalucía (IFAPA) que pertenece a la Junta de Andalucía. Esta base de datos es de libre acceso a través de la Web (Instituto de Investigación y Formación Agraria y Pesquera, 2009). Conociendo los valores de radiación acumulada diaria de la estación del IFAPA más cercana a la estación del GTER, el día descartado es reemplazado por otro día cercano cuya radiación acumulada diaria sea similar a la registrada para ese día en la estación del IFAPA (con una tolerancia de $\pm 5\%$). Este día cercano debe encontrarse en un rango de ± 5 días respecto al día descartado. Se establece un límite de ± 5 días porque los patrones meteorológicos a menudo permanecen sobre unos 5 días y la geometría solar no varía significativamente (Kumar et al., 2013). Si no existe un día cercano que reúna estos requisitos en su mismo año, se selecciona un día que los cumpla perteneciente a otro año.

4.2. Caracterización estadística de la radiación solar en Sevilla

La caracterización estadística del recurso solar en un emplazamiento es esencial para las diferentes fases de un proyecto de energía solar. Mientras que para las etapas más tempranas basta con una estimación aproximada de los valores anuales de irradiación (global o directa, dependiendo de la tecnología), a medida que avanza el proyecto aumentan los requisitos de precisión y detalle, exigiendo estimaciones a largo plazo que sólo pueden obtenerse a partir de un análisis estadístico de bases de datos extensas y completas. La base de datos del GTER cuenta con trece años de medidas de GHI y DNI completos y validados. Esto permite analizar las medidas de ambas variables en diferentes intervalos de tiempo.

La metodología que se presenta pretende caracterizar la distribución estadística de la GHI y la DNI en distintas resoluciones temporales, desde valores anuales a instantáneos. Con el objeto de poder considerar ambas variables con independencia del día del año al que pertenezcan, se usan dos índices adimensionales que eliminan

los efectos deterministas impuestos por la posición del Sol y la distancia Tierra-Sol. Estos índices son: el índice de claridad k_t (Liu and Jordan, 1960) y el índice de fracción de directa k_{b0} (Tovar et al., 1999).

El índice k_t se define como el cociente entre GHI y la irradiancia solar extraterrestre sobre superficie horizontal I_0 :

$$k_t = \frac{I_{g0}}{I_0} = \frac{I_{g0}}{I_{sc} \cdot \varepsilon_0 \cdot \cos \theta_z} \quad \text{Ec. (4-1)}$$

Donde I_{g0} es la irradiancia global horizontal, I_{sc} es la constante solar, ε_0 el factor de corrección de la distancia Tierra-Sol y θ_z es el ángulo cenital.

El índice k_{b0} se define como el cociente entre la irradiancia directa normal y la que se registraría en el caso de no haber atmósfera:

$$k_{b0} = \frac{I_{bn}}{I_n} = \frac{I_{bn}}{I_{sc} \cdot \varepsilon_0} \quad \text{Ec. (4-2)}$$

Donde I_{bn} es la irradiancia directa normal, I_n es la irradiancia extraterrestre normal, I_{sc} es la constante solar y ε_0 el factor de corrección de la distancia Tierra-Sol.

Ambos índices se han usado en la literatura en diferentes intervalos de agregación (instantáneos, horarios, diarios y mensuales). Dependiendo del intervalo de agregación Δt , los índices se simbolizan como $k_t^{\Delta t}$ o $k_{b0}^{\Delta t}$, y se calculan con las siguientes expresiones:

$$k_t^{\Delta t} = \frac{\int_{\Delta t} I_{g0} dt}{\int_{\Delta t} I_0 dt} \quad \text{Ec. (4-3)}$$

$$k_{b0}^{\Delta t} = \frac{\int_{\Delta t} I_{bn} dt}{\int_{\Delta t} I_n dt} \quad \text{Ec. (4-4)}$$

4.2.1. Caracterización estadística de los valores instantáneos, horarios y 10-min.

Varios autores han analizado la distribución de los valores instantáneos de k_t encontrando un resultado similar: la distribución muestra un carácter bimodal, que se acentúa cuando la masa óptica de aire decrece (Jurado et al., 1995; Skartveit and Olseth, 1992; Suehrcke and McCormick, 1988; Tovar et al., 1999, 1998; Varo et al., 2006; Woyte et al., 2007). Sin embargo, un estudio reciente, encuentra diferentes comportamientos en función del tipo de clima (Fernández-Peruchena and

Bernardos, 2015). Para identificarlos, utiliza la clasificación climática de Köppen-Geiger. En el mapa de clasificación de Köppen-Geiger, Sevilla se identifica por el símbolo "Csa", correspondiendo a un clima templado con veranos secos y calurosos. Este tipo específico de clima no se encuentra incluido en el estudio de Fernández-Peruchena and Bernardos (2015), aunque puede compararse con los estudios realizados por Tovar et al. (Tovar et al., 1999, 1998) y Varo et al. (Varo et al., 2006), ambos realizados en localizaciones identificadas por el mismo tipo de clima en la clasificación Köppen-Geiger.

Para la simulación del comportamiento de los sistemas de energía solar el uso de valores instantáneos suele reservarse para el análisis dinámico del sistema y de sus componentes, ya que para la evaluación o estimación de la producción los propios programas de simulación no contemplan la respuesta del sistema ante cambios de alta frecuencia. Por tanto, es habitual utilizar frecuencias mayores para estas simulaciones. El uso de frecuencias más elevadas para representar el comportamiento de los valores instantáneos puede incurrir en desviaciones en las estimaciones de producción si ambas series no son estadísticamente similares.

En el presente trabajo se analiza la función de densidad de probabilidad de los índices k_t y k_{60} instantáneos (registrados cada 5s), de los 13 años de medidas registradas en Sevilla por el GTER, en función de la masa óptica de aire y se compara con los resultados obtenidos por otros autores en emplazamientos con la misma clasificación climática de Köppen-Geiger. A continuación, se comparan las distribuciones de frecuencia de los índices k_t y k_{60} calculados a partir de las medidas registradas cada 5s con las promediadas cada 10 min y 1 h. Por último, se caracterizan cuantitativamente las medidas de GHI y DNI, comparando la distribución en frecuencia de los valores de ambas variables promediados cada 10 min y 1h, y representando los valores promedios horarios del periodo a lo largo del año.

4.2.2. Caracterización estadística de los valores diarios

El k^d_t se utiliza con mucha frecuencia en la bibliografía para la clasificación de días según el estado del cielo. Los autores proponen diferentes rangos de k^d_t para clasificar los días en: despejados, parcialmente nublados o completamente nublados (Iqbal, 1983; Kudish and Ianetz, 1996). Uno de los objetivos planteados en esta Tesis es desarrollar una metodología para clasificar días a partir de las características

intradíarias de la DNI. Previamente a su aplicación, se analiza la clasificación de los días únicamente en función de los valores de k^d , ya que, por su sencillez, es una metodología habitualmente empleada. Los valores límite considerados en este estudio para definir los días completamente cubiertos y los días despejados son 0.3 y 0.65 (Lam and Li, 1996).

Adicionalmente, se analiza la caracterización estadística de los valores diarios de cada año para el periodo completo a partir de una representación Box & Whisker. Este tipo de análisis ayuda a conocer la variabilidad de la energía diaria disponible de las variables radiométricas a lo largo del año y a lo largo del periodo.

4.2.3. Caracterización estadística de los valores mensuales y anuales

La evaluación del recurso solar para el análisis de viabilidad de los proyectos basados en el aprovechamiento de la energía solar, a veces, se lleva a cabo a partir de los valores mensuales de GHI y/o DNI (Pagola et al., 2010). Gracias a los procedimientos de rellenado de huecos aplicados a la base de datos, la evaluación de los valores mensuales y anuales de GHI y DNI puede realizarse con menor incertidumbre que en otros estudios. En el presente trabajo, se analizan los valores promedios diarios mensuales y anuales de ambas variables junto con sus valores extremos para evaluar el potencial del recurso solar en Sevilla.

4.3. Clasificación de días en función de las curvas de DNI

En este trabajo se presenta una metodología de clasificación de días a partir de las curvas de DNI orientada a las necesidades de las tecnologías termosolares. Desde el punto de vista del recurso, la producción de una central termosolar depende principalmente de la radiación directa normal acumulada diaria disponible y de la distribución de dicha radiación o energía a lo largo del día. La metodología propuesta se apoya en la definición de las propiedades de las curvas de DNI a partir de índices adimensionales. Estos índices ayudan a clasificar los días, facilitando la identificación de aquellos que posean características similares. Para realizar el proceso de clasificación se utilizan técnicas de *clustering* convencionales. Los índices propuestos caracterizan su energía (cuánto), su variabilidad intradiaria (cómo) y la distribución temporal de dicha energía (cuándo). Se han analizado distintos índices

para representar estas propiedades. Finalmente, se comparan dos grupos de índices para seleccionar el más adecuado a efectos de clasificación.

4.3.1. Modelo de cielo despejado envolvente

Los modelos de cielo despejado tratan de estimar la irradiancia disponible en ausencia de nubes reproduciendo la atenuación consecuencia de la composición de la atmósfera. En la bibliografía, se pueden encontrar numerosos modelos que dependen de diferentes parámetros relacionados con la composición atmosférica y su influencia en la atenuación de la radiación solar. Cuando esta información no está disponible se suelen utilizar modelos ajustados empíricamente con medidas del emplazamiento en días con condiciones de cielo despejado o días claros. Para definir lo que puede considerarse un “día claro” Gómez Camacho y Blanco Muriel (1990) introducen el concepto de día claro envolvente. El día claro envolvente es aquel cuya irradiancia directa normal predice cual sería el comportamiento del día si su atmósfera siguiera el comportamiento de sus máximos de irradiancia. De esta forma, se puede utilizar la información que se obtiene de los abundantes días con alternancia de nubes y claros para ampliar el rango de parámetros de entre los que se seleccionan los que definen la curva de cielo claro. El modelo de cielo despejado utilizado por los autores es similar al ampliamente referenciado modelo ASHRAE (Powell, 1984) de decrecimiento exponencial, cuyos dos coeficientes se obtienen a partir de los datos experimentales de irradiancia en función del tiempo:

$$I_{cs} = E_0 \cdot e^{\frac{-\beta}{\sin \alpha}} \quad \text{Ec. (4-5)}$$

donde I_{cs} es la irradiancia directa normal en condiciones de cielo despejado, α es la altura solar correspondiente y E_0 y β los parámetros que deben ser ajustados empíricamente. En la definición del modelo de ASHRAE sus parámetros pueden ser interpretados como E_0 irradiancia aparente extraterrestre y β como parámetro global de extinción (Fernández-Peruchena et al., 2015). La irradiancia aparente extraterrestre depende del día del año por la variación de la distancia entre la Tierra y el Sol. El uso de un único par de valores de ajuste seleccionado para definir la curva de cielo despejado en cualquier otro día del año podría ser discutible. Sin embargo, esta dependencia no afecta al ajuste de los parámetros si se define el modelo con la siguiente expresión:

$$I_{cs} = I_{sc} \cdot \varepsilon_0 \cdot A \cdot e^{\frac{-B}{\sin \alpha}} \quad \text{Ec. (4-6)}$$

donde I_{cs} es la irradiancia directa normal en condiciones de cielo despejado, I_{sc} es la constante solar, ε_0 es el factor de corrección de la distancia Tierra-Sol, α es la altura solar correspondiente, y A y B son los parámetros que deben ser ajustados empíricamente.

En esta Tesis, se aplica el concepto de día claro envolvente para ajustar el modelo a los instantes del día con menor atenuación atmosférica y usarlo como estado de referencia. Aunque en este estudio, el ajuste de los parámetros no se hará para cada día, sino para cada valor de altura solar. El recorrido del rayo es el mismo para una misma altura solar y la atenuación de la radiación únicamente depende de los componentes atmosféricos que se encuentre en su trayectoria. El único parámetro que es independiente del tipo de atmósfera que atraviese el rayo y que sí depende de la época del año, es la irradiancia extraterrestre. A efectos prácticos, la irradiancia incidente en dirección normal a la atmósfera puede considerarse el valor de la constante solar multiplicado por un factor que corrige su magnitud en función de la distancia Tierra-Sol. De modo que dividiendo la expresión que modela los valores de DNI en el emplazamiento por el factor de corrección de cada día, el modelo de cielo despejado únicamente depende de la altura solar. En el caso del modelo de ASHRAE, la expresión queda como se muestra en la Ec. (4-7).

$$\frac{I_{cs}}{\varepsilon_0} = I_{sc} \cdot A \cdot e^{\frac{-B}{\sin \alpha}} \quad \text{Ec. (4-7)}$$

Por tanto, se ajustan los parámetros A y B en función de los valores máximos observados de I_{bn}/ε_0 para cada valor de altura solar mayor de 5° de la base de datos (Gómez Camacho and Blanco Muriel, 1990). Así, sólo se necesita estimar un par de valores de A y B ($A=0.8187$ y $B=0.0840$) para representar las condiciones de máxima transmitancia atmosférica cualquier día del año. La Figura 4-2 muestra el ajuste de los instantes de mínima atenuación y las curvas resultantes para todos los días del año.

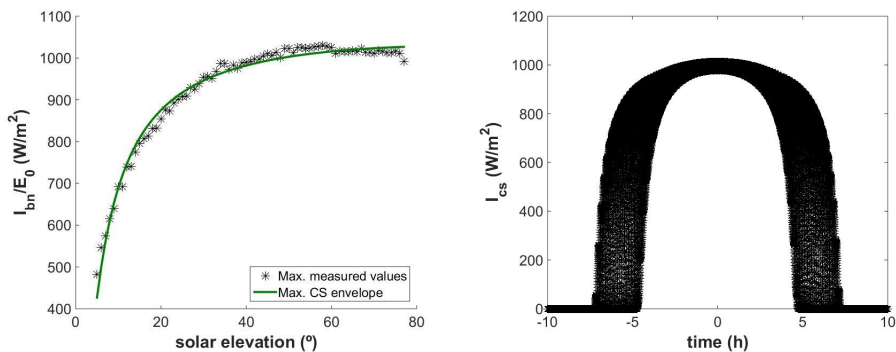


Figura 4-2. A la izquierda, valores máximos de I_{bn}/ϵ_0 frente a la altura solar y el modelo de ASHRAE ajustado. A la derecha, modelo de ASHRAE envolvente máximo para los 365 días del año.

4.3.2. Índices de clasificación

Varios índices previamente definidos en la bibliografía a partir de la componente global se han adaptado para su uso a partir de la componente directa. Estos, y otros índices definidos en la bibliografía para la DNI, se han analizado desde el punto de vista de las tres propiedades de las curvas de DNI objeto de representación. A continuación, se presentan los índices considerados como más adecuados para el objetivo propuesto.

4.3.2.1. Índice de caracterización de la energía

Para caracterizar la energía de la componente directa de la radiación, Skarveit y Olseth (1992) proponen un índice similar al habitualmente utilizado índice de claridad, denominado por algunos autores índice de transmitancia de la radiación directa o k_b . Este índice compara la irradiancia directa medida respecto a la que se recibiría si el cielo estuviera completamente despejado:

$$k_b = \frac{I_{bn}}{I_{cs}} \quad \text{Ec. (4-8)}$$

Donde I_{bn} es la irradiancia directa normal e I_{cs} es la irradiancia calculada mediante un modelo de cielo despejado (ver sección 2.1). Al igual que k_t y k_{to} (ver apartado 4.2.1),

k_b se define para valores instantáneos de la radiación, pero puede ser calculado para otros periodos de agregación (horario, diario o mensual) mediante la siguiente expresión:

$$k_b^{\Delta t} = \frac{\int_{\Delta t} I_{bn} dt}{\int_{\Delta t} I_{cs} dt} \quad \text{Ec. (4-9)}$$

Como se explica en el apartado 4.2.1 algunos autores utilizan una definición diferente del índice k_b , el índice k_{b0} . La definición del k_{b0} es más universal, ya que k_b necesita un modelo de cielo despejado. Los modelos de cielo despejado aun apoyándose en una serie de hipótesis que simplifican el comportamiento de la atmósfera, dependen de unos índices que se estiman con datos del emplazamiento en el que se ha realizado el ajuste. Sin embargo, en valores instantáneos, el valor de k_{b0} es función de la masa relativa de aire que atraviesa el rayo y ésta a su vez va variando a lo largo del día (Tovar-Pescador, 2008). Es decir, el valor de k_{b0} no es constante a lo largo de un día que se mantiene despejado y con una composición de atmósfera constante, observándose cómo disminuye al aumentar la atenuación debido al efecto del mayor espesor de atmósfera que recorre el rayo en las horas cercanas al orto y al ocaso. Este efecto es menor si se utiliza k_b . Como se observa en la Figura 4-3, al dividir por la irradiancia de un modelo de cielo despejado, se tiene en cuenta el efecto de la variación del espesor de atmósfera que recorre el rayo para un día de cielo despejado. Aun así, un aumento de la atenuación afecta en mayor proporción a los instantes en los que el rayo atraviesa mayor masa de aire, por lo que la curva de valores instantáneos de k_b no es totalmente constante para un día de cielo despejado a menos que corresponda a un día con valores de atenuación atmosférica similares a los utilizados en el ajuste del modelo. Aun así, respecto a la definición de k_{b0} presenta una mejor opción para adimensionalizar los valores instantáneos de DNI.

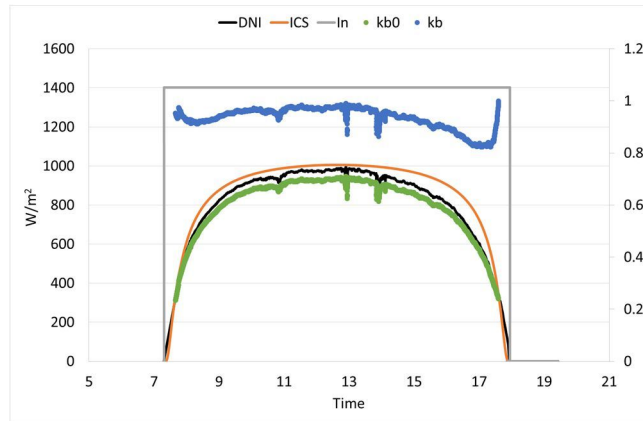


Figura 4-3. Representación de valores instantáneos del índice de transmitancia k_b , del índice de fracción de directa k_{b0} , de las medidas registradas de DNI y de los valores correspondientes del modelo envolvente máxima de cielo despejado de ASHRAE para un día en su mayoría despejado.

En este estudio se utiliza el índice diario k_{b^d} como índice de clasificación para representar la energía de las curvas de DNI y en valores instantáneos para la definición de los índices que representan la variabilidad, ya que k_b es menos dependiente de la masa de aire, como se muestra en la Figura 4-3. Para el cálculo de la DNI en condiciones de cielo despejado se utilizará el modelo de cielo despejado de envolvente máxima obtenido según se ha descrito en el apartado 4.3.1.

4.3.2.2. Índices de caracterización de la variabilidad

Para su uso en la metodología de clasificación se han considerado aquellos índices encontrados en la literatura capaces de caracterizar la variabilidad a partir de un índice diario, independientemente del día del año en el que sea evaluado,

A continuación, se presentan los índices seleccionados. Ambos índices se han usado en otros trabajos para caracterizar la variabilidad de medidas de GHI.

- **Índice de Variabilidad (VI):**

Stein et al. (2012) proponen el Índice de Variabilidad (VI) para identificar periodos y localizaciones con alta variabilidad de GHI. Este índice compara la longitud de las

curvas de GHI con la longitud de la curva de GHI del día correspondiente en condiciones de cielo despejado. El VI proporciona información sobre la variabilidad diaria, pero puede alcanzar valores similares en días muy despejados y días completamente cubiertos. Los autores utilizan el índice de claridad diario k_t junto con el VI para identificar cuatro tipos de día: claro, cubierto, mixto y altamente variable durante todo el día. En este trabajo, se aplica la misma definición a la DNI. En este caso, el VI se define como la relación entre la longitud de la curva de DNI y la longitud de la curva de DNI envolvente máxima de cielo despejado:

$$VI = \frac{\sum_{k=2}^n \sqrt{(I_{bn_k} - I_{bn_{k-1}})^2 + \Delta t^2}}{\sum_{k=2}^n \sqrt{(I_{cs_k} - I_{cs_{k-1}})^2 + \Delta t^2}} \quad \text{Ec. (4-10)}$$

Donde I_{cs} es la envolvente máxima de cielo despejado de radiación directa normal, Δt el tiempo entre dos datos consecutivos (10 min en nuestro caso) y n el número de intervalos diezminutales de cada día considerado.

- **Índice de Persistencia (POP_D):**

Kang and Tam (2013) presentan un método simple pero interesante llamado K-POP que clasifica los días según las condiciones del cielo a partir de medidas de GHI. Los autores utilizan el índice de claridad diario, k_t , y la probabilidad de persistencia diaria de los valores de k_t instantáneos, índice POP_D , para obtener una clasificación cuantitativa y cualitativa del perfil diario de la radiación global. El índice POP_D informa sobre la estabilidad del estado del cielo, de forma que un valor de POP_D cercano a la unidad corresponde a un día muy estable (despejado o cubierto) y viceversa.

El índice POP_D se define como el cociente del número de instantes en los que el índice instantáneo k_t se mantiene constante respecto a su predecesor y el número de instantes totales del día menos uno.

El índice POP_D se puede redefinir utilizando los valores instantáneos del índice k_b . Siguiendo una definición paralela, se calcula la persistencia del índice k_b como la ratio entre la suma de los instantes en los que k_b difiere en menos de 0.1 respecto a su predecesor y el total de instantes menos uno. En la Figura 4-4, se representa una curva de cielo despejado junto a otras curvas de DNI que mantienen un k_b constante a lo largo del día desde 0.1 a 0.9. Estas curvas nos ayudan a visualizar los transitorios computados con el índice POP_D .

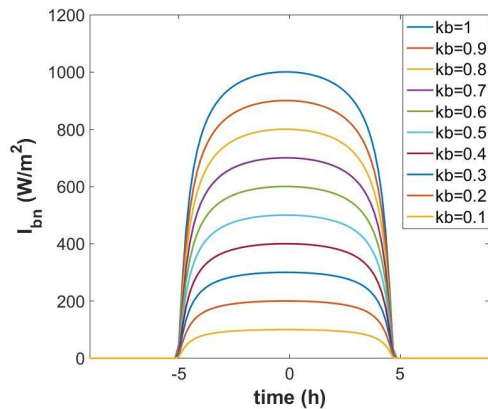


Figura 4-4. Curvas teóricas de cielo despejado de DNI con valores instantáneos de k_b desde 0.1 hasta 1.

4.3.2.3. Índices de caracterización de la distribución temporal de la energía

Los índices que se emplean habitualmente en la literatura caracterizan la radiación recibida según la energía acumulada diaria y, en ocasiones, la variabilidad con la que ésta se recibe, pero sin aportar información sobre su distribución en el tiempo. Para las centrales termosolares es útil conocer, en la medida de lo posible, la distribución de esta energía. En la primera fila de la Figura 4-5, se muestra la representación diaria de los valores de DNI de dos días típicos en los que una central sin almacenamiento térmico podría estar funcionando de forma ininterrumpida durante la mitad del día, aun teniendo un valor de k_b típico de un día de nubes y claros. En el presente trabajo se proponen un par de índices que informan al operador a tener información sobre la distribución en el tiempo de la radiación.

Una sencilla forma de representar la distribución a lo largo del día de la energía radiante es a través de la fracción de radiación acumulada durante la primera mitad del día o de la tarde respecto a la total. Estas relaciones se denominan fracción de mañana o fracción de tarde, según su definición. Se define la *fracción de mañana* como:

$$F_m = \frac{H_{bn,m}^d}{H_{bn}^d} \quad \text{Ec. (4-11)}$$

Donde $H_{bn,m}$ es la radiación directa normal recibida durante la primera mitad del día y H_{dn} es la DNI diaria del día completo.

De igual forma, el índice F_a o fracción de tarde puede ser definido mediante la ratio de la DNI acumulada durante la segunda mitad del día y la DNI acumulada del día completo. Resultando en un valor complementario respecto a la unidad y proporcionando la misma información para el objetivo propuesto.

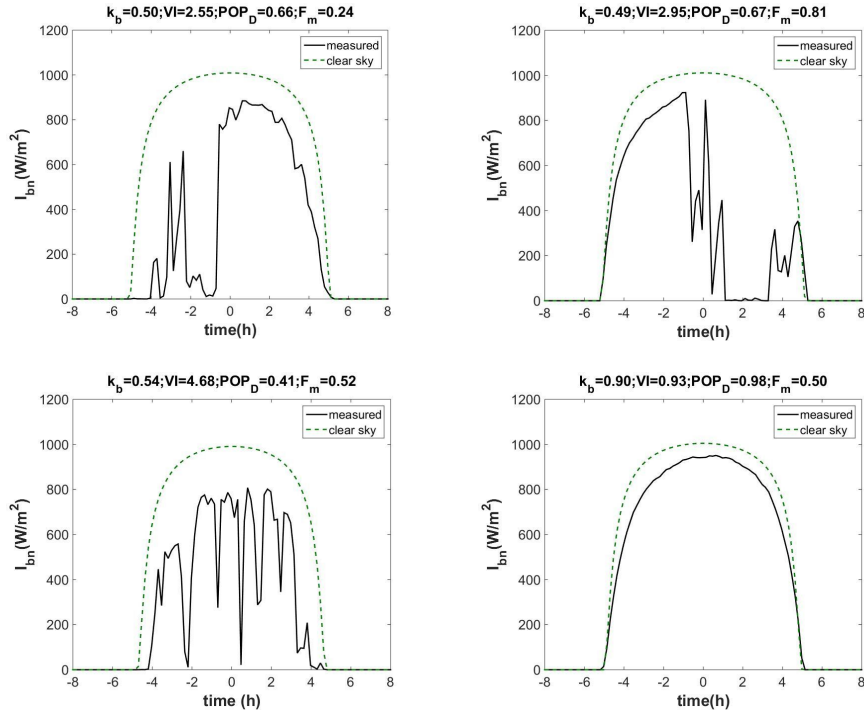


Figura 4-5. Representación de los valores de DNI registrados durante 4 días en Sevilla durante el 2010 junto con la envolvente de cielo despejado y sus correspondientes valores de k_b , VI, POP_D y F_m .

4.3.2.4. Evaluación de índices

En la Figura 4-5 se representa la DNI medida en Sevilla durante 4 días del año 2010 junto con la curva de cielo despejado envolvente correspondiente. También se incluye en cada figura los valores de los 4 índices propuestos para representar las

características de las curvas de DNI. En la primera fila, se representan dos días con energía y variabilidad similar, pero diferente distribución de la energía a lo largo del día. Ambos días tiene valores similares de k_t , POP_D y VI , informando de forma adecuada de la similitud en la energía y la variabilidad de ambos días, y diferentes valores de F_m , informando de en qué mitad del día se suministra la energía principalmente. El día representado en la segunda columna, tiene un valor muy alto de F_m , ya que la mayoría de la energía es recibida en la primera mitad del día. El día representado en la primera columna, es casi simétrico, por eso el valor de F_m es casi el complementario respecto a la unidad. En la segunda fila, se representan dos días con una distribución de energía homogénea en el tiempo con valores de F_m en torno a 0.5. Sin embargo, el día representado en la primera columna muestra una mayor variabilidad o una menor persistencia de los valores de DNI que queda reflejado en un mayor valor de VI y un valor bajo de POP_D . Merece la pena destacar, que VI es el único índice que no toma valores entre 0 y 1, mostrando más sensibilidad a los transitorios.

4.3.3. Procedimiento de clasificación

Los cuatro índices diarios descritos en el apartado 4.3.2 han sido calculados para los 13 años de medidas de DNI de la estación de Sevilla del periodo 2000-12. Dos combinaciones de índices son seleccionadas para analizar los resultados obtenidos al aplicar técnicas de agrupación sin supervisión. Ambas combinaciones incluyen un índice para representar la energía, la variabilidad y la distribución temporal de la DNI. La primera combinación la componen los índices k_t , VI y F_m . La segunda la componen los índices k_t , POP_D and F_m . Los índices VI y POP_D están claramente anticorrelacionados, pero representan de distinta forma la variabilidad, como se muestra en la Figura 4-6.

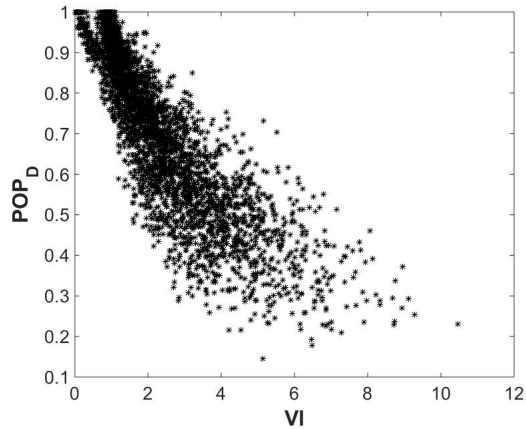


Figura 4-6. Representación de los valores del POP_D frente a los valores diarios del VI para todos los días del periodo 2000-12.

En esta sección, se describe el algoritmo de agrupación y la metodología aplicada para seleccionar el número más adecuado de grupos en cada caso. Finalmente, los grupos resultantes se comparan con una clasificación realizada por 5 expertos mediante la visualización de las curvas de DNI. Esta comparativa se utiliza como medida de validación para seleccionar la combinación de índices más adecuada.

4.3.3.1. Técnica de agrupación de datos sin supervisión: clustering

El objetivo de un algoritmo de agrupación o *clustering* basado en métodos de partición es clasificar una serie de datos en k grupos o clústeres: los datos de un mismo clúster deben compartir las mismas características, y éstas deberían ser diferentes respecto a las de otros clústeres. *k-means* es el algoritmo más popular de los métodos de partición sin supervisión. Este algoritmo proporciona buenos resultados para cierto tipo de datos, y su mayor ventaja es su simplicidad. Sin embargo, *k-means* es sensible a los *outliers*, o datos que alcanzan valores numéricamente distantes del resto de los datos, ya que busca los centros de sus agrupaciones mediante el promedio de la distancia media de los datos contenidos en cada uno de ellos. En ciertos casos, este algoritmo puede distorsionar la distribución de los datos en grupos. Para evitar este inconveniente, el algoritmo *k-*

medoid utiliza un punto de referencia o *medoide*, que corresponde al punto más centrado dentro del grupo (Han and Kamber, 2001). El algoritmo *k-medoid* trata de minimizar la suma de la distancia métrica entre cada dato de un grupo y su *medoide*. Este algoritmo permite obtener una buena caracterización de los grupos que no son demasiado alargados y hace posible aislar los *outliers* en la mayoría de los casos (Kaufman and Rousseeuw, 1990).

En este trabajo se aplica el algoritmo *k-medoid* "Partitioning Around Medoids" (PAM) (Han and Kamber, 2001) para tratar de agrupar días con características radiométricas similares. Este algoritmo es más robusto que el *k-means*, pero comparte dos inconvenientes: los centros iniciales se seleccionan de forma aleatoria, y el usuario tiene que especificar el número de clústeres que quiere obtener. La efectividad de los resultados, depende de la distancia entre los centros iniciales y finales de los clústeres. Por lo tanto, para buscar la agrupación más adecuada, se ejecuta el algoritmo varias veces para cada número de clúster que se pretende estudiar, barriendo un rango de número de clústeres adecuado para los datos analizados y el objetivo planteado.

4.3.3.2. Pre-procesamiento de los datos

Con el algoritmo de clasificación, se pretende agrupar los días en función de tres características diarias de las curvas de DNI representadas a partir de tres índices. Todos los índices seleccionados, excepto VI, toman valores entre 0 y 1. Para evitar que VI tenga mayor peso en el procedimiento de agrupación, se normaliza el índice dividiéndolo por el máximo valor registrado ($VI_{max} = 10.92$). Además, se segmenta el índice F_m para facilitar la agrupación de días de acuerdo a lo que resulta más interesante de conocer respecto a la distribución de la energía a lo largo del día. Como resultado, se presenta un nuevo índice llamado fracción de tiempo, F_t , que se define mediante la Ec. (4-12):

$$F_t = 1 \quad \text{if } F_m \leq 0.3$$

$$F_t = 0.5 \quad \text{if } 0.3 < F_m < 0.7 \quad \text{Ec. (4-12)}$$

$$F_t = 0 \quad \text{if } F_m \geq 0.7$$

Con este nuevo índice se identifican los días cuya energía está principalmente concentrada durante la mañana (0), durante la tarde (1) o durante ambos periodos (0.5). En la Figura 4-7 se muestran los valores de los dos grupos de índices que se van

a emplear en el procedimiento de identificación de tipos de días. Para facilitar la referencia a cada grupo de índices de aquí en adelante, se denomina el grupo de índices compuesto por VI , k_b y F_t como grupo VI , y el grupo compuesto por POP_D , k_b y F_t , como grupo POP_D .

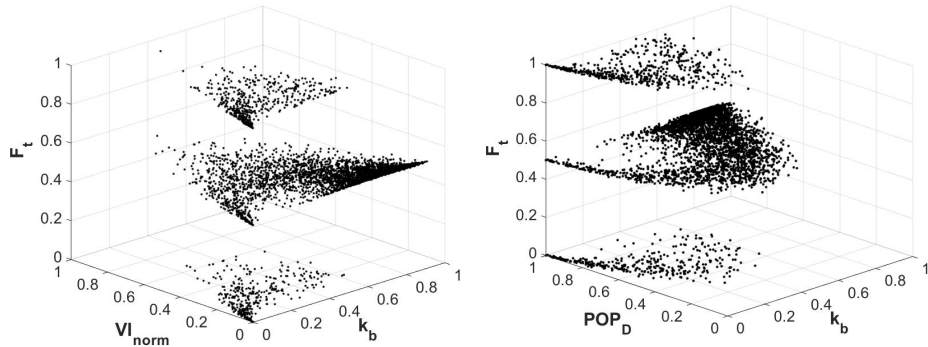


Figura 4-7. Representación de los dos grupos de índices calculados a partir de los 13 años de medidas de DNI utilizadas para clasificar los días en función del estado del cielo: k_b , VI y F_t (izquierda) y k_b , POP_D y F_t (derecha).

4.3.3.3. Determinación del número de clústeres

Las técnicas empleadas para determinar el número de grupos (o clústeres) más adecuado se clasifican en dos tipos, métodos de validación externos e internos. Los métodos externos de validación, se basan en el conocimiento previo que se tiene sobre la partición de los datos. Los métodos internos, consideran la información intrínseca de la estructura geométrica de los datos (Zagouras et al., 2013a). Estos métodos, se apoyan en una serie de índices que tratan de evaluar la similitud entre los miembros de un mismo clúster y la disimilitud entre los miembros de clústeres diferentes. En este trabajo se utilizan dos de los índices más comunes: Calinski-Harabasz (CH) (Calinski and Harabasz, 1974) y el índice Silhouette (SH) (Rousseeuw, 1987). Valores altos del índice CH corresponde a una partición de clústeres bien separados: donde los centros de los clústeres se encuentran separados en el espacio y los miembros de cada clústeres agrupados de forma compacta. El índice SH califica cada punto considerando su posición respecto a los puntos del

clúster al que pertenece y respecto a los puntos pertenecientes a otros clústeres. Este índice puede evaluar un único punto, los puntos de un clústeres o todos los puntos de la serie, tomando valores entre -1 y +1. Un valor de SH cercano a 1 de un punto, indica que ese punto pertenece a un clúster compacto y que se encuentra situado a la máxima distancia del clúster más cercano, y por lo tanto, está correctamente clasificado. Un valor negativo indica la posibilidad de que no esté correctamente clasificado.

En ocasiones la tendencia de los índices al aumentar el número de clústeres es monótona creciente o decreciente, dificultando la identificación de un máximo o un mínimo absoluto. Para ayudar a identificar el número de clústeres más adecuado a partir de los índices de validación interna, se utiliza el método gráfico L-method (Salvador and Chan, 2003; Zagouras et al., 2014, 2013b). Este método trata de identificar el punto de inflexión de la curva que representa el índice de validación frente al número de clústeres, buscando el par de rectas en el sentido de los mínimos cuadrados que mejor se ajustan a los puntos de ambos lados de la curva. Con este método, si la evolución del índice no muestra el óptimo buscado a través de un máximo o un mínimo absoluto, se busca el punto a partir del cual se produce un cambio de tendencia en la curva o "knee point".

4.3.3.4. Validación y selección de la combinación de índices apropiada

Para validar los resultados obtenidos y seleccionar el conjunto de índices más adecuado, una vez identificados los grupos o clústeres, se comparan los tipos de días de los años 2010 y 2011 para cada uno de los grupos de índices estudiados con una clasificación realizada por 5 expertos a partir de la visualización de las curvas de DNI. En el ejercicio de clasificación, los expertos han elegido entre 11 opciones de tipo de días, definidos en la Tabla 4-2. Se considera que estos 11 días abarcan todos los escenarios interesantes de analizar desde el punto de vista de las centrales termosolares, con las características de las curvas tenidas en cuenta. En la Figura 4-8 se muestra la representación gráfica diaria utilizada por los expertos como referencia para hacer su clasificación. La comparativa entre la clasificación realizada por los expertos y la obtenida al aplicar la metodología de clasificación propuesta en el presente trabajo, se utilizará para seleccionar el grupo de índices más adecuado para caracterizar los tipos de días.

Tabla 4–2. Tipos de días utilizados como referencia por los expertos para realizar la clasificación del periodo 2010-11. Los días han sido definidos de acuerdo a las características consideradas a través de los índices para aplicar el algoritmo de agrupación.

Tipo	Estado del cielo	Energía	Variabilidad	Distribución temporal de la energía
1	Totalmente despejado (sin nubes)	alta	baja	homogénea
2	En su mayoría despejado (nubes altas)	media	baja	mañana
3	En su mayoría despejado (nubes altas)	media	baja	homogénea
4	En su mayoría despejado (nubes altas)	media	baja	tarde
5	Nubes y claros (nubes medias/altas)	media	alta	mañana
6	Nubes y claros (nubes medias/altas)	media	alta	homogénea
7	Nubes y claros (nubes medias/altas)	media	alta	tarde
8	En su mayoría cubierto (nubes medias)	baja	media	mañana
9	En su mayoría cubierto (nubes medias)	baja	media	homogénea
10	En su mayoría cubierto (nubes medias)	baja	media	tarde
11	Totalmente cubierto (nubes bajas)	nula	baja	homogénea

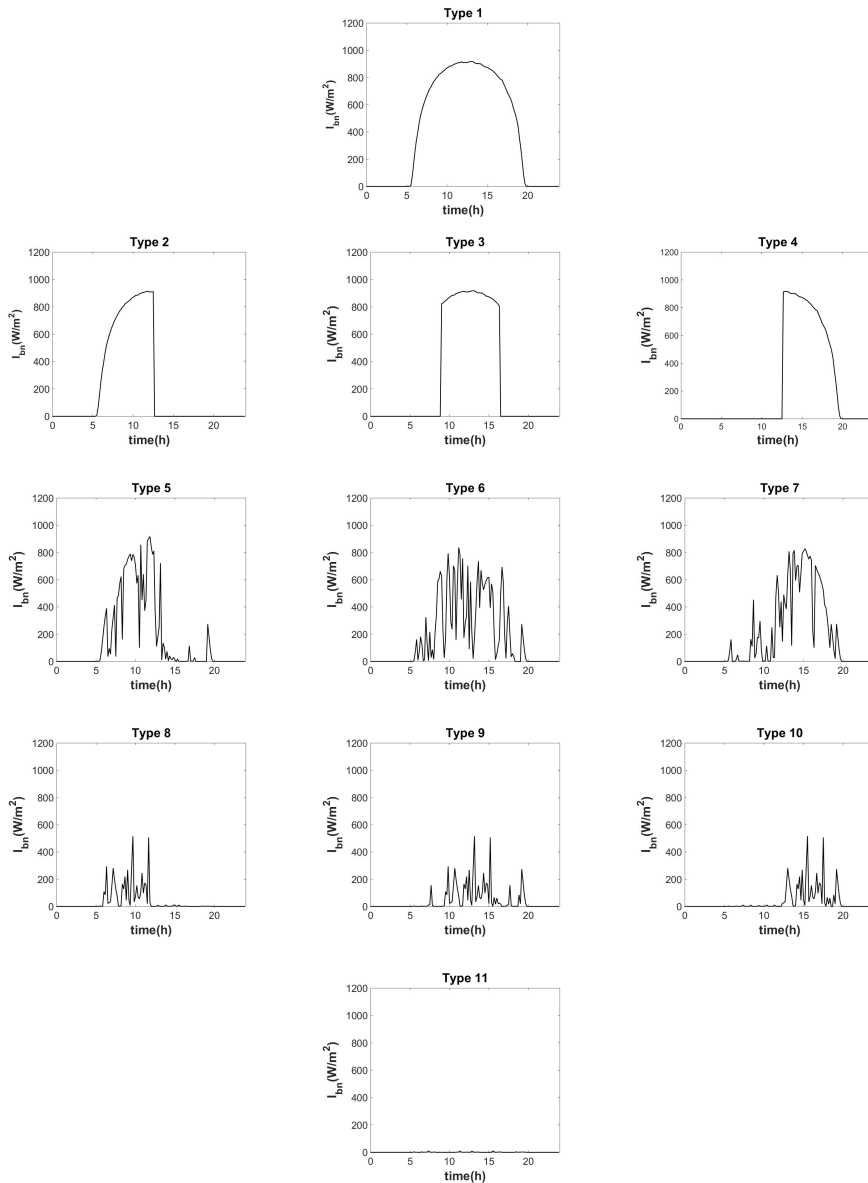


Figura 4-8. Ejemplos de las curvas correspondientes a los tipos de días definidos en la Tabla 4-2. . Estas figuras son las utilizadas por los expertos como referencia para realizar la clasificación visual.

4.4. Caracterización del tipo de día en función de la producción

Una vez clasificados los días a partir de la energía, la variabilidad y la distribución temporal de la DNI, se estudia cada tipo de día desde el punto de vista de la producción de una central termosolar cilindroparabólica (CCP). Para ello, se presenta un nuevo índice que permite caracterizar la producción de una central termosolar eliminando las dependencias estacionales (horas de sol y ángulo de incidencia). Este índice permite evaluar la capacidad de aprovechamiento del recurso de una central termosolar respecto al aprovechamiento del recurso por esa misma central en condiciones de cielo despejado. A partir de este índice, se analiza la influencia de las características (energía, variabilidad y distribución temporal) de las curvas de DNI en la producción de la central y se caracteriza la capacidad de aprovechamiento del recurso para producir en función de cada tipo de día obtenido al aplicar la metodología descrita en el apartado 4.3. Para dicho objeto, inicialmente, se presenta y se evalúa el índice k_p para 16 años de medidas registradas en la estación del GTER en Sevilla de dos centrales CCP con distinta configuración. A continuación, a partir de los 16 años de k_p se analiza la relación entre este índice y los tres índices empleados en el apartado 4.3 para representar a las propiedades de las curvas de DNI. Por último, se caracterizan estadísticamente los valores de k_p para cada tipo de día obtenido al aplicar la metodología descrita en el apartado 4.3.

4.4.1. Definición del índice de producción relativa de cielo despejado.

Se define el índice k_p como el cociente entre la producción de electricidad de la central termosolar durante un periodo (diario, mensual, anual) y la producción de electricidad de la central durante el mismo periodo (diario, mensual, anual) en condiciones de cielo despejado:

$$k_p^{\Delta t} = \frac{P^{\Delta t}}{P_{CS}^{\Delta t}} \quad \text{Ec. (4-13)}$$

donde $P^{\Delta t}$ es la energía eléctrica producida por la central durante el periodo Δt y $P_{CS}^{\Delta t}$ la energía eléctrica producida por la central durante el mismo periodo, pero en condiciones de cielo despejado. La definición adimensional de este índice elimina los efectos estacionales y permite comparar el aprovechamiento de la radiación desde el punto de vista de la producción de diferentes días del año o de centrales con diferente configuración. Este índice se puede definir utilizando la producción bruta

o la producción neta de la central. Incluso se podría definir un índice similar para evaluar la energía térmica captada por el campo solar. En este trabajo, se utiliza la energía eléctrica bruta para analizar las características del recurso solar desde el punto de vista de la producción de una central termosolar y comparar el aprovechamiento de la radiación de dos centrales CCPs con diferente configuración.

4.4.2. Datos de entrada

La base de datos utilizada para la simulación del comportamiento de las centrales CCPs abarca 16 años de medidas de DNI registrados en Sevilla por la estación radiométrica del GTER durante el periodo 2000-15. Las medidas han sido promediadas cada 10 min para su implementación. Los tres años adicionales utilizados en la fase de evaluación de la producción han pasado los mismos controles de calidad (Moreno-Tejera et al., 2015), y se han completado aplicando el mismo procedimiento (Moreno-Tejera et al., 2016) que el utilizado en la fase de clasificación para el período 2000-12 (apartados 4.1.1 y 4.1.2).

4.4.3. Modelo de centrales y simulación

Para modelar las centrales cuya producción se va a analizar, se seleccionan como referencia dos centrales CCPs, actualmente en operación, localizadas en el sur de España: Andasol 3 (NREL, 2013) y Solnova 1 (NREL, 2017). Ambas centrales tienen una capacidad instalada de 50 MWe de potencia nominal neta y utilizan la tecnología cilindroparabólica para captar la energía solar. Los principales parámetros utilizados para modelar ambas centrales se muestran en la Tabla 4-3.

Tabla 4–3. Principales parámetros técnicos utilizados en SAM para modelar las centrales semejantes a Solnova1 (PTP0) y Andasol3 (PTP7.5).

Parámetros	PTP0	PTP7.5
Potencia neta (MWe)	50	50
Tipo de colector	EuroTrough ET150	EuroTrough ET150
Tipo de receptor	Schott PTR70 2008	Schott PTR70 2008
Número de lazos	90	156
Colectores por lazo	4	4
Área de apertura del campo solar (m ²)	300 000	510 120
HTF	Therminol VP-1	Therminol VP-1
Temperatura de diseño a la salida del campo	391	391
Capacidad de almacenamiento (h equivalentes a plena carga)	0	7.5
Tipo de almacenamiento térmico	-	2 tanques
Fluido de almacenamiento térmico	-	Hitec Solar Salt

Para simular el comportamiento de las centrales modeladas durante el periodo 2000-2015 se utiliza el software libre de simulación SAM (versión 2017.1.17) (NREL, 2010) y los valores de DNI 10-min de la base de datos del GTER. De aquí en adelante, la central modelada como Andasol 3 se denominará PTP7.5 y la central modelada como Solnova 1 PTP0.

Para evaluar el índice diario k_p^d , se debe computar la energía eléctrica diaria producida por cada central. La estrategia de operación del sistema de almacenamiento de PTP7.5 se ha definido para producir a máxima potencia el máximo tiempo posible. El propósito de esta estrategia es utilizar toda la energía remanente en el sistema de almacenamiento después del ocaso con el objetivo de vaciar el sistema de almacenamiento antes del orto del día siguiente. La gran capacidad del almacenamiento de la central PTP7.5 permite seguir generando electricidad durante las primeras horas del día siguiente durante las estaciones más favorables (primavera, verano y otoño). Para computar sólo la energía producida con la radiación solar perteneciente a ese mismo día y no influir en la producción del día siguiente, se computa la energía producida diaria desde las 4:00 h de cada día hasta las 3:50 h del día siguiente. Como se muestra en la Figura 4-9, en el día más favorable, el 21 de junio, en condiciones de cielo despejado envolvente antes de las 4:00 h del día 22 se vacía el tanque caliente:

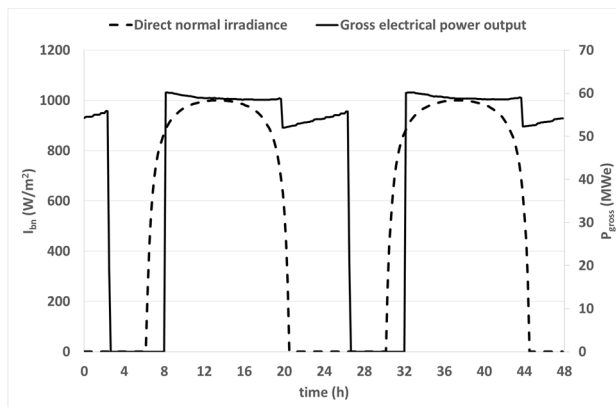


Figura 4-9. DNI y potencia eléctrica bruta producida por la central CCP PTP7.5 el 21 y 22 de junio en condiciones de cielo despejado envolvente.

4.4.4. Evaluación del índice de producción relativa de cielo despejado

A partir de los 16 años de medidas diez-minutal de DNI registrados en Sevilla se simula el comportamiento de las dos centrales CCPs definidas en el apartado 4.4.3, emplazadas en Sevilla. Adicionalmente, se construye un año con 365 días de cielo

despejado envolvente (apartado 4.3.1) para simular el comportamiento de ambas centrales en condiciones de cielo despejado. Finalmente, se calculan y se comparan los 16 años de valores diarios del índice de producción relativa de cielo despejado para las dos centrales CCPs con diferente configuración. El análisis de ambas series permite identificar las capacidades de este nuevo índice. En la Figura 4-10 se muestra una descripción esquemática del procedimiento seguido en este apartado.

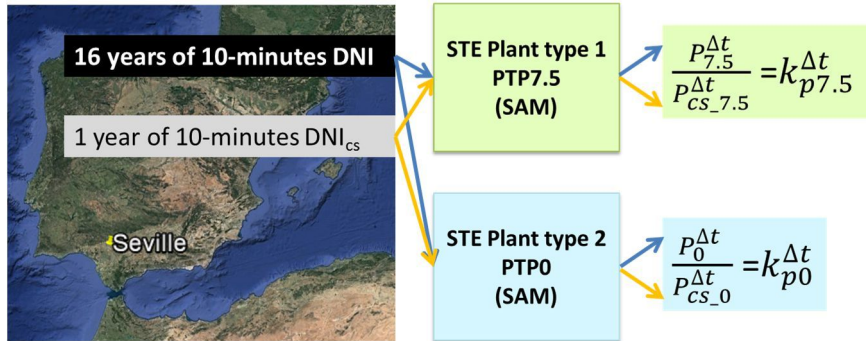


Figura 4-10. Descripción esquemática del procedimiento seguido para la evaluación del k_p diario: Primero se simula el comportamiento de un central CCP con y sin almacenamiento térmico durante 16 años a partir de las medidas de DNI diez-minutal de la base de datos del GTER y un año diez-minutal de valores de DNI en condiciones de cielo despejado (DNI_{cs}). A continuación, se calculan, se analizan y se comparan los k_p diarios de ambas centrales durante dicho periodo.

4.4.5. Caracterización del índice de producción relativa de cielo despejado de los tipos de días

En este apartado, se estudia la relación entre el índice k_p y los 3 índices seleccionados como más adecuados para clasificar los días (metodología descrita en el apartado 4.3) según la energía (k_b), la variabilidad (VI o POP_D) y la distribución en el tiempo de la DNI (F_m). Para ello se cuenta con 16 años de valores diarios de los tres índices de radiación seleccionados para representar estas propiedades y 16 años de valores diarios de k_p de las dos centrales CCPs modeladas. Además, se calculan los valores de F_m y k_p para ambas plantas a partir de una serie de 16 años de curvas de DNI

simétricas respecto al medio día solar a los 16 años de curvas registradas en el emplazamiento y se analiza la relación entre F_m y k_p de ambas series para evitar que los tipos de días característicos de la climatología del emplazamiento influyan en los resultados obtenidos. Finalmente, a partir de la clasificación de tipos de días que se obtiene al aplicar la metodología de clasificación propuesta en la presente Tesis, se caracteriza el índice k_p para cada tipo de día. La Figura 4-11 muestra una descripción esquemática del procedimiento seguido en el presente apartado.

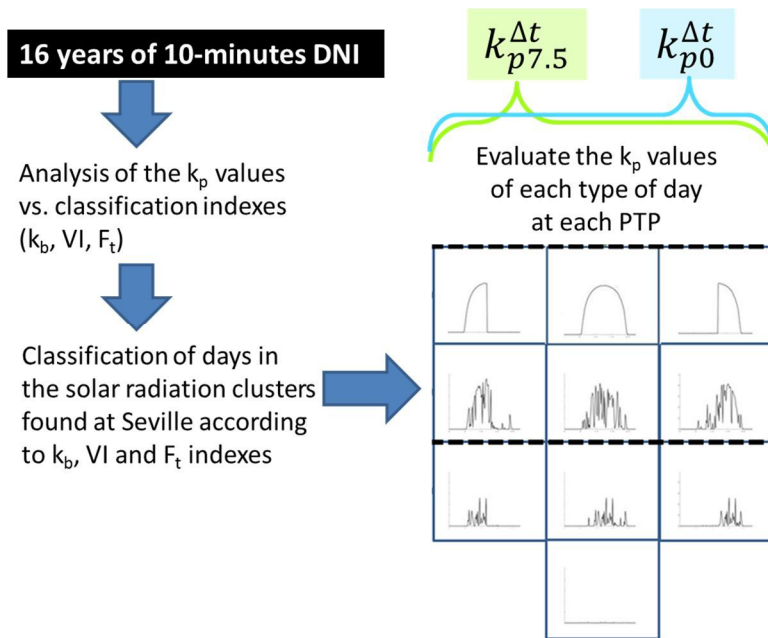


Figura 4-11. Descripción esquemática del procedimiento seguido para la caracterización de los tipos de días desde el punto de vista de la producción de las centrales: Primero, se estudia la relación entre k_p y k_b , VI y F_m . Estos índices son los utilizados para representar las características de las curvas de la DNI (energía, variabilidad y distribución temporal) en la metodología de clasificación. A continuación, se caracterizan los valores del índice k_p de cada uno de los tipos de días resultantes de aplicar la metodología de clasificación de acuerdo a estas tres características de la DNI.

5. RESULTADOS

En este apartado se presentan los resultados principales de los artículos que conforman la presente Tesis. Los resultados completos pueden consultarse en los artículos originales que se encuentran adjuntos en los *Anexos*.

5.1. Base de datos

Del control de calidad, la identificación de errores y el rellenado de huecos de la base de datos del GTER se obtienen los resultados que se muestran a continuación.

5.1.1. Control de calidad

La base de datos de medidas radiométricas del GTER se ha etiquetado en función de los procedimientos de control de calidad propuestos por la BSRN. Posteriormente, a partir de los dos tipos de gráficas propuestas en la metodología descrita en el apartado 4.1.1 se visualizan las etiquetas de los datos, con el objeto de localizar errores en el registro del tiempo y en las medidas, tratando de identificar el tipo de error (Moreno-Tejera et al., 2015). En la Figura 5-1, se muestran los resultados para las variables GHI y DNI del año 2001 de la base de datos del GTER. La gráfica de la izquierda de la Figura 5-1 representa los datos medidos de GHI frente a los calculados a partir de sus componentes. En la Figura se observan dos líneas negras que se utilizan como referencia para localizar las diferencias superiores al 8 %. Las otras dos figuras (la figura central y la situada a su derecha) representan las etiquetas del último procedimiento de la BSRN que cumple cada dato de cada variable. En estas dos figuras también se muestran en color azul unas líneas que representan el orto, el ocaso y el medio día solar de cada día del año. Estas líneas sirven como referencia para localizar los datos erróneos.

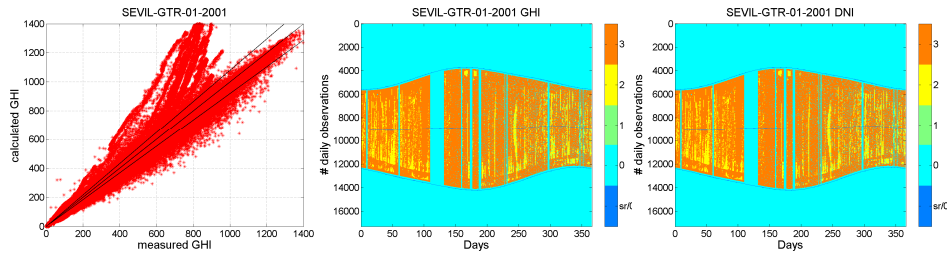


Figura 5-1. Resultados del control de calidad del año 2001 de la base de datos del GTER.

La figura más a la izquierda representa los valores medidos de GHI frente a los calculados a partir de sus componentes. La figura central y la situada a su derecha representan las etiquetas de los procedimientos del control de calidad de la BSRN de las medidas de GHI y DNI. Las líneas del orto el ocaso y el medio día solar (en color azul) ayudan a analizar los resultados.

Los ejemplos presentados en (Moreno-Tejera et al., 2015), muestran que los nuevos tipos de figuras propuestas son útiles para encontrar e identificar los errores frecuentes tales como los desfases de registros del tiempo, los cambios en la frecuencia de registro, el mal sombreado del piranómetro que mide la DHI, el desalineamiento del pirheliómetro, o los “offsets” nocturnos y diurnos.

5.1.2. Rellenado de huecos

Del total de días registrados durante el periodo 2000-12, sólo se clasificaron como descartados un 9% y un 19% como corregibles. El resto de días (72%), fueron registrados sin ninguna anomalía. La función de densidad de probabilidad de los índices k_{to} y k_t de la base de datos completada (después de aplicar el procedimiento de relleno de huecos) y de los días válidos de la base de datos original (sin completar) se muestran en la Figura 5-2 y en la Figura 5-3. Se observa cómo la serie de datos completada, conserva las características estadísticas de los datos registrados sin ninguna anomalía y sin mostrar diferencias apreciables en ambos índices.

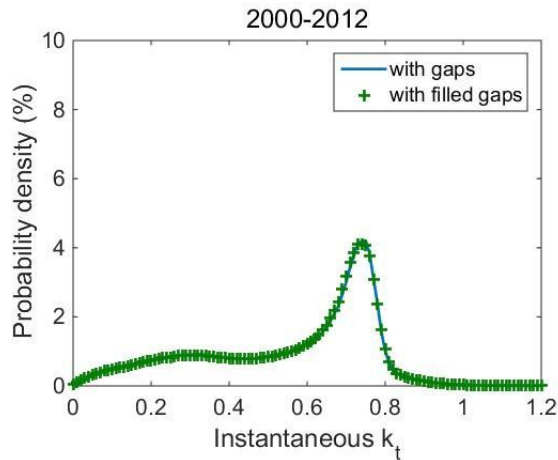


Figura 5-2. Función de densidad de probabilidad de los valores de k_t calculados a partir de la base de datos antes ("with gaps") y después ("filled") de aplicar el procedimiento de rellenado de huecos.

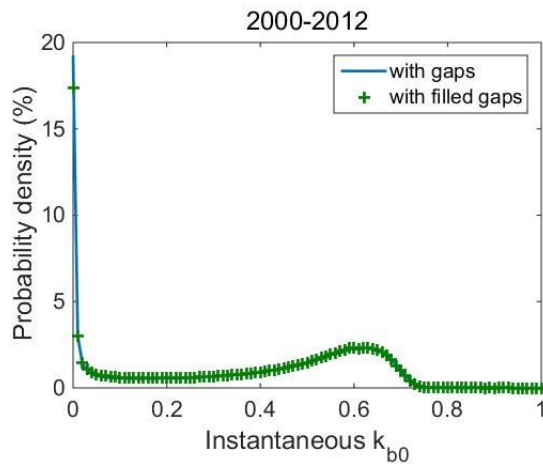


Figura 5-3. Función de densidad de probabilidad de los valores de k_{b0} calculados a partir de la base de datos antes ("with gaps") y después ("filled") de aplicar el procedimiento de rellenado de huecos.

5.2. Caracterización estadística de la radiación solar en Sevilla

A continuación, se presentan los resultados obtenidos del análisis estadístico de la base de datos radiométricos en las diferentes escalas temporales.

5.2.1. Caracterización estadística de los valores instantáneos, horarios y diez-minutal.

En la Figura 5-4 se representan las funciones de densidad de probabilidad de los valores instantáneos de k_t y k_b para diferentes rangos de masa óptica relativa de aire. La distribución de frecuencia de los valores instantáneos de k_t para los diferentes espesores de masa óptica de aire muestra dos estados predominantes, despejado y cubierto, en concordancia con los resultados de otros autores (Jurado et al., 1995; Suehrcke and McCormick, 1988). El valor máximo de k_t para los menores espesores de masa de aire estudiados es aproximadamente de 0.76. Este máximo decrece cuando la masa óptica de aire crece. El segundo estado predominante, que se obtiene en los valores bajos de k_t , está menos definido debido a la gran variedad de nubes que generan la componente difusa.

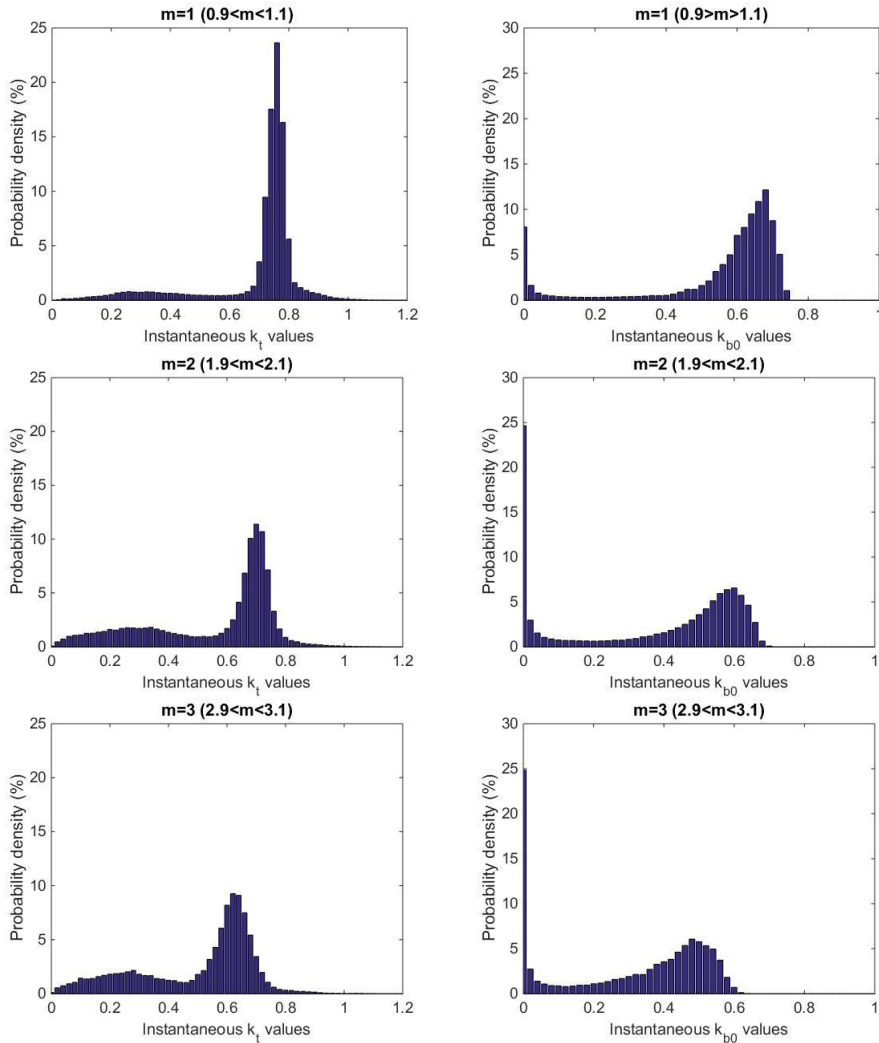


Figura 5-4. Funciones de densidad de probabilidad de los valores instantáneos de k_t y k_{b0} para diferentes rangos de masa óptica relativa del aire.

Las curvas de distribución de los valores de k_b instantáneos también muestran un comportamiento bimodal. En este caso, ambos picos de la distribución son pronunciados. El primer pico, se encuentra localizado entre los valores de k_{b0} 0 y 0.2,

correspondiendo a condiciones de cielo cubierto en los que la DNI suele ser nula o casi nula. Este pico es el más pronunciado de los dos, excepto para los valores de masa óptica de aire menores. El pico situado más a la derecha se encuentra localizado entre los valores altos de k_{b0} , correspondiendo a condiciones de cielo despejado. Este pico tiene una forma menos pronunciada, debido a las diferentes condiciones de turbidez atmosférica que se pueden encontrar en condiciones de cielo despejado; al igual que ocurría en k_t , este pico decrece cuando la masa de aire aumenta. El efecto de la turbidez atmosférica se incrementa al aumentar la masa de aire. Un efecto similar se observa en las condiciones de cielo cubierto, ya que el pico de la izquierda también aumenta al aumentar la masa óptica de aire debido al mayor efecto bloqueo de las nubes.

En la Figura 5.5 se representa la función de densidad de probabilidad de los índices k_t (a la izquierda) y k_{b0} (a la derecha) calculados con los valores de irradiancia registrados cada 5s, promediados cada 10 min y cada hora. Se observa como las funciones de distribución de los valores instantáneos y los promediados cada 10 min son muy similares, mostrando el mismo comportamiento bimodal con ligeras diferencias, que en el caso del índice k_{b0} son inapreciables. Sin embargo, las diferencias entre las funciones de distribución de los valores instantáneos y horarios son más significativas, especialmente en el caso del índice k_t , donde el primer pico de la función bimodal desaparece. Promediar los valores instantáneos en intervalos horarios resulta en un incremento de frecuencia de estados de cielos intermedios (ni completamente claro, ni completamente nublado), con una curva de distribución de forma triangular con un pico máximo menor que a mayores frecuencias.

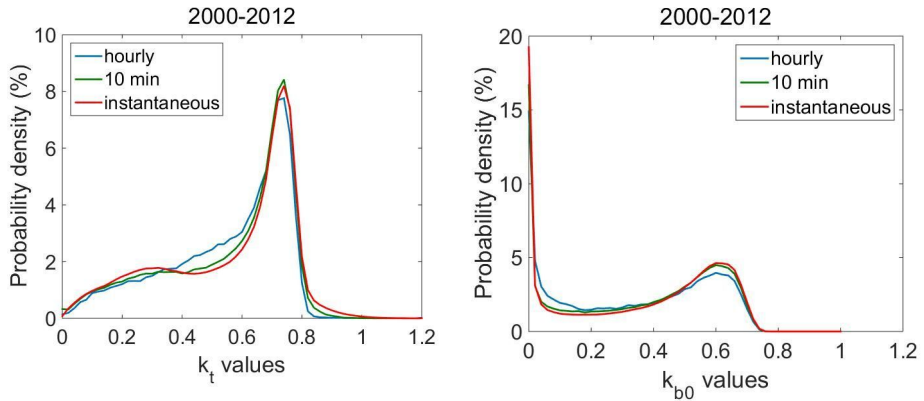


Figura 5-5. Función de densidad de probabilidad de los índices k_t (izquierda) y k_{b0} (derecha) obtenidos a partir de los valores de GHI y DNI cada 5s, promediados cada 10 min y cada hora.

Los valores mayores de k_t (mayores que 1.0 en algunos casos) pueden atribuirse al “efecto lupa” o “enhancement effect” (Tapakis and Charalambides, 2014), registrándose valores instantáneos de hasta 1583 W/m² de GHI. El valor del P99 (percentil 99) de la GHI es de 1009 W/m².

Las funciones de densidad de probabilidad de los valores de GHI y DNI promediados cada 10 min y una hora se representan en la Figura 5-6. Estas figuras, al igual que las correspondientes funciones de densidad acumulada, son útiles para seleccionar el punto de diseño de una central solar (generalmente correspondiente al valor P90 o P95). Los máximos de las funciones de densidad de GHI y DNI para los valores 10 min se encuentran localizados en los valores más bajos. En el caso de la DNI, se observa un segundo pico entorno a los 850 W/m². En Sevilla, cuando el día está totalmente despejado se registran valores de esta magnitud incluso a alturas solares de 20°. Es también destacable que más del 66% de los valores de DNI son mayores de 300 W/m², a menudo considerado como límite para la operación de las centrales termosolares.

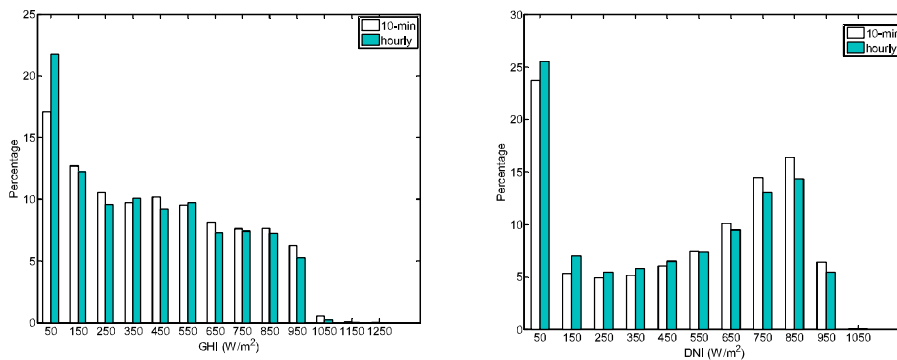


Figura 5-6. Función de densidad de probabilidad de los valores 10 min y horarios de GHI y DNI.

En la Figura 5-6, se observa que, en el caso de la GHI, las mayores diferencias entre ambas resoluciones temporales se encuentran en los valores bajos de irradiancia. Estas diferencias se compensan en los otros niveles sin una tendencia clara. Sin embargo, las mayores diferencias en la distribución de los valores de DNI se observan en los valores altos, donde los valores horarios muestra una menor frecuencia que los diez-minutal. El 47.4 % de los valores diez-minutal son superiores a 600 W/m^2 , mientras que para valores horarios esta fracción es del 42.4%. En general, el uso de valores horarios suaviza los gradientes de DNI.

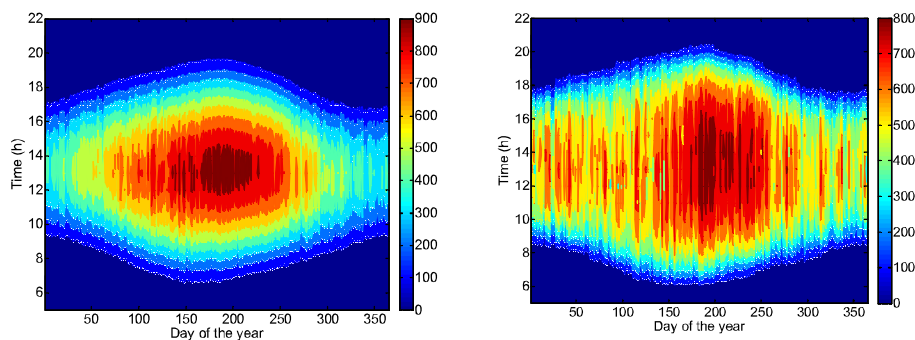


Figura 5-7. Valores medios horarios de GHI (izquierda) y DNI (derecha) para cada día del año.

En la Figura 5-7 se representan los valores medios horarios de toda la base de datos de GHI y DNI. Los mayores niveles de radiación en ambos casos se alcanzan en julio. El valor medio máximo horario de GHI es 964 W/m^2 y se alcanza al medio día solar en un día cercano al solsticio de verano, cuando la altura solar alcanza un máximo de 76.1° . En torno a este punto, los valores medios horarios de GHI parecen distribuirse casi gradualmente de forma simétrica respecto al medio día solar, decreciendo cuando los días se alejan del solsticio de verano. Este tipo de distribución está fuertemente relacionada con la evolución del ángulo cenital, ya que la radiación global se mide sobre superficie horizontal. Por otro lado, los valores horarios de DNI son menos homogéneos, alcanzando valores elevados en todas las estaciones, aunque observándose claramente una mayor frecuencia de valores altos cerca del solsticio de verano. También merece la pena destacar el número de horas al día con DNI por encima de los 600 W/m^2 (alrededor de 8 horas al día en los meses de verano).

1.1.1 Caracterización estadística de los valores diarios

Como se observa en la Figura 5-8 (a la izquierda), la función de densidad de probabilidad de k^d_t muestra un claro máximo entre los valores 0.65 y 0.72 que corresponde a más del 35 % de los días. Este máximo se localiza justo a partir del límite que diferencia los días despejados de los nublados en la clasificación de tipos de días a partir de los valores de k^d_t . En la Figura 5-8 (a la derecha), también se representa el número medio mensual de días nublados, parcialmente nublados y claros, considerando como límites para la clasificación los valores 0.3 y 0.65 (ver apartado 4.2.2). Según dicha metodología, el mayor número de días nublados (6 días) y parcialmente nublados (19 días), se alcanza en diciembre. En función de este resultado, este mes se considera el más adecuado para realizar una parada en la producción de una central solar para realizar operaciones de mantenimiento programadas. Por otro lado, junio, julio y agosto son los meses con mayor número de días claros, más de 20 días en todos los casos. Durante estos meses, los días nublados son poco frecuentes.

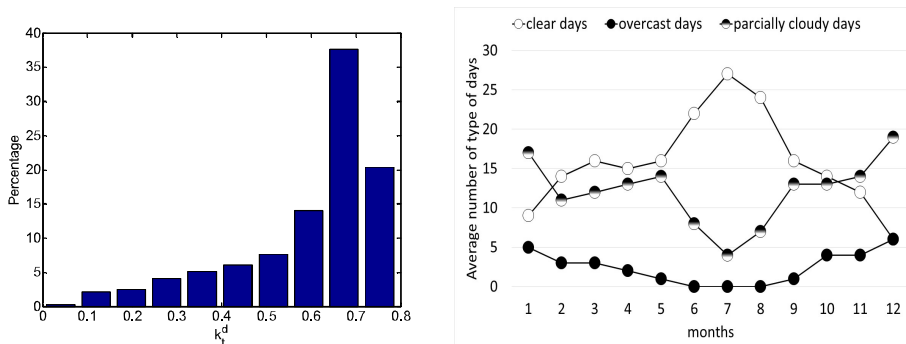


Figura 5-8. Función de densidad de probabilidad de k_t^d y número de tipos de día promedios mensuales de acuerdo al estado del cielo (despejado, cubierto, parcialmente cubierto).

En el Anexo B, se muestra la caracterización estadística de los valores diarios de cada año de DNI y GHI mediante una representación Box & Whisker.

5.2.2. Caracterización estadística de los valores mensuales y anuales

Los valores mensuales y anuales promedios diarios de GHI, DNI y k_t^d para los 13 años analizados, se muestran en la Tabla 5-1. La desviación máxima y mínima respecto al valor medio y la desviación estándar mensual también se incluyen en esta tabla.

El valor medio diario de la GHI para el periodo de análisis es muy estable, como se observa en la Tabla 5-1. La desviación máxima es aproximadamente del 5%. La desviación estándar de los valores medios mensuales es 2.05 kWh/m², lo que implica una alta variabilidad intermensual. El valor máximo medio diario de GHI se obtiene en julio (7.81 kWh/m²) y el mínimo en diciembre (2.17 kWh/m²) con una diferencia del 72% entre ellos respecto al valor mayor. Realizando el mismo análisis a los valores de k_t^d se obtienen diferencias menores, con un máximo de 0.69 y un mínimo de 0.51 para los mismos meses. El valor medio diario anual del periodo completo de k_t^d es 0.61. Esto refleja la predominancia de cielos despejados y parcialmente nublados en el clima de Sevilla. La causa principal de la alta variabilidad en los valores de GHI, ya comentada en el análisis de los valores horarios, es la variación de la altura solar y además, en el caso de los valores diarios, a la diferencia de horas de sol a lo largo del año.

Respecto a los valores anuales de DNI, el valor medio diario anual es 5.68 kWh/m², con mayores variaciones interanuales. El valor máximo medio mensual se obtiene en julio (8.76 kWh/m²) y el mínimo en diciembre (3.42 kWh/m²) con una desviación estándar mensual de 1.75 kWh/m². En este caso, la diferencia máxima entre los meses es del 61% con respecto al mayor valor, mostrando la importancia de la distribución mensual.

El valor medio mensual k^d_t excede el límite propuesto de cielo claro en el apartado 4.2.2 durante los meses de junio, julio y agosto. Estos meses son los más estables en Sevilla.

Tabla 5–1. Valores mensuales y anuales medios diarios de GHI, DNI y k^d_t en Sevilla. Las columnas “Diferencia” indican las diferencias mínimas y máximas con respecto al promedio

Meses	Rad. Global Horizontal			k^d_t	Rad. Directa Normal				
	Med (kWh/m ²)	Diferencia Mín (%) Máx (%)		Med	Diferencia Míni Máx			Diferencia Mín (%) Máx (%)	
Enero	2.56	-21%	21%	0.54	0.43	0.66	4.03	-32%	51%
Febrero	3.48	-21%	26%	0.56	0.44	0.71	4.62	-40%	61%
Marzo	4.58	-11%	11%	0.57	0.50	0.63	4.86	-21%	31%
Abril	5.80	-12%	13%	0.59	0.53	0.67	5.74	-28%	31%
Mayo	6.85	-14%	10%	0.62	0.53	0.68	6.59	-30%	21%
Junio	7.65	-6%	6%	0.66	0.62	0.70	7.88	-13%	17%
Julio	7.81	-3%	5%	0.69	0.67	0.73	8.76	-10%	8%
Agosto	6.96	-4%	7%	0.68	0.65	0.71	7.80	-7%	15%
Septiembre	5.33	-10%	8%	0.61	0.56	0.67	5.86	-27%	22%
Octubre	3.76	-14%	17%	0.56	0.48	0.61	4.61	-30%	25%
Noviembre	2.73	-13%	15%	0.54	0.46	0.61	3.93	-22%	28%
Diciembre	2.17	-23%	18%	0.51	0.39	0.60	3.42	-38%	39%
Anual	4.98	-3%	5%	0.61	0.59	0.64	5.68	-6%	8%
Desv. Est.	2.05	-	-	0.06	-	-	1.75	-	-

5.3. Clasificación de días en función de las curvas de DNI

En este apartado se presentan los resultados de aplicar la metodología de clasificación de días en función de la energía, la variabilidad y la distribución en el tiempo de la DNI descrita en el apartado 4.3 a las medidas de DNI registradas en Sevilla durante el periodo 2000-12.

5.3.1. Número y tipos de días en función de los índices

Como ya se mencionó en el apartado 4.3.3.1, el resultado final que se obtiene al aplicar el algoritmo PAM depende de la selección de los medoides iniciales. Para asegurar que se obtiene la solución óptima de cada número de clústeres, el algoritmo se ejecuta 1000 veces para cada caso y se evalúa el índice SH y CH de cada iteración. Este proceso se ha llevado a cabo barriendo un rango de número de clústeres desde 3 a 20. Como se discute en el apartado 4.3.3.4, 11 tipos de día parece, a priori, ser suficientes para identificar todos los casos de interés a partir de los índices de caracterización seleccionados. Con el objetivo de analizar la tendencia de los índices de validación y ayudar a localizar mejor el *knee point* se amplía el rango de estudio abarcando desde 3 hasta 20 tipos de días. En la Figura 5-9 se representan los valores máximos, mínimos y medianos de los índices SH y CH de los 1000 casos ejecutados para cada número de clúster y grupo de índices. El máximo valor de ambos índices indica el número óptimo de clústeres en términos de máxima similitud y compacidad. Los valores mínimos y medianos proporcionan información sobre la facilidad de encontrar el óptimo en un número de iteraciones o, lo que podría considerarse, la incertidumbre en encontrar el óptimo si se ejecuta el algoritmo una sola vez.

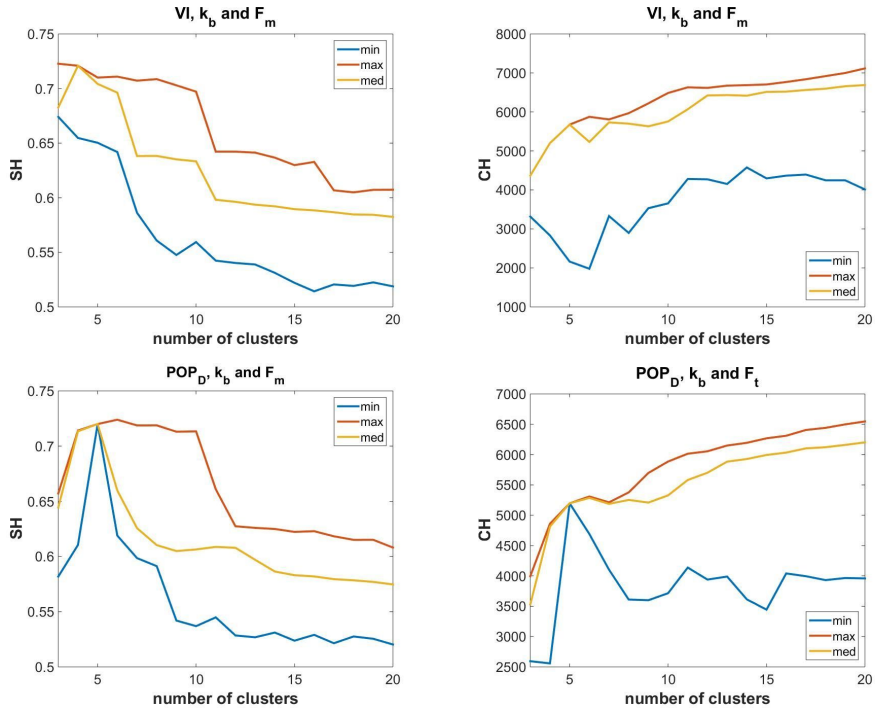


Figura 5-9. Representación de los valores máximos, mínimos y medianos de SH y CH de las 1000 iteraciones en función al número de clúster, para los dos grupos de índices considerados.

Como se observa en la Figura 5-9, ambos grupos de índices muestran un comportamiento monótono creciente del índice CH y casi un comportamiento monótono decreciente en el caso del índice SH para el grupo de índices VI. Estos resultados dificultan la localización del número óptimo de clústeres. También resulta difícil, encontrar el óptimo absoluto en el caso del índice SH del grupo de índices POP_D, ya que un rango de números de clústeres alcanza valores muy similares en torno al máximo. Por lo tanto, se utiliza el *L-method* para ayudar a seleccionar el número de tipos de día óptimo más razonable en cada caso. En la Figura 5-10, se observan las gráficas con todos los posibles pares de rectas ajustadas con los puntos localizados a ambos lados de cada posible candidato a ser el *knee point*, así como la evolución del RMSE_c para cada grupo de índices de la curva de valores máximos de CH y SH.

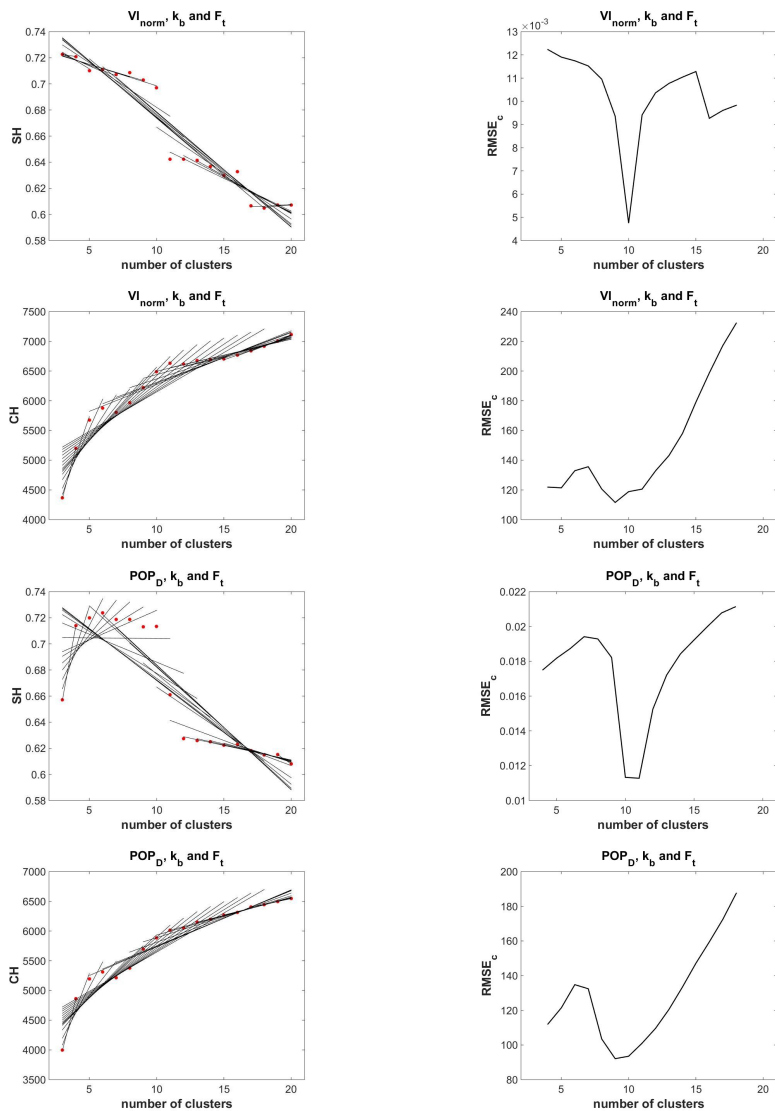


Figura 5-10. A la izquierda, representación de todos los posibles pares de rectas de ajuste de los valores máximos de SH y CH para los dos grupos de índices considerados para buscar el "knee point", o número de clúster más óptimo al aplicar el L-method. A la derecha, se representa el valor de RMSE_c respecto al número de clústeres.

En el caso del índice CH, el valor mínimo de $RMSE_c$ se localiza en 9 clústeres para el caso del grupo de índice VI y en 11 para el grupo de índices POP_D , aunque para 10 clústeres el valor de $RMSE_c$ es muy similar. En el caso del índice SH, ambos grupos de índices, alcanzan claramente el valor mínimo para 10 clústeres. Por tanto, se considera 10 clústeres como el número más razonable.

En la Figura 5-11 se representan los valores de los índices del periodo 2000-12 para los dos grupos de índices considerados. Los puntos que pertenecen al mismo clúster se simbolizan con el mismo color. Se observa que el número de clústeres para cada valor de F_i es diferente en cada grupo de índices. En el grupo POP_D , hay un reparto desequilibrado de número de clústeres entre los tipos de día con F_i igual a 1 (5 clústeres) y los días con F_i igual a 0 (2 clústeres). En el caso del grupo VI, el número de clústeres entre los tres tipos de distribución temporal definidos está más equilibrado, 3, 3 y 4 clústeres para F_i igual a 1, 0.5 y 0, respectivamente.

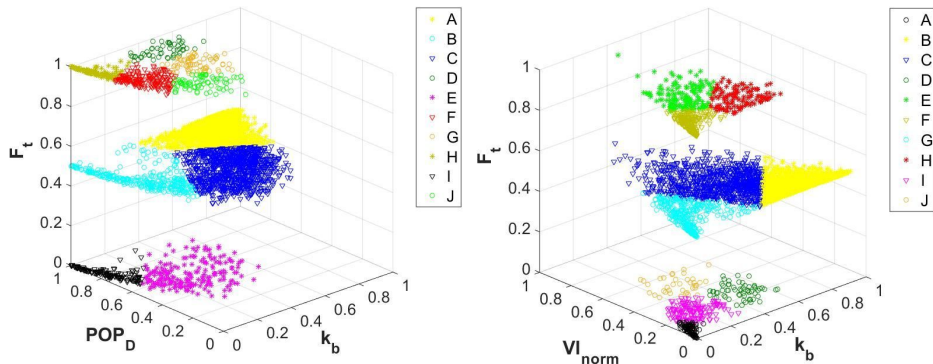


Figura 5-11. Representación de los valores del POP_D (a la izquierda) y de VI (a la derecha) frente a k_b y F_i del periodo 2000-13. Los diferentes símbolos identifican los diferentes clústeres encontrados en cada caso.

En la Figura 5-12 se representan las curvas diarias de DNI de los medoides del grupo de índices del POP_D . Comparando estas curvas con las utilizadas como referencia por los expertos para clasificar visualmente los días (ver Figura 4-8) se observan sólo algunas similitudes. Sin embargo, en el caso del grupo de índices VI las similitudes son bastante significativas (ver Figura 5.13). Con la excepción de dos tipos de días (tipo 2 y tipo 11), las características de los días de referencia pueden considerarse las

mismas: 3 días con energía media o alta, baja variabilidad y diferente distribución temporal; 3 días con energía media, alta variabilidad y diferente distribución temporal; y 3 días con baja energía, variabilidad media y diferente distribución temporal. Cabe destacar que en la Tabla 4–2, hay definidos 11 tipos de días, mientras que el número óptimo de días encontrado ha sido 10. En la clasificación obtenida a partir del algoritmo PAM con el grupo de índices VI, el tipo de día identificado por tener una alta energía y una baja variabilidad etiquetado en la Tabla 4–2 como, tipo “1”, ha sido absorbido por el clúster de tipo “B”, como se observa en la Figura 5-13. La otra excepción es el *medoide* que debería representar el día de tipo “11” (el *medoide* de tipo “A” de la Figura 5-13). Este *medoide* tiene muy baja energía y variabilidad, pero su F_i es 0, lo que implica que la energía principalmente se concentra por la mañana. Cuando la DNI es tan baja, la distribución de la energía en el tiempo no es significativa. De hecho, una decisión visual sobre la distribución del tiempo en este tipo de días no es sencilla y desde el punto de vista de las termosolares este tipo de día no es relevante. Por lo tanto, este satisfactorio resultado posiciona al VI como el mejor índice para definir la variabilidad de los días de las dos opciones planteadas, aplicando la metodología propuesta.

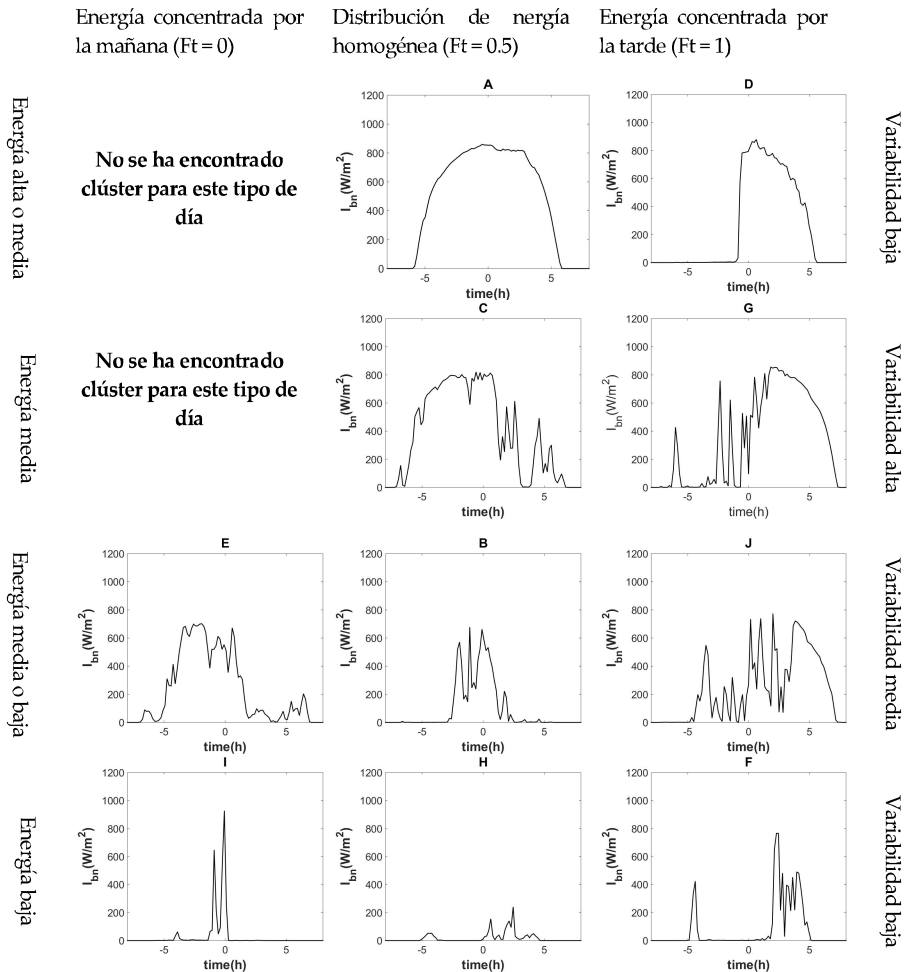


Figura 5-12. Curvas de DNI de los días correspondientes a los medoides de los clústeres del grupo de índices POP_D. Las etiquetas coinciden con las etiquetas de los clústeres de la Figura 5-11. Las figuras de la columna izquierda tienen distribución de energía concentrada principalmente por la mañana. Las de la columna central, energía homogénea a lo largo del día. Las de la columna derecha, energía concentrada principalmente por la tarde. La distribución de las figuras por filas está también relacionada con características similares de energía y variabilidad intradiaria de la DNI.

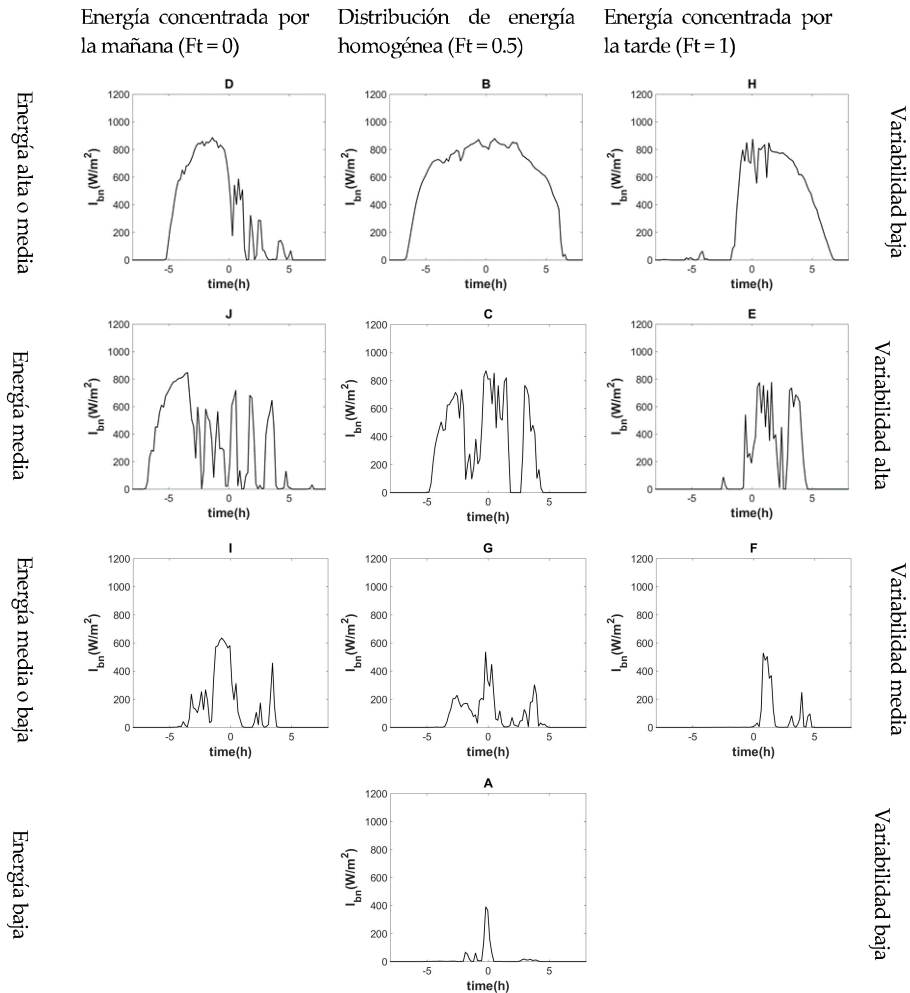


Figura 5-13. Curvas de DNI de los días correspondientes a los medoides de los clústeres del grupo de índices VI. Las etiquetas coinciden con las etiquetas de los clústeres de la Figura 5-11. Las figuras de la columna izquierda tienen distribución de energía concentrada principalmente por la mañana. Las de la columna central, energía homogénea a lo largo del día. Las de la columna derecha, energía concentrada principalmente por la tarde. La distribución de las figuras por filas está también relacionada con características similares de energía y variabilidad intradiaria de la DNI.

5.3.2. Validación de resultados

Una vez seleccionado el grupo de índices VI como el más adecuado para clasificar los días, se compara la clasificación de los días del periodo 2010-11 obtenida con este grupo de índices con la clasificación realizada por los expertos (ver apartado 4.3.3.4). Para la comparación, se selecciona una serie de referencia formada por los 702 días en que al menos tres de los expertos han coincidido en la clasificación.

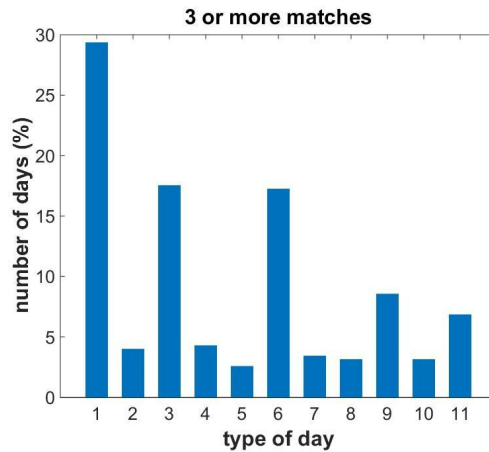


Figura 5-14. Porcentaje de días que han sido clasificados en el mismo tipo de día por 3 o más expertos.

Una vez seleccionada la serie de días que forman la *clasificación de referencia*, se compara la clasificación de dichos días con la clasificación de esos mismos días obtenida al aplicar la metodología de *clustering*. Para ello, inicialmente, se buscan similitudes entre las gráficas de los *medoides* de los clústeres (Figura 5-13) con los tipos de días entregados a los expertos como referencia (Figura 4-8) (B=1 y 3; D=2; H=4; J=5; C=6; E=7; I=8; G=9; F=10; A=11). Como los días del tipo 3 han sido absorbidos en los días del tipo 1 de la nueva clasificación, los tipos de día de la serie de referencia son re-etiquetados con un número menos a partir de tipo 3 para poder compararlos. En la Tabla 5-2, se muestra la matriz de confusión obtenida al comparar ambas clasificaciones. Cada columna de la matriz representa un tipo o clase de día de la clasificación de referencia, mientras que cada fila representa un tipo de clúster. El número de cada celda es el número de días que han sido clasificados

por los expertos en la clase correspondiente a la columna y pertenece al clúster de la fila correspondiente. Las celdas localizadas en la diagonal incluyen los días clasificados en grupos con similares características por ambos métodos (números en negrita). La última columna y la última fila muestran el porcentaje total de coincidencias de cada tipo de día. La última celda de la diagonal, muestra el total de coincidencias entre ambas clasificaciones. El 79.6 % de los días han sido clasificados en el tipo definido por las mismas características a partir de ambas metodologías. Este resultado es muy bueno, teniendo en cuenta que los expertos no sabían el límite utilizado para definir el tipo de día cuya energía está principalmente concentrada por la mañana o por la tarde utilizado para definir el F_i . Además, se observa que los clústeres donde el porcentaje de coincidencias es menor corresponden a los tipos de días con poca energía (tipos 7, 8 y 9; la distinción entre estos tipos de días no es relevante desde el punto de vista de la termosolar) o en los tipos de días con muy baja población (tipo 4).

Tabla 5-2. Matriz de confusión de los resultados de la clasificación obtenida con el procedimiento basado en las técnicas de clustering frente a la clasificación de referencia obtenida a partir de la clasificación visual realizada por los expertos.

Clases de Clustering	Clasificación de referencia										Total (%)
	1	2	3	4	5	6	7	8	9	10	
1	321	11	9	1	10	0	0	0	0	0	91.2
2	0	12	0	4	0	0	0	0	0	0	75.0
3	0	0	17	0	0	3	0	0	0	0	85.0
4	0	0	0	3	1	0	1	0	0	0	60.0
5	8	5	3	4	100	3	0	2	0	0	80.0
6	0	0	1	0	0	12	0	2	1	0	75.0
7	0	0	0	6	0	0	12	4	0	0	54.5
8	0	0	0	0	10	2	1	44	1	17	58.7
9	0	0	0	0	0	4	0	4	20	13	48.8
10	0	0	0	0	0	0	8	4	0	18	60.0
Total (%)	97.6	42.9	56.7	16.7	82.6	50.0	0.0	73.3	90.9	37.5	79.6

En la columna de la izquierda de la Figura 5-15, se representan los valores de k_t frente a los valores de V_t de los días del periodo 2000-12 en 3 figuras correspondientes a los 3 valores de F_t . Los puntos de cada clúster están representados con el mismo color y símbolo, y los *medoides*, están representados con el símbolo "X". Las líneas de puntos, limitan el área de cada clúster. En la columna derecha se muestran las mismas 3 figuras, representando los días de la *clasificación de referencia* junto con las líneas que en la columna de la izquierda limitan el área de cada clúster. Se observa que casi todos los puntos de los tipos 1 y 3 se encuentran localizados en el clúster etiquetado con la letra B. Los días con una distribución de la energía a lo largo del día homogénea tienen una buena concordancia con los límites de los clústeres. La mayoría de las discrepancias se identifican en los días con poca energía (o bajos valores de k_t) y en los días con distribuciones de tiempo no homogéneas (o F_t igual a 0 o 1). Como se apuntó anteriormente, cuando la energía es muy baja, su distribución en el tiempo no es relevante. Desde el punto de vista de las centrales termosolares, los tipos de día A, G, I y F podrían formar un único grupo.

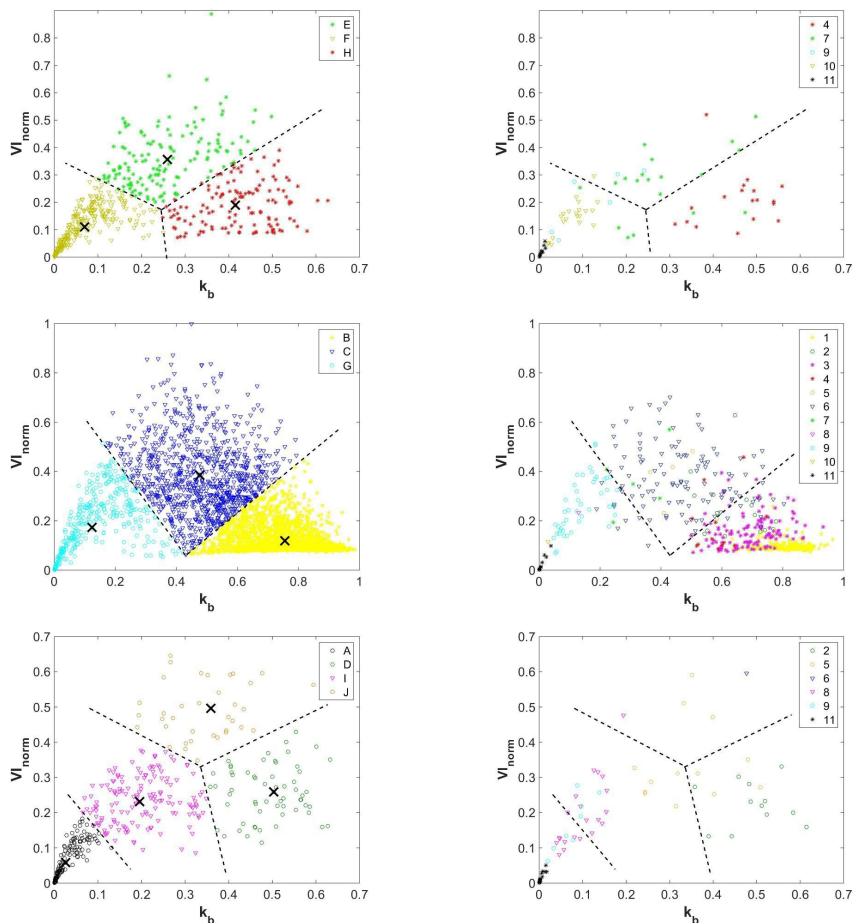


Figura 5-15. A la izquierda, se representa VI frente a k_b para el periodo 2000-12 y F_t igual a 1 (primera figura), 0.5 (segunda figura) y 0 (última figura). Las líneas de puntos definen los límites de cada clúster encontrado y cada medoide correspondiente es simbolizado con "X". A la derecha, se representan los valores de VI frente a k_b para F_t igual a 1 (primera figura), 0.5 (segunda figura) y 0 (última figura) de la serie de referencia elaborada a partir de la clasificación realizada por los expertos. En estas figuras, se incluyen las líneas que definen cada tipo de clúster.

Para facilitar la identificación de las características de cada tipo de día obtenido al aplicar las técnicas de *clustering*, se nombra cada uno de ellos con la primera letra del tipo de distribución temporal de su energía a lo largo del día ("M" mañana, "B"

Homogéneo y “A” tarde) y la primera letra de su nivel de energía (“H” alto, “M” media, “L” baja y “N” nula). La variabilidad correspondiente a cada a tipo de día se puede en la Figura 5-16. En esta figura se representan las curvas de DNI de los *medoides* de cada uno de los clústeres nombrados según este criterio.

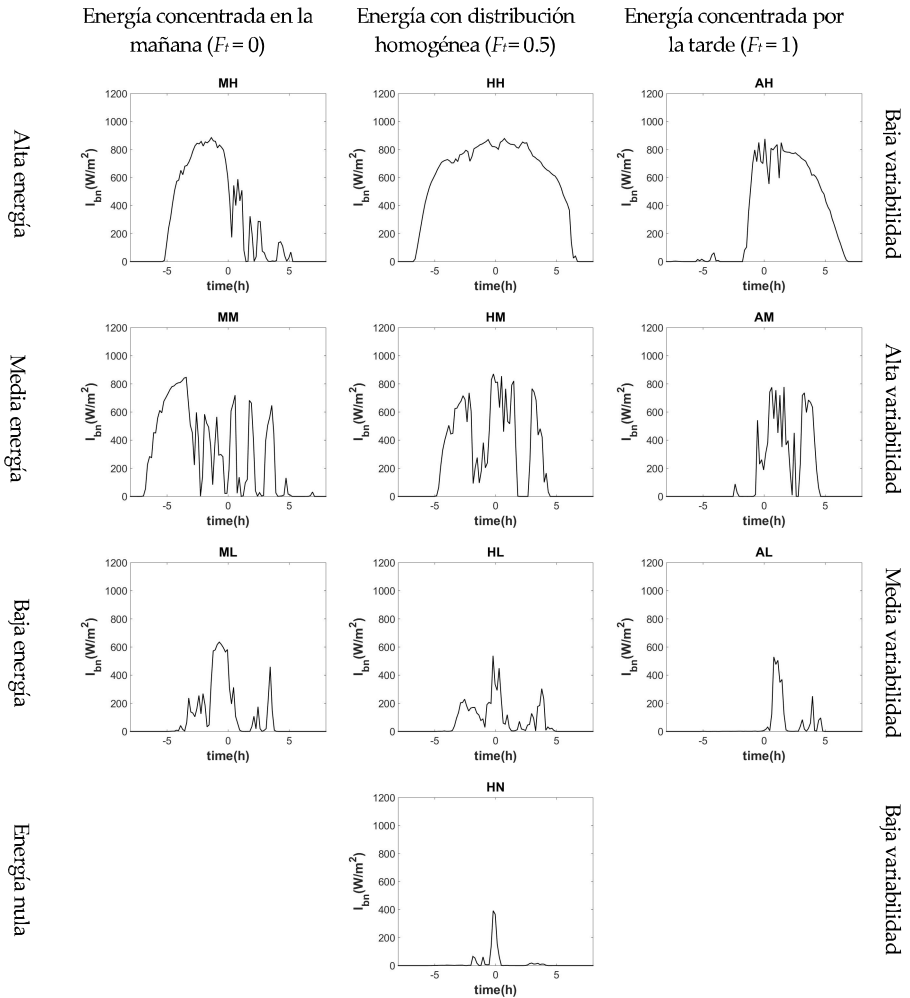


Figura 5-16. Representación de las curvas de DNI correspondientes a los *medoides* de los clústeres obtenidos finalmente como resultado de aplicar la metodología propuesta.

5.4. Caracterización de la producción en función del tipo de día

En este apartado, en primer lugar se evalúa el valor diario de k_p para PTP0 y PTP7.5 durante el periodo 2000-15 a partir de las medidas de DNI registradas en Sevilla. A continuación, se analiza la relación entre k_p y los tres índices de clasificación k_b , VI y F_m (grupo de índices considerado que proporciona el mejor resultado). Por último, se caracterizan estadísticamente los valores diarios de k_p para cada tipo de día definido en el apartado 4.4.1.

5.4.1. Evaluación del índice de producción relativa de cielo despejado diario

En la Figura 5.17 se representan los valores diarios de k_p para PTP0 (izquierda) y PTP7.5 (derecha) para los 16 años considerados. En ambas gráficas también se muestran los valores medianos de cada día del año. En el caso de PTP0, los valores de k_p cercanos a 1 se alcanzan principalmente en verano, aunque también en primavera y la primera parte del otoño. En invierno los valores máximos que se obtienen están en torno a 0.9, porque el modelo de cielo despejado se ha ajustado con los instantes de menor atenuación atmosférica en ausencia de sin nubes, y estos tienen lugar principalmente en primavera. La frecuencia de valores cercanos a 1 durante el verano, la primavera y el otoño es mayor para PTP7.5, aunque estas diferencias son menos significativas en invierno.

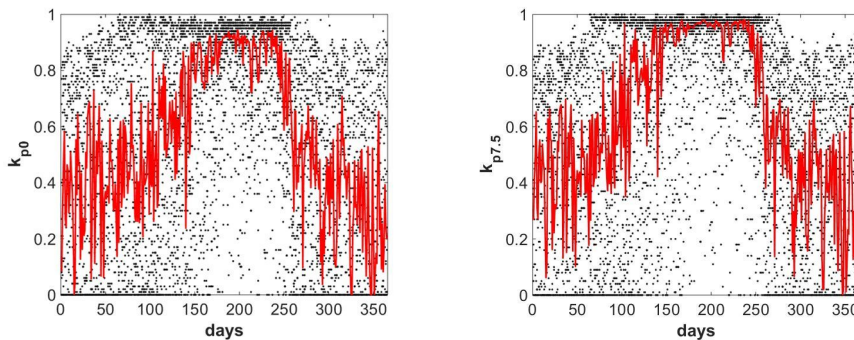


Figura 5-17. Valores diarios de k_p y su correspondiente valor mediano para cada día del año (línea roja) de PTP0 (izquierda) y PTP7.5 (derecha) durante los 16 años de medidas considerados.

A la izquierda de la Figura 5-18, se representan los valores de k_p de PTP7.5 frente a los de PTP0. Como se muestra en la figura, en casi todos los días los valores de k_p de PTP7.5 son mayores a los de PTP0, mostrando un mayor aprovechamiento del recurso disponible. Sin embargo, en algunos días se observan valores de k_p mayores para PTP0. En estos días PTP0 genera más energía respecto a la producción correspondiente a ese día en condiciones de cielo despejado envolvente que PTP7.5. Estos días corresponden en su mayoría a días de invierno con alta energía y baja variabilidad.

A la derecha de la Figura 5-18 se observa la distribución de frecuencia de k_p para los dos modelos de centrales. Ambas distribuciones, muestran un comportamiento bimodal con un máximo localizado en el rango 0-0.1 y otro máximo en el rango 0.9-1. El primer máximo representa la producción durante días nublados, cuando la energía producida es casi nula. El segundo máximo representa la producción en días despejados, muy frecuentes durante la estación de verano como se muestra en la Figura 5-17. Ambos máximos de la distribución bimodal tienen similares porcentajes de frecuencia para PTP0, sin embargo, el máximo localizado en el rango 0.9-1 para PTP7.5 es mayor. PTP7.5 produce un 30 % de días lo mismo que se produce en condiciones de cielo despejado envolvente, casi un 40% de días más que en el caso de PTP0.

El índice de producción relativa de cielo despejado k_p permite evaluar y comparar cómo aprovecha el recurso solar dos centrales termosolares con diferente configuración emplazadas en la misma localización.

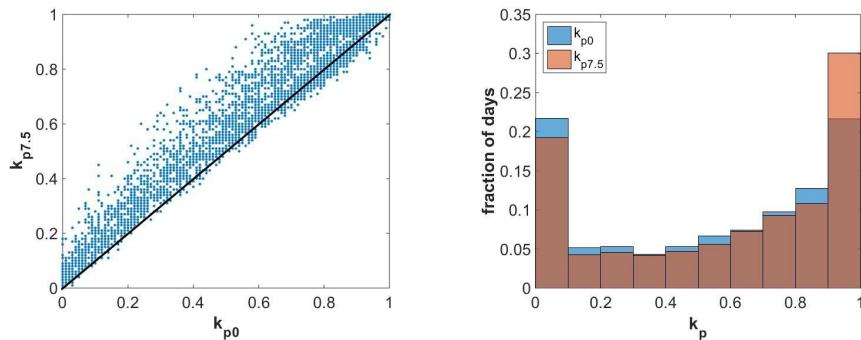


Figura 5-18. Representación de los valores de k_p de PTP7.5 frente a los de PTP0 durante el periodo 2000-15 (izquierda). Distribución de frecuencia de k_p para PTP0 (en azul) y para PTP7.5 (en rojo) durante el periodo 2000-15 (derecha).

5.4.2. El índice de producción relativa de cielo despejado frente a los índices de clasificación

En este apartado, se estudia la relación entre k_p y los índices k_b , VI y F_m comparando los resultados encontrados para la central CCP con y sin almacenamiento.

5.4.2.1. k_p vs. k_b

En la Figura 5-19 se representan los valores de k_p para PTP0 (izquierda) y para PTP7.5 (derecha) frente a los de k_b durante el periodo 2000-15. En ambas gráficas, se observa una correlación positiva entre k_b y k_p , aunque con dispersión. Para un valor de k_b se obtiene un rango de valores de k_p lo que indica que, aunque ambos índices están relacionados, k_p no depende sólo de k_b . Ya que la integración de la DNI para calcular los valores de k_b no pueden reflejar las fluctuaciones ni la distribución de la radiación solar, el único uso de k_b presenta limitaciones para el análisis de la producción (Kang and Tam, 2013b). Comparando ambas nubes de puntos, cabe destacar que el límite superior en el caso de PTP7.5 es mayor, alcanzando valores de k_p cercanos a 1 para días con valores de k_b menores que 0.6 e incrementando el número de días en los que k_p alcanza este valor.

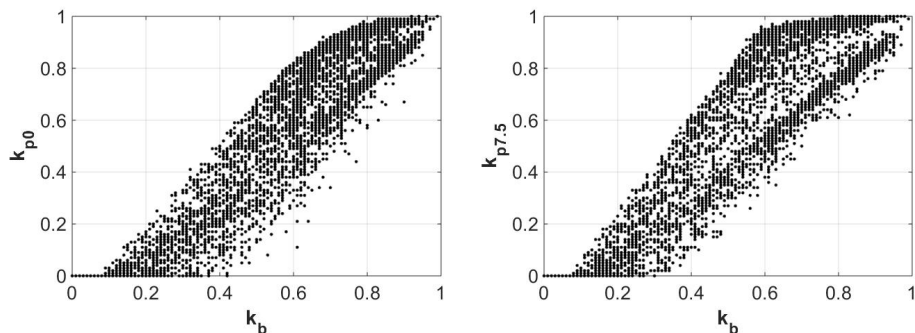


Figura 5-19. Representación de los valores de k_p de PTP0 (izquierda) y de PTP7.5 (derecha) frente a los valores de k_b durante el periodo 2000-15.

En la Figura 5-20 se representa la distribución de frecuencia de la ratio k_p y k_b para PTP0 y PTP7.5. Ambas distribuciones muestran un comportamiento bimodal con un máximo localizado en el rango de valores nulos de producción y el otro entre 0.9 y 1.2. Los valores de esta ratio mayores que 1 representan días en los que la ratio de

producción y DNI diaria es mayor que la ratio de producción y radiación de ese mismo día en condiciones de cielo despejado. Esto frecuentemente ocurre en días claros en los que la energía rechazada por el sistema (debido al desenfoco parcial del campo) es significativamente mayor bajo condiciones de cielo despejado; también, por la misma razón, en algunos días no totalmente despejados en los que la mayoría de la energía se recibe durante periodos de máxima eficiencia óptica.

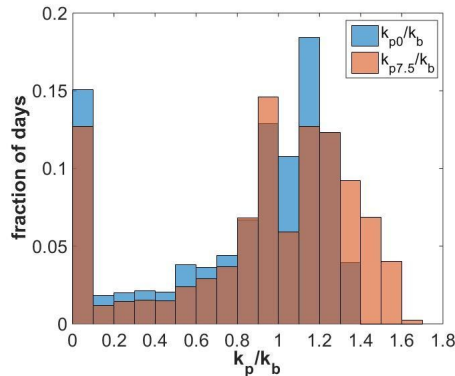


Figura 5-20. Distribución de frecuencia de la ratio k_p/k_b de PTP0 (azul) y de PTP7.5 (rojo) para el periodo 2000-2015.

5.4.2.2. k_p vs. VI normalizado

En la Figura 5-21 se representan los valores de k_p de PTP0 (izquierda) y de PTP7.5 (derecha) frente a los valores de VI normalizados del periodo 2000-15. Los puntos localizados en los mayores valores de k_p para valores de VI en torno a 0.1 corresponden a días de cielo despejado. Ambas nubes de puntos muestran una dispersión significativa, aunque con algunas diferencias. En el caso de PTP0 la nube de puntos tiene un contorno triangular definido por una vertical en VI igual a 0.1, el eje de abscisas y un límite superior con una pendiente decreciente. Esta forma muestra que para un valor dado de VI, k_p puede tomar un rango de valores, pero el valor máximo que puede alcanzar k_p decrece cuando VI aumenta, reflejando el efecto negativo que la variabilidad del recurso tiene en una central sin almacenamiento térmico. Para valores bajos de VI, el rango de valores de k_p se extiende desde 0 hasta 1 con una mayor frecuencia entre 0.6 y 1, indicando un mejor aprovechamiento del recurso cuando la variabilidad de la DNI es baja. En el caso de PTP7.5 la nube de

puntos tiene un contorno cuadrado: se alcanzan valores de k_p cercanos a 1 en días con valores de VI superiores a 0.6, ya que el almacenamiento térmico reduce los efectos negativos de la variabilidad y permite un mejor uso del recurso disponible.

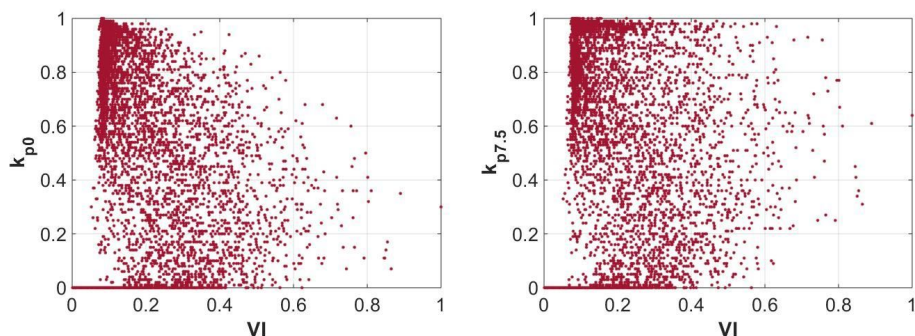


Figura 5-21. Representación de los valores de k_p de PTP0 (izquierda) y de PTP7.5 (derecha) frente a los valores normalizados de VI durante el periodo 2000-15.

5.4.2.3. k_p vs. F_m

Para estudiar la relación entre F_m y k_p , se calculan ambos índices para los 16 años de los valores de DNI medidos y para 16 años de días simétricos (curvas de DNI invertidas en el tiempo respecto al medio día solar). La evaluación de ambas series permite comparar los valores de k_p de días cuya energía principalmente se concentra por la mañana con días en los que la energía principalmente se concentra por la tarde con un perfil simétrico, evitando posibles influencias relacionadas con las características climáticas locales del emplazamiento.

En la Figura 5-22 se representan los valores de k_p de la serie medida y de la serie simétrica de PTP0 (izquierda) y PTP7.5 (derecha) frente a F_m . Se observa que ambas nubes de puntos son casi simétricas respecto a F_m igual a 0.5. Las nubes de puntos tienen una dispersión significativa – mostrando que k_p no es únicamente una función de F_m – aunque con un límite superior bien definido, con valores máximos de k_p iguales a 1 para valores de F_m en torno a 0.5 (el aparente eje de simetría). En el caso de PTP7.5 este límite se encuentra menos definido y decrece más lentamente cuando los valores de F_m se alejan de 0.5 (k_p alcanza valores de 0.9 para F_m igual a 0.7, mientras que para PTP0 k_p no supera el valor de 0.7 para el mismo valor de F_m). Las líneas de puntos verticales que se incluyen en las gráficas de la Figura 5-22 indican los límites

seleccionados en la metodología de clasificación propuesta en el apartado 4.3 para definir los días cuya energía principalmente se concentra por la mañana (*días de mañana*) y los días que principalmente tienen la energía concentrada por la tarde (*días de tarde*).

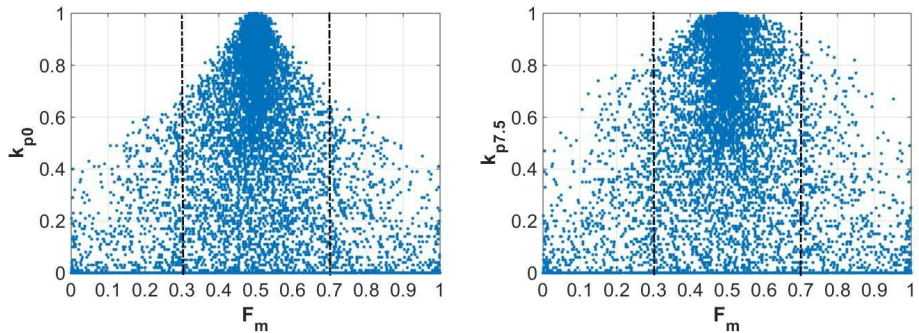


Figura 5-22. Representación de los valores de k_p de PTP0 (izquierda) y de PTP7.5 (derecha) frente a F_m de la serie de medidas de DNI y su serie simétrica. Las líneas de puntos verticales indican los valores límite de F_m seleccionados en la metodología de clasificación para definir los días de tarde ($F_m \leq 0.3$), los días homogéneos ($0.3 < F_m < 0.7$) y los días de mañana ($F_m \geq 0.7$).

Para comparar el comportamiento de los *días de mañana* (MH, MM, ML) con los *días de tarde* (AH, AM, AL), de cada central, en la Figura 5-23 se representan los valores de k_p de la serie medida y la serie simétrica correspondientes a los días considerados como días de tarde frente a los días de mañana (fila superior) y la distribución de frecuencia acumulada (CDF) de los valores de k_p para ambos tipos de días (fila inferior). En el caso de PTP0 ambas gráficas (gráficas representadas a la izquierda en la fila superior e inferior) muestran que en los días de tarde se alcanzan valores de k_p mayores que en los días de mañana. Sin embargo, en el caso de PTP7.5 la tendencia no está tan clara (gráficas representadas a la derecha de la Figura 5-23 en la fila superior e inferior). Para los valores altos de k_p los días de mañana parecen más favorable desde el punto de vista de la explotación del recurso, pero para los valores bajos la tendencia se invierte.

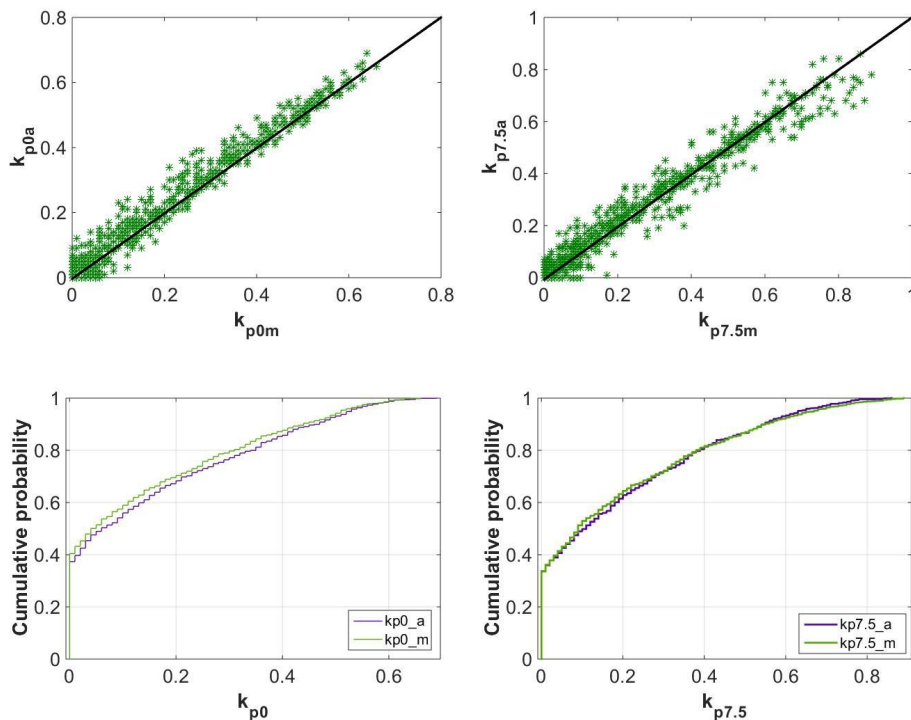


Figura 5-23. Fila superior: valores de k_p de los días de tarde frente a los valores de los días de mañana para la serie medida y la serie simétrica de PTP0 (izquierda) y PTP7.5 (derecha); fila inferior: CDF de los valores de k_p de los días de tarde (línea violeta) y los días de mañana (línea verde) para ambas series de PTP0 (izquierda) y PTP7.5 (derecha).

1.1.2 Caracterización del índice de producción relativa de cielo despejado de cada tipo de día

En la Figura 5-24 se representan los valores de k_p (periodo 2000-15) para cada tipo de día y cada tipo de central en gráficas "Box and Whisker". Excepto para los tipos de día BL, AL y BN, la mediana de los valores de k_p de todos los tipos de días, son mayores en el caso de PTP7.5. El sistema de almacenamiento térmico permite una mejor gestión de la variabilidad y el exceso de energía radiante en días despejados. Esto también se refleja en los valores máximos de k_p para PTP7.5. Sin embargo, los

mínimos de ambas centrales son similares. Estos días generalmente corresponden a días de invierno cuando el efecto del ángulo de incidencia penaliza el comportamiento de las centrales CCPs. Durante estos días, el desenfoque no es necesario y el sistema de almacenamiento no se puede utilizar eficientemente. Así, los valores de k_p durante los días de invierno son similares para ambas centrales.

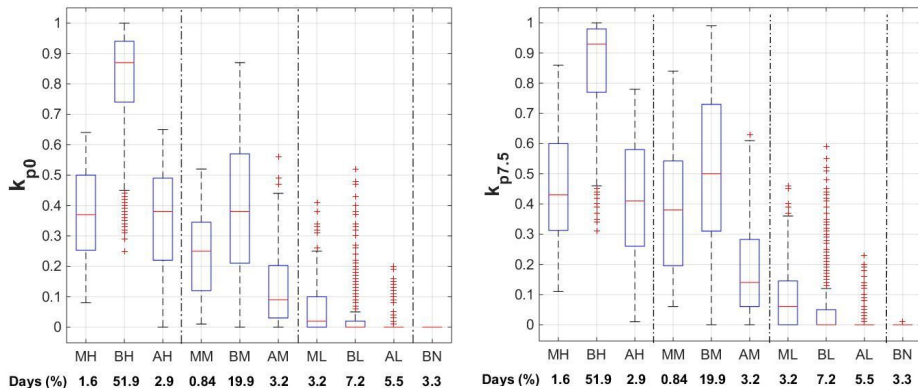


Figura 5-24. Gráficas “Box and Whisker” de los valores k_p de cada tipo de día de PTP0 (izquierda) y PTP7.5 (derecha) para el periodo 2000-15. El percentil 25 y 75 definen las cajas de cada tipo de día y el 99.3% de los valores se encuentran entre los límites de los “whiskers”. Las líneas rojas horizontales situadas en cada caja, corresponden al valor mediano de cada serie. Las líneas de puntos verticales que se muestran en ambas figuras separan los días clasificados con energía alta, media, baja y nula. El porcentaje de días de la serie que pertenece a cada tipo de día se muestra bajo la etiqueta de cada uno de ellos.

En ambas centrales, los valores mayores de k_p se obtienen para el tipo de día BH. Como se muestra en la Figura 5-24, más del 80% de los días clasificados en este clúster tienen un valor de k_p superior a 0.7. Los tipos de días MH y AH muestran cajas similares con pequeñas diferencias en los valores extremos. Sin embargo, los tipos de días MM y AM muestran diferencias claras. Para los tipos de días con baja energía se observan las mismas diferencias entre los días de mañana y de tarde: el tipo de día ML tiene valores de k_p mayores que cero en más de la mitad de los días, mientras que en el tipo AL sólo los “outliers” no son nulos. Los valores de k_p asociados a los tipos de día con variabilidad y DNI concentrada durante la mañana son mayores para nuestra localización. Este resultado puede atribuirse a un

comportamiento específico característico de la climatología de Sevilla, considerando los resultados obtenidos en el apartado 5.4.2.3 respecto a la relación entre k_p y F_m . También cabe destacar el alto rango de valores de k_p que abarca el tipo de día BM. La operación de las centrales –especialmente de PTP0- durante estos tipos de días es compleja. Dependiendo de la frecuencia y la duración de los transitorios la central debe parar la turbina o puede continuar operando, gracias a la inercia térmica de la central. Respecto a los tres últimos tipos de días (BL, AL, BN), la energía producida es básicamente inapreciable. El periodo del año en el que este tipo de días es más habitual es el mejor periodo para parar la central para operaciones de mantenimiento.

Para un mejor entendimiento de los resultados encontrados en nuestra localización, los valores medios y medianos de k_p para los 16 años simulados de PTP7.5 y PTP0 de cada tipo de día se presentan en la Tabla 5–3. Además, se incluyen los valores medios y medianos de la ratio $k_{p7.5}/k_{p0}$ para los 16 años. Los estadísticos de $k_{p7.5}$ son un 30 % mayor respecto a los de k_{p0} . En la Figura 5-25, se representan las distribuciones de frecuencia de cada tipo de día para PTP0 y PTP7.5. Como se observa en dicha figura, la distribución de los tipos de día BH y AM están lejos de ser una distribución normal. En ambas centrales, los valores de k_p más frecuentes para el tipo de día BH se encuentran localizados en el rango 0.9-1 y, como se muestra en la Tabla 5–3, más del 50% de los días del periodo analizado en Sevilla pertenecen a este tipo de día. La distribución de frecuencia del tipo de día BM representada en la Figura 5-25 muestra que este tipo de día abarca un rango amplio de valores de k_p y tiene una distribución cercana a una normal, principalmente en el caso de PTP0, en concordancia con el resultado observado en la Figura 5-24. Ya que este tipo de día es el segundo clúster con mayor número de días y que la operación de las centrales en los días con sus características es más compleja, dividir este grupo en más subgrupos podría ser interesante para la evaluación de la operación. Además, se confirma que los tipos de días BL, AL y BN podrían agruparse en un único clúster ya que en todos los casos tanto los valores medianos como los valores medios de k_p son nulos o cercanos a cero.

Tabla 5–3. Valores medios y medianos de k_p para cada tipo de día de los 16 años simulados de PTP7.5 y PTP0. En la última fila se muestran los porcentajes de días de cada clúster durante el periodo. Las líneas de puntos separan los tipos de días con energía alta (H), media (M), baja (L) y nula (N).

Tipo de día	MH	BH	AH	MM	BM	AM	ML	BL	AL	BN	
k_{p0}	Media	0.38	0.82	0.36	0.23	0.39	0.13	0.06	0.04	0.01	0.00
	Mediana	0.37	0.87	0.38	0.25	0.38	0.09	0.02	0.00	0.00	0.00
$k_{p7.5}$	Media	0.46	0.86	0.42	0.38	0.51	0.18	0.09	0.05	0.01	0.00
	Mediana	0.43	0.93	0.41	0.38	0.50	0.14	0.06	0.00	0.00	0.00
$k_{p7.5}/$	Media	1.21	1.05	1.17	1.65	1.31	1.38	1.50	1.25	1.00	0.00
k_{p0}	Mediana	1.16	1.07	1.08	1.52	1.32	1.56	3.00	0.00	0.00	0.00
N. días (%)	1.6	51.9	2.9	0.8	19.9	3.2	3.2	7.7	5.5	3.3	

Con las distribuciones de frecuencia, se confirma que en nuestro emplazamiento el recurso solar es peor en los días clasificados como días de tarde desde el punto de vista de la producción de una central CCP. En la Figura 5-25, la distribución de los valores de k_p para el tipo AM muestra un claro pico entre 0 y 0.1 (aproximadamente 50% de los valores para PTP0 y un 35% para PTP7.5). Además, k_p toma valores insignificantes en los días clasificados como de energía baja o nula, excepto en el caso de ML, donde aproximadamente el 35% de los valores para PTP7.5 y el 25% of los valores para PTP0 son mayores que 0.1. Por tanto, en nuestra localización, el comportamiento esperado de los clústeres clasificados en los niveles de energía baja o nula no se cumple si la energía radiante es recibida durante la primera mitad del día. En este caso, las centrales producen pequeñas cantidades de energía, pero dicha energía no es insignificante.

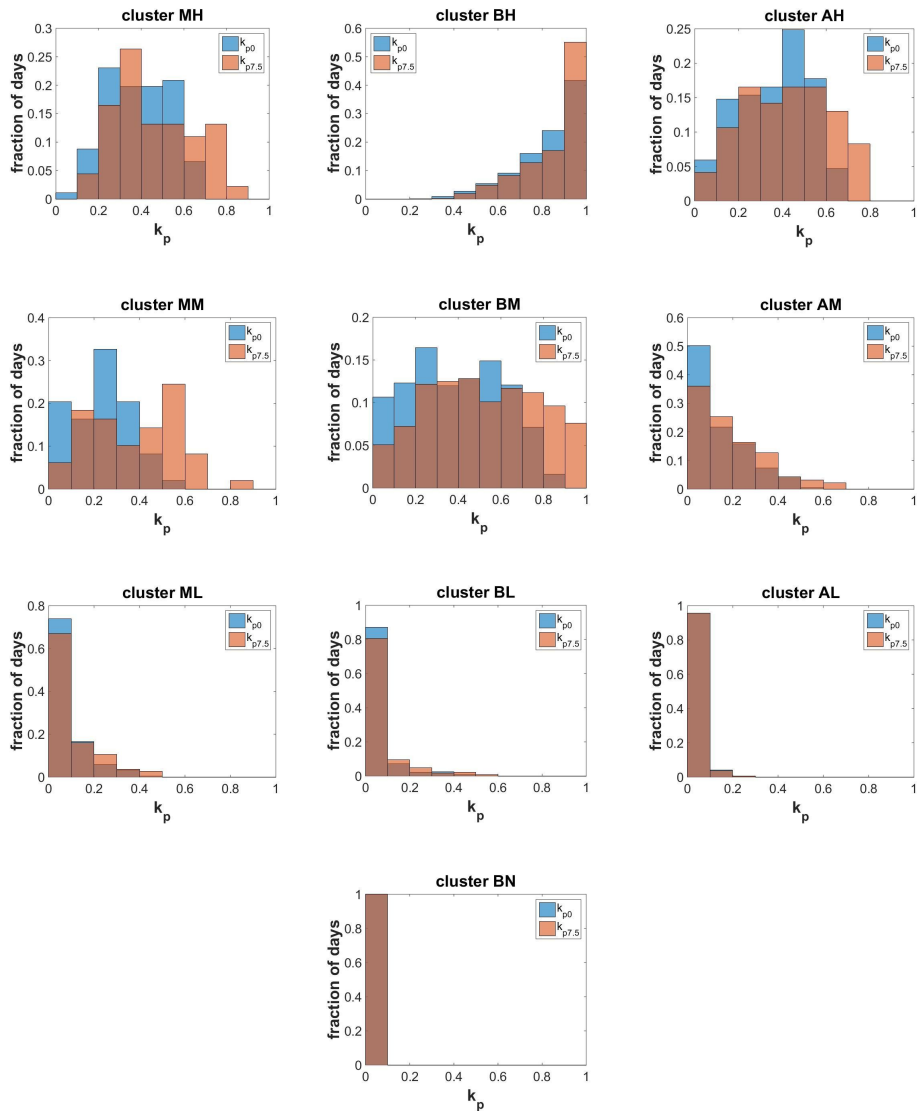


Figura 5-25. Histogramas de los valores de k_p para los tipos de días o clústeres a partir de la simulación de PTP0 (azul) y PTP7.5 (rojo) durante los 16 años de medidas.

En la Figura 5-25 se observan algunas similitudes entre las distribuciones de k_p de los diferentes tipos de días para PTP0 y PTP7.5. Para comparar la distribución de los valores de k_p de los diferentes tipos de días se aplica el test “two-sample Kolmogorov-Smirnov” (K-S test) a partir de la función de Matlab “kstest2” (MathWorks, 2016) a ambas centrales. Para cada par de tipos de días se analiza la decisión del test para la hipótesis nula (las dos series de datos tienen una misma distribución continua). El resultado es “false” si el test no puede rechazar la hipótesis nula con un 5 % de nivel de significancia, y “true” en el caso contrario. El resultado del K-S test para PTP0 se muestra en la Tabla 5-4 y para PTP7.5 en la Tabla 5-5. Como era de esperar, en ambos casos la hipótesis nula (las dos series de datos tienen la misma distribución continua) no puede ser rechazada para los mismos tipos de días, pero también se observa que no puede ser rechazada para los tipos MH y AH; aunque, si la rechaza en el caso de los tipos de días MM y AM (devolviendo un “true”). En el emplazamiento seleccionado, cuando la variabilidad es baja y la energía alta, los valores de k_p tienen la misma distribución independientemente de si la energía se concentra por la mañana o por la tarde, sin embargo para días con media o alta variabilidad las distribuciones difieren. En el caso de PTP7.5, Tabla 5-5, la hipótesis nula tampoco puede ser rechazada entre los tipos de días de mañana con energía alta y baja variabilidad y con energía media y alta variabilidad (las distribuciones de MM, MH y AH son similares). Estos resultados y algunos previos son recopilados en la Tabla 5-6 para facilitar su interpretación.

Tabla 5–6. Análisis cualitativo de los valores de k_p de cada clúster y central CCP. k_p toma valores insignificantes en los clústeres con energía nula y baja con la excepción del clúster ML (k_p en rojo). El clúster BH muestra los mayores valores de k_p (k_p en verde) para ambas centrales.

PTP0		PTP7.5	
Cluster	k_p	Cluster	k_p
BH	↑ k_p	BH	↑ k_p
MH & AH	k_p	MH & AH & MM	k_p
MM	k_p		
BM	k_p	BM	k_p
AM	k_p	AM	k_p
ML	↓ k_p	ML	↓ k_p
BL & AL & BN	null	BL & AL & BN	null

A partir del análisis de los valores de k_p realizado, se identifican 4 grupos con valores intermedios de k_p en el caso de PTP0: MH & AH, MM, BM y AM; mientras que solo identificamos 3 en el caso de PTP7.5: MH & AH & MM, BM y AM (Tabla 5-6). Considerando estos resultados, se concluye que hay solo 7 tipos de días desde el punto de vista de la producción de una central CCP sin almacenamiento y 6 para una con 7.5 horas de almacenamiento para el emplazamiento analizado.

La caracterización de los tipos de días desde la perspectiva de la producción de una central CCP puede usarse en las metodologías de predicción (Jiménez-Pérez and Mora-López, 2016). La predicción de los tipos de días ayuda a conocer el nivel más probable de aprovechamiento del recurso solar (k_p) y como resultado, para un día específico, el rango de producción más probable. Además, la predicción de los tipos de días de los días próximos puede ayudar a decidir el modo de operación a aplicar (Rohani et al., 2016). Desde el punto de vista del diseño de la central, este resultado podría utilizarse para optimizar el tamaño del sistema de almacenamiento,

definiendo estrategias de operación y técnicas de control adaptadas a las condiciones de las tarifas del mercado eléctrico (Casati et al., 2015) y a la distribución de tipos de día típica del emplazamiento.

6. DISCUSIÓN

Para caracterizar el recurso solar en un emplazamiento a partir de cualquier metodología es recomendable contar con una base de datos radiométricos extensa completa y validada. En la presente Tesis, los 16 años de datos radiométricos de la base de datos de Sevilla se han etiquetado siguiendo procedimientos de control de calidad propuestos por la BSRN. Para interpretar el etiquetado, localizar los errores y tratar de identificar el origen de los mismos, se presenta una nueva herramienta visual, compuesta por dos tipos de gráficas. El primer tipo de gráfica consiste en la representación de las etiquetas resultantes de aplicar el control de calidad para cada variable abarcando cualquier periodo de tiempo en una misma gráfica e independientemente de la frecuencia de registro de los datos. Este tipo de gráfica permite visualizar de forma general los resultados de aplicar cualquier etiquetado de un control de calidad para cualquier variable radiométrica. En el caso de haber registrado las tres variables, se propone enfrentar los valores de GHI medidos y calculados en una gráfica guiada por unas líneas que identifican las diferencias de $\pm 8\%$. El conjunto de ambos tipos de gráficas ayuda a evaluar la calidad de las medidas y a identificar el origen de algunos errores. Varios casos prácticos demuestran su utilidad para identificar los fallos más habituales de las tres componentes de la radiación como se muestra en el Anexo A.

Una vez identificados los errores mediante el control de calidad y la herramienta de visualización de las etiquetas de los datos, se ha implementado una nueva metodología de rellenado de huecos adaptada a cada variable, duración del error e instante en el que tiene lugar. Con la representación gráfica de los valores k_t y k_{b0} de los datos originales y la base de datos completada, se comprueba que una vez aplicado el procedimiento de rellenado de huecos y corrección de errores, las propiedades estadísticas de los datos de GHI y DNI prácticamente se conservan. Se observa una muy buena correspondencia entre la función de distribución de probabilidad de las series de datos válidos originales y los procesados.

Contar con esta extensa base de datos de valores registrados en su origen cada 5 s permite evaluar la distribución estadística de ambas variables en diferentes escalas temporales. En la presente Tesis, se han evaluado las series de valores instantáneos, diez-minutales, horarios, diarios, mensuales y anuales, escalas temporales más

relevantes para las distintas etapas de un proyecto de central termosolar.

Las distribuciones de los valores instantáneos de k_t y k_{ho} muestran un comportamiento bimodal, más pronunciado para los menores valores de masa óptica de aire, en concordancia con los resultados obtenidos por otros autores para localizaciones con el mismo clima de acuerdo a la clasificación Köppen-Geiger. Este comportamiento bimodal refleja dos estados predominantes, despejado y cubierto, más pronunciados en el caso de la componente directa.

De la comparativa de ambas variables en las escalas menores, cabe destacar que la distribución de frecuencia de los valores instantáneos y diez-minutales de los índices k_t y k_{ho} son muy similares, mientras que la distribución de los valores horarios muestra diferencias significativas respecto a los mismos. Estas diferencias también se observan al comparar los valores de GHI y DNI, principalmente en esta última. Las mayores diferencias en la distribución de DNI se observan en los valores altos, para los que la frecuencia de valores horarios es menor que la de los diez-minutales. En general, el uso de valores horarios suaviza los gradientes de DNI. Este resultado debería considerarse a la hora de seleccionar el paso de tiempo más adecuado para la simulación de sistemas de energía solar en los que los transitorios tienen un efecto significativo, como es el caso de las centrales termosolares. A continuación, se evalúa el recurso en las diferentes escalas obteniendo algunos resultados destacables. El valor más frecuente del índice de claridad instantáneo para los valores más bajos de masa óptica es 0.76, lo que demuestra un bajo nivel de turbidez atmosférica. Clasificando los días a partir del k^d , más de un 50 % de días en valor promedio anual se clasifican como días de cielo claro y menos de un 8 % como días completamente cubierto, aunque el valor límite que define cada tipo de día puede variar considerablemente el porcentaje de días claros computados, ya que este valor se encuentra entre los valores máximos en la distribución en frecuencia del k^d . En promedio, la DNI es superior a $600 \text{ W}\cdot\text{m}^{-2}$ durante aproximadamente 8 horas en los días de verano, estación que reúne las mejores propiedades desde el punto de vista de las centrales. Tanto el mes más variable, (febrero), como el de menor promedio de DNI diaria (diciembre) son meses de invierno. El valor medio diario anual de GHI es $4.98 \text{ kWh}\cdot\text{m}^{-2}$ y el de la DNI, $5.68 \text{ kWh}\cdot\text{m}^{-2}$. Ambas variables tienen una variabilidad inter-anual relativamente baja (entre -3% y 5% para GHI y -6% a +8% para DNI) y una variabilidad inter-mensual significativa, con dependencia estacional.

Los resultados de la clasificación de días en función de índices representativos de la

energía, la variabilidad y la distribución temporal predominante de esta energía muestran que los dos conjuntos de índices analizados ($\{k_t, VI, F_m\}$ y $\{k_t, POP_D, F_m\}$) proporcionan un número óptimo de grupos igual a 10, aunque la incertidumbre del grupo que incluye el índice VI es menor a la hora de encontrar el caso óptimo. La comparación de la coincidencia entre la clasificación de referencia realizada por expertos y la resultante de la metodología de clasificación también es más favorable al mismo grupo. La concordancia de los días con distribución de energía homogénea es muy buena, y la mayoría de las discrepancias se localizan en los días con baja energía (o bajo k_t), y con energía principalmente concentrada por la mañana o por la tarde (F_t igual a 0 o a 1). Los días con baja energía no son relevantes desde el punto de vista de las centrales termosolares. Por tanto, la discriminación de estos grupos no resulta de utilidad. Para este fin, parece razonable pensar que podría definirse un único clúster con la unión de todos los días con baja energía, tipos de días ML, HL, AL y HN. De esta forma, 7 tipos de días (en lugar de 10) podrían ser suficientes para su aplicación en el análisis de la producción de esta tecnología.

La caracterización de la producción de una central termosolar en función de los tipos de días identificados con la nueva metodología de clasificación se lleva a cabo mediante un nuevo índice, denominado índice de producción relativa de cielo despejado o índice k_p . Este índice informa de la proporción de energía eléctrica obtenida por la central respecto a la que podría haber obtenido en condiciones de cielo despejado durante un cierto periodo. Su determinación requiere definir un modelo de central termosolar y un modelo de cielo despejado.

El análisis de los valores diarios de k_p de las dos centrales de tecnología CCP analizadas (PTP0 y PTP7.5) durante los 16 años de medidas muestra que:

- Los valores diarios de k_p toman valores mayores en verano y menores en invierno, y se muestran muy variables durante primavera y otoño.
- Los valores de k_p de PTP7.5 son normalmente –aunque no siempre– mayores.
- Las distribuciones de frecuencia de k_p son bimodales, con el máximo situado en los valores mínimos para PTP0 y en los valores máximos para PTP7.5.

Los resultados obtenidos son un indicativo de las ventajas que supone poder contar con un sistema de almacenamiento de alta capacidad para el aprovechamiento del recurso solar.

El estudio de la relación de k_p con los índices finalmente considerados como más adecuados para representar a las características del recurso en la metodología de clasificación (k_b , VI , y F_m) muestra que:

- Para ambas centrales, se observa una buena correlación entre k_b y k_p , con unos límites superior e inferior crecientes bien definidos, aunque con cierta dispersión indicativa de que se requieren otros parámetros para relacionar ambos índices. Al comparar las nubes de puntos de PTP0 y PTP7.5, se identifica un límite superior de mayor pendiente en el caso de PTP7.5. Además, se observan valores de k_p cercanos a 1 para días con k_b menores que 0.6 y el número de días que alcanzan este valor aumenta.
- En el caso de VI , se distingue un rango de valores de k_p entre 0 y 1 para valores bajos de VI con mayor frecuencia en los valores cercanos a 1. El límite superior de la nube de puntos, decrece cuando VI aumenta. Así, la nube de puntos se encuentra limitada por una frontera lineal decreciente que muestra que mayores valores de VI implican menores valores de k_p . Este comportamiento sugiere, de nuevo, que se requieren otros parámetros para relacionar estos índices. La influencia de la variabilidad es menos relevante en el caso de PTP7.5. El comportamiento de una central de tecnología CCP con almacenamiento puede duplicar el valor de k_p respecto a una central sin almacenamiento en días con alta energía y alta variabilidad. Esta ventaja no es tan significativa en los días con baja energía y alta variabilidad.
- Los mayores valores de k_p se encuentran en el rango de días con distribuciones de energía homogénea, correspondientes también a los días de mayor energía. El rango que abarca los valores de k_p cercanos a 1, es mayor para PTP7.5; alcanzando valores de k_p iguales a 0.9 para valores de F_m de 0.7, mientras que en PTP0 no se superan valores de k_p de 0.7 para los mismos valores de F_m . Se han comparado los valores de k_p para días considerados como *días de mañana* (F_m mayor o igual a 0.7) y los considerados como *días de tarde* (F_m menor o igual a 0.3) para ambas centrales. En el caso de PTP0, los *días de tarde* tienen, en promedio, mayores valores de k_p que los de *días de mañana*. Para PTP7.5 este comportamiento no está bien definido, sugiriendo que no se le puede atribuir diferencias significativas a este factor.

Finalmente, se caracteriza la producción de las dos centrales para cada uno de los tipos de días definidos a partir de la metodología de clasificación mediante el índice

de producción relativa de cielo despejado, k_p . Todos los tipos de días excepto los días con energía media, alta variabilidad y distribución de energía homogénea de la DNI (tipo BM) tienen una distribución de k_p característica. Los valores de k_p estadísticamente mayores, pertenecen al tipo de día BH (alta energía, distribución en el tiempo homogénea y variabilidad baja). Los valores de k_p de los días con la energía concentrada principalmente por la mañana o por la tarde, tienen distribuciones similares si la variabilidad es baja; en caso contrario, las distribuciones son claramente diferentes, alcanzando mayores valores de k_p cuando la energía se concentra por la mañana. Teniendo en cuenta los resultados obtenidos al comparar los días de mañana con los días de tarde de la unión de la serie medida y su simétrica, se concluye que, en nuestro emplazamiento, los días de mañana tienen mejores características radiométricas que los días de tarde desde la perspectiva del comportamiento de una central de tecnología CCP. Los días con energía baja o casi nula (BL, AL, ML y BN) muestran distribuciones de frecuencia similares desde el punto de vista de la producción, sugiriendo la posibilidad de unir estos tipos de días en uno sólo (con la posible excepción de ML para la localización de este estudio). Por tanto, los 10 tipos de día identificados en función de la variabilidad, la energía y la distribución de la energía de la DNI, pueden reducirse a 6 desde el punto de vista de la producción de una central CCP sin almacenamiento y a 7 desde el punto de vista de una central CCP con 7.5h de almacenamiento emplazada en Sevilla.

La relación entre los tipos de días y el índice de producción relativo al cielo despejado tiene una aplicación directa para la definición de la estrategia de operación de una central y la predicción de la energía generada a partir de la predicción cualitativa del tipo de día.

7. CONCLUSIONES

Con el objetivo de mejorar la caracterización del recurso para el análisis de la producción de centrales termosolares, en esta Tesis se presentan dos metodologías de caracterización del recurso solar con diferentes enfoques, así como nuevas herramientas y procedimientos que facilitan su aplicación y validación. La primera metodología, caracteriza estadísticamente la radiación solar en diferentes escalas temporales, analizando los aspectos que pueden ser de utilidad en el desarrollo de un proyecto de una central termosolar. La segunda metodología, clasifica días en función de las propiedades de la radiación que influyen en la operación y producción de estas centrales aplicando técnicas de *clustering* sin supervisión. Finalmente, la representatividad de los tipos de días identificados se evalúa desde el punto de vista de la producción de una central de tecnología CCP con y sin almacenamiento, a partir de un nuevo índice. La información que aporta esta clasificación y su relación con la producción es valiosa para el diseño, el análisis de viabilidad y la operación de las centrales.

Ambas metodologías se han aplicado a la caracterización del recurso solar en Sevilla a partir de 16 años completos de medidas validadas. La validación de dichas medidas implica el control de calidad y el rellenado de los huecos consecuencia del registro de datos erróneos, tarea que se ha realizado con la ayuda de una nueva herramienta de visualización a partir de dos tipos de gráficas. Varios casos prácticos demuestran su utilidad para identificar los fallos más habituales de las tres componentes de la radiación. Además, se propone un procedimiento para el rellenado de huecos y corrección de datos erróneos, que aplica diferentes criterios en función de la variable, el periodo a rellenar y el tipo de día según el estado del cielo. De la aplicación a los datos de Sevilla se puede concluir que este procedimiento conserva las propiedades estadísticas de los datos originales.

Una vez completada y validada la base de datos radiométricos de Sevilla, se aplica la metodología de caracterización estadística en las diferentes escalas temporales, de la que se concluye lo siguiente:

- La distribución en frecuencia de los valores de k_t y k_{to} muestran comportamientos bimodales, reflejando dos estados predominantes, despejado y cubierto, más pronunciados en el caso de la componente

directa.

- El análisis de la base de datos de Sevilla en distintas escalas temporales muestra que las distribuciones estadísticas de los valores instantáneos y diez-minutales son semejantes entre sí, pero muy distintas de las de los valores horarios. Este resultado debe tenerse en cuenta a la hora de seleccionar la escala temporal adecuada en la simulación de centrales termosolares.
- La clasificación de tipos de días en función únicamente de los valores diarios de k^d_t puede variar considerablemente en función de los valores límites fijados. En el caso de Sevilla, puede ser muy influyente en el cómputo de días claros ya que los valores de k^d_t que normalmente se emplean para la consideración de un día como día claro se encuentran entre los valores más frecuentes de su distribución en frecuencia.
- Sevilla presenta una alta frecuencia de instantes de cielo despejado con baja turbidez atmosférica, unas 8 horas de valores de DNI superiores a 600 Wm^{-2} durante los meses de verano, una DNI media diaria anual de 5.68 kWhm^{-2} y una variabilidad interanual relativamente baja (-6% y $+8\%$), aunque con una significativa variabilidad inter-mensual, con dependencia estacional.

De la aplicación de la metodología de clasificación en función de la energía, la variabilidad y la distribución en el tiempo de la DNI a los datos de Sevilla se concluye que:

- La clasificación de tipos de días a partir de la visualización de curvas tiene un cierto carácter subjetivo, incluso teniendo una guía de tipos de días previamente definidos.
- La metodología propuesta para clasificar días a partir de los índices k_t , VI y F_t aplicada a la base de datos de Sevilla obtiene unos buenos resultados respecto a la serie de referencia definida a partir de las coincidencias en la clasificación visual de 3 o más expertos.
- El número óptimo de tipos de días obtenido al aplicar la metodología de clasificación basada en técnicas de partición sin supervisión es 10, resultado óptimo desde el punto de vista matemático. Este resultado puede refinarse en función de la aplicación (en este caso, el análisis de la producción de centrales termosolares) unificando o dividiendo grupos mediante nueva

aplicación de las técnicas de partición sobre algún grupo.

De la caracterización de los tipos de días desde el punto de vista de la producción de las centrales de tecnología CCP con y sin almacenamiento se concluye que:

- El índice de producción relativa de cielo despejado k_p permite evaluar las características del recurso solar y comparar la capacidad de aprovechamiento del recurso solar respecto a otras centrales con diferente configuración.
- La variabilidad y la distribución en el tiempo de las curvas de DNI tienen una alta influencia en las centrales sin almacenamiento. Esta influencia disminuye considerablemente cuando la central cuenta con un sistema de almacenamiento térmico de alta capacidad.
- Todos los tipos de días excepto los días con energía media, alta variabilidad y distribución de energía homogénea de la DNI (tipo BM) tienen una distribución de k_p característica. El tipo de día BM incluye un alto número de días abarcando un rango amplio de valores de k_p sin una tendencia clara. Este tipo de día resulta de interés desde la perspectiva de la operación de la central, por lo que una subdivisión del grupo acorde a la característica más interesante de este tipo de día (alta variabilidad) puede resultar muy útil para su aplicación en la optimización de las estrategias de operación.
- Los 10 tipos de días identificados en la metodología de clasificación a partir de las propiedades de las curvas de DNI se reducen en el emplazamiento estudiado a 7 tipos de días desde el punto de vista de la producción de una CCP sin almacenamiento, y a 6 para una central con 7.5h de almacenamiento térmico.
- La caracterización de los tipos de días respecto al índice de producción relativa de cielo despejado permite entregar al operador información sobre la producción más probable de la central, con un rango de incertidumbre dependiente del tipo de día que resulte de los modelos de predicción.

Finalmente se puede concluir que la introducción de la variabilidad y la distribución temporal a escala diaria deben tenerse en cuenta en la caracterización del recurso solar para centrales termosolares de concentración, ya que afectan de manera significativa a la operación de estas centrales.

La aplicación de estas metodologías en futuros trabajos a otras bases de datos de

emplazamientos característicos de otros climas y a otras variantes tecnológicas (centrales de torre, captadores lineales de Fresnel, discos parabólicos...) permitirá generalizar y/o matizar las conclusiones obtenidas.

Otras futuras líneas de trabajo que se proponen a partir de los resultados obtenidos son:

- El estudio de la incorporación en las metodologías de evaluación del recurso solar a largo plazo el uso de estos tres índices para tratar que los días de la serie anual considerada como representativa de la DNI, tenga porcentajes mensuales o estacionales de tipos de días similares a los de la serie completa.
- Estudio de nuevas metodologías de dimensionado de centrales termosolares basadas en la distribución estacional de los distintos tipos de días.
- La incorporación en los sistemas de predicción de DNI de la predicción de los tipos de días y a partir de esta predicción proporcionar una estimación de la producción basada en k_p y en la incertidumbre asociada al tipo día.

Por último, merece la pena mencionar una aplicación de la clasificación de días que actualmente está en curso. La clasificación de tipos de días se está empleando para desarrollar una metodología de generación de datos de DNI a alta frecuencia a partir de datos horarios como parte de otra Tesis Doctoral. Esta metodología contempla la clasificación en tipos de días para desarrollar un algoritmo específico para la generación de datos a alta frecuencia en función del tipo de día. Esta incorporación ha supuesto una considerable mejoría en los resultados obtenidos.

REFERENCIAS

- Bachour, D., Perez-Astudillo, D., 2014. Ground measurements of Global Horizontal Irradiation in Doha, Qatar. *Renew. Energy* 71, 32–36.
- Bright, J.M., Smith, C.J., Taylor, P.G., Crook, R., 2015. Stochastic generation of synthetic minutely irradiance time series derived from mean hourly weather observation data. *Sol. Energy* 115, 229–242.
- Calbó, J., González, J.A., Pegès, D., 2001. A method for sky-condition classification from ground-based solar radiation measurements. *J. Appl. Meteorol.* 40, 2193–2199. doi:10.1175/1520-0450(2001)040<2193:AMFSCC>2.0.CO;2
- Calinski, T., Harabasz, J., 1974. A dendrite method for cluster analysis. *Commun. Stat. Methods* 3, 1–27. doi:10.1080/03610917408548446
- Casati, E., Casella, F., Colonna, P., 2015. Design of CSP plants with optimally operated thermal storage. *Sol. Energy* 116, 371–387. doi:10.1016/j.solener.2015.03.048
- Cebecauer, T., Suri, M., 2015. Typical Meteorological Year Data: SolarGIS Approach. *Energy Procedia* 69, 1958–1969. doi:10.1016/j.egypro.2015.03.195
- Fernández-Peruchena, C.M., Bernardos, A., 2015. A comparison of one-minute probability density distributions of global horizontal solar irradiance conditioned to the optical air mass and hourly averages in different climate zones. *Sol. Energy* 112, 425–436.
- Fernández-Peruchena, C.M., Blanco, M., Gastón, M., Bernardos, A., 2015. Increasing the temporal resolution of direct normal solar irradiance series in different climatic zones. *Sol. Energy* 115, 255–263. doi:10.1016/j.solener.2015.02.017
- Gastón-Romeo, M., Leon, T., Mallor, F., Ramírez-Santigosa, L., 2011. A Morphological Clustering Method for daily solar radiation curves. *Sol. Energy* 85, 1824–1836. doi:10.1016/j.solener.2011.04.023
- Gómez Camacho, C., Blanco Muriel, M., 1990. Estimación de la atmósfera estándar de radiación solar a partir del concepto de día claro envolvente. Aplicación a la Plataforma Solar de Almería. *Era Sol.* 40, 11–14.
- Han, J., Kamber, M., 2001. *Data Mining. Concepts and Techniques.* Chapter 8. Cluster analysis, 1st Editio. ed. Morgan Kaufmann Publishers.
- Harrouni, S., Guessoum, a., Maafi, a., 2005. Classification of daily solar irradiation by

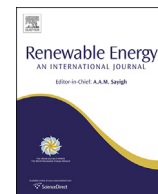
- fractional analysis of 10-min-means of solar irradiance. *Theor. Appl. Climatol.* 80, 27–36. doi:10.1007/s00704-004-0085-0
- Hirsch, T., Schenk, H., 2010. Dynamics of oil-based Parabolic Trough plants - a detailed transient simulation model. *SolarPaces Conf.*
- Hoff, T.E., Perez, R., 2010. Quantifying PV power Output Variability. *Sol. Energy* 84, 1782–1793. doi:10.1016/j.solener.2010.07.003
- Hoyer-Klick, C., Hustig, F., Schwandt, M., Meyer, R., 2009. Characteristic meteorological years from ground and satellite data, in: *SolarPaces Conference*. Berlin, Germany.
- Instituto de Investigación y Formación Agraria y Pesquera, 2009. Estaciones Agroclimáticas [WWW Document]. URL <http://www.juntadeandalucia.es/agriculturaypesca/ifapa/ria/servlet/FrontController> (accessed 7.1.14).
- Iqbal, M., 1983. *An introduction to solar radiation*. Academic Press.
- Islam, M.D., Alili, A.A., Kubo, I., Ohadi, M., 2010. Measurement of solar-energy (direct beam radiation) in Abu Dhabi, UAE. *Renew. Energy* 35, 515–519.
- Islam, M.D., Kubo, I., Ohadi, M., Alili, A.A., 2009. Measurement of solar energy radiation in Abu Dhabi, UAE. *Appl. Energy* 86, 511–515.
- Jiménez-Pérez, P.F., Mora-López, L., 2016. Modeling and forecasting hourly global solar radiation using clustering and classification techniques. *Sol. Energy* 135, 682–691. doi:10.1016/j.solener.2016.06.039
- Jurado, M., Caridad, J.M., Ruiz, V., 1995. Statistical Distribution of the Clearness Index With. *Sol. Energy* 55, 469–473.
- Kang, B.O., Tam, K.S., 2013a. A new characterization and classification method for daily sky conditions based on ground-based solar irradiance measurement data. *Sol. Energy* 94, 102–118. doi:10.1016/j.solener.2013.04.007
- Kang, B.O., Tam, K.S., 2013b. A new characterization and classification method for daily sky conditions based on ground-based solar irradiance measurement data. *Sol. Energy* 94, 102–118. doi:10.1016/j.solener.2013.04.007
- Kaufman, L., Rousseeuw, P.J., 1990. *Finding Groups in Data: An Introduction to Cluster Analysis*. N J.
- König-Langlo, G., Sieger, R., Schmithüsen, H., Bücken, A., Richte, F., Dutton, E.G., 2013. *Baseline Surface Radiation Network (BSRN) Update of the Technical Plan for BSRN Data Management October 2013*.

- Kudish, A.I., Ianetz, A., 1996. Analysis of daily clearness index, global and beam radiation for Beer Sheva, Israel: Partition according to day type and statistical analysis. *Energy Convers. Manag.* 37, 405–416.
- Kumar, A., Gomathinayagam, S., Giridhar, G., Mitra, I., Vashistha, R., Meyer, R., Schwandt, M., Chhatbar, K., 2013. Field experiences with the operation of solar radiation resource assessment stations in India. *Energy Procedia* 49, 2351–2361.
- Lam, J.C., Li, D.H., 1996. Study of solar radiation data for Hong Kong. *Fuel Energy Abstr.* 37, 201.
- Liu, B.Y.H., Jordan, R.C., 1960. The interrelationship and characteristic distribution of direct, diffuse and total solar radiation. *Sol. Energy* 4, 1–19.
- Long, C.N., Dutton, E.G., 2004. BSRN Global Network recommended QC tests, V2.0 1–3.
- MathWorks, 2016. kstest2 [WWW Document]. URL <https://es.mathworks.com/help/stats/kstest2.html>
- Meyer, R., 2010. Recommendations for Bankable Meteorological Site Assessments for Solar Thermal Power Plants, in: *SolarPaces Conference*. Perpignan, France.
- Moreno-Tejera, S., Ramírez-Santigosa, L., Silva-Pérez, M.A., 2015. A proposed methodology for quick assessment of timestamp and quality control results of solar radiation data. *Renew. Energy* 78, 531–537. doi:10.1016/j.renene.2015.01.031
- Moreno-Tejera, S., Silva-Pérez, M.A., Lillo-Bravo, I., Ramírez-Santigosa, L., 2016. Solar resource assessment in Seville, Spain. Statistical characterisation of solar radiation at different time resolutions. *Sol. Energy* 132, 430–441. doi:10.1016/j.solener.2016.03.032
- Moreno-Tejera, S., Silva-Pérez, M.A., Ramírez-Santigosa, L., Lillo-Bravo, I., 2017. Classification of days according to DNI profiles using clustering techniques. *Sol. Energy* 146, 319–333. doi:10.1016/j.solener.2017.02.031
- Moreno, S., 2009. Obtención de modelos kt-kd a partir de medidas registradas en Cáceres y análisis de metodologías de estimación de valores diarios de irradiación directa normal a partir de medidas de irradiación global horizontal diarias. Universidad de Sevilla.
- Morf, H., 2013. A stochastic solar irradiance model adjusted on the Angström-Prescott regression. *Sol. Energy* 87, 1–21.
- Muselli, M., Poggi, P., Notton, G., Louche, A., 2000. Classification of typical meteorological days from global irradiation records and comparison between two Mediterranean coastal sites in Corsica Island. *Energy Convers. Manag.* 41, 1043–1063.

- Ngoko, B.O., Sugihara, H., Funaki, T., 2014. Synthetic generation of high temporal resolution solar radiation data using Markov models. *Sol. Energy* 103, 160–170.
- NREL, 2017. Concentrating Solar Power Projects. Solnova-1 [WWW Document]. URL https://www.nrel.gov/csp/solarpaces/project_detail.cfm/projectID=21
- NREL, 2013. Concentrating Solar Power Projects. Andasol-3 [WWW Document]. URL https://www.nrel.gov/csp/solarpaces/project_detail.cfm/projectID=117
- NREL, 2010. System Advisor Model (SAM) [WWW Document]. URL <https://sam.nrel.gov/>
- Ogunsola, O.T., Song, L., 2014. Restoration of long-term missing gaps in solar radiation. *Energy Build.* 82, 580–591.
- Pagola, I., Gastón, M., Fernández-Peruchena, C., Moreno, S., Ramírez, L., 2010. New methodology of solar radiation evaluation using free access databases in specific locations. *Renew. Energy* 35, 2792–2798.
- Perez, R., Kivalov, S., Schlemmer, J., Hemker, K., Hoff, T., 2011. Parameterization of site-specific short-term irradiance variability. *Sol. Energy* 85, 1343–1353. doi:10.1016/j.solener.2011.03.016
- Powell, G.L., 1984. The clear sky model. *ASHRAE J.* 26, 27–29.
- Rohani, S., Fluri, T., Dinter, F., 2016. Modelling , Simulation and Data Validation of a Solar Thermal Parabolic Trough Plant with Storage. *Conf. Proc. SolarPACES 2015*.
- Rousseeuw, P.J., 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* 20, 53–65. doi:10.1016/0377-0427(87)90125-7
- Salvador, S., Chan, P., 2003. Determining the Number of Clusters / Segments in Hierarchical Clustering / Segmentation Algorithms, in: *Tools with Artificial Intelligence, 2004. ICTAI 2004. 16th IEEE International Conference on.* pp. 575–584. doi:10.1109/ICTAI.2004.50
- Schwandt, M., Chhatbar, K., Meyer, R., Fross, K., Mitra, I., Vashistha, R., Giridhar, G., Gomathinayagam, S., Kumar, A., 2014. Development and Test of Gap Filling Procedures for Solar Radiation Data of the Indian SRRM Measurement Network. *Energy Procedia* 57, 1100–1109.
- Skartveit, A., Olseth, J.A., 1992. The probability density and autocorrelation of short-term global and beam irradiance. *Sol. Energy* 49, 477–487.
- Soubdhan, T., Emilion, R., Calif, R., 2009. Classification of daily solar radiation distributions using a mixture of Dirichlet distributions. *Sol. Energy* 83, 1056–1063. doi:10.1016/j.solener.2009.01.010

- Stein, J., Hansen, C., Reno, M., 2012. The variability index: A new and novel metric for quantifying irradiance and PV output variability. *World Renew. Energy* 1–7.
- Stoffel, T., Renné, D., Myers, D., Wilcox, S., 2010. *CONCENTRATING SOLAR POWER Best Practices Handbook for the Collection and Use of Solar Resource Data*. Contract 146. doi:NREL/TP-550-47465
- Suehrcke, H., McCormick, P.G., 1988. The frequency distribution of instantaneous insolation values. *Sol. Energy* 40, 413–422.
- Tapakis, R., Charalambides, a. G., 2014. Enhanced values of global irradiance due to the presence of clouds in Eastern Mediterranean. *Renew. Energy* 62, 459–467.
- Tovar-Pescador, J., 2008. Modelling the Statistical Properties of Solar Radiation and Proposal of Technique Based on Boltzmann Statistics, in: *Modeling Solar Radiation at the Earth Surface*. Springer, pp. 55–91.
- Tovar, J., Olmo, F.J., Alados-Arboledas, L., 1998. One-minute global irradiance probability density distributions conditioned to the optical air mass. *Sol. Energy* 62, 387–393.
- Tovar, J., Olmo, F.J., Batlles, F.J., 1999. One minute kb and kd probability density distributions conditioned to the optical air mass. *Sol. Energy* 65, 297–304.
- Varo, M., Pedrós, G., Martínez-Jiménez, P., Aguilera, M.J., 2006. Global solar irradiance in Cordoba: Clearness index distributions conditioned to the optical air mass. *Renew. Energy* 31, 1321–1332.
- Wilcox, S., Marion, W., 2008. Users manual for TMY3 data sets, National Renewable Energy Laboratory, Technical Report NREL/TP-581-43156. doi:NREL/TP-581-43156
- Woyte, A., Belmans, R., Nijs, J., 2007. Fluctuations in instantaneous clearness index: Analysis and statistics. *Sol. Energy* 81, 195–206.
- Zagouras, A., Inman, R.H., Coimbra, C.F.M., 2014. On the determination of coherent solar microclimates for utility planning and operations. *Sol. Energy* 102, 173–188. doi:10.1016/j.solener.2014.01.021
- Zagouras, a., Kazantzidis, a., Nikitidou, E., Argiriou, a. a., 2013a. Determination of measuring sites for solar irradiance, based on cluster analysis of satellite-derived cloud estimations. *Sol. Energy* 97, 1–11. doi:10.1016/j.solener.2013.08.005
- Zagouras, a., Kazantzidis, a., Nikitidou, E., Argiriou, a. a., 2013b. Determination of measuring sites for solar irradiance, based on cluster analysis of satellite-derived cloud estimations. *Sol. Energy* 97, 1–11. doi:10.1016/j.solener.2013.08.005
- Zawilska, E., Brooks, M.J., 2011. An assessment of the solar resource for Durban, South Africa. *Renew. Energy* 36, 3433–3438.

ANEXO A



A proposed methodology for quick assessment of timestamp and quality control results of solar radiation data



S. Moreno-Tejera ^{a,*}, L. Ramírez-Santigosa ^b, M.A. Silva-Pérez ^a

^a Department of Energy Engineering, University of Seville, Spain

^b Solar Radiation Unit, Renewable Energy Division, CIEMAT, Spain

ARTICLE INFO

Article history:

Received 23 May 2014

Accepted 15 January 2015

Available online 6 February 2015

Keywords:

Quality control

Timestamp assessment

Solar radiation measurements

ABSTRACT

To evaluate the solar resource at a site, the period of measurements analyzed must be as long as possible. In solar radiation database, a quality control that identifies errors and labels the data by means of different flags or indicators is fundamental. Reading and interpretation of flagged data can usually be tedious due to the large numbers of data that have to be handled.

This article presents a new type of graphical representation that facilitates the identification and interpretation of data quality by using their flagged values. These graphs represent the results of quality control (QC) for up to one year of measurements with any recording frequency on the same graph, making it easier to identify frequent errors such as incorrect timestamp. The utility of this visual tool to identify the most common errors found in quality control of solar radiation data is exemplified by applying it to the QC performed to 4 databases registered at different locations in Spain. The quality control process followed the recommendations of the Baseline Solar Radiation Network (BSRN).

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

The assessment of the solar resource, an essential activity in any solar energy project be it commercial or not, relies on the availability of long time-series of solar radiation and other meteorological data (up to 30 years, according to the WMO recommendations [1]). Ideally, these time-series should be the result of on-site measurements. In many cases they are also generated from satellite images using well-established methodologies and techniques, but these techniques frequently require the use of on-site measured data for model adjustment.

Measuring solar radiation is not an easy task. In addition to the uncertainty associated to the instruments used for this purpose, solar radiation measurement requires careful operation and maintenance (O&M) of the measurement station to avoid or minimize the effect of other sources of error. In this context, quality control and assessment of solar radiation measurements is essential, but this task can be tedious and time-consuming when processing large amounts of data if it is not as automated as possible.

Two main types of errors can affect the measurement of solar radiation: specific sensor errors, mainly related to device quality, and operation and maintenance errors. Muneer and Younes [2,3] describe both types of errors and consider the most usual sources of error to be associated to the sensor itself and its construction. These inherent and sometimes systematic sensor errors add uncertainty to the measurements. These authors cite studies estimating these uncertainties. Other authors estimate the uncertainty associated with improper sensor maintenance. Geuder and Quaschnig [4] provide experimental values for the effect of soiling on sensors, suggesting that the uncertainty depends on the sensor and site-specific characteristics. In some cases the uncertainty associated to some measurement errors cannot be propagated to the measurements and these errors must be identified and corrected because their effect could be of the same order of magnitude as the measurement itself. Errors of this type are usually due to improper use or incorrect operation of the sensor or the measurement station. Incorrect timestamp of recorded measurements is a frequent error that falls in this category. If these errors are not identified and, if possible, corrected, they could affect the evaluation of the solar resource resulting in significant deviations. Small differences in the solar resource assessment can be decisive for considering a project economically feasible or not [5]. Thus, it is essential to perform a quality assessment of solar radiation data before using them.

* Corresponding author. Department of Energy Engineering, Escuela Técnica Superior de Ingeniería, University of Seville, Camino de los Descubrimientos, s/n. 41092, Seville, Spain.

E-mail address: smoreno2@etsi.us.es (S. Moreno-Tejera).

There are many procedures and tools for quality assessment of solar radiation data. Some of the procedures described in the literature concentrate on filtering global horizontal radiation data only [6–9], but there are also procedures that include diffuse [2,3] and direct radiation [10–15]. The best known of them use tests based on the comparison with the physical limits or extreme observed values of the solar radiation components under clear and overcast skies [6,14,16], and filters based on the physical relationship between global, direct and diffuse values [11–14] when they are available. Other tests proposed by some authors are based on the graphic representation of indices defined as ratios between one component and its physical maximum as the clearness index, the diffuse ratio or the beam transmittance index [2,3,10,14,15]. These tests usually graph a beam or diffuse ratio versus the clearness index, defining envelopes that identify outliers based on statistical analysis or the comparison with models that simulate different sky conditions. Additionally, the visual inspection of the graphical representation of the data registered versus time (daily curves) by an experienced user permits the identification of possible errors. Most of procedures propose labeling every record with flags, which provide information about the results of the tests. Handling of this information is often cumbersome, especially when huge amounts of data are to be processed.

This work presents a quick way to display and process the flags resulting from quality controls by means of a graphical representation. For this purpose, the most significant results by applying the BSRN test [17] at four different sites are shown. This visual tool makes it possible to quickly acquire a precise overall idea of the quality of up to one year's measurements regardless of their recording frequency. It also facilitates identification of errors from improper use of equipment (such as timestamp errors or incorrect illumination of sensors) in just one figure. The figures proposed illustrate the results found in the different stages of QC procedures graphically using a color code (in web version).

The results shown are part of the validation studies done for the series generation procedures proposed in the context of the development of the standard, “Solar Thermal Power Plants; Procedure for Generating a Representative Solar Year,” elaborated by the Spanish Association for Standardization and Certification (AENOR) and presented to the IEC/TC-117 as a proposal for the standard. According to this standard, the data for construction of a Representative Solar Year (RSY) must undertake a prior QC

procedure. The data maintenance plan recommended by the Baseline Solar Radiation Network published in September 2013 [17] was selected from the experiences in quality control reviewed.

2. Databases

Data from four measurement stations located at different sites throughout Spain (Fig. 1) were used. Table 1 shows the location and environment, frequency of recording, maintenance characteristics and recording period of each station. Solar radiation sensors were all thermopile-based.

3. Methodology

Data processing was done with specific programs developed in Matlab following a similar sequence in all cases but adapted specifically to the characteristics of the different data files. The procedure included preprocessing of data for format homogenization and detection of recording anomalies. The QC was applied to the database and every data was flagged automatically. The results were analyzed by means of a useful type of graph that helps to speed up the process.

3.1. Quality control of original data

The quality control procedure applied in this work follows the BSRN recommendations published by the World Meteorological Organization [17]. This procedure proposed by Long and Dutton [18] consists of a sequence of three tests to check whether the data are within the limits set. Data that do not comply with the restrictions of any of the limits are flagged with the number of the previous test and the following is not applied to them.

The first test is called “Physically Possible Limits”, and is applied independently to each of the three variables global horizontal irradiance (I_{g0}), diffuse horizontal irradiance (I_{d0}) and direct normal irradiance (I_{bn}). This consists of imposing some basic not-very-strict physical limits:

$$-4 \text{ W/m}^2 < I_{g0} < I_{CS} \cdot \varepsilon_0 \cdot 1.5 \cdot (\cos\theta_z)^{1.2} + 100 \text{ W/m}^2 \quad (1)$$

$$-4 \text{ W/m}^2 < I_{d0} < I_{CS} \cdot \varepsilon_0 \cdot 0.95 \cdot (\cos\theta_z)^{1.2} + 50 \text{ W/m}^2 \quad (2)$$

$$-4 \text{ W/m}^2 < I_{bn} < I_{CS} \cdot \varepsilon_0 \quad (3)$$

where I_{CS} is the solar constant, ε_0 the Sun–Earth distance correction factor and θ_z the solar zenith angle.

The data which fail to pass this test, should not be considered valid and are flagged with “0” (zero). No other test is applied to these flagged data.



Fig. 1. Location of the meteorological stations.

Table 1
Description of data processed.

Province	Source	Environment	Maintenance	Recording interval	Period
Navarre	CENER	Urban	HIGH	1-min	01/01/10 to 12/31/12
Soria	CIEMAT	Agricultural	LOW	5-min	10/14/05 to 04/30/13
Seville	DEVELOPER	Agricultural	HIGH	1-min	01/01/11 to 12/31/12
Almeria	DLR	Agricultural	MEDIUM	1-min	06/26/02 to 12/31/12

The second test is called “Extremely Rare Limits”. This test has more restrictive limits than the one above:

$$-2 \text{ W/m}^2 < I_{g0} < I_{CS} \cdot \varepsilon_0 \cdot 1.2 \cdot (\cos\theta_z)^{1.2} + 50 \text{ W/m}^2 \quad (4)$$

$$-2 \text{ W/m}^2 < I_{d0} < I_{CS} \cdot \varepsilon_0 \cdot 0.75 \cdot (\cos\theta_z)^{1.2} + 30 \text{ W/m}^2 \quad (5)$$

$$-2 \text{ W/m}^2 < I_{bn} < I_{CS} \cdot \varepsilon_0 \cdot 0.95 \cdot (\cos\theta_z)^{0.2} + 10 \text{ W/m}^2 \quad (6)$$

The data outside of these limits are flagged with “1” and must be analyzed before being discarded, since they could be the result of occasional phenomena.

The third and most restrictive test is called, “Comparisons”. This test evaluates two relationships. In the first relationship, measured direct global irradiance is compared with the value calculated from its measured components. The tolerance in this test depends on the value of the zenith angle, as shown in the equations below.

$$|I_{g0}/(I_{d0} + I_{bn} \cdot \cos\theta_z)| < 1.08 \text{ for } \theta_z < 75^\circ \quad (7)$$

$$|I_{g0}/(I_{d0} + I_{bn} \cdot \cos\theta_z)| < 1.15 \text{ for } 93 > \theta_z > 75^\circ \quad (8)$$

This condition can only be applied to records in which the calculated global irradiance is over 50 W/m^2 . In this study, data below 50 W/m^2 were considered to comply with the third test if and only if all three variables had complied with the second.

The second relationship checks that the diffuse percentage of global radiation is not over specific limits:

$$I_{d0}/I_{g0} < 1.05 \text{ for } \theta_z < 75^\circ \quad (9)$$

$$I_{d0}/I_{g0} < 1.10 \text{ for } 93 > \theta_z > 75^\circ \quad (10)$$

This condition can only be applied to records in which the measured global irradiance is over 50 W/m^2 . When lower, the condition imposed is that both GHI and DHI data must have passed the second test to be able to pass the third.

Data that comply these conditions are flagged with “3”, otherwise with “2”. All data flagged with 3 are considered valid. The others should be analyzed to identify the problem.

BSRN quality control tests 1 and 2 impose independent limits on each variable. However, the third procedure imposes two different conditions in which more than one component is used. In this study, GHI and DNI measurements are considered to comply with the third test when they are within the limits defined in Equations (7) and (8). The third test cannot be applied if any of the three variables has not passed the second test. For the DHI measurements to comply with the third test they must be within the limits set in Equations (9) and (10).

3.2. Timestamp test and quality control flags reading

The uncertainty associated to solar radiation measurements is mainly related to the quality of the sensor. There are also other issues related to inadequate installation, operation and maintenance of the sensors that may add uncertainty to the measurement. Some of them, like incorrect timestamp or deficient installation of wiring could be as large as the measurement itself. This type of errors cannot be associated with uncertainty in measurement, but must be identified and corrected or eliminated.

Correct timestamping of solar radiation records is essential to their later use. A slow or fast clock may affect quality control results and any later calculation to be performed to simulate solar system

performance. To identify incorrect timestamps, Ineichen [15] suggests plotting the global clearness index and the beam clearness index versus the solar height on clear days, using a different color for records before and after solar noon. Thus, correct timestamp is evaluated by checking the symmetry of the points. This method is very sensitive to shifts of a few minutes, but when representing a complete year, the identification of the period when the shift was recorded is an awkward task.

A different method is proposed in this study to identify incorrect timestamps (shifts) and at the same time have an overall idea of the quality of the data. This method consists of graphing the flags resulting from the quality control of every variable for a large period in a single graph such as the example shown in Fig. 2.

In this figure, the flags of every day are represented by a vertical line with as many points as number of observations have been recorded. The lines of sunrise, sunset and solar noon throughout the year are included for reference. This lines help to identify the timestamp errors in the recorded data, as will be seen in Section 4.5. This type of graph allows for representation of any number of days. In this work, the graphs display periods of one year, as shown in the horizontal axis.

This example shows the results for the GHI measurements recorded during 2010 at the CENER station. The recording interval at this station was 1 min, which means 1440 total daily observations (see vertical axis). The color of every point informs about the last test passed by every data. In Fig. 2, not all the points are in orange (flag = 3) meaning that not all the data pass the three BSRN tests. One period flagged in yellow and other flagged in turquoise during the nighttime can be appreciated. Data from the yellow period have passed until the second test (flag = 2) and data from the turquoise one haven't passed any test (flag = 0). The horizontal axis scale helps to identify the periods with different flags. This figure will be analyzed more deeply in Section 4.3.

In addition, if all three variables are available, it is suggested that the GHI measured values be plotted over GHI calculated from its components (see first graph in Fig. 3). Two lines representing deviations of $\pm 8\%$ between measured and calculated values are plotted for reference. This graph assists in identifying and quantifying some errors such as will be seen below.

4. Results

The results from the quality control for every station were represented and analyzed. Some of the most interesting graphics

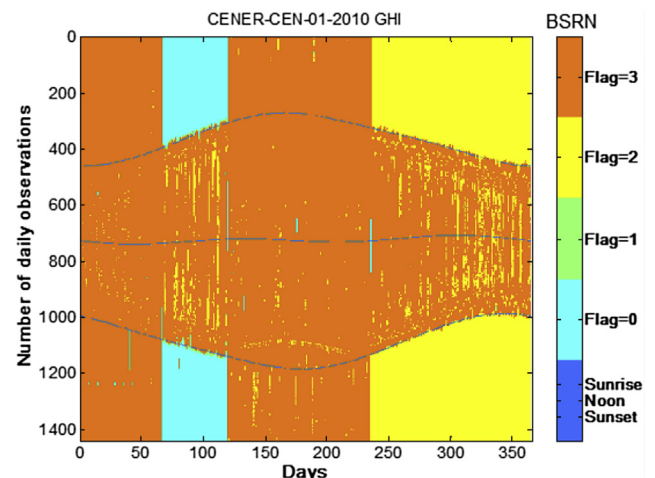


Fig. 2. “At a glance” representation of BSRN-based quality control results.

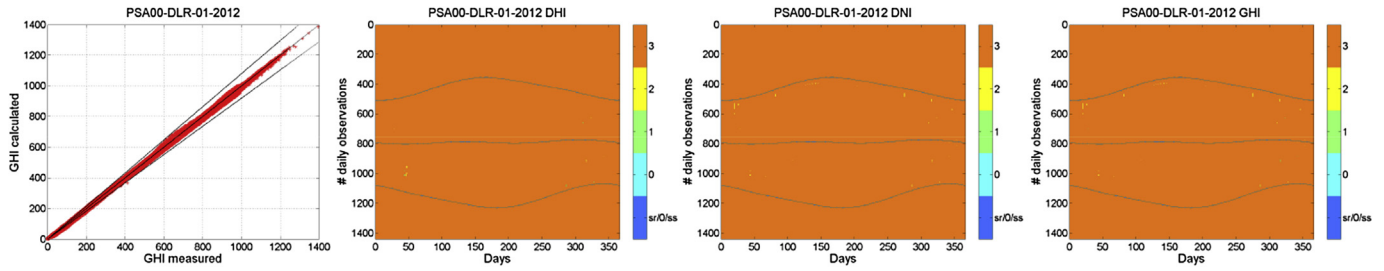


Fig. 3. Annual plots of quality control results for the Almería station data in 2012.

illustrating the identification of frequent errors are presented and discussed in this section.

4.1. Good maintenance

The four graphs proposed for a quick evaluation of the QC results of a solar radiation database are displayed in Fig. 3. In this case, the QC results of the PSA data are analyzed. The first graph represents the GHI measured values in 2012 versus GHI calculated from its components. The other three graphs (one per each variable) show the results of the QC of the PSA database in 2012 with the format and color-coding described in Section 3.2. This station was correctly and carefully maintained and operated. As results show, practically all of the data from the three variables pass the three tests and GHI values measured are very similar to estimated (Fig. 3).

4.2. Dust on sensors

The results of the quality control of the Seville data for 2011 (Fig. 4) show that most of the data pass all three tests, except in isolated cases. Most of these occur during two different periods: a period of daytime data in August and another one near sunset

during the months of September and October. During these periods, Test No. 3 fails. The observation of the graphic that compares estimated GHI to measured GHI in Fig. 4, suggests that the errors in these periods are not very large.

A visual inspection of the daily curves for August (Fig. 5) suggests a possible cause for the errors during this period. The data for August 17 (first graph) pass all 3 tests. This was a very hazy day, due to very high content of aerosols (dust) in the air that results in increased DHI and reduced DNI. Soiling on the sensors usually increases during these episodes, affecting the measurements. This is illustrated with the graph for August 30, where many of the records fail to pass Test No. 3. A few days later (September 5) the sensors had been likely cleaned and all the variables pass all the QC tests.

4.3. Offsets

Fig. 6 shows the quality control results for data from the CENER station in Pamplona (Navarre) for the year 2010. It may be observed how in different periods, DHI and GHI appear flagged as physically impossible (flag = 0) during nighttime. For instance, in the case of DHI and according to Eq. (2), this representation informs us that

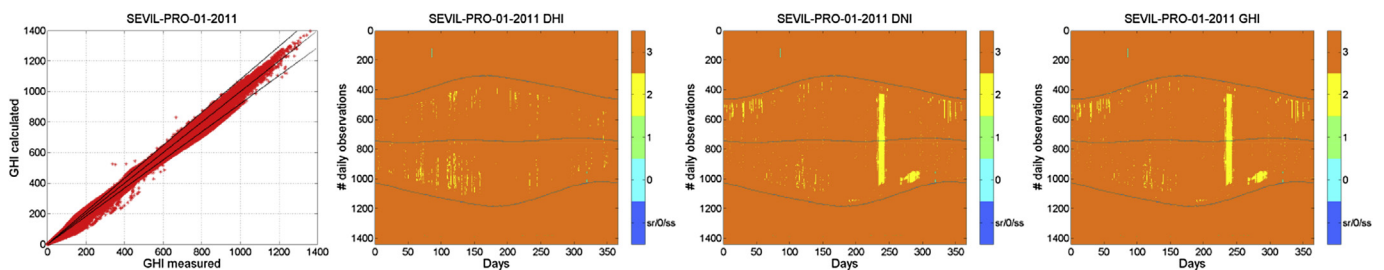


Fig. 4. Annual plots of quality control results for Seville station data in 2011.

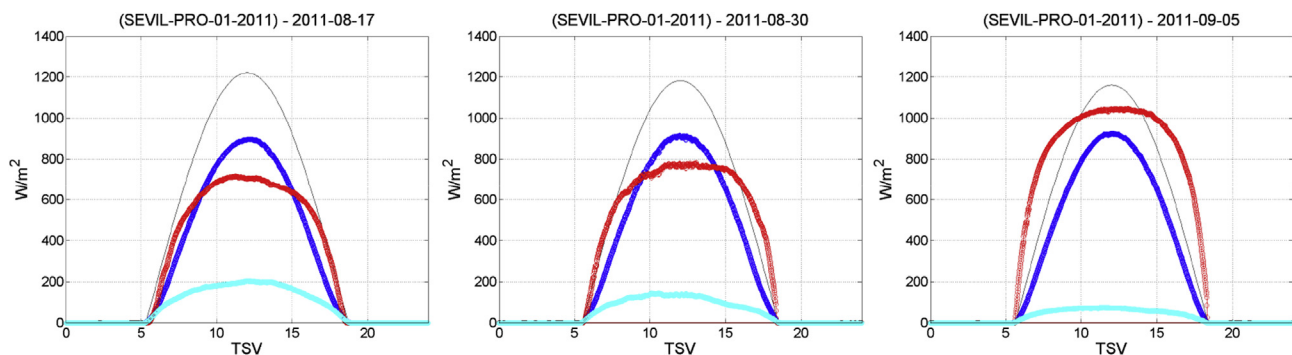


Fig. 5. Daily graphs for the Seville station before, during and after a dust episode. All figures show GHI (blue), DNI (red), DHI (turquoise) and extraterrestrial irradiance (black) over time. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

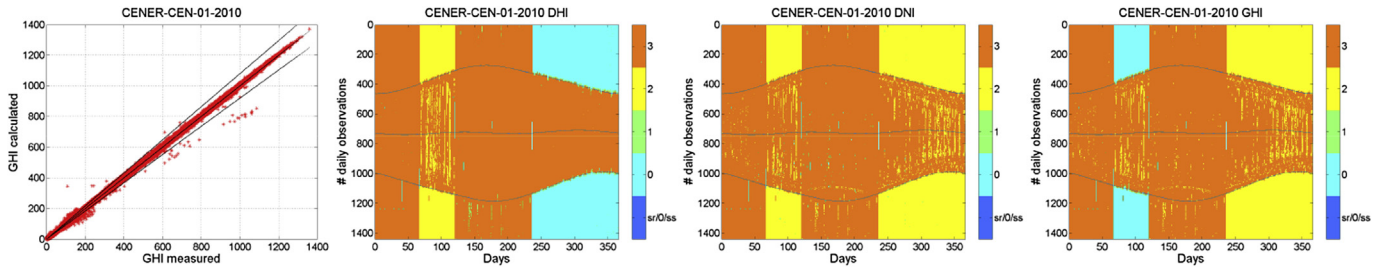


Fig. 6. Annual plots of quality control results for Navarre station data in 2010.

values below -4 W/m^2 or over 50 W/m^2 are being recorded during the period (turquoise colored area in the graph).

A review of the numerical values of DHI for this period shows that they are negative and slightly lower than the limits set in Test 1. These offsets may be due to infrared irradiance losses in the pyranometers during the night [20]. There are procedures in the bibliography for correcting this type of error [21]. As a consequence of the application of the procedure described above, Test No. 3

(“Comparisons”) cannot be applied to GHI or DNI data during these periods.

4.4. Changes in recording frequency

Fig. 7 shows the most significant results of quality control applied to the Soria station data. Of all the stations, the Soria station is the one that had the least maintenance and this is reflected in its

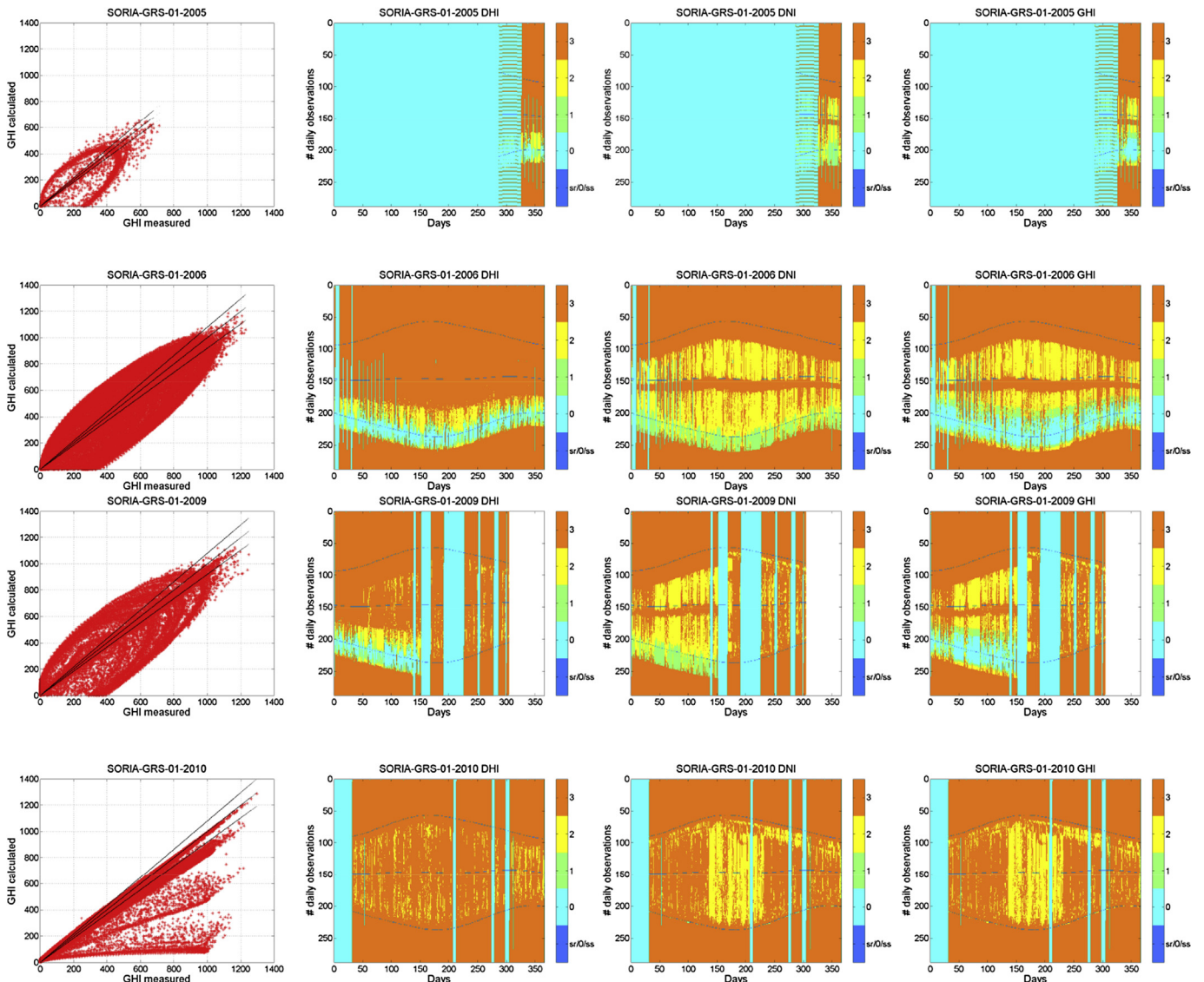


Fig. 7. Annual plots of quality control results for the Soria station data.

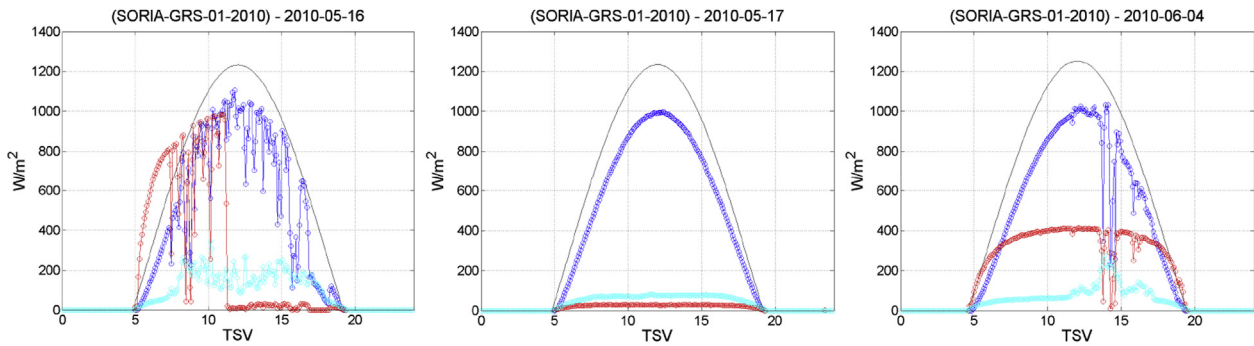


Fig. 8. Daily graphs for the Soria station showing likely tracking errors in the pyrheliometer. In each of the figures GHI (blue), DNI (red), DHI (turquoise) and extraterrestrial irradiance (black) are plotted over time. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

results. In the first row of this figure (2005), thick horizontal turquoise stripes appear among the orange valid data (thin orange lines) in the first month, showing that at those times there was nothing recorded. During the first month, data were recorded every 30 min and then every 5 min. These turquoise stripes disappear with the change in recording interval to 5 min. This kind of graph allows for representation of quality control information for data recorded at varying intervals.

4.5. Timestamp error

The most outstanding observation in the graphics from Soria station is a two-hour time shift from the beginning of the period to mid-2009. As a result of this shift, the data up to 2009 show an asymmetric image of its results shifted two hours with respect to solar noon. These images show how all the GHI, DNI and DHI data do not pass Test No. 3 most of the days, DNI data from around sunset do not pass the second test either, and DHI and GHI data do not pass any of the tests. The data up to mid-2009 were recorded in UTC + 2 and afterwards in UTC.

Once the clock is corrected this asymmetry disappears, and from there to the end of the period, most of the data for the three variables pass at least the second test.

4.6. Tracker alignment error

Other frequent error is detected in the Soria results during the middle months of 2010. In this period, the graphs of Fig. 7 show that GHI and DNI values don't pass Test No. 3. In addition, the comparison of estimated GHI to measured GHI informs us on that the estimated GHI (calculated from the DNI and DHI values) is much lower than measured. In this case, Fig. 8 shows that the

pyrheliometer is not measuring correctly due to incorrect tracking since approximately midday of May 16. This situation was only corrected on August 20.

5. Discussion of results

The examples presented in this work show that the most typical error can be detected with this new type of flags representation from the BSRN quality control. This graph allows for a quick assessment of a large database, whatever the period or quality control procedure, as long as the records are labeled with flags. Furthermore, measurements recorded with different frequency can be assessed in the same graph making easier this task.

In relation to the BSRN procedure two issues should be stressed. The first one is that most of the records affected by the “lens effect” pass the second test, since the upper limit imposed for GHI allows for values of 1400 W/m^2 at times near solar noon. This effect, also known as the magnifying glass or enhancement effect [19], occurs when scattered clouds passing near the apparent location of the Sun reflect part of the radiation onto the sensor. In these situations, global radiation recorded could be higher than the extraterrestrial irradiance. This phenomenon, which has been widely reported at latitudes near the tropics, may also be present at higher latitudes [19]. Thus, these values are not considered “extremely rare”. Fig. 9 shows several examples.

The condition imposed by the third BSRN procedure is based on a relative comparison. As a result, this comparison is stricter for low values of GHI, making it the more restrictive test on very cloudy days on which DNI is practically zero. The pyranometers have to be very well calibrated to comply with Test No. 3 under these conditions.

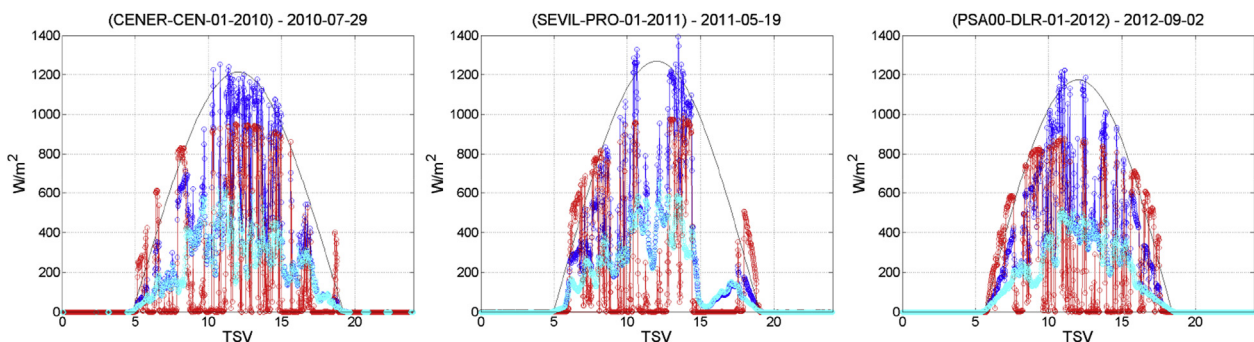


Fig. 9. Daily plots showing the “magnifying glass effect”. All figures show GHI (blue), DNI (red), DHI (turquoise) and extraterrestrial irradiance (black) over time. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

6. Conclusions

This article proposes and presents a new type of graphical representation of the results of quality control of solar radiation measurements. This method is based on plotting color-coded quality control flags for each variable during a certain period of time on the same graph, regardless of the recording frequency. This overall representation of the results facilitates the identification of errors that cannot be associated with the uncertainty of the measurements, and therefore, must be corrected or eliminated. This visual tool was used to examine the results of the quality control to four databases recorded at four stations in Spain, applying the procedures recommended by the BSRN. The examples presented in this paper show that the proposed representation is useful to find and identify frequent errors such as time shifts (incorrect time-stamps of records), changes in recording frequency, misalignment of pyrhemometers, or nighttime and daytime offsets.

This type of graph makes it possible to represent the results of any quality control procedure that labels the data by means of flags, regardless of the variables considered.

Acknowledgments

The work described here was partially developed under the Training Collaboration Agreement 4939 between CIEMAT and Carlos III University of Madrid. The authors are grateful to S. Wilbert (DLR), R. Gómez (CEDER), X. Olano (CENER) and CSP Promoter for data supplied.

The results of this study form part of the validation work done on procedures proposed for the Spanish Association for Standardization and Certification (AENOR) standard, “Solar thermal power plants. Procedure for generating a representative solar year.” The authors wish to thank AENOR and Protermosolar for promoting these standardization studies.

Glossary

I_{g0}	global horizontal irradiance
I_{d0}	diffuse horizontal irradiance
I_{bn}	direct normal irradiance
I_{CS}	solar constant
θ_z	Zenith angle
ϵ_0	Sun–Earth distance correction factor

References

- [1] World Meteorological Organization. Technical regulations. Basic documents N°2. Volume I – General meteorological standards and recommended practices (WMO-No 49). 2011.
- [2] Muneer T, Fairouz F. Quality control of solar radiation and sunshine measurements – lessons learnt from processing worldwide databases. *Build Serv Eng Res Technol* 2002;23(3):151–66.
- [3] Younes S, Claywell R, Muneer T. Quality control of solar radiation data: present status and proposed new approaches. *Energy* 2005;309:1533–49.
- [4] Geuder N, Quaschnig V. Soiling of irradiation sensors and methods for soiling correction. *Sol Energy* 2006;80:1402–9.
- [5] Ratings Fitch. Rating Criteria for solar power projects. Utility-scale photovoltaic, concentrating photovoltaic, and concentrating solar power. 2013. Global Infrastructure & Project Finance.
- [6] Geiger M, Diabate L, Menard L, Wald L. A web service for controlling the quality of measurements of global solar irradiation. *Sol Energy* 2002;73:474–80.
- [7] Moradi I. Quality control of global solar radiation using sunshine duration hours. *Energy* 2009;34(1):1–6.
- [8] Tang W, Yang K, He J, Qin J. Quality control and estimation of solar radiation in China. *Sol Energy* 2010;84(3):466–75.
- [9] Mirás-Avalos JM, Rodríguez-Gómez BA, Meizoso-López MC, Sande-Fouz P, González-García MA, Paz-González A. Data quality assessment and monthly stability of ground solar radiation in Galicia (NW Spain). *Sol Energy* 2012;86:3499–511.
- [10] Maxwell E, Wilcox S, Rymes M. Users manual for SERI QC software, assessing the quality of solar radiation data. Report No. NREL-TP-463–5608. 1617 Cole Boulevard, Golden, Colorado: National Renewable Energy Laboratory; 1993.
- [11] Hay JE. Solar radiation data: validation and quality control. *Renew Energy* 1993;3:349–55.
- [12] Molineaux B, Ineichen P. Automatic quality control of daylight measurements: software for IDMP stations. Technical report: guide to recommended practice of daylight measurement. Vienna: Commission internationale de l'eclairage; 1994. p. 34–42.
- [13] Kendrick D, et al. Guide to recommended practice of daylight measurement. Report no. CIE-108. Wein, Austria: International Commission on Illumination (CIE); 1994.
- [14] Journé M, Bertrand C. Quality of solar radiation data within the RMI solar measurements network. *Sol Energy* 2011;85(1):72–86.
- [15] Ineichen P. Five satellite products deriving beam and global irradiance validation on data from 23 ground stations. University of Geneva; 2011.
- [16] Shi GY, Hayasaka T, Ohmura A, Chen ZH, Wang B, Zha JQ, et al. Data quality assessment and the long-term trend of ground solar radiation in China. *J Appl Meteorol Climatol* 2008;47:1006–16.
- [17] Update of Technical Plan for BSRN Data Management. Baseline Surface Radiation Network. Global climate observing system. 2013. Available at: <http://www.wmo.int/pages/prog/gcos/Publications/gcos-174.pdf> [accessed 21.05.14].
- [18] Long CN, Dutton EG. BSRN Global Network recommended QC tests. V2.0. BSRN Technical Report. 2002. Available at: http://bsrn.awi.de/fileadmin/user_upload/redakteur/Publications/BSRN_recommended_QC_tests_V2.pdf [accessed 21.05.14].
- [19] Tapakis R, G Charalambides A. Enhanced values of global irradiance due to the presence of clouds in Eastern Mediterranean. *Renew Energy* 2014;62:459–67.
- [20] Philipona R. Underestimation of solar global and diffuse radiation measured at Earth's surface. *J Geophys Res* 2002;107(22):15–21. 15–8.
- [21] Younkin K, Long CN. Improved correction of IR Loss in diffuse shortwave measurements: an ARM value added product. Atmospheric radiation measurement program technical report, ARM TR-009. 2004. Available at: https://www.arm.gov/publications/tech_reports/arm-tr-009.pdf [accessed 21.05.14].

ANEXO B



Solar resource assessment in Seville, Spain. Statistical characterisation of solar radiation at different time resolutions

S. Moreno-Tejera^{a,*}, M.A. Silva-Pérez^a, I. Lillo-Bravo^a, L. Ramírez-Santigosa^b

^a Department of Energy Engineering, University of Seville, Spain

^b Solar Radiation Unit, Renewable Energy Division, CIEMAT, Spain

Received 30 July 2015; received in revised form 14 March 2016; accepted 17 March 2016

Available online 4 April 2016

Communicated by: Associate Editor Frank Vignola

Abstract

The characterisation of the solar resource of a site is essential for different phases of solar energy projects. While only rough estimates of yearly levels of solar irradiation (global or direct, depending on the technology) are needed in their very early stages, the required depth of the assessment increases as the project advances, including long-term estimates that can only be obtained through a statistical analysis of a continuous and long-term database of solar radiation measurements. This paper provides the results of a statistical analysis of thirteen years of Global Horizontal Insolation (GHI) measurements and Direct Normal Insolation (DNI) measurements from Seville, Spain (37.4°N, 6.05°W) at different time resolutions, i.e. from annual to nearly instantaneous (5-s). In addition, a new methodology for gap-filling is proposed which keeps the frequency distribution of the original dataset and reduces the uncertainty of the aggregated values (hourly, daily, monthly, yearly) due to the gaps. Some relevant results of this analysis are: (a) the instantaneous values of GHI and DNI have bimodal distributions, although of different characteristics, in agreement with the results of some works developed in similar climate locations; (b) the frequency distributions of the instantaneous and 10-min clearness index (k_t) and beam fraction index (k_b) are almost identical, suggesting 10 min as a good time resolution for the simulation of Concentrated Solar Power (CSP) systems oriented to feasibility analyses; (c) the distributions of hourly k_t and k_b values, show significant differences with respect to the instantaneous ones; (d) the difference between the percentile 99 (P99) of the instantaneous GHI and its maximum value is very high, because of the enhancement effect due to the cloud reflection, while for the DNI the corresponding values are much closer. The comparison with the results of other locations of similar climates suggest that these results can be extrapolated, at least, to other locations of similar climates. Other, more site-specific, results are: (a) the number of typical overcast days in summer is extremely low, while it takes its maximum value in December, suggesting this month as the best for maintenance operations that require halting the operation of CSP plants; (b) the annual mean daily values are 4.98 kW h m⁻² for GHI and 5.68 kW h m⁻² for DNI, with a low inter-annual variability and a greater monthly variability which depends on the season. The monthly and yearly average values from Seville have been compared with three long-term databases derived from satellite images. The best concordance in GHI values is found with NASA's Surface Meteorology and Solar Energy (NASA SSE), but NASA SSE provides significantly higher DNI values compared to the Seville database. A comparison of one year of DNI and GHI measurements recorded at two locations, Durban (South Africa) and Abu Dhabi (United Arab Emirates), with high solar potential is also addressed.

© 2016 Elsevier Ltd. All rights reserved.

Keywords: Solar resource assessment; GHI; DNI; Clearness index

* Corresponding author at: Department of Energy Engineering, Escuela Técnica Superior de Ingeniería, University of Seville, Camino de los Descubrimientos s/n, 41092 Seville, Spain.

E-mail address: smoreno2@etsi.us.es (S. Moreno-Tejera).

Nomenclature

k_t	clearness index	I_{g0}	horizontal global irradiance
$k^{\Delta t}_t$	clearness index of the aggregation time interval Δt	I_o	horizontal extraterrestrial irradiance
k^d_t	daily clearness index	I_{cs}	solar constant
$k^{\Delta t}_b$	beam fraction index of the aggregation time interval Δt	ε_0	Sun-Earth distance correction factor
k_b	beam fraction index	θ_z	solar zenith angle
		I_{bn}	normal direct irradiance
		I_n	normal extraterrestrial irradiance

1. Introduction

Characterising the solar resource of a site is essential for different phases of solar energy projects. While only rough estimates of yearly levels of solar irradiation (global or direct, depending on the technology) are needed in their very early stages, the required depth of the assessment increases as the project advances, including long-term estimates that can only be obtained through a statistical analysis. The statistical characterisation of the resource requires long-term series of solar radiation measurements –between 6 and 30 years, depending on the researcher (Meyer, 2010; Stoffel et al., 2010)– with different requirements in terms of time resolution depending on the specific use or application of the measurements. The time resolution of the solar radiation data used for simulations may have a significant effect on the results of feasibility analyses, depending on the dynamics of the solar energy system under consideration (Hirsch and Schenk, 2010). In the case of CSP plants, where the most relevant variable is the DNI, some authors (Cebecauer & Suri, 2015) suggest that the use of 10-min averaged values is an appropriate choice for simulation oriented to feasibility analyses. High-resolution instantaneous data are required for the detailed analysis of transient processes associated to cloud passages and certain operational procedures, like plant start-up and shut-down.

Availability of ground-measured DNI databases with long and continuous time coverage is scarce. Thus, it is a common practice to derive DNI values from GHI measurements or satellite-based estimates, which are more frequently available. The synthetic generation of DNI series from GHI should take into account the differences with regards to the statistical distribution of these variables in different time intervals (Boland et al., 2013; Dugaria et al., 2015).

Different researchers have studied the statistical characteristics of the solar radiation in different time intervals. Early studies were focused on the analysis of daily values mainly. The most relevant work was that by Liu and Jordan (1960). Some other relevant studies analysed and/or compared the distributions of hourly and instantaneous values (Jurado et al., 1995; Olseth and Skartveit, 1989; Skartveit and Olseth, 1992; Suehrcke and McCormick, 1988). Complete literature reviews about the statistical

characteristics of the solar radiation components in monthly, daily, hourly and instantaneous resolution have been published by Tovar-Pescador (2008) and Fernández-Peruchena and Bernardos (2015). In the last years, the analysis of the instantaneous behaviour of solar radiation attracts special attention in the context of solar radiation component modelling for simulation of solar systems (Bright et al., 2015; Morf, 2013; Ngoko et al., 2014). On the other hand, the analysis of monthly and annual solar radiation values is addressed for quantifying and qualifying the solar resource at a location. Examples of such studies that analyse regions with promising solar potential are presented in (Bachour and Perez-Astudillo, 2014; Islam et al., 2010, 2009; Zawilska and Brooks, 2011). Many of these studies are based on data collected over relatively short periods (less than 5 years).

Solar energy performance simulations need continuous time-series of radiation data. It is very common to find gaps, wrong time-stamps or improperly measured data in a long solar radiation database (Kumar et al., 2013). Designing and applying a good gap filling procedure is not an easy task. A common procedure to estimate monthly and annual values from time series with gaps is to fill the missing days with the average values of the available days of the same month. This procedure may result in inaccurate estimates, depending on the number of missing days. Ogunsola and Song (2014) compares different approaches based on the analysis of time series of solar radiation and other meteorological variables for restoring hourly solar radiation data. Other authors suggest a simple linear interpolation when gaps are up to 3 h and filling gap with neighbouring data or data of the same day from other years when the gaps are greater. The use of the solar radiation components or data from a nearby station is considered the best option (Hoyer-Klick et al., 2009; Schwandt et al., 2014).

In this work, a novel procedure for gap-filling, which keeps the frequency distribution of the original dataset is proposed. This procedure has been applied to 13 years of DNI and GHI measurements from the radiometric station of the Group of Thermodynamics and Renewable Energies (GTER) of the University of Seville (Spain).

An analysis of the solar resource in Seville based on this database is presented. The processed database is

statistically characterised in instantaneous, 10 min, hourly and daily time resolutions, providing results of hourly, daily and annual average values as well as interannual variability of GHI and DNI. The study addresses some issues of great interest for the feasibility analysis of CSP plants and other solar energy projects. A comparison of the distributions of hourly, 10-min and instantaneous values is addressed to assess the differences and its potential use in simulations of solar plants. In the case of monthly and annual values, the aggregated values from the database are compared with long-term datasets –based on satellite images– provided by other public sources (CIEMAT, 2012; NASA, 2012; Sancho et al., 2012) including NASA SSE database. Finally, a comparison of one year of DNI and GHI measurements from Durban (South Africa) and Abu Dhabi (United Arab Emirates) is presented (Bachour and Perez-Astudillo, 2014; Islam et al., 2010, 2009; Zawilska and Brooks, 2011).

2. Experimental data

The data assessed in this study have been recorded by GTER during the period 2000–2012¹, at the meteorological station sited at the Engineering School of Seville with geographical coordinates 37.40°N, 6.01°W. The station collects DNI and GHI as well as data of other variables data in 5-s intervals. The GHI measurements are recorded with a Kipp & Zonen CM21 pyranometer and the DNI measurements with an Eppley Normal Incidence Pyrheliometer (NIP) mounted on the Kipp & Zonen 2AP 2-axis tracker. Both are secondary standard devices according to ISO specifications. The hourly and daily uncertainty specified by the manufacturer is $\pm 2\%$ in the case of the GHI measurements and $\pm 1\%$ in DNI measurements assuming a proper maintenance of the instruments. The GTER station follows a maintenance and calibration procedure according to the recommendations of the instruments manufacturers. Solar radiation sensors are calibrated every two years, approximately. Other measured radiometric and meteorological variables are: diffuse horizontal irradiance (DHI), global irradiance on a tilted plane (37°), air temperature, relative humidity, speed and direction wind, but they will not be analysed in this study.

3. Methodology

3.1. Quality control

A quality control process is essential before using any solar radiation database for assessing the resource at a location. The first step is the application of a visual tool (Moreno-Tejera et al., 2015) to identify incorrect time-stamps, gaps or incorrect measurements. This tool displays

the flags that result from applying the quality control procedures recommended by the Baseline Solar Radiation Network (BSRN). The second step is to identify and characterise the gaps and a daily visual inspection of the graphical representations of DNI and GHI with the extraterrestrial horizontal irradiance (as a reference). With the results of this step, the time stamp is corrected, if necessary. Finally, the days are classified depending on the cloudiness level (Lam and Li, 1996; Li and Lam, 2001; Maafi and Harrouni, 2003).

3.2. Gap filling

Gaps and incorrect data (including time stamp) are usual in this kind of database (Kumar et al., 2013). However, detailed description about the process of identification and replacement of missing or incorrect data is not very common in the literature. The usual solution is to discard these data in the analysis (Bachour and Perez-Astudillo, 2014; Islam et al., 2010, 2009; Zawilska and Brooks, 2011), although other options are available (Ogunsola and Song, 2014; Schwandt et al., 2014). In this work, data collected during 13 years were checked and the gaps were filled following a procedure developed by GTER, resulting in a complete high quality time series of DNI and GHI. The gap filling procedure starts with a classification of the days fulfilling to the data quality and cloudiness level. The procedure is detailed in Table 1.

Correctable days are those where at least one of the main variables (GHI or DNI) has been correctly recorded or shows a gap of less than one hour in the last hour before sunset or in the first hour after sunrise. These days have been corrected using clear sky models mainly (Powell, 1984) and empirical decomposition models developed from GTER measurements. These models are based on widely recognised relations between the clearness index and the beam fraction index (Iqbal, 1983; Skartveit and Olseth, 1992).

Discarded days are those that cannot be considered as correctable after applying the criteria described above; so measurements from nearby stations are used as a reference. The database used for this purpose in this work is provided by the Institute for Research and Training in Agriculture and Fisheries (IFAPA) of the Junta de Andalucía (Regional Administration). This database is free and available on the Web (Instituto de Investigación y Formación Agraria y Pesquera, 2009). Thus, knowing the daily global radiation from the closest IFAPA station, the discarded day is replaced with another *neighbouring* day from the GTER database with similar daily global irradiation (with a tolerance of $\pm 5\%$). This neighbouring day must be within a range of ± 5 days from the discarded day. The limit of ± 5 days is set, because weather patterns often remain constant over a period of 5 days and the sun position does not deviate significantly in such a short period (Kumar et al., 2013). If no neighbouring day from the same year meets

¹ Measurements from the period 2013–2014 are in the process of gap filling. A high number of incidences occurred in the station during this period. These measurements will be taken in account in futures studies.

Table 1
Classification of days according to the level of correction and gap filling procedure applied in each case.

Correctable days	Gap filling procedure
CASE 1. Correct measurements of GHI and DNI with a gap of 1 h or less before sunset/after sunrise	CASE 1A. Cloudy day: linear interpolation
CASE 2. Correct and completed measurements of GHI. Wrong/Not existing measurements of DNI	CASE 1B. Clear day: clear sky model
CASE 3. Correct measurements of GHI with a gap of 1 h or less before sunset/after sunrise. Wrong/Not existing measurements of DNI	CASE 2. Empirical k_t - k_b decomposition model developed from GTER measurements, if DHI is not available
CASE 4. Correct and completed measurements of DNI. Wrong/Not existing measurements of GHI	CASE 3. Combination of cases 1 and 2
Discarded days	CASE 4. Empirical decomposition model developed from GTER measurements, if DHI is not available
CASE 1. File not created	Gap filling procedure
CASE 2. Wrong measurements of GHI and DNI	The discarded day in the GTER database is replaced by another day's solar radiation from the station, assuming same accumulated daily radiations at both GTER and IFAPA stations
CASE 3. Gap longer than 1 h close to the sunrise/sunset in GHI and DNI measurements	
CASE 4. Gap longer than 30 min	

the irradiation condition, a day from another year is chosen.

3.3. Statistical characterisation

The analysis of the GTER database pursues two objectives: characterising the distribution of these variables in high and medium time resolution and quantifying the solar potential in low time resolution. The probability density function of instantaneous datasets is compared with the results of similar works published by other authors at other locations. Then, a comparison between the PDFs of instantaneous, 10-min and hourly datasets is made. Two dimensionless indexes are used in order to avoid the effect of the deterministic patterns imposed by Sun position and the Sun-Earth distance. The sky-clearness index k_t and the beam fraction index k_b help characterise the state-of-the-sky.

The k_t index (Liu and Jordan, 1960) is defined as the ratio of the GHI to the solar extraterrestrial irradiance on horizontal surface I_0 :

$$k_t = \frac{I_{g0}}{I_0} = \frac{I_{g0}}{I_{cs} \cdot \varepsilon_0 \cdot \cos \theta_z} \quad (1)$$

where I_{g0} is the horizontal global irradiance, I_{cs} is the solar constant, ε_0 is the Sun-Earth distance correction factor and θ_z is the solar zenith angle.

The beam fraction index, k_b , is based on a similar concept. This index is defined as the ratio of DNI I_{bn} to the maximum irradiance available at the top-of-the-atmosphere in normal direction I_n :

$$k_b = \frac{I_{bn}}{I_n} = \frac{I_{bn}}{I_{cs} \cdot \varepsilon_0} \quad (2)$$

Both indexes have been used in the literature in different aggregation intervals (instantaneous, hourly, daily and monthly). Depending on the aggregation interval Δt , the indexes are denominated $k_t^{\Delta t}$ or $k_b^{\Delta t}$, with the following definitions:

$$k_t^{\Delta t} = \frac{\int_{\Delta t} I_{g0} dt}{\int_{\Delta t} I_0 dt} \quad (3)$$

$$k_b^{\Delta t} = \frac{\int_{\Delta t} I_{bn} dt}{\int_{\Delta t} I_n dt} \quad (4)$$

Following other authors (Iqbal, 1983; Kudish and Ianetz, 1996; Lam and Li, 1996), the daily clearness index k_t^d is used to classify the days according to the state of the atmosphere (clear, cloudy, overcast).

Annual and monthly solar radiation values are often used in the first steps of feasibility analyses of solar energy projects. The gap-filling methodology helps reduce the uncertainty due to the gaps in the datasets of the annual and monthly solar radiation values. The monthly and annual mean daily values of GHI and DNI for the period 2000–2012 have been obtained and compared with datasets from others locations.

4. Results

4.1. Quality control

The results of the application of the BSRN quality control procedures are displayed and analysed by means of a visual tool described in (Moreno-Tejera et al., 2015). This methodology permits to identify incorrect time-stamps, gaps or measurements that do not pass the BSRN tests with the help of only two types of graphs for each year. The left most graph in Fig. 1 represents the measured GHI data versus the calculated GHI values from its components, before gap filling, for the year 2001. The middle and the rightmost graphs represent the value of the last BSRN test passed by the GHI and the DNI data, respectively, for the same year.

4.2. Gap filling

From the total recorded days during the period 2000–2012, only 9% were classified as discarded and 19%

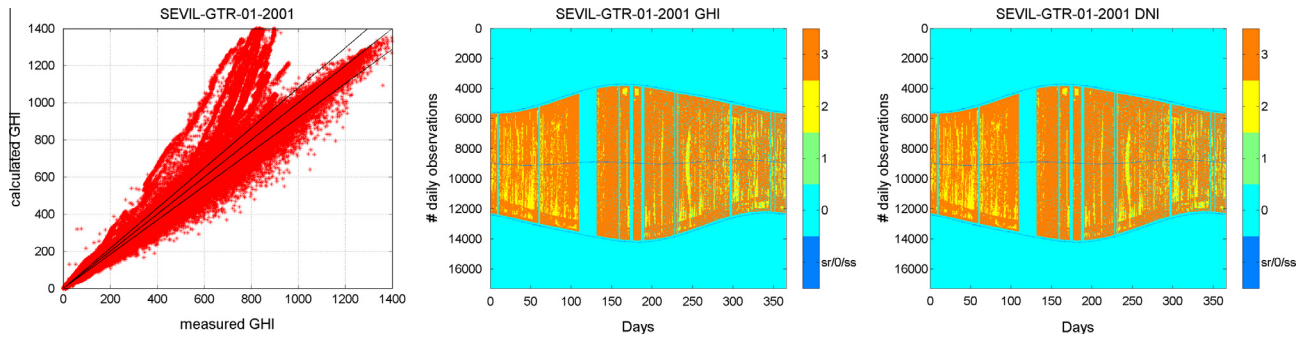


Fig. 1. Quality control results of the year 2001 from the GTER database. The leftmost graph represents the measured GHI data versus the calculated GHI values from its components. The middle and the rightmost graphs display the flags from the BSRN quality-control procedure of the GHI and DNI measurements. The lines of the sunrise, solar noon and sunset (in blue colour) help analyse the results. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

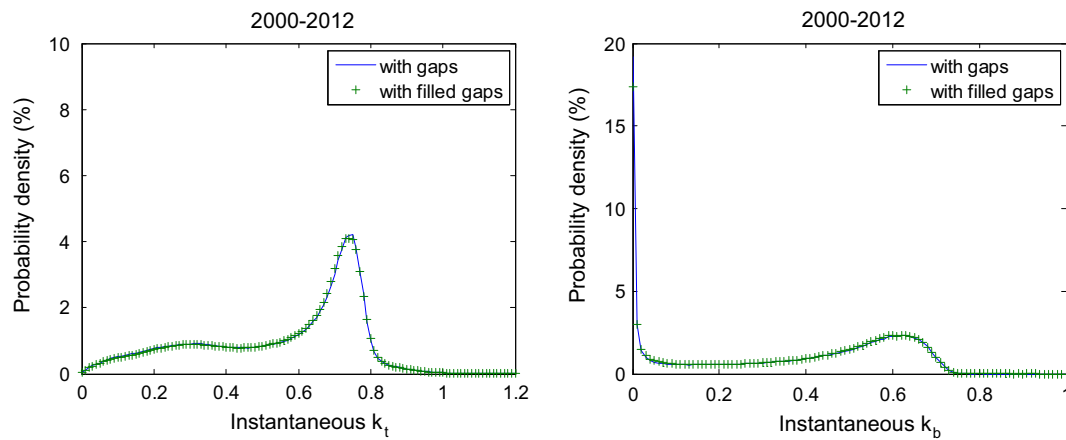


Fig. 2. Probability density function of the clearness index and the direct transmittance index calculated from the database before and after the gap-filling process.

as correctable. The rest of the days (72%), were recorded without any anomalies. The analysis of the probability density functions of the k_t and k_b indexes calculated from the completed (after the application of the gap-filling procedure) and the original databases shows (Fig. 2) that the completed database keeps the statistical characteristics of the original one, with no visible differences in either of the indexes.

4.3. Statistical characterisation of instantaneous values

Several authors (Jurado et al., 1995; Skartveit and Olseth, 1992; Suehrcke and McCormick, 1988; Tovar et al., 1999, 1998; Varo et al., 2006; Woyte et al., 2007) have analysed the instantaneous k_t index, concluding to similar results. The distribution shows a bi-modal character, more pronounced as the optical air mass decreases. Notwithstanding, similar studies developed in other locations show an opposite behaviour (da Assunção et al., 2003; Gansler et al., 1995). A recent work (Fernández-Peruchena and Bernardos, 2015) analyses 9 locations representative of the main Köpen-Geiger classification climate (Peel et al., 2006), revealing different behaviour as a function of the climate.

In the Köpen-Geiger classification map, Seville is identified by the symbols “Csa”, corresponding to a temperate climate with dry and hot summer. This specific type of climate is not included in the Fernández-Peruchena and Bernardos (2015) study. The behaviour of the instantaneous k_t respect to the air mass in Seville is shown in Fig. 3. The instantaneous k_t distributions for different air masses show two predominant states, clear and cloudy, in agreement with the results of (Jurado et al., 1995; Suehrcke and McCormick, 1988). The maximum peak of k_t is approx. 0.76 for the lowest optical air mass studied, but decreases when the optical air mass increases. The lower peak, observed at low k_t values, is less defined due to the great variety of clouds that generate the diffuse component.

A bi-modal behaviour is also identified in the k_b distribution curves. In this case, both observed peaks are pronounced (Fig. 3, right side). The left peak is located in the 0–0.02 interval, corresponding to overcast conditions when the DNI is close to zero. This peak is the highest one, except for low values of the optical air mass. The right peak is located in high k_b values, corresponding to clear sky conditions. The shape of the right peak is smoother because of the different atmospheric turbidity conditions

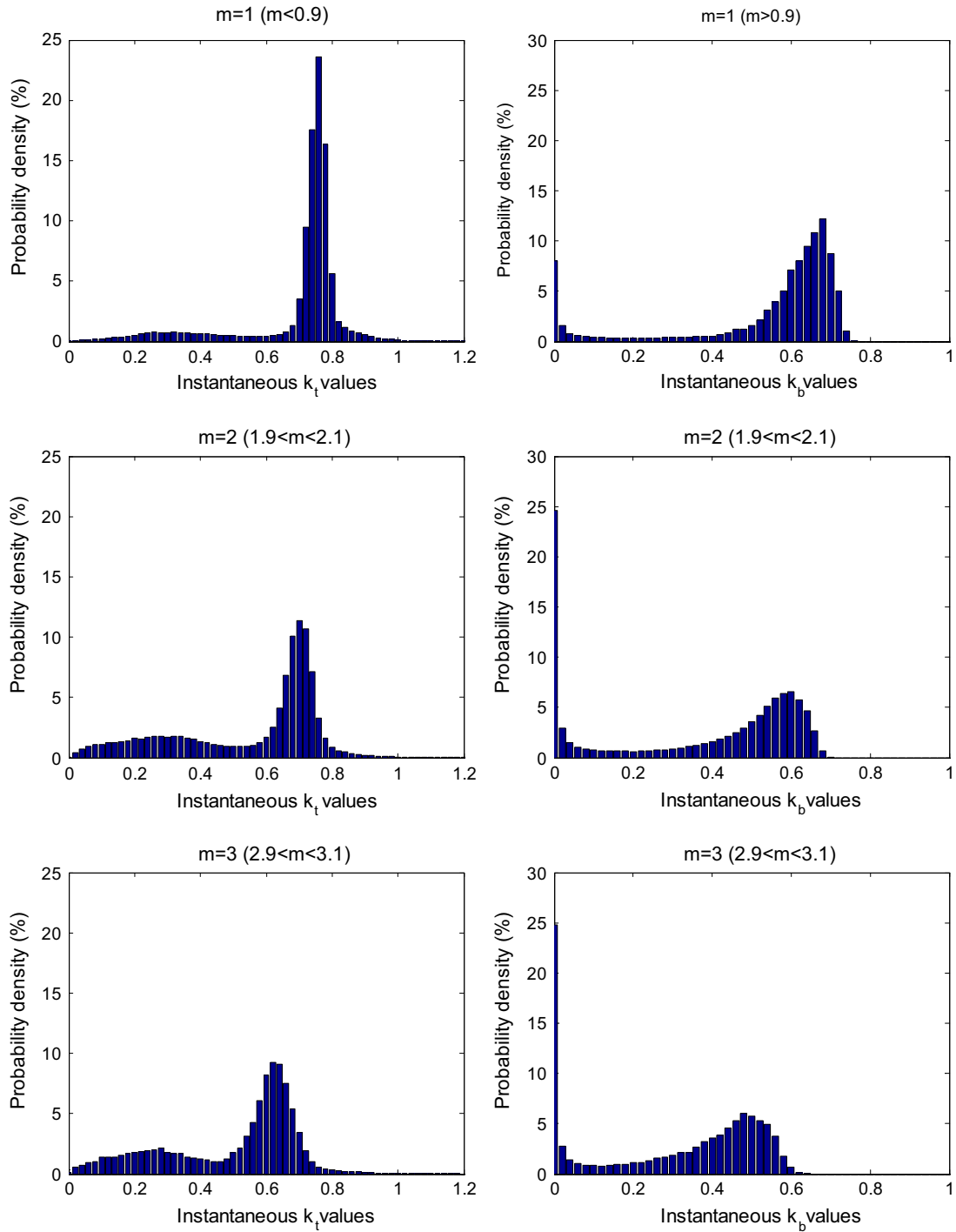


Fig. 3. Instantaneous k_t and k_b probability density functions for different air masses.

for clear skies; similarly to k_t , the right peak of the k_b distribution decreases when the air mass increases. The effect of atmospheric turbidity increases with increasing optical air mass. This effect is also identified in cloudy conditions: the left peak is higher when the optical mass increases due to the higher blocking effect of the clouds. These results are in concordance with the studies of Tovar et al. (1999, 1998) performed with a 1-min dataset from Granada and of Varo et al. (2006) performed with a 5-min dataset from Córdoba.

Both sites are located in the South of Spain and identified as “Csa” climates according to the Köppen-Geiger classification. In all these cases, the maximum values of k_t (0.75–0.77) and k_b (0.69) indexes indicate the predominance of clear-sky conditions with low atmospheric turbidity.

According to Tovar-Pescador (2008), the distribution of 10-min values can be considered as instantaneous, although most authors (Jurado et al., 1995; Skartveit and Olseth, 1992; Suehrcke and McCormick, 1988; Tovar

et al., 1999, 1998; Varo et al., 2006), study only the distributions of 1-min and 5-min values, concluding that these time resolutions can also be considered as instantaneous. However, hourly values have a different distribution (Gansler et al., 1995; Suehrcke and McCormick, 1988).

The probability density function of the k_t and k_b indexes recorded every 5 s, averaged every 10 min and hourly intervals are graphed in Fig. 4. The distributions of the instantaneous values and the 10-min averages are very similar, showing the same characteristic bimodal behaviour with slight differences, which are negligible in the case of k_b . However, the differences between hourly and instantaneous indexes are more significant, especially for k_t , where the left peak is not formed. Averaging the instantaneous values in hourly intervals results in an increase of the frequency of intermediate sky states (neither totally clear nor totally cloudy) and a ‘triangular’ shape of the distribution of hourly k_t values.

The highest values of k_t (greater than 1.0 in some cases) can be attributed to the ‘enhancement effect’ of the solar irradiance due to the reflection from the base of the clouds (Tapakis and Charalambides, 2014) producing instantaneous records of up to 1583 W m^{-2} in GHI. The highest DNI values (up to 1038 W m^{-2}) are recorded in partially cloudy days (see Fig. 5), suggesting a possible participation of the clouds in these high values. The corresponding P99 (percentile 99) values are 1009 W m^{-2} and 960 W m^{-2} , respectively. The difference between the maximum observed and the P99 values is much greater in the case of GHI.

The probability density functions of the levels of GHI and DNI are graphed for 10-min and hourly averages in Fig. 6. These figures, as well as the corresponding cumulative density functions, are useful to select the design point of a solar plant (usually corresponding to the P90 or P95 value). The maxima of GHI and DNI in the 10 min distributions are located in the low values. A second peak is observed in the DNI distribution around 850 W m^{-2} . When a day in Seville is totally clear, values of this level are recorded even for solar altitudes lower than 20° . It is

also remarkable that more than 66% of the DNI values are higher than 300 W m^{-2} , often considered as the threshold for the operation of solar thermal power plants.

4.4. Statistical characterisation of hourly values

Perhaps 1-h time series are the most frequently used in feasibility analyses of solar power plants. The 10-min and hourly averaged values of irradiance are compared in Fig. 6. In the case of GHI, the higher differences between both temporal resolutions are observed at low levels. However, in the distributions of DNI values, high hourly values are less frequent than the 10-min ones. 47.4% of the 10-min values are higher than 600 W m^{-2} , while for the hourly values this fraction is only 42.4%. This is a consequence of the 1-h averaging of instantaneous values.

Fig. 7 shows the daily mean of GHI and DNI hourly values. The higher levels of solar radiation are reached in July, in both cases. The mean maximum value of hourly GHI is 964 W m^{-2} and is achieved at the solar noon close to the summer solstice, when the solar altitude reaches a maximum of 76.1° . Around this point, the mean hourly values seem to spread symmetrically respect to the solar noon, decreasing as the day of the year goes more and more away from the summer solstice. This trend is strongly linked to the evolution of the zenith angle, because the global radiation is measured on horizontal surface. The frequency of high hourly DNI values is also highest near the summer solstice, although they are observed in all seasons. The number of hours with DNI higher than 600 W m^{-2} (almost 8 h in all summer days) is also remarkable.

4.5. Statistical characterisation of daily values

The k_t^d is used in the literature to help classify the days according to the state-of-the-sky. Several boundaries have been proposed to classify the sky conditions in clear, partially cloudy and cloudy skies (Iqbal, 1983; Kudish and Ianetz, 1996). In this work, k_t^d values of 0.3 and 0.65 are

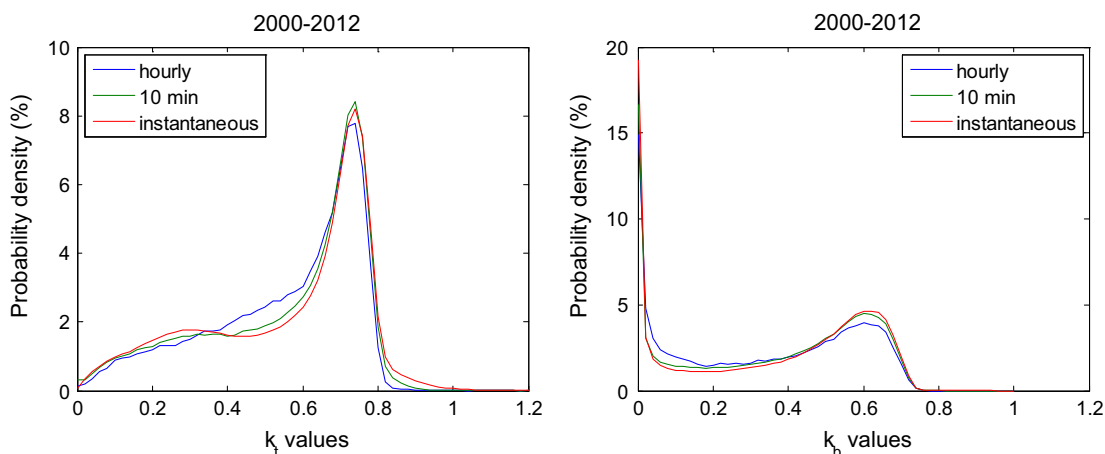


Fig. 4. Probability density function of the k_t (left) and k_b (right) indexes recorded every 5 s, averaged every 10 min and hourly intervals.

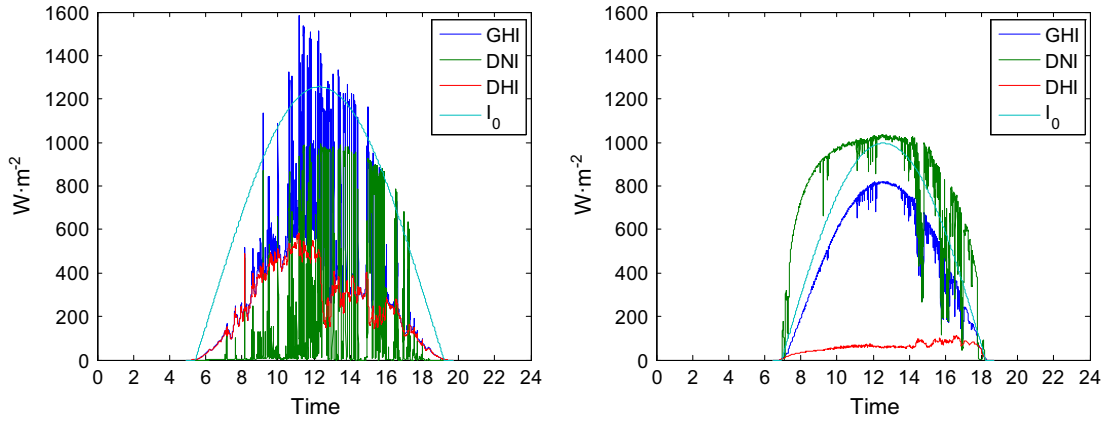


Fig. 5. Daily graphs from Seville station for the days when maximum GHI (left) and DNI instantaneous values (right) were recorded.

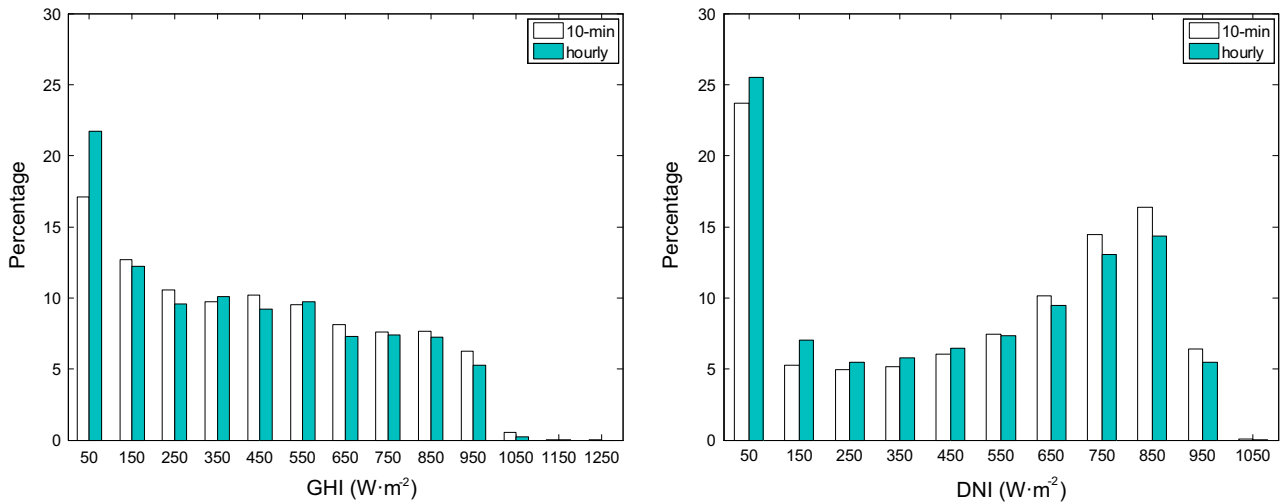


Fig. 6. Probability density function of the 10 min and hourly GHI and DNI values.

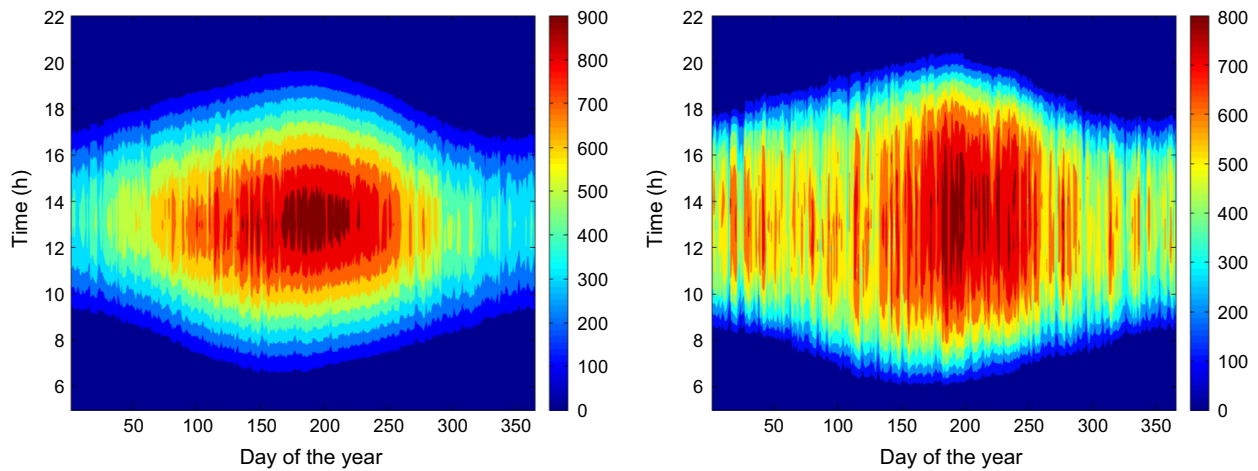


Fig. 7. Mean hourly values of GHI (left) and DNI (right) for all days in the period 2000–2012.

used to define cloudy and clear skies (Lam and Li, 1996). As shown in Fig. 8, the probability density function of k_t^d has a clear maximum between 0.65 and 0.72 corresponding to more than the 35% of the days. The average numbers of

overcast, part-cloudy and clear days every month is graphed in Fig. 8. The highest number of cloudy (6 days) and partially cloudy (19 days) days is reached in December. According to this statistic, this month could be suggested

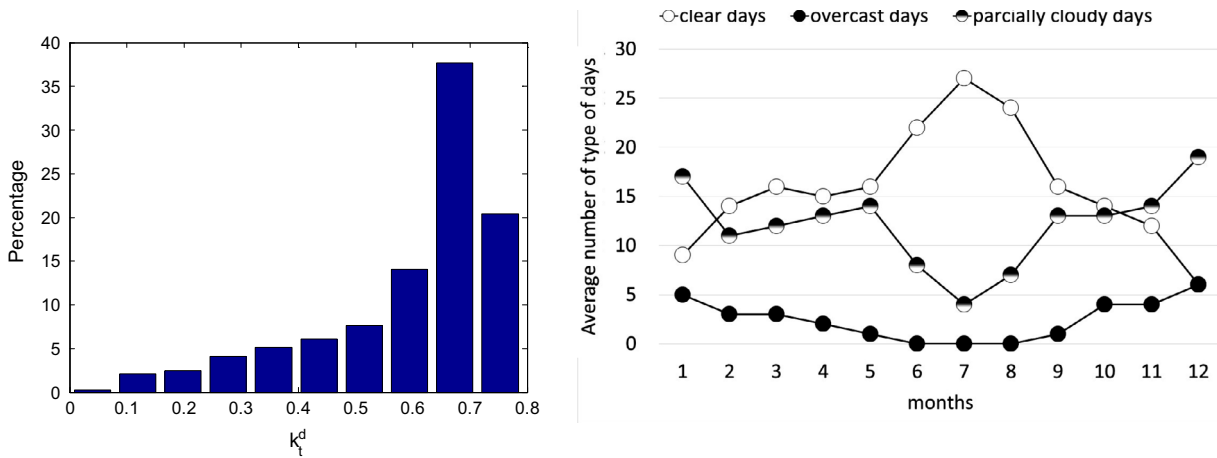


Fig. 8. Probability density function of k_t^d and monthly average of number of days according to the state of the sky.

to perform scheduled maintenance operations that require halting the production in a Solar Thermal Power (STP) plant. On the other hand, June, July and August are the months with higher number of clear days, more than 20 in all cases. In these months, cloudy days are very rare.

The annual averages of the daily GHI and DNI values for the period 2000–2012 are represented in the respective box and whisker plots (see Fig. 9). Both graphs show a very stable range of daily values. The mean value of GHI is very close to 5 kW h m^{-2} with small variations along the period, while the daily average of DNI –which shows greater changes from year to year– is around 5.7 kW h m^{-2} . The highest DNI values were registered in 2012 and 2005.

4.6. Statistical characterisation of monthly and annual values

The assessment of the solar resource for feasibility analyses of solar energy projects is frequently based on the distribution of monthly and annual values of GHI and/or DNI (Pagola et al., 2010). Thanks to the gap-filling procedure applied to the database in this study the monthly values of GHI and DNI have been assessed with a low uncertainty compared to other analyses. The monthly

and annual averages of daily GHI, DNI and k_t^d from the 13 years are shown in Table 2. The maximum and minimum deviation from the average and the standard monthly deviation are also provided in this Table.

The annual mean daily value of GHI for the period of analysis is very stable around 4.98 kW h m^{-2} with a range of interannual variability of -3% to $+5\%$, as shown in Table 2 and in Fig. 9. The standard deviation of the average monthly values is 2.05 kW h m^{-2} , which denotes a high variability between the different months of the year. The maximum GHI monthly mean daily value is obtained in July (7.81 kW h m^{-2}) and the minimum in December (2.17 kW h m^{-2}) with a difference of 72% between them with respect to the highest value. If the same analysis is done with the k_t^d values, the difference is less than 26%, with a maximum of 0.69 and a minimum of 0.51 for the same months. The annual mean of k_t^d is 0.61. This indicates the predominance of clear and partially cloudy days in the Seville climate. The main cause of the high variability of the GHI values is the variation of the solar altitude and the daylight duration along the year.

Regarding DNI, the annual mean daily value is 5.68 kW h m^{-2} , with a higher range of interannual

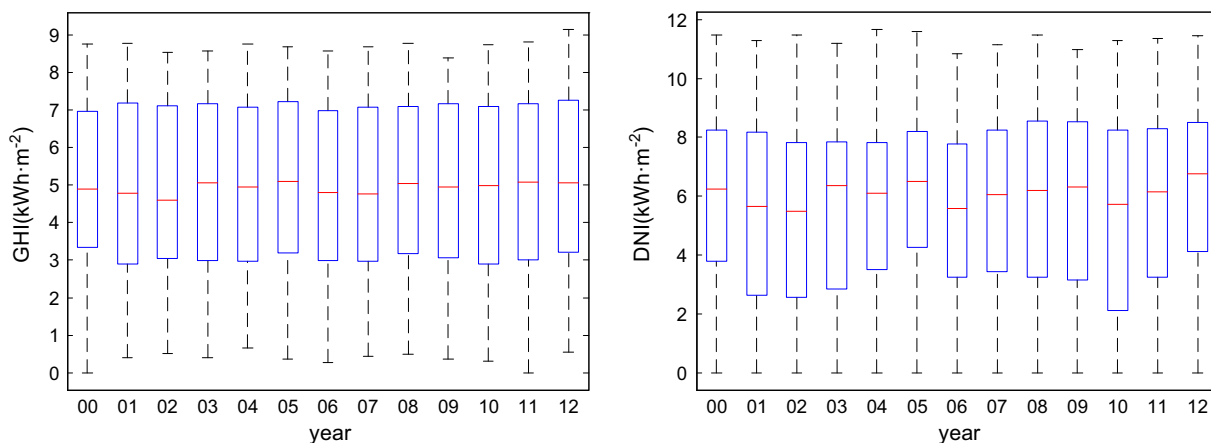


Fig. 9. Box and whisker plots of GHI (left) and DNI (right) daily values of each year in the period 2000–2012.

Table 2
Monthly and annual mean daily values of GHI, DNI and k_t^d in Seville.

Month	Global Horizontal Radiation Diff From Monthly Av.			k_t^d			Direct Normal Radiation Diff From Monthly Av.		
	Average (kW h m ⁻²)	Min. (%)	Max. (%)	Average	Min.	Max.	Average (kW h m ⁻²)	Min. (%)	Max. (%)
January	2.56	-21%	21%	0.54	0.43	0.66	4.03	-32%	51%
February	3.48	-21%	26%	0.56	0.44	0.71	4.62	-40%	61%
March	4.58	-11%	11%	0.57	0.50	0.63	4.86	-21%	31%
April	5.80	-12%	13%	0.59	0.53	0.67	5.74	-28%	31%
May	6.85	-14%	10%	0.62	0.53	0.68	6.59	-30%	21%
June	7.65	-6%	6%	0.66	0.62	0.70	7.88	-13%	17%
July	7.81	-3%	5%	0.69	0.67	0.73	8.76	-10%	8%
August	6.96	-4%	7%	0.68	0.65	0.71	7.80	-7%	15%
September	5.33	-10%	8%	0.61	0.56	0.67	5.86	-27%	22%
October	3.76	-14%	17%	0.56	0.48	0.61	4.61	-30%	25%
November	2.73	-13%	15%	0.54	0.46	0.61	3.93	-22%	28%
December	2.17	-23%	18%	0.51	0.39	0.60	3.42	-38%	39%
Annual avg.	4.98	-3%	5%	0.61	0.59	0.64	5.68	-6%	8%
st.dev	2.05	-	-	0.06	-	-	1.75	-	-

variability (-6% to +8%) and a monthly standard deviation of 1.75 kW h m⁻². The maximum monthly mean daily value is obtained in July (8.76 kW h m⁻²) and the minimum in December (3.42 kW h m⁻²). In this case, the maximum difference between months is 61% with respect to the highest value, showing the importance of knowing the monthly distribution.

The mean monthly value of k_t^d exceeds the lower limit of clear-sky state proposed in Section 4.5 during June, July and August. These months are the steadier.

4.7. Comparison with other sources

The main network of meteorological stations equipped with pyranometers and pyrhemometers in Spain, belongs to the Meteorology State Agency (AEMET). The distance from Seville to the closest AEMET station is 75 km. AEMET also provides a Solar Radiation Atlas in its website (Sancho et al., 2012), based on satellite images estimates obtained from the EUMETSAT European meteorological satellites. A specific solar resource map for Spain (ADRASE) is also available on the website of CIEMAT, although this tool only provides monthly and annual GHI values. Another source frequently used is NASA's Surface Meteorology and Solar Energy (SSE) database. This database provides monthly averages of GHI, DNI and k_t^d from 22 years of satellite images estimates (NASA, 2012). The mean values supplied by these sources are compared in Table 3.

4.7.1. Global Horizontal Insolation

Compared to the average values of the GTER database, NASA SSE provides the most similar monthly values with an annual difference of -1.6%. The AEMET values are higher than those of GTER in all months with monthly differences between 2% and 10% (larger differences for months with higher variability). The average annual value provided by AEMET is close to the maximum annual value

recorded by GTER. The values of ADRASE are similar to AEMET with small deviations in most months that compensate to result in a similar annual value. The monthly mean k_t^d from NASA is in good agreement with the results of GTER, although these are slightly higher.

4.7.2. Direct Normal Insolation

The comparison between GTER and NASA SSE shows larger differences in monthly DNI values. In March, the difference is 24.2%. For this month, NASA SSE provides an average value which is close to the maximum value recorded in Seville by GTER in that month (see Table 2). The differences are smaller in the summer months, but the mean annual value provided by NASA SSE also exceeds the maximum annual value recorded by GTER.

4.8. Comparison with other locations

The results for Seville are compared to one year of GHI and DNI measurements recorded in Durban (South Africa, latitude 29° 58'S) in 2007 (Zawilska and Brooks, 2011) as well as with results from a similar study carried out in Abu Dhabi (UAE, latitude 24° 26'N) for the same year (Islam et al., 2010, 2009). Fig. 10 shows the comparison of the monthly and annual mean daily values of GHI and DNI recorded in 2007. In general, the monthly distributions are very different. Regarding GHI values, Abu Dhabi and Durban show a symmetrical monthly distribution with respect to the 5 kW h m⁻² line, obviously conditioned by the respective latitudes. The distribution of monthly values for Seville is similar to Abu Dhabi but with larger differences between summer and winter, and more extreme values in Seville. In annual values, Abu Dhabi shows the highest GHI mean value, although the annual maximum DNI is recorded in Seville (5.7 kW h m⁻², see Table 4). Durban and Abu Dhabi show more regular monthly DNI values -between 4 and 6 kW h m⁻². Seville shows a higher variability with values in the range

Table 3
Monthly and annual mean daily values of GHI, DNI and k_t^d in Seville.

Month	GHI (kW h m^{-2})				DNI (kW h m^{-2})		k_t^d	
	GTER	AEMET	NASA SSE	ADRASE	GTER	NASA SSE	GTER	NASA SSE
January	2.56	2.72	2.50	2.5	4.03	4.44	0.54	0.53
February	3.48	3.66	3.44	3.7	4.62	5.18	0.56	0.55
March	4.58	5.03	4.70	5.1	4.86	6.03	0.57	0.58
April	5.80	6.14	5.63	6.2	5.74	6.11	0.59	0.56
May	6.85	6.99	6.47	7.1	6.59	6.65	0.62	0.58
June	7.65	7.88	7.59	8.0	7.88	8.5	0.66	0.65
July	7.81	8.10	7.79	7.8	8.76	9.23	0.69	0.68
August	6.96	7.20	6.93	7.1	7.80	8.46	0.68	0.67
September	5.33	5.78	5.32	5.7	5.86	6.71	0.61	0.61
October	3.76	4.02	3.59	4.1	4.61	4.81	0.56	0.53
November	2.73	2.92	2.62	3.0	3.93	4.25	0.54	0.51
December	2.17	2.33	2.12	2.3	3.42	3.76	0.51	0.49
Annual avg.	4.98	5.23	4.90	5.2	5.68	6.18	0.61	0.58

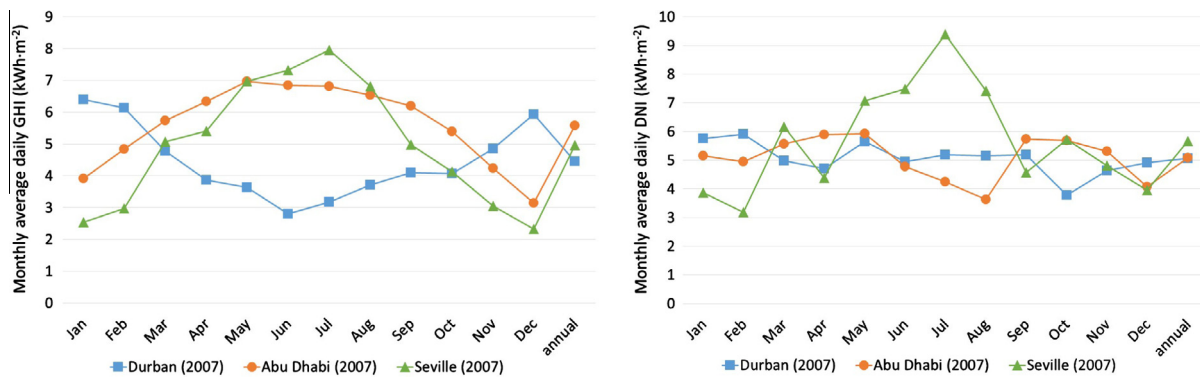


Fig. 10. Monthly and annual mean daily values of GHI (left) and DNI (right) recorded in 2007 in Durban, Abu Dhabi and Seville and average of the 13-year measurements recorded in Seville.

Table 4
Annual mean daily values of GHI and DNI recorded in 2007 in Durban, Abu Dhabi and Seville.

	Durban	Abu Dhabi	Seville
GHI (kW h m^{-2})	4.45	5.58	4.96
DNI (kW h m^{-2})	5.07	5.08	5.66

3–9 kW h m^{-2} . This variability has a significant effect on the design of solar energy systems that include energy storage; thus the need to assess the seasonal distribution of the solar resource and not only the annual values.

5. Conclusions

This work presented the results of the analysis of 13 years of GHI and DNI measurements recorded in Seville, Spain. The time series have previously been checked and corrected using a novel gap-filling procedure described in this paper which conserves the statistical distribution of the original series. There is a perfect correspondence between the probability density functions of the processed dataset of instantaneous values and the original one.

The analysis of this large database of measurements permits to assess the statistical distributions of GHI and DNI in different time resolutions, as well as providing results of hourly, daily and annual average values and interannual variability of GHI and DNI which are of great interest for the assessment of solar energy projects. The distributions of the instantaneous values of k_t and k_b show a bimodal character, more pronounced at low values of the optical air mass, in accordance with the results obtained by other researchers for locations with the same climate, according to the Köppen-Geiger classification. A relevant result is that the frequency distributions of the instantaneous and 10-min values of k_t and k_b are very similar, while the distribution of hourly values shows significant differences with respect to them. These differences are also observed in the distribution of solar radiation values, mainly in the case of DNI, and should be considered when selecting the time step for the simulation of solar energy systems where transients have a significant effect.

The most frequent value of the clearness index for the lowest optical air-mass studied is approx. 0.76, indicating a high frequency of clear sky conditions with low atmospheric turbidity. Seville has an average of about 6 overcast days per month and more than 20 clear-sky days in June,

July and August. On average, the DNI is higher than 600 W m^{-2} for approx. 8 h during summer days. The rest of the months are located in the partially cloudy interval. The most variable results (February) and the lowest mean daily values (December) are found in the winter months.

The annual mean daily value of GHI is 4.98 kW h m^{-2} and 5.68 kW h m^{-2} for DNI. Both variables have a relatively low inter-annual variability (ranges of -3% to 5% for GHI and -6% to $+8\%$ for DNI) and a significant inter-monthly variability, which depends on the season. A comparison with the annual and monthly values from other sources reveals a good agreement with NASA SSE in GHI and clearness index series, while the DNI values provided by NASA SSE are 8.8% higher than those recorded by GTER.

References

- Bachour, D., Perez-Astudillo, D., 2014. Ground measurements of Global Horizontal Irradiation in Doha, Qatar. *Renew. Energy* 71, 32–36.
- Boland, J., Huang, J., Ridley, B., 2013. Decomposing global solar radiation into its direct and diffuse components. *Renew. Sustain. Energy Rev.* 28, 749–756.
- Bright, J.M., Smith, C.J., Taylor, P.G., Crook, R., 2015. Stochastic generation of synthetic minutely irradiance time series derived from mean hourly weather observation data. *Sol. Energy* 115, 229–242.
- Cebecauer, T., Suri, M., 2015. Typical Meteorological Year data: SolarGIS approach. *Energy Procedia* 69, 1958–1969.
- CIEMAT, 2012. ADRASE [WWW Document]. URL <<http://www.adrase.es/>> (accessed 7.1.14).
- da Assunção, H.F., Escobedo, J.F., Oliveira, A.P., 2003. Modelling frequency distributions of 5 minute-averaged solar radiation indexes using Beta probability functions. *Theor. Appl. Climatol.* 75, 213–224.
- Dugaria, S., Padovan, A., Sabatelli, V., Del Col, D., 2015. Assessment of estimation methods of DNI resource in solar concentrating systems. *Sol. Energy* 121, 103–115.
- Fernández-Peruchena, C.M., Bernardos, A., 2015. A comparison of one-minute probability density distributions of global horizontal irradiance conditioned to the optical air mass and hourly averages in different climate zones. *Sol. Energy* 112, 425–436.
- Gansler, R.a., Klein, S.a., Beckman, W.a., 1995. Investigation of minute solar radiation data. *Sol. Energy* 55, 21–27.
- Hirsch, T., Schenk, H., 2010. Dynamics of oil-based Parabolic Trough plants - a detailed transient simulation model. In: *SolarPaces Conference*, Perpignan, France.
- Hoyer-Klick, C., Hustig, F., Schwandt, M., Meyer, R., 2009. Characteristic meteorological years from ground and satellite data. In: *SolarPaces Conference*. Berlin, Germany.
- Instituto de Investigación y Formación Agraria y Pesquera, 2009. Estaciones Agroclimáticas [WWW Document]. <<http://www.juntadeandalucia.es/agriculturaypesca/ifapa/ria/servlet/FrontController>> (accessed 7.1.14).
- Iqbal, M., 1983. *An Introduction to Solar Radiation*. Academic Press.
- Islam, M.D., Alili, A.A., Kubo, I., Ohadi, M., 2010. Measurement of solar-energy (direct beam radiation) in Abu Dhabi, UAE. *Renew. Energy* 35, 515–519.
- Islam, M.D., Kubo, I., Ohadi, M., Alili, A.A., 2009. Measurement of solar energy radiation in Abu Dhabi, UAE. *Appl. Energy* 86, 511–515.
- Jurado, M., Caridad, J.M., Ruiz, V., 1995. Statistical distribution of the clearness index with. *Sol. Energy* 55, 469–473.
- Kudish, A.I., Ianez, A., 1996. Analysis of daily clearness index, global and beam radiation for Beer Sheva, Israel: partition according to day type and statistical analysis. *Energy Convers. Manage.* 37, 405–416.
- Kumar, A., Gomathinayagam, S., Giridhar, G., Mitra, I., Vashistha, R., Meyer, R., Schwandt, M., Chhatbar, K., 2013. Field experiences with the operation of solar radiation resource assessment stations in India. *Energy Procedia* 49, 2351–2361.
- Lam, J.C., Li, D.H., 1996. Study of solar radiation data for Hong Kong. *Fuel Energy Abstr.* 37, 201.
- Li, D.H.W., Lam, J.C., 2001. An analysis of climatic parameters and sky condition classification. *Build. Environ.* 36, 435–445.
- Liu, B.Y.H., Jordan, R.C., 1960. The interrelationship and characteristic distribution of direct, diffuse and total solar radiation. *Sol. Energy* 4, 1–19.
- Maafi, A., Harrouni, S., 2003. Preliminary results of the fractal classification of daily solar irradiances. *Sol. Energy* 75, 53–61.
- Meyer, R., 2010. Recommendations for bankable meteorological site assessments for solar thermal power plants. In: *SolarPaces Conference*. Perpignan, France.
- Moreno-Tejera, S., Ramírez-Santigosa, L., Silva-Pérez, M.a., 2015. A proposed methodology for quick assessment of timestamp and quality control results of solar radiation data. *Renew. Energy* 78, 531–537.
- Morf, H., 2013. A stochastic solar irradiance model adjusted on the Angström-Prescott regression. *Sol. Energy* 87, 1–21.
- NASA, 2012. Surface Meteorology and Solar Energy [WWW Document]. <<https://eosweb.larc.nasa.gov/sse/RETScreen/>> (accessed 7.1.14).
- Ngoko, B.O., Sugihara, H., Funaki, T., 2014. Synthetic generation of high temporal resolution solar radiation data using Markov models. *Sol. Energy* 103, 160–170.
- Ogunsola, O.T., Song, L., 2014. Restoration of long-term missing gaps in solar radiation. *Energy Build.* 82, 580–591.
- Olseth, J.A., Skartveit, A., 1989. Observed and modelled hourly luminous efficacies under arbitrary cloudiness. *Sol. Energy* 42, 221–233.
- Pagola, I., Gastón, M., Fernández-Peruchena, C., Moreno, S., Ramírez, L., 2010. New methodology of solar radiation evaluation using free access databases in specific locations. *Renew. Energy* 35, 2792–2798.
- Peel, M.C., Finlayson, B.L., McMahon, T.A., 2006. Updated world map of the Köppen-Geiger climate classification. *Meteorol. Zeitschrift* 15, 259–263.
- Powell, G.L., 1984. The clear sky model. *ASHRAE J.* 26, 27–29.
- Sancho, J., Riesco, J., Jiménez, C., 2012. Atlas de Radiación Solar en España utilizando datos del SAF de Clima de EUMETSAT.
- Schwandt, M., Chhatbar, K., Meyer, R., Fross, K., Mitra, I., Vashistha, R., Giridhar, G., Gomathinayagam, S., Kumar, A., 2014. Development and test of gap filling procedures for solar radiation data of the Indian SRRA measurement network. *Energy Procedia* 57, 1100–1109.
- Skartveit, A., Olseth, J.A., 1992. The probability density and autocorrelation of short-term global and beam irradiance. *Sol. Energy* 49, 477–487.
- Stoffel, T., Renné, D., Myers, D., Wilcox, S., 2010. CONCENTRATING SOLAR POWER best practices handbook for the collection and use of solar resource data. Contract 146, NO: NREL/TP-550-47465.
- Suehrcke, H., McCormick, P.G., 1988. The frequency distribution of instantaneous insolation values. *Sol. Energy* 40, 413–422.
- Tapakis, R., Charalambides, a.G., 2014. Enhanced values of global irradiance due to the presence of clouds in Eastern Mediterranean. *Renew. Energy* 62, 459–467.
- Tovar, J., Olmo, F.J., Alados-Arboledas, L., 1998. One-minute global irradiance probability density distributions conditioned to the optical air mass. *Sol. Energy* 62, 387–393.
- Tovar, J., Olmo, F.J., Battles, F.J., 1999. One minute kb and kd probability density distributions conditioned to the optical air mass. *Sol. Energy* 65, 297–304.
- Tovar-Pescador, J., 2008. Modelling the Statistical Properties of Solar Radiation and Proposal of Technique Based on Boltzmann Statistics. In: *Modeling Solar Radiation at the Earth Surface*. Springer, pp. 55–91.
- Varo, M., Pedrós, G., Martínez-Jiménez, P., Aguilera, M.J., 2006. Global solar irradiance in Cordoba: Clearness index distributions conditioned to the optical air mass. *Renew. Energy* 31, 1321–1332.
- Woyte, A., Belmans, R., Nijs, J., 2007. Fluctuations in instantaneous clearness index: analysis and statistics. *Sol. Energy* 81, 195–206.
- Zawilska, E., Brooks, M.J., 2011. An assessment of the solar resource for Durban, South Africa. *Renew. Energy* 36, 3433–3438.

ANEXO C



Classification of days according to DNI profiles using clustering techniques



S. Moreno-Tejera^{a,*}, M.A. Silva-Pérez^a, L. Ramírez-Santigosa^b, I. Lillo-Bravo^a

^a Department of Energy Engineering, University of Seville, Spain

^b Solar Radiation Unit, Renewable Energy Division, CIEMAT, Spain

ARTICLE INFO

Article history:

Received 13 December 2016

Received in revised form 6 February 2017

Accepted 20 February 2017

Keywords:

State of sky

DNI

Transmittance index

Clustering techniques

ABSTRACT

A methodology to classify days as a function of the state of the sky for Concentrated Solar Power (CSP) plant operation is proposed. For this purpose, three indexes are used to characterize the energy, variability and time distribution of the DNI and to define the type of days by means of clustering techniques. Two sets of indexes are tested and compared. The energy of days is represented by the transmittance index, k_b . Two indexes are used to characterize the variability of the DNI: persistence index of the instantaneous k_b values (POP_D) and Variability Index (VI). Equivalent indexes have been previously used to classify the types of days using Global Horizontal Irradiation (GHI). A novel index to define the time distribution of the DNI daily energy is introduced. Clustering analysis is applied to thirteen years (2000–2012) of 10-min DNI measurements recorded in Seville (37.40°N, 6.01°W) by the Group of Thermodynamics and Renewable Energy (GTER) at the University of Seville. The k-medoids algorithm is used for cluster analysis. Through the use of well-known internal validity indexes and with the help of the L-method, the optimum number of clusters (types of days) is found to be 10. The results are compared with the assessment carried out by five experts on a reference set composed of DNI daily curves from two years (2010 and 2011). This comparison reveals a better coincidence when the clustering is performed using VI.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

The solar resource assessment for a feasibility analysis of a Concentrated Solar Power (CSP) plant is usually based on annual, monthly and, in some cases, daily Direct Normal Irradiation (DNI) data (Cebecauer and Suri, 2015; Pagola et al., 2010; Wilcox and Marion, 2008). The aim of these studies is to characterize statistically the solar resource of the location in these temporal scales. This information is used at different stages of a CSP plant project, from the feasibility analysis to the daily or hourly electricity generation forecast. However, the production of a CSP plant is not directly proportional to the energy radiation received; it depends on when (time of the day) and how (with a low or high variability) is received. A day with highly variable DNI because of the frequent passage of clouds can have the same energy as another with very stable DNI during a fraction of the day and yet the electricity generation can be very different because the operation of the plant will be very different, too. Therefore, a methodology to classify the days

that takes into account the intraday variability and temporal distribution, besides the daily energy, could help to better characterize the solar resource from a CSP plant standpoint. Since temporal distribution and variability affect the performance of CSP plants, this characterization can be useful from the design stage to the definition of the operation strategy of the plant regarding different aspects like the thermal storage dispatch or the commercialization of the electricity generated. For example, Kraas et al. (2013) have shown that the DNI forecasting error increases with variability, resulting in penalties derived from incorrect power generation forecasts in the electricity market.

The classification of the days according to the solar radiation can be approached from a qualitative or quantitative standpoint (Kang and Tam, 2013). From a quantitative standpoint, the quantity of daily energy recorded in a location is evaluated. This approach follows the TMY3 methodology (Wilcox and Marion, 2008) and the majority of methodologies used to characterize the long-term behaviour of the solar radiation for feasibility analyses of CSP plants (Cebecauer and Suri, 2015; Fernández Peruchena et al., 2016; Pagola et al., 2010). The qualitative classification has been studied for decades by means of the clearness index. Liu and Jordan introduced this concept in 1960 (Liu and Jordan, 1960). The clearness index is the ratio between the global and

* Corresponding author at: Department of Energy Engineering, Escuela Técnica Superior de Ingeniería, University of Seville, Camino de los Descubrimientos s/n, 41092 Seville, Spain.

E-mail address: smoreno2@etsi.us.es (S. Moreno-Tejera).

Nomenclature

k_b	transmittance index	I_{CS}	direct normal irradiance in clear sky conditions
k_t	clearness index	I_{sc}	solar constant
H_{bn}^d	daily direct normal irradiation	e_0	the Earth-Sun distance correction factor
H_{CS}^d	daily direct normal irradiation in clear sky conditions	α	solar elevation angle
I_{bn}	direct normal irradiance		

the extra-terrestrial radiation on a horizontal surface. This index has been widely used in different intervals of temporal integration (Perez et al., 2011; Skartveit and Olseth, 1992; Tovar-Pescador, 2008). The daily clearness was initially used as a classification criterion to categorize the days into three types: clear days, cloudy days, and overcast days (Lam and Li, 1996; Li and Lam, 2001). This classification proved insufficient for solar energy applications. The operation and the production of two days with similar k_t can be very different.

With the development of solar technologies, new classification methods introduce the combination of the daily clearness index with other parameters or parameters based on the clearness index in other temporal scales. The latter parameters provide more qualitative information to classify the days by considering the state of the sky. Muselli et al. (2000) characterize classes of typical days according to some parameters elaborated from hourly clearness index profiles. One parameter estimates the perturbation state of the hourly k_t values by considering the integral of the squared second derivative of the hourly clearness profile. Harrouni et al. (2005) uses the fractal dimension of Global Horizontal Irradiance (GHI) curves along with the daily clearness index as classification indexes. Rahim et al. (2004) and Baharuddin and Rahim (2010) classify the days using the cloud fraction and the sunshine duration applied to building design, finding similar results in both cases.

With the inclusion of new parameters, some studies propose increasing the number of the types of days or the states of the sky. Soubdhan et al. (2009) identify 4 different types, analysing instantaneous clearness indexes by means of the combination of Dirichlet distributions that model the clearness index distribution curves. Calbó et al. (2001) try to classify the days into 5 and 9 groups, testing different parameters derived from GHI and Diffuse Horizontal Irradiance (DHI) measurements. The best results are found for the 5 types of days compared with human observations. Umemiya and Kanou (2008) integrate the 15 classes proposed by the CIE standard in 7 classes, using the atmospheric turbidity, the clearness index, the brightness and the normalized global illuminance as classification parameters. Kang and Tam (2013) assess the probability of persistence of the instantaneous clearness index and, with the help of the daily clearness index, suggest classifying the days into 10 types. In the literature, there is neither a clear consensus about the number of types of the states of the sky, nor the most proper parameters to identify them.

Most of the parameters used to classify the days are calculated from the global component. Calbó et al. (2001) suggest that the diffuse component is a better discriminant factor than the global radiation. The DNI seems to be a better discriminant factor, too, given its higher sensitivity to cloud passages and the absence of geometrical effects. Perez et al. (2011) use the transmittance index, calculated from hourly DNI values, as an input to characterize the intrahourly variability of the GHI. M Gastón-Romeo et al. (2011) classify the days into 4 classes through clustering techniques that use the morphology of the DNI curves. This methodology is oriented to establish a classification of days useful for CSP. The use of the Mathematical Morphology-based technique maintains the dynamic pattern of the DNI curves,

which is what helps characterize the variability of the DNI, but the temporal distribution of the energy is lost.

In this work, we propose a methodology to classify the state of the sky from DNI measurements, which can be useful for CSP plant analysis and operation. For this purpose, two combinations of parameters that represent the energy, the variability and the temporal distribution of the DNI are tested. Conventional clustering techniques are used to define the number and identify the features of the types of days. The parameters are calculated from 13 years (2000–2012) of DNI measurements recorded in Seville (Spain). Both options are validated by comparison with the visual classification carried out by 5 experts from two years of DNI curves.

2. Experimental data

The database used for this work corresponds to thirteen years (2000–2012) of DNI measurements recorded in Seville (37.40° N, 6.01° W) by the Group of Thermodynamics and Renewable Energies (GTER) of the University of Seville. These measurements were recorded every 5 s with an Eppley Normal Incidence Pyrheliometer (NIP) mounted on a Kipp & Zonen 2AP 2-axis tracker. This pyrheliometer is considered a secondary standard device according to ISO specifications. The hourly and daily uncertainty specified by the manufacturer is $\pm 1\%$, assuming proper maintenance of the instruments. The GTER station follows a maintenance and calibration procedure according to the recommendations from the instrument manufacturers. The database passed a quality control check (Moreno-Tejera et al., 2015), and it was filled by applying gap filling techniques (Moreno-Tejera et al., 2016). The DNI database was averaged every 10 min for this implementation.

3. Methods

3.1. Classification indexes

From a CSP plant point of view, there are three useful features of the DNI to take into account for grouping the days: the daily energy, the distribution over time of this energy and the variability (high frequency changes) of the instantaneous values caused by the passage of clouds throughout the course of the day. The aim of this work is to identify a good combination of indexes that help classify the days as a function of these features.

3.1.1. Daily energy characterization index

The first methods proposed in the literature to classify the state of the sky were based on the daily clearness index, k_t (Iqbal, 1983; Kudish and Ianetz, 1996). The k_t index is defined (Liu and Jordan, 1960) as the ratio of the global radiation and the solar extra-terrestrial radiation on the horizontal surface in instantaneous values or for an interval of time. If we are interested only in the direct component, we can find in the literature a similar index that relates the daily DNI with the value of this component in the absence of atmosphere (Tovar-Pescador, 2008) or in an ideal clear sky condition. This is the definition of the transmittance index (k_b) introduced by Skartveit and Olseth (1992):

$$k_b = \frac{H_{bn}^d}{H_{cs}^d}. \quad (1)$$

where H_{bn}^d is the daily DNI and H_{cs}^d is the daily irradiation in clear sky conditions. Both daily irradiation values are obtained as the time integral of 10-min irradiance values.

To calculate the DNI in clear sky conditions we use a concept introduced by Gómez Camacho and Blanco Muriel (1990), the enveloping clear day. The authors propose the empirical adjustment of a clear sky model with the instances of lowest atmospheric attenuation for every day of the year. There are several simple models to calculate the direct normal irradiance under clear sky conditions. Most of these models make use of a short number of parameters that are usually estimated from empirical measurements under clear sky situations. Gómez Camacho and Blanco Muriel (1990) use the ASHRAE clear sky model (Powell, 1984):

$$I_{cs} = I_{sc} \cdot \varepsilon_0 \cdot A e^{\frac{B}{\sin^2 \alpha}} \quad (2)$$

where I_{cs} is the direct normal irradiance in clear sky conditions, I_{sc} is the solar constant, ε_0 is the Earth-Sun distance correction factor, α is the solar elevation angle, and A and B are the parameters to be empirically adjusted.

In this work, we also use the ASHRAE clear sky model but adjusting the parameters with the instances of lowest atmospheric attenuation for every solar elevation. For this purpose, we re-write Eq. (2) as follows:

$$\frac{I_{cs}}{\varepsilon_0} = I_{sc} \cdot A \cdot e^{\frac{B}{\sin^2 \alpha}} \quad (3)$$

If the irradiance of clear sky is divided by ε_0 , the parameters A and B are independent of the day of the year and the equation only depends on the solar elevation angle. The parameters A and B are estimated by adjusting the maximum observed values of I_{bn}/ε_0 for every value of solar elevation higher than 5° from the database. Thus, we only need to estimate a single pair of values for A and B that are valid for all the days of the year ($A = 0.8187$ and $B = 0.0840$). Fig. 1 shows the fitting and the resulting curves for all the days of the year.

3.1.2. Variability characterization index

The characterization of the DNI variability is usually conducted by means of complex methodologies that analyse the DNI curves. Muselli et al. (2000) classify the days in terms of hourly k_t profiles,

and Gastón-Romeo et al. (2011) use Cumulative Distribution Functions (CDF) that define the morphology of the curves. Kraas et al. (2013) compute the number of DNI direction changes during the day to describe the variability or volatility of hourly DNI. The use of this simple concept at 10-min scale could result, however, in computing an excessive number of direction changes associated to small DNI gradients.

In this work, we aim to characterize the variability with an index to facilitate its use on other applications, such as the methodologies for characterization of the solar resource's long-term behavior. We have considered two simple indexes:

3.1.2.1. Variability index. Stein et al. (2012) propose the “Variability Index” (VI) to identify periods and locations with high GHI variability. This index compares the length of the GHI curves with the length of the corresponding clear sky GHI curves. VI provides information about the variability of the day, but it can reach similar values on very clear and overcast days. The authors use the daily clearness index k_t together with VI to identify four types of days: clear, overcast, mixed and highly variable all day. Huang et al. (2014) define the daily variability index (DVI) in a similar way.

In this work, we apply the same definition to characterize the variability of the DNI. Our VI is defined as the ratio between the length of the DNI curve and the length of the maximum enveloping clear day curve calculated in Section 3.1:

$$VI = \frac{\sum_{k=2}^n \sqrt{(I_{bnk} - I_{bnk-1})^2 + \Delta t^2}}{\sum_{k=2}^n \sqrt{(I_{csk} - I_{csk-1})^2 + \Delta t^2}} \quad (4)$$

where I_{cs} is the enveloping clear sky direct normal irradiance every 10 min, Δt refers to an interval of 10 min, and n is the number of 10-min intervals of the considered day.

3.1.2.2. Persistence index. Kang and Tam (2013) present a simple but useful method called the K-POP method to classify the days from GHI measurements. With this method, they characterize quantitatively and qualitatively the state of the sky from the daily k_t values and the daily probability of k_t instantaneous persistence index (POP_D). POP_D index provides information about the stability of the state of the sky in such a way that a high POP_D value corresponds to a very stable day (clear or overcast) and vice versa.

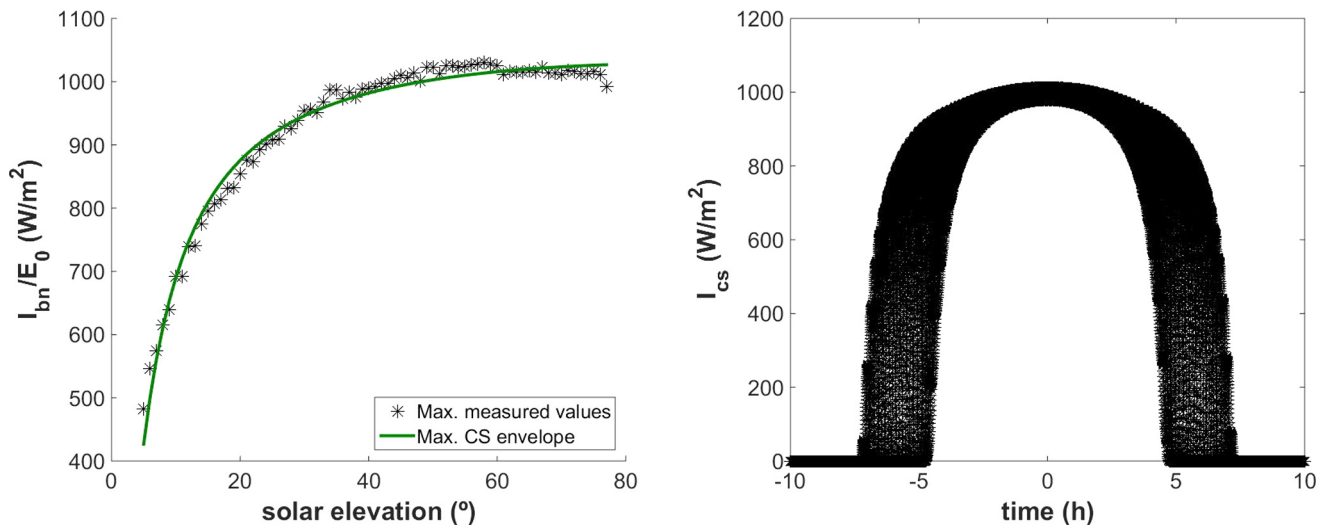


Fig. 1. On the left, maximum values of I_{bn}/ε_0 with respect to the solar elevation and the ASHRAE model fit. On the right, fitted curves of the ASHRAE model for the 365 days of the year.

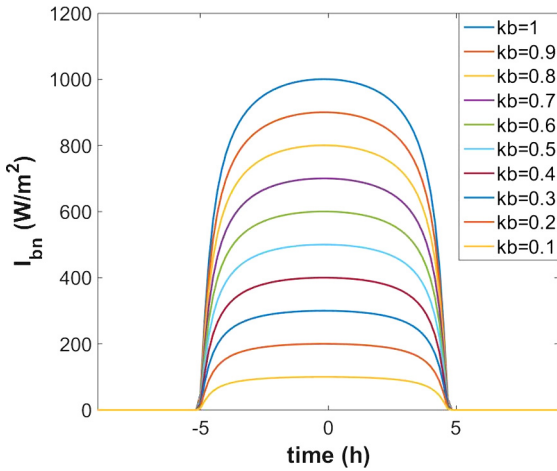


Fig. 2. Theoretical curves of DNI with instantaneous k_b from 0.1 up to 1, which corresponds to a clear sky day.

The POP_D index can be re-defined using the instantaneous k_b index instead of k_t . Following a parallel approach, the persistence of the k_b index is calculated in this work as the ratio between the

sums of instances in which the k_b index differs less than 0.1 from its predecessor and the total number of instances. In Fig. 2, a clear sky curve is plotted along with DNI curves that hold a constant k_b index from 0.1 to 0.9. These curves help us to visualize the transients that are computed with the POP_D index.

3.1.3. Index to characterize the temporal DNI distribution

The indexes usually used in the literature to characterize the state of the sky are focused mainly on the daily energy and sometimes on the variability of the DNI but not on its temporal distribution. For CSP plants, the temporal distribution of this energy is relevant. In the first row of Fig. 3, we show two examples of days, with a daily k_b value usually related to a cloudy day when a CSP plant could generate electricity continuously for a few hours. In this work we introduce the *morning fraction* index, F_m , to help identify this type of day. The *morning fraction* is defined as the ratio between the accumulated DNI in the first half of the day and the accumulated DNI for the whole day. With this index we can determine if the daily radiation is concentrated in the first or the second part of the day:

$$F_m = \frac{H_{bn,m}^d}{H_{bn}^d} \tag{5}$$

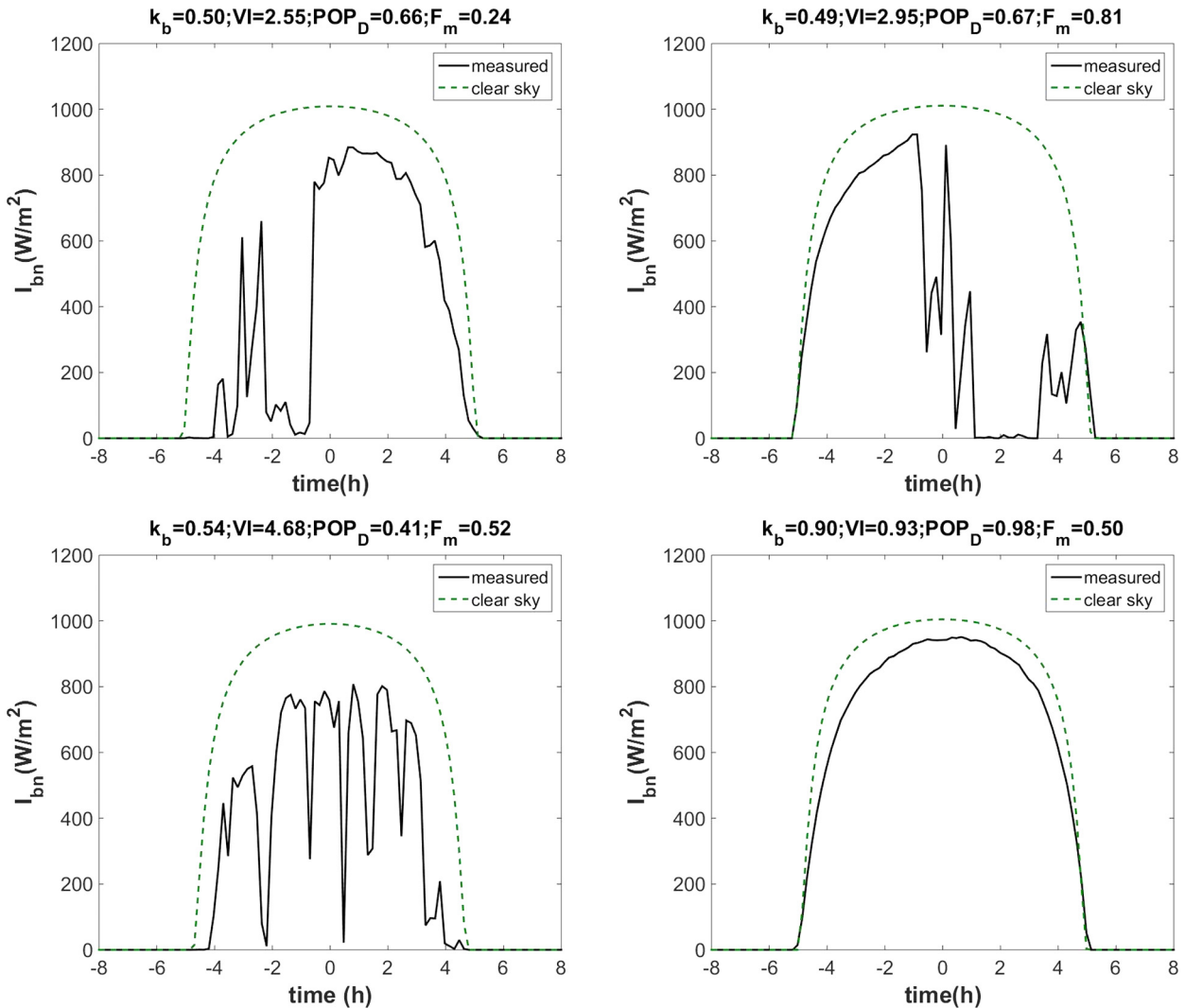


Fig. 3. Representation of the DNI values recorded in Seville and the enveloping clear sky DNI for four days in 2010 and their corresponding values of k_b , VI , POP_D and F_m .

Table 1
Spearman correlation coefficient from the analysed indexes.

	k_b	VI	POP _D	F_m
k_b	1	0.08	0.20	-0.02
VI	-	1	-0.93	0.1
POP _D	-	-	1	-0.13
F_m	-	-	-	1

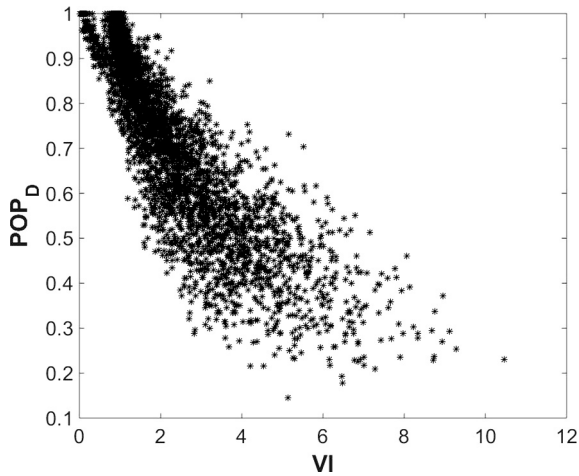


Fig. 4. Representation of POP_D values versus VI values for all the data from the period 2000–12.

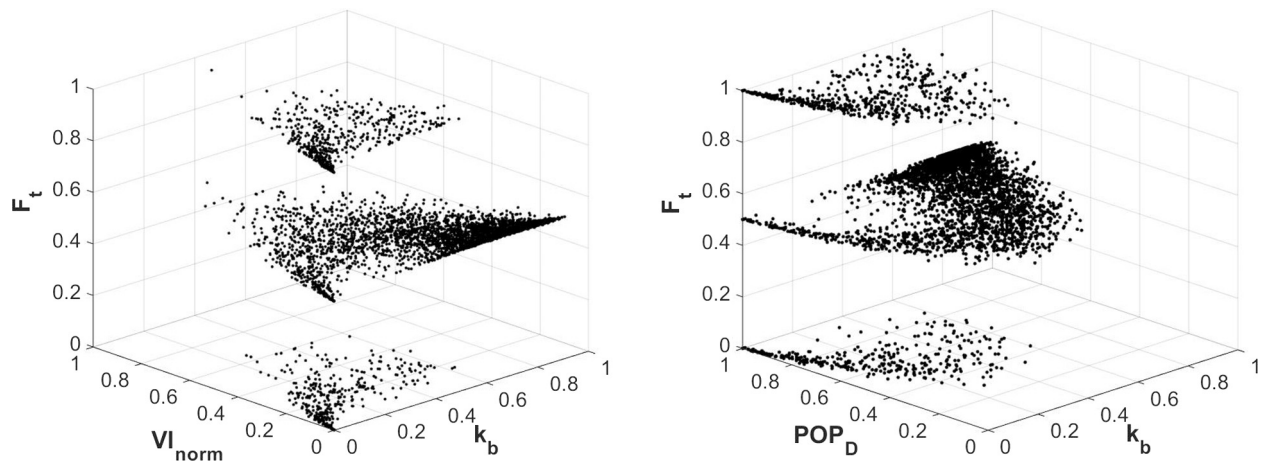


Fig. 5. Representation of the two groups of indexes calculated from 13 years of DNI measurements that are used to classify the days as a function of the state of the sky: k_b , VI and F_t (on the left) and k_b , POP_D and F_t (on the right).

Table 2

Day types used as reference by the experts to carry out the classification of the days from the years 2010 and 2011. Day types are defined according to the features taken into account to apply the clustering algorithm.

Type	Sky condition	Energy	Variability	Energy time distribution
1	Totally clear (no clouds)	High	Low	Balanced
2	Mostly clear (high clouds)	Medium	Low	Morning
3	Mostly clear (high clouds)	Medium	Low	Balanced
4	Mostly clear (high clouds)	Medium	Low	Afternoon
5	Partly covered (mid/high clouds)	Medium	High	Morning
6	Partly covered (mid/high clouds)	Medium	High	Balanced
7	Partly covered (mid/high clouds)	Medium	High	Afternoon
8	Mostly covered (mid clouds)	Low	Medium	Morning
9	Mostly covered (mid clouds)	Low	Medium	Balanced
10	Mostly covered (mid clouds)	Low	Medium	Afternoon
11	Totally covered (low clouds)	Null	Low	Balanced

where H_{bn-m}^d is the DNI recorded from the sunshine to the solar noon and H_{bn}^d is the daily DNI. This index can be defined, obviously, for the second part of the day, too, but without adding more information.

3.2. Analysis of the indexes

To avoid overlapping in the features that every index provides, we calculate the Spearman correlation coefficient for the proposed indexes:

As shown in Table 1, there is a high anticorrelation between VI and POP_D. This is an expected result, since both characterize the transients of DNI caused by the clouds. However, they express opposite concepts: the variability and the persistence. In Fig. 4, both indexes are represented, and they show a grade of dispersion that, despite its high anticorrelation, justifies the study of these parameters separately. Additionally, we observe that the other indexes present a low Spearman correlation coefficient, so they can contribute to the definition of the type of days without duplication.

In Fig. 3, the representations of the DNI and the maximum clear sky DNI from four different types of days are shown. The values of the four indexes proposed for this work are included. In the first row, two days with similar energies and variabilities but different time distributions are represented. Both days present similar k_b , VI and POP_D values and very different F_m values. For the day presented on the right, F_m value is very high, since the majority of the energy is received in the first half of the day. The day presented on the left is almost symmetrical, and this effect is reflected in its F_m value. For both days of the second row, the energy is spread uniformly

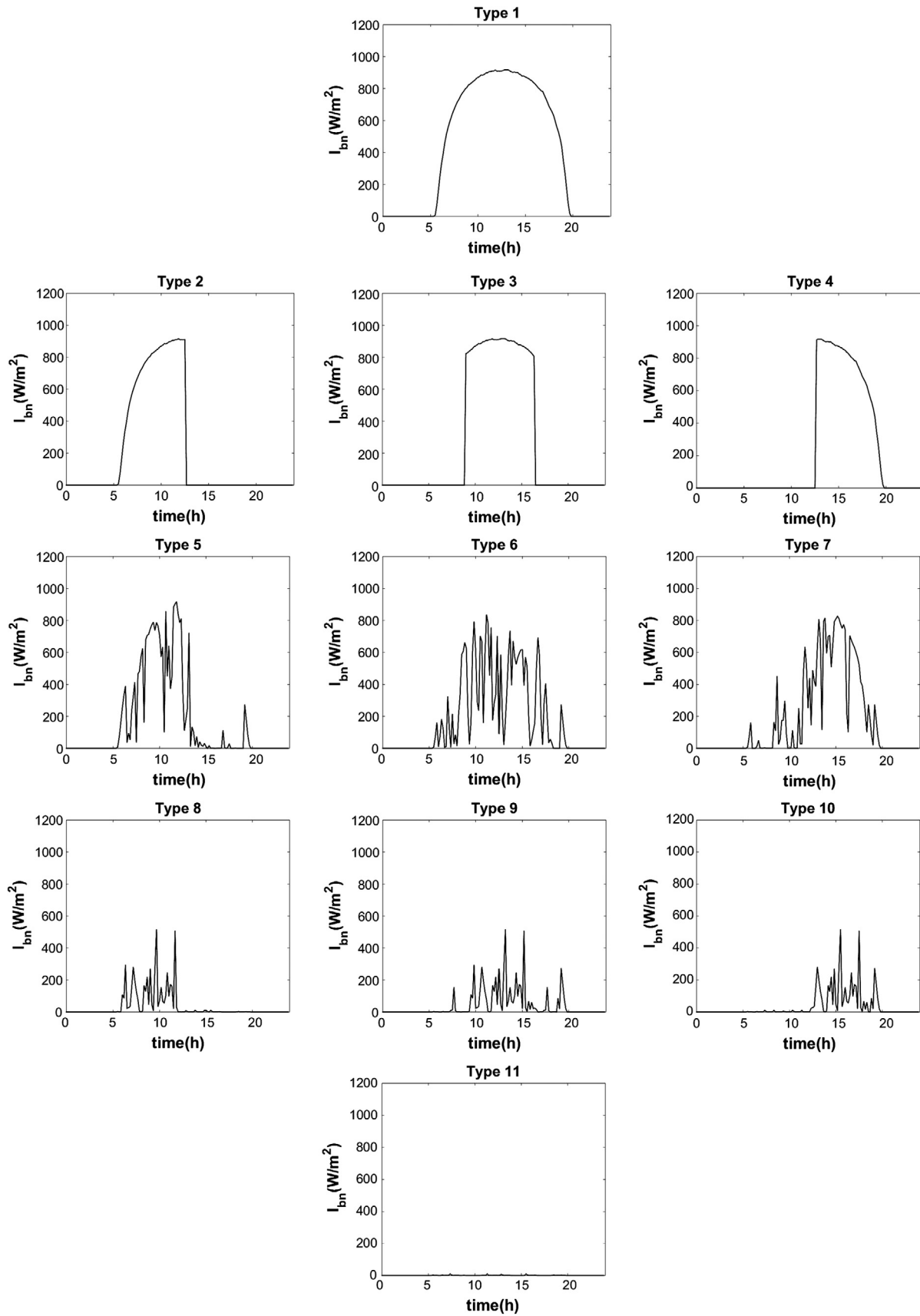


Fig. 6. Graphical examples of the types of days defined in Table 2. This illustration was used by the experts as a reference to accomplish the visual classification.

in time, with F_m approximately 0.5. However, the day on the left shows a higher variability and a lower persistence of its DNI, which is depicted in the higher VI values and the lower POP_D value. It is worth highlighting that VI is the only index that does not range between 0 and 1, showing more sensitivity to transients.

3.3. Classification procedure

The four indexes presented in Section 4 are calculated for the thirteen years of DNI data from the Seville station. Two combinations of indexes are analysed by clustering techniques. Both combinations include indexes to represent the energy, the variability and the temporal distribution of the DNI. The first one is composed of k_b , VI and F_m . The second one is composed of k_b , POP_D and F_m . In this section, the clustering algorithm and the methodology applied to choose the proper number of clusters is described. Finally, the clusters are validated by comparison with a classification developed by 5 experts in the field. This comparison is also explained in this section.

3.3.1. Classification technique

The aim of a clustering algorithm based on a partitioning method is to classify a set of data into k groups or clusters: the data of the same cluster should share the same features, whereas data from different clusters should be dissimilar. k -Means is the most popular among the partitioning methods. This algorithm provides good results for a certain type of data, and its best advantage is its simplicity. However, the k -means algorithm is sensible to “outliers” because it searches the centres of the clusters using the average distance between the data located in it. This method can distort the data distribution in groups (Han and Kamber, 2001). To avoid this drawback, the k -medoid algorithm uses a reference point, the “medoid,” which is the most centrally located object or point in the cluster (Han and Kamber, 2001). The k -medoid algorithm tries to minimize the sum of the distance metric between every object of a group and its medoid. It allows a good characterization of all clusters that are not too elongated and makes it possible to isolate outliers in most situations (Kaufman and Rousseeuw, 1990).

From a recent comparative analysis between fuzzy c -means, k -means and k -medoids that aimed to capture the patterns of the

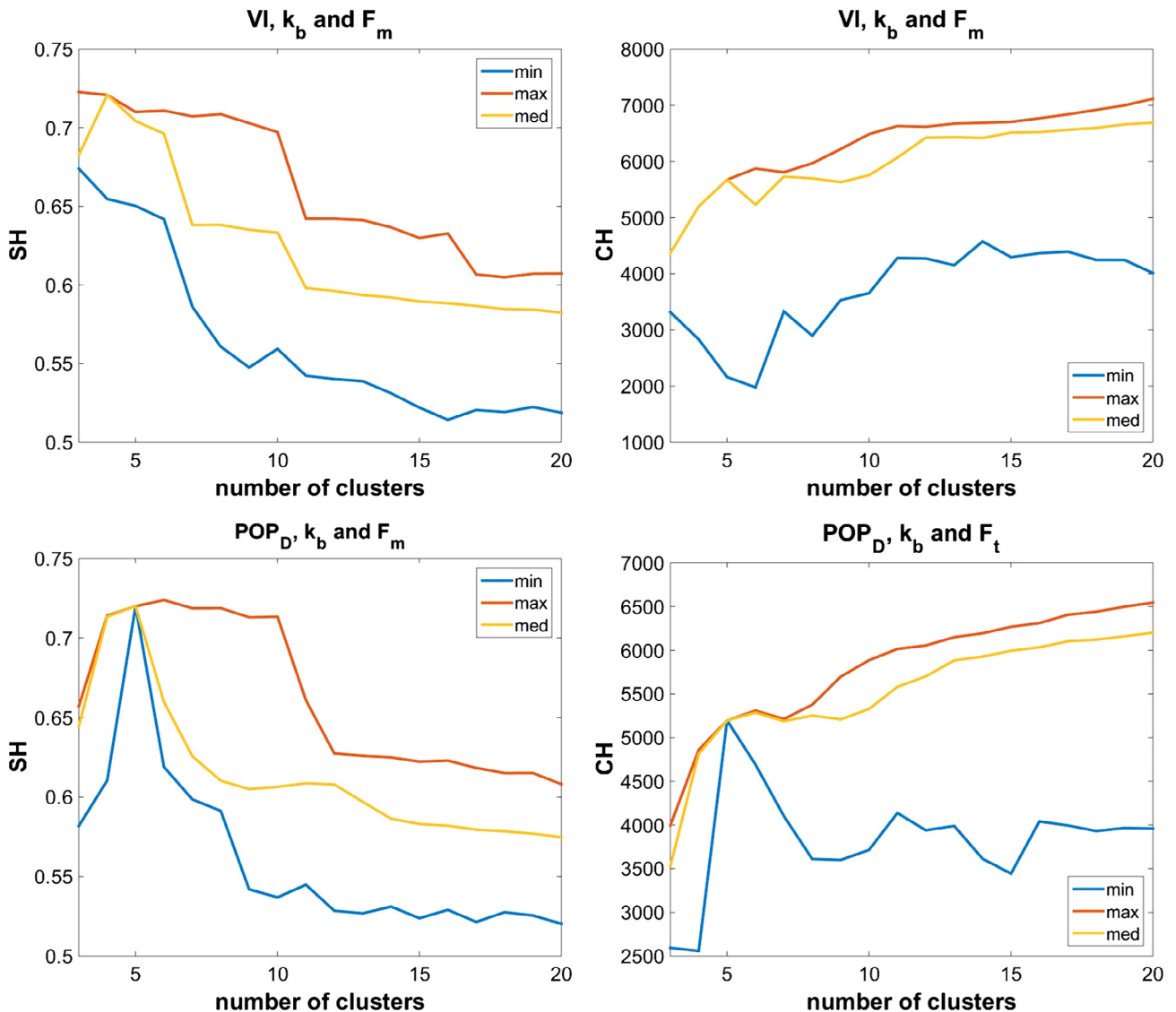


Fig. 7. Representation of the maximum, minimum and median values of the SH and CH values from the 1000 iterations with respect to the number of clusters for the two groups of indexes studied.

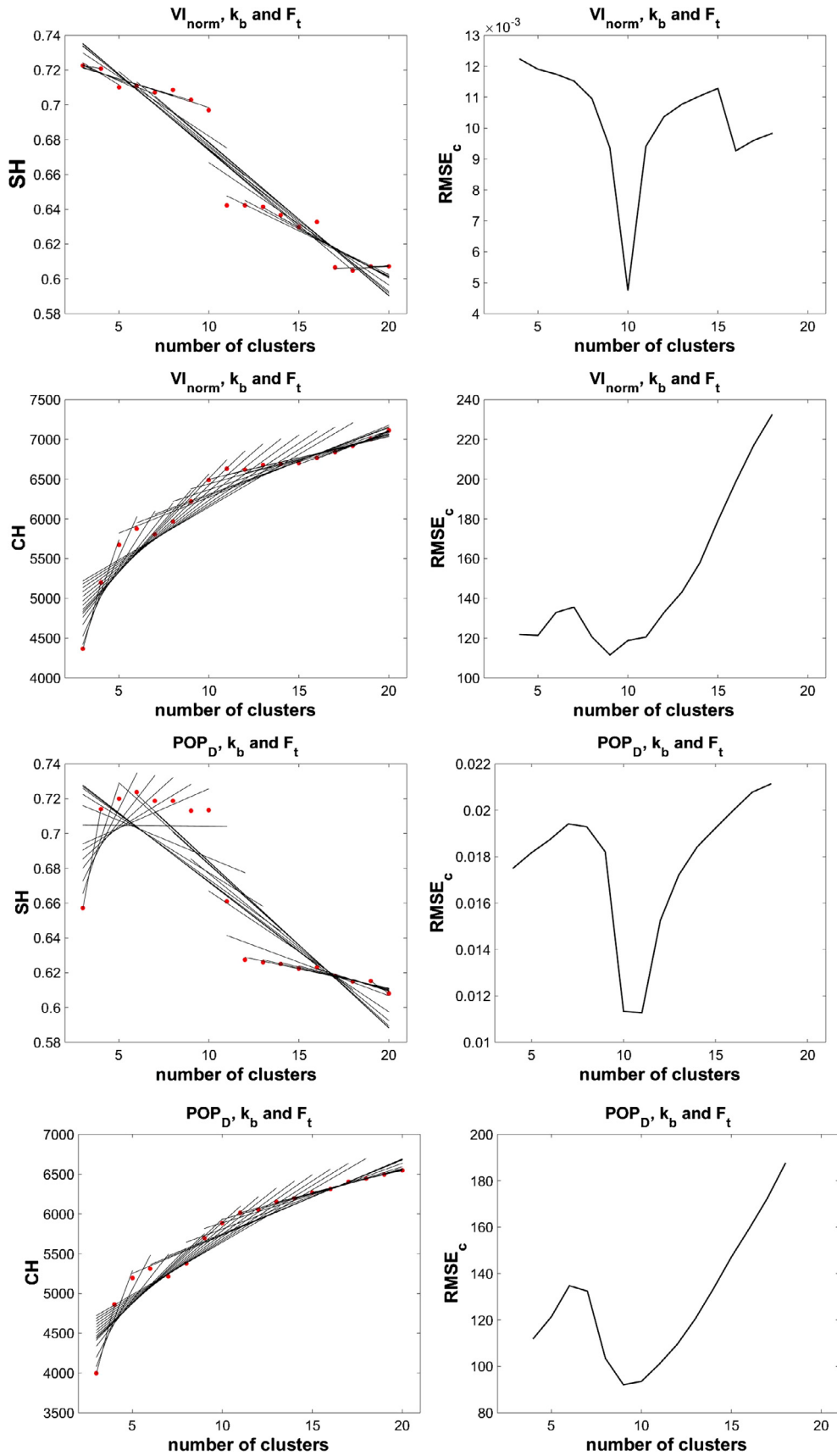


Fig. 8. On the left: representations of all possible pairs of fit lines to the maximum values of CH and SH for the two groups of indexes for finding the “knee point,” or the best number of clusters for applying the L-method. On the right: the corresponding RMSE_c values are graphed with respect to the number of clusters.

wake effect of the wind, it can be concluded that *k-medoid* is the most appropriate algorithm for this type of data (Al-Shammari et al., 2016).

The *k-medoid* algorithm applied in this work, *Partitioning Around Medoids (PAM)* (Han and Kamber, 2001), is more robust than the *k-means* algorithm, but it shares two weaknesses: the start centre points are selected randomly, and the user has to specify the number of clusters to apply the algorithm. Therefore, the effectiveness of the result depends on the distance between the initial and the final centres of the clusters. The algorithm is applied several times for each number of clusters to overcome these obstacles. Moreover, the algorithm is applied by sweeping a range of numbers of clusters and analysing the results through several validation indexes that are described in Section 3.3.3.

3.3.2. Pre-processing data

Three different features of the DNI are assessed with three indexes. All the indexes, except for *VI*, range between 0 and 1, so *VI* is normalized dividing all values by the maximum *VI* obtained ($VI_{max} = 10.92$) to avoid assigning a greater weight to this index in the clustering process. Furthermore, F_m index is segmented to optimize and facilitate the clustering of the data. The resulting new index is called F_t and defined by Eq. (6).

$$\begin{aligned}
 F_t &= 1 \text{ if } F_m \leq 0.3 \\
 F_t &= 0.5 \text{ if } 0.3 < F_m < 0.7 \\
 F_t &= 0 \text{ if } F_m \geq 0.7
 \end{aligned} \tag{6}$$

Thus, we help the algorithm to identify the days whose accumulated energy is received mainly during the afternoon (0), during the morning (1) or during both periods (0.5). In Fig. 5, the two groups of indexes to be clustered are shown. Hereinafter, the group of indexes composed of *VI*, k_b and F_t is called the *VI* group, and the group of indexes composed of POP_D , k_b and F_t is called the *POP_D* group.

3.3.3. Determination of the number of clusters

The techniques for the determination of the appropriate number of clusters are classified into two types, internal and external validation methods. The external validation methods are based on previous knowledge about the partition of the data. The internal methods consider the intrinsic information of the data's geometric structure (Zagouras et al., 2013). These methods consider several indexes to evaluate the similitude among the members of the same

cluster and the dissimilitude among members of different clusters, thus helping to optimize the number of clusters. Arbelaiz et al. (2013) developed an extensive comparative study, providing orientations to select the proper index as a function of the data type. Two of the indexes that achieve successful results in this study are used in this work: the Calinski-Harabasz index (CH) (Calinski and Harabasz, 1974) and the Silhouette index (SH) (Rousseeuw, 1987). CH is one of the most popular and robust internal validity indexes (Zagouras et al., 2013). It is based on the positions of the cluster centres in the dataset, as these are used to define the proximity of all the data points. Larger values of the CH index are related to cluster partitions composed of well-separated clusters, where the cluster centres lie distantly in space and sustain compact intra-cluster distances. The SH index qualifies every point by considering its position with respect to the other points of the cluster to which it belongs and its position with respect to points from other clusters. SH could be evaluated for a single point, for the points of a cluster, or for all the points, taking values from -1 to +1. A SH value close to 1 for a point indicates that this point belongs to a compact cluster and that it is located at a maximal distance from its nearest cluster; thus, the point is properly classified. A negative value indicates a probability that the point is misclassified. In a study conducted by Arbelaiz, the SH index was shown to yield the best results in most cases (Arbelaiz et al., 2013).

The identification of the optimal number of clusters by means of these indexes is not an easy task. Sometimes, the results show monotonic increases or decreases, providing a better clustering quality as the number of clusters increases, or relative optimums. In this work, a graphical method, called the L-method, is used to help select the optimum number of clusters (Salvador and Chan, 2003; Zagouras et al., 2014, 2013). This method attempts to identify the inflection point of the curve that represents the number of clusters vs. the evaluation index, and it detects the pair of straight lines that best fit either side of the curve. If the validation index evolution does not show an optimum number of clusters, the point of maximum curvature of the curve (the *knee point*) is considered a reasonable number of clusters. Following a repeated sequence, all possible pairs of best-fit lines for each point *c* are evaluated. L_c and R_c are, respectively, the left and the right sequences of data points partitioned at *c*, the number of clusters to evaluate. Each line must contain at least two points. The Root Mean Squared Error (RMSE) of each pair of best-fit lines is calculated by:

$$RMSE_c = \frac{c-1}{b-1} RMSE_{L_c} + \frac{b-c}{b-1} RMSE_{R_c} \tag{7}$$

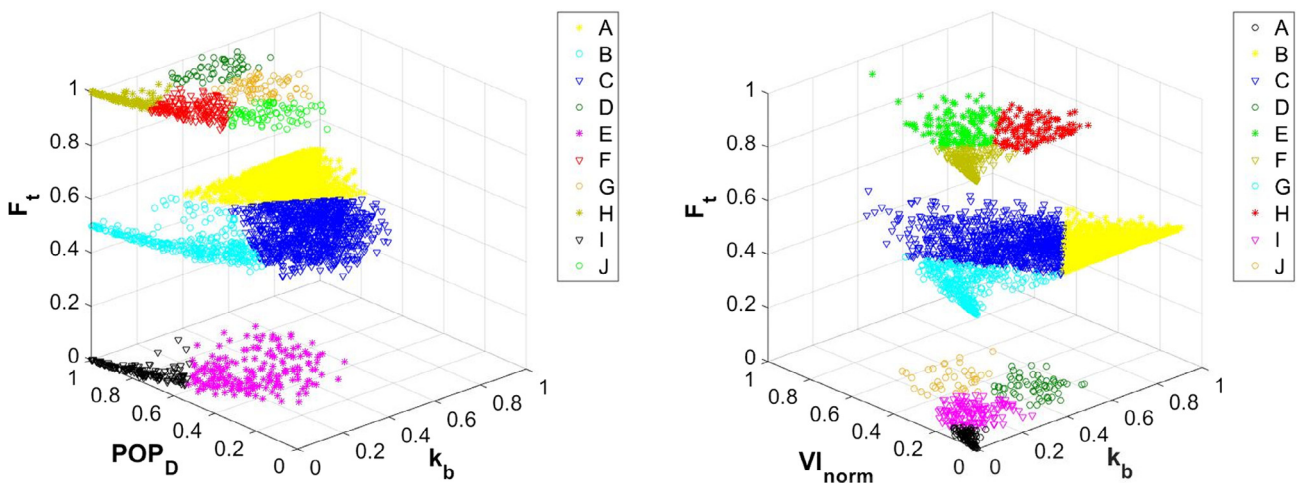


Fig. 9. Representation of *POP_D* (left) and *VI_{norm}* (right) values versus k_b and F_t from the whole database. The markers and labels identify the clusters.

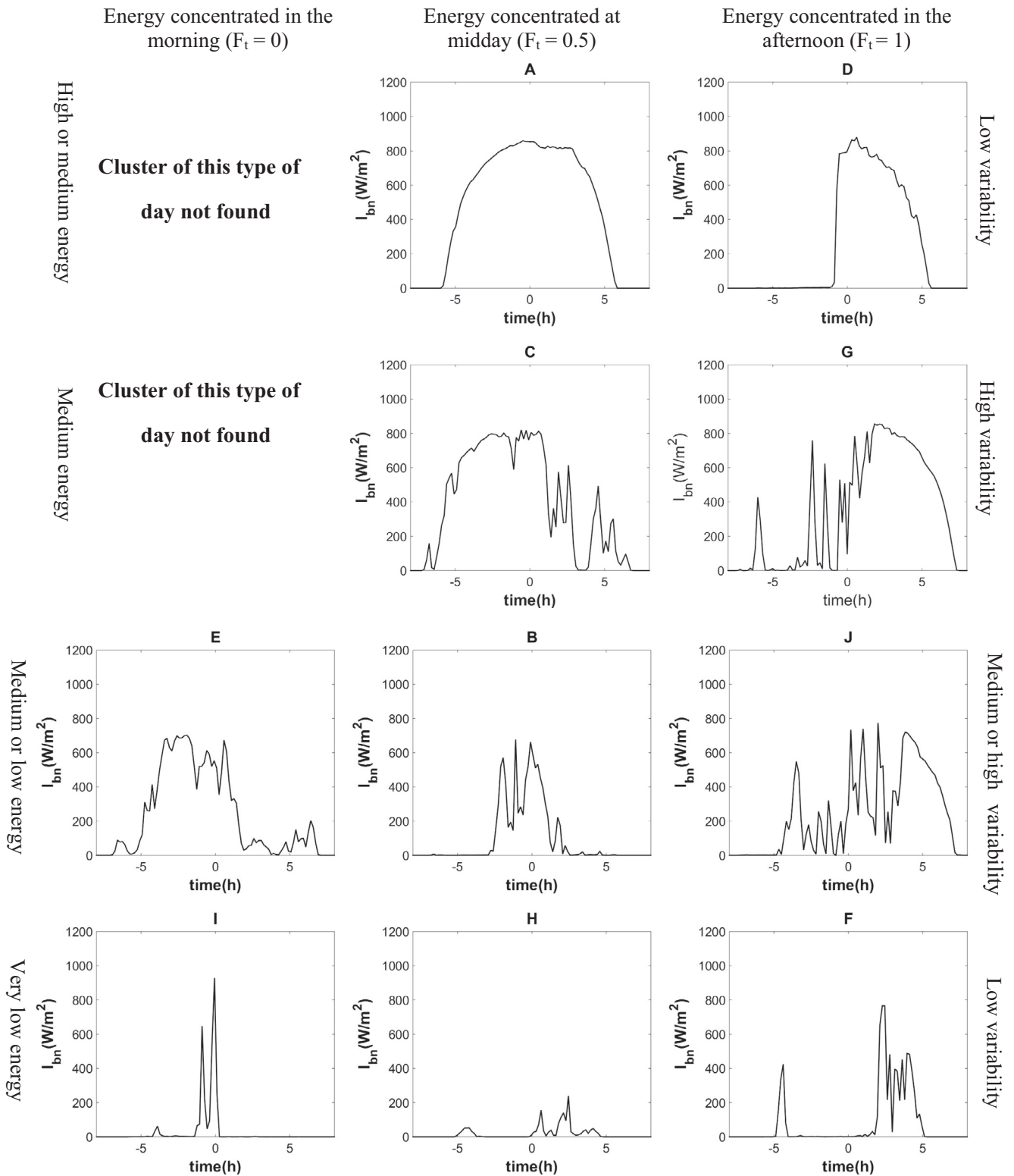


Fig. 10. Representation of the days corresponding to the cluster medoids in the study of the POP_D group. The labels of the medoids match the cluster labels of Fig. 9. The location of the plots follows a similar order to that used in Fig. 6. The plots located in the left column have energy distributions concentrated in the morning. The plots in the central column have a balanced energy distribution between morning and afternoon, and the plots in the right column have distributions concentrated in the afternoon. The distribution of rows is also related with similar characteristics of energy and intraday variability of DNI.

where c is the vertical projection over the x axis of the last point belonging to L_c ; $RMSE_{L_c}$ and $RMSE_{R_c}$ are, respectively, the RMSE values of the left and right fit lines; and b is the maximum number of clusters analysed. The optimum number of clusters is the one with the minimum $RMSE_c$ value.

4. Validation and selection of the appropriate combination of indexes

To select the most appropriate combination of indexes and validate the results, the clusters obtained with the proposed

methodology for the years 2010 and 2011 are compared with a visual classification of 5 experts in the field. These experts were provided with a reference table (Table 2) that defines 11 types of days. We consider that these days cover all the options that could be interesting to identify from a CSP standpoint and

according to the indexes taken into account. In Fig. 6, an example of each type of day is illustrated. In the following sections, the results of the visual classification are compared with the clusters obtained through the methodology proposed in this study.

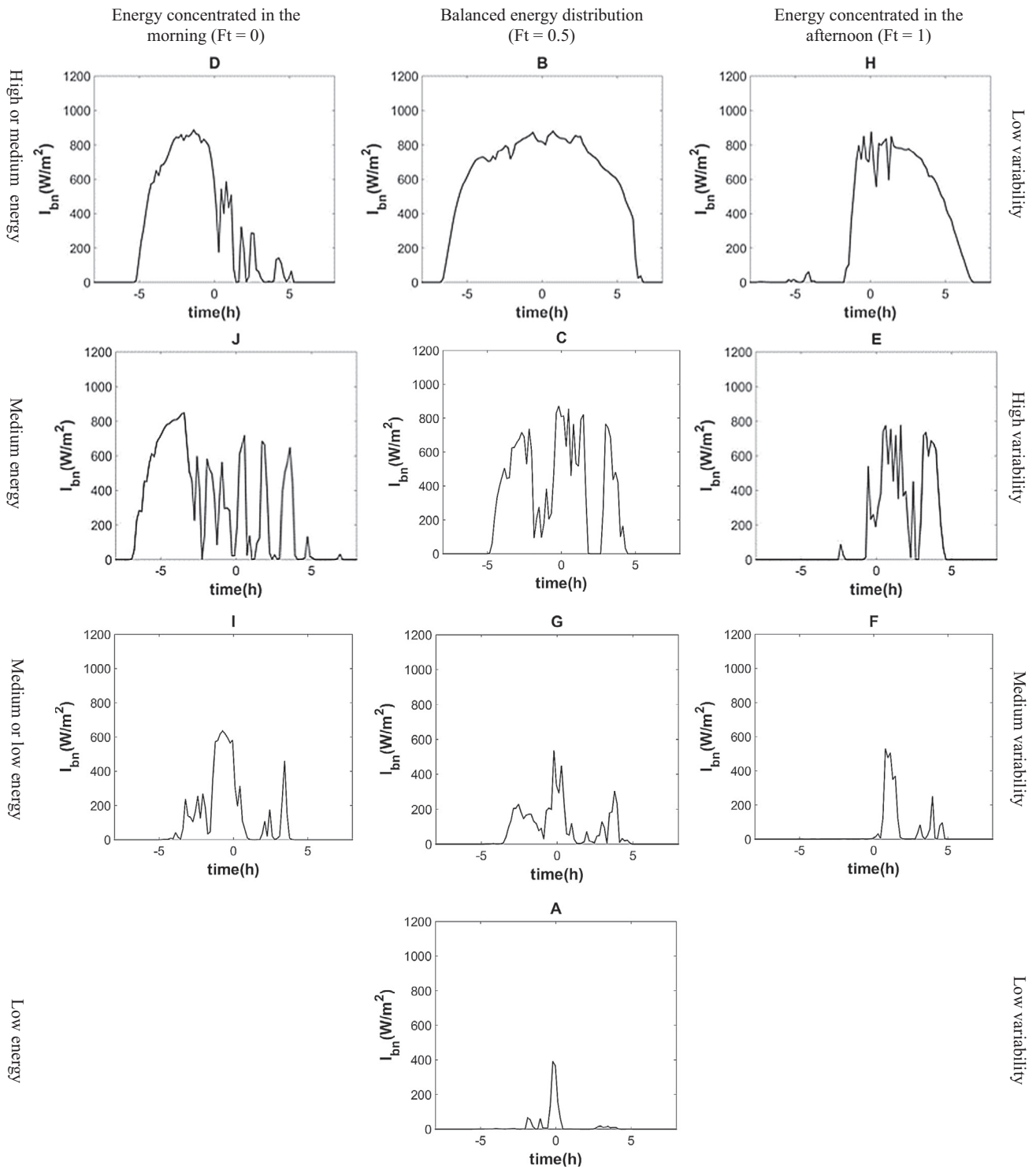


Fig. 11. Representation of the days corresponding to the cluster medoid in the study of the VI group. The labels on the medoids match the cluster labels of Fig. 9. The location of the plots follows a similar order to that used in Fig. 6. The plots located in the left column have an energy distribution concentrated in the morning. The plots located in the centre column have a balanced distribution, and those in the right column have a distribution concentrated in the afternoon. The distribution of rows is also related with similar characteristics of energy and intraday variability of DNI.

5. Results and discussion

5.1. Classification of states of the skies

As mentioned in the previous sections, the results of a clustering algorithm depend on the initial selection of the medoids. To ensure that the optimum solution is reached for each number of clusters, the algorithm is run 1000 times for each case and evaluated by using the SH and CH indexes. We have found in the literature classifications that consider a number of types of days, ranging from 3 to 10 (Calbó et al., 2001; Gastón-Romeo et al., 2011; Harrouni et al., 2005; Kang and Tam, 2013; Lam and Li, 1996; Li and Lam, 2001; Muselli et al., 2000; Rahim et al., 2004; Soubdhan et al., 2009; Umemiya and Kanou, 2008), although none of the methodologies consider the time distribution of the solar radiation. From our point of view and as shown in Section 5.4, 11 types of days seem to be enough to identify all the cases of interest from the indexes used. Even so, a range from 3 to 20 clusters is explored with the aim of analysing the tendencies of the validity indexes and helping to identify the “knee point.” The maximum, median and minimum values of the SH and CH indexes from the 1000 cases for each number of clusters and group of indexes are shown in Fig. 7. The maximum values of both indexes drive to the optimum number of clusters in terms of similarity and compactness. The median and minimum values provide information about the ability to find that optimum in a number of iterations or, in other words, the uncertainty in finding the optimum case if the algorithm is run only one time.

As seen from Fig. 7, both groups of indexes show monotonically increasing maximum CH values and an almost monotonic decrease in the case of the SH index for the VI group. These results make it difficult to localize the optimum number of clusters. It is also difficult to find the absolute optimum in the case of SH for the POP_D group, since a range of numbers of clusters reach similar values, close to the maximum, in several cases. Thus, the L-method is applied to help select the most reasonable optimum in each case. The plots of all possible pairs of fit lines and the evolution of the RMSE_c for each group of indexes and for the maximum values of CH and SH are represented with respect to each possible “knee point” candidate in Fig. 8.

In the case of CH index, the minimum is found in 9 clusters for the VI group and in 11 clusters in the POP_D group, but the value for 10 clusters is very similar. In the case of SH index for both the VI and POP_D groups, the minimum values are clearly reached for 10 clusters. Thus, we consider 10 clusters as the most reasonable option.

The values of the indexes for the whole database are plotted in Fig. 9. The points that belong to the same clusters are represented by the same colour. It is observed that the number of clusters for every value of F_t is different in each group of indexes analysed. In the POP_D group, 5 clusters are obtained for F_t equal to 1, and an unbalanced number of 2 clusters for F_t equal to 0 is found. In the case of the VI group, the number of groups is more balanced, with 3, 3 and 4 clusters for F_t values equal to 1, 0.5 and 0, respectively.

In Fig. 10, the DNI curves of the medoids of the cluster from the POP_D group are illustrated. If these curves are compared with the

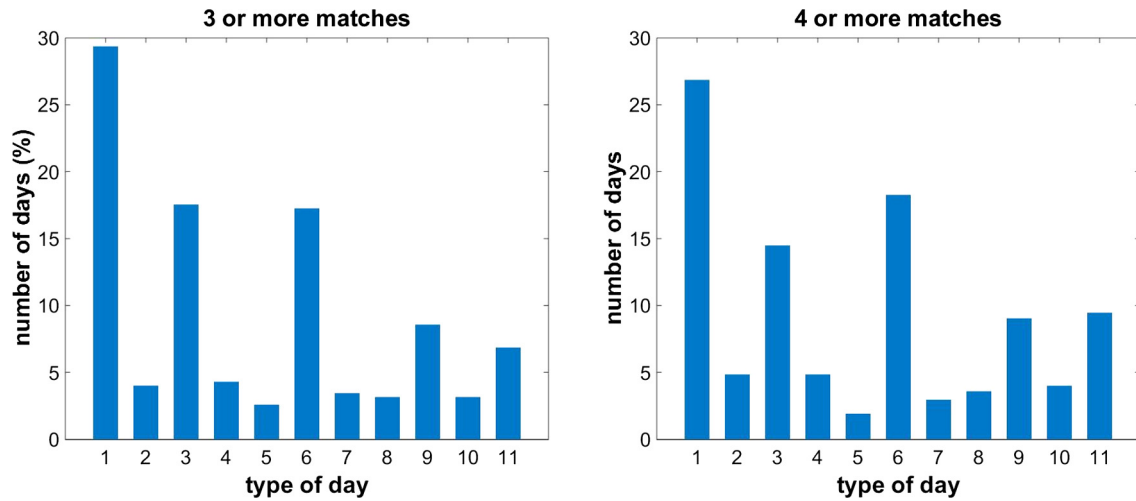


Fig. 12. Percentage of days that have been classified as the same type by 3 or more experts (left) and by 4 or more experts (right).

Table 3 Confusion matrix of the results of the clustering process versus the experts' classification.

Clustering class	Reference class										Total (%)
	1	2	3	4	5	6	7	8	9	10	
1	321	11	9	1	10	0	0	0	0	0	91.2
2	0	12	0	4	0	0	0	0	0	0	75.0
3	0	0	17	0	0	3	0	0	0	0	85.0
4	0	0	0	3	1	0	1	0	0	0	60.0
5	8	5	3	4	100	3	0	2	0	0	80.0
6	0	0	1	0	0	12	0	2	1	0	75.0
7	0	0	0	6	0	0	12	4	0	0	54.5
8	0	0	0	0	10	2	1	44	1	17	58.7
9	0	0	0	0	0	4	0	4	20	13	48.8
10	0	0	0	0	0	0	8	4	0	18	60.0
Total (%)	97.6	42.9	56.7	16.7	82.6	50.0	0.0	73.3	90.9	37.5	79.6

curves used as reference by the experts to visually classify the days (see Fig. 6) only some similarities are found. However, in the case of the VI group the similarities are quite significant (see Fig. 11). With the exception of two types of days (type 2 and type 11), the features of the reference days can be considered the same (see Table 2): 3 days with medium or high energy, low variability and different time distributions; 3 days with medium energy, high

variability and different time distributions; and 3 days with low energy, medium variability and different time distributions. It should be noted that in Table 2, 11 types of days are defined, while the optimum number of clusters identified by applying the methodology is 10. This means that the case of high energy with low variability, case “1”, is merged with case “2” in cluster B, as shown in Fig. 11. The other exception is the medoid that should

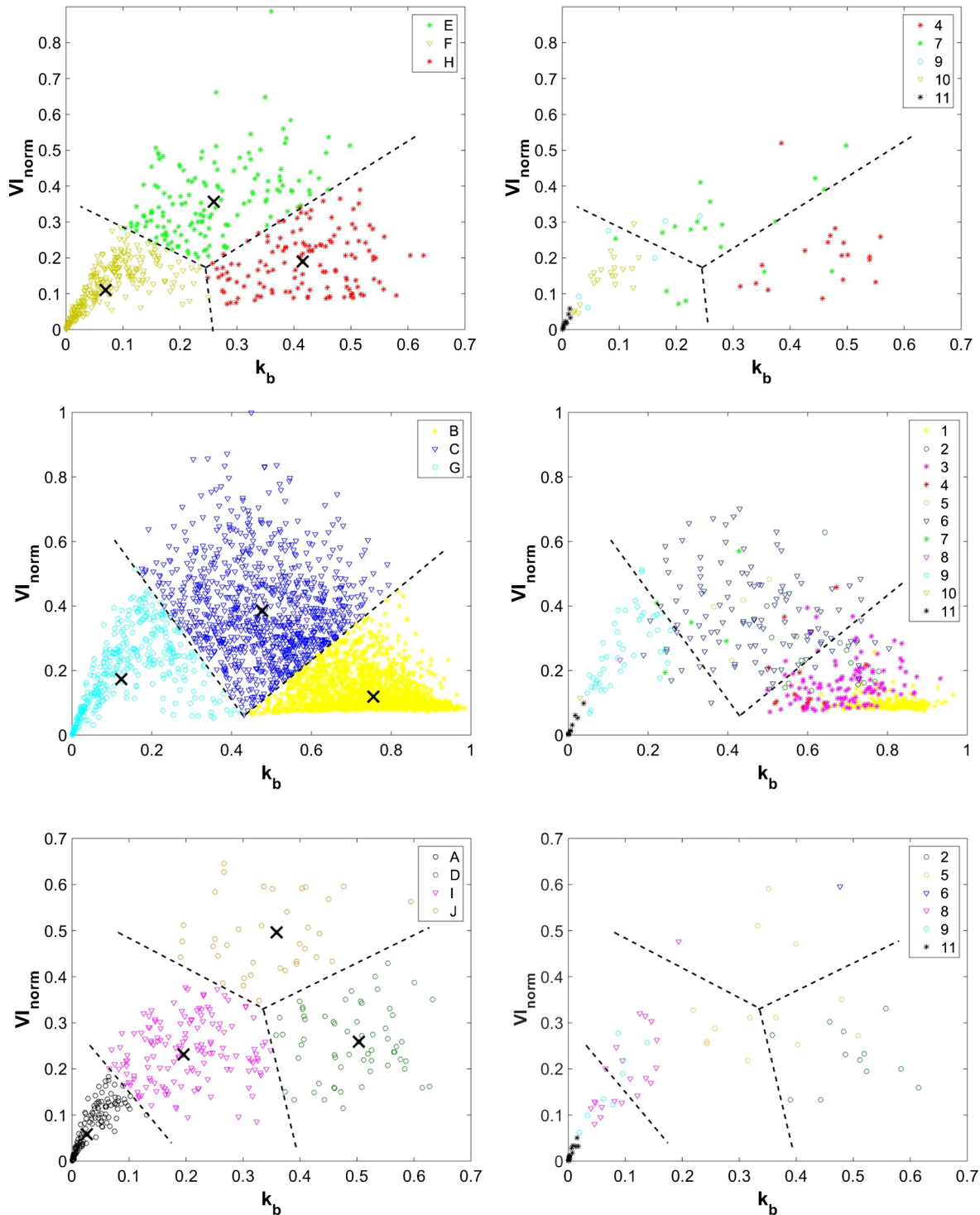


Fig. 13. On the left side, the representations of VI_{norm} vs. k_b for the whole database and for F_t equal to 1 (first plot), 0.5 (second plot) and 0 (last plot) are illustrated. The dotted lines define the limits of each cluster and the medoids of each are marked with “X” symbols. On the right side, representations of VI_{norm} vs. k_b for F_t equal to 1 (first plot), 0.5 (second plot) and 0 (last plot) from the reference set elaborated with the experts’ classifications are represented. The lines defining the limits between clusters have been added.

represent type 11 (the medoid of cluster A from Fig. 11). This medoid has very low energy and variability, but its F_t is 0, which means that the energy is mainly from the morning, when the DNI is so low that the distribution of the energy is not significant. In fact, a visual decision about the time distribution of these types of days is not easy and from a CSP standpoint is not relevant. This satisfactory result reveals VI index as a better way to define the variability of the days following the methodology proposed.

5.2. Validation of results

Once the VI group was selected as the most appropriate set of indexes, the resulting classification of the days from the period 2010–11 through the clustering technique was compared with the classification realized by the experts (see Section 4). For this purpose, a reference set was built from the classification developed by the experts. A total of 477 days from the selected period have been classified as the same types of days by 4 or more experts and a total of 702 days by 3 or more. In Fig. 12, the percentages of days of these two options are graphed for every type of day. Both options show similar percentages of days for the 11 types of days. Thus, bearing in mind that the number of days available for the validation increases by 47% when using the matches of 3 or more experts, this option is considered as the *reference classification*. It is worth to highlight that the percentage of days that belong to the types of days with energy concentrated mainly in the morning or in the afternoon are significantly lower than the types of days with balanced energy time distribution.

The reference classification is compared with the classification of the same days resulting from the methodology proposed in this study. To this end, the medoids of the clusters are matched to the types of days proposed to the experts to search for similitudes between the plots displayed in Figs. 6 and 11 ($B = 1$ and 3 ; $D = 2$; $H = 4$; $J = 5$; $C = 6$; $E = 7$; $I = 8$; $G = 9$; $F = 10$; $A = 11$). Because type 3 has been merged with type 1, the clusters have been relabelled, starting from type 4 with 1 number less. The confusion matrix obtained after this modification is shown in Table 3. Each column of the matrix represents a type or class of day from the reference classification while each row represents a cluster type. The number in each cell is the number of days that have been classified by the experts in the class corresponding to the column and pertain to the cluster corresponding to the row. The cells located in the diagonal indicate the days classified in groups with similar features by both methods (numbers in bold). The last column and the last row show the total percentage of coincidences. The lower right corner cell displays the total of coincidences between both classifications. A total of 79.6% of the matches was found. This is a very good result considering that the numerical limit for defining a day with the energy mainly concentrated on the first or second mid-day was not indicated to the experts. Moreover, the clusters with the highest number of disagreements are in the types of days with low energy (types 7, 8 and 9; the distinction between these types of days with very low energy is not relevant from the CSP standpoint) or in a type of day with a very low population (type 4).

In the left column of Fig. 13, the points of the whole database are graphed in 3 plots that correspond to the three values of F_t . The points of each cluster are marked with the same symbol and colour, and the medoids are marked with an "X" symbol. The dotted lines limit the area of every cluster. The classification of the reference set of days is plotted in the right column, and lines are added to define the clusters. Almost all points from types 1 and 3 are located in the cluster labelled with the letter B. The points with a balanced time distribution have a good concordance with the cluster limits, and most disagreements are found in the definition of the clusters with low energies (or low k_b values) and unbalanced time distributions (or F_t equal to 0 or 1). As noted above, when the

energy is very low, the time distribution is not relevant. From the standpoint of a CSP plant, types A, G, I and F could be merged into only one cluster.

6. Conclusions

A new methodology is proposed to classify days according to the state of the sky based on three features of the DNI curves: variability, energy and temporal distribution of this energy. These indexes are used to group the days by means of known clustering techniques that define the number of groups and the limits between them. Two different groups of indexes are evaluated with internal validation methods. The optimum number of clusters found using both groups of indexes is 10 clusters or types of days, although the VI group shows a slightly lower uncertainty in finding the best clusters than the POP_D group. A comparison of the visual classifications of a reference set of days from five experts has been used for the external validation of the clustering outcomes. A higher degree of coincidence is achieved with the VI group. The concordance in days with balanced time distributions is very good, and most of the disagreements are found in the definition of the clusters with low energies (or low k_b values) and unbalanced time distributions (F_t equal to 0 or 1). Since the time distribution of days with very low DNI is not relevant for CSP plants, the discrimination of these groups is not useful. A unique cluster that includes all low energy days could be defined by merging the groups A, G, I and F. Then, only seven types of days or clusters (instead of 10) would be sufficient. A new index related to the production of a CSP plant could help to find a solution by means of partitioning techniques. For future works, the characterization of the production of a CSP plant for every cluster and the application of the methodology in other locations are proposed. Moreover, it could be interesting to analyse the possibility to include the statistical characterization of the types of days according to these new features in the methodologies of generation of representative solar resource sets.

References

- Al-Shammari, E.T., Shamshirband, S., Petković, D., Zalnezhad, E., Yee, P.L., Taher, R.S., Cojbašić, Ž., 2016. Comparative study of clustering methods for wake effect analysis in wind farm. *Energy* 95, 573–579.
- Arbelaitz, O., Gurrutxaga, I., Muguerza, J., Pérez, J.M., Perona, I., 2013. An extensive comparative study of cluster validity indices. *Pattern Recognit.* 46, 243–256.
- Baharuddin, S.S.Y., Lau, R., 2010. Daylight availability in Hong Kong: classification into three sky conditions. *Archit. Sci. Rev.* 53, 396–407.
- Calbó, J., González, J.A., Pegès, D., 2001. A method for sky-condition classification from ground-based solar radiation measurements. *J. Appl. Meteorol.* 40, 2193–2199.
- Calinski, T., Harabasz, J., 1974. A dendrite method for cluster analysis. *Commun. Stat. Methods* 3, 1–27. <http://dx.doi.org/10.1080/03610917408548446>.
- Cebecauer, T., Suri, M., 2015. Typical meteorological year data: SolarGIS approach. *Energy Procedia* 69, 1958–1969.
- Fernández Peruchena, C.M., Ramírez, L., Silva-Pérez, M.A., Lara, V., Bermejo, D., Gastón, M., Moreno-Tejera, S., Pulgar, J., Liria, J., Macías, S., Gonzalez, R., Bernardos, A., Castillo, N., Bolinaga, B., Valenzuela, R.X., Zarzalejo, L.F., 2016. A statistical characterization of the long-term solar resource: towards risk assessment for solar power projects. *Sol. Energy* 123, 29–39.
- Gastón-Romeo, M., Leon, T., Mallor, F., Ramírez-Santigosa, L., 2011. A Morphological Clustering Method for daily solar radiation curves. *Sol. Energy* 85, 1824–1836.
- Gómez Camacho, C., Blanco Muriel, M., 1990. Estimación de la atmósfera estándar de radiación solar a partir del concepto de día claro envolvente. Aplicación a la Plataforma Solar de Almería. *Era Sol.* 40, 11–14.
- Han, J., Kamber, M., 2001. *Data Mining. Concepts and Techniques*. Chapter 8. Cluster analysis, 1st Edition. ed. Morgan Kaufmann Publishers.
- Harrouni, S., Guessoum, A., Maafi, A., 2005. Classification of daily solar irradiation by fractional analysis of 10-min-means of solar irradiance. *Theor. Appl. Climatol.* 80, 27–36.
- Huang, J., Troccoli, A., Coppin, P., 2014. An analytical comparison of four approaches to modelling the daily variability of solar irradiance using meteorological records. *Renew. Energy* 72, 195–202.
- Iqbal, M., 1983. *An Introduction to Solar Radiation*. Chapter 11. Solar Radiation Incident on Tilted Planes on the Earth's Surface.

- Kang, B.O., Tam, K.S., 2013. A new characterization and classification method for daily sky conditions based on ground-based solar irradiance measurement data. *Sol. Energy* 94, 102–118.
- Kaufman, L., Rousseeuw, P.J., 1990. *Finding Groups in Data: An Introduction to Cluster Analysis*. N.J.
- Kraas, B., Schroedter-Homscheidt, M., Madlener, R., 2013. Economic merits of a state-of-the-art concentrating solar power forecasting system for participation in the Spanish electricity market. *Sol. Energy* 93, 244–255.
- Kudish, A.I., Ianetz, A., 1996. Analysis of daily clearness index, global and beam radiation for Beer Sheva, Israel: partition according to day type and statistical analysis. *Energy Convers. Manage.* 37, 405–416.
- Lam, J.C., Li, D.H., 1996. Study of solar radiation data for Hong Kong. *Fuel Energy Abstr.* 37, 201.
- Li, D.H.W., Lam, J.C., 2001. An analysis of climatic parameters and sky condition classification. *Build. Environ.* 36, 435–445.
- Liu, B.Y.H., Jordan, R.C., 1960. The interrelationship and characteristic distribution of direct, diffuse and total solar radiation. *Sol. Energy* 4, 1–19.
- Moreno-Tejera, S., Ramírez-Santigosa, L., Silva-Pérez, M.A., 2015. A proposed methodology for quick assessment of timestamp and quality control results of solar radiation data. *Renew. Energy* 78, 531–537.
- Moreno-Tejera, S., Silva-Pérez, M.A., Lillo-Bravo, I., Ramírez-Santigosa, L., 2016. Solar resource assessment in Seville, Spain. Statistical characterisation of solar radiation at different time resolutions. *Sol. Energy* 132, 430–441.
- Muselli, M., Poggi, P., Notton, G., Louche, A., 2000. Classification of typical meteorological days from global irradiation records and comparison between two Mediterranean coastal sites in Corsica Island. *Energy Convers. Manage.* 41, 1043–1063.
- Pagola, I., Gastón, M., Fernández-Peruchena, C., Moreno, S., Ramírez, L., 2010. New methodology of solar radiation evaluation using free access databases in specific locations. *Renew. Energy* 35, 2792–2798.
- Perez, R., Kivalov, S., Schlemmer, J., Hemker, K., Hoff, T., 2011. Parameterization of site-specific short-term irradiance variability. *Sol. Energy* 85, 1343–1353.
- Powell, G.L., 1984. The clear sky model. *ASHRAE J.* 26, 27–29.
- Rahim, R., Baharuddin, Mulyadi, R., 2004. Classification of daylight and radiation data into three sky conditions by cloud ratio and sunshine duration. *Energy Build.* 36, 660–666.
- Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* 20, 53–65.
- Salvador, S., Chan, P., 2003. Determining the number of clusters/segments in hierarchical clustering/segmentation algorithms. In: *Tools with Artificial Intelligence*, 2004. ICTAI 2004. 16th IEEE International Conference on. pp. 575–584.
- Skartveit, A., Olseth, J.A., 1992. The probability density and autocorrelation of short-term global and beam irradiance. *Sol. Energy* 49, 477–487.
- Soubdhan, T., Emilion, R., Calif, R., 2009. Classification of daily solar radiation distributions using a mixture of Dirichlet distributions. *Sol. Energy* 83, 1056–1063.
- Stein, J., Hansen, C., Reno, M., 2012. The variability index: a new and novel metric for quantifying irradiance and PV output variability. *World Renew. Energy*, 1–7.
- Tovar-Pescador, J., 2008. Modelling the statistical properties of solar radiation and proposal of technique based on Boltzmann statistics. In: *Modeling Solar Radiation at the Earth Surface*. Springer, pp. 55–91.
- Umemiya, N., Kanou, T., 2008. Classification of sky conditions by the ranges considering CIE standard for general sky. *J. Light Vis. Environ.*
- Wilcox, S., Marion, W., 2008. Users manual for TMY3 data sets. *Renew. Energy* 51.
- Zagouras, A., Inman, R.H., Coimbra, C.F.M., 2014. On the determination of coherent solar microclimates for utility planning and operations. *Sol. Energy* 102, 173–188.
- Zagouras, A., Kazantzidis, A., Nikitidou, E., Argiriou, A.A., 2013. Determination of measuring sites for solar irradiance, based on cluster analysis of satellite-derived cloud estimations. *Sol. Energy* 97, 1–11.

ANEXO D

Manuscript Number: SE-D-17-01162

Title: The clear sky yield index: a new indicator for qualifying the solar resource for CSP plants

Article Type: Research paper

Section/Category: Solar resources & forecasting, Meteorology, Grid Integration

Keywords: solar radiation; sky conditions; parabolic trough; CSP plant

Corresponding Author: Ms. SARA MORENO, M.D.

Corresponding Author's Institution: University of Seville

First Author: SARA MORENO, M.D.

Order of Authors: SARA MORENO, M.D.; Manuel A Silva-Pérez, Ph. D.; Lourdes Ramírez-Santigosa, Ph. D.; Isidoro Lillo-Bravo, Ph. D.

Abstract: A novel index called clear sky yield index or k_p is introduced to characterize the production of each day removing seasonal effects (daylight time and incidence angle). This index is defined as the quotient between the production of a plant for a period and the production of the plant for the same period under the clearest sky conditions. The k_p index allows to assess the solar resource features from the perspective of the production of a Concentrated Solar Power (CSP) plant and to compare the exploitation of the solar resource by CSP plants with different configurations in the same location. In this work, the electrical production of two Parabolic Trough Plants (PTP) with and without Thermal Energy Storage (TES) system and an installed capacity of 50 MWe is estimated using System Advisor Model (SAM). Sixteen years of 10-min Direct Normal Irradiance (DNI) measurements from a radiometric station sited in Seville (Spain) are used as input. Results of daily k_p in Seville for both types of PTPs are assessed as well as its relationship with three indexes (k_b , VI and F_m) that represent three features of the DNI profiles: daily energy (direct normal irradiation), variability and temporal distribution. The comparison between both plants in terms of k_p shows the benefits of the TES system, especially in days with relatively high energy and high variability. The dimensionless k_p index allows to compare the utilization of the solar resource on a location of CSP plants with different configuration. This property can be used to optimize the configuration of the CSP plants during the design stage. Additionally, every day of this database is classified in 10 different day types according to these three features of the DNI following an unsupervised clustering procedure developed in a previous work. Finally, the distribution functions of the daily k_p for each type of day and plant configuration are assessed revealing that some day types can be grouped from the production of the PTPs viewpoints: in 7 day types for the PTP without TES and in 6 for the PTP with TES. These results can help to forecast the most probably range of production for a specific day.

Suggested Reviewers: Frank Dinter
frankdinter@sun.ac.za

Manuel Blanco
m.blanco@cyi.ac.cy

Carlos Fernández-Peruchena
cfernandez@cener.com

Christian Gueymard
Chris@SolarConsultingServices.com

The dimensionless clear sky yield index, or k_p index, is introduced
 k_p index allows to assess solar resource features from the CSP plant yield standpoint
The use of the solar resource of a PT plant with and without TES are compared
The comparison highlights the benefits of TES in days with high variability
The relation between day types according to sky conditions and k_p is studied

1 **The clear sky yield index: a new indicator for qualifying the solar resource for**
2 **CSP plants.**

3 S. Moreno-Tejera^a, M.A. Silva-Pérez^a, L. Ramírez-Santigosa^b, I. Lillo-Bravo^a

4 ^a *Department of Energy Engineering, University of Seville, Spain*

5 ^b *Solar Radiation Unit, Renewable Energy Division, CIEMAT, Spain*

6 **Corresponding author**

7 *Address: Department of Energy Engineering, Escuela Técnica Superior de Ingeniería, University of Seville, Camino*
8 *de los Descubrimientos s/n. 41092, Seville, Spain.*

9 *E-mail: smoreno2@etsi.us.es*

10 **ABSTRACT**

11 A novel index called clear sky yield index or k_p is introduced to characterize the production of
12 each day removing seasonal effects (daylight time and incidence angle). This index is defined as
13 the quotient between the production of a plant for a period and the production of the plant for
14 the same period under the clearest sky conditions. The k_p index allows to assess the solar
15 resource features from the perspective of the production of a Concentrated Solar Power (CSP)
16 plant and to compare the exploitation of the solar resource by CSP plants with different
17 configurations in the same location.

18 In this work, the electrical production of two Parabolic Trough Plants (PTP) with and without
19 Thermal Energy Storage (TES) system and an installed capacity of 50 MWe is estimated using
20 System Advisor Model (SAM). Sixteen years of 10-min Direct Normal Irradiance (DNI)
21 measurements from a radiometric station sited in Seville (Spain) are used as input. Results of
22 daily k_p in Seville for both types of PTPs are assessed as well as its relationship with three
23 indexes (k_b , VI and F_m) that represent three features of the DNI profiles: daily energy (direct
24 normal irradiation), variability and temporal distribution. The comparison between both plants
25 in terms of k_p shows the benefits of the TES system, especially in days with relatively high
26 energy and high variability. The dimensionless k_p index allows to compare the utilization of the
27 solar resource on a location of CSP plants with different configuration. This property can be
28 used to optimize the configuration of the CSP plants during the design stage. Additionally,
29 every day of this database is classified in 10 different day types according to these three features
30 of the DNI following an unsupervised clustering procedure developed in a previous work.
31 Finally, the distribution functions of the daily k_p for each type of day and plant configuration are
32 assessed revealing that some day types can be grouped from the production of the PTPs

33 viewpoints: in 7 day types for the PTP without TES and in 6 for the PTP with TES. These
34 results can help to forecast the most probably range of production for a specific day.

35

36 Keywords: Solar energy, solar radiation, day types, sky conditions, parabolic trough,
37 Concentrated Solar Power (CSP) plant, System Advisory Model (SAM).

38 **Nomenclature**

39 *DNI*: Direct Normal Irradiance (or Irradiation)

40 *DNI_{CS}*: Direct Normal Irradiance (or Irradiation) under clear sky conditions

41 *GHI*: Global Horizontal Irradiance (or Irradiation)

42 *ANI*: Aperture Normal Irradiance (or Irradiation). The product of the DNI times the cosine of
43 the incidence angle

44 *CSP*: Concentrating Solar Power

45 *PTP*: Parabolic Trough Plant

46 *TES*: Thermal Energy Storage

47 k_p^{At} clear sky yield index for the period Δt

48 P^{At} electric energy produced by the plant during the period Δt

49 P_{CS}^{At} electric energy produced by the plant during the period Δt under clear sky conditions

50 I_{bn} direct normal irradiance

51 I_{cs} direct normal irradiance in clear sky conditions

52 k_b transmittance index

53 *VI* variability index

54 F_m morning time fraction

55 F_t time fraction

56

57 **1. Introduction**

58 The electricity yield of a given Concentrated Solar Power (CSP) plant depends mainly on the
59 amount of DNI available. Its temporal distribution also has a significant impact, not only due to

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93

time-related geometrical effects (incidence angle, daylight time), but also to the interaction between the dynamics of the DNI and of the plant itself. The knowledge of relevant features of the solar resource can be useful to assess the performance of a CSP plant or to define operating strategies based on the prediction of these features. For example, Rohani et al. (2016) analysed operational data of Andasol 3 (a 50 MW PTP with a 2-tank TES of 7.5 full-load hour capacity located in Guadix, Spain) from different seasons considering three day types (clear, cloudy and partly cloudy) in order to evaluate a simulation model. The operating strategy of the model was adapted according to the operating data of the real plant. From the analysis of the measured data, they conclude that the most important parameter to decide the operation mode of the plant is based on the forecast of the available DNI of the next days. So, the TES system operating strategy implemented in the model is based on the sky conditions of the two next days.

In the literature, several dimensionless indexes have been studied to represent solar radiation features (Kang and Tam, 2013; Liu and Jordan, 1960; Moreno-Tejera et al., 2017; Skartveit and Olseth, 1992; Stein et al., 2012). The solar radiation features can be used to classify days according to the sky conditions in different types, but the classification criteria use to be application-related. The classification of days according to the state of the sky has been addressed by many authors (Kang and Tam, 2013; Muselli et al., 2000; Soubdhan et al., 2009; Stein et al., 2012) using Global Horizontal Irradiance (GHI) as input variable with the aim of analysing and improving the integration of PV plants into the electrical grid. Classification methods are useful for forecasting methodologies (Jiménez-Pérez and Mora-López, 2016). Gaston-Romeo et al. (2011) propose a classification methodology based on the morphology of the DNI curves oriented to establish a classification of days useful for CSP plants. But, the relation between the types of days obtained and the production of a CSP is not addressed.

The objective of this work is to present a new index that could help in the characterization of the most significant qualitative features of solar radiation from a CSP plant operation perspective. These features are used in classification methods, but can be used for others purposes, as the assessment of the solar resource on a location or in solar radiation forecast methods (Kang and Tam, 2015). The new index is called clear sky yield index or k_p and is defined as the relation between the production of a CSP plant in a specific period (day, month, season) and the production of the same plant for the same period but under clear sky conditions. This dimensionless index allows assessing the types of days classified according to the daily DNI features relevant for the operation of a CSP plant. Moreover, for a specific location and the same solar resource, this index also makes possible the comparison of the utilization of the solar resource of CSP plants with different configurations.

94 In a previous work, Moreno-Tejera et al. (2017) apply unsupervised clustering techniques to
95 classify thirteen years (2000-12) of 10-min DNI measurements recorded in Seville. Each day is
96 classified according to its daily DNI profile using three indexes: k_b (Skartveit and Olseth, 1992),
97 VI (Stein et al., 2012) and F_m (Moreno-Tejera et al., 2017), which represent three features of the
98 DNI: energy, time distribution and variability. In this work, the relation between these indexes
99 and the clear sky yield index is studied. Moreover, the characterization of the k_p index for each
100 day type defined according to these features is addressed for two Parabolic Trough Plants
101 (PTPs).

102 This work is organized in two main parts. In the first part, the new index is defined, calculated
103 and characterized at daily scale for two PTPs with different configuration sited in Seville
104 (Spain). To this end, the electrical production of these PTPs for a long period is estimated using
105 System Advisor Model (NREL, 2010) (Michael J. and Gilman, 2011), a publicly available
106 renewable energy systems simulation program developed by the National Renewable Energy
107 Laboratories (NREL) and used by other authors for similar purposes (Casati et al., 2015;
108 Meybodi et al., 2017; Polo et al., 2016). Sixteen years of 10-min DNI measurements from a
109 radiometric station sited in Seville are used as input. The PTPs modelled are similar to two real
110 plants currently in operation: Andasol 3 (Guadix, Spain) and Solnova 1 (Sanlúcar La Mayor,
111 Spain). Both plants have an installed net capacity of 50 MWe; the first one has a 7.5 hours TES
112 system while the second has no TES. In the second part, the daily indexes k_b , VI and F_m are
113 calculated for the same period and the relationship of daily k_p versus each of these indexes is
114 addressed. Finally, using the selected indexes and applying the unsupervised clustering
115 technique described in (Moreno-Tejera et al., 2017), the k_p behaviour at each cluster is assessed.

116 2. The clear sky yield index. Definition and calculation

117 The clear sky yield index k_p is defined as the ratio between the electricity production of a CSP
118 plant during a certain period (daily, monthly, annual) and the electricity production of the plant
119 if there had been clear sky conditions all the time during the same period (daily, monthly,
120 annual):

$$121 \quad k_p^{\Delta t} = \frac{P^{\Delta t}}{P_{CS}^{\Delta t}} \quad (1)$$

122 where $P^{\Delta t}$ is the electric energy produced by the plant during the period Δt and $P_{CS}^{\Delta t}$ the electric
123 energy produced by the plant during the same period, but under clear sky conditions. The
124 dimensionless definition of this index removes seasonal effects derived from the sun-earth
125 geometry and allows the comparison of the exploitation of the solar radiation of different days
126 of the year for a specific plant or the exploitation of the radiation of a specific period from CSP
127 plants with different configurations. This index can be defined for the gross or for the net

128 electricity production. A similar index could also be defined to evaluate the thermal energy
 129 collected by the solar field. In this work, the k_p index is calculated with the gross power
 130 produced and used to compare the utilization of the solar resource on a location of two parabolic
 131 trough CSP plants with different configurations.

132 3. Methodology

133 In this work, the value of the daily k_p index of two PTPs with and without TES for a specific
 134 location is assessed in a first step. To this end, 16 years of DNI measurements are used to
 135 simulate the performance of each plant and calculate the k_p index for the period (Fig 1). In a
 136 second step, the types of days classified according to the sky conditions defined by Moreno-
 137 Tejera et al. (2017), are statistically characterized from the perspective of the production of both
 138 PTPs. This classification is based on three solar radiation indexes that represent the variability,
 139 the energy and the time distribution of the daily DNI curves. To help characterize the relation
 140 between k_p index and day types, the relation between k_p and the three classification indexes is
 141 evaluated and then the frequency distribution of the k_p index for each type of day is obtained
 142 (Fig. 2).

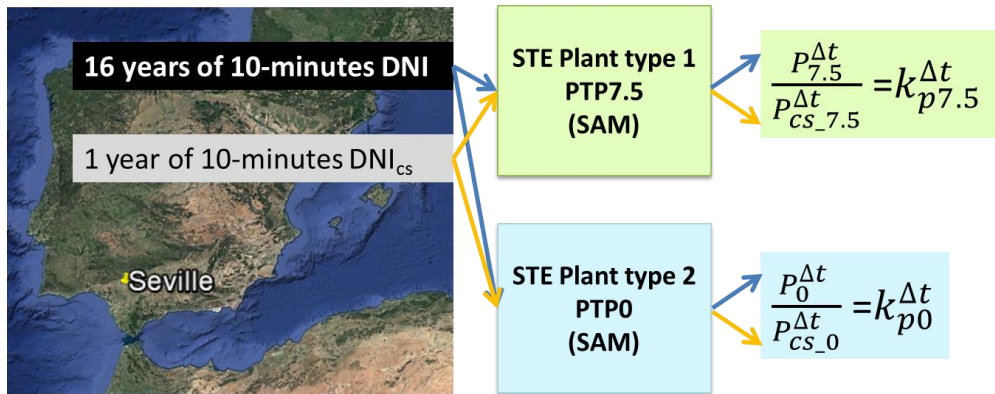


Fig. 1. Schematic description of the process followed in the first step of this work: First, 16 years of 10-min DNI values and a year of 10-min DNI under clear sky conditions (DNI_{CS}) values are used to simulate the performance of a PTP without TES and a PTP with 7.5h of TES using SAM. Then, the daily k_p index is calculated and assessed for the same 16 years for the plants with and without TES.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

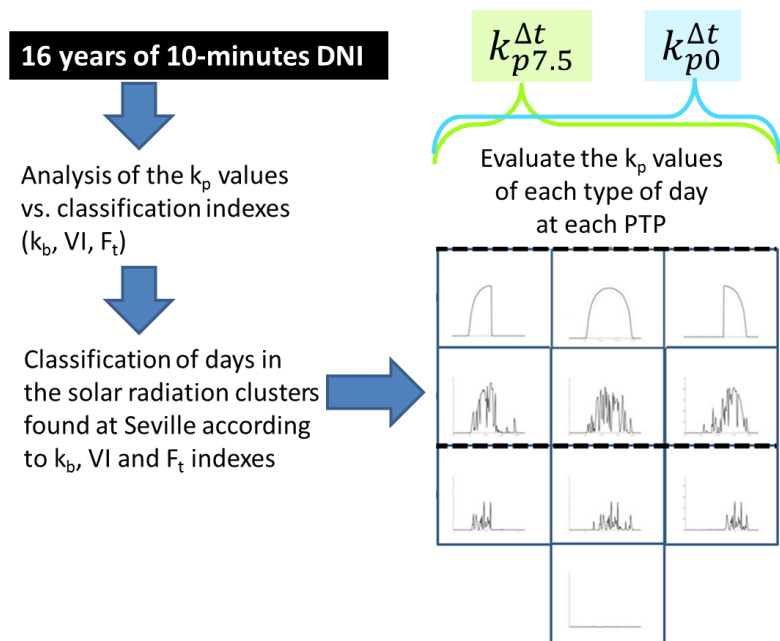


Fig. 2. Schematic description of the process followed in the second step of this work: First, the relations between k_p and k_b , VI and F_m , are studied. These indexes are related to three features of the DNI curves (energy, variability and temporal distribution). Then, the k_p index of each of the 10 day types defined in a previous work according to these three features of the DNI curves is characterized.

3.1. Meteorological data

The meteorological database used for the simulation covers sixteen years (2000-2015) of DNI measurements recorded in Seville (37.40° N, 6.01° W) by the Group of Thermodynamics and Renewable Energy (GTER) of the University of Seville. These measurements were recorded every 5 seconds with an Eppley Normal Incidence Pyrheliometer (NIP) mounted on a Kipp & Zonen 2AP 2-axis tracker. This pyrheliometer is a secondary standard device according to ISO specifications. The hourly and daily uncertainty specified by the manufacturer is $\pm 1\%$, assuming proper maintenance of the instruments. The GTER station follows a regular maintenance and calibration procedure according to the recommendations from the instrument manufacturers. The database passed a quality control check (Moreno-Tejera et al., 2015) and it was gap filled when necessary by applying gap filling techniques described in (Moreno-Tejera et al., 2016). The DNI database was averaged every 10 minutes for this implementation.

3.2. Plant models and simulation

Two PTPs similar to operational plants sited in the South of Spain and currently in operation have been selected as reference to model the plants that are going to be analysed: Andasol 3

160 (NREL, 2013) and Solnova 1 plant (NREL, 2017). Both plants have an installed capacity of 50
 161 MWe of net nominal power and use parabolic trough technology to collect the solar energy. The
 162 main characteristics of these plants are summarized in Table 1.

Table 1. Main technical data used in SAM to model the plants. PTP0 and PTP7.5 models are based on the characteristics of SOLNOVA 1 and ANDASOL 3, respectively, both plants located in Spain.

Parameter	PTP0	PTP7.5
Net output at design (MWe)	50	50
Collector type	EuroTrough ET150	EuroTrough ET150
Receiver type	Schott PTR70 2008	Schott PTR70 2008
Number of loops	90	156
Collectors per loop	4	4
Solar field aperture area (m ²)	300,000	510,120
HTF	Therminol VP-1	Therminol VP-1
Design loop outlet Temp. (°C)	391	391
Thermal Storage Capacity (full-load equivalent hours)	0	7.5
TES type and medium	-	Two-tank, Hitec Solar Salt

163
 164 The Andasol 3 and Solnova 1 models have been used to estimate the electricity generation of
 165 both plants during the period 2000 – 2015 in 10-minute time-steps with SAM, (version
 166 2017.1.17). Hereinafter, the plant modelled similar to Andasol 3 is called PTP7.5 and the plant
 167 modelled similar to Solnova 1 is called PTP0.

168 To assess the daily k_p^d value, the electric energy produced by the plant every day must be
 169 computed. The operation strategy of the PTP7.5 TES has been defined to provide full power
 170 output the maximum time as possible. The purpose of defining this operating strategy is to use
 171 all the remaining energy in the storage tanks after the sunset, so that the TES is empty at
 172 sunrise. The large capacity of the PTP7.5 TES permits the generation of electricity during the
 173 first hours of the next day. To avoid the effect of the energy collected during a given day on the
 174 assessment of the next day, the electricity yield is computed from 4:00 a.m. to 3:50 a.m. of the

175 next day. As shown Fig. 3, where the electricity generation of PTP7.5 during the most
 176 favourable day, June 21st, is simulated using the enveloping clear day (Moreno-Tejera et al.,
 177 2017) as input, the hot tank is empty before 4:00 a.m. of June 22nd.

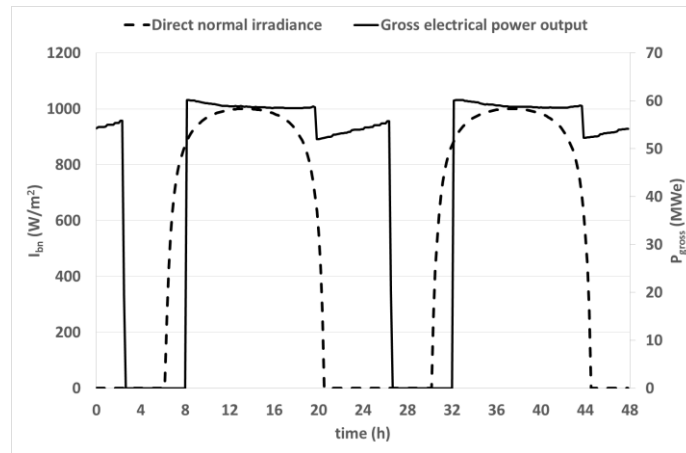


Fig. 3. DNI and gross electrical power produced by the PTP7.5 plant on the 21th and 22th of June in envelope clear sky conditions.

3.3. Assessment of the daily clear sky yield index

The daily clear sky yield index is calculated and statistically characterized for two different PTPs sited at a specific location. To this end, 16 years of 10-min DNI measurements are used to simulate the performance of two different PTPs models. Moreover, a year of 365 envelope clear sky days is built to simulate the performance of the two PTPs models in clear sky conditions. This year is necessary to calculate the k_p indexes. To build these DNI curves, a clear sky model is adjusted with the moments of less turbidity recorded at the site as function of the solar elevation angle (Moreno-Tejera et al., 2017). Thus, 16 years of daily k_p values of two PTPs with different configuration are used to study and show the capacities of this new index. Fig. 1 shown a schematic description of the process followed in this section.

3.4. Relation between the clear sky yield index and the day-type clusters

In a previous work, the DNI data from the GTER station of the University of Seville (37.40°N, 6.01°W) for the period 2000-12 were used to test a classification of days methodology based on three features of the daily DNI curves: the energy, the variability and the temporal distribution of energy (Moreno-Tejera et al., 2017). Three indexes were used to represent each feature: the *clearness index*, k_t (Liu and Jordan, 1960); the *variability index*, VI (Stein et al., 2012) and the *morning fraction*, F_m (Moreno-Tejera et al., 2017). The relation between these three indexes and k_p is analysed for the 16 years of measurements and the two PTPs models outlined in the section 3.2. Then, the 16 years are classified according to the methodology proposed by Moreno-Tejera

197 et al. (2017) and the k_p values for each type of day are assessed. Fig. 2 shows a schematic
 198 description of the process followed in this part of the work.

199 **3.4.1. Evaluation of clear sky yield index vs. classification indexes**

200 Several indexes have been used in the literature to try to classify days according to the sky
 201 condition. Moreno-Tejera et al. (2017) proposed a classification method based on three
 202 dimensionless indexes defined from DNI measurements. These indexes take into account three
 203 useful features of the DNI curves from a CSP viewpoint: the daily energy, the distribution over
 204 time of this energy and the variability (high frequency changes) of the instantaneous values
 205 caused by the passage of clouds throughout the course of the day.

206 The index used to characterize the daily energy of the DNI is the transmittance index (k_b)
 207 introduced by Skartveit and Olseth (1992). The definition is shown in Eq (2):

$$208 \quad k_b = \frac{H_{bn}^d}{H_{cs}^d} \quad (2)$$

209 where H_{bn}^d is the daily DNI and H_{cs}^d is the daily irradiation in clear sky conditions. Both daily
 210 irradiation values are obtained as the integration of the 10-min irradiance values.

211 In (Moreno-Tejera et al., 2017) two indexes, previously used for the GHI component, are
 212 evaluated to characterize the variability of the DNI. As result, the Variability Index (VI) was
 213 found as the most proper index for the classification method proposed. This index was
 214 introduced by Stein et al. (2012) to identify periods and locations with high GHI variability. VI
 215 compares the length of the GHI curves with the length of the corresponding clear sky GHI
 216 curves. The VI used by Moreno-Tejera et al. (2017) is defined as the ratio between the length of
 217 the DNI curve and the length of the maximum enveloping clear sky day curve:

$$218 \quad VI = \frac{\sum_{k=2}^n \sqrt{(I_{bn_k} - I_{bn_{k-1}})^2 + \Delta t^2}}{\sum_{k=2}^n \sqrt{(I_{cs_k} - I_{cs_{k-1}})^2 + \Delta t^2}} \quad (3)$$

219 where I_{cs} is the maximum DNI clear sky average every 10 minutes, Δt refers to an interval of 10
 220 minutes, and n is the number of 10-minute intervals of the considered day. This index was
 221 normalized dividing by the maximum VI obtained for using in the classification method.

222 Characterizing the time distribution of a curve by mean of an index is not an easy task. Moreno-
 223 Tejera et al. (2017) proposed a new index that inform about the fraction of energy concentrated
 224 for the morning. The *morning fraction* is defined as the ratio between the accumulated DNI in
 225 the first half of the day and the accumulated DNI for the whole day. This index allows knowing
 226 if the daily radiation is concentrated in the first or the second part of the day:

$$F_m = \frac{H_{bn_m}^d}{H_{bn}^d} \quad (4)$$

where $H_{bn_m}^d$ is the DNI recorded from the sunshine to the solar noon and H_{bn}^d is the daily DNI.

In the classification method, this index was segmented in three values to optimize and facilitate the clustering of the data. As result, a new index called *time fraction* F_t is defined by Eq. (5).

$$F_t = 1 \quad \text{if } F_m \leq 0.3$$

$$F_t = 0.5 \quad \text{if } 0.3 < F_m < 0.7$$

$$F_t = 0 \quad \text{if } F_m \geq 0.7 \quad (5)$$

The relation between the k_b , VI and F_m with the k_p for the 16 years of measurements and the two PTPs models is addressed in this work. Moreover, 16 years with symmetric DNI curves to the 16 years of recorded DNI measurements respect to the solar noon are built and simulated. The relation between k_p and F_m for this *symmetric set* is also evaluated. The analysis of both sets avoids conclusions related to the typical days from the location because of the climate.

3.4.2. Clear sky yield index vs. the day types

In (Moreno-Tejera et al., 2017) the k_b , VI and F_t indexes were used to group the days for the period 2000-12 from the Seville station using unsupervised clustering techniques that defined the number of groups and their features. Results were compared with a visual classification carried out by 5 experts according to the 11 day types represented in Fig. 4. The best result classifies the days in a total of 10 clusters or day types. The *medoids* (or centres of the clusters) of each cluster were matched to the day types used as reference by the experts to find similitudes between plots, resulting in a merging between the days with very high and high energy, balanced and with low variability. The reference plots used by the experts and the resulting merging between day types are illustrated in Fig. 4. Moreno-Tejera et al. (2017) identified 10 day types from the solar radiation viewpoint, suggesting that the 4 low-level energy classes could be merged for the application to CSP plants. In Fig 4, the day types as classified as per (Moreno-Tejera et al., 2017) are grouped by similar levels of energy and variability in rows and similar temporal distribution in columns. To help to identify each type of day with its features, each cluster is called with the first letter of its type of energy temporal distribution (“M” morning, “B” balanced and “A” afternoon) and the first letter of its level of energy (“H” high, “M” medium, “L” low and “N” null). The relation of each day with the level of variability is shown in Fig. 4.

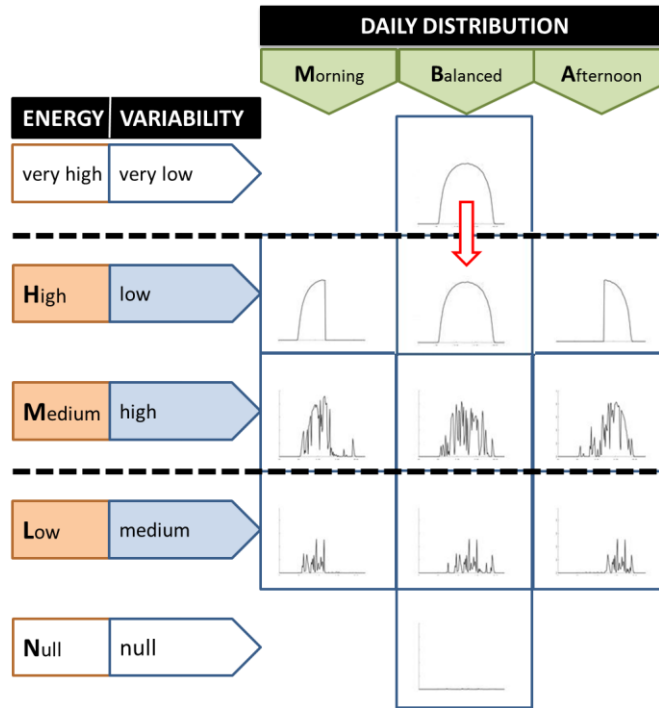


Fig. 4. DNI representation of the days corresponding to the reference plots used by the experts to carry out the visual classification. The experts proposed these eleven classes of day types initially, but the unsupervised classification found that the class on the top of the figure is classified among the high energy days.

256 The days from the period 2013 – 2015 have been automatically classified into the same 10
 257 classes or day types. The data were quality-controlled and gap filled when necessary using the
 258 methodology described in (Moreno-Tejera et al., 2016). Thus, the resulting database consists of
 259 the meteorological database and the classification of all the days for the period 2000 – 2015.

260

261 4. Results and discussion

262 In this section, the daily values of the k_p index of PTP0 and PTP7.5 for the 16 years of
 263 measurements of Seville is assessed in a first step. In a second step, the relation between k_p and the
 264 three classification indexes k_b , VI and F_m is studied and the statistical characterization of the k_p
 265 index for each type of day defined by Moreno-Tejera et al. (2017) is obtained.

266 4.1. Assessment of the clear sky yield index

267 In Fig 5 the daily k_p values for the PTP0 (left side) and the PTP7.5 (right side) for the 16 years
 268 are represented. The median value of each day of the year for the period is also shown in both
 269 graphs (red lines). In the case of PTP0, k_p values close to 1 are reached mainly in summer, but
 270 also in spring and the first part of autumn. In winter, the maximum values obtained are around
 271 0.9, because the clear sky model is adjusted with the clearest instances of cloudless sky, which

272 are mainly found in spring. The frequency of days with k_p close to 1 for summer, spring and
 273 autumn is higher in PTP7.5. In general, the median k_p values are higher for PTP7.5 although
 274 these differences are less significant in winter season.

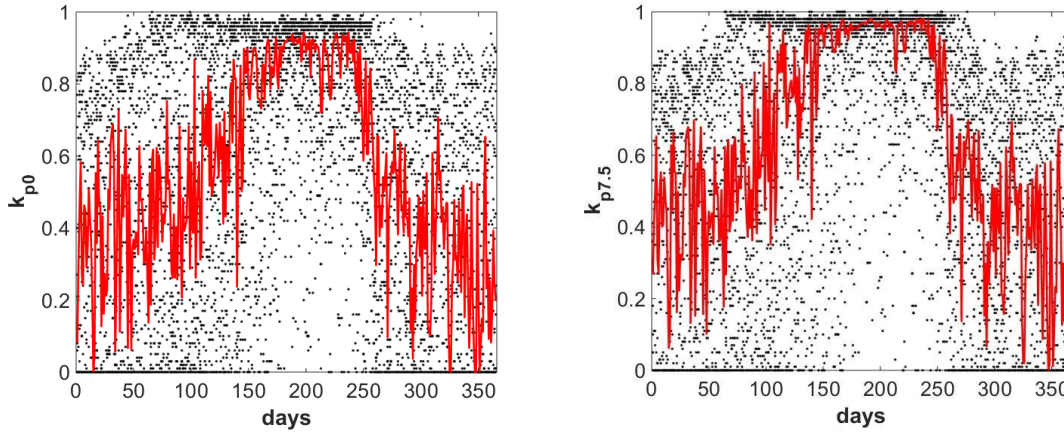


Fig. 5. k_p values and median k_p values of each day of the year (red line) of PTP0 (left side) and PTP7.5 (right side) for the 16 years considered.

275
 276 On the left side of Fig. 6, the k_p values of PTP7.5 are represented versus the k_p values of PTP0.
 277 As shown, in almost all days k_p values of PTP7.5 are higher than for PTP0, showing a better
 278 utilization of the solar resource available. However, some days with higher values of k_p from
 279 PTP0 are observed. In these days the PTP0 generated more energy with respect to the
 280 production of the corresponding enveloping clear sky day than the PTP7.5. These days are
 281 winter days with high energy and low variability in most of cases.

282 The right side of Fig. 6 shows the frequency distributions of the k_p values for the two PTP
 283 models. Both distributions have a bimodal character, with a maximum located in the range 0-0.1
 284 and the other in the range 0.9-1. The first maximum represents the production of cloudy days
 285 where the energy yield is almost null. The second maximum represents the production of clear
 286 days, very frequent in summer season as shown Fig 5. Both maxima of the bimodal distribution
 287 are similar in frequencies for PTP0, however, the maximum located in the range of 0.9-1 is
 288 higher for PTP7.5. In more than 30 % of the days, PTP7.5 yields as much energy as it would
 289 have yielded in the corresponding enveloping clear sky day, almost 40 % of days more than in
 290 the case of PTP0.

291 The clear sky yield index k_p can be used to evaluate and compare the exploitation of the solar
 292 resource by two CSP plants with different configurations sited in the same location. This
 293 property can be used to optimize the design of the plant.

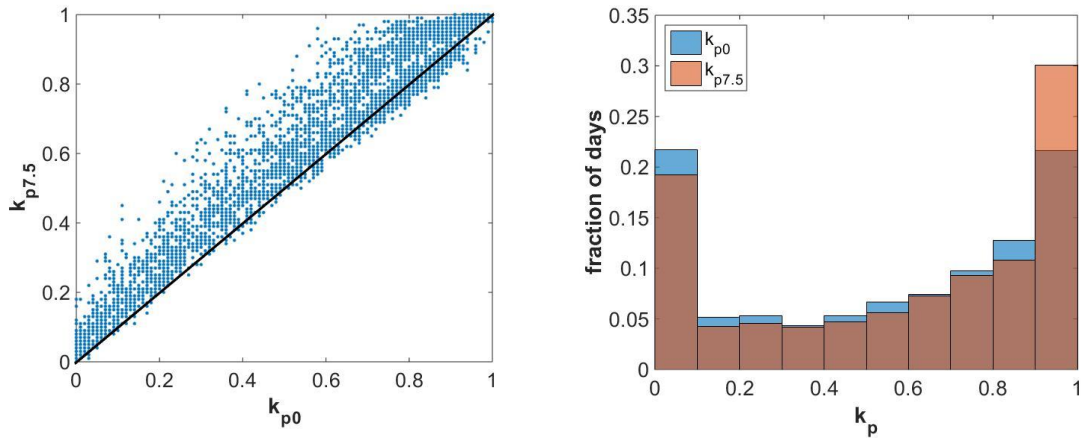


Fig. 6. Representation of k_p values of PTP7.5 versus PTP0 for the period 2000-15 (left side). Frequency distribution function of k_p of PTP0 (blue) and of PTP7.5 (red) for the period 2000-15 (right side).

294

4.2. Evaluation of clear sky yield index vs. classification indexes

In this section, the relation between k_p and the k_b , VI and F_m indexes for the 16 years of measurements and the PTP with and without TES system is presented and discussed.

4.2.1. k_p vs. k_b

Fig. 7 shows the representation of k_p values for PTP0 (left side) and PTP7.5 (right side) versus k_b values for the period 2000-15. In both graphs, a positive correlation between k_b and k_p is observed, but with a significant dispersion. For a value of k_b , a range of k_p values can be reached indicating that, although closely related, k_p does not depend on k_b only. Since the integration of the DNI to calculate the k_b value cannot reflect the fluctuation and the distribution of solar radiation, the stand-alone use of k_b has limitation for the production analysis (Kang and Tam, 2013). Comparing both scatter plots, it is worth noting that the upper limit of PTP7.5 is higher, showing k_p values close to 1 for days with k_b values lower than 0.6 and increasing the number of days that reach this value of k_p .

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

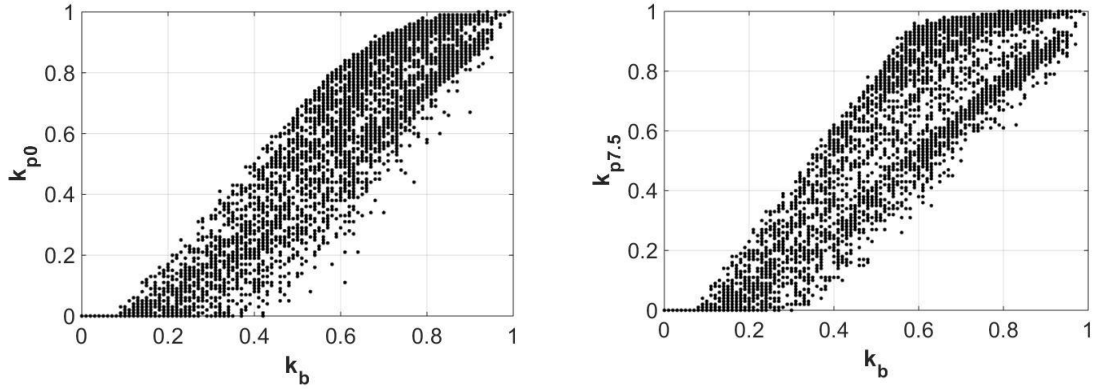


Fig. 7. Representation of k_p values of PTP0 (left side) and of PTP7.5 (right side) versus k_b values for the period 2000-15.

308

309 The frequency distributions of the ratio k_p to k_b are shown in Fig. 8. Both distributions show a
 310 bimodal behavior with a maximum located in the range of null production and the other
 311 between 0.9 and 1.2. The values higher than 1 represent days where the ratio of production to
 312 daily DNI is higher than the ratio of production to daily DNI for the same day in clear sky
 313 conditions. This frequently occurs in clear days when the thermal energy dumped by the system
 314 (due to partial solar field defocusing) is significantly greater under clear sky conditions than
 315 under actual conditions; it also occurs, for the same reason, in some non-clear days when most
 316 of the energy is received during periods of highest optical efficiency.

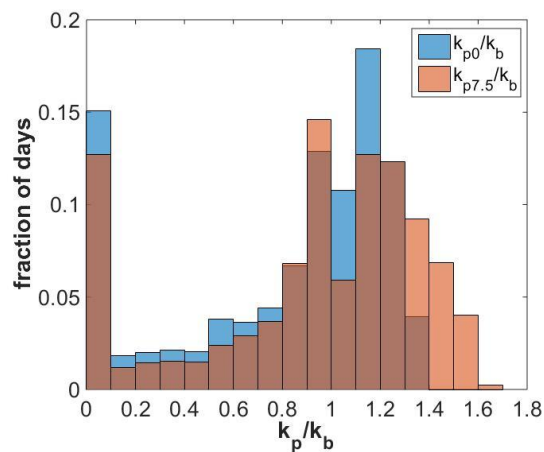


Fig. 8. Frequency distribution function of the ratio k_p/k_b of the PTP0 (in blue) and of the PTP7.5 (in red) for the period 2000-2015.

317

318 4.2.2. k_p vs. normalized VI

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

319 The k_p values of PTP0 (left side) and of PTP7.5 (right side) versus the normalized VI for the
 320 period 2000-15 are represented in Fig. 9. The points located in the higher values of k_p for
 321 normalized VI around 0.1 correspond to clear sky days. Both scatter plots show significant
 322 dispersion, although with some differences. In the case of PTP0 the scatter plot has a triangular
 323 shape defined by a vertical in the value of 0.1 of the VI, the abscissa axis and an upper limit
 324 with a downward slope. This shape shows that for a given value of VI, k_p can take a range of
 325 values, but the upper limit of this range decreases when the VI increases, reflecting the negative
 326 effect that the variability of the resource has in a PTP without TES. For low values of VI, the
 327 range of k_p values extends from 0 to 1 with higher frequency in the range between 0.6 and 1,
 328 indicating a better exploitation of the resource when the variability of the DNI is low. In the
 329 case of PTP7.5 the scatter plot limits have a *squarer* shape: values of k_p close to 1 are reached
 330 on days with VI higher than 0.6, because the TES system reduces the negative effect of
 331 variability and allows for a better use of the available solar resource.

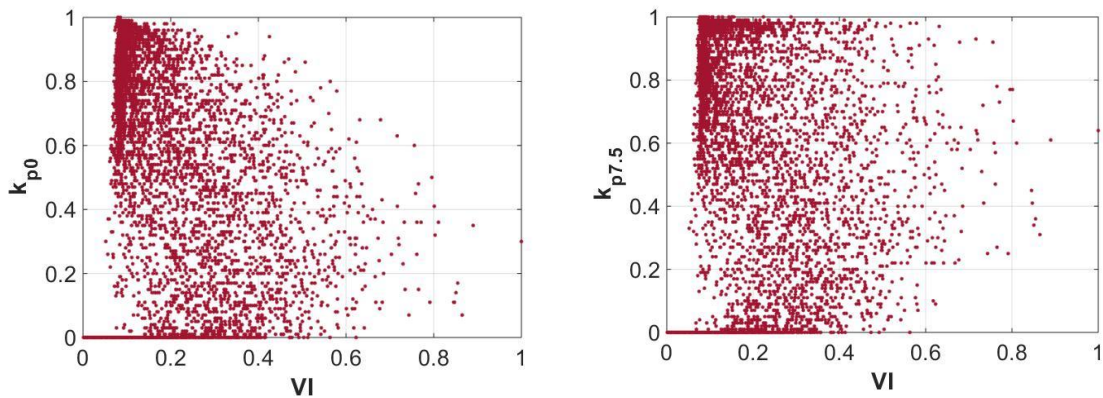


Fig. 9. Representation of k_p values of PTP0 (left side) and of PTP7.5 (right side) versus the normalized VI for the period 2000-15.

332

333 4.2.3. k_p vs. F_m

334 To study the relation between F_m and k_p , both indexes are calculated for the 16 years of DNI
 335 measurements and for 16 years of symmetrical days (DNI curves time-inverted with respect to
 336 the solar noon). The evaluation of both sets allows us to compare the values of k_p in days with
 337 energy mainly concentrated in the morning and days with energy mainly concentrated in the
 338 afternoon that have symmetric profiles, avoiding the effect of possible local climatic
 339 characteristics.

340 In Fig. 10 the k_p values for the measured set and symmetric set of PTP0 (left side) and PTP7.5
 341 (right side) are plotted versus F_m . Both scatter plots are almost symmetrical with respect to F_m
 342 equal to 0.5. The scatter plots have a significant dispersion –showing that k_p is not a function of

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

343 F_m only- but a well-defined upper limit, with maximum values of k_p equal to 1 for F_m values
 344 around 0.5 (the apparent symmetry axis). In the case of the PTP7.5, this limit is less defined and
 345 decreases more slowly as F_m diverges from 0.5 (k_p reaches 0.9 for F_m equal to 0.7, while for
 346 PTP0 k_p does not exceed 0.7 for the same value of F_m). The dotted vertical lines shown in Fig.
 347 10 limit the values selected in the classification method to define the days with energy mostly
 348 concentrated on the morning (*morning days*) and mostly concentrated in the afternoon
 349 (*afternoon days*).

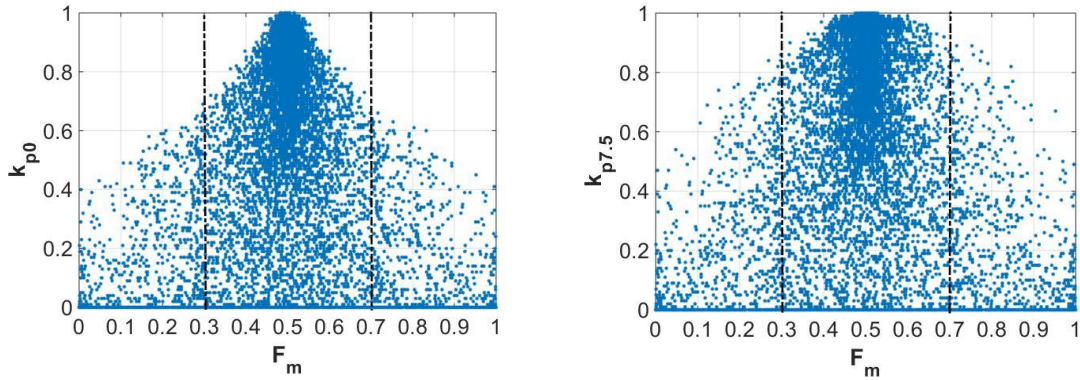


Fig. 10. Representation of k_p values of PTP0 (left side) and of PTP7.5 (right side) versus F_m for the measured DNI set and its symmetric set. The dotted vertical lines indicate the F_m values selected in the classification method to define *afternoon days* ($F_m \leq 0.3$), *balanced days* ($0.3 < F_m < 0.7$) and *morning days* ($F_m \geq 0.7$).

350

351 To compare the behaviour of the PTP plants in *morning days* (MH, MM, ML) and *afternoon*
 352 *days* (AH, AM, AL), Fig. 11 shows the representation of the k_p values of the measured DNI set
 353 and its symmetric for *afternoon days* vs. *morning days* (upper row) and the cumulative
 354 distribution function (CDF) of k_p values for *morning days* and *afternoon days* (lower row). In
 355 the case of PTP0 both representations (upper and lower left side graphs) show that the afternoon
 356 days have mostly higher k_p values than morning days. However, in the case of PTP7.5 the
 357 tendency is not so clear (Fig. 11, upper and lower right side graphs). For higher k_p values, the
 358 morning days seem to be more favourable for the exploitation of the solar resource, but for
 359 lower values this trend is inverted.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

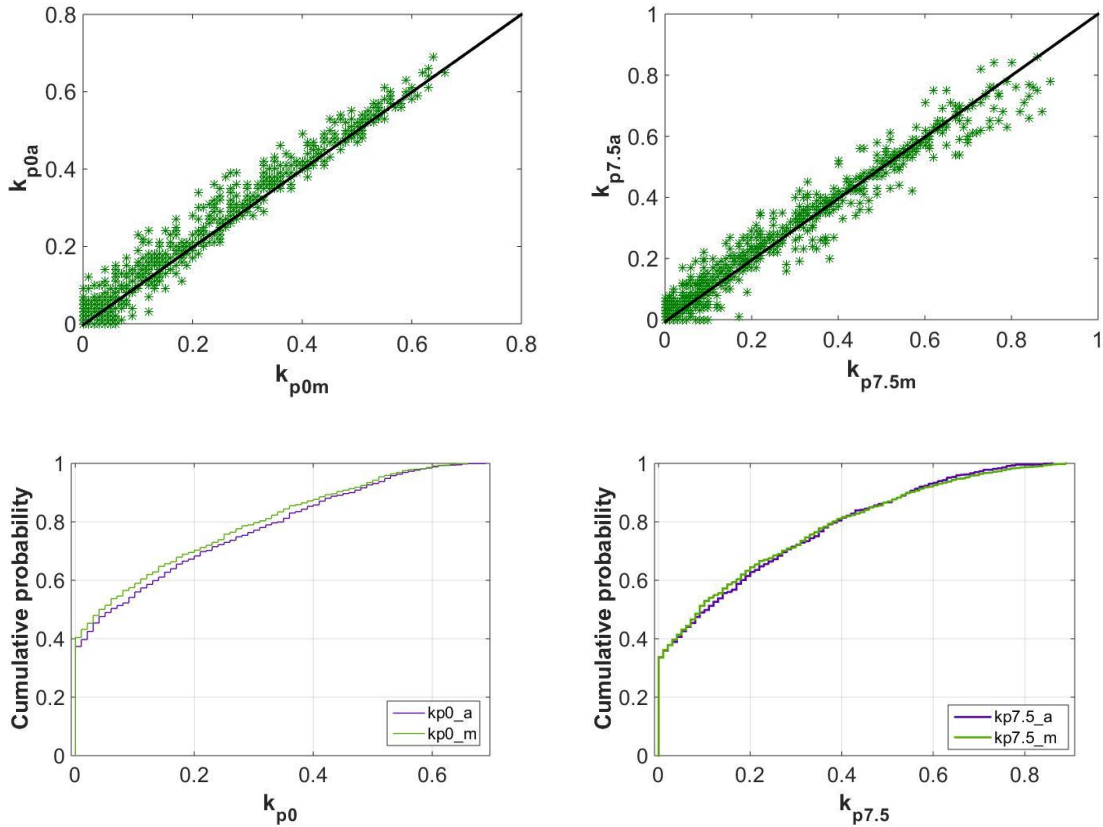


Fig. 11. Upper row: k_p values of *afternoon days* versus *morning days* for the measured DNI set and its symmetric set for PTP0 (left side) and PTP7.5 (right side); second row: CDF plots of k_p values of *afternoon days* (violet line) and *morning days* (green line) for PTP0 (left side) and PTP7.5 (right side).

360

361 4.3. Clear sky yield index vs the day type clusters

362 Fig. 12 shows box and whisker plots of the values of k_p (period 2000-15) for each day type and
 363 for PTP0 and PTP7.5. Except for the BL, AL and BN day types, the median k_p values of all
 364 types of days are higher in the case of the PTP7.5. The TES system permits a better
 365 management of variability and the excess of radiation in clear days. This is also reflected in the
 366 maximum values of k_p for PTP7.5. However, the minimum values are quite similar. These are
 367 often found for winter days, when the effect of the incidence angle penalizes the performance of
 368 these PTPs. During these days, defocussing is not required and the TES is not efficiently used.
 369 So, in winter days k_p values are similar in both plants.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

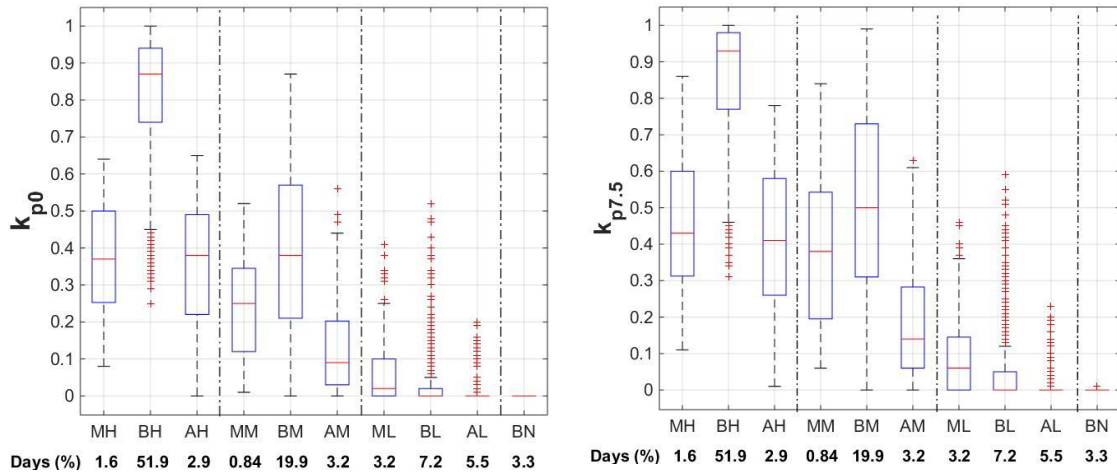


Fig. 12. Box and whisker plots of the k_p values of each type of day for the plant with (right side) and without TES (left side) for the period 2000-15. The boxes are defined by the percentiles 25 and 75 and 99.3% of the values (considering normal distribution of the set) are within the whiskers limits. The red lines within the boxes correspond to the median values of the sets. The day types with high, medium, low and null energy are separated with a dotted line. The percentage of days from the period belonging to that type of day is shown under each day type label.

370

371 For both plants, the higher values of k_p are obtained for the BH type. More than 80% of the days
 372 classified in this cluster have a k_p higher than 0.7, as shown Fig. 12. The day types MH and AH
 373 show similar boxes with small differences in the extreme values. However, the types MM and
 374 AM show clear differences. The same difference between morning and afternoon days can be
 375 observed for the types of days with low energy: the type ML has k_p values higher than zero in
 376 more than half of days while in the type AL only outliers are not nulls. The day types with high
 377 variability and DNI mainly concentrated during the morning are associated with higher k_p for
 378 this location. This result can be attributed to specific climatological characteristics of Seville,
 379 considering the results regarding the relation between k_p and F_m presented in section 4.2.3.
 380 Noteworthy is the high range of values of k_p found in the day type BM. The operation of the
 381 plants –especially PTP0- during these types of days is very complex. Depending on the
 382 frequency and duration of the transients, the plant must stop the turbine or can continue the
 383 operation thanks to the thermal inertia of the plant. In the last three types of day (BL, AL, BN)
 384 the energy production is basically negligible. The period of the year when these day types are
 385 more frequent is the best period to shut the plant down for maintenance.

386 For a better understanding of the results found in our location, the average and median values of
 387 k_p from the sixteen years of simulation of PTP7.5 and PTP0 for each type of day are shown in

388 Table 2. In addition, the average and median values of $k_{p7.5}/k_{p0}$ from the sixteen years of
 389 simulation are also included. The types MH and AH have very similar statistics, however the
 390 types MM and AM show a different median and average, in agreement with the discussion on
 391 Fig 12. In both cases (with and without TES), the k_p values are higher for the MM type than for
 392 the AM type. Except for BL, AL, BN, all median and average k_p values are higher for PTP7.5
 393 (when the TES is available) but the differences are more notable in the types of days with high
 394 variability (MM, BM and AM). This result can be observed in the average and median values of
 395 $k_{p7.5}/k_{p0}$: $k_{p7.5}$ statistics are more than 30 % higher respect to k_{p0} . The types catalogued with low
 396 or null energy have averages and median values null or almost null. In all cases, the median
 397 values are close to the average except for the BH and AM day types. The frequency
 398 distributions of the k_p values for all types of days are displayed in Fig. 13 for PTP0 and PTP7.5.
 399 The distribution for the day types BH and AM are far from being a normal distribution. In both
 400 plants, the more frequent k_p values for the BH type are located between 0.9 and 1 and on
 401 average, as shown in Table 2, more than 50% of the days in Seville are from this type. Fig. 13
 402 shows that BM type covers a wide range of values of k_p and has a distribution close to normal,
 403 mainly for PTP0, in concordance with the result observed in Fig 12. Since the BM cluster has
 404 the second highest number of days and the plant operation is more complex in these day types,
 405 dividing this group in two or more subgroups could be interesting for the assessment of the
 406 operation. Moreover, the day types BL, AL and BN could be grouped in a single cluster since in
 407 all cases the median and the average values of k_p are null or very close to zero.

Table 2. Average and median values of k_p from the sixteen years of simulation of PTP7.5 and PTP0 for each day type. The last row shows the percentage of days of each cluster. Dotted lines are placed to separate the days with high (H), medium (M), low (L) and null (N) energy.

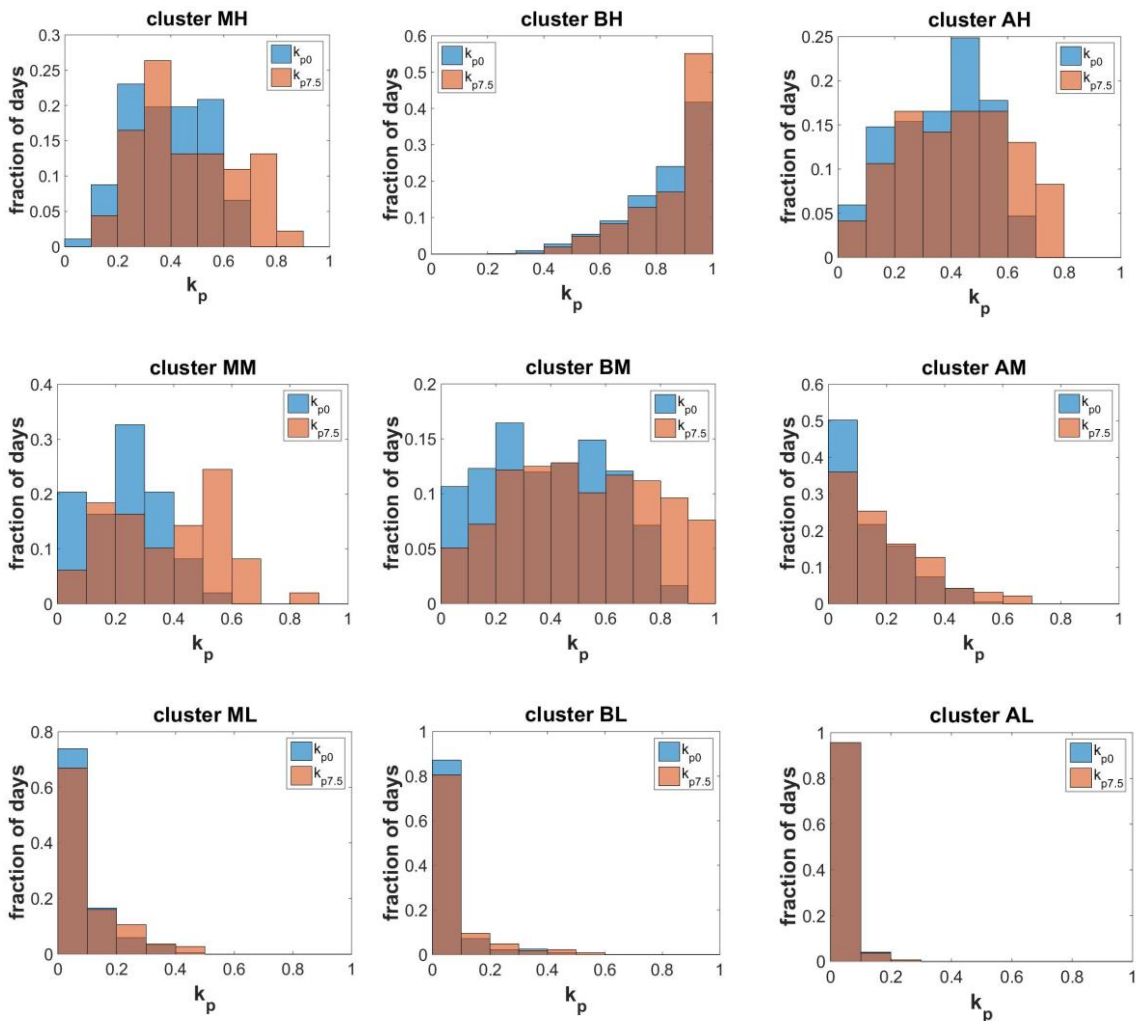
Type of day	MH	BH	AH	MM	BM	AM	ML	BL	AL	BN	
k_{p0}	Average	0.38	0.82	0.36	0.23	0.39	0.13	0.06	0.04	0.01	0.00
	Median	0.37	0.87	0.38	0.25	0.38	0.09	0.02	0.00	0.00	0.00
$k_{p7.5}$	Average	0.46	0.86	0.42	0.38	0.51	0.18	0.09	0.05	0.01	0.00
	Median	0.43	0.93	0.41	0.38	0.50	0.14	0.06	0.00	0.00	0.00
$k_{p7.5}/k_{p0}$	Average	1.21	1.05	1.17	1.65	1.31	1.38	1.50	1.25	1.00	0
	Median	1.16	1.07	1.08	1.52	1.32	1.56	3.00	0	0	0

N. days (%)	1.6	51.9	2.9	0.8	19.9	3.2	3.2	7.7	5.5	3.3
-------------	-----	------	-----	-----	------	-----	-----	-----	-----	-----

408

409 For our location, the features of the days classified as afternoon types have worse features from
 410 the standpoint of a PTP yield. In Fig. 13, the distributions of the k_p values for the AM type show
 411 a clear peak for values between 0 and 0.1 (approx. 50% of values for PTP0 and 35% for
 412 PTP7.5). Moreover, k_p takes negligible values at clusters with null or low values of expected
 413 energy except in the case of ML cluster, where approx. 35% of values for PTP7.5 and 25% of
 414 values for PTP0 are greater than 0.1. Thus, in our location, the expected similar behaviour in
 415 clusters with low or null energy levels does not occur if the incoming energy is collected during
 416 the first part of the day. In this case, small but not negligible energy can be produced.

417 Some similitudes can be observed between the distributions of k_p of different types of days for
 418 PTP0 and PTP7.5.



1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

BM	-	-	-	-	false	true	true	true	true	true
AM	-	-	-	-	-	false	true	true	true	true
ML	-	-	-	-	-	-	false	true	true	true
BL	-	-	-	-	-	-	-	false	true	true
AL	-	-	-	-	-	-	-	-	false	true
BN	-	-	-	-	-	-	-	-	-	false

437

Table 4. Results of the K-S test between the distributions of k_p values of each type of day obtained with the simulation of the period 2000-15 of PTP7.5.

	MH	BH	AH	MM	BM	AM	ML	BL	AL	BN
MH	false	true	false	false	true	true	true	true	true	true
BH	-	false	true	true	true	true	true	true	true	true
AH	-	-	false	false	true	true	true	true	true	true
MM	-	-	-	false	true	true	true	true	true	true
BM	-	-	-	-	false	true	true	true	true	true
AM	-	-	-	-	-	false	true	true	true	true
ML	-	-	-	-	-	-	false	true	true	true
BL	-	-	-	-	-	-	-	false	true	true
AL	-	-	-	-	-	-	-	-	false	true
BN	-	-	-	-	-	-	-	-	-	false

438

439 Table. 5. Qualitative analysis of the k_p values per cluster and PTP. k_p takes negligible values at
440 clusters with null or low values of energy except in the case of ML cluster (red k_p). BH cluster
441 shows the highest k_p values (green) at both PTP.

PTP0		PTP7.5	
Cluster	k_p	Cluster	k_p
BH	$\uparrow k_p$	BH	$\uparrow k_p$
MH & AH	k_p	MH & AH & MM	k_p
MM	k_p		
BM	k_p	BM	k_p
AM	k_p	AM	k_p
ML	$\downarrow k_p$	ML	$\downarrow k_p$
BL & AL & BN	null	BL & AL & BN	null

442

443 A result from the previous analysis and Table 5 is that 4 groups have intermediate values of k_p
444 in the case of PTP0: MH & AH, MM, BM and AM; while there are only three groups in the
445 case of PTP7.5: MH & AH& MM, BM and AM. Considering this result, it can be stated that
446 there are only 7 day-types from the production standpoint for PTP0 and 6 types for PTP7.5.

447 The characterization of day types in a location from the perspective of the production of PTPs
448 can be used in forecasting methods (Jiménez-Pérez and Mora-López, 2016). The forecast of the
449 day types helps to know the most likely level of solar resource utilization (k_p) and as result for a
450 specific day the most probably range of production. Moreover, the prediction of the day types
451 for the next days can help to decide the best operation mode to apply (Rohani et al., 2016).
452 From the design point of view, this result could also be used to optimize the TES size by
453 defining operating strategies and control techniques adapted to the market and tariff conditions
454 (Casati et al., 2015) and the typical distribution of day types.

455

456 5. Conclusions

457 The clear sky yield index k_p is introduced. k_p is defined as the quotient of the energy generated
458 by a solar plant to the energy that the same plant would generate under clear sky conditions
459 during a certain period. Thus, a solar plant model and a clear sky model are required to calculate
460 k_p . In this work, the concept of clear sky yield index is applied to CSP plants, using the daily
461 gross electricity generated. For the assessment of the features and possible applications of k_p ,
462 two parabolic trough CSP plants models (PTP0, without TES, and PTP7.5, with 7.5 full-load
463 hour capacity TES), a clear sky model and 16 years of DNI measurements recorded in Seville
464 (Spain) have been used.

465 The analysis of the results shows that: 1) The daily k_p values for both PTP0 and PTP7.5 take
466 higher values in summer and lower in winter, with greater variability during spring and autumn;
467 2) the values of k_p are normally –but not always- higher for PTP7.5; 3) the frequency
468 distributions of k_p are bimodal with the maximum placed at low k_p values for PTP0 and at high
469 k_p values for PTP7.5. All these features are indicative of the advantages in the utilization of the
470 solar resource provided by the high capacity TES of PTP7.5.

471 The relation of k_p with other indexes proposed for the characterization of the solar resource
472 features (transmittance index k_b , variability index VI , morning factor F_m) has also been
473 investigated. A good correlation between k_b and k_p has been found, with well-defined,
474 increasing upper and lower limits, though the spread indicates that other parameters are required
475 to relate both indexes. If the shapes of the scatterplot of PTP0 and PTP7.5 are compared, a
476 higher slope of the upper limit is identified in the case of PTP7.5, showing k_p values close to 1

477 for days with k_b values lower than 0.6 and increasing the number of days that reach this value of
1 478 k_p . In the case of the VI, a range of k_p values from 0 to 1 is observed for lower VI values with
2 more points close to 1. The upper limit of this range decreases when VI increases. So, the
3 479 scatterplot is limited by a decreasing linear boundary showing that higher variability involves
4 480 lower k_p . This behaviour suggests, again, that other parameters are needed to relate both
5 481 indexes. The influence of variability is less relevant in the case of PTP7.5. The performance of a
6 482 PTP with TES can double the k_p value with respect to a plant without TES in days with high
7 483 energy and high variability. This advantage of TES is not so significant in days with low energy
8 484 and high variability. Additional sets of 16 years of DNI synthetic data with symmetric profiles
9 485 with respect to the solar noon and the corresponding daily k_p values have been elaborated to
10 486 investigate the relation between F_m and k_p . The higher values are always found for days with a
11 487 well-balanced distribution of the solar resource during the day –which are days with high
12 488 energy, too. The range where these higher values are found is wider for PTP7.5; reaching values
13 489 of k_p equal to 0.9 for values of F_m of 0.7 while in PTP0 k_p 0.7 is not exceeded for the same value
14 490 of F_m . The comparison of the k_p values for the days considered as *morning days* (F_m higher or
15 491 equal to 0.7) and as *afternoon days* (F_m lower or equal to 0.3) are compared for both plants. In
16 492 the case of PTP0, *afternoon days* have, on average, higher k_p values than *morning days*. For
17 493 PTP7.5 the behaviour is not well defined, suggesting that no significant differences can be
18 494 attributed to this factor.
19 495

20 496 As an application of the clear sky yield index, the relation between the day types, classified by
21 497 means of an unsupervised clustering method based on the variability, energy and time
22 498 distribution of the DNI, and the electricity yield, is characterized by means of the clear sky yield
23 499 index, k_p . All day types, except for days with medium energy, high variability and balanced
24 500 temporal distribution of the solar resource (BM type), have characteristic frequency
25 501 distributions of k_p . The k_p values are statistically higher for the BH –high energy, balanced time
26 502 distribution and low variability- day type. The k_p values for days with the energy concentrated
27 503 mainly in the morning or in the afternoon have similar distributions if the variability is low;
28 504 otherwise, the distributions are clearly different and k_p reaches higher values when the energy is
29 505 concentrated in the morning rather than in the afternoon. The already commented results
30 506 obtained when the k_p values for morning and afternoon days are compared for the measured set
31 507 and the symmetric set suggest that the morning days from our location have better features than
32 508 afternoon days from the standpoint of the PTP performance. The days with low or negligible
33 509 energy (BL, AL, ML and Null) show very similar frequency distribution, thus suggesting the
34 510 possibility to merge all these day types (with the possible exception of ML for the location of
35 511 this study) in one only type. Therefore, the 10 day types identified in a previous work can be
36 512 reduced to 6 (for PTP7.5) or 7 (for PTP0). So, with the characterization of the 10 day-types
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

513 from the production of a PTP without TES viewpoint the number of types can be reduced to 7
514 and from the production of a PTP with TES to 6.

515 The relation between day types and clear sky yield index has a direct application to the
516 definition of operating strategies and forecasting of the energy generation based on the
517 qualitative prediction of the type of day. Another potential application is the optimization of the
518 configuration of CSP plants during the design stage based not only on the solar resource
519 magnitude, but also on the knowledge of the typical distribution of day types for the location.

520 In future works, the comparison of different CSP technologies and different locations based on
521 the clear sky yield index will be addressed.

522

523 REFERENCES

524 Casati, E., Casella, F., Colonna, P., 2015. Design of CSP plants with optimally operated thermal
525 storage. *Sol. Energy* 116, 371–387. doi:10.1016/j.solener.2015.03.048

526 Gastón-Romeo, M., Leon, T., Mallor, F., Ramírez-Santigosa, L., 2011. A Morphological
527 Clustering Method for daily solar radiation curves. *Sol. Energy* 85, 1824–1836.
528 doi:10.1016/j.solener.2011.04.023

529 Jiménez-Pérez, P.F., Mora-López, L., 2016. Modeling and forecasting hourly global solar
530 radiation using clustering and classification techniques. *Sol. Energy* 135, 682–691.
531 doi:10.1016/j.solener.2016.06.039

532 Kang, B.O., Tam, K.-S., 2015. New and improved methods to estimate day-ahead quantity and
533 quality of solar irradiance. *Appl. Energy* 137, 240–249.
534 doi:10.1016/j.apenergy.2014.10.021

535 Kang, B.O., Tam, K.S., 2013. A new characterization and classification method for daily sky
536 conditions based on ground-based solar irradiance measurement data. *Sol. Energy* 94,
537 102–118. doi:10.1016/j.solener.2013.04.007

538 Liu, B.Y.H., Jordan, R.C., 1960. The interrelationship and characteristic distribution of direct,
539 diffuse and total solar radiation. *Sol. Energy* 4, 1–19.

540 MathWorks, 2016. kstest2 [WWW Document]. URL
541 <https://es.mathworks.com/help/stats/kstest2.html> (accessed March 2017)

542 Meybodi, M.A., Ramírez-Santigosa, L., Beath, A.C., 2017. A Study on the Impact of Time
543 Resolution in Solar Data on the Performance CSP Plants. *Renew. Energy* 109, 551–563.
544 doi:10.1016/j.renene.2017.03.024

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

545 Michael J., W., Gilman, P., 2011. Technical manual for the SAM physical trough model. NREL
546 Tech. Rep. 303, 275–3000.

547 Moreno-Tejera, S., Ramírez-Santigosa, L., Silva-Pérez, M.A., 2015. A proposed methodology
548 for quick assessment of timestamp and quality control results of solar radiation data.
549 Renew. Energy 78, 531–537. doi:10.1016/j.renene.2015.01.031

550 Moreno-Tejera, S., Silva-Pérez, M.A., Lillo-Bravo, I., Ramírez-Santigosa, L., 2016. Solar
551 resource assessment in Seville, Spain. Statistical characterisation of solar radiation at
552 different time resolutions. Sol. Energy 132, 430–441. doi:10.1016/j.solener.2016.03.032

553 Moreno-Tejera, S., Silva-Pérez, M.A., Ramírez-Santigosa, L., Lillo-Bravo, I., 2017.
554 Classification of days according to DNI profiles using clustering techniques. Sol. Energy
555 146, 319–333. doi:10.1016/j.solener.2017.02.031

556 Muselli, M., Poggi, P., Notton, G., Louche, A., 2000. Classification of typical meteorological
557 days from global irradiation records and comparison between two Mediterranean coastal
558 sites in Corsica Island. Energy Convers. Manag. 41, 1043–1063.

559 NREL, 2017. Concentrating Solar Power Projects. Solnova-1 [WWW Document]. URL
560 https://www.nrel.gov/csp/solarpaces/project_detail.cfm/projectID=21 (accessed January
561 2017)

562 NREL, 2013. Concentrating Solar Power Projects. Andasol-3 [WWW Document]. URL
563 https://www.nrel.gov/csp/solarpaces/project_detail.cfm/projectID=117 (accessed January
564 2017)

565 NREL, 2010. System Advisor Model (SAM) [WWW Document]. URL <https://sam.nrel.gov/>
566 (accessed January 2017)

567 Polo, J., Téllez, F.M., Tapia, C., 2016. Comparative analysis of long-term solar resource and
568 CSP production for bankability. Renew. Energy 90, 38–45.
569 doi:10.1016/j.renene.2015.12.057

570 Rohani, S., Fluri, T., Dinter, F., 2016. Modelling , Simulation and Data Validation of a Solar
571 Thermal Parabolic Trough Plant with Storage. Conf. Proc. SolarPACES 2015.

572 Skartveit, A., Olseth, J.A., 1992. The probability density and autocorrelation of short-term
573 global and beam irradiance. Sol. Energy 49, 477–487.

574 Soubdhan, T., Emilion, R., Calif, R., 2009. Classification of daily solar radiation distributions
575 using a mixture of Dirichlet distributions. Sol. Energy 83, 1056–1063.
576 doi:10.1016/j.solener.2009.01.010

577 Stein, J., Hansen, C., Reno, M., 2012. The variability index: A new and novel metric for
578 quantifying irradiance and PV output variability. *World Renew. Energy* 1–7.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65



ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA
Universidad de Sevilla
Camino de los Descubrimientos,s/n
41092 SEVILLA



Dr. Jan Kleissl
University of California at San Diego (UCSD),
La Jolla, California,
USA

Dear Dr. Kleissl,

It is my pleasure to enclose for your consideration and possible publication in this Journal the article:

“The clear sky yield index: a new indicator for qualifying the solar resource for CSP plants”

Co-authored by Sara Moreno-Tejera, Manuel A. Silva-Pérez, Lourdes M. Ramirez-Santigosa and Isidoro Lillo-Bravo. I'm a Ph.D. student at the Department of Energy Engineering, University of Seville, where I'm developing my Thesis on the topic of solar resource assessment for Concentrating Solar Power (CSP) plants with the supervision of Dr. Silva-Pérez. Characterization the solar resource on site is a basic issue for a feasibility analysis of any solar technology. The majority of the methodologies used to characterize the long-term behaviour of the solar resource for CSP plants, only consider the daily energy of the Direct Normal Irradiation (DNI). However, the production of a CSP plant is not directly proportional to the daily energy radiation received; it depends also on its variability and temporal distribution. A methodology to classify the days according to their daily DNI, intraday variability and temporal distribution was proposed in a previous work, published in Solar Energy Journal. The objective of this work is to explore the relation between the daily DNI profiles –classified in the previous study- and the electricity yield of Parabolic Trough (PT) CSP plants with and without Thermal Energy Storage (TES). This relation could help to better characterize the solar resource of a location from a CSP plant point of view. A novel index, the clear sky yield index or k_p is introduced to assess the solar resource features from the perspective of the operation of a CSP plant and to compare the exploitation of the solar resource by CSP plants with different configurations in the same location.

The research significance of this paper is the characterization of each type of day from the standpoint of the production of a PT plant with and without TES, revealing that some day types can be grouped. The relation between day types and clear sky yield index has a direct application to the definition of operating strategies and forecasting of the energy generation based on the qualitative prediction of the type of day.

Best regards,

Sara Moreno Tejera