

Daylight modeling for energy efficiency and visual comfort in buildings

PhD Dissertation

Diego Granados López

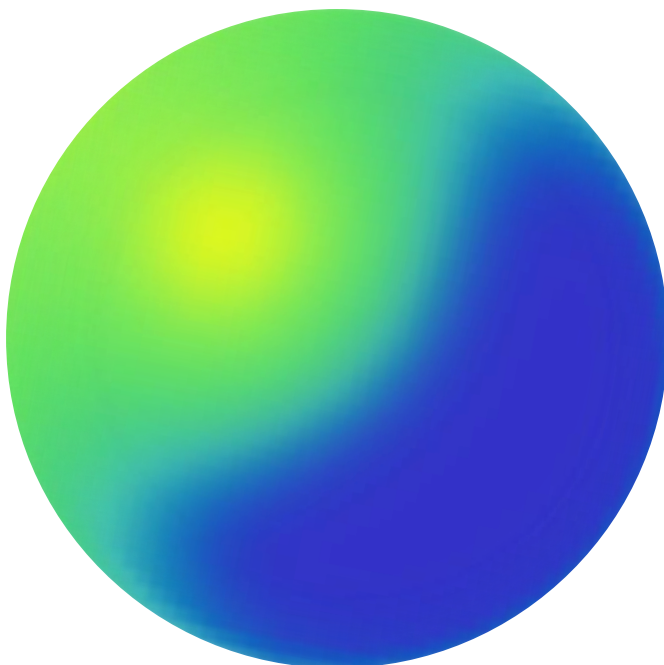
Program:

Energy Efficiency and
Sustainability in Engineering
and Architecture

Supervisors:

Cristina Alonso Tristán

Montserrat Díez Mediavilla



UNIVERSIDAD
DE BURGOS

Universidad de Burgos

**Modelado de la iluminación natural
para la eficiencia energética y el
confort visual en edificios**

Tesis doctoral

Diego Granados López

Programa de doctorado:
Eficiencia Energética y
Sostenibilidad en Ingeniería
y Arquitectura

Dirigida por:

Cristina Alonso Tristán y
Montserrat Díez Mediavilla

Burgos, 2022

Diego Granados López, 2022



A mis seres queridos.

*No se puede negar que a veces los sueños
significan algo.*

San Alberto Magno

Preface

This doctoral dissertation is presented as compendium of publications, and fully complies with the requirements established by the Doctoral School of the University of Burgos and the Doctoral Program “Energy Efficiency and Sustainability in Engineering and Architecture”, Interuniversity Doctoral Program of the Universities of Burgos, Vigo and Basque Country/Euskal Herriko Unibertsitatea (RD 99/2011).

The publications demonstrate the results obtained in the different research works developed to achieve the objectives of the thesis. The referenced publications are listed in order of relation to the objectives in the Section: List of Papers, and constitute Sections I to VI of this thesis dissertation. Two complementary papers, directly related with objectives and in which the author of this work has collaborate, have been included in the List of Papers. All articles have been published in journals indexed in the Journal of Citation Reports (JCR), in the first and second quartiles of the Energy&Fuels, Mathematics, Multidisciplinary Applications or Engineering, Multidisciplinary categories

This doctoral thesis has been supported thanks to the funding of the PROGRAMA DE FORMACIÓN DE PROFESORADO UNIVERSITARIO (PIRTU ORDEN EDU/556/2019), the Mobility Grant for Doctoral Students Stays of the University of Burgos, Program (2021), and the following competitive funding research projects:

1. Análisis Espectral de la Radiación Solar: Aplicaciones Climáticas, Energéticas y Biológicas (RTI-2018-098900-B-I00). Ministerio de Universidades e Investigación Programa Estatal De I+D+i Orientada a los Retos de la Sociedad. IP: Cristina Alonso Tristán y Montserrat Díez Mediavilla. 1/01/2019-30/09/2022.
2. Valoración técnica de los niveles de exposición a radiación solar en trabajos de exterior: identificación de grupos de riesgo y medidas de prevención. (INVESTUN/19/BU/004) Junta de Castilla y León.

Dirección General de Trabajo y Prevención de riesgos laborales. IP: Montserrat Díez Mediavilla. 01/01/2019-30/09/2021.

3. Metodología para la rehabilitación energética de edificios de uso público en Castilla y León mediante integración fotovoltaica (BU021G19). Junta de castilla y León. Programa de Apoyo a los Grupos de Investigación Reconocidos de Universidades públicas de Castilla y León. 01/01/2019-31/12/2021. IP: Montserrat Díez Mediavilla.
4. Medida y modelización de la iluminación solar para la optimización de técnicas de iluminación natural en la edificación (ENE2014-54601-R), Ministerio de Economía y Competitividad. RETOS DE LA SOCIEDAD. IP: Montserrat Díez Mediavilla. 01/12/2015-31/12/2018.

Acknowledgments

Por ser unos pilares fundamentales a lo largo de estos años, por su gran e indispensable apoyo y confianza, me gustaría agradecer de manera especial y sincera a mis tutoras Montserrat Díez Mediavilla y Cristina Alonso Tristán.

Quiero expresar también mi más profundo agradecimiento al grupo G, y a todos mis compañeros del grupo SWIFT, ya que sin su continuo apoyo el camino recorrido no sería el mismo.

Me gustaría agradecer a mi amigo Carlos, gracias a sus consejos e inestimable ayuda.

Sin duda, sobre todo, el agradecimiento más profundo va para mi familia y personas queridas, ya que, sin su constante amor incondicional, apoyo y paciencia me habría sido imposible realizar esta tesis doctoral.

Universidad De Burgos - Escuela De Doctorado


Cristina Alonso Tristán y Monserrat Díez Mediavilla hacen constar:

Que el presente trabajo, titulado “**Daylight modeling for energy efficiency and visual comfort in buildings**”, que presenta Diego Granados López para la obtención del título de doctor ha sido realizado bajo su supervisión. en el Programa de Doctorado en Eficiencia Energética y Sostenibilidad en Ingeniería y Arquitectura por la Universidad de Burgos, la Universidad de Vigo y la Universidad del País Vasco/Euskal Herriko Unibertsitatea (5600956; RD 99/2011).

Burgos, 4 de abril de 2022

A handwritten signature in blue ink, appearing to read 'Cristina Alonso', enclosed within a large, loopy circular flourish.

Cristina Alonso Tristán

A handwritten signature in blue ink, appearing to read 'M. D. M.', with a long horizontal line extending to the left from the end of the signature.

Montserrat Díez Mediavilla

Abstract

Searching and selecting an adequate methodology for daylight modeling is essential in the design of energy efficient buildings that guarantee the visual, physical and psychological comfort of their occupants. The first step in determining the indoor building illuminance lies in knowing the outdoor illuminance. This dissertation addresses this key aspect through different strategies such as luminous efficacy models and the determination of the angular distribution of the sky's luminance.

Daylight is strongly determined by sky conditions. The CIE/ISO standard provides a good general framework to represent the real conditions of the sky, covering the entire probable spectrum of skies, and has been used as a reference throughout this work. The characterization of the skies according to the CIE standard requires experimental measurements of the luminance distribution of the sky, scarcely recorded in terrestrial meteorological facilities. The thesis proposes, as alternatives for the classification of skies according to the CIE taxonomy, the use of meteorological indices, sky images and algorithms based on artificial intelligence. The structure and efficiency of the machine learning algorithms used, both neural networks and decision trees, have been optimized through feature selection procedures in the case of the use of meteorological indices and through image pre-processing techniques, as a step prior to using the classification algorithm. The thesis has also developed a new locally calibrated luminous efficacy model, with excellent results both when used for all-sky types and for clear, overcast and partially overcast sky conditions.

Keywords: CIE standard sky classification, illuminance, daylight, meteorological indices, machine learning

Resumen

La búsqueda y elección de una metodología adecuada para el modelado de la iluminación natural es fundamental en el diseño de edificios energéticamente eficientes y que garanticen el confort visual, físico y psicológico de sus ocupantes. El primer paso para la determinación de la iluminación en el interior de un edificio reside en el conocimiento de la iluminación exterior. La tesis doctoral aborda este aspecto fundamental a través de diferentes estrategias como son los modelos de eficacia luminosa y la determinación de la distribución angular de la luminancia del cielo.

La iluminación natural está fuertemente determinada por las condiciones de cielo. El estándar CIE/ISO proporciona un buen marco general para representar las condiciones reales del cielo cubriendo todo el espectro probable de cielos, por lo que se ha seleccionado como referencia a lo largo de este trabajo. La caracterización de los cielos según el estándar CIE requiere de medidas experimentales de la distribución de luminancia del cielo, escasamente registradas en las instalaciones meteorológicas terrestres. La tesis propone como alternativas para la clasificación de cielos según la taxonomía CIE, la utilización de índices meteorológicos, imágenes del cielo y algoritmos basados en inteligencia artificial. La estructura y la eficacia de los algoritmos de aprendizaje automático empleados, redes neuronales y árboles de decisión, se han optimizado mediante procedimientos de selección de variables en el caso de la utilización de índices meteorológicos y mediante técnicas de pre-procesamiento de imágenes, como paso previo a la utilización del algoritmo de clasificación. La tesis ha desarrollado también un nuevo modelo de eficacia luminosa, calibrado localmente, con excelentes resultados tanto al utilizarlo para todos los tipos de cielo como para condiciones de cielo claro, cubierto y parcialmente cubierto.

Palabras clave: iluminancia, iluminación natural, clasificación de cielos estándar CIE, inteligencia artificial, índices meteorológicos.

List of Papers

Paper I

Suárez-García, A., Granados-López, D., González-Peña, D., Díez-Mediavilla, M. and Alonso-Tristán, C., 2018. Seasonal characterization of CIE standard sky types above Burgos, northwestern Spain. *Sol. Energy* 169, 24–33. DOI: 10.1016/j.solener.2019.11.060

Citations: March (2022)

Citations Google Scholar	Citations Scopus	Field-Weighted Citation Impact
11	6	0.72

Rank by Journal Impact Factor: 4.674 (2018)

JCR Category	Rank in Category	Quartile in Category
Energy & Fuels	24/103	Q1

SCImago Journal Rank (SJR): 1.593 (2018)

Category	Rank in Category	Quartile in Category
Renewable Energy, Sustainability and the Environment	27/580	Q1

Paper II

Granados-López, D., Díez-Mediavilla, M., Dieste-Velasco, M.I., Suárez-García, A. and Alonso-Tristán, C, 2020. Evaluation of the vertical sky component without obstructions for daylighting in Burgos, Spain. Appl. Sci. 10 (9), 3095. DOI:10.3390/app10093095

Citations: March (2022)

Citations Google Scholar	Citations Scopus	Field-Weighted Citation Impact
3	2	0.4

Rank by Journal Impact Factor: 2.679 (2020)

JCR Category	Rank in Category	Quartile in Category
Engineering, Multidisciplinary	38/90	Q2

Rank by Journal SCImago Journal Rank (SJR): 0.435 (2020)

Category	Rank in Category	Quartile in Category
Instrumentation	67/922	Q2

Paper III

Granados-López, D., Suárez-García, A., Díez-Mediavilla, M. and Alonso-Tristán, C., 2021. Feature selection for CIE standard sky classification. Sol. Energy 218, 95-107. DOI: 10.1016/j.solener.2019.11.060

Citations: March (2022)

Citations Google Scholar	Citations Scopus	Field-Weighted Citation Impact
2	2	0.9

Rank by Journal Impact Factor: 5.742 (2020)

JCR Category	Rank in Category	Quartile in Category
Energy&Fuels	38/114	Q2

SCImago Journal Rank (SJR): 1.337 (2020)

Category	Rank in Category	Quartile in Category
Renewable Energy, Sustainability and the Environment	36/486	Q1

Paper IV

Granados-López, D., García-Rodríguez, A., García-Rodríguez, S., Suárez-García, A., Díez-Mediavilla, M., Alonso-Tristán, C., 2021. Pixel-based image processing for CIE standard sky classification through ANN. Complexity 2021, 1-15. DOI: <https://doi.org/10.1155/2021/2636157>

Citations: March (2022)

Citations Google Scholar	Citations Scopus	Field-Weighted Citation Impact
0	0	0

Rank by Journal Impact Factor: 2.833 (2020)

JCR Category	Rank in Category	Quartile in Category
Mathematics, Interdisciplinary Applications	31/108	Q2

SCImago Journal Rank (SJR): 0.447 (2020)

Category	Rank in Category	Quartile in Category
Multidisciplinary	27/135	Q1

Complementary Paper - CP I

Dieste-Velasco, M.I., Díez-Mediavilla, M., Granados-López, D., González-Peña, D., Alonso-Tristán, C., 2019. Performance of global luminous efficacy models and proposal of a new model for daylighting in Burgos, Spain. *Renewable Energy* 133, 1000-1010. DOI: 10.1016/j.renene.2018.10.085

Citations: March (2022)

Citations Google Scholar	Citations Scopus	Field-Weighted Citation Impact
9	7	0.41

Rank by Journal Impact Factor: 6.274 (2019)

JCR Category	Rank in Category	Quartile in Category
Energy & Fuels	19/112	Q1

SCImago Journal Rank (SJR): 2.052 (2019)

Category	Rank in Category	Quartile in Category
Renewable Energy, Sustainability and the Environment	19/494	Q1

Complementary Paper - CP II

Suárez-García, A., Díez-Mediavilla, M., Granados-López, D., González-Peña, D., Alonso-Tristán, C. 2020. Benchmarking of meteorological indices for sky cloudiness classification. *Sol. Energy* 195, 499–513. DOI: 10.1016/j.solener.2019.11.060

Citations: March (2022)

Citations Google Scholar	Citations Scopus	Field-Weighted Citation Impact
11	6	0.72

Rank by Journal Impact Factor: 5.742 (2020)

JCR Category	Rank in Category	Quartile in Category
Energy & Fuels	38/114	Q2

SCImago Journal Rank (SJR): 1.337 (2020)

Category	Rank in Category	Quartile in Category
Renewable Energy, Sustainability and the Environment	36/486	Q1

Contents

Preface	III
Acknowledgments.....	V
Abstract.....	VII
Resumen	VIII
List of Papers.....	IX
Contents.....	XV
List of Figures.....	XVI
List of Tables	XVIII
Nomenclature	XIX
List of abbreviations.....	XXII
Chapter 1: Introduction and objectives	1
1.1 Introduction	1
1.2 General and partial objectives	5
1.3 Experimental Data	9
1.4 Summary of papers	11
Chapter 2: Results by objectives	13
2.1 Obtaining the CIE standard sky classification using different strategies: Seasonal characterization of CIE standard sky types (Paper I).....	13
2.2. Development of models for global and diffuse illuminance (I): Vertical Sky Component (VSC) for Daylight availability (Paper II)	23
2.3. Development of models for global and diffuse illuminance (II): Luminous efficacy model (CP I).....	30
2.4. Establishment of alternative methods to sky scanner for CIE standard sky classification from the use of meteorological indices (MIs) and sky images (I): Benchmarking of Meteorological Indices (CP II)	35
2.5. Establishment of alternative methods to sky scanner for CIE standard sky classification from the use of meteorological indices (MIs) and sky images (II): Feature Selection for CIE standard sky classification (Paper III).....	42
2.6. Establishment of alternative methods to sky scanner for CIE standard sky classification from the use of meteorological indices (MIs) and sky images (III): Pixel-Based Image Processing for CIE Standard Sky Classification through ANN	51
Chapter 3: Conclusions and further work	65
Chapter 4: Bibliography.....	75

List of Figures

<i>Figure 1. Spectral distribution of the SR and main phenomena related to its energy. (Source: own elaboration).</i>	1
<i>Figure 2. Relation of the General and partial objectives with the publications.</i>	7
<i>Figure 3. Radiometric variables recorded at the SWIFT experimental facility.</i>	10
<i>Figure 4. Angles defining the position of the sun and a sky element p [43].</i>	14
<i>Figure 5. Relative gradation and indicatrix function for the CIE sky-classification[40].</i>	15
<i>Figure 6. Comparison of the modeled CIE luminance distribution with the real one obtained by the sky-scanner. (Source: own elaboration).</i>	16
<i>Figure 7. Comparative characterizations of the skies over Burgos [44].</i>	19
<i>Figure 8. Seasonal CIE sky types histogram over Burgos calculated using the NR method [44].</i>	20
<i>Figure 9. Seasonal CIE sky types histogram over Burgos calculated using the Tregenza Method [44].</i>	21
<i>Figure 10. Confusion matrix comparing the Tregenza and the NR cloudiness classifications [44].</i>	22
<i>Figure 11. Geometry description: Angles of the Sun and the sky element p.</i>	25
<i>Figure 12. Sky patch's projection [51].</i>	25
<i>Figure 13. VSC dependence with the scattering angle for all CIE skies. The labels one (overcast) to fifteen (clear) design a specific CIE sky typology [51].</i>	27
<i>Figure 14. Distribution of VSC values by intervals calculated in Burgos, Spain, between September 2016 and January 2019 [51].</i>	29
<i>Figure 15. RMSE variation (%): Best score of the reviewed model (underlined model) versus the proposed model in each category, described in Table 8.</i>	34
<i>Figure 16. Interval adaptation procedure of the original MIs [69].</i>	36
<i>Figure 17. Multiclass to dichotomous transformation [69].</i>	39
<i>Figure 18. Workflow of Supervised Machine-Learning tool [27].</i>	43
<i>Figure 19. CIE standard sky classification tree (MIs selected with the Pearson FS method) [27].</i>	47
<i>Figure 20. CIE standard sky classification tree (MIs selected with the PI FS method) [27].</i>	47
<i>Figure 21. CIE standard sky classification tree (MIs selected with the RFE FS method) [28].</i>	47
<i>Figure 22. CIE standard sky classification tree. (MIs selected with the Boruta FS method) [27].</i>	47

<i>Figure 23. Precision, Recall, and f1 indices calculated for Clear-Partial-Overcast classification each classification tree versus CIE standard.</i>	49
<i>Figure 24. Results of the image-processing methods applied to clear, partial, and overcast CIE standard sky types [28].</i>	53
<i>Figure 25. Workflow for image processing. Example with RGB space [28].</i>	54
<i>Figure 26. ANN system architecture [28].</i>	54
<i>Figure 27. Improvement in ANN accuracy, $\Delta(A)$, for CIE standard sky classification, using the results of each image-processing method as input, over ANN accuracy obtained with the original RGB images as input [28].</i>	57
<i>Figure 28. Improvement in ANN accuracy, $\Delta(A)$, for CIE standard sky classification in three sky categories: overcast, partial, and clear conditions, using as input the image processed by each image-processing method, over ANN accuracy obtained with the RGB images as input [28].</i>	58
<i>Figure 29. Confusion matrix for ANN-calculated RGB-CIE sky classification into fifteen categories and into three categories: Overcast (CIE standard sky types 1 to 5); Partial (CIE standard sky types 6 to 10); Clear (CIE standard sky types 1 to 5). Critical refers to cases classified out of category. The color scale shows the number of coincidences in each category [28].</i>	60
<i>Figure 30. Confusion matrices color space image processing methods-CIE standard sky classification [28].</i>	61
<i>Figure 31. Confusion matrices spectral feature image processing methods-CIE standard sky classification [28].</i>	62
<i>Figure 32. Confusion matrices for texture filter image processing methods-CIE standard sky classification [28].</i>	62
<i>Figure 33. Accuracy for the CIE standard sky classification through sky images and ANN, using the different image-processing methods recorded in Table 15 [28].</i>	64
<i>Figure 34. FOC of the illuminance level higher than 1000 lux and 300 lux, respectively, in the measurement points.</i>	69
<i>Figure 35. Left side: ASE annual hours distribution on the top floor, right side: picture of the housing under renovation.</i>	71
<i>Figure 36. VSC dependence on solar altitude.</i>	72
<i>Figure 37. Confusion matrices of the Machine learning predictions. Classification tree (left), and ANN (right), for the CIE sky classification into five categories.</i>	72

List of Tables

<i>Table 1. Radiometric variables recorded at the SWIFT experimental facility. The table includes each sensor model, and the commissioning date. Table includes reference of the Paper in which the recorded data were used.</i>	10
<i>Table 2. Parameters of CIE standard Sky types [40].</i>	15
<i>Table 3. Statistical indicators RMSE and MBE [40].</i>	22
<i>Table 4. rRMSE and rMBE calculated for the vertical diffuse illuminance, L_{dv}, and the horizontal diffuse illuminance, L_{dH}. The reference values of L_{dV} were calculated from (Eq. 10). The reference values of L_{dH} were measured in the experimental facility, described in Table 1 [51].</i>	28
<i>Table 5. rRMSE and rMBE results from the comparison between the VSC values calculated with the different approaches [51].</i>	28
<i>Table 6. Summary of the global luminous efficacy models reviewed in this work. Literature reference of the original model, year, sky type classification, input parameters used in the models and the original place of development of the model[9].</i>	31
<i>Table 7. List of coefficients of the proposed model.</i>	32
<i>Table 8. Summary of the results obtained with the testing data for the reviewed and proposed models. The reviewed model with the lowest RMSE has been underlined.</i>	33
<i>Table 9. Summary of the MIs used to classify the skies, the original intervals used to define the clear, partial and cloudy sky conditions and the adapted ones used in this work [69].</i>	37
<i>Table 10. Brief description of the goodness indicators.</i>	39
<i>Table 11. Ranking average summary [69].</i>	41
<i>Table 12. Definition of the 43 MIs reviewed as candidates for sky classification. L_0 is the Luminous solar constant (133.8 klx) and I_{sc} is the solar constant (1361.1 W/m^2) [27].</i>	44
<i>Table 13. Summary of several FS by category. (UR: Universal results).</i>	45
<i>Table 14. Summary of the MIs selected by each one of the FS algorithms [27].</i>	46
<i>Table 15. Summary of pixel image-processing methods [28].</i>	52
<i>Table 16. Accuracy and structure (number of neurons in the hidden layer) of the selected ANN for each image-processing method [28].</i>	57
<i>Table 17. Accuracy and structure (number of neurons in the hidden layer) of the ANN selected for each image-processing method [28].</i>	59
<i>Table 18. Summary of the results [28].</i>	63

Nomenclature

L_p	Absolute luminance of the sky point
A	Accuracy
$f(X_i^n)$	Activation function
ρ	Albedo
W	Atmospheric precipitable water
ϕ_p	Azimuth angle of the sky point
ϕ_s	Azimuth of the sun
ϕ_{N_r}	Azimuth vertical surface angle
k_k	Battles Clearness index
B	Bias
k_c	Clear sky index
F	Clearness function
k_t	Clearness index
CC	Cloud Cover
Cle	Cloudless index
κ	Cohen Kappa
k_d	Diffuse fraction
k_b	Direct fraction
Ra_{BN}	Direct normal irradiance
α_p	Elevation angle of the sky point
p_e	Expected agreement by chance
f_1	F1 factor
D_p	Geometrical factor
φ	Gradation function
Lx_{BH}	Horizontal beam illuminance
Lx_{DH}	Horizontal diffuse illuminance
Ra_{DH}	Horizontal diffuse irradiance
Ra_{BH}	Horizontal direct irradiance
Lx_{GH}	Horizontal global illuminance
Ra_{GH}	Horizontal global irradiance

S_i	Igawa sky index
S_i	Igawa's sky index
C_V	Illuminance cloud ratio
$\Delta(A)$	Improvement in ANN accuracy
$f(\chi)$	Indicatrix function
X_i^n	Input of the activation function
X^{n-1}	Input variables
C_e	Irradiance cloud ratio
JI	Jaccard index
FK	Klucher's clearness index
NR	Littlefair normalization ratio
K	Luminous efficacy
L_0	Luminous solar constant
t_v	Luminous Turbidity index
M	Matthews
MBE	Mean Bias Error
P_e	Modelled direct solar irradiance
X_i^0	Neurons of the input layer
$X_i^{2'}$	Neurons of the output layer
$EVSM$	Normalized beam illuminance
$EVDM$	Normalized diffuse illuminance
$EVGM$	Normalized global illuminance
M_v	Optical mass
OFP	Original Perraudau's Nebulosity index
$X_i^{n'}$	Output of the activation function
p_0	Overall agreement
Δ	Perez sky's brightness
ϵ	Perez sky's clearness
$PERV$	Permeability
FP	Perraudau's Nebulosity index
Pr	Precision
P	Pressure

C_2	Ratio diffuse illuminance
C_1	Ratio global illuminance
$LERT$	Ratio zenith illuminance to horizontal diffuse illuminance
Re	Recall
Ω	Relative Heaviness
RH	Relative Humidity
l_p	Relative luminance of a sky point
$rMBE$	Relative Mean Bias Error
$rRMSE$	Relative Root Mean Square Error
$RMSE$	Root Mean Square Error
χ	Scattering angle
$L_{pred,p,sc}$	Sky patch luminance normalized by NR
$L_{pr,sc}$	Sky patch luminance normalized by Tregenza
α_s	Solar altitude
I_{sc}	Solar constant
Ces	Standard Cloud Ratio
T	Temperature
N	Total number of cases
$TURV$	Turbidity
$CLDV$	Umemiya's Cloud Ratio
$LxDV$	Vertical diffuse illuminance
$LxGV$	Vertical global illuminance
$RaGVS$	Vertical global irradiance facing South
VSC	Vertical Sky Component
W_n	Weighting matrix
WD	Wind direction
WS	Wind Speed
k_{t2}	Zenith Angle Independent Clearness index
Z_p	Zenith angle of a sky element
Z_s	Zenith angle of the sun
L_z	Zenith luminance

List of abbreviations

ANN	Artificial Neural Networks
ARGD	Adjusted Red Green Difference
B	Blue channel
BIPV	Building Integrated Photovoltaics
BOR	Boruta (FS)
C1	Combination of RBR, RBD and NRBR
C2	Combination of RBR, ARGD and NRBR
CIE	International Commission on Illumination
CP	Complementary paper
CPO	Clear-partial-overcast
D	Dark channel
EM	Entropy Matrix
EU	European Union
FM	Filter Method
FN	False Negative
FOC	Frequency of occurrence
FOV	Field of view
FP	False Positive
FS	Feature Selection
G	Green channel
GS	Gray scale color space
H	Hue channel
HDR	High Dynamic Range
HSV	Hue Saturation Value (color space)
IEA	International Energy Agency
IR	Infrared Radiation
L	Bright channel
LR	Local Range
MI	Meteorological Indice
NIR	Near Infrared Radiation

NR	Littlefair normalization ratio
NRBR	Normalized Red-Blue Ratio
P	Pearson (FS)
PAR	Photosynthetically Active Radiation
PePSS	Photomultiplier-equipped portable sky scanner
PI	Permutation importance (FS)
PT	Photothermal
PV	Photovoltaic
R	Red channel
RAS	Removal Atmospheric Scattering
RBD	Red-Blue Difference
RBR	Red Blue Ratio
RFE	Recursive feature elimination (FS)
RGB	Red Green Blue (color space)
S	Saturation channel
SCG	Scaled Conjugate Gradient method
SML	Supervised machine learning
SR	Solar Radiation
SSLD	Standard Sky Luminance Distribution
STD	Local Standard Deviation
SWIFT	Solar and Wind Feasibility Technologies
TN	True Negative
TP	True Positive
UR	Universal results
UVR	Ultraviolet Radiation
UV-A	Ultraviolet Radiation - A
UV-B	Ultraviolet Radiation - B
V	Value channel
VIS	Visible light
WM	Wrapper Method
Y	Panchromatic channel

Chapter 1: Introduction and objectives

1.1. Introduction

The exploitation of as many natural energy sources as possible has never been so relevant as nowadays. Climate change is getting more impact every year on ecosystems and biodiversity. Being aware of the dimensions of the problem, several international entities like the European Union (EU) through the European Green Deal show their ambition to develop a climate adaptation as a new international mission [1]. Their agenda aims to protect living beings, the Earth planet, and its sustainability. The European Parliament established an action plan in which the implementation of renewable sources constitutes important necessary measures, as it reduces the greenhouse gas emissions and the Union's energy dependency [2].

Solar radiation (SR) is the main driver of the planetary energy balance, and it is essential in various areas of human activity (Figure 1). Due to the filtering effect of the atmosphere, the solar radiation that reaches the Earth's surface is composed largely for the frequencies within the infrared (IR) and the visible light (VIS) that constitute respectively 45% and 50% of the SR electromagnetic spectrum. The remaining 5% corresponds to ultraviolet radiation (UVR).

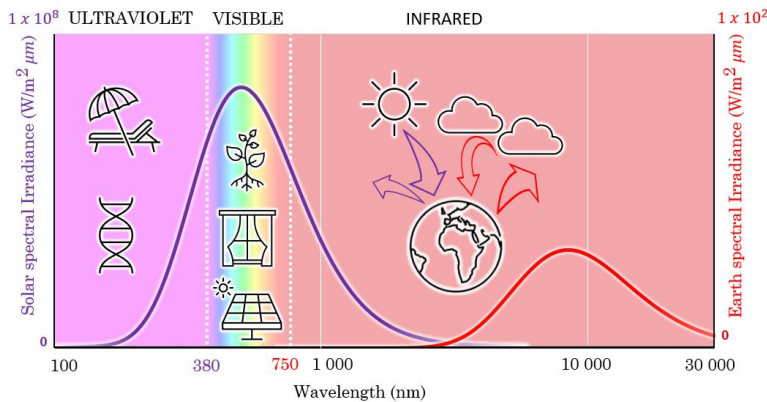


Figure 1. Spectral distribution of the SR and main phenomena related to its energy. (Source: own elaboration).

Photosynthetically Active Radiation (PAR, 400-700 nm) is the energy source to trigger photosynthesis which makes possible the production of food and biomass through leaf photosynthesis [3], and has become central to determination of the impact of deforestation and climate change on agriculture [4]. Atmospheric downwelling longwave radiation, or near-infrared (NIR, 780 nm-3000 nm), is an important component of the terrestrial energy budget since it is strongly related with the greenhouse effect and affects the climate remarkably [5]. NIR is beneficial to building energy savings in winter but undesirable in summer [6]. UVR (UV-A, 315-400 nm; UV-B, 280-215 nm) is responsible for a variety of photochemical reactions [7] and, especially its shorter wavelengths, has deleterious effects in many biological systems.

In the energy context, particularly in renewable energies, the broadband solar radiation is converted into useful heat or electricity, through solar thermal, photovoltaic (PV) or thermal (PT) installations. The interaction of the wavelengths of the visible spectrum of SR (daylighting or illuminance, L, 400-780 nm) with the photoreceptors of the human eye, allows to perceive the world and its colours, and it influences on mood and circadian rhythms [8]. The knowledge and management of daylight is essential to improve the energy efficiency of buildings and visual comfort [9].

The global radiation reaching the Earth's surface can be divided into two components according to their different nature, i.e., direct radiation and diffuse radiation. While direct radiation does not undergo any attenuation on its journey through the atmosphere, diffuse radiation is caused by the interaction of SR with the molecules and particles of the atmosphere. Simple and multiple scattering and absorption phenomena take place, simultaneously or not, at every wavelength. These effects modify the intensity and the spectral distribution of the incident radiation and redistribute the energy in different directions. As a result, diffuse radiation is highly non-uniform and anisotropic [10].

Accurate knowledge of the angular distribution of radiance and luminance in the sky vault is essential when determining the diffuse component received by a given tilted surface, particularly if this surface is in a complex or obstructed environment, such as a city. That is why the modeling of the angular distribution of radiance and luminance in the sky has been the subject of study since the first half of the 20th century.

In general, the first step in approaching such modeling is the knowledge of the sky conditions and their classification into different clusters with similar characteristics that facilitate their independent treatment. Some of the commonly used meteorological indices (MIs) for characterizing the sky condition are the sky's clearness (ϵ) and sky's brightness (Δ) proposed by Perez et al. [11] or the sky index used by Igawa et al. [12], (Si).

However, sky classification is not a trivial task, due in part to the abstract conceptual definitions as clear, intermediate, and overcast conditions, as well as other intermediate ranges. In the case of homogeneous skies, overcast and clear conditions have similar characteristics easily modeled. Thus, the first models of angular distribution of luminance in the sky vault were dedicated to a single sky type. This is the case of the Moon and Spencer [13], model for overcast skies, and the Kittler [14], model for clear skies, that resulted in two standards adopted by the International Commission on Illumination (CIE, Commission internationale de l'éclairage) [15, 16].

However, clear and overcast skies represent only the extremes of a wide range of variability of the real skies. In order to deal with this reality, a second group of angular distribution models for all sky conditions emerged, including the models developed by Perraudau [17], Matsura and Iwata [18], Perez et al. [11], Brunger and Hooper [19], Perez et al. [20], Igawa et al. [12] and Igawa [21].

For their part, Kittler et al. [22, 23] proposed a set of 15 sky standards whose luminance distributions, called standard sky luminance distributions (SSLD), were described in the SSLD catalog. In 2004, this

proposal was consolidated as CIE Standard: CIE S 011/E:2003, and recognized by the ISO Council: ISO 15469:2004 [24] that incorporated the existing CIE standard skies.

Sky classification according to the 15 types established in the ISO/CIE standard has consumed a great deal of research effort. In fact, more than 400 research articles have been published on this subject so far. In this regard, several authors had proposed different methodologies for the characterization of the sky according to the ISO/CIE standard [25]. In this sense, machine learning for meteorological and atmospheric applications has already advanced significantly [26]. According to the goal, researchers should decide the most potentially effective model among a wide variety. Feature selection (FS) relies on statistics and machine learning to identify redundant information and detect the input variables most related to the desired output [27]. The capacity of supervised machine learning tools to identify patterns and classified data allows computers to perform specific tasks autonomously. This science has gained relevance in recent years due to the increase in computing capacity and the data boom, and it is presented as new alternative for sky classification. Artificial neural network (ANN) highlights for sky classification from sky images [28] or thought MIs [29].

It should be noted that the angular distribution of luminance model proposed by the ISO/CIE standard, as well as other widely used angular distribution models such as the All-weather [20] or the All-sky [12, 21], consider a homogeneous distribution of luminance in the sky vault and symmetrical with respect to the solar meridian. However, the natural sky luminance is neither monotonic nor smooth function of zenith/azimuth angle, primarily due to the presence of broken cloud arrays or single clouds scattered over the whole sky vault [30].

The most used and precise way to obtain the angular luminance distribution in the sky is its direct measurement using the sky scanner devices. Some of these devices are prototypes built to be used in other specific studies, as it is the case of the work by Perez et al. [11] and adapted

for their use as sky-scanner. Also, in the field of experimental devices, it is worth mentioning the portable sky scanner for measuring extremely low night-sky brightness (PePSS) described by Kocifaj et al. [30]. However, commercial sky scanners, such as the EKO MS-321L, are the most widely used. This device measure the radiance and luminance corresponding to the 145 patches of sky hemisphere, according to the CIE Guide [31].

In general, commercial sky scanners have relatively low temporal and spatial resolutions. On the one hand, they last more than four minutes to make a complete scan of the sky. This time is considerable in relation to the possible variability in sky conditions. On the other hand, sky scanners generally provide the radiance and/or luminance measurements of the aforementioned 145 sky patches that cover only approximately 2/3 of the sky vault, with a very limited spatial resolution determined by the relatively large field of view (FOV) of their radiance and luminance sensors. In addition, the number of routinely operating sky scanners in the world is very small [31–34].

Currently, the use of luminance-calibrated sky cameras has become one of the most promising techniques for the accurate measurement of the angular distribution of luminance in the sky [35, 36]. In this sense, the use of hemispherical sky images overcomes the main drawbacks of commercial sky scanners. The spatial resolution increases considerably. In this case, it depends on the resolution of the sensor itself. In addition, the measurement time is significantly reduced as well their cost. Thus, the use of hemispheric images of the sky is emerging as one of the most accurate methods for the characterization of sky conditions.

1.2. General and partial objectives

As presented above, the importance of modeling luminance and radiance patterns for taking advantage of renewable energy is a global concern. Thus, the main objective of this doctoral dissertation is **Modeling daylighting for visual comfort and energy efficiency in buildings**. This purpose covers several fields of knowledge: machine learning,

statistics, photometric, and geometry. The dissertation pursues these partial objectives:

1. Obtaining of CIE standard sky classification using different strategies.
2. Developing new models for global and diffuse illuminance.
3. Establishment of alternative methods to sky scanner for CIE standard sky classification from the use of meteorological indices (MIs) and sky images.

The dissertation develops the three partial objectives in six papers, four main papers and two complementary papers, published in JCR indexed journals, included in Q1 or Q2 of the Energy&Fuels, Applied Physics, or Multidisciplinary Mathematics categories. Each paper describes the theoretical basis, the experimental data acquisition, the data processing, and the procedure to carry out the partial objective. The quality control applied to the experimental data follows the international recommendations [37], and it is consistent along the six publications. Figure 2 represents the relationship among the partial and general objectives.

The first objective of the thesis, the characterization of skies according to the CIE standard, is addressed in the first paper, based on the study of the skies in Burgos, Spain. To do this, the standard procedure is followed using a sky scanner. Two different luminance normalization methods, the Tregenza and the NR, were compared. This paper established the methodology to obtain de CIE standard sky classification used as reference for the following works included in this thesis.

Once proposed alternative to the use of sky scanner to obtain the CIE standard sky classification is the calculation of vertical sky component, VSC (Paper II), which directly refers to the daylight availability in façades of buildings, and the luminous efficacy, and allows obtaining illuminance values from irradiance data (CP-I).

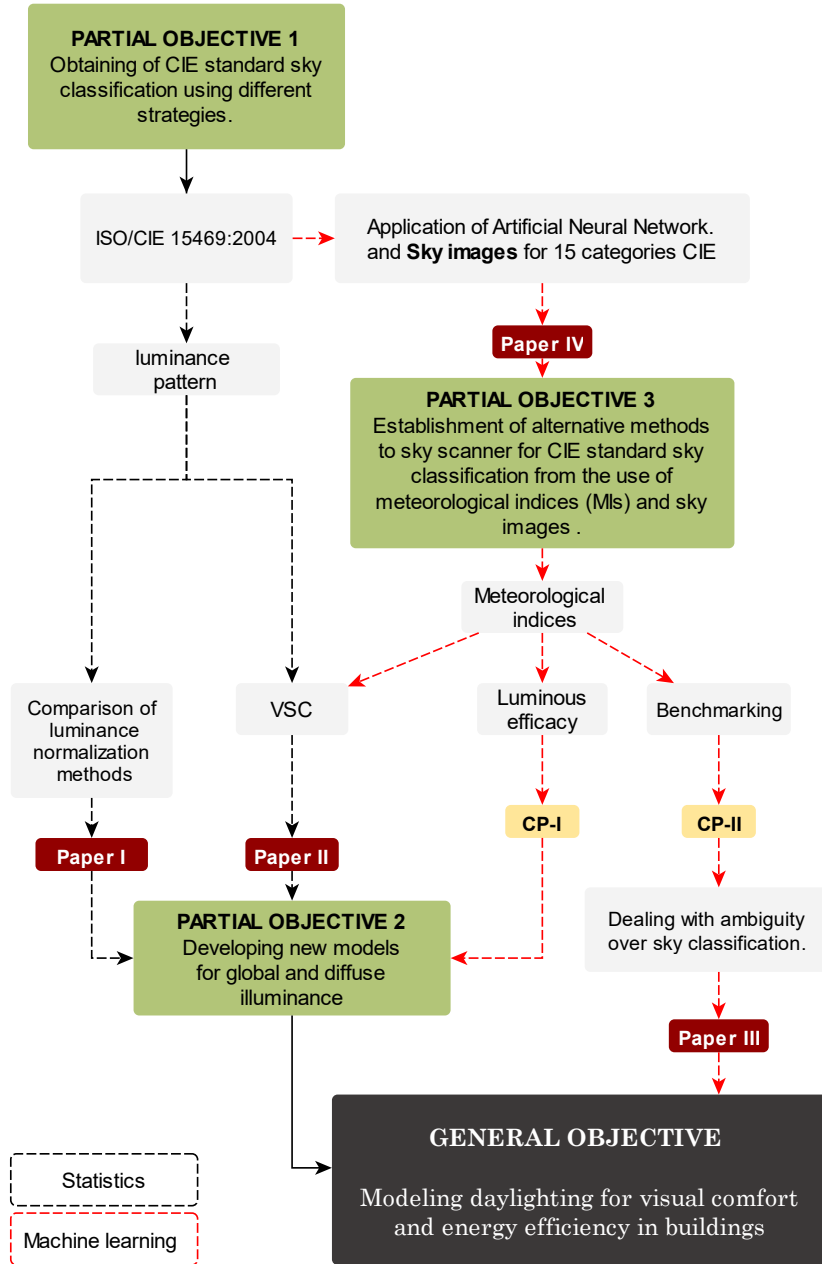


Figure 2. Relation of the General and partial objectives with the publications.

Although the information on the angular distribution of luminance in the sky is valuable, its measurement is not very common. Therefore, this work focused on alternative methods to obtain the CIE standard sky classification that avoid the use of a sky scanner. One of the suggested

method is the use of Meteorological Indices (MIs), and machine learning algorithms. The MIs are meteorological variables, directly measured or calculated through well-known mathematical expressions that can represent, individually or by combination of some of them, the sky conditions. However, the great number and diversity of these indicators, defined by different authors along time, together with the frequent unavailability of the CIE sky classification, has led to a marked heterogeneity of criteria and definition of concepts of clear, partial, and cloudy skies. Papers III and CP-II compare a sample of the sky cloudiness classification through classic MIs with the CIE standard sky classification to establish correlations between them. Specifically, CP-II uses benchmarking to evaluate each author's original criteria against the standard CIE. Paper III compares an extended the MIs set to analyse their contribution to the derived classification cloudiness of the skies as overcast, partial, and clear. Finally, Paper IV directly faces the problem of not having a sky scanner by using artificial intelligence image processing as a procedure to obtain sky classifications without using any meteorological index or sky scanner.

The dissertation structure is as follows. *Chapter 1: Introduction and Objectives*, sets the general description of the dissertation and introduces the social and scientific interest of this work. *Chapter 2: Results by objectives*, describes each of the partial objectives and provides a detailed description of the achieved results, summarizing the articles that develop the related research. *Chapter 3: Conclusions*, summarizes and describes the most significant conclusions and scientific advances achieved by this dissertation. Finally, the *Future Lines of Research* are proposed.

As Annexe 1, the six published papers, ordered by date of publication, are included. Other works, related to a greater or lesser extent to the work and presented at different scientific congresses are also included in the Annexe 2.

1.3. Experimental Data

All experimental data used in this work have been obtained in the meteorological and radiative station of the research group Solar and Wind Feasibility Technologies (SWIFT) at the University of Burgos. The facility is located on the roof at the Escuela Politécnica Superior building. (42°21'04"N; 3°41'20"O; above mean sea level 856 m). The installation, shown in Figure 3, measures the following meteorological variables: pressure (P), ambient temperature (T), relative humidity (RH), and wind speed (WS), and direction (WD). Regarding the SR spectral components, the facility records irradiance, illuminance, PAR, UV (A,B,E), and net IR. For irradiance and illuminance global and diffuse components are measured on the horizontal plane and on the vertical one, north, south, east, and west orientations. Beam component is also measured for all spectral components. Each data is measured every 30 seconds and the average value is recorded every ten minutes. Experimental data were analyzed and then filtered using conventional quality criteria [37]. Indeed, the installation has a sky scanner that measures the angular distribution of radiance and luminance in the 145 sectors of the sky recommended by the CIE standard [24], and a sky camera (All Sky-camera). The sky scanner takes 4 minutes to record the complete skydome, and the timing is every 10 minutes. The sky-camera takes a picture of the sky every 15 seconds in RAW, png, and jpg format.

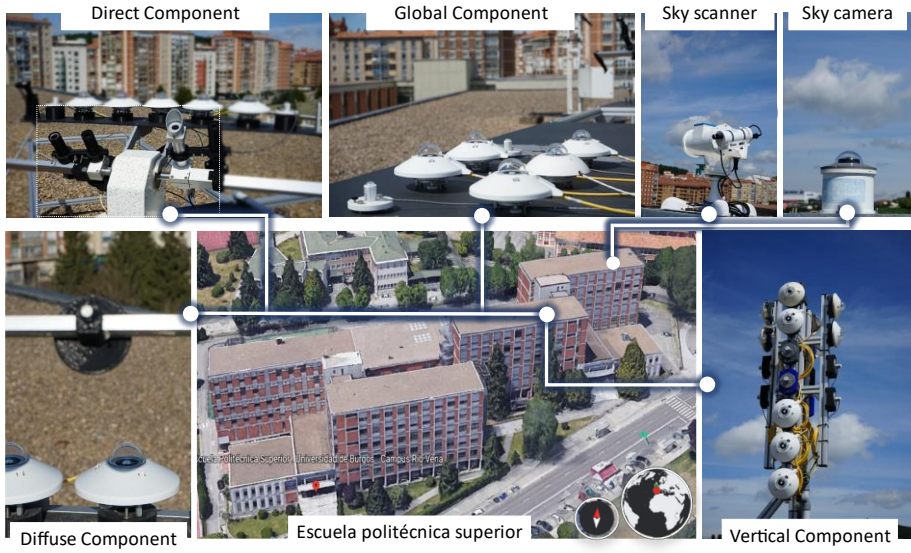


Figure 3. Radiometric variables recorded at the SWIFT experimental facility.

Table 1. Radiometric variables recorded at the SWIFT experimental facility. The table includes each sensor model, and the commissioning date. Table includes reference of the Paper in which the recorded data were used.

Variable (Units/Format)	Type	Sensor	From Date	Paper				CP	
				I	II	III	IV	I	II
Horizontal Irradiance ($W \cdot m^{-2}$)	Diffuse	Hukseflux	Jan. (2014)	-	-	x	-	x	x
	Global	SR12-T205		-	-	x	-	x	x
	Direct	Hukseflux DR01-05		-	-	x	-	x	x
Horizontal Illuminance (lx)	Diffuse	EKO ML-020S-O	Apr. (2016)	-	x	x	-	x	x
	Global			-	x	x	-	x	x
	Direct			-	x	x	-	x	x
Vertical Illuminance (lx)	Global:N	EKO ML-020S-O	Apr. (2016)	-	x	-	-	-	-
	Global:S			-	x	-	-	-	-
	Global:E			-	x	-	-	-	-
	Global:W			-	x	-	-	-	-
Luminance pattern ($kcd \cdot m^{-2}$)	-	EKO MS-321LR	Jun. (2016)	x	x	x	x	x	-
Cloud Cover (%)	-	SONA202U	Sept. (2016)	-	-	x	x	-	-
Sky images (RAW, png, jpg)	-			-	-	x	-	-	
Temperature ($^{\circ}C$)	-	Campbell Scientific CS215	Apr. (2016)	-	-	x	-	x	-
Relative Humidity (%)	-			-	-	x	-	-	-
Wind Speed ($m \cdot s^{-1}$)	-	Campbell Scientific 03002 Wind	(2016)	-	-	x	-	-	-

1.4. Summary of papers

Next, a short summary of the publications that support this thesis is presented. The publications appear according to the position of the partial objectives to which they correspond.

- **PAPER-I:** *Suárez-García, A., Granados-López, D., González-Peña, D., Díez-Mediavilla, M. and Alonso-Tristán, C., 2018. Seasonal characterization of CIE standard sky types above Burgos, northwestern Spain. Sol. Energy 169, 24–33. DOI: 10.1016/j.solener.2019.11.060.* This paper compared two methodologies for sky classification in Burgos, Spain. Both methods used the international standard ISO 15469:2004 of the Commission International of illumination (CIE) for the sky classification and differ in the luminance normalization method.
- **PAPER-II.** *Granados-López, D., Díez-Mediavilla, M., Dieste-Velasco, M.I., Suárez-García, A. and Alonso-Tristán, C, 2020. Evaluation of the vertical sky component without obstructions for daylighting in Burgos, Spain. Appl. Sci. 10 (9), 3095. DOI: 10.3390/app10093095.* The second paper in the dissertation used the standard ISO 15469:2004 (CIE) to obtain the sky luminance pattern and obtained through numerical integration, the horizontal and vertical diffuse illuminance. The VSC, defined as the ratio of the vertical and diffuse illuminance, was used to estimate daylight availability for lighting studies in buildings.
- **PAPER-III.** *Granados-López, D., Suárez-García, A., Díez-Mediavilla, M. and Alonso-Tristán, C., 2021. Feature selection for CIE standard sky classification. Sol. Energy 218, 95-107. DOI: 10.1016/j.solener.2019.11.060.* This paper deepened in the use of MIs for cloudiness sky classification, using feature selection (FS) to analyse the relationship of an extended set of MIs (43 traditional used MIs for sky conditions classification). The FS process avoided redundant information and highlighted the MIs that are strongly related to sky conditions using the CIE standard taxonomy as reference.

-
- **PAPER-IV.** *Granados-López, D., García-Rodríguez, A., García-Rodríguez, S., Suárez-García, A., Díez-Mediavilla, M., Alonso-Tristán, C., 2021. Pixel-based image processing for CIE standard sky classification through ANN. Complexity 2021, 1-15. DOI: 10.1155/2021/2636157.* Paper IV dealt with the CIE standard sky classification through sky images using artificial neural networks (ANN) as alternative to the use of sky scanner devices. In addition, it compared several image processing methods to optimize the ANN model.

Additionally, the following two publications complement actively the thesis.

- **CP-I.** *Dieste-Velasco, M.I., Díez-Mediavilla, M., Granados-López, D., González-Peña, D., Alonso-Tristán, C., 2019. Performance of global luminous efficacy models and proposal of a new model for daylighting in Burgos, Spain. Renewable Energy 133, 1000-1010. DOI: 10.1016/j.renene.2018.10.085.* This paper aimed to estimate global illuminance through the global luminous efficacy. It compared eighteen classic global illuminance models, which are applicable exclusively for certain sky conditions (clear, partial, overcast, or all-sky conditions). Besides, it proposed a new model of global luminous efficacy, that resulted very effective for all-sky conditions.
- **CP-II.** *Suárez-García, A., Díez-Mediavilla, M., Granados-López, D., González-Peña, D., Alonso-Tristán, C.2020. Benchmarking of meteorological indices for sky cloudiness classification. Sol. Energy 195, 499-513. DOI: 10.1016/j.solener.2019.11.060.* CP-II paid attention to the lack of criteria for determining the cloudiness of skies using individually the traditional MIs. The original criteria used by the MIs to define the number and the sky conditions depends on the author and no uniform criteria exist. This work analysed and compared the sky classification capability of the most relevant MIs using ISO/CIE standard (ISO 15469: 2004), as reference.

Chapter 2: Results by objectives

2.1 Obtaining the CIE standard sky classification using different strategies: Seasonal characterization of CIE standard sky types (Paper I)

The availability of natural light is highly recommended for reasons of energy efficiency, visual comfort, and the physical and mental well-being of building occupants. Architects and engineers need quantitative information on illumination levels and solar irradiance absorbed on surfaces at different inclinations for the incorporation of daylighting in the design of energy-efficient buildings and for suitable dimensioning of both the cooling and the heating systems. It requires an accurate estimation of the amount of available outdoor illuminance and, of course, the availability of daylight is mainly influenced by the levels and the patterns of luminance in the sky. To obtain sky luminance distribution, empirical models of homogeneous skies represent a low-cost approach. Many of these methods [38] aim at estimating daylight availability.

In 2003, the ISO/CIE standard defined 15 sky types [24] each of them with the same well-defined sky luminance pattern. Once the sky type is identified, the solar irradiance and daylight illuminance on the surfaces of interest can be obtained through simple mathematical expressions [39]. The luminance distribution for each standard sky type can help arrive at accurate determinations of daylight illuminance [22].

In this regard, Paper I aims at defining the daylight conditions in Burgos according to the ISO/CIE categorization. The ISO/CIE standard describes the relative luminance of a sky point (l_p), normalized by the sky's zenith luminance (L_z), as the product of the relative gradation function $\varphi(Z_p)/\varphi(0)$ and the indicatrix function $f(\chi)/f(Z_s)$, (Eq. 1).

$$l_p = \frac{L_p}{L_z} = \frac{\varphi(Z_p)}{\varphi(0)} \frac{f(\chi)}{f(Z_s)} \quad (\text{Eq. 1})$$

where L_p is the absolute luminance of the sky point, Z_p is its zenith angle of the sky point and χ is the scattering angle, that is, the angular distance between the sun and the sky vault point (see Figure 4). Z_s is the angle of elevation of the sun.

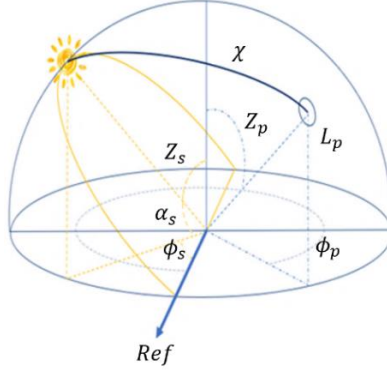


Figure 4. Angles defining the position of the sun and a sky element p [43].

The gradation function, $\varphi(Z_p)$, describes the luminance variation in the sky dome from the horizon to the zenith. (Eq. 2) defines the relative gradation function as the quotient of the gradation corresponding to a given zenith angle, $\varphi(Z_p)$, to the one corresponding to zenith, $\varphi(0)$.

$$\frac{\varphi(Z_p)}{\varphi(0)} = \frac{1 + a \cdot \exp(b/\cos Z_p)}{1 + a \cdot \exp(b)} \quad (\text{Eq. 2})$$

The indicatrix function, $f(\chi)$, models how the luminance changes from the sun position to the sky point. (Eq. 3) describes the relative indicatrix function.

$$\frac{f(\chi)}{f(Z_s)} = \frac{1 + c \cdot [\exp(d\chi) - \exp(d\pi/2)] + e \cdot \cos^2 \chi}{1 + c \cdot [\exp(dZ_s) - \exp(d\pi/2)] + e \cdot \cos^2 Z_s} \quad (\text{Eq. 3})$$

Coefficients a , b , c , d , and e adopt different values according to the sky conditions, as Table 2 shows. A graphical representation of the relative gradation function, $\varphi(Z_p)/\varphi(0)$, and the indicatrix function, $f(\chi)$, are presented in Figure 5.

Table 2. Parameters of CIE standard Sky types [40].

Type	a	b	c	d	e	Description
I.1	4.0	-0.70	0	-1.0	0.00	Overcast with a steep gradation and azimuthal uniformity
I.2	4.0	-0.70	2	-1.5	0.15	Overcast with a steep gradation and slight brightening toward sun
II.1	1.1	-0.80	0	-1.0	0.00	Overcast with a moderate gradation and azimuthal uniformity
II.2	1.1	-0.80	2	-1.5	0.15	Overcast with a moderate gradation and slight brightening toward sun
III.1	0.0	-1.00	0	-1.0	0.00	Overcast, foggy or cloudy, with overall uniformity
III.2	0.0	-1.00	2	-1.5	0.15	Partly cloudy with a uniform gradation and slight brightening toward sun
III.3	0.0	-1.00	5	-2.5	0.30	Partly cloudy with a uniform gradation and a brighter circumsolar effect
III.4	0.0	-1.00	10	-3.0	0.45	Partly cloudy, rather uniform with a clear solar corona
IV.2	-1.0	-0.55	2	-1.5	0.15	Partly cloudy with a shaded sun position
IV.3	-1.0	-0.55	5	-2.5	0.30	Partly cloudy with brighter circumsolar effect
IV.4	-1.0	-0.55	10	-3.0	0.45	White-blue sky with a clear solar corona
V.4	-1.0	-0.32	10	-3.0	0.45	Very clear / unturbid with a clear solar corona
V.5	-1.0	-0.32	16	-3.0	0.30	Cloudless polluted with a broader solar corona
VI.5	-1.0	-0.15	16	-3.0	0.30	Cloudless turbid with a broader solar corona
VI.6	-1.0	-0.15	24	-2.8	0.15	White-blue turbid sky with a wide solar corona effect

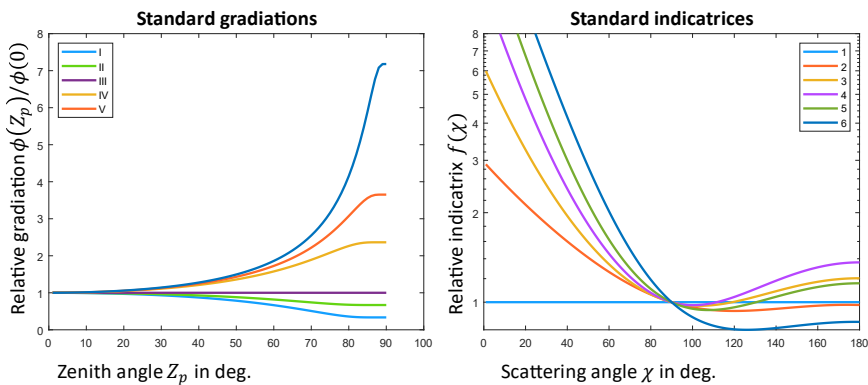


Figure 5. Relative gradation and indicatrix function for the CIE sky-classification[40].

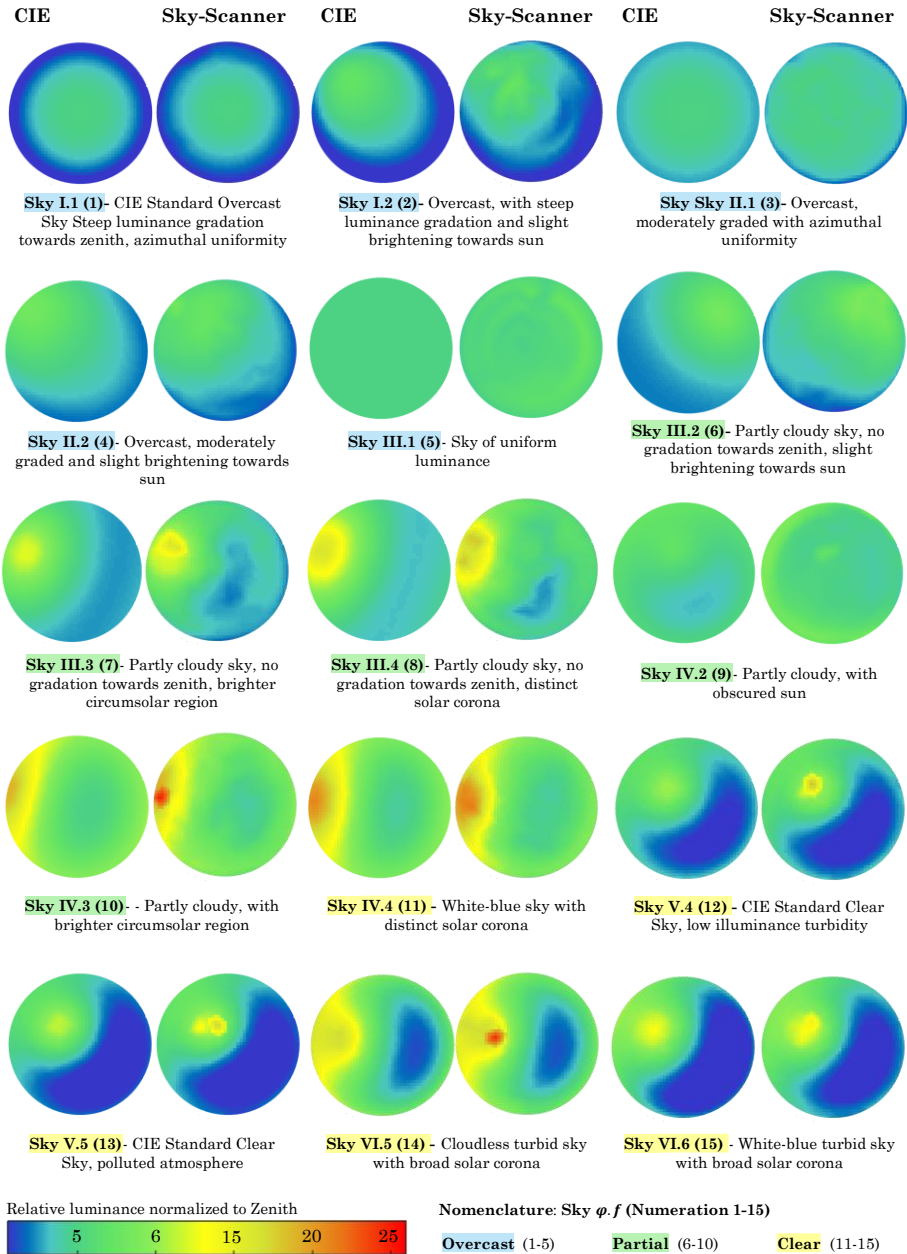


Figure 6. Comparison of the modeled CIE luminance distribution with the real one obtained by the sky-scanner. (Source: own elaboration)

Figure 6 shows the horizontal projection of the angular distribution of relative luminance normalized to zenith for each of the 15 CIE sky types. The corresponding real luminance maps obtained from the processing of

hemispherical sky images are shown next to each of the theoretical luminance distributions. These images were acquired at the UBU meteorological station by a EKO sky-scanner (Table 1).

The first proposed criterion for sky classification according to the CIE was the Standard Sky Luminance Distribution method (SSLD) [22]. Here, the sky classification derives from a family of 15 curves, one for each standard sky, parameterized by the ratio of the zenith luminance to the horizontal diffuse illuminance. Nonetheless, this method is not valid when the solar altitude exceeds 35° [39]. Unfortunately, these solar altitudes are very common in middle latitudes, such as Spain, especially during the summertime. The solution to this non-negligible problem was addressed by later proposals that based the sky type selection on the comparison of standard and real angular distribution of normalized sky luminances. Paper I compares two widely used classification approaches:

- The Tregenza method [41] that normalizes the sky patch luminance ($L_{pr,sc}$) with the horizontal diffuse illuminance ($LxDH$), according to (Eq. 4).

$$L_{pr,sc} = \frac{L_p}{LxDH} \quad (\text{Eq. 4})$$

$LxDH$ integrates the luminance contribution of all patches (L_p) over the sky dome. Patches adjacent to the sun position are excluded since the sky scanner sensor, used to measure the sky luminance distribution, saturates above 50 kcd/m^2 . This value is several orders of magnitude lower than the luminance of the sun.

- The Littlefair normalization ratio (NR) method [42] implements a factor for normalizing the luminance of each sky patch modelled from the ISO/CIE with the experimental luminance pattern, (Eq. 5).

$$NR = \frac{\sum L_p \cos \phi_p \sin \phi_p d\phi_p d\alpha_p}{\sum l_p \cos \phi_p \sin \phi_p d\phi_p d\alpha_p} \quad (\text{Eq. 5})$$

where ϕ_p and α_p are the elevation angle and the azimuth of the sky patch p , respectively. Hence, the normalized sky luminance results from the product of the relative theoretical luminance (l_p) and NR, as follows, (Eq. 6):

$$L_{pred_{p,sc}} = l_p NR \quad (\text{Eq. 6})$$

Both methods follow the same procedure to classify the sky condition according to the ISO/CIE standard, except for the parameter used for luminance normalization. The comparison between the normalized luminances, measured and modelled, for each CIE sky type is done by means of a goodness indicator. The used indicator differs depending on the classification method considered, Root-mean-square error ($rms_{sc,st}$) in the Tregenza method (Eq. 7), and Relative Root Mean Square Error ($rRMSE$) in the NR method (Eq. 8). Fifteen values of the indicator are obtained according to the fifteen ISO/CIE standard sky types. Finally, the sky category whose indicator performs best is selected.

$$rms_{sc,st} = \sqrt{\frac{\sum (L_{pr,sc} - L_{pr,st})^2}{N}} \quad (\text{Eq. 7})$$

$L_{pr,st}$, is the normalized luminance of a sky patch by the Tregenza Method, corresponding to a (CIE) standard sky, $L_{pr,sc}$, is the normalized luminance of a sky patch by the Tregenza Method, corresponding to an experimental measurement (Eq. 4) N is the number of measurements, excluding the empty patches. Finally, sc refers to an experimental type and st is the particular standard sky type that is tested. $L_{pred_{p,sc}}$ is the normalized luminance obtained using the NR factor (Eq. 6), L_p is the luminance measured in kcd/m^2 by the Sky scanner, and p refers to the path of the sky.

$$rRMSE = \sqrt{\frac{1}{N} \sum \left(\frac{L_{pred_{p,sc}} - L_p}{L_p} \right)^2} \quad (\text{Eq. 8})$$

The data used in Paper I for the comparison of the two classification methods came from the sky scanner EKO MS-321 Luminance measurements recorded between June 2016 and May 2017 at the UBU meteorological station, described in Table 1.

Regarding the seasonal classification of skies in Burgos, results in Figure 7 show the relative frequency of occurrence (FOC, %) of each sky type over the period of study, calculated using both the Tregenza and the NR method. As can be seen, all CIE sky types can be found in Burgos. The lowest frequency is for type I.2, corresponding to overcast with the steep gradation and slight brightening toward the sun, and the highest frequency is for type V.5 (cloudless polluted with a broader solar corona). Both methods present very few differences and are almost equivalent in the II.2, IV.3, and V.5 sky types, as shown in Figure 7. The biggest differences in the classification were found in types IV.2 and VI.5.

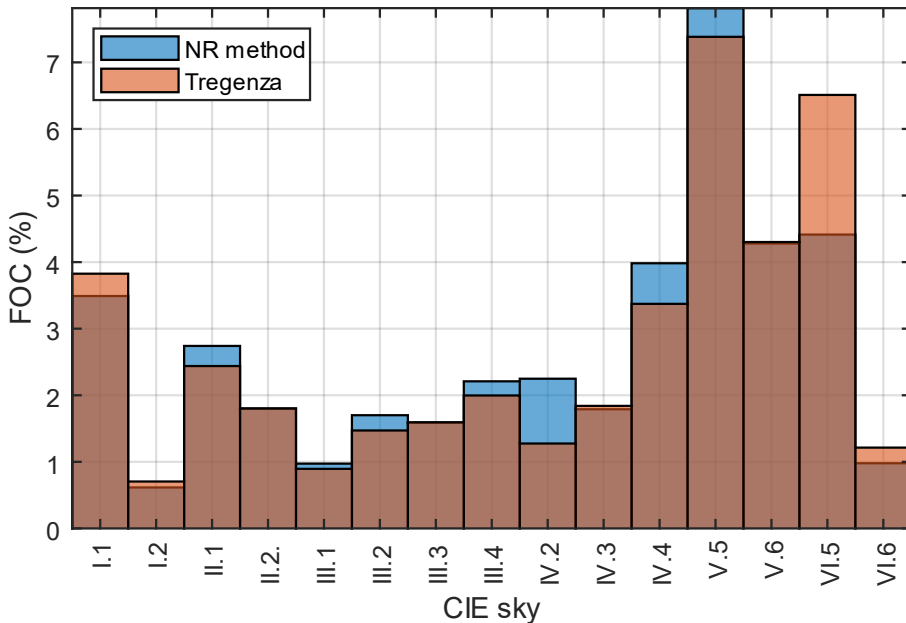


Figure 7. Comparative characterizations of the skies over Burgos [44].

The seasonal classification obtained by both methods is shown in Figure 8 and Figure 9. Sky types I.1, I.2, V.5. and V.6 are predominant in winter (almost 10% frequency each). Clear sky types (IV.4, V.5., V.6, and

VI.5) prevail in spring and summer time. In summer, type IV.2, corresponding to a partly cloudy sky is also frequent. Autumn is a clear sky season too. Grouping the CIE types by seasons clearly profiles the switch between the IV.2 and VI.5 sky types. As can be seen in Figure 8, the NR method classifies 15% of the skies in summer as IV.2 and almost 22% as VI.5. In contrast, the Tregenza method classified 25% of the skies as VI.5, in the same season, as shown in Figure 9. It also labelled 6% of the recorded skies as IV.2. Sky type VI.5. appeared in 23% of cases when using the Tregenza method. It is evident that the mismatch is limited to 9% of all records. The differences between both methods in the other seasons of the year are insignificant.

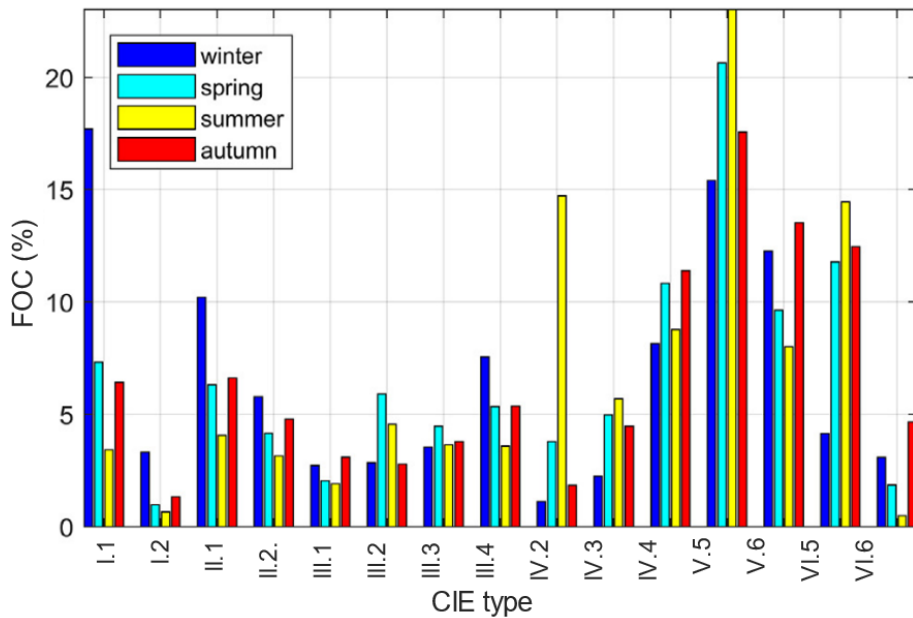


Figure 8. Seasonal CIE sky types histogram over Burgos calculated using the NR method [44].

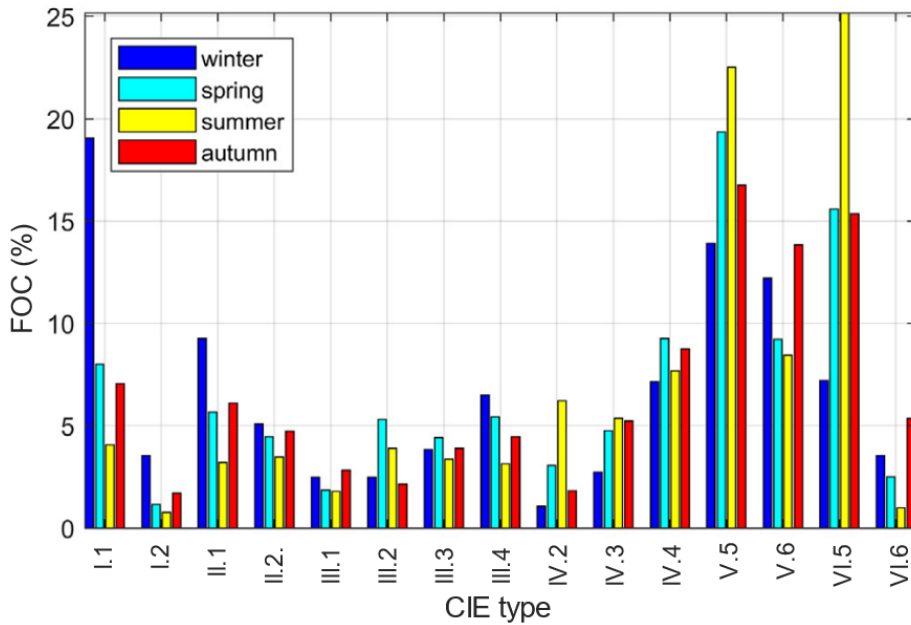


Figure 9. Seasonal CIE sky types histogram over Burgos calculated using the Tregenza Method [44].

Sky type matches obtained by both the Tregenza and the NR methods were compared in a confusion matrix (Figure 10). Each cell of the matrix shows the number of coincidences and their corresponding percentages. The upper-left 3×3 matrix corresponds to the raw comparison, coding the cells either in green, if Tregenza and NR agree, or in red, if otherwise. The gray cells are the percentages that count the total cases in each row or column. Finally, the blueish cell to the lower-right shows the extent of global matching. As it can be seen, the global coincidences for the classification amount to 94.3%.

As explained, both analyzed methods define the type of sky based on a slightly different goodness indicator. So, the relative Root Mean Square Error ($rRMSE$) and the relative Mean Bias Error ($rMBE$) were used to compare the goodness of fit between measured and modeled luminance values. As can be seen in Table 3, both statistical indicators yielded similar results. The low obtained $rRMSE$ values show that both methods tend to get a very acceptable agreement between predictions and measured values,

so both methods can be used with high confidence at the latitude of Burgos. However, this confidence decreases at the highest solar altitudes.

Confusion Matrix

CIE Tregenza	cloudy	1686 23.2%	42 0.6%	0 0.0%	97.6% 2.4%
	partial	34 0.5%	1375 18.9%	53 0.7%	94.0% 6.0%
	clear	0 0.0%	288 4.0%	3785 52.1%	92.9% 7.1%
		98.0% 2.0%	80.6% 19.4%	98.6% 1.4%	94.3% 5.7%
	cloudy	partial	clear	CIE NR Method	

Figure 10. Confusion matrix comparing the Tregenza and the NR cloudiness classifications [44].

Table 3. Statistical indicators RMSE and MBE [40].

-	$\alpha_s \in (0, \pi/2)$		$\alpha_s < 35^\circ$		$\alpha_s > 35^\circ$	
	<i>rRMSE</i>	<i>rMBE</i>	<i>rRMSE</i>	<i>rMBE</i>	<i>rRMSE</i>	<i>rMBE</i>
Tregenza	36.7%	14.1%	34.9%	13.1%	40.1%	16.0%
NR	36.0%	13.3%	33.4%	11.4%	40.8%	17.0%

Therefore, the results obtained in the Paper I have met the Partial Objective 1, that is, obtaining the CIE sky classification. The methodology followed in Paper I will be used henceforth for obtaining the CIE standard sky classification, as reference for establishing the sky conditions.

2.2. Development of models for global and diffuse illuminance (I): Vertical Sky Component (VSC) for Daylight availability (Paper II)

As result of the current renovation and sustainability wave, European countries expressed their strategies to reduce energy consumption and enhance building renovation [1, 43]. According to the International Energy Agency (IEA) electrical lighting in buildings makes up 14% of electrical consumption in the European Union and 19% worldwide [44]. The use of natural light can reduce significantly the energy consumption in buildings offering the occupants comfort and health benefits [45].

A fundamental step towards studying the lighting profile of a building for energy-effective design is compiling information of outdoor conditions [46]. Recording the illuminance on vertically oriented surfaces is relevant for daylight availability, particularly in a high-rise building where the glazed surfaces are large extensions [47]. In addition, the vertical component of solar radiation serves to model the performance of Building Integrated Photovoltaics (BIPV), because the vertical facades of modern cities occupy larger areas than roof surfaces and usually present better maintenance conditions for photovoltaic panels [48].

Nevertheless, despite having very high interest, both diffuse outdoor solar irradiance and illuminance data for the surfaces of interest are not likely to be available in numerous locations around the world [13]. In particular, Paper II studies the outdoor daylight availability through the Vertical Sky Component (VSC) [49]. VSC is defined as the ratio of the vertical diffuse illuminance (L_{xDV}) to the unobstructed horizontal diffuse illuminance (L_{xDH}), in (Eq. 9).

$$VSC = \frac{L_{xDV}}{L_{xDH}} \quad (\text{Eq. 9})$$

VSC offers an overall knowledge about the external conditions to make decisions about room design factors, such as the dimensions and shape of the glazing and rooms or the uses of the indoor spaces. High VSC values

usually head to further illumination in interior spaces. In addition, several entities concerning building construction establish minimum VSC levels for more effective daylighting [46, 49, 50]. Roughly, VSC values that exceed 27% indicate good daylight availability. Furthermore, VSC levels lower than 10% do not guarantee the availability of direct light from the sky. It is noteworthy that this VSC target works on low-density suburban housing models. Thus, the daylight and sunlight review states that, in an inner-city urban environment, 20% of VSC values are reasonably good. Knowledge of Vertical Sky Component (VSC) allows the calculation of daylighting availability for buildings at any cardinal orientation for energetic and visually efficient building and city design. This work describes different alternatives for VSC calculation using a complete experimental characterization of the VSC in an extensive case study carried out in Burgos, Spain.

When experimental values of $LxDH$ and $LxDV$ are available, VSC can be calculated directly. As $LxDV$ data were not available at the experimental facility during the experimental campaign, from September 2016 to January 2019, these data were obtained, considering the similar characteristics of both solar irradiance and illuminance using the well know transposition model, as (Eq. 10) shows:

$$LxDV = LxGV - \frac{\rho}{2} \cdot LxGH - \left(\frac{LxBH}{\sin \alpha_s} \right) \cdot \cos \alpha_s \cdot \cos(\phi_s - \phi_{Nr}) \quad (\text{Eq. 10})$$

Where $LxGV$ and $LxGH$ are vertical, and horizontal global illuminance, respectively, and $LxBH$ is the horizontal beam illuminance; ρ is the albedo; ϕ_{Nr} is the azimuth vertical surface angle; α_s the solar elevation, and ϕ_s is the azimuth of the sun as Figure 11 shows. $\frac{\rho}{2}LxGH$ refers to the reflected illuminance. Therefore, VSC was calculated using (Eq. 11):

$$VSC = \frac{LxGV - \frac{\rho}{2}LxGH - \left(\frac{LxBH}{\sin \alpha_s} \right) \cdot \cos \alpha_s \cdot \cos(\phi_s - \phi_{Nr})}{LxDH} \quad (\text{Eq. 11})$$

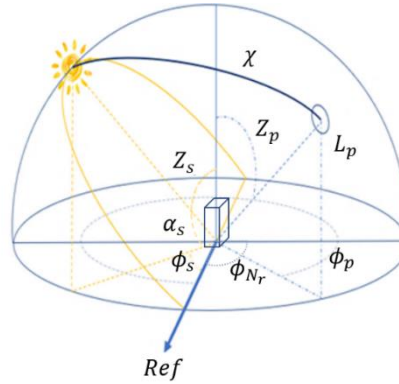


Figure 11. Geometry description: Angles of the Sun and the sky element p .

With a well-defined luminance distribution, the daylight on any surface can be estimated, instead the experimental determination, by integrating the luminance distribution of the sky dome over each surface. Two alternatives to obtain the sky luminance distribution were used in this work: its direct measurement using a sky scanner and the theoretical luminance pattern that the ISO/CIE assigns univocally to each of the sky type.

The sky scanner measures the luminance on a limited number of sky elements (patches) of finite angular size, typically 145, and continuously scan the luminance data corresponding to each patch, L_p . Figure 12 shows the traditional split into 145 patches of the skydome and the geometric components.

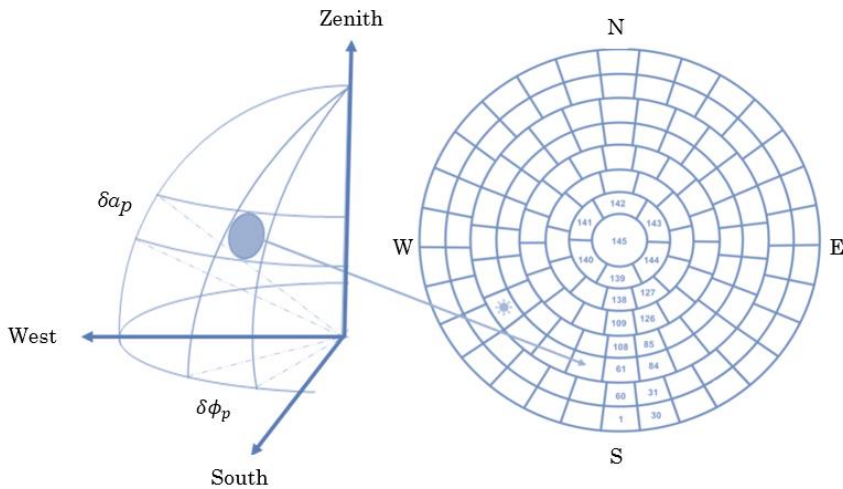


Figure 12. Sky patch's projection [51].

If the CIE standard sky classification is known, the luminance in each sky patch, L_p , can be obtained using (Eq. 12).

$$L_p = \left(1 + c \cdot \left[e^{d\chi} - e^{\frac{d \cdot \pi}{2}}\right] + e \cdot \cos^2 \chi\right) \left(1 + a \cdot e^{\frac{b}{\cos Z_p}}\right) \quad (\text{Eq. 12})$$

Coefficients a, b, c, d , and e , are defined by CIE [24] as functions of the sky type, as shown in Table 2. Z_p is the sky element zenith angle and χ is the dispersion angle, calculated from (Eq. 13):

$$\chi = \arccos(\cos Z_s \cos Z_p + \sin Z_s \sin Z_p \cos|\phi_p - \phi_s|) \quad (\text{Eq. 13})$$

Where ϕ_p is the azimuth angle of the sky element, p , and Z_s and ϕ_s are the zenith and azimuth angles of the sun. The χ represents the shortest angular length between the sky element, p , and the sun, as is shown in Figure 11. Once the luminance in each sky element is known, equations (Eq. 14) to (Eq. 16) can be applied to calculate $LxDH$, $LxDV$ and, therefore, VSC.

$$LxDH = \sum_{p=1}^{145} L_p \sin \alpha_p \cos \alpha_p \delta \alpha_p \delta \phi_p \quad (\text{Eq. 14})$$

$$LxDV = \sum_{p=1}^{145} D_p L_p \cos \alpha_p \delta \alpha_p \delta \phi_p \quad (\text{Eq. 15})$$

D_p is a geometrical factor which projects only the sky patches that apport illuminance to the vertical surface [52], (Eq. 16):

$$D_p = \begin{cases} \cos \alpha_p \cos(\phi_p - \phi_{N_r}) & \text{if } 0 \leq |\phi_p| \leq 90^\circ \\ 0 & \text{Otherwise} \end{cases} \quad (\text{Eq. 16})$$

As can be seen, sky types 1, 3, and 5 present constant VSC values, regardless of χ . Therefore, the orientation of the vertical surface has no effect on the level of illumination. The predicted VSC values for these sky types (1, 3, and 5) were 38.5%, 45% and 50%, respectively, see Figure 13.

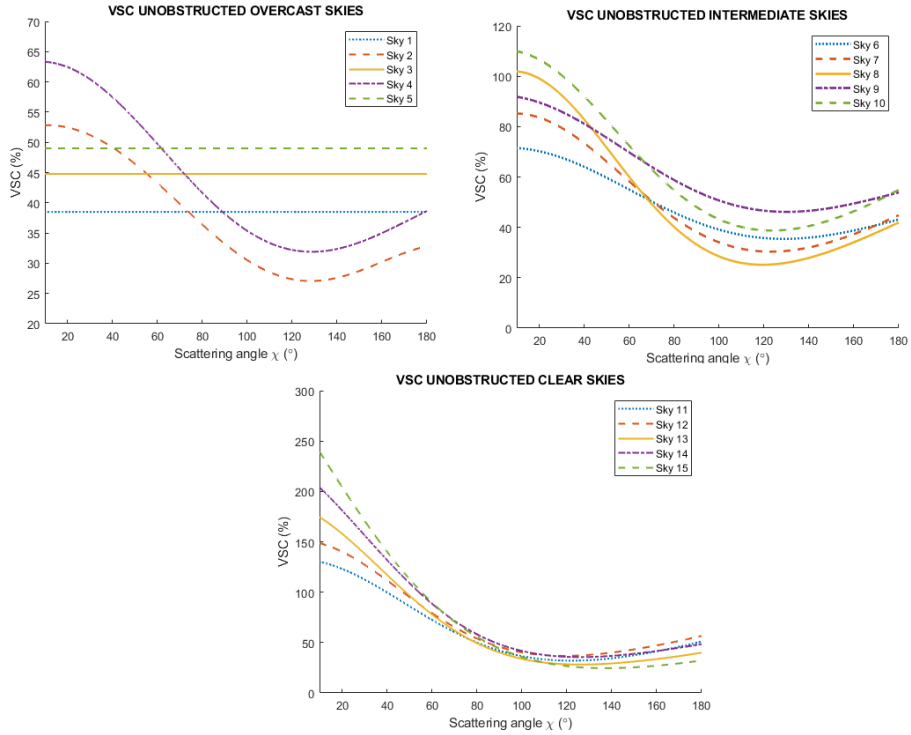


Figure 13. VSC dependence with the scattering angle for all CIE skies. The labels one (overcast) to fifteen (clear) design a specific CIE sky typology [51].

The experimental data of Lx_{DV} obtained from (Eq. 10) on the four cardinal facing surfaces, were compared to those calculated as projections on the same surface using the sky-scanner luminance pattern. Both the $rRMSE$ and the $rMBE$ values are recorded in Table 4 and, as can be appreciated, are comparable to those obtained for horizontal diffuse illuminance. It can, therefore, be concluded from this study that the use of the sky-scanner measurements, to determine the diffuse illuminance on any horizontal or tilted surface, had an intrinsic $rRMSE$ due, mainly, to the technical specifications of the experimental device, near 30%, agreeing with other work [47].

Table 4. $rRMSE$ and $rMBE$ calculated for the vertical diffuse illuminance, L_{dv} , and the horizontal diffuse illuminance, L_{dH} . The reference values of L_{dV} were calculated from (Eq. 10). The reference values of L_{dH} were measured in the experimental facility, described in Table 1 [51].

Orientation		$rRMSE$ (%)	$rMBE$ (%)
Vertical	South surface	27.31	-3.55
	North surface	21.46	-0.31
	East Surface	31.19	-11.87
	West Surface	27.69	4.63
Horizontal		27.14	-3.80

VSC values calculated from the different procedures were compared, taking the VSC values obtained by (Eq. 11) as a reference. Both the $rRMSE$ and the $rMBE$ parameters are shown in Table 5. As can be seen, for the different vertical surfaces facing the cardinal orientations, the statistical indices ranged between 23% and 32% for the $rRMSE$ and between -1% and 16% for the $rMBE$. Both procedures underestimated the VSC, as shown by the negative $rMBE$ values. The highest discrepancies between the different approaches were for the east orientation. This observation agreed other works, where the results of different approaches for VSC calculation varied with the different surface orientations [52].

Table 5. $rRMSE$ and $rMBE$ results from the comparison between the VSC values calculated with the different approaches [51].

VSC	$rRMSE$ (%)	$rMBE$ (%)	$rRMSE$ (%)	$rMBE$ (%)
	Sky-Scanner		CIE	
South surface	24.46	-8.47	25.56	-11.25
North surface	23.46	-4.48	23.53	-5.56
East Surface	31.93	-15.51	29.85	-15.38
West Surface	22.99	-1.14	24.84	-2.90

The experimental values of the VSC calculated from (Eq. 11) were obtained for the city of Burgos. Figure 14 shows the VSC values, classified by intervals, and for the four vertical cardinal orientations. As can be seen

from Figure 14, VSC values lower than 20% were practically non-existent on surfaces facing the four cardinal orientations. During the campaign, VSC was unlikely to have a lower value than 20% in all orientations, the range 20–40% was a little more likely. Range 40–60% has a frequency of occurrence (FOC) that reaches the 30% in almost every cardinal direction. Besides, the south and east orientation highlighted over the rest orientations in the interval 60–100% with FOC equal to 20% and 10%, respectively. Values higher than 100% were present in all directions.

On the other hand, international recommendations [46, 49] refer to the interval $VSC > 27\%$, as the representative value for acceptable daylight levels. During the experimental campaign, FOC of $VSC > 27\%$ varies from 89–90%. North orientation yields good results and, therefore, its energetic and luminous profile should not be underestimated.

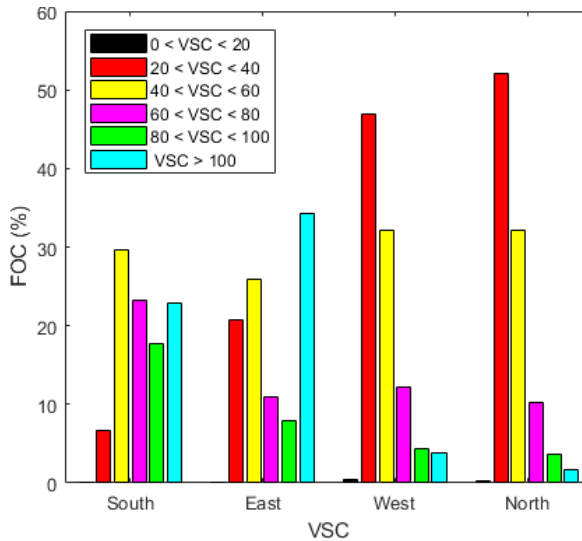


Figure 14. Distribution of VSC values by intervals calculated in Burgos, Spain, between September 2016 and January 2019 [51].

Therefore, values of diffuse illuminance on horizontal, vertical, and VSC cardinal-oriented surfaces, from the luminance pattern of the sky scanner and CIE standard classification for homogeneous skies, showed comparable results. Hence, Paper II contributes to the General Objective of the thesis: Development of models for global and diffuse illuminance.

2.3. Development of models for global and diffuse illuminance (II): Luminous efficacy model (CP I)

Several approaches and international regulations consider the integration of daylight into the building design [53, 54]. However, sometimes finding suitable illuminance measures can be difficult [55]. As a result, daylight prediction often starts from the derivation of solar irradiance measurements [56], since this variable is more frequently recorded at ground meteorological stations.

The luminous efficacy is defined as the ratio between solar irradiance and illuminance. Then, if $LxGH$ is the horizontal global illuminance (lx) and $RaGH$ (W/m^2) is the horizontal global irradiance, the global luminous efficacy is given in lm/W by (Eq. 17).

$$K = \frac{LxGH}{RaGH} \quad (\text{Eq. 17})$$

Numerous proposals in the literature aim at modeling luminous efficacy. Whereas some model it as a function of different variables, others consider it a constant. Also, these models specify or not into sky typologies. CP-I reviewed and compared eighteen classic models of global luminous efficacy: six for all skies, five for clear, three for partly cloudy, and four for the overcast skies. A summary of these models and its main specifications can be seen in Table 6.

In this work, the reviewed models were locally adapted and a new luminous efficacy model, described by (Eq. 18), was developed for both, all-sky conditions, and clear, partial, and overcast skies. The experimental data for this study were gathered at the meteorological and radiometric facility of the research group SWIFT, described in Chapter 1, Table 1. The experimental campaign extended from 1st October, 2016 to 31st May, 2018. The original dataset was divided into two subsets: subset 1 (01/10/16 to 31/03/18) fitted the coefficients of each model, and period 2 (01/04/18 to 31/05/18) validated the calibrated models. The sky classification was done using the Perez's clearness index, ϵ , and Perez sky's brightness, Δ , as

follows: Clear skies ($\epsilon > 5$, and $\Delta < 0.12$), Partial skies – ($1.2 < \epsilon < 5$), and Overcast skies ($\epsilon < 1.2$) [57].

Table 6. Summary of the global luminous efficacy models reviewed in this work. Literature reference of the original model, year, sky type classification, input parameters used in the models and the original place of development of the model[9].

Ref.	Year	Authors	Sky types	Model parameters	Location
[11]	1990	Perez et al.	All	Δ, Z_S, W	USA and Europe
			Clear	α_S	
[58]	1992	Chung	Overcast	α_S, Ω	China
			Partial	α_S, k_d	
			Clear	α_S, k_d	
[59]	1996	Lam and Li	Overcast	116.2 lm/W	China
			Partial	α_S, CC, k_d	
[60]	1998	Munner and Kinghorn	All	k_t	UK
[61]	2000	Robledo and Soler (Model A), and Robledo and Soler (Model B)	Clear	α_S	Spain
[62]	2001	Ruiz et al.	All	α_S, k_t	Spain
[57]	2001	Robledo et al.	Partial	α_S, Δ	Spain
		Robledo et al. (model A) and Robledo et al. (Model B)	Overcast		
[63]	2006	De Souza et al.	Clear	α_S	Brazil
[64]	2011	Fraka et al.	All	121.5 lm/W	Reunion Island
[65]	2011	Mahdavi and Dervishi	All	k_t, T	Austria
[66]	2013	Chaiwiwatworakul and Chirarattananon	All	Z_S, ϵ	Thailand

The new model depends on the clearness index (k_t) and the solar altitude (α_s). The model coefficients (p_0, p_1, p_2) fitted for and All-types of skies, and for clear, partial, and overcast skies, are shown in Table 7.

$$K = p_0 \exp^{p_1 \cdot k_t \cdot \sin(p_2 \cdot \alpha_s^2)} \quad (lm/W) \quad (\text{Eq. 18})$$

Table 7. List of coefficients of the proposed model.

Sky type	Interval	p_0 (lm/W)	p_1 (Dim)	p_3 (rad ⁻²)
Clear	$\epsilon > 5$ and $\Delta < 0.12$	108.591	-0.111	1.031
Partial	$1.2 < \epsilon < 5$	109.152	-0.100	1.013
Overcast	$\epsilon < 1.2$	111.693	-0.103	1.241
CPO	-	111.616	-0.127	1.232

The goodness-of-fit of the models was calculated by means of the statistical indicators *MBE* (%) (Mean Bias Error), and *RMSE* (%) (Root Mean Square Error) [64, 67], using experimental data of two additional months (01/10/16 and 31/03/18).

Table 8 shows the results obtained with the testing data for the reviewed and proposed models. In the case of all sky conditions, the lower value of *RMSE* was obtained by Ruiz et al. model followed by the new proposal. However, the *MBE* obtained with the proposed model was ten times lower than the one obtained with the Ruiz et al. model.

Table 8. Summary of the results obtained with the testing data for the reviewed and proposed models. The reviewed model with the lowest RMSE has been underlined.

	Model	RMSE (%)	MBE (%)
All sky models	<u>Ruiz et al.</u>	2.57	-0.10
	Proposed model - CPO	2.66	-0.01
	Chaiwiwatworakul and Chirarattananon	2.81	1.23
	Mahdavi and Dervishi	2.94	0.94
	Perez et al.	2.98	1.31
	Muneer and Kinghorn	3.22	0.36
	Fakra et al.	3.64	-2.52
Clear Sky models	Proposed model - CPO	0.66	-0.26
	Proposed model - Clear	1.40	-1.21
	<u>Robledo and Soler (Model A)</u>	1.53	-1.25
	Robledo and Soler (Model B)	1.65	-1.32
	Lam and Li	2.07	-0.2
	Chung	2.30	-1.84
	De Souza et al.	3.43	-3.07
Partial cloudy sky models	<u>Robledo et al.</u>	2.43	0.51
	Proposed model - Partial	2.46	-0.09
	Chung	2.67	0.93
	Proposed model - CPO	2.80	0.27
	Lam and Li	3.44	2.25
Overcast sky models	Proposed model - Overcast	2.30	0.65
	Proposed model - CPO	2.48	-0.21
	<u>Robledo and Soler (Model A)</u>	2.52	0.71
	Robledo and Soler (Model B)	2.65	0.86
	Chung	2.76	1.27
	Lam and Li	4.16	2.56

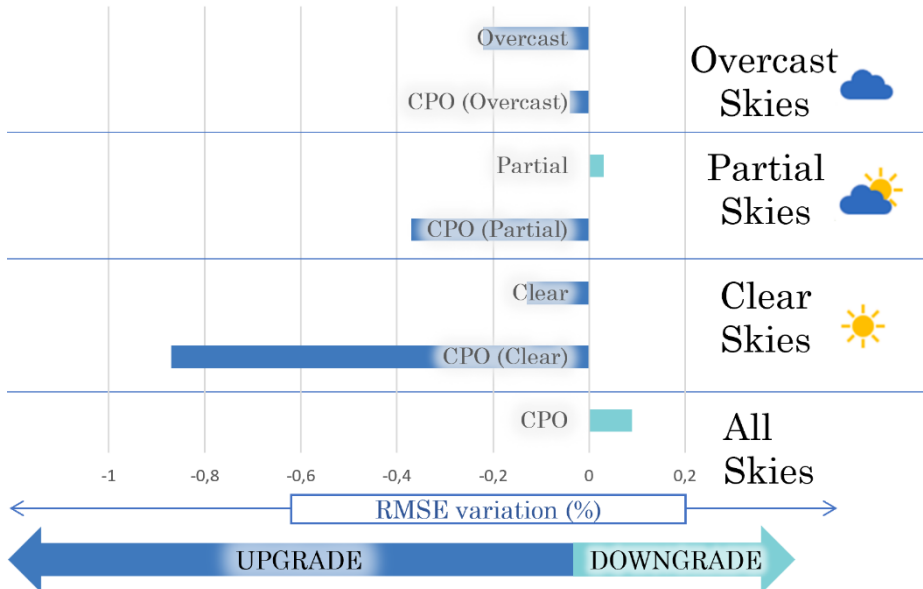


Figure 15. *RMSE variation (%)*: Best score of the reviewed model (underlined model) versus the proposed model in each category, described in Table 8.

As can be appreciated in Table 8 and Figure 15, the new model of global luminous efficacy adequately fitted the illuminance data, both if the data were clustering according to the cloudiness of skies as for all sky conditions. Even the all-sky proposed model fitted well when has been used for clear, overcast, and partial skies. Regarding the *RMSE* value, the new proposed model is in the first positions in the ranking for all studied cases. The calculated *RMSE* values are comparable to the rest of the models and, sometimes, including the *MBE* value, significantly improved the prediction. In addition, it is observed that the *MBE* of all models in clear category underestimated the global illuminance values. As conclusion, the proposed model proved to be effective for modeling the global illuminance in all-sky conditions. Therefore, CP I contributes to the partial Objective II: Developing new models for global and diffuse illuminance.

2.4. Establishment of alternative methods to sky scanner for CIE standard sky classification from the use of meteorological indices (MIs) and sky images (I): Benchmarking of Meteorological Indices (CP II)

As it has been highlighted previously, the determination of the sky conditions using the CIE/ISO standard sky classification requires the use of sky scanner devices for obtaining the radiance and luminance pattern of the sky. The scarcity of these devices in the meteorological ground stations, encourages searching for alternative methods. A complete review of the proposed alternatives can be found in the work of Li et al [25].

A classic alternative may be the use of the meteorological indices (MIs). The meteorological indices are variables, usually recorded at ground meteorological stations or easily derivable from such variables, which collect characteristics of the atmosphere that can be related, to a greater or lesser extent, with the distribution of luminance and radiance of the sky. Complementary Paper II carries out an extensive review of these variables, defined and used, alone or in groups, by different authors to establish sky conditions, generally, but not exclusively, in three categories, clear, partial and overcast, with different objectives. Among these objectives, the most common is the definition of new models of illuminance and irradiance. The definition of the MI is carried out either through direct measurement, or it is calculated through mathematical expressions that use these magnitudes directly measured at the meteorological facilities. The interpretation of sky conditions, in three or more categories, is determined by numerical intervals of the MI. In the literature review carried out in the work, discrepancies were observed between different authors both in the number of intervals proposed for the same index, and in the numerical values of these intervals. In this work, the ability for sky classification of 10 traditionally used meteorological indices has been analyzed, taking as

reference the CIE standard classification in a period of a complete year, from June 2016 to May 2017.

One of the problems addressed in the Complementary Paper II is the homogenization of the sky categories in the same number of classes without changing the original limits established by their authors. This follows a similar approach to that in the work of Gueymard et al. [68] where the authors compared different MI models in an attempt to classify clear skies. Different intervals were tested, in order to mitigate the semantic effects and to maximize MI performance, keeping the class limits that the respective authors set in their original works. Figure 16 shows the possible combinations of the limits of the intervals for the different MIs studied, keeping the ordinal relationship between cloudy, partial, and clear-sky conditions. For some of the MIs, no adaptation was needed.

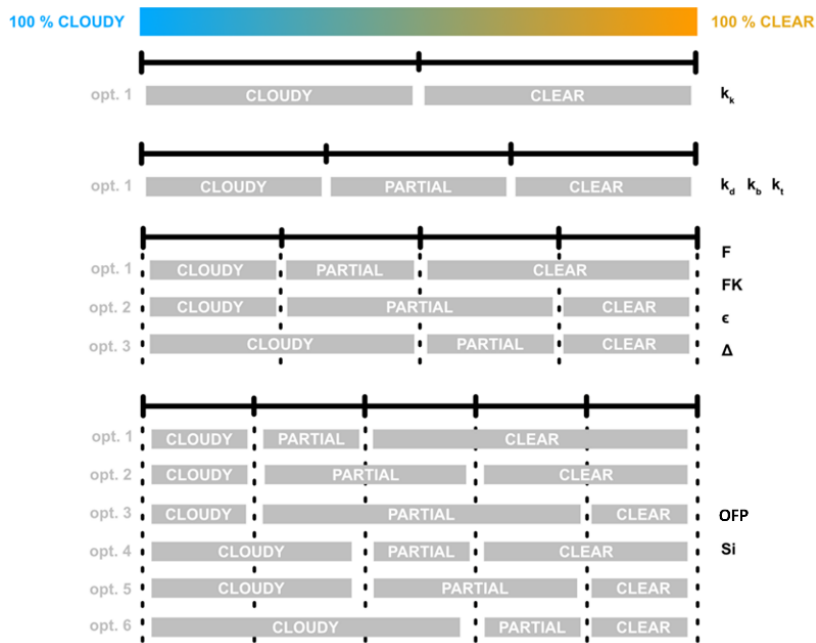


Figure 16. Interval adaptation procedure of the original MIs [69].

When more than three intervals were defined in the original work, all possible combinations for merging the intervals were tested. The adapted interval is the one that maximized the classification metrics for each MI. The original and the adapted values of the intervals selected by each index

are summarized in Table 9. The MIs are calculated using global horizontal irradiance, $RaGH$, diffuse horizontal irradiance, $RaDH$, and direct normal irradiance, $RaBN$.

Table 9. Summary of the MIs used to classify the skies, the original intervals used to define the clear, partial and cloudy sky conditions and the adapted ones used in this work [69].

Symbol	MI		Original		Adapted
k_d	Diffuse fraction [70]	[0.00, 0.33]	clear	[0.00, 0.33]	clear
		(0.33, 0.8)	partial	(0.33, 0.8)	partial
		[0.8, 1)	cloudy	[0.8, 1)	cloudy
k_b	Direct fraction			[0.66, 1)	clear
				(0.33, 0.66)	partial
				(0, 0.33]	cloudy
k_t	Clearness Index [71]	[0.65, 1)	clear	[0.65, 1)	clear
		(0.35, 0.65)	partial	(0.65, 0.35)	partial
		(0, 0.35]	cloudy	(0, 0.35]	cloudy
F	Clearness Function [72]	[0.61, 1.00]	completely clear		
		[0.51, 0.61)	clear	[0.51, 1.00)	clear
		[0.18, 0.51)	partial	[0.18, 0.51)	partial
k_k	Batlles Clearness Index [73]	$k_t > k_{tt}$	clear	$k_t > k_{tt}$	clear
		$k_d < k_k$		$k_d < k_k$	
FK	Klucher Clearness Index [74]	[0.61, 1.00)	completely clear	[0.51, 1.00)	clear
		[0.51, 0.61)	clear	[0.18, 0.51)	partial
		[0.18, 0.51)	partial	(0.00, 0.18)	cloudy
		[0.00, 0.18)	completely cloudy		
ϵ	Perez sky's brightness [11, 75]	[6.20, ∞)	completely clear		
		[2.80, 6.20)	clear	[2.4, ∞)	clear
		[1.50, 2.80)	partial	[1.50, 2.4)	partial
		[1.065, 1.50)	cloudy	[1.00, 1.50)	cloudy
Δ	Perez sky's brightness [11, 75]	[0.48, ∞)	very bright	[0.30, ∞)	clear
		[0.30, 0.48)	bright	[0.10, 0.30)	partial
		[0.10, 0.30)	partial	[0.00, 0.10)	cloudy
		[0.00, 0.10)	very dark		
$OFFP$	Original Nebulosity Index [76]	[0.90, 1.00]	blue sky		
		[0.70, 0.90)	partial	[0.70, 1.00]	clear
		[0.20, 0.70)	partially blue	[0.05, 0.70)	partial
		[0.05, 0.20)	partially cloudy	[0.00, 0.05)	cloudy
		[0.00, 0.05)	totally cloudy		
S_i	Igawa Sky Index [12]	[1.70, ∞)	clear	[1.70, ∞)	clear
		[1.50, 1.70)	almost clear	(0.30, 1.70)	partial
		[0.60, 1.50)	partially clear		
		[0.30, 0.60)	partially cloudy	[- ∞ , 0.30]	cloudy
		[- ∞ , 0.30]	totally cloudy		

Confusion matrices were used for measuring the performance of classification algorithms. Comparing the prediction with reality, there are four possible scenarios: the predicted positive will agree with the actual one (True Positive or TP), the predicted positive will not agree with the actual one (False Positive or FP), the predicted negative will agree with the actual one (True Negative or TN) and the predicted negative will not agree with the real one (False Negative or FN). Each row of the matrix represents the instances of the predicted class, while each column represents instances of the reference class [77]. The matrix was so-named because it easily visualizes whether the algorithm is confusing or mislabeling two classes. In the present work, the reference labels are established by the CIE methodology for sky classification and the predicted labels by the MIs. Hence, the latter are the algorithms or predictive models under analysis.

As mentioned above, the CIE defines fifteen types of sky that can be grouped into three types of cloudiness: clear, partial, or cloudy. The MIs distinguish between the features in an attempt to define these three classes. This task can be analyzed as a multiclass problem decomposed in a multiple dichotomous classification where each cloudiness categorization is predicted against the other remaining ones (Figure 17). At the end of the process, the dichotomous classification of the MIs is evaluated using several performance ratings, Accuracy, Jaccard Index, Cohen Kappa, and Matthews, whose brief description are shown in *Table 10*.

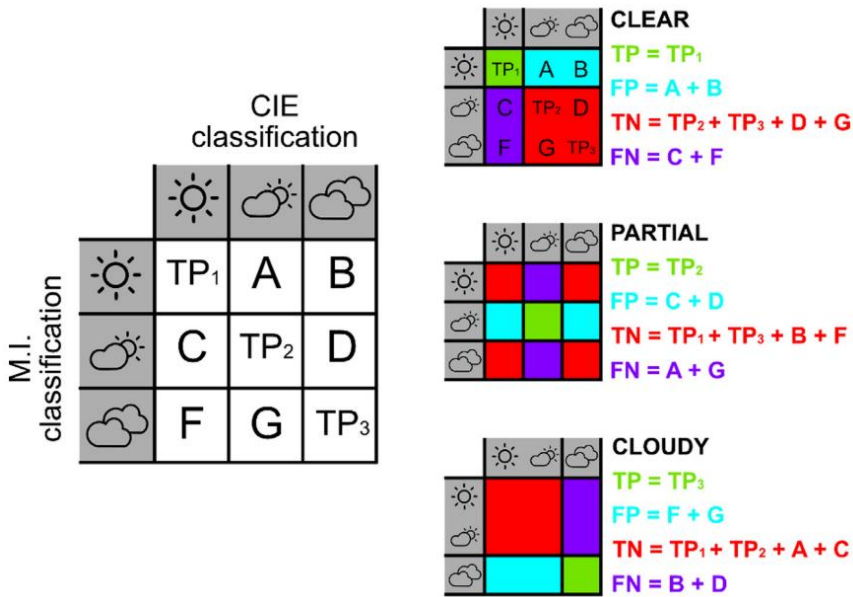


Figure 17. Multiclass to dichotomous transformation [69].

Table 10. Brief description of the goodness indicators.

Goodness indicator	Description	Formulation
Accuracy	Ratio of correct predictions amongst all the cases evaluated.	$A = \frac{TP + TN}{N}$
Jaccard Index	Ratio of correct predictions amongst all the cases evaluated.	$JI = \frac{TP}{TP + FP + FN}$
Cohen Kappa	It reflects the possibility that the algorithm agrees with the reference by chance.	$\kappa = \frac{A - p_e}{1 - p_e}$ <p>p_e is the expected agreement by chance. N is the total number of cases ($N = TP + TN + FP + FN$), and $p_e = \frac{(TP+FN)(TP+FN) + (TN+FP)(TN+FN)}{N^2}$</p>
Matthews	Correlation coefficient between the reference and the predicted classification.	$M = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FN)(TP + FP)(TN + FP)(TN + FN)}}$

All of the metrics are attempts to sum up the confusion matrix associated with the algorithm using only one number. Inevitably, the process is associated with a loss of information, because a four dimensional matrix is collapsed into one number. Each dimension attempts to highlight one aspect of interest. However, they are correlated in some way because the starting data are all the same. In fact, the ranking offered by each one is very similar.

The different rankings produced by all the metrics are combined to arrive at the best MI for classifying sky cloudiness. At the end of the process, four rankings are thus obtained, one per confusion matrix. The one designated as best model holds the best positions in all of them. So, by using all of them simultaneously, the deficiencies of the metric and its biases are avoided.

Individual analysis of sky classification with the MIs shows that practically no index is able to identify the high percentage of clear skies indicated by the CIE classification. The Klucher clearness index (FK) is the one that identifies a higher percentage of clear skies, and on the opposite extreme, the Batlles clearness index identifies practically all cases as overcast. Δ , Si , OFP , and F , identify most of the cases as partial cloudiness.

Table 11 summarizes the average ranking obtained by the MIs. ϵ was in first place in the Jaccard Index, κ and Matthews evaluations. It was only surpassed by the FK index in the accuracy ranking. Bearing in mind the similarity of the numerical values obtained by ϵ , FK and OFP in the evaluation indices, these three MIs may be identified as the best classifiers of sky cloudiness. When studying only the capacity of the MIs to identify clear skies, FK obtained the first position in all the indicators. None of the MIs under analysis have simultaneously shown good results in the rating variables, for the identification of either partial or cloudy skies.

Table 11. Ranking average summary [69].

-	k_d	k_b	k_t	F	k_k	FK	ϵ	Δ	OFP	S_i
Accuracy	5	7	4	9	10	1	2	8	3	6
Jaccard	5	7	4	9	10	2	1	8	3	6
Cohen Kappa	5	7	4	8	9	2	1	10	3	6
Matthews	5	7	6	9	10	3	1	8	2	4
Total	5	7	4	9	10	2	1	8	3	6

As main conclusions of the work, the metrics for the cloudiness classification have shown that the performance of the MIs may, at best, be considered as “fair”. Classification accuracy reached values of over 70% for cloudy skies. However, this metric can lead to misunderstandings, as it accounts for the correct classification of true positives and true negatives. As Table 11 shows, the best MI for the sky classification was ϵ followed closely by FK and OFP indices. k_k and F showed the worst results for the classification of cloudiness of the skies into three categories. No correlation was observed between the mathematical complexity of the MIs or the number of input parameters and the classification result of the MIs. As can be observed, the original definition of the intervals is one of the main factors that justifies the difference in the performance of the MIs. The experimental data of the present study were taken from a different location other than the one used in the original studies for the definition of each MI. However, if subtle language differences are neglected, the definition of “clear” or “cloudy” sky should be independent of location. Attending to the metrics, regardless of the accuracy, the results were quite similar. The average accuracy almost reached 70 % in FK , and was above 50 % in all the MIs. However, this result is overoptimistic because of the accuracy paradox. The average value was below 33% for the Jaccard Index and below 25% for Cohen’s Kappa and Matthews correlation coefficient.

As demonstrated, the isolated use of the MIs would not be advisable for the cloudiness skies classification.

2.5. Establishment of alternative methods to sky scanner for CIE standard sky classification from the use of meteorological indices (MIs) and sky images (II): Feature Selection for CIE standard sky classification (Paper III)

CP-II focused on the lack of homogeneity in the definition of sky categories by MI ranges and concluded that the performance of the MIs, each individually, for sky classification using CIE standard as reference, may, at best, be considered as “fair”, even in a reduced classification with only three sky types: the clear-partial-overcast (CPO) classification. Paper III studies the possibility to combine various MIs to improve the sky classification through meteorological indices. For that, a previous and detailed study on comparing and analyzing the information that each MI offers, was mandatory to remove irrelevant and redundant information for the task of sky types of definition [78].

Machine learning algorithms were used for this task. If the dataset has enough samples, supervised machine learning (SML) tools can identify patterns and relationships between inputs and outputs. So far, SML techniques have shown great success by the scientific community for sky classification, as Artificial Neural Networks (ANNs) [38] or classification trees [79].

The workflow of the SML tools is described in Figure 18. Firstly, the quality control follows a process of filtering and analysis. Secondly, a Feature Selection (FS) procedure detects irrelevant or less important features. Its objective is to remove the inputs that contribute little or nothing to the SML task. Following acceptable FS, the algorithm is trained using part of the input data set (training set), typically 80% of the total, using the remaining 20% for validation tests. Training set data and test set data are exchanged as many times as necessary, until the algorithm is considered validated.

FS is one of the core concepts of ML that will impact on the performance of the developed model, improving its precision and reducing its complexity and overfitting as well as its runtime.

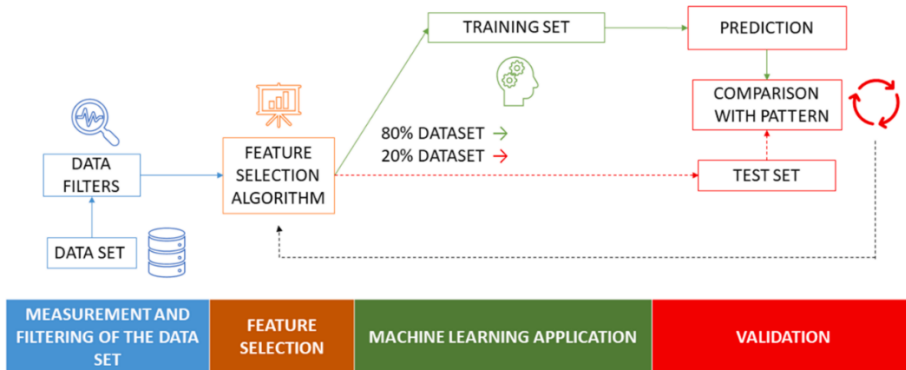


Figure 18. Workflow of Supervised Machine-Learning tool [27].

The main objective of this study is to determine, through a FS procedure, the most suitable MIs and their precise number for the optimization of the sky classification algorithms. Forty-three MIs were included (Table 12), and calculated from half-hourly experimental data records collected at Burgos, Spain, between September 2016 and December 2019. The following FS criteria were selected: Pearson [80], Permutation Importance [81], Recursive Feature Elimination [82], and Boruta [83].

The study reports an extensive review of the MIs that define different sky conditions and features that are suitable for sky classification. Structured and rigorous FS procedures can determine the usefulness of the information in these indices, with a high degree of success, for the problem of sky classification, the informative equivalence between some of the MIs, and the number of MIs that may be needed for sky classification in line with the CIE standard. The classification tree algorithm was selected, above other ML classification algorithms, due to the transparency of the results it can obtain. The classification tree produces a diagram that can be more easily understood than those produced by other ML techniques such as, Support Vector Machines, Neural Networks, Random Forest, and Gradient Boosting, traditionally known as “black boxes”.

Table 12. Definition of the 43 MIs reviewed as candidates for sky classification. L_0 is the Luminous solar constant (133.8 klx) and I_{SC} is the solar constant (1361.1 W/m²) [27].

Ratio Zenith Illuminance to horizontal diffuse Illuminance	Ratio Global Illuminance	Ratio Diffuse Illuminance	Luminous Turbidity index	Vertical Component	Sky
$LERT = \frac{L_z}{L_{xDH}}$	$C_1 = \frac{L_{xGH}}{L_{0h}}$	$C_2 = \frac{L_{xDH}}{L_{0h}}$	$t_v = \frac{\ln(\frac{L_{0h}}{L_{xBH}})}{A_V M_V}$	$VSC = \frac{L_{xDV}}{L_{xDH}}$	
Normalized Global Illuminance	Normalized Beam Illuminance	Normalized Diffuse Illuminance	Cloudless Index	Igawa's Sky Index	
$EVGM = M_V \frac{L_{xGH}}{L_0}$	$EVSM = M_V \frac{L_{xBH}}{L_0}$	$EVDM = M_V \frac{L_{xDH}}{L_0}$	$Cle = \frac{1 - k_d}{1 - Ces(M)}$	$S_i = \frac{RaGH}{0.84 \frac{I_{SC}}{M_V} e^{-0.0675 M_V} + \sqrt{Cle}}$	
Direct Fraction	Cloud Cover	Illuminance Cloud Ratio	Irradiance Cloud Ratio	Standard Cloud Ratio	
$k_b = \frac{RaBH}{RaGH}$	$CC(\% \text{ Clouds})$	$C_V = \frac{L_{xDH}}{L_{xDH} + L_{xBH}}$	$C_e = \frac{RaDH}{RaDH + RaBH}$	$Ces = 0.01299 + 0.07698 M_V - 0.003857 M_V^2 + 0.0001054 M_V^3 - 0.000001031 M_V^4$	
Umeyiya's Cloud Ratio	Relative Heaviness	Clear Sky Index	Clearness Index	Zenith Angle Independent Clearness Index	
$CLDV = \frac{L_{xDH}}{L_{xGH}}$	$\Omega = \frac{L_{xGH}}{\sin \alpha_s}$	$\frac{k_c = \frac{L_{xGH}}{0.84 \frac{I_{SC}}{M_V} e^{-0.0675 M_V}}}{L_{xGH}}$	$k_t = \frac{RaGH}{I_0 \sin \alpha_s}$	$k_{t2} = k_t (1.031 e^{-\frac{1.4}{0.9+9.4M_V}} + 0.1)^{-1}$	
Luminous Efficacy	Perez sky's brightness	Perez's Clear sky index	Original Perraudau's Nebulosity Index	Perraudau's Nebulosity Index	
$K = \frac{L_{xGH}}{RaGH}$	$\Delta = \frac{RaDH M_V}{I_{sc} \epsilon_0 \sin \alpha_s}$	$\epsilon = \frac{(RaDH + RaBH)}{(\frac{RaDH}{+ 1.04 Z^3}) / (1 + 1.04 Z^3)}$	$OFP = (1 - k_D) / (1 - E_{clear} / (E_{clear} + RaBN))$	$FP = \frac{(1 - k_D)}{(1 - 0.12037 \cdot (\sin Z_s)^{-0.82})}$	
Klucher's Clearness Index	RaBH, RaDH, RaGH, RaGVS, LxGH	Optical Mass	Scattering Angle	Turbidity	
$FK = 1 - k_D^2$	-	$M_V = (\sin \alpha_s + 0.50572(\alpha_s + 6.07995)^{-1.6364})$	$\chi = \arccos(\cos Z_s \cos Z_p + \sin Z_s \sin Z_p \cos \phi_p - \phi_s)$	$TURV = \frac{1 + 0.0045 M_V}{0.1 M_V \ln(\frac{L_0}{L_{xBH}})}$	
T, RH, WS, L_z, α_s	Diffuse fraction	Clearness Function	Modelled direct solar irradiance	Permeability	
-	$k_d = \frac{RaDH}{RaGH}$	$F = \frac{RaGH - RaBH}{I_{sc} \epsilon_0 \sin \alpha_s}$	$P_e = \frac{RaGH - RaDH}{\sin(\alpha_s)}$	$PERV = M_V \sqrt{\frac{L_{xBH}}{L_0}}$	

There are two different types of FS algorithms. If the FS works without any clustering algorithms, that means, only with a statistical index, the FS typology is known as Filter Methods (FM). Otherwise, when the process relies on clustering algorithms, the FS is a Wrapper Method (WM). Consequently, FM is independent of the SML afterward [84, 85]. On the other hand, WM executes machine learning to generate a ranking and perform a global evaluation of the entire set. Therefore, the score is not universal and does not suit any other ML algorithm [86]. It means that WM will generate for each ML algorithm a different ranking. A brief description of the implemented FS methods is shown in Table 13.

Table 13. Summary of several FS by category. (UR: Universal results).

	Method	Short Description	UR
FM	Pearson correlation criterion (P)	A statistical score, a near 0 value implies a weak or null correlation between two variables: X_i and X_j . In contrast, a near value to 1 or -1 implies a strong correlation: $r(X_i, X_j)$	Yes
	Permutation importance (PI)	It analyzes how the score of the prediction model decreases when the data of a single variable is randomly permuted (generating random noise).	No
WM	Recursive feature elimination (RFE)	It removes the weakest features until a specified number of variables is reached.	No
	Boruta (BOR)	It determines if a feature is a truly important attribute but does not aim to avoid redundant information.	No

Using the FS methods described above, the MIs to be included in the classification algorithm have been selected from the variables summarized in Table 14. After this, one decision tree with each of the set of MI's selected by each of the FS methods was programmed. The analysis of the

performance of the decision tree is carried out through confusion matrices, using the goodness indices Precision (Pr), Recall (Re), and $f1$ factor ($f1$), defined in equations (Eq.19) to (Eq.21), to analyze the performance of algorithm in the classification task. Pr is the probability that a positive prediction was correct. Re is the percentage of correctly detected positive cases. Both indices are independent, and $f1$ factor merges them.

$$Pr = \frac{TP}{TP + FP} \quad (\text{Eq.19})$$

$$Re = \frac{TP}{TP + FN} \quad (\text{Eq.20})$$

$$f1 = \frac{2}{1/Pr + 1/Re} \quad (\text{Eq.21})$$

Table 14. Summary of the MIs selected by each one of the FS algorithms [27].

Feature Selection	MI selected	Number
P	FP, CC	2
PI	$CC, CLDV, FP, Lz, PERV, RaGVS, TURV, VSC, k_d, k_t, t_v$	11
RFE	$CC, LxGH, VSC, WS, OFP, C_{le}, k_t, K, \chi$ $C_1, C_2, CC, CLDV, C_e, \chi, Cle, C_v, \epsilon, EVDm, EVGM, EVSM$	9
BOR	$F, FK, FP, LERT, L_z, OFP, \Omega, PERV, P_e, RaBH, RaDH$ $RaGVS, \Delta, S_i, TURV, VSC, k_b, k_c, k_d, k_t, k_{t2}, t_v$	34

The classification trees for CIE Standard Sky Classification from the MIs selected by Pearson, Permutation Importance, RFE and Boruta FS procedures are shown in Figures: Figure 19-Figure 22. Starting in the main left node, if the condition is met, the path of the upper branch is followed and, if not, the path followed is the one indicated by the lower branch. Evaluating each node consecutively, the sky conditions are obtained. The number inside the nodes represents the number of samples inside each partition. The number of binary partitions or levels of the classification tree is a previously set parameter. In this work all the classification trees have four levels. An increased number of levels might

increase the precision of the classification algorithm in the same way as complexity. The starting MI and the number of levels of the classification tree were selected following the Gini [87] and the Entropy [88] criteria.

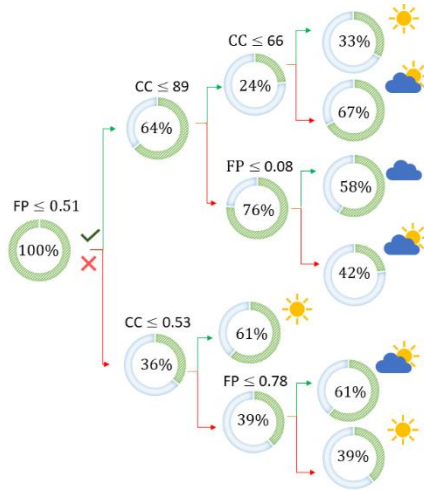


Figure 19. CIE standard sky classification tree (MIs selected with the Pearson FS method) [27].

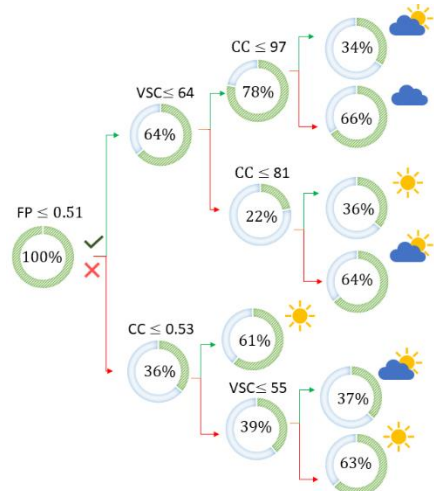


Figure 20. CIE standard sky classification tree (MIs selected with the PI FS method) [27].

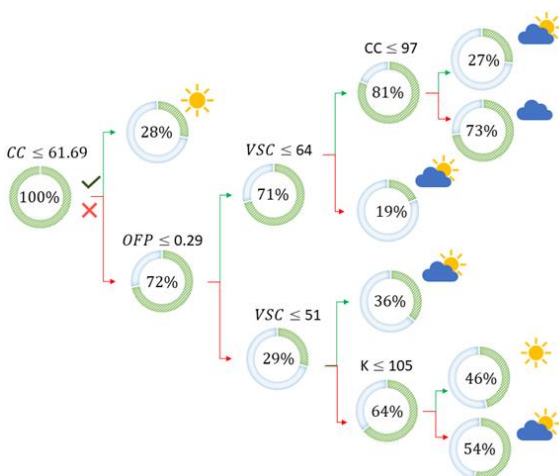


Figure 21. CIE standard sky classification tree (MIs selected with the RFE FS method) [28].

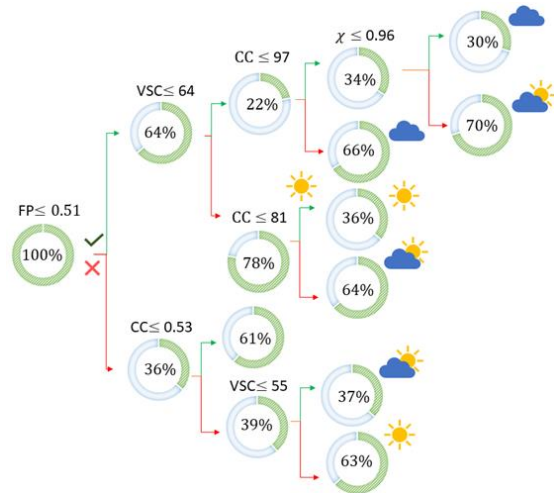


Figure 22. CIE standard sky classification tree. (MIs selected with the Boruta FS method) [27].

FP and *CC* are MIs selected by the Pearson FS method for the classification tree (Figure 19). Both MIs are related to the cloud conditions, through the diffuse horizontal fraction (ratio diffuse horizontal irradiation to global horizontal irradiation) and the percentage of sky covered by clouds, respectively. The CIE Standard decision tree obtained from the variables selected by the Pearson FS method identifies the clear sky type by one of these cases: a) $FP > 0.51$, and $CC \geq 0.53$; b) $FP > 0.78$; c) $FP \leq 0.51$ and $CC \leq 0.66$.

Although the PI FS methods selected 11 MIs for the CIE sky classification (Figure 20), only three were used in the four-level classification tree: *FP*, *CC*, and *VSC*. The Vertical Sky Component, *VSC*, linked the classification to the daylighting. The classification tree obtained with the MIs selected by the RFE FS method (Figure 21) started with the *CC*, a variable which directly classified the skies as clear if $< 61.7\%$. On the second and third levels of the classification tree, original Perraudeau's Index, *OFFP*, and *VSC*, were evaluated. At the last level, the MI selected to fit the classification was luminous efficacy, *K*. Again, two of the MIs were related to daylighting (*VSC*, *K*) and *OFFP* and *CC* were related to cloud coverage. Boruta FS methods selected 34 MIs for the CIE standard sky classification (Figure 22), but four were necessary to build the four-level classification tree. The sky classification started by evaluating Perraudeau's nebulosity index, *FP*. At the second level, cloud cover, *CC*, and the vertical sky component, *VSC*, were introduced. Finally, the scattering angle, χ , a geometrical variable, was investigated.

Figure 23 shows the results of the *Pr*, *Re*, and *f1* metrics obtained for each of the decision trees shown in figures (Figure 19-Figure 22) for CIE standard cloudiness classification. The dispersion of the value of the metrics varies with the sky conditions. For clear conditions, *f1* ranged from 65% to 87% from Pearson and Boruta method, respectively, but all indices exceeded 65%. The case of partial skies identification was the worst overall as the values of all indices felt between 55-70%. The irregularity of partial skies causes high variability in the MIs valued, and the dependency of

cloud cover with respect to the Sun might explain this fact for partially covered conditions. Specifically, the RFE method yielded the closest values of the three metrics. In addition, the weighted-average $f1$ of the four classification trees yielded results in the range 74-77%, in which the highest score corresponds to the RFE method, quite close to the BOR and PI FS procedure with similar $f1$. The Pearson method, which only selected 2 MIs, showed the lowest $f1$ value. Nevertheless, the results highlighted no significant advantages to the proposal of the others FS models. The best results for all metrics were obtained in overcast sky conditions, with Pr and Re above 85% and reaching 90% in the case of the classification tree that applied the four MIs selected for the RFE method (CC, OFP, VSC, K). The classification trees that used the MIs selected by both PI (FP, VSC, CC), and BOR (FP, VSC, CC, χ) obtained the same results. The simplest classification tree, from the two MIs selected by the P method, (CC, FP), yielded worse $f1$ metrics.

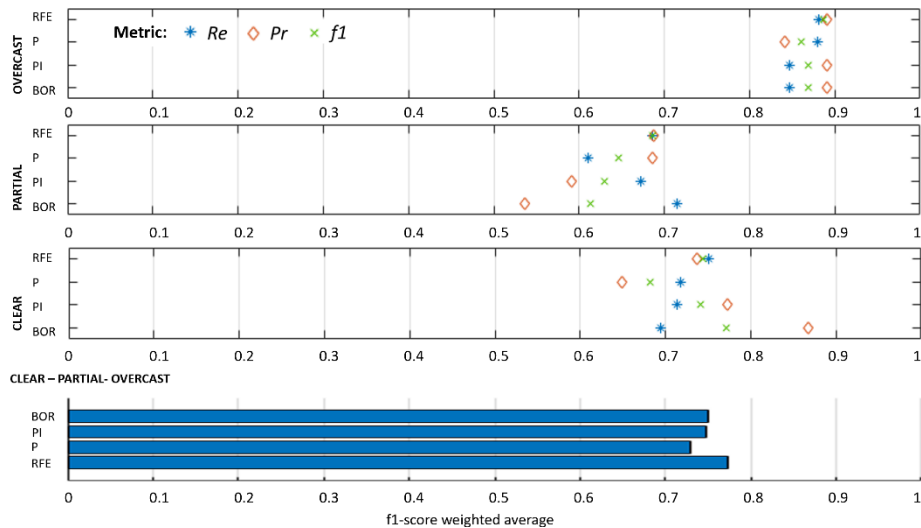


Figure 23. Precision, Recall, and $f1$ indices calculated for Clear-Partial-Overcast classification each classification tree versus CIE standard.

This study has highlighted the usefulness of the FS procedure for adequate determination of MIs for sky classification in accordance with the CIE Standard classification, as an alternative to the use of sky-scanner

devices. The maximum number of MIs can be identified with FS for use as an input for the ML algorithm, avoiding the introduction of redundant and useless information. The FS results, processed in a classification tree to test their validity, confirmed that the intervals of definition of the MIs for each sky type were close to the intervals that were established in each study for the individual use of each MI.

2.6. Establishment of alternative methods to sky scanner for CIE standard sky classification from the use of meteorological indices (MIs) and sky images (III): Pixel-Based Image Processing for CIE Standard Sky Classification through ANN

A new alternative to sky scanner devices is proposed in Paper IV: the use of sky images and Artificial Neural Networks (ANNs). Currently, high interest has been expressed in calibrated sky luminance maps for sky classification and cloud detection [89–92]. A digital camera equipped with a fisheye lens can map at a higher resolution than commercial sky scanners, and High Dynamic Range (HDR) images can capture the full sky luminance range [93]. Paper IV studies the use of digital sky images for the classification of sky conditions in accordance with the CIE Standard General Sky Guide and analyzes adequate image-processing methods that highlight key image information before applying ANN classification algorithms.

Image-processing methods can help to improve the efficiency of ANN for sky classification, overcoming misclassification due to cloud cover and simplifying the ANN structure. Some studies have focused on color space, i.e., the RGB (red, green, and blue), as basic standard for computer images, and the Hue Saturation Value (HSV), which adapts the image to the color perception of the human eye. The focus of others has been on the modification and combination of the original monochromatic channels, known as the spectral features. A third alternative, texture filters, uses the gray distribution of pixels and their spatial neighborhoods to identify objects and regions. Table 15 classifies, into the categories Color Space, Spectral and Texture feature, all the reviewed image-processing methods. Figure 24 contextualizes these methods of image processing and shows an example of the results of the application of these methods applied to images of sky conditions classified as clear, partial, and overcast, following the CIE taxonomy.

Table 15. Summary of pixel image-processing methods [28].

Type	Name	Purpose	Formulation	Ref.
Color Space	RGB	Image visualization based on primary colors	R: Red channel. G: Green channel. B: Blue channel.	[94]
	R, G, B*	R: Red channel; G: Green channel; B: Blue channel.		
	GS	Gray Scale. Gray-Scale intensity image (calculated from the RGB image)	$rgb2gray^1$	[95]
	HSV	Image visualization based on the perceptions of the human eye	H: Hue channel. S: Saturation channel. V: Value channel.	[89]
	H, S, V*	H: Hue channel; S: Saturation channel; V: Value channel.		
Spectral feature	RAS	Image processing for removing atmospheric scattering	$RAS = Y - (L - D)$ $L = \max(R, G, B)$ bright. $D = \min(R, G, B)$ dark of the channel; $Y = 0.299 \cdot R + 0.587 \cdot G + 0.114 \cdot B$	[96, 97]
	Y, L, D*	Y: Panchromatic channel; L: Bright channel; D: Dark channel		
	RBR	Background due to atmospheric scattering (Red Blue Ratio).	$RBR = R/B$	[97]
	RBD	Difference between Red and Blue channels.	$RBD = R - B$	[98]
	NRBR	Blueness of the sky. High robustness to noise (Normalized Red-Blue Ratio).	$NRBR = \frac{R - B}{B + R}$	[98]
	C1	Combination of RBR, RBD and NRBR channels.	$RBR - RBD - NRBR$	[90]
	ARGD	Reduce sunlight interference. (Adjusted Red Green Difference)	$ARGD = k \cdot R - G$ $k = 1.7$ (weight of the red channel)	[99]
	C2	Combination of RBR, ARGD and NRBR channels.	$RBR - RBD - ARGD$	[99]
Texture feature	LR	For distinguishing edges and contours (Local Range)	Rangefilt1	
	STD	Obtained from the standard deviation in each neighborhood.	Stdfilt1	[100]
	EM	Randomness of the image (Entropy Matrix).	Entropyfilt1	

¹ Matlab function ([100]), * Subchannels

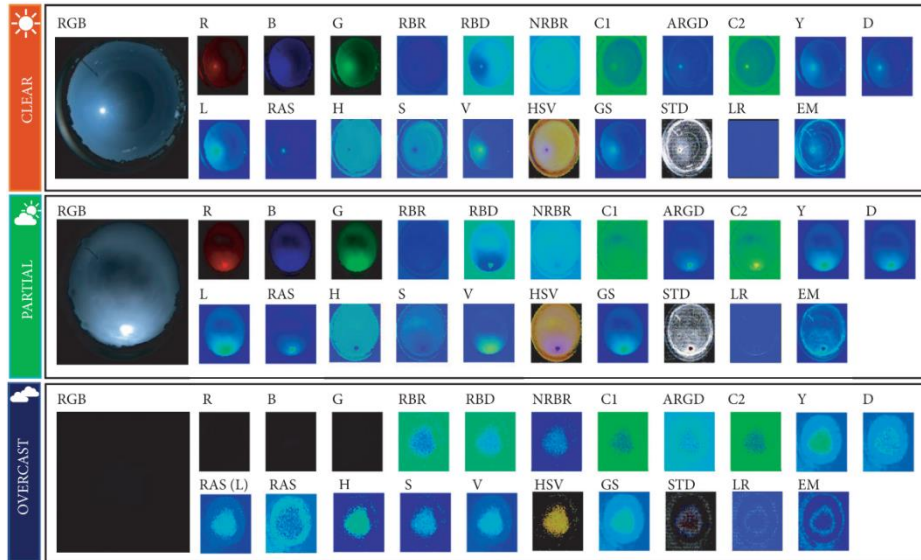


Figure 24. Results of the image-processing methods applied to clear, partial, and overcast CIE standard sky types [28].

The workflow followed for the analysis of image-processing algorithms for CIE standard sky classification is described in Figure 25. Firstly, experimental data acquisition took place at the meteorological station described in Table 1, from November 1, 2016, to March 31, 2020. 1,500 images were selected from the experimental dataset (more than 80,000 sky images), 100 from each CIE sky category, which were characterized by greater concordance with the CIE pattern for that category. The experimental dataset was therefore composed of one hundred sky images cataloged as CIE standard sky categories using the methodology described in Paper I. This sky classification was used as a reference for the sky conditions. Secondly, in the step of Data Processing, the original sky images were processed using the twenty-two different image processing methods chosen for the study and summarized in Table 15. Finally, in the step of ANN, the image set modified by each image-processing algorithm, was introduced in ANN for CIE Standard Sky Classification.

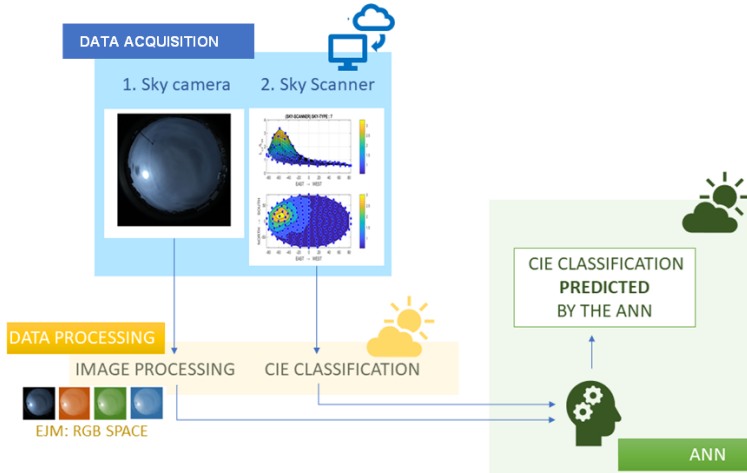


Figure 25. Workflow for image processing. Example with RGB space [28].

ANNs classification models serve to classify input information into certain categories or targets. A Supervised Machine Learning (SML) neural network is required for CIE standard sky classification when the sky types are previously known [29]. The model works efficiently when the prediction matches the target. Modelled on the biological concept of neurons, ANN is a very powerful technique for classification problems. Figure 26 shows a conventional ANN structure, which consists of an the input layer, a set of several hidden layers, and the output layer.

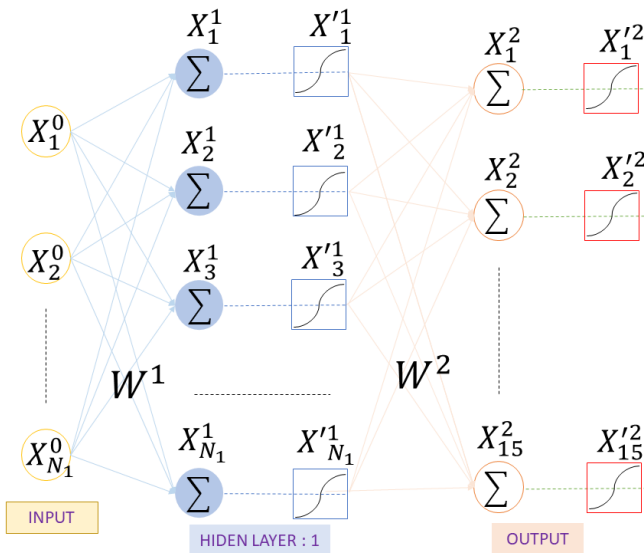


Figure 26. ANN system architecture [28].

The information from the neurons of the input layer (X_i^0) crosses the hidden layers (one in this work), following unidirectional connections, to the output layer that has one neuron ($X_i^{2'}$) per target. Each processing center or neuron is adjusted to the other neurons through an interactive process, using (Eq. 22). The Scaled Conjugate Gradient method (SCG) [101] fits the weights for each iteration.

$$X_i^n = W^n X^{n-1} + B \quad (\text{Eq. 22})$$

Where W_n is the weighting matrix, X^{n-1} are the input variables, and B is the Bias. X_i^n is the input of the activation function. The neuron generates the output, $X_i^{n'}$, through the activation function, $f(X_i^n)$, (Eq. 23), given by the hyperbolic tangent sigmoid transfer function in this study, as shown in [102].

$$X_i^{n'} = f(X_i^n) = \frac{2}{1 + \exp(-2 \cdot X_i^n)} - 1 \quad (\text{Eq. 23})$$

SML requires three datasets: training, validation, and test datasets. The training group is used to determine the weighted matrix and the bias in an iterative process. The training is over when the results of the performance of the resulting model, calculated using the validation set, reach the desired quality. The test data group is used to calculate the performance of the model. Random dataset division is crucial to achieve a reliable performance. A conventional training dataset is randomly selected and consists of 70% of the total data, while the validation set and the test set each represent 15%, respectively.

There is no standardized procedure for establishing the most effective number of neurons and hidden layers [103], so experimentation or tuning is needed. In this study, several trials were performed in which the number of neurons (1–100) was varied, searching for the best accuracy of the ANN. The accuracy index, A , is defined as (Eq. 24).

$$A = \frac{TP + TN}{N} \quad (\text{Eq. 24})$$

Figure 27 shows the improvement in ANN accuracy, $\Delta(A)$, for CIE standard sky classification, using the results of each image-processing method as input, as summarized in Table 16, over ANN accuracy obtained with the original RGB images as input. Figure 27 illustrates that HSV is a slightly better color space than RGB for CIE standard sky classification using images, with a small improvement in accuracy (0.66%) over RGB image processing. The GS color space and the RGB space were equally accurate. The use of the R, G, and B monochromatic channels also improved the accuracy of the ANN for CIE standard sky classification, being the G channel the most suitable for this task. The accuracy of the ANN fitted using the individual channels, H and S, worsened over the RGB color space, while the V channel significantly improved ANN accuracy. In the spectral feature category, the RAS processing method worsened the sky classification accuracy of the ANN. However, channels Y and L showed better behavior for sky classification, although they used more neurons in the hidden layer. Among the rest of spectral feature channels, only RBD and C1 significantly improved ANN accuracy. With regard to the texture filters, EM showed little or no advantages over the use of the RGB color space and the other two filters, LR and STD, impaired the accuracy of the resulting neural network.

Table 16 shows the accuracy and structure (number of neurons in the hidden layer) of the selected ANN for each image-processing method. As can be observed from Table 16, the number of neurons in the hidden layer, never increased the accuracy of the ANN, as can be seen from the use of image-processing methods Y and V.

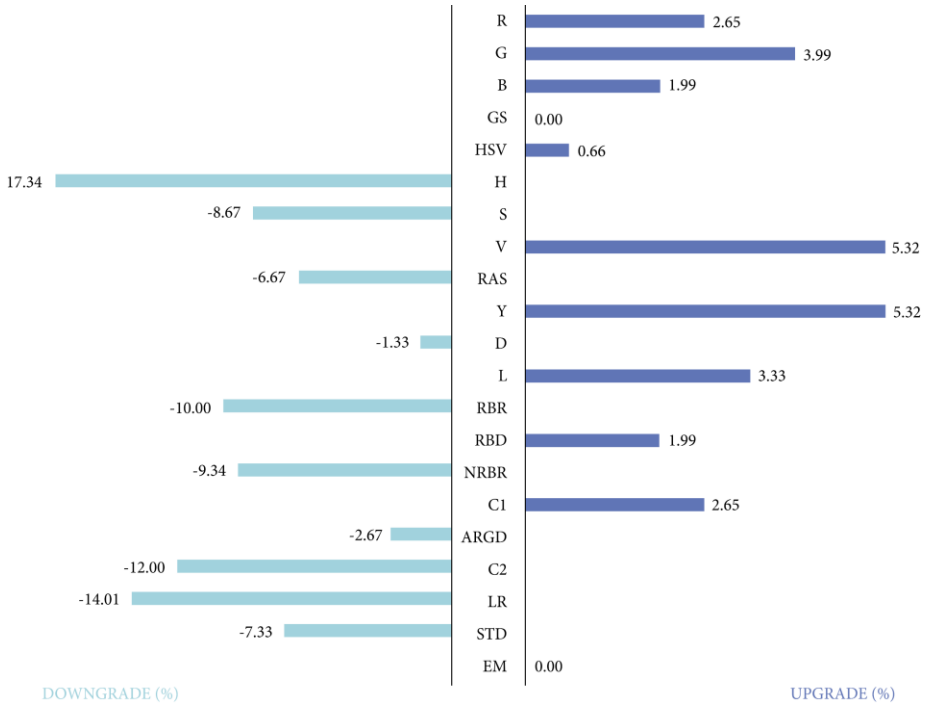


Figure 27. Improvement in ANN accuracy, $\Delta(A)$, for CIE standard sky classification, using the results of each image-processing method as input, over ANN accuracy obtained with the original RGB images as input [28].

Table 16. Accuracy and structure (number of neurons in the hidden layer) of the selected ANN for each image-processing method [28].

Method	A (%)	Number of Neurons	Method	A (%)	Number of Neurons
RGB	66.67	84	D	65.78	21
R	68.44	66	L	68.89	74
G	69.33	58	RBR	60	45
B	68	36	RBD	68	98
GS	66.67	58	NRBR	60.44	27
HSV	67.11	40	C1	68.44	73
H	55.11	23	ARGD	64.89	17
S	60.89	27	C2	58.67	68
V	70.22	58	LR	57.33	39
RAS	62.22	25	STD	61.78	52
Y	70.22	95	EM	66.67	58

Figure 27 shows the results of each ANN classifying the skies into the fifteen CIE standard sky categories. A simpler classification into three categories (clear, overcast, and partial conditions) is often enough for many applications, such as luminous efficacy calculations [9] and lighting design in buildings [104]. The fitted results of the ANN sky classification for three categories are shown in Figure 28 and Table 17. Differences in accuracy are lower and only the G, the B, and the GS monochromatic channels and the spectral features L and C1 improved ANN accuracy. In all these classification cases, the number of neurons in the hidden layer was lower.

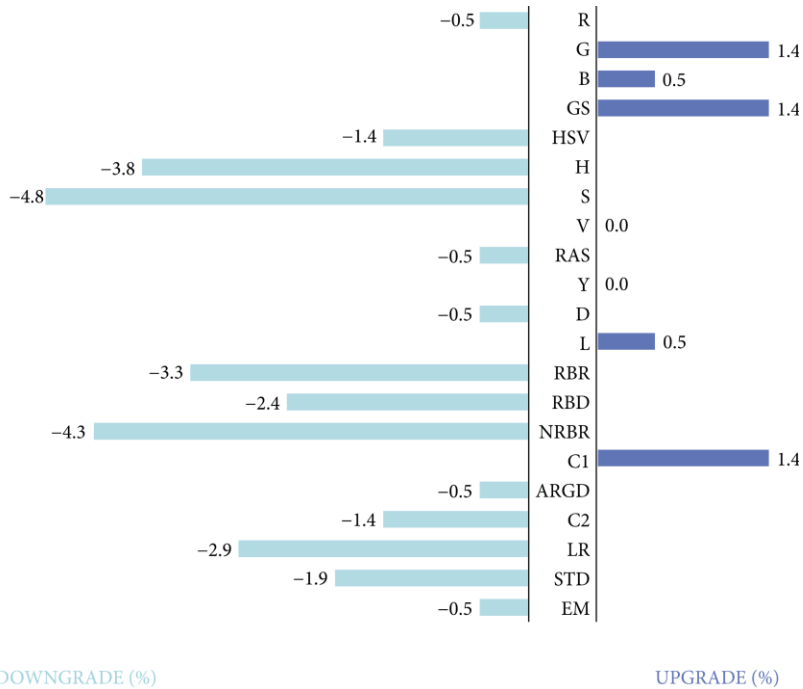


Figure 28. Improvement in ANN accuracy, $\Delta(A)$, for CIE standard sky classification in three sky categories: overcast, partial, and clear conditions, using as input the image processed by each image-processing method, over ANN accuracy obtained with the RGB images as input [28].

Table 17. Accuracy and structure (number of neurons in the hidden layer) of the ANN selected for each image-processing method [28].

Method	A (%)	Number of Neurons	Method	A (%)	Number of Neurons
RGB	93.33	71	D	92.89	15
R	92.89	48	L	93.78	15
G	94.67	52	RBR	90.22	23
B	93.78	52	RBD	91.11	43
GS	94.67	52	NRBR	89.33	23
HSV	92	11	C1	94.67	18
H	89.78	23	ARGD	92.89	19
S	88.89	73	C2	92	41
V	93.33	15	LR	90.67	81
RAS	92.89	90	STD	91.56	7
Y	93.78	21	EM	92.89	68

The accuracy index was used to group the goodness of fit of the ANN in all categories, although the fitted quality in each individual category was not processed. Figures: (Figure 29 – Figure 32) represent the confusion matrices corresponding to the 15 types of CIE standard skies. Figure 29 shows the confusion matrix of the ANN, calculated RGB-CIE sky classification for the test set (15% of the total dataset). It can be seen that the RGB-CIE classification with machine learning misclassified cloudy and partial skies: few matches are visible in the boxes along the diagonal line. In Figure 29, the CIE standard sky classification into three categories (clear, partial, and overcast sky conditions) is also presented. Those cases classified outside the corresponding category were designated as critical, i.e., clear skies were classified as either partial or overcast or vice versa.

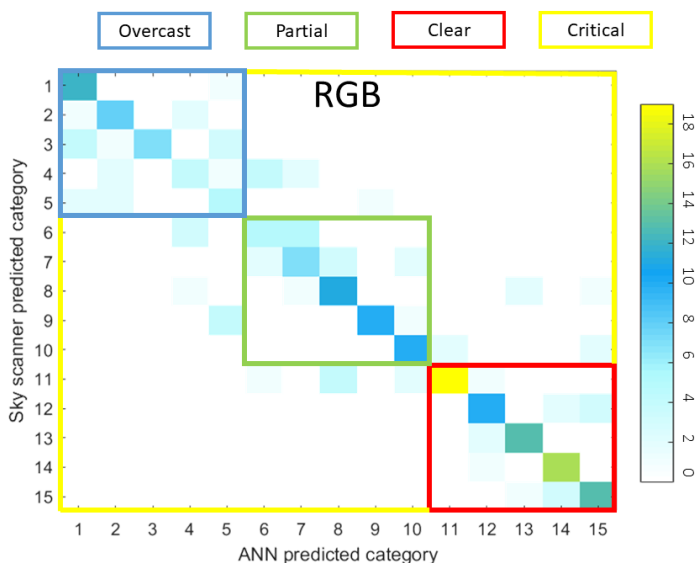


Figure 29. Confusion matrix for ANN-calculated RGB-CIE sky classification into fifteen categories and into three categories: Overcast (CIE standard sky types 1 to 5); Partial (CIE standard sky types 6 to 10); Clear (CIE standard sky types 1 to 5). Critical refers to cases classified out of category. The color scale shows the number of coincidences in each category [28].

The same information is shown in Figure 30 for CIE standard sky classification corresponding to the other color space processing methods under analysis. The red, the green, and the blue channels showed a similar behavior to the RGB color space. The red channel adequately classified CIE standard sky types 7 to 15, in other words, all clear skies and some partial sky types. The HSV color space showed a similar performance in all categories, in contrast to the RGB color space, in which the classification of clear sky types may be highlighted. Hue and saturation channels introduced too much noise, but the value channel showed good performance.

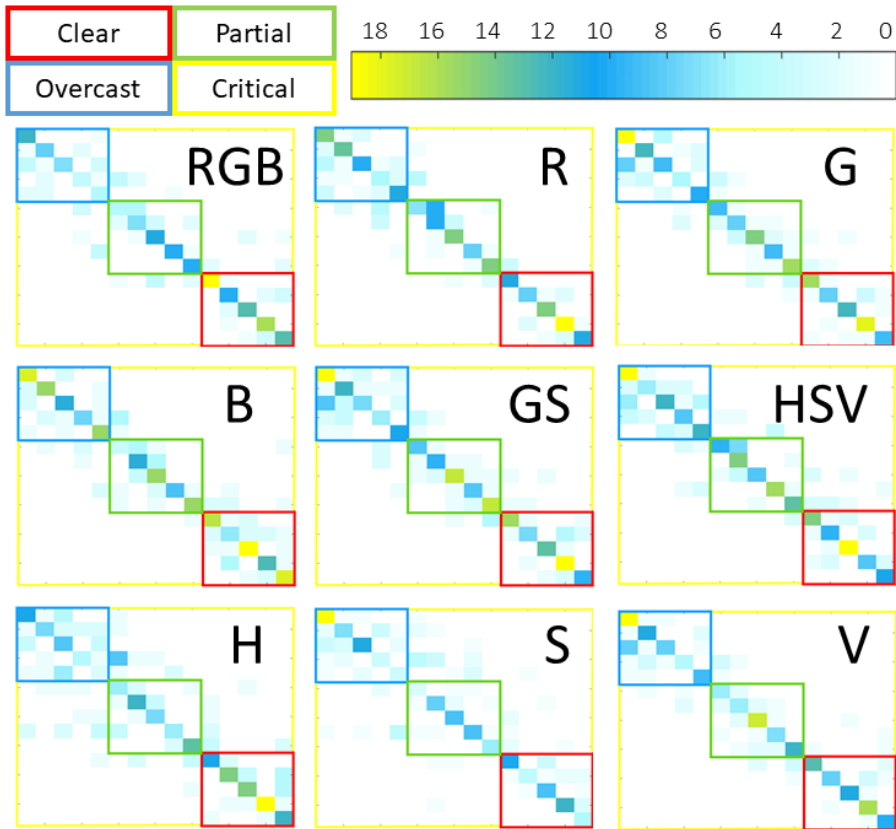


Figure 30. Confusion matrices color space image processing methods-CIE standard sky classification [28].

In Figure 31, the confusion matrices are shown for the CIE standard sky classification according to the spectral feature image-processing methods. The RBR and NRBR spectral features introduced noise, but the resultant combination, C1, reduced misclassification, improving the traditional RGB color space. Therefore, it appears to be an adequate alternative image-processing method for CIE standard sky classification using sky camera images. The RAS channel theoretically removed atmospheric scattering, but the confusion matrix never reflected a better performance than the RGB color space. The confusion matrix has demonstrated that it cannot distinguish the CIE sky types 1, 3, and 5. The RAS method also introduced too much noise in cloudy-to-partial sky types.

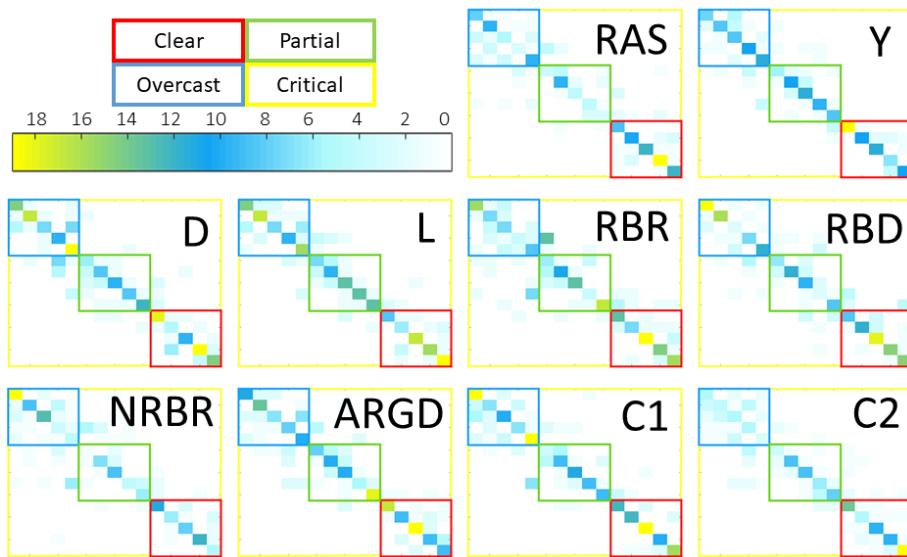


Figure 31. Confusion matrices spectral feature image processing methods-CIE standard sky classification [28].

Finally, the confusion matrices are shown in Figure 32 for texture filter processing methods-CIE standard sky classification with ANN. As can be seen, all texture channels performed well, especially the EM channel, while LR largely failed for CIE standard sky classifications partial and overcast.

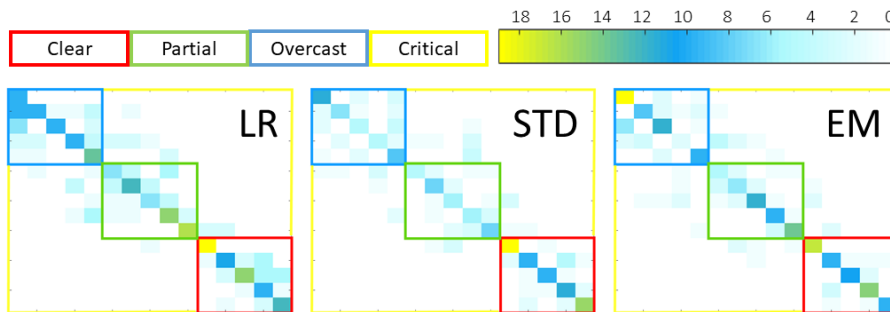


Figure 32. Confusion matrices for texture filter image processing methods-CIE standard sky classification [28].

A detailed study for the CIE standard sky classification into three categories is presented in Figure 33, where the confusion matrices presented in Figure 29 to Figure 32 are divided into four submatrices: overcast (CIE standard sky types 1 to 5), partial (CIE standard sky types 6 to 10), clear (CIE standard sky types 1 to 5), and critical, that refers to cases classified out of category. The red line indicates the RGB result, taken as a baseline for accuracy improvements, $\Delta(A)$. RBD, D, and B showed the best performance for the detection of overcast skies, increasing the performance of each respective ANN. G, S, and GS achieved better results for the detection of partial skies and clear skies were also in the same category in which the conventional RGB color space achieved its best performance. Some channels highlighted certain sky types but drastically failed to classify other types. The blue channel saturated in clear skies, to such a point that its performance was almost the worst for clear skies detection. This behavior was also noted for the D channel.

Unfortunately, no image preprocessing method drastically improved the RGB classification in the three subcategories simultaneously (clear, partial, and overcast conditions), as Table 17 showed. However, Y, green, red, RBD, V, and EM processing methods were prominent in one or two categories and their results were acceptable in all other categories, as shown in Table 18.

Table 18. Summary of the results [28].

CIE categories	One category	Two categories	Three categories
Clear	R, C1, Y, EM, and V		
Partial	G, S, HSV, and GS		
Overcast	R, G, B, RBD, NRBR, C1, ARGD, Y, D, L, S, V, and EM	R, G, C1, Y, S, V, and EM	R, G, Y, V, and EM
Critical	R, G, B, C1, C2, Y, D, L, RAS, V, HSV, GS, STD, LR, and EM		

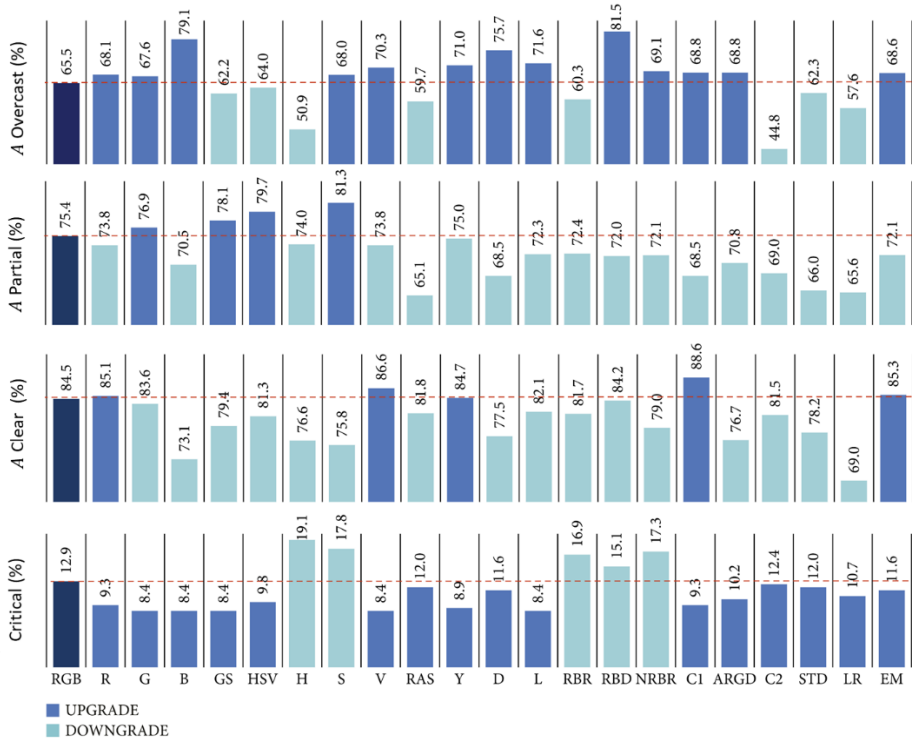


Figure 33. Accuracy for the CIE standard sky classification through sky images and ANN, using the different image-processing methods recorded in Table 15 [28].

Regarding the partial objective “Establishment of alternative methods to sky scanner for CIE standard sky classification from the use of meteorological indices (MIs) and sky images”, digital cameras equipped with fisheye lens can be used as alternatives to sky scanner devices for ANN-assisted CIE standard sky classification. The accuracy of the classification algorithm can be improved with adequate preliminary image processing that highlights the sky image information and optimizes the algorithmic structure.

Chapter 3: Conclusions and further work

Daylighting is a design concept of buildings recognized as a key strategy in reducing the energy consumption and improving visual comfort and well-being of their occupants [105]. In addition to improve the energy efficiency of buildings, it has been demonstrated that daylighting has positive effects on human health and productivity [106]. Energy standards strongly recommend incorporating daylighting strategies into building design [53, 54, 107, 108]. However, daylighting design is complex due to its changing nature. Likewise improving the occupants' life, daylight implementation is also related to technical and architectural solutions such as heating and cooling. Therefore, it is necessary to balance the positive health effects and energy savings in artificial lighting with the possible thermal and visual discomfort caused by excessive sunlight penetrations and glare [109].

The first step to model the interior illuminance in a building lies in the characterization of the exterior illuminance. This doctoral thesis addresses this key aspect through different strategies, such as luminous efficacy models and the determination of the luminance distribution of the sky.

The doctoral thesis "*Daylight modeling for energy efficiency and visual comfort in buildings*", was carried out through a compendium of articles. These publications actively contribute to the knowledge of daylight modeling and incorporate novel techniques that apply artificial intelligence for sky conditions characterization. Each paper carried out one specific objective of the doctoral thesis.

Paper I compared two different methods for sky classification according to the international standard ISO 15469:2004 (CIE). Several works have reported that the CIE standard provides a good overall framework to represent the actual sky conditions and covers the whole probable spectrum of skies found in nature [32, 41, 56, 110–113]. The CIE

classification has served throughout the doctoral thesis as the reference for sky conditions. It was demonstrated that the sky conditions in Burgos were predominantly clear along the year in several experimental campaigns. This result is highly relevant because traditional daylight modeling for indoor estimation would consider overcast conditions [46].

The sky luminance pattern, accessible throughout the CIE standard, allows obtaining both the vertical diffuse illuminance and the horizontal diffuse illuminance [114]. Their ratio, known as VSC, is used by city planners and developers to assess the impact of early-stage built constructions on the access to daylighting [52]. Paper II highlighted the good daylighting conditions in Burgos, in terms of VSC. VSC values higher than 100% were found in all cardinal orientations, more probably on east, and south-facing vertical surfaces, showing the high energetic and daylight potential of those orientations without underestimating the potential of the north and west facades, usually less considered for an energy-efficient design of buildings.

Complementary Paper I suggested a more considered traditional method for global illuminance characterization by luminous efficacy models. Thus, it evaluated eighteen classic models and, proposed a new one. This new model was fitted for either all sky types or particular sky type (clear, partly cloudy, and overcast). It employed the solar altitude, α_s , and the clearness index, k_t , as independent variables. *RMSE* values calculated for the proposed luminous efficacy model were comparable to other literature models and sometimes, including the *MBE* value, significantly improved the prediction, both for all sky types and for clear, partial, and overcast skies.

The problem of CIE standard sky classification through alternative methods to the sky scanner was addressed in the rest of the papers. The first alternative proposal focused on modeling from meteorological indices, MIs. Complementary Paper II (CP II) analyzed the capacity of 43 traditional MIs to classify skies according to the CIE standard and implemented machine learning algorithms for the classification using MIs

as inputs. The main result was the actual inability of the meteorological indices when used independently to perform a sky classification according to the CIE taxonomy, even if it only considers three categories for cloudiness classification. Therefore, Paper III addressed the possibility of combining several traditional MIs using artificial intelligence techniques for sky classification. In this regard, using feature selection (FS) procedures, the original set of 43 MIs was notably reduced, allowing the construction of different classification trees with high performance for the CIE standard sky conditions classification. Hence, FS found the most relevant variables and rejected the redundant ones. The MIs selected by the FS algorithms can be classified into three types: variables related to the cloud conditions, (FP, CC, OFP), others related to daylighting, brightness, or clearness conditions of the skies (VSC, K), and geometrical variables, such as α_s and χ .

As a second alternative to sky scanner for CIE standard sky classification, Paper IV addressed the use of digital sky images for the classification of sky conditions in accordance with the CIE Standard General Sky Guide. It also analyzed adequate image-processing methods that highlight key image information before applying ANN classification algorithms. As main result, Paper IV demonstrated that the accuracy of the classification algorithm can be improved with adequate preliminary image processing that highlights the sky image information and optimizes the algorithmic structure. For classifying the skies based on the images into the fifteen CIE standard sky types HSV and the monochromatic channels R, G, and B color spaces stood out over the traditional RGB. Only the V individual channel of HSV worked better than both HSV and RGB. Spectral feature channels Y and L showed better behavior for sky classification than the RGB color space, but they used more neurons in the hidden layer. Among the rest of the spectral feature channels, only RBD and C1 significantly improved ANN accuracy. Texture filters added no significant advantages over the RGB color space.

The results obtained in this doctoral thesis open a series of new work lines related to the sky classification using machine-learning algorithms, modeling illuminance, and irradiance and improved daylight calculations for visual comfort and energy efficiency in buildings, some of these lines already in course. These lines include the following:

- **Line 1: Modeling indoor illuminance and analyzing the relationship between SR on energy efficiency and visual comfort through dynamic metrics.** Average Daylight Factor (ADF) has been, traditionally, used for established the relationship between the indoor and outdoor illuminance [46, 49, 115]. Daylight factor, DF, is the ratio between indoor illuminance and outdoor illuminance, measured for a specific point or for an average of a space. By definition, DF should be calculated only under the CIE overcast sky condition. Therefore, ADF allows a stable characterization of the luminous environment by eliminating the dependence of the temporal variable and the orientation of the study. However, the annual average number of sunny days in Europe varies enormously. Other studies [105, 116] aims to consider the variability of natural lighting, using dynamic metrics, like the Annual Sunlight Exposure (ASE) or the spatial Daylight Autonomy (sDA). ASE is defined as the FOC at a point on the working plane that receives sunlight greater than a threshold value. SDA is the FOC at a point on the working plane where a minimum illuminance threshold can be maintained by daylight.

The procedures for indoor illumination calculations are traditionally classified into two types [117]: (a) radiosity calculation algorithms, which characterize each surface by its reflection coefficient and vision factors with respect to the rest of the surfaces and the working plane. SUPERLITE [118], INLUX [119], DIALUX [120], and RELUX [121] software tools implement this approximation; (b) ray-tracing algorithms, which develop the light trajectory that reaches a point giving each ray an intensity or weight that depends on the reflexions on every surface up to the measurement point. RADIANCE [122], and ECOTEC [123] are the most known ray-tracing algorithms.

DeLight is a compact daylight simulation tool [124], categorized as a radiosity calculation system [125, 126]. This tool calculates the indoor illuminance on the surfaces of interest, e.g. on the horizontal surface at a desk level. DeLight needs the sky luminance pattern, directly measured by a sky-scanner device or modeled by a sky luminance model, like Perez all-weather sky model [20, 127] or directly derived from the CIE standard sky [24].

A first work, already developed, analyzes the efficacy of the DeLight algorithm. An experimental campaign extended from 01 July 2021 to 21 August 2021 has been used for recording experimental indoor illuminance data in five points inside a scaled building. Figure 34 (left) shows the FOC (%) of the illuminance greater than 300 lux, that links with the recommended illuminance levels for human activity [53]. Figure 34 (right) is the FOC (%) of the illuminance values greater than 1000 lux, threshold linked to glare and the visual discomfort [53]. The inner circle refers to experimental data, the middle one to DeLight predictions from the experimental sky luminance pattern recorded by the sky scanner, and finally, the outer one refers to DeLight predictions from CIE luminance pattern

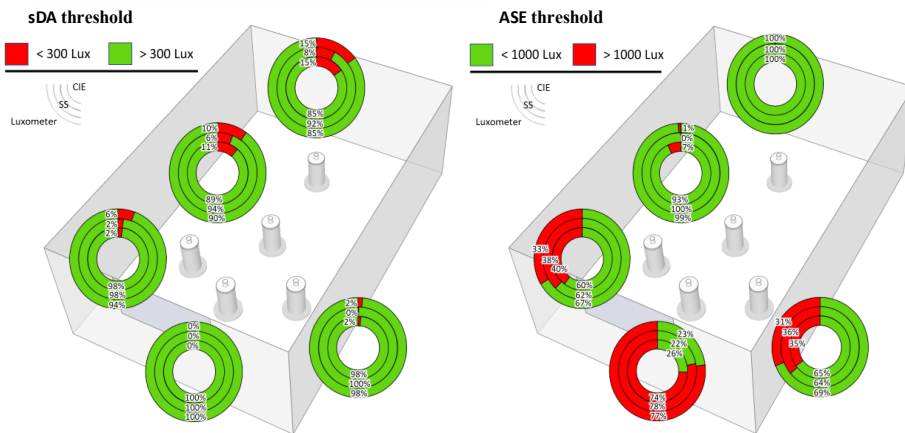


Figure 34. FOC of the illuminance level higher than 1000 lux and 300 lux, respectively, in the measurement points

This paper has been already accepted at the international congress: International Conference on Engineering Thermodynamics (12CNIT), in Madrid, Spain from June 29 to July 1, 2022, as an oral communication. Further details of the full paper are shown in Annex 2.

- **Line 2: Visual comfort characterization and energy efficiency in Buildings using Design-Builder.** As a result of the research stay carried out at the University of Malta, a case study based on a residential building under renovation in the European project Interreg Europe ZeroCO₂ [128, 129] has been done. A calibrated model of the building has been developed for analysing the visual comfort and energy efficiency following the international standards and improvement measures have been proposed. This project was carried out by a ray-tracing calculation system (Radiance), through the commercial software DesignBuilder [130]. The annual daylight modelling option of DesignBuilder runs through the Daysim simulation tool [122], a validated software for daylight analysis [130].

The following dynamic metrics are applied to improve visual comfort characterization: Useful Daylight Illuminance (UDI) [131] to evaluate the effectiveness of natural light that enters the building for different human activities; UDI autonomous (UDI-A) [131] to analyze the possibility of avoiding the use of artificial lighting; Annual Sunlight Exposure (ASE) [132] to analyze the intensity of lighting levels causing visual discomfort, and Daylight Autonomy (sDA) to summarize the frequency in which an illuminance threshold can be maintained by natural light on its own. Figure 35 shows the number of hours that the luminance reaches 1000 lux at each point of the area on the top floor of the building under study. If this occurrence exceeds 250 hours, the affected area can have visual discomfort by glare [133].

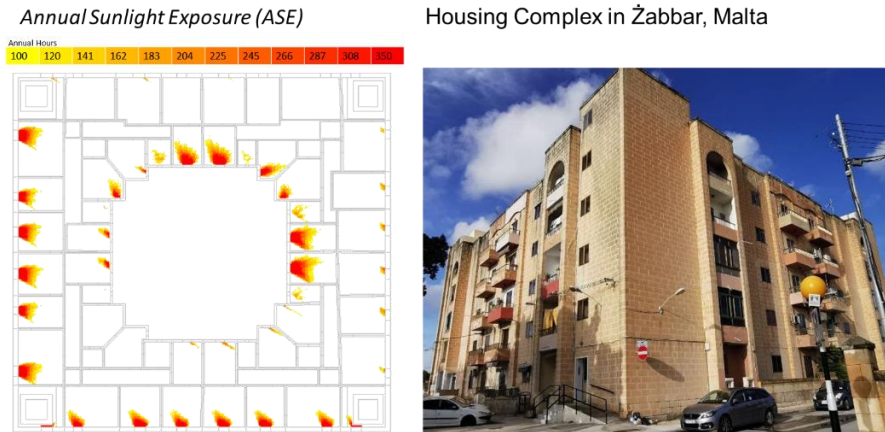


Figure 35. Left side: ASE annual hours distribution on the top floor, right side: picture of the housing under renovation.

Besides, the additional effects that the lighting optimization produces, heating and cooling loads, are also analyzed. This publication is currently under review.

Line 3: New proposals of CIE Standard sky characterization using MIs and Machine Learning algorithms. The work in progress in this field focuses on a new clustering of sky types using the dependence of VSC on solar altitude, as Figure 36 shows. ANN with the Scaled conjugate gradient backpropagation, and decision trees are used. The preliminary results are shown in Figure 37.

This paper has been accepted at the XVIII Congreso Ibérico y XIV Congreso Iberoamericano de Energía Solar (CIES2022), that will be developed in Mallorca, Spain from June 20 to June 22, 2022, as an oral communication. Further details of the full paper are shown in Annex 2.

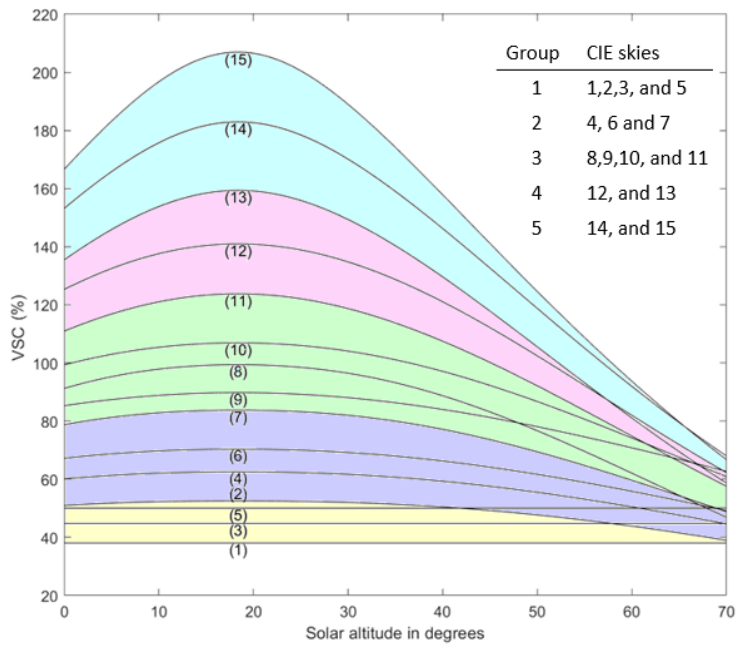


Figure 36. VSC dependence on solar altitude.

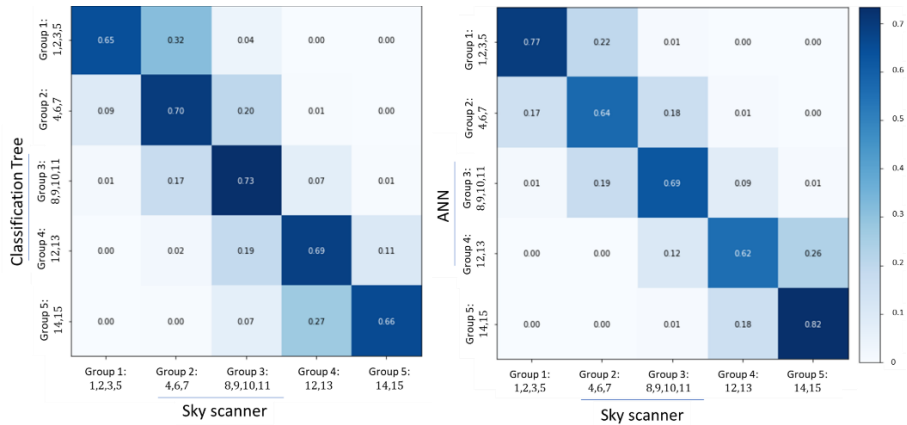


Figure 37. Confusion matrices of the Machine learning predictions. Classification tree (left), and ANN (right), for the CIE sky classification into five categories.

- **Line 4: New Proposals for CIE Standard sky characterization using sky images and Machine Learning Algorithms.** This is a new step in the research line initiated with paper IV and proposes the use of convolutional neural networks, specifically developed for image recognition [134, 135]. The objective is to improve the quality of the CIE standard sky classification by the comparison of several artificial neural network typologies.

- **Line 5: Development of new sky scanning devices.** As explained in Section 2.4 (Chapter 2), although commercial sky scanners are valuable devices that have led to a significant increase in the knowledge of the angular distribution of luminance and radiance in the sky, they have significant drawbacks. Among these are their low spatial and temporal resolution, not to mention their high cost. For this reason, we propose a future line of research focused on the development of two new types of devices for measuring the angular distribution of celestial luminance and radiance to solve the mentioned drawbacks:

1. To increase the temporal resolution: a sky scanner consisting of numerous low-cost sensors (145) will allow simultaneous measurement of the angular distribution of radiance over the entire celestial vault. The development of this equipment will involve a precise design and execution of the frame that will house the sensors and the geometric and photometric calibration of the sensor array.
2. To increase the spatial resolution: design, characterization, calibration, and validation of a device for capturing high dynamic range (HDR) high-resolution hemispherical HDR images. In this way, the spatial resolution of the measurement will improve substantially. The equipment will consist of a commercial camera equipped with a fisheye lens. The camera-lens assembly will be characterized from the geometric and photometric point of view. Specifically, the spectral characterization of the device will allow correlating the digital

levels of the image with different ranges of the solar radiation spectrum: photosynthetically active radiation, illuminance, and irradiance. These devices' implementation also means a decrease in the time required to scan the sky, although it does not reach the high temporal resolution of the equipment mentioned in the previous point.

In both cases, the cost of the equipment is significantly lower than that of commercial sky scanners.

Chapter 4: Bibliography

- [1] European Commission. The European Green Deal. *European Commission, COM(2019) 640 final Communication*. 2019.
- [2] Directive 2010/31/EU. Directive 2010/31/EU of the European parliament and of the council of 19 may 2010 on the energy performance of buildings (recast). *Official Journal of the European Union*. 2010.
- [3] Alados, Inmaculada, Inmaculada. Foyo-Moreno and Lucas Alados-Arboledas. Photosynthetically active radiation: measurements and modelling. *Agricultural and Forest Meteorology*. 1996, **78**(1–2), 121–131. DOI: [https://doi.org/10.1016/0168-1923\(95\)02245-7](https://doi.org/10.1016/0168-1923(95)02245-7)
- [4] Pei, Fengsong, Xia Li, Xiaoping Liu and Chunhua Lao. Assessing the impacts of droughts on net primary productivity in China. *Journal of Environmental Management*. 2013, **114**, 362–371. DOI: [10.1016/j.jenvman.2012.10.031](https://doi.org/10.1016/j.jenvman.2012.10.031)
- [5] Feng, Chunjie, Xiaotong Zhang, Yu Wei, Weiyu Zhang, Ning Hou, Jiawen Xu, Shuyue Yang, Xianhong Xie and Bo Jiang. Estimation of Long-Term Surface Downward Longwave Radiation over the Global Land from 2000 to 2018. *Remote Sensing*. 2021, **13**(9), 1848. DOI: [10.3390/rs13091848](https://doi.org/10.3390/rs13091848)
- [6] Duan, Qiuhua, Yanxiao Feng and Julian Wang. Estimating Hourly Solar NIR Irradiance Using Meteorological Data for Sustainable Building Design and Engineering. *ASES Solar 2020 Proceedings, June 24-25 (virtual conference)*. 2020. DOI: [10.18086/solar.2020.01.12](https://doi.org/10.18086/solar.2020.01.12)
- [7] Jacovides, C.P., F.S. Tymvios, D.N. Asimakopoulos, N.A. Kaltsounides, G.A. Theoharatos and M. Tsitouri. Solar global UVB (280–315nm) and UVA (315–380nm) radiant fluxes and their relationships with broadband global radiant flux at an eastern Mediterranean site. *Agricultural and Forest Meteorology*. 2009,

-
- 149(6–7), 1188–1200. DOI:
<https://doi.org/10.1016/j.agrformet.2009.02.009>
- [8] Rea, MS, MG Figueiro, A Bierman and R Hamner. Modelling the spectral sensitivity of the human circadian system. *Lighting Research & Technology*. 2012, 44(4), 386–396. DOI: <https://doi.org/10.1177/1477153511430474>
- [9] Dieste-Velasco, M. I., M. Díez-Mediavilla, D. Granados-López, D. González-Peña and C. Alonso-Tristán. Performance of global luminous efficacy models and proposal of a new model for daylighting in Burgos, Spain. *Renewable Energy*. 2019, 133, 1000–1010. DOI: <https://doi.org/10.1016/j.renene.2018.10.085>
- [10] Torres, José Luis and Luis Miguel Torres. Modeling Solar Radiation at the Earth’s Surface. Angular Distribution of Sky Diffuse Radiance and Luminance. *Modeling Solar Radiation at the Earth’s Surface: Recent Advances. Springer Berlin Heidelberg, Berlin, Heidelberg*. 2008, 427–448. DOI: 10.1007/978-3-540-77455-6_17
- [11] Perez, Richard, Pierre Ineichen, Robert Seals, Joseph J. Michalsky and Ronald Stewart. Modeling daylight availability and irradiance components from direct and global irradiance. *Solar Energy*. 1990, 44(5), 271–289. DOI: [https://doi.org/10.1016/0038-092X\(90\)90055-H](https://doi.org/10.1016/0038-092X(90)90055-H)
- [12] Igawa, Norio, Yasuko Koga, Tomoko Matsuzawa and Hiroshi Nakamura. Models of sky radiance distribution and sky luminance distribution. *Solar Energy*. 2004, 77(2), 137–157. DOI: 10.1016/j.solener.2004.04.016
- [13] Moon, P. and D.E. Spencer. Illumination from a non-uniform sky. *Transactions of the Illuminating Engineering Society*. 1942, 37, 707–726.
- [14] Kittler, Richard. Standardisation of outdoor conditions for the calculation of daylight factor with clear skies. In: *Proceedings of the CIE Intersessional Conference on Sunlight in Buildings, Bouwcentrum, Netherlands*. 1967, p. 273–285.

- [15] CIE. Natural Daylight. Official Recommendation. *Compte Rendu, CIE 13th Session, Committee E-3.2, vol. II, parts 3– 2, II–IV&35–37*. 1955, **2**(3– 2).
- [16] CIE. *Standardization of luminance distribution on clear skies. Technical Report CIE 022:1973*. 1973.
- [17] Perraudon, M. Luminance models. In: *National Lighting Conference and Daylighting Colloquium*. 1988, p. 291–292.
- [18] Matsuura, K and T Iwata. A model of daylight source for the daylight illuminance calculations on the all weather conditions. In: *Proceedings of 3rd International Daylighting Conference*. 1990.
- [19] Brunger, Alfred P. P. and Frank C. C. Hooper. Anisotropic sky radiance model based on narrow field of view measurements of shortwave radiance. *Solar Energy*. 1993, **51**(1), 53–64. DOI: [https://doi.org/10.1016/0038-092X\(93\)90042-M](https://doi.org/10.1016/0038-092X(93)90042-M)
- [20] Perez, Richard, Robert Seals and Joseph J. Michalsky. All-weather model for sky luminance distribution-Preliminary configuration and validation. *Solar Energy*. 1993, **50**(3), 235–245. DOI: [https://doi.org/10.1016/0038-092X\(93\)90017-I](https://doi.org/10.1016/0038-092X(93)90017-I)
- [21] Igawa, Norio. Improving the All Sky Model for the luminance and radiance distributions of the sky. *Solar Energy*. 2014, **105**, 354–372. DOI: <http://dx.doi.org/10.1016/j.solener.2014.03.020>
- [22] Kittler, Richard, Richard Perez and Stanislav Darula. A new generation of sky standards. In: *Lux Europa. Proceedings of the Eight European Lighting Conference*. 1997, p. 359–373.
- [23] Kittler, Richard, Richard Perez and Stanislav Darula. A set of standard skies characterizing daylight conditions for computer and energy conscious design. *American - Slovak grant project US - SK 92 052*. 1998. DOI: 10.13140/RG.2.1.4798.7048
- [24] ISO 15469:2004(E)/CIE S 011/E:2003. Spatial distribution of daylight-CIE standard general sky. *International standard*. 2004.
- [25] Li, Danny H.W., T. C. Chau and Kevin K.W. Wan. A review of the

-
- CIE general sky classification approaches. *Renewable and Sustainable Energy Reviews*. 2014, **31**, 563–574. DOI: <https://doi.org/10.1016/j.rser.2013.12.018>
- [26] Watt-Meyer, Oliver, Noah D. Brenowitz, Spencer K. Clark, Brian Henn, Anna Kwa, Jeremy McGibbon, W. Andre Perkins and Christopher S. Bretherton. Correcting Weather and Climate Models by Machine Learning Nudged Historical Simulations. *Geophysical Research Letters*. 2021, **48**(15), 1–10. DOI: <https://doi.org/10.1029/2021GL092555>
- [27] Diego Granados-López , Andrés Suárez-García , Montserrat Díez-Mediavilla, Cristina and Alonso-Tristán. Feature selection for CIE standard sky classification. *Solar Energy*. 2021, **218**, 95–107. DOI: <https://doi.org/10.1016/j.solener.2021.02.039>
- [28] Granados-López, Diego., Ana. García-Rodríguez, Sol. García-Rodríguez, Andrés Suárez-García, Montserrat Díez-Mediavilla and Cristina Alonso-Tristán. Pixel-Based Image Processing for CIE Standard Sky Classification through ANN. *Complexity*. 2021, **2021**, 1–15. DOI: <https://doi.org/10.1155/2021/2636157>
- [29] Li, Danny H.W., H. L. Tang, Eric W.M. Lee and Tariq Muneer. Classification of CIE standard skies using probabilistic neural networks. *International Journal of Climatology*. 2010, **30**(2), 305–315. DOI: <https://doi.org/10.1002/joc.1891>
- [30] Kocifaj, Miroslav. Unified model of radiance patterns under arbitrary sky conditions. *Solar Energy*. 2015, **115**, 40–51. DOI: <https://doi.org/10.1016/j.solener.2015.02.019>
- [31] García-Rodríguez, Ana, Sol García-Rodríguez, Montserrat Díez-Mediavilla and Cristina Alonso-Tristán. Photosynthetic active radiation, solar irradiance and the cie standard sky classification. *Applied Sciences (Switzerland)*. 2020, **10**(22), 8007. DOI: <https://doi.org/10.3390/app10228007>
- [32] Torres, J. L., M. De Blas, A. García, A. Gracia and A. De Francisco. Sky luminance distribution in Pamplona (Spain) during the summer

- period. *Journal of Atmospheric and Solar-Terrestrial Physics*. 2010, **72**(5–6), 382–388. DOI: <https://doi.org/10.1016/j.jastp.2009.12.005>
- [33] Alshaibani, Khalid, Danny Li and Emmanuel Aghimien. Sky luminance distribution models: A comparison with measurements from a maritime desert region. *Energies*. 2020, **13**(20), 5455. DOI: <https://doi.org/10.3390/en13205455>
- [34] Li, Danny H.W., K. L. Cheung, H. L. Tang and Charles C.K. Cheng. Identifying CIE standard skies using vertical sky component. *Journal of Atmospheric and Solar-Terrestrial Physics*. 2011, **73**(13), 1861–1867. DOI: <https://doi.org/10.1016/j.jastp.2011.04.015>
- [35] Tohsing, Korntip, Michael Schrempf, Stefan Riechelmann, Holger Schilke and Gunther Seckmeyer. Measuring high-resolution sky luminance distributions with a CCD camera. *Applied Optics*. 2013, **52**(8), 1564–1573. DOI: <https://doi.org/10.1364/AO.52.001564>
- [36] Inanici, Mehlika. Evaluation of high dynamic range image-based sky models in lighting simulation. *LEUKOS - Journal of Illuminating Engineering Society*. 2010, **7**(2), 69–84. DOI: <https://doi.org/10.1582/LEUKOS.2010.07.02001>
- [37] Gueymard, Christian A. and Jose A. Ruiz-Arias. Extensive worldwide validation and climate sensitivity analysis of direct irradiance predictions from 1-min global irradiance. *Solar Energy*. 2016, **128**, 1–30. DOI: <https://doi.org/10.1016/j.solener.2015.10.010>
- [38] Li, Danny H W. A review of daylight illuminance determinations and energy implications. *Applied Energy*. 2010, **87**(7), 2109–2118. DOI: <https://doi.org/10.1016/j.apenergy.2010.03.004>
- [39] Li, Danny H.W., Natalie T.C. Chau and Kevin K.W. Wan. Predicting daylight illuminance and solar irradiance on vertical surfaces based on classified standard skies. *Energy*. 2013, **53**, 252–258. DOI: <https://doi.org/10.1016/j.energy.2013.02.049>
- [40] Suárez-García, Andrés, Diego Granados-López, David González-Peña, Montserrat Díez-Mediavilla and Cristina Alonso-Tristán.

-
- Seasonal characterization of CIE standard sky types above Burgos, northwestern Spain. *Solar Energy*. 2018, **169**, 24–33. DOI: <https://doi.org/10.1016/j.solener.2018.04.028>
- [41] Tregenza, Peter Roy. Analysing sky luminance scans to obtain frequency distributions of CIE Standard General Skies. *Lighting Research and Technology*. 2004, **36**(4), 271–281. DOI: 10.1191/1477153504li1170a
- [42] Li, D. H.W. and H. L. Tang. Standard skies classification in Hong Kong. *Journal of Atmospheric and Solar-Terrestrial Physics*. 2008, **70**(8–9), 1222–1230. DOI: 10.1016/j.jastp.2008.03.004
- [43] Directive(EU). Directive (EU) 2018/844 of the European Parliament and of the Council of 30 May 2018 amending Directive 2010/31/EU on the energy performance of buildings and Directive 2012/27/EU on energy efficiency (Text with EEA relevance). *Official Journal of the European Union, L 156/75*. 2018, 75–91. Available at: <https://eur-lex.europa.eu/eli/dir/2018/844/oj>
- [44] IEA. Tracking Buildings 2021. *IEA*. 2021. Available at: <https://www.iea.org/reports/tracking-buildings-2021>.
- [45] DiLaura, David L., Kevin W. Houser, Richard G. Mistrick and Gary R. Steffy. The lighting handbook - Reference and Application. In: *Illuminating Engineering Society*. Tenth Edit. 2011, p. 1–1331. ISBN 978-0-87995-241-9.
- [46] Littlefair, Paul. Site Layout planning for daylight and sunlight. In: *Building Research Establishment (BRE), IHS BRE Press, Garston, Watford*. 2011. ISBN 978-1-84806-178-1.
- [47] Kómar, Ladislav and Miroslav Kocifaj. Uncertainty of daylight illuminance on vertical building façades when determined from sky scanner data: A numerical study. *Solar Energy*. 2014, **110**, 15–21. DOI: <https://doi.org/10.1016/j.solener.2014.09.010>
- [48] Chen, Wenqiang, Danny H.W. Li and Siwei Lou. Estimation of irregular obstructed vertical sky components under various CIE

- skies. *Energy Procedia, 10th International Conference on Applied Energy (ICAE2018), August 22-25, Hong Kong, China.* 2019, **158**, 309–314. DOI: 10.1016/j.egypro.2019.01.094
- [49] Olina, Ance and Nevila Zaimi. Daylight prediction based on the VSC - DF relation. *Energy and Building Design Department of Architecture and Built Environment, Lund University, Thesis: EEBD - xx/18.* 2018. Available at: <https://www.lunduniversity.lu.se/lup/publication/8955344>
- [50] DETR. Desktop guide to daylighting - for architects. *Good practice 245, DERT Environment Transport Regions.* 1998. Available at: <https://www.cibse.org/getmedia/19b5aa2a-309f-4e8d-8e72-07b5c4c4cf72/GPG245-Desktop-Guide-to-Daylighting-for-Architects.pdf.aspx>
- [51] Granados López, Diego, Montserrat Díez Mediavilla, M Isabel Dieste Velasco, Andrés Suárez García and Cristina Alonso Tristán. Evaluation of the Vertical Sky Component without Obstructions for Daylighting in Burgos , Spain. *Applied sciences.* 2020, **10**(9), 3095. DOI: <https://doi.org/10.3390/app10093095>
- [52] Ng, Edward, Vicky Cheng, Ankur Gadi, Jun Mu, Max Lee and Ankit Gadi. Defining standard skies for Hong Kong. *Building and Environment.* 2007, **42**(2), 866–876. DOI: <https://doi.org/10.1016/j.buildenv.2005.10.005>
- [53] Andersen, M., J. Ashmore, L. Beltran, J. Bos, D. Brentrup, K. Cheney, N. Digert, D. Eijadi, L. Fernandes and D. Glase. Approved Method: IES Spatial Daylight Autonomy (sDA) and Annual Sunlight Exposure (ASE). In: *Illuminating Engineering Society of North America, IES LM-83-12.* 2012. ISBN 978-0-87995-272-3.
- [54] EN 12464-1:2011. BSI Standards Publication. Light and lighting — Lighting of work places Part 1: Indoor work places. In: *European committee for standardization.* 2011. ISBN 978 0 580 68495 1.
- [55] Mayhoub, Mohammed. S. and David J. Carter. Methods to estimate global and diffused luminous efficacies based on satellite data. *Solar*

-
- Energy*. 2011, **85**(11), 2940–2952. DOI: <https://doi.org/10.1016/j.solener.2011.08.033>
- [56] Li, Danny H.W., Tony N.T. Lam, K. L. Cheung and H. L. Tang. An analysis of luminous efficacies under the CIE standard skies. *Renewable Energy*. 2008, **33**(11), 2357–2365. DOI: <https://doi.org/10.1016/j.renene.2008.02.004>
- [57] Robledo, L., A. Soler and E. Ruiz. Luminous efficacy of global solar radiation on a horizontal surface for overcast and intermediate skies. *Theoretical and Applied Climatology*. 2001, **69**, 123–134. DOI: <https://doi.org/10.1007/s007040170039>
- [58] Chung, T. M. A study of luminous efficacy of daylight in Hong Kong. *Energy and Buildings*. 1992, **19**(1), 45–50. DOI: [https://doi.org/10.1016/0378-7788\(92\)90034-E](https://doi.org/10.1016/0378-7788(92)90034-E)
- [59] Lam, Joseph C. and Danny H.W. Li. Luminous efficacy of daylight under different sky conditions. *Energy Conversion and Management*. 1996, **37**(12), 1703–1711. DOI: [https://doi.org/10.1016/0196-8904\(96\)00021-0](https://doi.org/10.1016/0196-8904(96)00021-0)
- [60] Muneer, T. and D. Kinghorn. Luminous efficacy models—evaluation against UK data. *Journal of the Illuminating Engineering Society*. 1998, **27**(1), 163–170. DOI: <https://doi.org/10.1080/00994480.1998.10748221>
- [61] Robledo, Luis and Alfonso Soler. Luminous efficacy of direct solar radiation for clear skies. *Energy Conversion & Management*. 2000, **25**(16), 1769–1779. DOI: [https://doi.org/10.1016/S0196-8904\(00\)00019-4](https://doi.org/10.1016/S0196-8904(00)00019-4)
- [62] Ruiz, Enrique, Alfonso Soler and Luis Robledo. Assessment of Muneer’s Luminous Efficacy Models in Madrid and a Proposal for New Models Based on His Approach. *Journal of Solar Energy Engineering*. 2001, **123**(3), 220–224. DOI: <https://doi.org/10.1115/1.1385200>
- [63] De Souza, Roberta G., Luis Robledo, Fernando O.R. Pereira and

- Alfonso Soler. Evaluation of global luminous efficacy models for Florianópolis, Brazil. *Building and Environment*. 2006, **41**(10), 1364–1371. DOI: <https://doi.org/10.1016/j.buildenv.2005.07.010>
- [64] Fakra, A. H., H. Boyer, F. Miranville and D. Bigot. A simple evaluation of global and diffuse luminous efficacy for all sky conditions in tropical and humid climate. *Renewable Energy*. 2011, **36**(1), 298–306. DOI: [10.1016/j.renene.2010.06.042](https://doi.org/10.1016/j.renene.2010.06.042)
- [65] Mahdavi, Ardeshir and Sokol Dervishi. A comparison of luminous efficacy models based on data from Vienna, Austria. *Building Simulation*. 2011, **4**(3), 183–188. DOI: [10.1007/s12273-011-0021-z](https://doi.org/10.1007/s12273-011-0021-z)
- [66] Chaiwiwatworakul, Pipat and Surapong Chirarattananon. Luminous efficacies of global and diffuse horizontal irradiances in a tropical region. *Renewable Energy*. 2013, **53**, 148–158. ISSN 09601481. DOI: [10.1016/j.renene.2012.10.059](https://doi.org/10.1016/j.renene.2012.10.059)
- [67] López, Gabriel and Christian A. Gueymard. Clear-sky solar luminous efficacy determination using artificial neural networks. *Solar Energy*. 2007, **81**(7), 929–939. DOI: <https://doi.org/10.1016/j.solener.2006.11.001>
- [68] Gueymard, Christian A., Jamie M. Bright, David Lingfors, Aron Habte and Manajit Sengupta. A posteriori clear-sky identification methods in solar irradiance time series: Review and preliminary validation using sky imagers. *Renewable and Sustainable Energy Reviews*. 2019, **109**, 412–427. DOI: <https://doi.org/10.1016/j.rser.2019.04.027>
- [69] A. Suárez García, C. Alonso Tristán and M. Díez Mediavilla D. Granados López, D. González Peña. Benchmarking of meteorological indices for sky cloudiness classification. *Solar Energy*. 2020, **195**, 499–513. DOI: <https://doi.org/10.1016/j.solener.2019.11.060>
- [70] Hyo Joo Kong, Jeong Tai Kim. A Classification of real sky conditions for Yongin, Korea. *Sustainability in Energy and Buildings. Smart Innovation, Systems and Technologies, Springer. Berlin, Heidelberg*. 2013, **22**, 1025–1032. DOI: <https://doi.org/10.1007/978-3-642-36645->

-
- [71] Wang, Lunche, Wei Gong, Yingying Ma, Bo Hu, Wenling Wang and Miao Zhang. Analysis of ultraviolet radiation in Central China from observation and estimation. *Energy*. 2013, **59**, 764–774. DOI: <https://doi.org/10.1016/j.energy.2013.07.017>
- [72] Muneer, T., C. Gueymard and H. Kambezidis. Solar Radiation and Daylight Models. In: *Elsevier Butterworth-Heinemann*. Second Edition. 2004. ISBN 978-0-7506-5974-1.
- [73] Batlles, F. J., F. J. Olmo, J. Tovar and L. Alados-Arboledas. Comparison of cloudless sky parameterizations of solar irradiance at various Spanish midlatitude locations. *Theoretical and Applied Climatology*. 2000, **66**, 81–93. DOI: <https://doi.org/10.1007/s007040070034>
- [74] Klucher, T. M. Evaluation of models to predict insolation on tilted surfaces. *Solar Energy*. 1979, **23**(2), 111–114. DOI: [https://doi.org/10.1016/0038-092X\(79\)90110-5](https://doi.org/10.1016/0038-092X(79)90110-5)
- [75] Perez, Richard, Robert Seals, Pierre Ineichen, Ronald Stewart and David Menicucci. A new simplified version of the perez diffuse irradiance model for tilted surfaces. *Solar Energy*. 1987, **39**(3), 221–231. DOI: [https://doi.org/10.1016/S0038-092X\(87\)80031-2](https://doi.org/10.1016/S0038-092X(87)80031-2)
- [76] Kambezidis, H. D., E. M. Katevatis, M. Petrakis, S. Lykoudis and D. N. Asimakopoulos. Estimation of the Linke and Unsworth-Monteith turbidity factors in the visible spectrum: Application for Athens, Greece. *Solar Energy*. 1998, **62**(1), 39–50. DOI: [https://doi.org/10.1016/S0038-092X\(97\)00079-0](https://doi.org/10.1016/S0038-092X(97)00079-0)
- [77] Powers, David M. W. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *International Journal of Machine Learning Technology*. 2011, **2**, 37–63. DOI: <https://doi.org/10.48550/arXiv.2010.16061>
- [78] Yiming, Yang and Pedersen Jan O. A Comparative Study on Feature Selection in Text Categorization. *Proceeding ICML '97 Proceedings*

- of the Fourteenth International Conference on Machine Learning, July 8 - 12, San Francisco, USA.* 1997, 412–420. Available at: <https://dl.acm.org/doi/10.5555/645526.657137>
- [79] Lou, Siwei, Danny H.W. Li and Joseph C. Lam. CIE Standard Sky classification by accessible climatic indices. *Renewable Energy*. 2017, **113**, 347–356. DOI: <https://doi.org/10.1016/j.renene.2017.06.013>
- [80] Biesiada, Jacek and Włodzisław Duch. Feature selection for high-dimensional data - A pearson redundancy based filter. *Computer Recognition Systems 2, Springer, Berlin, Heidelberg*. 2007, **45**, 242–249. DOI: https://doi.org/10.1007/978-3-540-75175-5_30
- [81] Gregorutti, Baptiste, Bertrand Michel and Philippe Saint-Pierre. Correlation and variable importance in random forests. *Springer Science and Business Media*. 2016, **27**(3), 659–678. DOI: <https://doi.org/10.48550/arXiv.1310.5726>
- [82] Kursa, Miron Bartosz. Robustness of Random Forest-based gene selection methods. *BMC Bioinformatics*. 2014, **15**(8). DOI: <https://doi.org/10.1186/1471-2105-15-8>
- [83] Degenhardt, Frauke, Stephan Seifert and Silke Szymczak. Evaluation of variable selection methods for random forests and omics data sets. *Briefings in Bioinformatics*. 2019, **20**(2), 492–503. DOI: <https://doi.org/10.1093/bib/bbx124>
- [84] Mitra, Pabitra, C. A Murthy and Sankar K Pal. Unsupervised Feature Selection Using Feature Similarity. *IEEE, Transactions On Pattern Analysis And Machine Intelligence*. 2002, **24**(4), 301–312. DOI: <https://doi.org/10.1109/34.990133>
- [85] Yu, Lei and Huan Liu. Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution. *Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003), Washington (USA)*. 2003. Available at: <https://www.aaai.org/Papers/ICML/2003/ICML03-111.pdf>
- [86] Wald, Randall, Taghi M. Khoshgoftaar and Amri Napolitano.

-
- Optimizing wrapper-based feature selection for use on bioinformatics data. *Proceedings of the 27th International Florida Artificial Intelligence Research Society Conference, FLAIRS 2014*. 2014, 288–293.
- [87] D’Ambrosio, Antonio and Valerio A. Tutore. Conditional classification trees by weighting the gini impurity measure. *New Perspectives in Statistical Modeling and Data Analysis. Studies in Classification, Data Analysis, and Knowledge Organization*. Springer, Berlin, Heidelberg. 2011. DOI: https://doi.org/10.1007/978-3-642-11363-5_31
- [88] Witten, Ian H., Eibe Frank and Mark A. Hall. Data mining: practical machine learning tools and techniques. 2011. DOI: <https://doi.org/10.1016/C2009-0-19715-5>
- [89] Alonso-Montesinos, J. and F. J. Batlles. The use of a sky camera for solar radiation estimation based on digital image processing. *Energy*. 2015, **90**, 377–386. DOI: <https://doi.org/10.1016/j.energy.2015.07.028>
- [90] Chauvin, R., J. Nou, S. Thil, A. Traoré and S. Grieu. Cloud Detection Methodology Based on a Sky-imaging System. *Energy Procedia*. 2015, **69**, 1970–1980. DOI: <https://doi.org/10.1016/j.egypro.2015.03.198>
- [91] Cheng, Hsu Yung and Chih Lung Lin. Cloud detection in all-sky images via multi-scale neighborhood features and multiple supervised learning techniques. *Atmospheric Measurement Techniques*. 2017, **10**, 199–208. DOI: <https://doi.org/10.5194/amt-10-199-2017>
- [92] Piderit, Maria Beatriz, Coralie Cauwerts and Muriel Diaz. Definition of the CIE standard skies and application of high dynamic range imaging technique to characterize the spatial distribution of daylight in Chile. *Revista de la Construccion*. 2014, **13**(2), 22–30. DOI: <http://dx.doi.org/10.4067/S0718-915X2014000200003>
- [93] García, Ignacio, Marian De Blas, José Luis Torres, Begoña Hernández, Carlos Sáenz and Mikel Ormazábal. Evaluation of two

- procedures for selecting the CIE standard sky type using high dynamic range images. *Proceedings of the ISES Solar World Congress 2019 and IEA SHC International Conference on Solar Heating and Cooling for Buildings and Industry 2019, November 03 - 07*, Santiago, Chile. 2019, 1958–1968. DOI: 10.18086/swc.2019.42.03
- [94] Tuominen, Sakari and Anssi Pekkarinen. Performance of different spectral and textural aerial photograph features in multi-source forest inventory. *Remote Sensing of Environment*. 2005, **94**(2), 256–268. DOI: <https://doi.org/10.1016/j.rse.2004.10.001>
- [95] Zhao, Xingying and Hong Zheng. Cloud Detection of Gray Photographs by Cloud Region Detection and Optimal Region Grow Threshold. *Proceedings of the 2nd International Conference on Intelligent Computing and Cognitive Informatics, ICICCI 2015*. 2015, 91–95. DOI: <https://doi.org/10.2991/icicci-15.2015.20>
- [96] Wan, Xinrui and Juan Du. Cloud classification for ground-based sky image using random forest. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLIII-B3-2020, 2020 XXIV ISPRS Congress (2020 edition)*. 2020, 835–842. DOI: <https://doi.org/10.5194/isprs-archives-XLIII-B3-2020-835-2020>
- [97] Yang, Jun, Qilong Min, Weitao Lu, Ying Ma, Wen Yao and Tianshu Lu. An RGB channel operation for removal of the difference of atmospheric scattering and its application on total sky cloud detection. *Atmospheric Measurement Techniques*. 2017, **10**, 1191–1201. DOI: <https://doi.org/10.5194/amt-10-1191-2017>
- [98] Yang, J., Q. Min, W. Lu, W. Yao, Y. Ma, J. Du and T. Lu. An automated cloud detection method based on green channel of total sky visible images. *Atmospheric Measurement Techniques Discussions*. 2015, **8**(5), 4581–4605. DOI: <https://doi.org/10.5194/amtd-8-4581-2015>
- [99] Li, Xin, Zhiying Lu, Qingxia Zhou and Zhengyang Xu. A Cloud

-
- detection algorithm with reduction of sunlight interference in ground-based sky images. *Atmosphere*. 2019, **10**(11), 640. DOI: <https://doi.org/10.3390/atmos10110640>
- [100] Gonzalez, R. C., R. E. Woods, and S. L. Eddins. Image Processing Toolbox for use with MATLAB. *MathWorks Inc, User's Guide Version 5*. 2004. Available at: <https://es.mathworks.com/help/images/>
- [101] Møller, Martin Fodslette. A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks*. 1993, **6**(4), 525–533. DOI: [https://doi.org/10.1016/S0893-6080\(05\)80056-5](https://doi.org/10.1016/S0893-6080(05)80056-5)
- [102] Kazanasmaz, Tuğçe, Murat Günaydin and Selcen Binol. Artificial neural networks to predict daylight illuminance in office buildings. *Building and Environment*. 2009, **44**(8), 1751–1757. DOI: <https://doi.org/10.1016/j.buildenv.2008.11.012>
- [103] Lima, Francisco J.L., Fernando R. Martins, Enio B. Pereira, Elke Lorenz and Detlev Heinemann. Forecast for surface solar irradiance at the Brazilian Northeastern region using NWP model and artificial neural networks. *Renewable Energy*. 2016, **87**, 807–818. DOI: <http://dx.doi.org/10.1016/j.renene.2015.11.005>
- [104] Kambezidis, HD, ThI Oikonomou and D Zevgolis. Daylight climatology in the Athens urban environment: Guidance for building designers. *Lighting Research and Technology*. 2002, **34**(4), 297–309. DOI: <https://doi.org/10.1191/1365782802li037oa>
- [105] Tregenza, Peter. and John Mardaljevic. Daylighting buildings: Standards and the needs of the designer. *Lighting Research and Technology*. 2018, **50**(1), 63–79. DOI: <https://doi.org/10.1177/1477153517740611>
- [106] MacNaughton, Piers, May Woo, Brandon Tinianov, Mohamed Boubekri and Usha Satish. Economic implications of access to daylight and views in office buildings from improved productivity. *Journal of Applied Social Psychology*. 2021, **51**(12), 1176–1183. DOI: <https://doi.org/10.1111/jasp.12764>

- [107] CEN/TR15193-2:2017. Energy performance of buildings - Energy requirements for lighting - Part 2: Explanation and justification of EN 15193-1, Module M9. *Technical Report Rapport Technique Technischer Bericht*. 2017. Available at: <https://tienda.aenor.com/norma-bsi-pd-cen-tr-15193-2-2017-00000000030292861>
- [108] EN 15193-1:2017. EN 15193-1 April. Energy performance of buildings - Energy requirements for lighting - Part 1: Specifications, Module M9. *European Standard*. 2017. Available at: <https://tienda.aenor.com/norma-une-en-15193-1-2019-n0061428>
- [109] Gago, E. J., T. Muneer, M. Knez and H. Köster. Natural light controls and guides in buildings. Energy saving for electrical lighting, reduction of cooling load. *Renewable and Sustainable Energy Reviews*. 2015, **41**, 1–13. DOI: <http://dx.doi.org/10.1016/j.rser.2014.08.002>
- [110] Alshaibani, K. Finding frequency distributions of CIE Standard General Skies from sky illuminance or irradiance. *Lighting Research and Technology*. 2011, **43**(4), 487–495. DOI: <https://doi.org/10.1177/1477153511404999>
- [111] Li, Danny H.W. and G. H.W. Cheung. Average daylight factor for the 15 CIE standard skies. *Lighting Research and Technology*. 2006, **38**(2), 137–152. DOI: <https://doi.org/10.1191/1365782806li165oa>
- [112] Li, Danny H.W., H. L. Tang, S. L. Wong, Ernest K.W. Tsang, Gary H.W. Cheung and Tony N.T. Lam. Skies classification using artificial neural networks (ANN) techniques. In *Proceedings of 6th International Conference on Indoor Air Quality, Ventilation and Energy Conservation in Buildings: Sustainable Built Environment, IAQVEC 2007, October 28 - 31, Sendai, Japan*. 2007, **1**, 61–68. Available at: [https://scholars.cityu.edu.hk/en/publications/publication\(82e6bcfa-346c-4fc5-bb14-d3fbc2f33eb3\).html](https://scholars.cityu.edu.hk/en/publications/publication(82e6bcfa-346c-4fc5-bb14-d3fbc2f33eb3).html)
- [113] Torres, J.L., M. De Blas, A. García, A. Gracia and A De Francisco.

-
- Sky luminance distribution in the North of Iberian Peninsula during winter. *Journal of Atmospheric and Solar-Terrestrial Physics*. 2010, **72**(16), 1147–1154. DOI: <https://doi.org/10.1016/j.jastp.2010.07.001>
- [114] Kittler, R. and S. Darula. The method of aperture meridians: A simple calculation tool for applying the ISO/CIE Standard General Sky. *Lighting Research and Technology*. 2006, **38**(2), 109–122. DOI: <https://doi.org/10.1191/1365782806li163oa>
- [115] Bonaiuti, S Pagani Guazzugli and M Wilson. An analysis of the (BRE) average daylight factor and limiting depth guidelines as design criteria. *2nd PALENC Conference and 28th AIVC Conference on Building Low Energy Cooling and Advanced Ventilation Technologies in the 21st Century, September, Crete island, Greece*. 2007, **2**, 739–745. Available at: https://www.new-learn.info/about/doc/pagani_adf_2007.pdf
- [116] Darula, Stanislav, Jens Christoffersen and Marta Malikova. Sunlight and insolation of building interiors. *Energy Procedia, 6th International Building Physics Conference, IBPC 2015, JUN 14-17, Torino, Italy*. 2015, **78**, 1245–1250. DOI: <http://dx.doi.org/10.1016/j.egypro.2015.11.266>
- [117] Yu, Xu and Yuehong Su. Daylight availability assessment and its potential energy saving estimation -A literature review. *Renewable and Sustainable Energy Reviews*. 2015, **52**, 494–503. DOI: <https://doi.org/10.1016/j.rser.2015.07.142>
- [118] SUPERLITE 2.0, predicting daylighting and lighting performance in complex building spaces. *USA: Lawrence Berkeley Laboratory*; 1994.
- [119] Ferraro, V., N. Igawa and V. Marinelli. INLUX-DBR - A calculation code to calculate indoor natural illuminance inside buildings under various sky conditions. *Energy*. 2010, **35**(9), 3722–3730. DOI: <https://doi.org/10.1016/j.energy.2010.05.021>
- [120] DIALUX 4.0. *Germany: DIAL GmbH; 2005*. 2005.
- [121] RELUX professional, version 2007.5. *Switzerland: Relux Informatik*

- AG. 2007.
- [122] DAYSIM. Official Web-site. 2021. Available at: <https://www.radiance-online.org/pipermail/radiance-daysim/2013-February/000026.html>
- [123] ECOTECT, version 5.50. U.K: *SQUARE ONE*. 2006.
- [124] Vartiainen, Eero. Daylight modelling with the simulation tool Delight. In: *Helsinki University of Technology Publications in Engineering Physics, TKK-F-A799*. 2000. ISBN 951-22-5071-3.
- [125] Yoon, Yeo Beom, Rashmi Manandhar and Kwang Ho Lee. Comparative study of two daylighting analysis methods with regard to window orientation and interior wall reflectance. *Energies*. 2014, 7(9), 5825–5846. DOI: <https://doi.org/10.3390/en7095825>
- [126] Yun, Su-in and Kang-Soo Kim. Sky Luminance Measurements Using CCD Camera and Comparisons with Calculation Models for Predicting Indoor Illuminance. *Sustainability*. 2018, 10(5), 1556. DOI: <https://doi.org/10.3390/su10051556>
- [127] Richard Perez, Robert Seals, Joseph Michalsky. Erratum to all-weather model for sky luminance distribution — preliminary configuration and validation. *Solar Energy*. 1993, 51(5), 423.
- [128] Gatt, Damien. and C. Yousif. Interreg Europe zeroco2 project - action plan for Malta to promote near zero co2 emission buildings. In: *The ESSE '18 Conference, May 18, St. Paul's Bay, Malta*. 2018. ISBN 9789995785321.
- [129] Gatt, Damien and Charles Yousif. Policy measures addressing nearly zero-energy buildings in the small island state of Malta. In: *Handbook of Governance in Small States, Routledge*. B.m.: Handbook of governance in small states, 2020, p. 117–140. ISBN 9780429061356. Available at: <https://www.um.edu.mt/library/oar/handle/123456789/65132>
- [130] DesignBuilder. helpv7.0. 2021. Available at: <https://designbuilder.co.uk/helpv7.0/>

-
- [131] Mardaljevic, John, Marilyne Andersen, N Roy and Jens Christoffersen. Daylighting metrics: is there a relation between useful daylight illuminance and daylight glare probability? *First IBPSA-England conference on "Building Simulation and Optimization"*. September 10-11, Loughborough, United Kingdom. 2012. Available at: <http://www.ibpsa.org/bsa-2012-proceedings/>
- [132] Basurto, Chantal, Magali Bodart, Jan De Boer, Marie-Claude Dubois, David Geisler-Moroder and Jérôme Kaempf. Energy audit and inspection procedures. A Technical Report of IEA SHC Task 50. *Solar Heating and Cooling Programme. International Energy Agency*. 2016. Available at: https://task50.iea-shc.org/Data/Sites/1/publications/Technical_Report_T50_C4_final.pdf
- [133] AL-Dossary, Ali Mohammed and Daeung Danny Kim. A study of design variables in daylight and energy performance in residential buildings under hot climates. *Energies*. 2020, **13**(21), 5836. DOI: <https://doi.org/10.3390/en13215836>
- [134] Xie, Wanyi, Dong Liu, Ming Yang, Shaoqing Chen, Bengge Wang, Zhenzhu Wang, Yingwei Xia, Yong Liu, Yiren Wang and Chaofang Zhang. SegCloud: A novel cloud image segmentation model using a deep convolutional neural network for ground-based all-sky-view camera observation. *Atmospheric Measurement Techniques Discussions*. 2020, **13**(4), 1953–1961. DOI: <https://doi.org/10.5194/amt-2019-356>
- [135] Zhang, Jinglin, Pu Liu, Feng Zhang and Qianqian Song. CloudNet: Ground-Based Cloud Classification With Deep Convolutional Neural Network. *Geophysical Research Letters*. 2018, **45**(16), 8665–8672. DOI: <https://doi.org/10.1029/2018GL077787>

Annex I:

Papers



**UNIVERSIDAD
DE BURGOS**

La Investigadora **SOL GARCÍA RODRÍGUEZ**, con DNI 71278921N, como coautor del siguiente trabajo:

Granados-López, D., García-Rodríguez, A., García-Rodríguez, S., Suárez-García, A., Díez-Mediavilla, M., Alonso-Tristán, C., 2021. Pixel-based image processing for CIE standard sky Classification through ANN. Complexity 2021, 1-15. DOI: <https://doi.org/10.1155/2021/2636157>

Expreso mediante el presente escrito mi renuncia a la utilización de la anterior publicación para la presentación de una tesis por compendio de artículos.

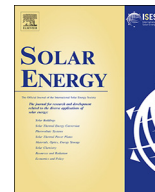
En Burgos, a 4 de abril de 2022

Fdo. Sol García Rodríguez

Paper I

Suárez-García, A., Granados-López, D., González-Peña, D., Díez-Mediavilla, M. and Alonso-Tristán, C., 2018. *Seasonal characterization of CIE standard sky types above Burgos, northwestern Spain*. *Sol. Energy* 169, 24–33. DOI: 10.1016/j.solener.2019.11.060

Abstract: Outdoor daylight conditions in Burgos, Spain, are studied throughout a full year. The CIE standard sky type is selected in accordance with the lowest RMSD (Root Mean Square Deviation) following the comparison of both the theoretical and the experimental luminance distributions in the sky hemisphere. The selection is based on luminance distribution data, recorded every 30 min, from 145 patches of the sky hemisphere. The original criterion to determine the type of sky, the SSLD (the Standard Sky Luminance Distribution), is difficult to apply in certain places and at times when the solar elevation is higher than 35°. In consequence, two alternative procedures are used and compared in this study: the Tregenza method and the Normalization Rate (NR) introduced by Littlefair. The selection was taken from luminance distribution data of 145 patches of the sky hemisphere recorded between June 2016 and May 2017. The most frequent sky type observed in Burgos was V.5. (cloudless polluted with a broad solar corona), with a frequency of occurrence close to 20%. Notwithstanding that observation, the group of clear skies exhibited a higher frequency (in almost 50% of the cases under study, using both methods). The skies above Burgos were of an overcast sky type in less than 25% of cases, a situation with a higher likelihood in winter and in autumn, while in spring and summer the skies tended to be clear and cloud free. Both of the methodologies showed similar results in percentage terms and in confusion matrixes with almost insignificant differences when compared on a monthly, a seasonal, and an annual basis. Nevertheless, some mismatches were located in the highest solar elevation values.



Seasonal characterization of CIE standard sky types above Burgos, northwestern Spain



Andrés Suárez-García^{a,b}, Diego Granados-López^b, David González-Peña^b,
Montserrat Díez-Mediavilla^b, Cristina Alonso-Tristán^{b,*}

^a Centro Universitario de la Defensa, Marín, Pontevedra, Spain

^b Research Group Solar and Wind Feasibility Technologies (SWIFT), Electromechanical Engineering Department, Universidad de Burgos, Spain

ARTICLE INFO

Keywords:

CIE standard skies
Luminance
Daylight
Sky luminance distribution

ABSTRACT

Outdoor daylight conditions in Burgos, Spain, are studied throughout a full year. The CIE standard sky type is selected in accordance with the lowest RMSD (Root Mean Square Deviation) following the comparison of both the theoretical and the experimental luminance distributions in the sky hemisphere. The selection is based on luminance distribution data, recorded every 30 min, from 145 patches of the sky hemisphere. The original criterion to determine the type of sky, the SSLD (the Standard Sky Luminance Distribution), is difficult to apply in certain places and at times when the solar elevation is higher than 35°. In consequence, two alternative procedures are used and compared in this study: the Tregenza method and the Normalization Rate (NR) introduced by Littlefair. The selection was taken from luminance distribution data of 145 patches of the sky hemisphere recorded between June 2016 and May 2017. The most frequent sky type observed in Burgos was V.5. (cloudless polluted with a broad solar corona), with a frequency of occurrence close to 20%. Notwithstanding that observation, the group of clear skies exhibited a higher frequency (in almost 50% of the cases under study, using both methods). The skies above Burgos were of an overcast sky type in less than 25% of cases, a situation with a higher likelihood in winter and in autumn, while in spring and summer the skies tended to be clear and cloud free. Both of the methodologies showed similar results in percentage terms and in confusion matrixes with almost insignificant differences when compared on a monthly, a seasonal, and an annual basis. Nevertheless, some mismatches were located in the highest solar elevation values.

1. Introduction

Energy efficiency and sustainability are increasingly important issues in the field of architecture. Lighting often has the highest electrical consumption and cost in buildings with no air-conditioning systems and could account for over 40% of electricity costs in naturally ventilated offices. Daylighting is recognized as a key strategy in reducing energy consumption. The availability of natural light is highly recommendable for reasons of energy efficiency, visual comfort, and the physical and mental well-being of building occupants (Hwang and Jeong, 2011; Torrington and Tregenza, 2007). Consequently, the recommendations of energy standards and green building rating systems strongly advise architects to incorporate daylighting strategies in their building designs (Aalto University School of Science and Technology, 2010). Architects and engineers need quantitative information on illumination levels and solar irradiance absorbed on surfaces at different inclinations for the incorporation of daylighting in the design of energy-efficient buildings

and for suitable dimensioning of both the cooling and the heating systems. It requires an accurate estimation of the amount of available outdoor illuminance and of course the availability of daylight is mainly influenced by the levels and the patterns of luminance in the sky. To obtain sky luminance distribution, empirical models of homogeneous skies represents a low cost approach. Many of these methods (Li, 2010) are aimed at estimating daylight availability.

In 2003, the CIE categorization defined 15 standard sky types (Uetani et al., 2003). Sky types of the same category have the same well-defined sky luminance pattern. Once the sky types are identified, the basic solar irradiance and daylight illuminance on the surfaces of interest can be obtained through simple mathematical expressions (Li et al., 2013). The luminance distribution for each standard sky type can help arrive at accurate determinations of daylight illuminance (Kittler et al., 1997). The classification includes five types of clear sky, five intermediate types, and five with cloud-cover. The distribution is characterized by continuous mathematical expressions to calculate

* Corresponding author.

E-mail address: catristan@ubu.es (C. Alonso-Tristán).

Nomenclature		
γ, Z	angle of elevation, angle from zenith	$L_{s,p}$ luminance relative to the Zenith (<i>dim</i>)
α	angle of azimuth	$L_{pr,st}$ normalized luminance of a sky patch, corresponding to a standard sky (CIE)
α_s, γ_s	solar azimuth, solar elevation	$L_{pr,sc}$ normalized luminance of a sky patch, corresponding to an experimental measurement
n_p	number of patches in band b	E_h horizontal diffuse illuminance (kcd/m ²)
b_p	reference number of each band	E_{hp} horizontal illuminance from sky patch p (kcd/m ² sr)
p	reference number of a scanned sky patch	NR luminance normalization ratio (kcd/m ²)
L_p	luminance measured by the Sky-scanner of a sky patch (kcd/m ²)	$L_{pred_{p,sc}}$ normalized luminance (NR method) (kcd/m ²)

smooth variations in luminance from the horizon to the zenith and in accordance with the angular distance from the sun. The general formula for defining the relative pattern of luminance for any sky type is a combination of a gradation function, dependent on two parameters, *a* and *b*, and the indicatrix function, which considers the scatter of luminance with regard to the direction of sunrays, which is modelled as a function of three adjustable parameters: *c*, *d*, and *e*. The gradation function modifies the luminance value between the horizon and the local zenith, assigning the highest luminance value to the zenith with cloudy skies and in reverse to clear skies, as shown in Fig. 1. The indicatrix function shows the dispersion in the atmosphere of sunlight, as represented in Fig. 2. The maximum luminance appears near the solar position, decreasing rapidly with the distance to the sun. Each of the functions takes six different forms and the combination yields 36 sky types from which 15 were selected: five overcast, five partly cloudy, and five clear sky types, as shown in Table 1.

The sky type must first be known, in order to apply the CIE standard general sky type as per ISO 15469:2004 CIE S 011/E:2003 (2004) for determining luminance distribution. The determination of the sky type at each location and time is a complex problem, due to the high fluctuation of the luminance magnitude and the influence of zenith luminance, that is determined with difficulty at low latitude locations. The original criterion to define the sky type, known as the SSLD method (Kittler et al., 1997) (Standard Sky Luminance Distribution), uses a theoretical assemblage of curves that represent the relation between the zenith luminance/diffuse illuminance (L_z/D_v) ratio and the solar elevation angle. These curves converge at solar elevation values higher than 35°, making it difficult to apply this method in certain areas and times when the solar elevation angle is higher than 35°, as can be seen in Fig. 3. Several alternatives have been proposed involving the ratios of horizontal global illuminance and extra-terrestrial illuminance (G_v/E_v), horizontal sky diffuse illuminance and extra-terrestrial illuminance (D_v/E_v), the turbidity index, T_v (Li et al., 2014) and different climatic and atmospheric parameters (Kocifaj, 2011; Li et al., 2004). Machine learning algorithms and other progressive methods has also been used in the CIE standard sky classification (Li et al., 2010; Lima et al., 2016; Lou et al., 2017)

The application of the above-mentioned calculations to the successive sky conditions yields the statistical distribution of the General Sky types that best fit the sky luminance patterns at a given location and that consequently define the daylight climate of a given site. Notwithstanding its validity for calculating daylight availability, the empirical ISO/CIE model of homogeneous skies could be inaccurate when interpreting illuminances and irradiances on arbitrarily oriented surfaces under cloudy conditions (Kocifaj and Kómar, 2016) due to heterogeneity of cloud field.

Despite the high interest in these measurements, very few studies at only a handful of European (Bratislava, Athens, South England, Spain) and Asian (Honk Kong, Japan, Singapore) locations have been conducted to characterize the sky under the CIE standard (Markou et al., 2005; Markou et al., 2004; Torres et al., 2010a,b; Tregenza, 1999) (Chaiwiwatworakul and Chirattananon, 2004; Li et al., 2003; Tregenza, 1999; Wittkopf et al., 2007) and over certain time periods

(Torres et al., 2010a,b).

The main objective of this work is to define the daylight conditions in Burgos, northwestern Spain. A seasonal classification of the sky conditions was performed with a full year of data recordings. The SSLD method is not applicable to the location under study as the solar elevation of 35° is surpassed, especially in summer. Various procedures to circumvent this issue are proposed, using various methods of normalization, two of which were selected for comparison in this study. The first one is the widely accepted method proposed by Tregenza (2004), consisting of a horizontal-based illuminance estimation. The second method, initially proposed by Littlefair (1994a,b), deals with the Normalization Ratio (NR) obtained by direct comparison between the CIE theoretical luminance and the experimental luminance levels; a method successfully applied to data recorded in Hong-Kong (Li et al., 2004; Li and Tang, 2008).

A Matlab code was developed for this purpose and a complete comparison of both methods is shown in a set of confusion matrixes. At the end, more than seven thousand samples were used. A complete year of measurement data was sufficient for a seasonal characterization of the skies over Burgos, referenced by the hour of the day. This study represents the first classification of this type in a Spanish city. Although both methodologies have been proposed for solar elevations higher than 35°, the NR approach is simpler and easier to compute than the Tregenza method. The results obtained by both methodologies in this study are perfectly comparable for the location under study and the solar elevation that is observed.

The study is organized as follows: Section 2 describes the

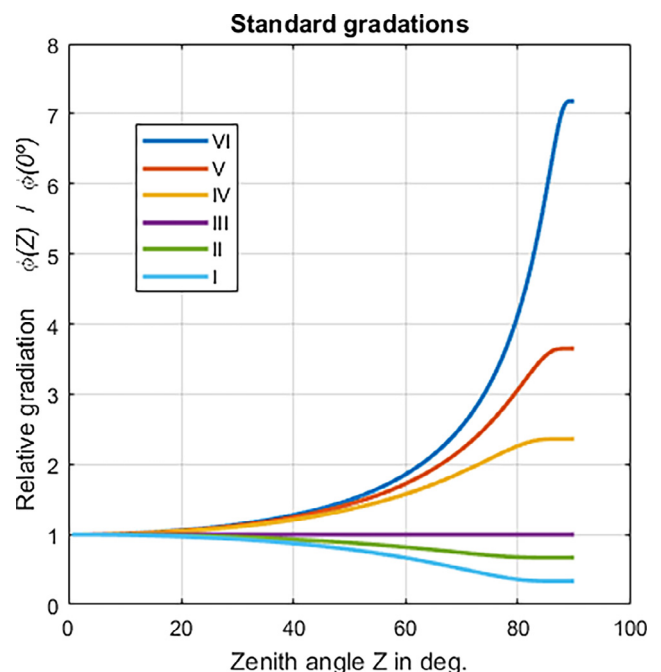


Fig. 1. Relative gradation function.

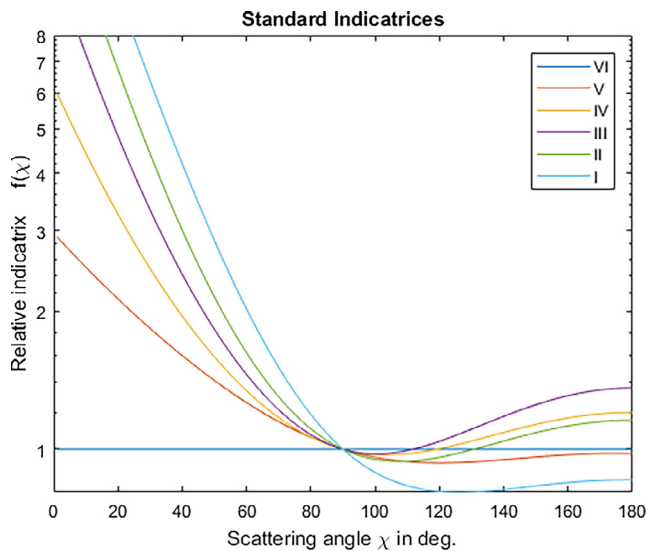


Fig. 2. Relative indicatrix function.

experimental facility used in this work, the meteorological features of the location under study and the experimental procedure. Section 3 summarizes the peculiarities of both methods applied for the CIE standard skies classification. Section 4 compares the sky classification according to both methodologies and presents the results in both graphical and numerical terms. Finally, Section 5 summarizes the principal observations and the contributions of the study.

2. Experimental section

CIE sky modelling and the application of the two methodologies was done using a code developed in Matlab.

2.1. Experimental facility

The experimental equipment used in this work is a commercial Sky-scanner model MS-321LR. The apparatus was installed on the roof of a building at Burgos University (42°21'04"N; 3°41'20"O; above mean sea level 856 m). Fig. 4 shows the sky scanner equipment and its geographical location.

The period of time under analysis was in general warm and dry. Burgos has an average of 575 mm of precipitation and an average annual global irradiance of 1500 kWh/m², as can be seen in a Typical

Table 1
Parameters of CIE standard Sky types.

Type	a	b	c	d	e	Description
I.1	4.0	-0.70	0	-1.0	0.00	Overcast with a steep gradation and azimuthal uniformity
I.2	4.0	-0.70	2	-1.5	0.15	Overcast with a steep gradation and slight brightening toward sun
II.1	1.1	-0.80	0	-1.0	0.00	Overcast with a moderate gradation and azimuthal uniformity
II.2	1.1	-0.80	2	-1.5	0.15	Overcast with a moderate gradation and slight brightening toward sun
III.1	0.0	-1.00	0	-1.0	0.00	Overcast, foggy or cloudy, with overall uniformity
III.2	0.0	-1.00	2	-1.5	0.15	Partly cloudy with a uniform gradation and slight brightening toward sun
III.3	0.0	-1.00	5	-2.5	0.30	Partly cloudy with a uniform gradation and a brighter circumsolar effect
III.4	0.0	-1.00	10	-3.0	0.45	Partly cloudy, rather uniform with a clear solar corona
IV.2	-1.0	-0.55	2	-1.5	0.15	Partly cloudy with a shaded sun position
IV.3	-1.0	-0.55	5	-2.5	0.30	Partly cloudy with brighter circumsolar effect
IV.4	-1.0	-0.55	10	-3.0	0.45	White-blue sky with a clear solar corona
V.4	-1.0	-0.32	10	-3.0	0.45	Very clear/unturbid with a clear solar corona
V.5	-1.0	-0.32	16	-3.0	0.30	Cloudless polluted with a broader solar corona
VI.5	-1.0	-0.15	16	-3.0	0.30	Cloudless turbid with a broader solar corona
VI.6	-1.0	-0.15	24	-2.8	0.15	White-blue turbid sky with a wide solar corona effect

Meteorological Year (TMY) over the last twenty years, compiled by the Spanish State Meteorology Agency (AEMET) (ITACYL-AEMET, 2013). However, annual rainfall of 510 mm was recorded in the year under analysis (12% less than the TMY) and annual solar irradiance of 1650 kWh/m² (10% higher than the TMY) as shown in Fig. 5. These data might bias the analysis, by giving the impression of a higher percentage of clear skies.

According to the sky-scanner specifications, shown in Table 2, the sky is divided into 145 patches or sectors (p) that cover the whole dome. The sectors are grouped into eight bands, named b_p , and by their solar altitude ($\frac{\pi}{2}-Z$), where Z is the zenith angle. Fig. 6 shows the location of the sectors in the whole dome. A luminance measurement (kcd/m²) of each patch is taken four times per hour. Half-hourly and hourly measurements taken between June 2016 and May 2017 were used for this study. Continuous scanning yielded luminance data corresponding to the 145 patches (see Fig. 6) recommended for the CIE in the Guide to Daylight Measurements, which were measured and registered. Likewise, the luminance corresponding to each of the commonly considered 15 standard sky types presented in Table 1 was calculated at the same time and for the same 145 patches. The standard sky type ascribed to each recorded moment showed the lowest RSM (Root Mean Square Deviation) between the 145 normalized luminance values that were measured and calculated.

2.2. Experimental procedure

Half-hourly and hourly measurements taken from June 2016 to May 2017 were used in this study. The sky scanner is monthly adjusted to measure from the sunrise to the sunset. The first and last measurement of the day (solar elevation angle equal or lower than 5°) are discarded, as well measurements higher than 50 kcd/m² and lower than 0.1 kcd/m², following the specification of the equipment. Data pre-processing in Matlab code was performed, to avoid incorrect measurements. The number of available hourly data were classified by months, as shown in Fig. 7. The abnormal number of available measurements in summer 2016, with the longest days of the year, is caused by modifications in the device.

3. CIE sky characterization

As mentioned in the introduction section, the two different methods applied in this study are both explained in this section.

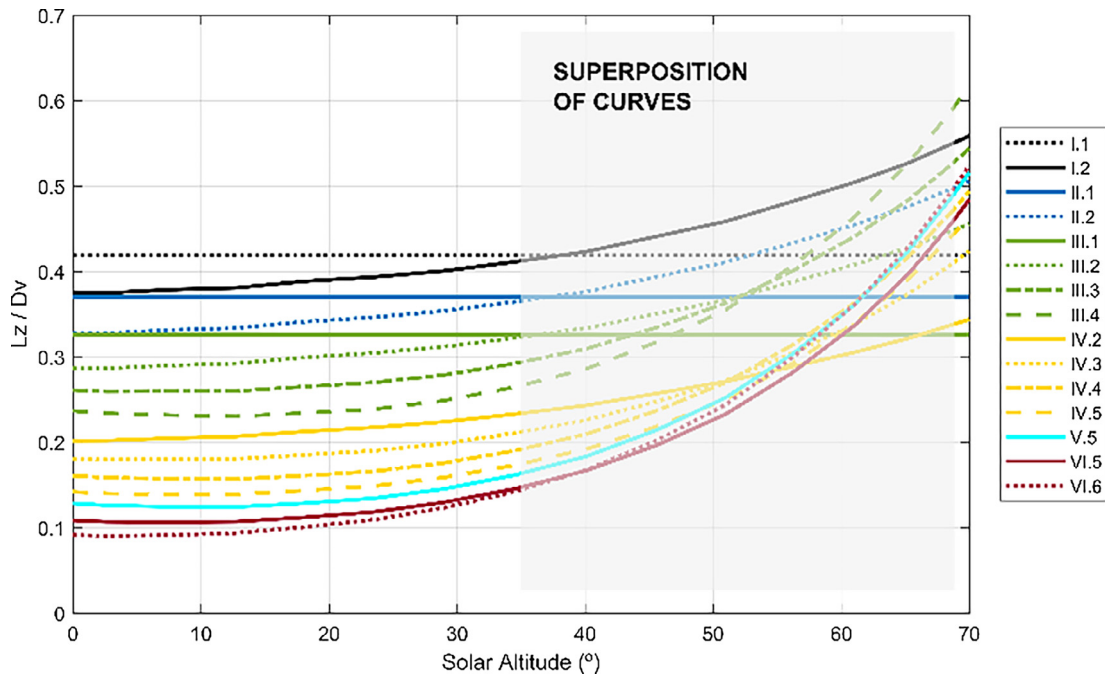


Fig. 3. Graphical representation of the ratio zenith luminance/diffuse illuminance (L_z/D_v) vs solar altitude, where the convergence of the curves for values higher than 35° can be appreciated.

3.1. The Tregenza method

The Tregenza method (Tregenza, 2004) calculates the horizontal diffuse illuminance, E_h (Eq. (1)) in each scan. This value is the sum of the luminance from the different parts of the sky hemisphere. E_{hp} (Eq. (2)) is the contribution of the measured luminance, L_p , coming from the patch p , to the horizontal diffuse illuminance. F_c is a correction factor that must be included because, in practice, the sum does not extend to the 145 patches. A few of them are discarded for the analysis due to two causes: (a) according to the specifications of the instrument, luminance values lower than 0 kcd/m^2 or higher than 50 kcd/m^2 are outside the range of its measurements; and, (b) patches close to the position of the sun should also be excluded, in order to avoid direct luminance. As can

be seen in Fig. 6, the patches excluded for the calculation are those bordering the patch corresponding to the position of the sun.

$$E_h = F_c \sum_{p=1} E_{hp} \tag{1}$$

$$E_{hp} = \int_{\gamma_0}^{\gamma} \int_{\alpha_0}^{\alpha} L_p \sin(\gamma) \cos(\gamma) d\gamma d\alpha = \frac{L_p}{2} (\sin^2(\gamma) - \sin^2(\gamma_0)) (\alpha - \alpha_0) = L_p F_{g,p} \tag{2}$$

where, γ_0 is the sky patch's basis, and γ is the top. $(\alpha - \alpha_0)$ is the azimuthal distance between patch's limits. $F_{g,p}$ is a geometrical factor characteristic to each patch. Eq. (3) is used for calculating the geometrical factor of a particular patch, taking into consideration the

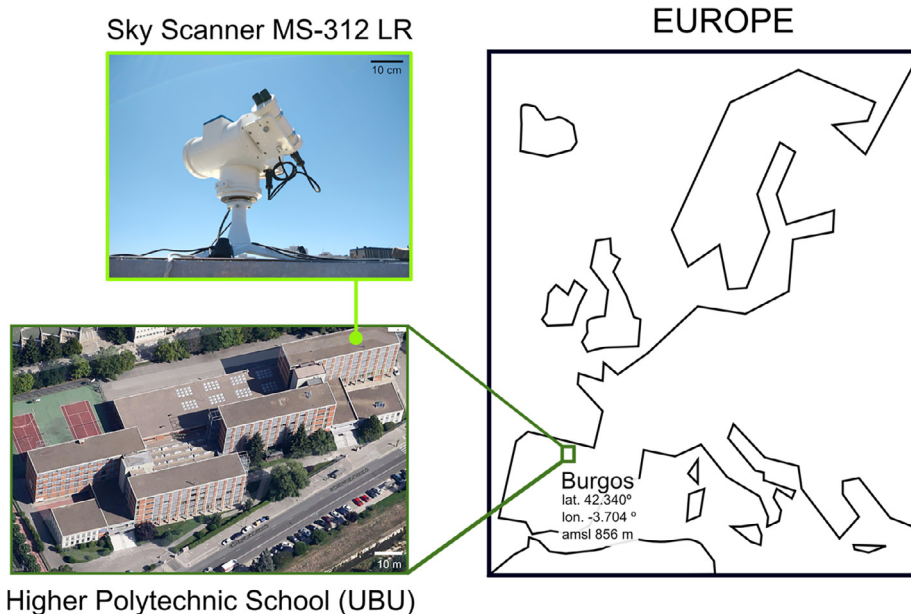


Fig. 4. Location of the experimental apparatus on the roof of the Higher Polytechnic School building at the University of Burgos, Spain.

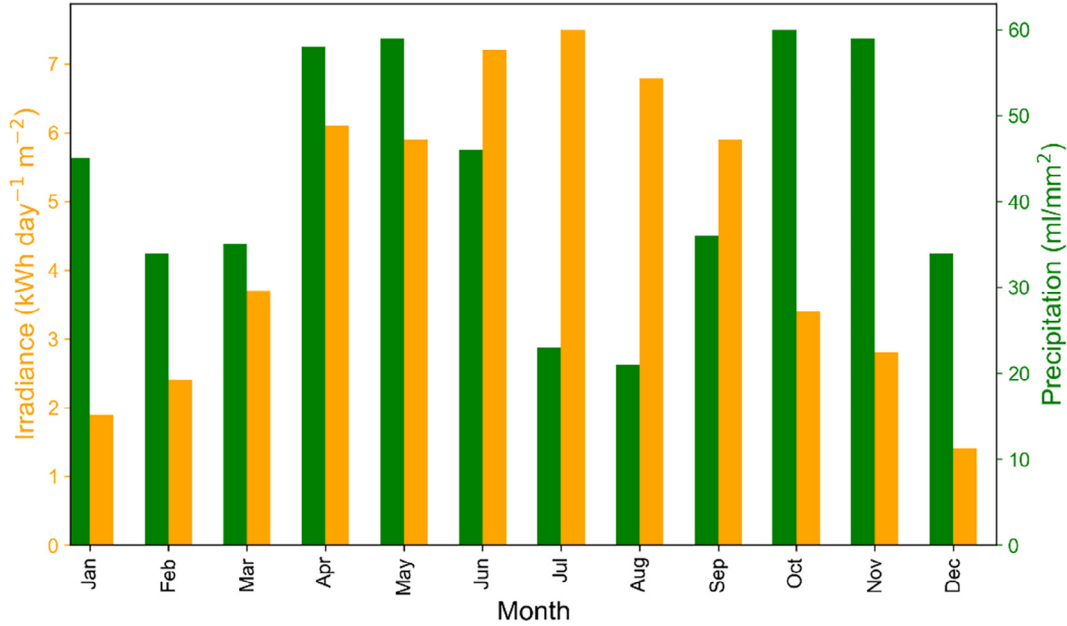


Fig. 5. Global irradiance and precipitation in Burgos (Ministerio de Agricultura y Pesca) (June 2016 to May 2017).

Table 2

Sky Scanner specifications.

Model	MS-321LR Sky Scanner
Manufacturer	EKO Instruments
Dimensions (W × D × H)	430 mm × 380 mm × 440 mm
Mass	12.5 kg
Aperture	11°
Illuminance	0–50 kcd/m ²
Radiance	0–300 W/m ² /sr
A/D convertor	16 bit
Calibration error	2%

distribution of the 145 patches shown in Fig. 6.

$$F_{g,p} = \frac{\pi}{n_p} \left\{ \sin^2 \left(\frac{\pi b_p}{15} \right) - \sin^2 \left(\frac{\pi (b_p - 1)}{15} \right) \right\} \text{ for } 1 \leq b_p \leq 7 \quad (3)$$

$$F_{g,p} = \pi \left\{ 1 - \sin^2 \left(\frac{\pi (b_p - 1)}{15} \right) \right\} \text{ for } b_p = 8.$$

The correction factor, F_C , is calculated as:

$$F_C = \frac{\pi}{\sum_p F_{g,p}} \quad (4)$$

where, the additions correspond only to the patches really considered for each scan. The normalized luminance distribution given for each patch, and for each sky type is given by:

$$L_{pr,sc} = \frac{L_p}{E_h} \quad (5)$$

where, $L_{pr,sc}$ is the previously normalized sky patch luminance, corresponding to one experimental measure. Additionally, the Tregenza method requires an estimate of the mean luminance of the Standard Sky type across each patch, p , of the scanning pattern, in order to make an accurate comparison with the measured values (Tregenza, 2004). So, we should take the mean of the luminances at the corners of the patch. The mean luminance of each Standard Sky type of the first 144 elements, is the average luminance obtained in the four corners by each patch, given by the coordinates $Z = \left\{ Z_p - \frac{\pi}{30}, Z_p + \frac{\pi}{30} \right\}$ and $\alpha = \left\{ \alpha_p - \frac{\pi}{n_b}, \alpha_p + \frac{\pi}{n_b} \right\}$, where Z_p is the zenithal angle of each center

patch, α_p is the azimuth angle, as shown in Fig. 8, and n_b is the number of patches in the band (see Fig. 6). Note that, for b_1 , (patches from 1 to 30) only two corners are used in the average. The area around the zenith is split into six triangles, so the resulting average is the sum of each triangle calculated within its vertices.

$$L_{s145} = \frac{1}{9} \left\{ 3L_{sZenith} + \sum_{a=1}^6 L_{sZenith} \left[\frac{\pi}{30}, \frac{a\pi}{3}, \frac{\pi}{6} \right] \right\} \quad (6)$$

It is also possible to obtain normalized theoretical measurements, by using Eq. (2) to calculate the corresponding horizontal illuminance. Starting from that point, if the luminance of each patch is divided by the previously calculated horizontal illuminance, $L_{pr,st}$, fifteen sets of average normalized luminance are obtained, one for each type of standard sky. Finally, each standard sky type will be compared with the previously measured and normalized ($L_{pr,st}$) experimental sky type.

The RMSE is obtained by comparing the measured patch luminance with standard sky type luminance, excluding the empty patches from the sum under the square root. The type of sky is obtained by picking out the lowest $RMS_{sc, st}$ from the fifteen possible types.

$$rms_{sc,st} = \sqrt{\frac{\sum (L_{pr,sc} - L_{pr,st})^2}{n}} \quad (7)$$

$L_{pr,st}$ is the normalized luminance of a sky patch, corresponding to a (CIE) standard sky, $L_{pr,sc}$ is the normalized luminance of a sky patch, corresponding to an experimental measurement; N is the number of measurements, excluding the empty patches. Finally, sc refers to an experimental type and st is the particular standard sky type that is tested.

3.2. The Littlefair normalization ratio (NR)

Luminance can be normalized according to the Normalization Ratio introduced by Littlefair (1994a,b) and described by Li and Tang (2008). This method was developed to be applied at locations where high solar altitudes dominate. The luminance normalization ratio (NR) is given by:

$$NR = \frac{\sum L_p \cos(\gamma) \sin(\gamma) d\gamma d\alpha}{\sum L_{s,p} \cos(\gamma) \sin(\gamma) d\gamma d\alpha} \quad (8)$$

where, L_p is the luminance (kcd/m²) measured by the Sky scanner,

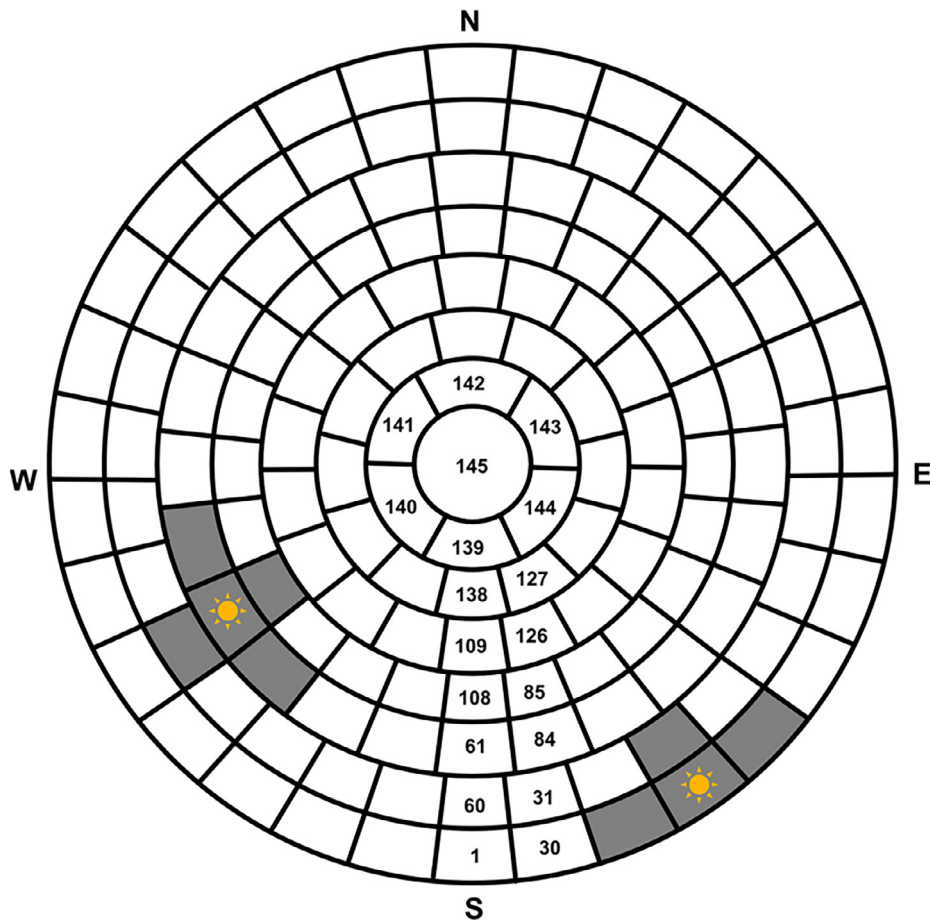


Fig. 6. Sky divided into 145 sectors (p) grouped into 8 bands (b_p). The number of patches per band (n_b) is shown in the figure. Patches adjacent to the position of the sun are excluded from the luminance calculation.

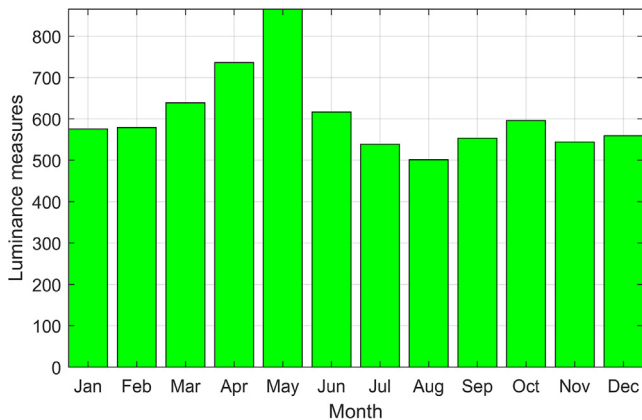


Fig. 7. Number of available luminance measurements for the study.

excluding luminance higher than 50 kcd/m^2 and lower than 0.1 kcd/m^2 and direct luminance values; $L_{s,p}$ is the predicted sky path luminance in a relative form given by CIE (Eq. (2)); γ is the angle of elevation of the sky point above the horizon (radians) and α is the azimuth of a sky patch (radians).

Finally, the predicted sky luminance value is calculated by multiplying the relative theoretical luminance by the Zenith and the normalization ratio, as follows:

$$L_{pred,p,sc} = L_{s,p}NR \tag{9}$$

After normalizing the luminance, the sky type is chosen from the best fit with the 15 CIE Standard skies, in accordance with the lowest

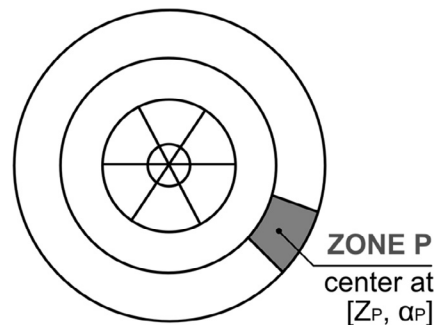


Fig. 8. Scheme of the average done to obtain the luminance of each patch.

$rRMSE_{sc}$ value:

$$rRMSE_{sc} = \sqrt{\frac{1}{n} \sum_i \left(\frac{L_{pred,sc} - L_p}{L_p} \right)^2} \tag{10}$$

N is the number of non-empty readings, p refers to the path of the sky or the number i , $L_{pred,p,sc}$ is the normalized luminance obtained using the NR factor and L_p is the luminance measured in kcd/m^2 by the Sky scanner.

4. Results

The classifications of the two previously described normalization procedures are compared in this study. In view of the differences, the overall annual comparison was disaggregated into hourly periods. First,

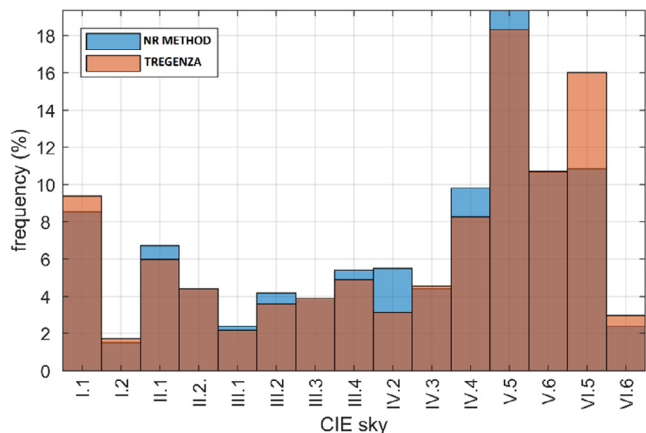


Fig. 9. Comparative characterizations of the skies over Burgos.

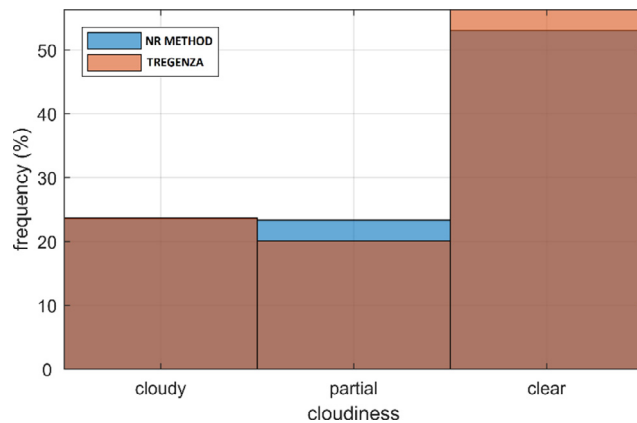


Fig. 11. Comparative study of cloudiness classifications in Burgos.

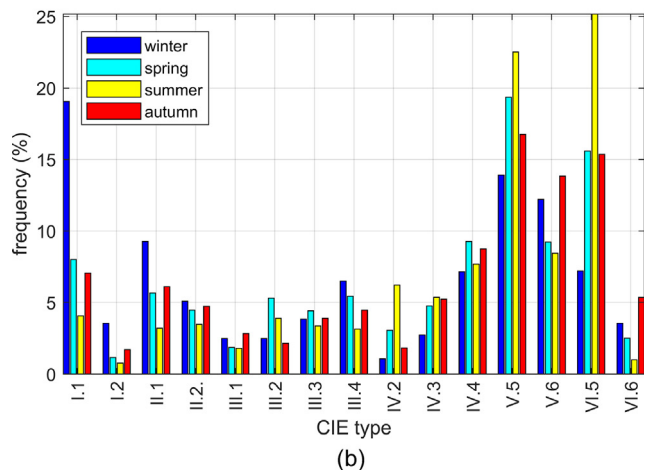
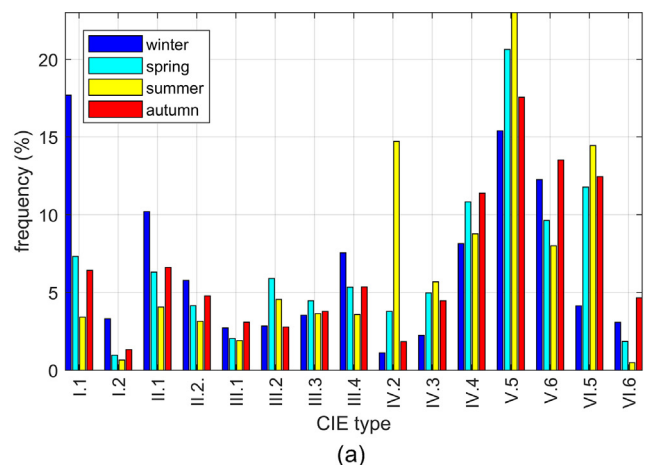


Fig. 10. Seasonal CIE sky types histogram over Burgos calculated using the NR method (a) and the Tregenza Method (b).

an annual comparison of the classifications of both algorithms was completed; second, the CIE and cloudiness types were grouped by seasons (summer, autumn, winter and spring) to identify the times of year with larger differences. The hourly relative differences of both models were estimated, by selecting the months of the season with the biggest discrepancies. Finally, all of the cloudiness models were compared in a confusion matrix. All of those analyses contribute to the characterization of the differences between both methods.

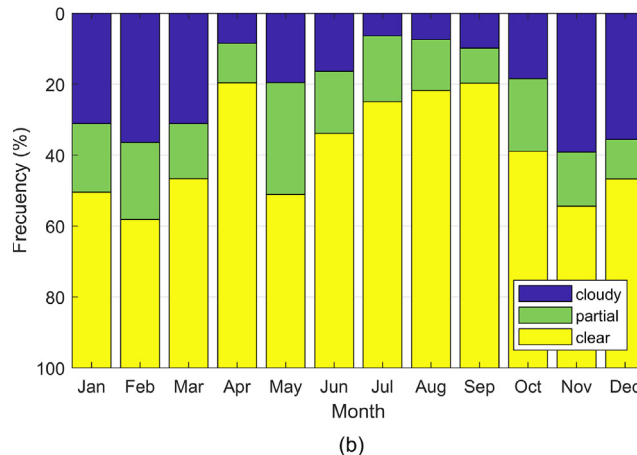
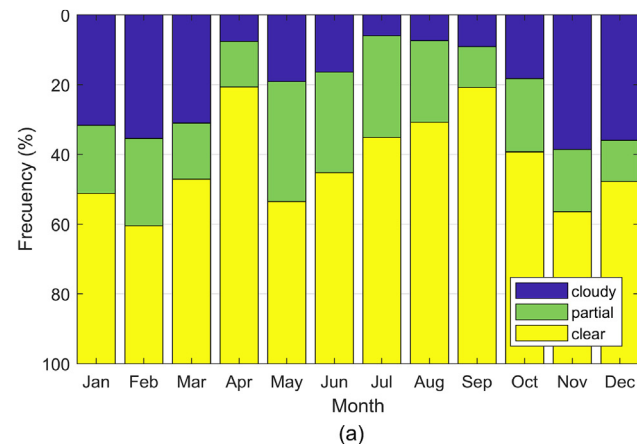


Fig. 12. Burgos sky cloudiness grouped by months using NR method (a) and Tregenza method (b).

4.1. Seasonal classification of skies in Burgos

Fig. 9 shows the relative frequency of each sky type over the period of study, calculated using both the Tregenza and the NR method. As can be seen in both figures, all sky types of the CIE classification, shown in Table 1, can be found in Burgos, from overcast to very clear. The lowest frequency is for type I.2, corresponding to Overcast with the steep gradation and slight brightening toward the sun, and the highest frequency is for type V.5. (Cloudless polluted with a broader solar corona). Both methods present very few differences and are almost equivalent in the II.2, IV.3 and V.5 types, as shown in Fig. 9. The biggest differences in the classification were found in types IV.2 and VI.5. These

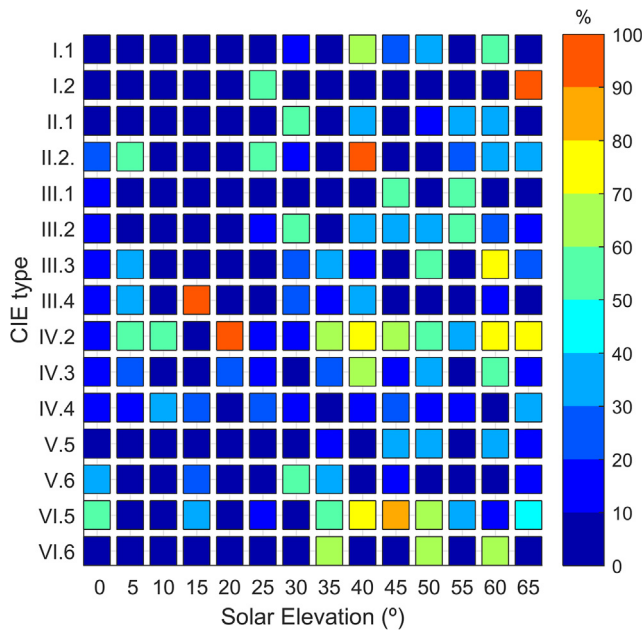


Fig. 13. Relative difference, d_r , calculated using Eq. (11), between the NR and the Tregenza sky classifications, for the summer months.

		cloudy	partial	clear	
CIE Tregenza	cloudy	1686 23.2%	42 0.6%	0 0.0%	97.6% 2.4%
	partial	34 0.5%	1375 18.9%	53 0.7%	94.0% 6.0%
	clear	0 0.0%	288 4.0%	3785 52.1%	92.9% 7.1%
		cloudy	partial	clear	CIE NR Method
		98.0% 2.0%	80.6% 19.4%	98.6% 1.4%	94.3% 5.7%

Fig. 14. Confusion matrix comparing the Tregenza and the NR cloudiness classifications.

Table 3
Statistical indicators RMSE and MBE, calculated using Eqs. (12) and (13).

	$\alpha \in (0, \frac{\pi}{2})$		$\alpha < 35^\circ$		$\alpha > 35^\circ$	
	rRMSE	rMBE	rRMSE	rMBE	rRMSE	rMBE
Tregenza method	36.7%	14.1%	34.9%	13.1%	40.1%	16.0%
NR method	36.0%	13.3%	33.4%	11.4%	40.8%	17.0%

results were grouped into different time intervals, in order to find a pattern that produces the aforementioned gap.

One characteristic of Burgos is that all types of CIE skies classification can be found throughout the year. This fact can be observed in

the monthly classification of the results of both methods shown in Fig. 10(a) and (b). Sky types I.1, I.2, V.5. and V.6 are predominant in winter (almost 10% frequency each). clear types of sky (IV.4, V.5.V.6 and VI.5) prevail in spring and summer time. In summer, type IV.2, corresponding to a partly cloudy sky is also frequent. Autumn is a clear sky season too.

Grouping the CIE types by seasons clearly profiles the switch between the IV.2 and VI.5 sky types. As can be seen in Fig. 10a, the NR method classifies 15% of the skies in summer as IV.2 and almost 22% as VI.5. In contrast, the Tregenza method classified 25% of the skies as VI.5, in the same season, as shown in Fig. 10b. It also labelled 6% of the recorded skies as IV.2. Sky type VI.5. appeared in 23% of cases when using the Tregenza method. It is evident that the mismatch is limited to 9% of all records. The differences between both methods in the other seasons of the year are insignificant.

The values of E_h would coincide for different types of sky in specific positions of the sun in the hours of sunrise and/or sunset (Kocifaj, 2012). In these cases, this magnitude, which is what the Tregenza method uses to perform the classification (Eq. (3)), can lead to an inadequate result. The NR method uses L_p , a magnitude that differs more in these sky conditions.

4.2. Cloudiness classification

Cloudiness labelling was done with the CIE sky types: I.1 to III.1 were classified as cloudy, III.2 to IV.3 as partially cloudy, and IV.4 to VI.6 as clear skies. These three categories, represented in Fig. 11, reflect the characteristically clear skies that it is predominant in Burgos. According to both methods, a clear sky type was present every month in almost 50% of cases, as can be seen in Fig. 12(a) and (b). An overcast sky type was observed in Burgos in less than 25% of cases, a situation more probably in winter and autumn, while in spring and summer the skies were mainly clear.

The differences in the cloudiness classifications of the methods are clearly visible in the months of June, July and August. The NR method (Fig. 12a) classified a lower percentage of clear skies than the Tregenza method (Fig. 12b) in the three aforementioned months. The percentages of the skies classified as cloudy, partially cloudy, and clear, over the remaining months are almost identical.

4.3. Analysis of summer time by daytime hour

The CIE relative differences in the classifications grouped by daytime hours during the summer months was prepared to examine the differences in greater detail. The results are shown in Fig. 13. There are several forms of estimating the relative difference between both magnitudes (Bennett and Briggs, 2008). In the present work, neither algorithm can be considered superior, because there are no qualitative differences between either one. Eq. (11) was used to estimate the relative difference between the frequencies grouped by daytime hours.

$$d_r = \frac{|x-y|}{\max(|x|,|y|)} \tag{11}$$

It may be easily noted that the main divergence between CIE sky types IV.2 and VI.5 are at midday hours. There are also other points where the relative difference is above 50%. However, their weight in the global percentages is irrelevant, as Fig. 13 shows. In summary, the main differences between the Tregenza and the NR methods are at highest solar altitudes of the year that take place in the central hours of the day during summer.

4.4. Confusion matrix of Tregenza and NR methods

The results of both the Tregenza and the NR methods were compared in a confusion matrix; an indicatrix of the matches between two series of values. The confusion matrix is shown in Fig. 14. In this case,

the cloudiness classifications of the two methods are compared step by step. Each square of the table shows the number of coincidences and their corresponding percentages. The upper-left 3×3 matrix corresponds to the raw comparison, coding the cells either in green, if Tregenza and NR agree, or in red, if otherwise. The gray cells are the percentages that count the total cases in each row or column. Finally, the blueish cell to the lower-left shows the extent of global matching. As it can be seen, the global coincidences for the cloudiness classification amount to 94.3%.

4.5. Comparison between the goodness indicators of each method

As previously explained, both methods used in this study define the type of sky based on a goodness indicator, the RMSE (Root Mean Square Error), but the definition of this statistical indicator is different for each method. While the NR method uses the luminance normalized ratio NR defined by Eq. (10), Tregenza uses the horizontal diffuse illuminance to normalize the values. The RMSE values calculated for both methods are therefore not comparable values.

One solution to this issue would be a new definition of the relative *rRMSE* coupled with another widely used statistical indicator, the relative Mean Bias Error (*rMBE*). Both indices are defined in Eqs. (12) and (13). The *rMBE* provides information on the grade of dispersion relating to the center of the distribution and is a good dispersion indicator of the model versus the reality (Bennett and Briggs, 2008). The comparison was done using the theoretical CIE Luminance that refers to the Zenith, without normalization, $L_{sp,st}$, and the experimental Luminance, L_{sp} , calculated using the above method:

$$rMBE_{sc} = \frac{1}{N} \sum_p \left(\frac{L_{sp} - L_{sp,st}}{L_{sp}} \right) \quad (12)$$

$$rRMSE_{sc} = \sqrt{\frac{1}{N} \sum_p \left(\frac{L_{sp} - L_{sp,st}}{L_{sp}} \right)^2} \quad (13)$$

N is the number of measurements, excluding the empty patches. Table 3 shows the results obtained using this new criterion. As can be seen, the unification of the normalization criterion permits the numerical comparison of both methods. Both methodologies, as mentioned throughout the study, are applicable to the area under study and both statistical indicators yielded similar results, offering low MBE values at high solar elevations.

5. Conclusions

Two different methodologies, Tregenza and NR, to define the CIE standard sky types in the skies over Burgos, Spain, have been applied and compared. Both methods are recommended for use in areas where the latitude is higher than 35° , which is the case of Burgos. The best-fitting sky types and their frequency of occurrence have been studied over a complete year. The low value of the RMSE index shows that both methods tend to get a very acceptable agreement between predictions and measured values, so both methods can be used with high confidence at the latitude of Burgos. However, this study shows that this confidence will decrease at the highest solar altitudes. In addition, this study supports the fact that, despite the crucial aspect of normalized luminance, some effective and very different methods exist. The NR method uses a statistical parameter (NR) while the Tregenza method uses the horizontal illuminance, although the results are realistic, accurate, and very similar. As the confusion matrix has shown, the matches between both models were very good, so it comes as no surprise that the frequency distribution was likewise very similar. The smooth differences in the frequency distribution of the sky type found can be explained by the consideration of homogenous skies inherent of the CIE standard classification. The luminance of a partly cloudy sky can vary over a wide range even if the cloud fraction is stable for a long time and

CIE standard classification does not account for such variability.

The aim of this research work has been to determine the frequency distribution of each Sky type, so as to obtain quantitative information on the levels of illumination and solar radiation on surfaces. Both methods confirm that the most frequent sky type observed in Burgos is V.5. (cloudless polluted with a broad solar corona), with a frequency of occurrence close to 20%. Nevertheless, the group of clear skies has a higher frequency (in almost 50% of the cases under study for both methods). The skies over Burgos are of an overcast type in less than 25% of cases, a situation with a greater likelihood in both winter and autumn, while in spring and summer the sky is predominantly clear. The summer skies over Burgos are very clear and the winter is quite cloudy. Although those results show clear skies over Burgos during the year that is under study, additional years of measurement will be needed to arrive at a clear picture of global behavior that excludes years of drought and excessive rainfall.

The generalization of the results will be given the more cases are performed by the scientific community. It can be expected seasonal behavior in the measurements and results obtained. Hence, it would be necessary a multi-year analysis in order to avoid the aforementioned bias. However, as it was said, given the amount of samples analyzed, the differences detected cannot be deprecated. Not only the development of new sky models is important in the field of science that concerns us, but also, their comparison and the discern of their strengths in order to use them in an optimum way.

Acknowledgements

The authors gratefully acknowledge the financial support provided by the Regional Government of Castilla y León (Ref. BU034U16), under European Regional Development Fund, and the Spanish Ministry of Economy, Industry and Competitiveness, under the I+D+i State Programme Challenges for the Society (Ref. ENE-2014-54601-R). David González Peña would also like to thank the Junta de Castilla-León for economic support (PIRTU Program, ORDEN EDU/301/2015). Finally, Andrés Suárez García extends his thanks to the International Excellence Triangular-E³ for financing his stay under the Centrally Managed Programme of the Ministry of Education, Culture & Sports (BOE 295, 10th December 2015).

References

- Aalto University School of Science and Technology, I.E.A., 2010. Guidebook on Energy Efficient Electric Lighting for Buildings. IEA publications, Raisio.
- Bennett, J.O., Briggs, W.L., 2008. Using and Understanding Mathematics: A Quantitative Reasoning Approach.
- Chaiwiwatworakul, P., Chirarattananon, S., 2004. Distribution of sky luminance in tropical climate. In: Proceedings of the Joint International Conference on Sustainable Energy and Environment, Hua Hin, pp. 530–537.
- Hwang, T., Jeong, T.K., 2011. Effects of indoor lighting on occupants' visual comfort and eye health in a green building. *Indoor Built Environ.* 20 (1), 75–90.
- ITACYL-AEMET, 2013. Atlas Agroclimático de Castilla y León <http://atlas.itacyl.es>. (accessed January, 2017).
- Kittler, R., Perez, R., Darula, S., 1997. A new generation of sky standards. In: Proceedings of the Eighth European Lighting Conference, Amsterdam, pp. 359–373.
- Kocifaj, M., 2011. CIE standard sky model with reduced number of scaling parameters. *Sol. Energy* 85 (3), 553–559.
- Kocifaj, M., 2012. Angular distribution of scattered radiation under broken cloud arrays: an approximation of successive orders of scattering. *Sol. Energy* 86 (12), 3575–3586.
- Kocifaj, M., Kómar, L., 2016. Modeling diffuse irradiance under arbitrary and homogeneous skies: comparison and validation. *Appl. Energy* 166, 117–127.
- Li, D.H.W., 2010. A review of daylight illuminance determinations and energy implications. *Appl. Energy* 87 (7), 2109–2118.
- Li, D.H.W., Chau, N.T.C., Wan, K.K.W., 2013. Predicting daylight illuminance and solar irradiance on vertical surfaces based on classified standard skies. *Energy* 53, 252–258.
- Li, D.H.W., Chau, T.C., Wan, K.K.W., 2014. A review of the CIE general sky classification approaches. *Renew. Sustain. Energy Rev.* 31, 563–574.
- Li, D.H.W., Lau, C.C.S., Lam, J.C., 2003. A study of 15 sky luminance patterns against hong kong data. *Archit. Sci. Rev.* 46 (1), 61–68.
- Li, D.H.W., Lau, C.C.S., Lam, J.C., 2004. Standard skies classification using common climatic parameters. *J. Sol. Energy Eng. Trans. ASME* 126 (3), 957–964.
- Li, D.H.W., Tang, H.L., 2008. Standard skies classification in Hong Kong. *J. Atmos. Sol.-*

- Terr. Phys. 70 (8), 1222–1230.
- Li, D.H.W., Tang, H.L., Lee, E.W.M., Muneer, T., 2010. Classification of CIE standard skies using probabilistic neural networks. *Int. J. Climatol.* 30 (2), 305–315.
- Lima, F.J.L., Martins, F.R., Pereira, E.B., Lorenz, E., Heinemann, D., 2016. Forecast for surface solar irradiance at the Brazilian Northeastern region using NWP model and artificial neural networks. *Renew. Energy* 87, 807–818.
- Littlefair, P.J., 1994. A comparison of sky luminance models with measured data from Garston, United Kingdom. *Sol. Energy* 53 (4), 315–322.
- Littlefair, P.J., 1994b. The luminance distributions of clear and quasi-clear skies. In: *Proceedings of the CIBSE National Lighting Conference*, Cambridge, UK, pp. 267–283.
- Lou, S., Li, D.H.W., Lam, J.C., 2017. CIE Standard Sky classification by accessible climatic indices. *Renew. Energy* 113, 347–356.
- Markou, M.T., Kambezidis, H.D., Bartzokas, A., Katsoulis, B.D., Muneer, T., 2005. Sky type classification in Central England during winter. *Energy* 30(9 SPEC. ISS.), 1667–1674.
- Markou, M.T., Kambezidis, H.D., Katsoulis, B.D., Muneer, T., Bartzokas, A., 2004. Sky type classification in South England during the winter period. *Build. Res. J.* 52 (1), 19–30.
- Ministerio de Agricultura y Pesca, A.Y.M.A., SIAR: Sistema de Información Agroclimática para el Regadío. <http://www.mapa.es/siar/Informacion.asp> (accessed January, 2018).
- Torres, J.L., de Blas, M., García, A., Gracia, A., de Francisco, A., 2010. Sky luminance distribution in Pamplona (Spain) during the summer period. *J. Atmos. Sol.-Terr. Phys.* 72 (5–6), 382–388.
- Torres, J.L., de Blas, M., García, A., Gracia, A., de Francisco, A., 2010. Sky luminance distribution in the North of Iberian Peninsula during winter. *J. Atmos. Sol.-Terr. Phys.* 72 (16), 1147–1154.
- Torrington, J.M., Tregenza, P.R., 2007. Lighting for people with dementia. *Light. Res. Technol.* 39 (1), 81–97.
- Tregenza, P.R., 1999. Standard skies for maritime climates. *Light. Res. Technol.* 31 (3), 97–106.
- Tregenza, P.R., 2004. Analysing sky luminance scans to obtain frequency distributions of CIE Standard General Skies. *Light Res. Technol.* 36 (4), 271–279.
- Uetani, Y., Aydinli, S., Joukoff, A., Kendrick, J., Kittler, R., Koga, Y., 2003. *Spatial Distribution of Daylight-CIE Standard General Sky*. Vienna, Austria.
- Wittkopf, S.K., Soon, L.K., Ng, E.Y.Y., 2007. Analysing sky luminance scans and predicting frequent sky patterns in Singapore. *Light Res. Technol.* 39 (1), 31–51.

Paper II

Granados-López, D., Díez-Mediavilla, M., Dieste-Velasco, M.I., Suárez-García, A. and Alonso-Tristán, C, 2020. *Evaluation of the vertical sky component without obstructions for daylighting in Burgos, Spain*. *Appl. Sci.* 10 (9), 3095. DOI:10.3390/app10093095

Abstract: Daylight availability knowledge is the first step for an energetic and visually efficient building and city design. It can be estimated with the Vertical Sky Component (VSC), which is defined as the ratio of the vertical diffuse illuminance over the unobstructed horizontal diffuse illuminance, simultaneously measured at the same point. These illuminance magnitudes are obtained from luxmeter measurements but these data are scarce. Alternatively, VSC can be obtained from prior knowledge of the sky illuminance distribution, which can be measured with a sky scanner device or by reference to the CIE (Commission Internationale de L'Éclairage) Standard classification for homogeneous skies. Both approaches are compared in this study. The coherence of the results obtained for the four cardinal orientations are analyzed by applying classical statistical parameters and luxmeter measurements as references for the results. The measurement campaign was completed between September 2016 and January 2019 in Burgos (Spain), as representative case study and specific contribution of this work. It was observed that the VSC values were higher than 100 in many cases: 21.94% for the south- and 33.6% for the east-facing vertical surfaces. The study highlights the good daylighting conditions in Burgos, mainly due to the predominance of clear skies over much of the year. This fact implies high daylight availability that, with efficient city planning and building design, could potentially lead reduction energy consumption of buildings, improvements in visual comfort, and the well-being of occupants.

Article

Evaluation of the Vertical Sky Component without Obstructions for Daylighting in Burgos, Spain

Diego Granados-López ¹, Montserrat Díez-Mediavilla ¹, M. Isabel Dieste-Velasco ¹,
Andrés Suárez-García ^{1,2}  and Cristina Alonso-Tristán ^{1,*} 

¹ Research Group Solar and Wind Feasibility Technologies (SWIFT), Electromechanical Engineering Department, Universidad de Burgos, 09006 Burgos, Spain; dgranados@ubu.es (D.G.-L.); mdmr@ubu.es (M.D.-M.); midieste@ubu.es (M.I.D.-V.); andres.suarez@tud.uvigo.es (A.S.-G.)

² Centro Universitario de la Defensa, Escuela Naval Militar de Marín, 36920 Marín, Pontevedra, Spain

* Correspondence: catristan@ubu.es or cristinaalonso.tristan@gmail.com; Tel.: +0034-947-258-853

Received: 23 March 2020; Accepted: 22 April 2020; Published: 29 April 2020



Featured Application: Knowledge of Vertical Sky Component (VSC) allows the calculation of daylighting availability for buildings at any cardinal orientation for energetic and visually efficient building and city design. This work describes different alternatives for VSC calculation and a complete experimental characterization of the VSC in an extensive case study carried out in Burgos, Spain.

Abstract: Daylight availability knowledge is the first step for an energetic and visually efficient building and city design. It can be estimated with the Vertical Sky Component (VSC), which is defined as the ratio of the vertical diffuse illuminance over the unobstructed horizontal diffuse illuminance, simultaneously measured at the same point. These illuminance magnitudes are obtained from luxmeter measurements but these data are scarce. Alternatively, VSC can be obtained from prior knowledge of the sky illuminance distribution, which can be measured with a sky scanner device or by reference to the CIE (Commission Internationale de L'Éclairage) Standard classification for homogeneous skies. Both approaches are compared in this study. The coherence of the results obtained for the four cardinal orientations are analyzed by applying classical statistical parameters and luxmeter measurements as references for the results. The measurement campaign was completed between September 2016 and January 2019 in Burgos (Spain), as representative case study and specific contribution of this work. It was observed that the VSC values were higher than 100 in many cases: 21.94% for the south- and 33.6% for the east-facing vertical surfaces. The study highlights the good daylighting conditions in Burgos, mainly due to the predominance of clear skies over much of the year. This fact implies high daylight availability that, with efficient city planning and building design, could potentially lead reduction energy consumption of buildings, improvements in visual comfort, and the well-being of occupants.

Keywords: VSC; daylighting; diffuse illuminance; CIE standard sky classification

1. Introduction

Daylighting is a design concept of buildings used over the years with varying degrees of success. It has been proven that buildings with good daylighting have positive effects on the well-being and health of occupants, reducing stress levels, and improving mood and photobiological effects [1–5]. The International Energy Agency (IEA) promoted the use of daylight as a means of reducing electricity consumption on lighting [6], and increasing the energy efficiency of buildings [7,8]. International Net

Zero Energy Building (NZEB) standards and regulations recommend the incorporation of natural lighting strategies in their design and define minimum standards [9,10].

A fundamental step towards building design that illuminates interiors with daylight is the compilation of information on daylight availability outside the building. Illuminance on vertically oriented surfaces is, therefore, important for modeling the daylight availability, especially for high-rise buildings with substantial areas of glazed surfaces [11]. Solar radiation data on vertical surfaces can also be used to evaluate the performance of Building Integrated Photovoltaics (BIPV), because the vertical facades of modern cities occupy larger areas than roof surfaces and usually present better maintenance conditions for photovoltaic panels [12]. However, the basic outdoor solar irradiance and illuminance data for the surfaces of interest are not usually available in many parts of the world [13].

The Vertical Sky Component (VSC) is defined as the ratio of the vertical diffuse illuminance ($E_{v,d}$) over the unobstructed horizontal diffuse illuminance ($E_{h,d}$) simultaneously measured at the same point [14]. City planners and developers use the VSC to assess the impact of newly built constructions on the access to daylighting. Its potential in early stage design decisions is widely recognized [15]. The VSC is an important daylight parameter, which has been also used for different applications: Sky classification [16,17], unobstructed sky irradiance and illuminance determinations [18], and calculations of indoor illuminance [19]. Global and diffuse outdoor illuminance on a horizontal surface and vertical global illuminance on the four (north, east, south, and west) cardinal orientations have been used to compute the VSC. However, as experimental data of L_{dv} and L_{dh} are scarce, different alternatives are used to calculate VSC [11,17,20].

With a well-defined luminance distribution, the daylight illuminance on various inclined surfaces, such as vertical planes facing towards different directions, can be estimated by integrating the luminance distribution of the sky dome over each surface [21]. The usual instrument for measuring sky luminance distribution is by means of a sky scanner [8]. On the other hand, empirical models of homogeneous skies represent a low-cost approach to obtain sky luminance distribution. Generally, sky conditions can be subdivided into overcast, partly cloudy, and clear skies. In 2003, CIE [22] and the International Organization for Standardization (ISO) [23] adopted 15 Standard Sky Luminance Distributions (SSLD) as the most versatile definition of skylight in various localities and daylight climate regions, making it possible to simulate annual daylight profiles in absolute units based on typical luminance sky patterns. Despite the high interest in those measurements, very few studies at only a handful of European [24–29] and Asian [15,30,31] locations have been conducted to characterize the sky under the CIE standard, mainly due to scarce sky luminance data obtained from sky scanner devices.

Different alternatives to the use of sky scanners have been proposed for classifying the skies [32] including the use of different climatic parameters [33–36], vertical [12,20] and horizontal illuminance [37,38], and satellite data [39]. VSC has been proposed as an appropriate method, both for categorizing the CIE Standard General Skies and for determining daylight illuminance on inclined surfaces with various orientations [17,32]. The VSC facing a given orientation at a given time under an individual CIE Standard General Sky has its own features and stands for a unique value. So, the use of the CIE standard classification could represent an alternative means of estimating VSC.

As it has been pointed out, the determination of daylight availability is a complex issue due to scarce experimental data needed for all methodologies proposed. Therefore, it is mandatory the establishment of the equivalence among the different methodologies for obtaining daylighting through any of the proposed methods and even the possible combination between them when necessary.

This study focused on the determination of VSC in Burgos, Spain, for which purpose different alternative proposals for its calculation are compared. Three methods are generally used to calculate VSC: Available experimental data of horizontal global, diffuse, and beam illumination; through the data from experimental sky scanner measurements; and the CIE standard classification for homogeneous skies. The results obtained under different sky conditions were tested in a long-term experimental test campaign over 29 months. The VSC data obtained from the experimental data of global horizontal, diffuse, and beam illuminance were taken as the reference data and two statistical indicators, root

mean square deviation (RMSD) and mean bias deviation (MBD), were used as quality indicators of the equivalence of the different approaches. Finally, the daylight conditions in Burgos, Spain, were determined with the VSC calculations for surfaces facing all four cardinal orientations.

The paper will be structured as follows. The experimental facility and the measurement campaign as well the quality filters applied to the experimental data will be described in Section 2. The different approaches used for the VSC calculations will be defined in Section 3 and the results of all three methodologies with their statistical indicators will be compared in Section 4. In Section 5, the VSC in Burgos will be calculated for all four cardinal orientations and the results will be compared with the existing knowledge of the CIE standard sky type, analyzing both frequency of occurrence and seasonal and monthly distributions. Finally, the main results and the conclusions of the study will be summarized in Section 6.

2. Experimental Data

The experimental data for this study were recorded at a meteorological weather station located on the roof of the Higher Polytechnic School building of Burgos University ($42^{\circ}21'04''$ N, $3^{\circ}41'20''$ W, 856 m above mean sea level). The flat roof of this five-story building, in an area with no other buildings of comparable height, is ideal for recording measurements with no external obstructions or reflections from other surfaces. The experimental equipment is shown in Figure 1.



Figure 1. Location of the experimental equipment on the roof of the Higher Polytechnic School building at University of Burgos, Spain.

The daylight measurements included the global, diffuse, and beam outdoor illuminance on a horizontal plane; the vertical global components facing the four cardinal orientations (north, east, south, and west); and the luminance distributions for the whole sky, recorded with seven lumeters model ML-020S-O luxmeters, and a commercial MS-321LR sky scanner both from EKO Instruments (EKO Instruments Europe B.V. Den Haag, Netherlands). The diffuse luxmeter was obscured from direct sunlight by a shadow hat. The direct solar illuminance (L_{bh}) sensor was installed on a sun tracker (model SunTracker-3000, Geónica, Madrid, Spain) facing the sun. The technical specifications of the sky scanner and luxmeter are shown in Tables 1 and 2, respectively.

Illuminance data were recorded every 10 min (averaging recorded scans of 30 s). The illuminance data (horizontal global, L_{gh} , diffuse, L_{dh} , and beam, L_{bh}) were analyzed and then filtered using CIE quality criteria, shown in Table 3. The sky scanner was adjusted on a monthly basis for taking measurements from sunrise to the sunset. It completed a full scan in four minutes and started a

new scan every 15 min. The first and last measurement of the day ($\alpha_s \leq 5^\circ$) were discarded, as well measurements higher than 50 kcd/m² and lower than 0.1 kcd/m², following the specifications of the equipment. Only half-hourly and hourly sky scanner measurements were used in this study, to match simultaneous records of illuminance data. If the illuminance data failed to pass the quality criteria, then all the simultaneous data sets were rejected.

Table 1. Sky scanner specifications.

Model	MS-321LR Sky Scanner
Dimensions (W × D × H)	430 mm × 380 mm × 440 mm
Mass	12.5 kg
FOV	11°
Luminance	0 to 50 kcd/m ²
Radiance	0 to 300 W/m ²
A/D Convertor	16 bits
Calibration Error	2%

Table 2. Luxmeter technical specifications.

Model	ML-020S-O
Irradiance Range	0 to 150,000 lux
Output	0 to 30,000 μV
Impedance	280 Ω
Operating temperature range	−10 °C to 50 °C
Temperature response	0.4%

Table 3. Quality criteria for experimental illuminance measurements. The experimental data were discarded if any of the following conditions were met. L₀ is the luminous solar constant (133.8 kLx).

(a) $L_{gh} > 1.2 \cdot L_0$	(e) L_{gh}, L_{dh}, L_{bh} if $\alpha_s < 5^\circ$
(b) $L_{dh} > 0.8 \cdot L_0$	(f) $L_{dh} > 1.15 \cdot L_{gh}$
(c) $L_{bh} > L_0$	(g) $L_{gh} = (L_{bh} \cdot \cos Z_s + L_{dh}) \pm 15\%$
(d) $L_{gh}, L_{dh}, L_{bh} < 20 \text{ Lux}$	

The measurement campaign ran from September 2016 up until January 2019. A total of 3991 datasets of illuminance (horizontal global, diffuse, and beam and vertical global in the four cardinal orientations) and sky scanner measurements passed the quality criteria.

3. Determination of VSC

The Vertical Sky Component (VSC) is defined as the ratio of the vertical diffuse illuminance, L_{dv} , over the unobstructed horizontal diffuse illuminance, L_{dh} , simultaneously measured at the same point, as shown in Equation (1).

$$VSC = \frac{L_{dv}}{L_{dh}} \tag{1}$$

In this study, the measured global, beam, and diffuse illuminance on a horizontal surface and vertical global illuminance on vertical surfaces facing the four cardinal orientations (north, east, south and west) were used to compute the VSC for subsequent analysis. As L_{dv} data were not available at the experimental facility described in the Section 2, these data will be obtained, considering the similar characteristics of both solar irradiance and illuminance. The models developed for solar irradiance prediction on a vertical surface can, therefore, also be employed for daylight calculations [12]. The illuminance incident on a vertical surface, L_{gv} , was evaluated as the sum of beam, L_{bv} , sky-diffuse, L_{dv} , and ground-reflected, L_{rv} , components. It can be written as:

$$L_{gv} = L_{bv} + L_{dv} + L_{rv} \tag{2}$$

Calculation of beam illuminance is quite straightforward, given the position of the sun and the orientation of the surface, as shown in Equation (3):

$$L_{bv} = \left(\frac{L_{bh}}{\sin \alpha_s} \right) \cdot \cos \alpha_s \cdot \cos(\phi_s - \phi_{Nr}), \tag{3}$$

where α_s is the solar elevation, ϕ_s is the azimuth angle of the sun, and, ϕ_{Nr} is the azimuth vertical surface angle (which is 0 for the south-facing surfaces, $-\frac{\pi}{2}$ for the vertical east surface, $+\frac{\pi}{2}$ for the vertical west surface, and π for the north surface). It was assumed when estimating the ground-reflected illuminance, L_{rv} , that vertical surfaces will receive half of the total illuminance reflected isotropically from the ground, ignoring the inter-reflection between the vertical and the ground surfaces. L_{rv} depends on the albedo, ρ , and the global horizontal illuminance, L_{gh} , as it is shown in Equation (4). In this work, the value of $\rho = 0.1519$ was experimentally obtained [40].

$$L_{rv} = \frac{\rho}{2} \cdot L_{gh}. \tag{4}$$

L_{dv} can then be calculated with Equation (5). The VSC was obtained from the experimental data of horizontal global, diffuse, and beam illuminance, as shown in Equation (6):

$$L_{dv} = L_{gv} - \frac{\rho}{2} \cdot L_{gh} - \left(\frac{L_{bh}}{\sin \alpha_s} \right) \cdot \cos \alpha_s \cdot \cos(\phi_s - \phi_{Nr}). \tag{5}$$

$$VSC = \frac{L_{gv} - \frac{\rho}{2} L_{gh} - \left(\frac{L_{bh}}{\sin \alpha_s} \right) \cos \alpha_s \cdot \cos(\phi_s - \phi_{Nr})}{L_{dh}}. \tag{6}$$

If the luminance distribution in the sky dome is known, one alternative to determine the VSC is through integration. With a well-defined luminance distribution, the daylight on any surface can be estimated by integrating the luminance distribution of the sky dome over each surface. Traditionally, sky luminance distribution has been measured by sky-scanner instruments. These instruments divide the sky hemisphere into a limited number of sky elements (patches) of finite angular size and continuously scan the luminance data corresponding to each patch, L_p . Each sky patch will, it is expected, be treated as point sources with negligible error, but integration over all sky luminance scanning patches is only an approximation of total illuminance. Figure 2 shows a typical sky hemisphere sectorization into 145 patches and the geometric magnitudes that characterize each patch.

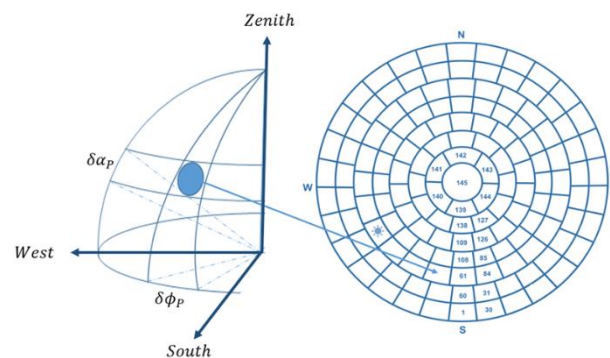


Figure 2. Angles defining the position of a sky element and the pattern of the 145 sky patches.

When sky scanner measurements are available, the horizontal and vertical diffuse illuminance can be obtained from Equations (7) and (8), respectively [37,41,42]:

$$L_{dh} = \sum_{p=1}^{145} L_p \sin \alpha_p \cdot \cos \alpha_p \cdot \delta \alpha_p \cdot \delta \phi_p, \tag{7}$$

$$L_{dv} = \sum_{p=1}^{145} \Delta E_{LvDp} \tag{8}$$

In Equation (7), L_p is the luminance of patch p , α_p is the angle of elevation of a patch above the horizon, $\delta\alpha_p$ is the altitude length of patch p ($\frac{\pi}{28}$ for all patches), and $\delta\phi_p$ is the azimuth length of patch p . In Equation (8), ΔE_{LvDp} represents the contribution of each patch to the vertical diffuse illuminance, calculated with Equation (9):

$$\Delta E_{LvDp} = D_p \cdot L_p \cdot \cos \alpha_p \cdot \delta\alpha_p \cdot \delta\phi_p \tag{9}$$

D_p is a geometrical factor which depends on the patch position in the sky dome, defined by Equation (10) [32]:

$$D_p = \begin{cases} \cos \alpha_p \cdot \cos(\phi_p - \phi_{Nr}) & \text{if } 0 \leq |\phi_p| \leq 90^\circ \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

where ϕ_p is the azimuth of the patch p and ϕ_{Nr} is the azimuth of the vertical surface.

Knowledge of the sky luminance distribution can be obtained through the CIE standard sky classification, which establishes a biunivocal relationship between the sky type and the sky luminance pattern. If the CIE standard sky classification is known, the luminance in each sky patch, L_p , can be obtained using Equation (11):

$$L_p = \left(1 + c \cdot \left[e^{d\chi} - e^{\frac{d\pi}{2}} \right] + e \cdot \cos^2 \chi \right) \left(1 + a \cdot e^{\frac{b}{\cos Z_p}} \right) \tag{11}$$

Coefficients $a, b, c, d,$ and $e,$ are defined by CIE [23] as functions of the sky type, as shown in Table 4. Z_p is the sky element zenith angle and χ is the dispersion angle, calculated from Equation (12):

$$\chi = \arccos(\cos Z_s \cos Z_p + \sin Z_s \sin Z_p \cos|\phi_p - \phi_s|) \tag{12}$$

where ϕ_p is the azimuth angle of the sky element, $p,$ and Z_s and ϕ_s are the zenith and azimuth angles of the sun. The χ represents the shortest angular length between the sky element, $p,$ and the sun, as is shown in Figure 3. Once the luminance in each sky element is known, Equations (7)–(10) can be applied to calculate L_{dh}, L_{dv} and, therefore, VSC. Figure 4 shows the VSC calculation for each CIE standard sky type as a function of the dispersion angle, $\chi.$ As can be seen, sky types 1, 3, and 5 present constant VSC values, regardless of $\chi.$ Therefore, the orientation of the vertical surface has no effect on the level of illumination. The predicted VSC values for these sky types (1, 3, and 5) were 38.5%, 45% and 50%, respectively.

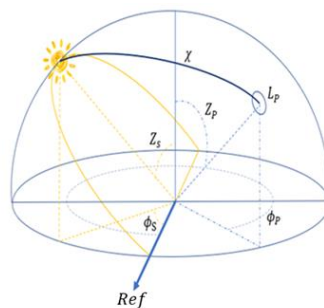


Figure 3. The angles of the sun and the sky element $p.$

The VSC showed a great dependency on the scattering angle of the other CIE standard sky types, with a minimum value of around 30% and a scattering angle close to $120^\circ,$ as shown in Figure 4.

Table 4. Parameters of CIE standard sky types [22].

Sky	a	b	c	d	e	Sky Description
1	4	-0.7	0	1	0	CIE Standard Overcast Sky,
2	4	-0.7	1	-1.5	0.5	Overcast, with steep luminance gradation and slight brightening towards the sun
3	1.1	-0.8	0	-1	0	Overcast, moderately graded with azimuthal uniformity
4	1.1	-0.8	2	-1.5	0.15	Overcast, moderately graded and slight brightening towards the sun
5	0	-1	0	-1	0	Sky of uniform luminance
6	0	-1	2	-1.5	0.15	Partly cloudy sky, no gradation towards zenith, slight brightening towards the sun
7	0	-1	5	-2.5	0.3	Partly cloudy sky, no gradation towards zenith, brighter circumsolar region
8	0	-1	10	-3	0.45	Partly cloudy sky, no gradation towards zenith, distinct solar corona
9	-1	0.55	2	-1.5	0.15	Partly cloudy, with the obscured sun
10	-1	0.55	5	-2.5	0.3	Partly cloudy, with brighter circumsolar region
11	-1	0.55	10	-3	0.45	White-blue sky with distinct solar corona
12	-1	0.32	10	-3	0.45	CIE Standard Clear Sky with low polluted atmosphere
13	-1	0.32	16	-3	0.3	CIE Standard Clear Sky, polluted atmosphere
14	-1	0.15	16	-3	0.3	Cloudless turbid sky with broad solar corona
15	-1	0.15	24	-2.8	0.15	White-blue turbid sky with broad solar corona

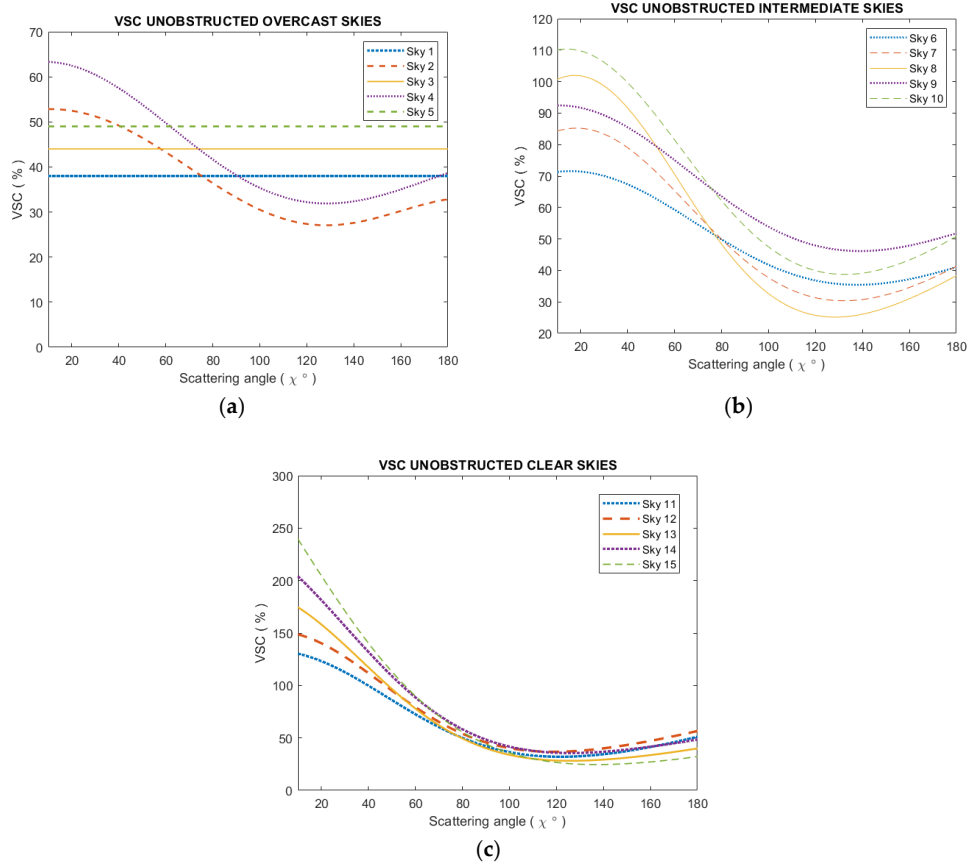


Figure 4. Vertical Sky Component (VSC) values for the different CIE sky types as a function of the scattering angle, χ : (a) CIE standard sky types 1–5, (b) CIE standard sky types 6–10, (c) CIE standard sky types 11–15.

4. Comparative Study of the Different Approaches to Obtain the VSC

Three different procedures to calculate the VSC have been described. In this section, the results of the different approaches to calculate the VSC in Burgos during the experimental campaign will be compared, taking the VSC calculated by Equation (6) as a reference. The comparison was conducted on a half-hourly basis. The two widely used statistical parameters, root mean square deviation, RMSD, and mean bias deviation, MBD, calculated in Equations (13) and (14), respectively, were chosen as indices to assess any differences between the procedures:

$$\text{RMSD} = 100 \cdot \sqrt{\frac{1}{n} \sum_i \left(\frac{X_{\text{estimation}} - X_{\text{measured}}}{X_{\text{measured}}} \right)^2} \tag{13}$$

$$\text{MBD} = 100 \cdot \frac{1}{n} \sum_i \frac{X_{\text{estimation}} - X_{\text{measured}}}{X_{\text{measured}}} \tag{14}$$

Information on the long-term performances of the estimation is given by the MBD. The RMSD indicates the data scattering around the estimation and gives information on the short-term performance.

The calculation of the diffuse illuminance on surfaces of any inclination and orientation is carried out by means of Equations (7)–(10), from the measurements of the sky scanner. An estimation of the error in the calculation of the diffuse illuminance can be obtained comparing the experimental data of horizontal diffuse illuminance, L_{dh} , and the one that is calculated from the projection on the horizontal surface of the sky luminance pattern obtained with the sky scanner data. The RMSD and the MBD, both for the city of Burgos and for all experimental results of L_{dh} recorded during the experimental campaign, were 27.14% and -3.8% , respectively. Hence, the use of the sky scanner data from the device used in this study (EKO, model MS-321LR) to determine illuminance on inclined surfaces facing any orientation had an implicit error higher than 25% in the city of Burgos. This error was caused by three main factors. First, the field of view (FOV) of the sky scanner (11° as indicated in Table 1). Secondly, the main assumption of the measurement with the sky scanner: As shown in Figure 2, it was assumed that the luminance of a sky patch was the value measured in the circumference inscribed within that patch. Thirdly, every scanning time was about 4 min and the measurements were taken every 15 min. Therefore, important variations in sky illuminance may occur between records.

The experimental data of L_{dv} were obtained from Equation (5). Those values, calculated on the four cardinal facing surfaces, were compared to those calculated as projections on the same surface using the sky-scanner luminance pattern. Both the RMSD and the MBD values are recorded in Table 5 and, as can be appreciated, are comparable to those obtained for horizontal diffuse illuminance. It can, therefore, be concluded from this study that the use of the sky-scanner measurements, to determine the diffuse illuminance on any horizontal or tilted surface, had an intrinsic RMSD due, mainly, to the technical specifications of the experimental device, near 30%, agreeing with other works [11].

Table 5. RMSD and MBD calculated for the vertical diffuse illuminance, L_{dv} , and the horizontal diffuse illuminance, L_{dh} . The reference values of L_{dv} were calculated from Equation (5). The reference values of L_{dh} were measured in the experimental facility, described in Section 2.

	Orientation	RMSD (%)	MBD (%)
Vertical	South surface	27.31	-3.55
	North surface	21.46	-0.31
	East Surface	31.19	-11.87
	West Surface	27.69	4.63
Horizontal		27.14	-3.80

As has been previously stated, when CIE standard classification is available, the sky luminance distribution can be obtained from Equation (11), and Equations (7)–(10) can be applied to calculate L_{dh} , L_{dv} and, therefore, VSC. In this work, the CIE standard sky type in Burgos was determined between September 2016 and January 2019, following the procedure described in a previous paper [26]. The frequency of occurrence (FOC) of each sky type during the period under study is shown in Figure 5. As can be seen, all types of CIE standard skies can be found in Burgos. The most frequent sky type was sky 7, (partly cloudy sky, no gradation towards zenith, brighter circumsolar region), with a FOC of almost 13%. Sky types 12, 13, and 11, corresponding to CIE standard clear sky categories, had FOCs of around 11%, similar to the FOC of sky type 8 (partly cloudy sky, no gradation towards zenith, distinct solar corona). Only a category of cloudy sky, sky type 4 (overcast, moderately graded, and slight brightening towards the sun), had a FOC close to 10%, while the FOC values of the other overcast categories were very low.

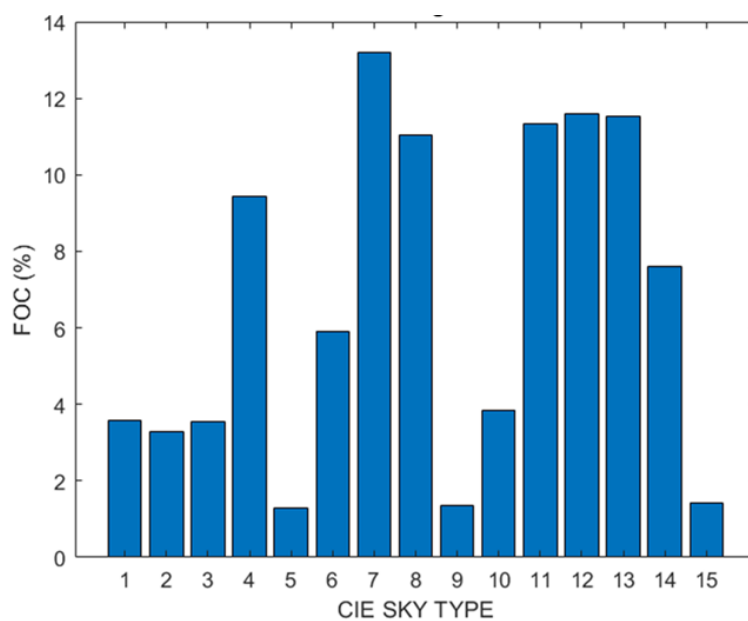


Figure 5. Frequency of occurrence (FOC, %) of CIE standard skies in Burgos, Spain, between September 2016 and January 2019.

VSC values calculated from the different procedures were compared, taking the VSC values obtained by Equation (6) as a reference. Both the RMSD and the MBD parameters are shown in Table 6. As can be seen, for the different vertical surfaces facing the cardinal orientations, the statistical indices ranged between 23% and 32% for the RMSD and between -1% and 16% for the MBD. Both procedures underestimated the VSC, as shown by the negative MBD values. The highest discrepancies between the different approaches were for the east orientation. This observation agreed other works, where the results of different approaches for VSC calculation varied with the different surface orientations [15].

Table 6. RMSD and MBD results from the comparison between the VSC values calculated with the different approaches.

VSC	RMSD (%)		MBD (%)	
	Sky-Scanner		CIE	
South surface	24.46	−8.47	25.56	−11.25
North surface	23.46	−4.48	23.53	−5.56
East Surface	31.93	−15.51	29.85	−15.38
West Surface	22.99	−1.14	24.84	−2.90

5. Experimental Characterization of the VSC in Burgos

The experimental values of the VSC calculated from Equation (6) were obtained for the city of Burgos. Figure 6 shows the VSC values, classified by intervals, and for the four vertical cardinal orientations. As can be seen from Figure 6, VSC values lower than 20% were practically nonexistent on surfaces facing the four cardinal orientations. VSC values within the interval between 20% and 40% had high FOC, ranging between 6.74% (south-facing surface) and 52.53% (north-facing surface). The following VSC interval, from 40% to 60%, presented FOC of almost 30% in all four cardinal orientations. VSC values between 60% and 80% could also be found in all cardinal orientations with FOC of 23.4% for south-facing facades and 10.91% and 10.16% for east- and north-facing facades, respectively. VSC values higher than 100% were present in all directions, with greater frequency on the south- (21.94%) and the east-facing surfaces (33.68%).

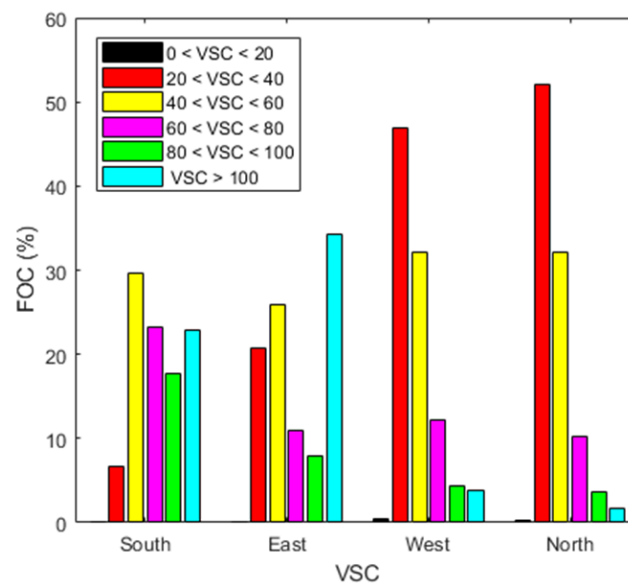


Figure 6. Distribution of VSC values by intervals calculated in Burgos, Spain, between September 2016 and January 2019.

Building design must take into account the availability of daylighting for proper building and room design factors, such as depth, floor area, glazing surface, and window heights and widths, among others. High VSC values usually result in a greater illumination in interior spaces. Different buildings standards [43–45] establish minimum levels of VSC for more effective daylighting [46]. If the obstruction angle of external obstructions is no higher than 25° above the horizon, then VSC values $>27\%$ usually indicate good daylight availability [14]. It must be noted that the VSC target value of 27% is a figure based on low-density suburban housing models. The daylight and sunlight review states that in an inner-city urban environment, VSC values in excess of 20% should be considered as reasonably good. However, whenever the VSC value falls below 10%, then the availability of direct light from the sky will be poor. As pointed out in Figure 6, high VSC values are predominant in Burgos in all cardinal orientations. Table 7 shows the frequency of occurrence of VSC values higher than 27% for surfaces facing the four cardinal orientations.

This study, therefore, shows that the availability of daylight in Burgos is very favorable for integration into the lighting design of buildings. Detailed plans for new city developments accept higher urban density, thereby increasing the difficulties associated with the required levels of daylight. The impact of site layout on daylight conditions, especially in dense urban areas at high latitudes, has been remarked upon by several authors [47]. High VSC levels mean that even dense urban areas can have good daylighting conditions.

Table 7. Frequency of occurrence (FOC) of VSC values higher than 27% for surfaces facing the four cardinal orientations.

Orientation of Surface	FOC (%)
SOUTH	99.76
EAST	96.84
WEST	87.41
NORTH	90.22

A seasonal study of Burgos VSC conditions was completed using the CIE standard classification for homogeneous skies. Knowledge of the sky type permits the determination of VSC on any surface regardless of its orientation, due to the univocal relationship between the sky type and the illuminance distribution. This approach presents the advantage that CIE standard classification can be done using different procedures that avoid the use of sky scanner devices [32]. Firstly, the seasonal distribution of sky types in Burgos was determined, as presented in Figure 7. As can be seen, the sky types corresponding to clear sky conditions (11 to 15 CIE standard sky types) presented FOCs ranging from 71.72% in summer to 34.45% in winter. Sky type 12 was the most likely CIE standard sky type, with FOCs of 18.19% in winter, 21.21% in spring, 27.84% in summer, and 17.61% in autumn. Those figures imply VSC values higher than 45%, regardless of the scattering angle, with acceptable daylighting conditions throughout the year for dense urban design.

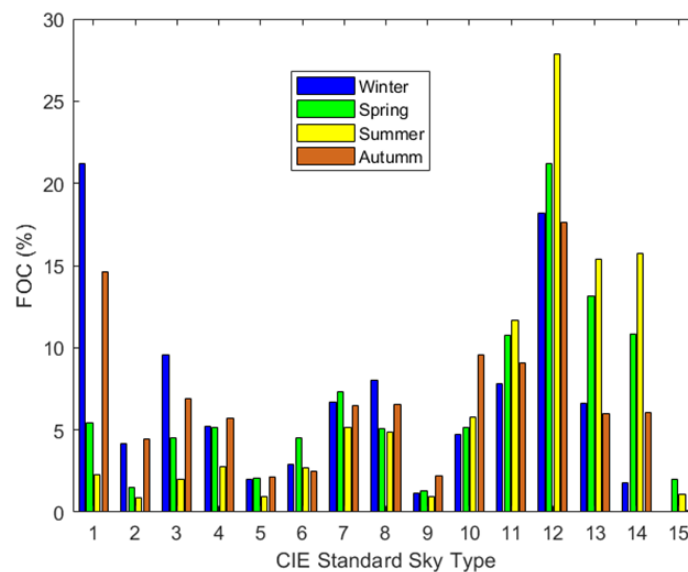


Figure 7. Seasonal FOC of the different CIE standard sky types in Burgos, Spain, between September 2016 and January 2019.

Sky types corresponding to cloudy sky conditions (CIE standard 1 to 5) are more probable in winter and autumn. Sky type 1 has the highest FOC (21.17% in winter and 14.63% in autumn) in the cloudy sky type group, while sky type 5 shows FOC lower than 2.5% throughout the year. There is, therefore, a very low probability that the VSC will be lower than 25% (minimum value of VSC corresponding to sky type 2 with a scattering angle of 120°). The FOC of partly cloudy sky conditions (CIE standard type from 6 to 10) was homogeneous throughout the year, ranging from 19.33% in summer to 27.29% in autumn. This result confirms the availability of daylighting in Burgos and agrees with the previous VSC calculations.

The monthly CIE standard sky classification is presented in Figure 8. The FOC of each clear sky type is highlighted in Figure 8, mainly from April to September, confirming that VSC values higher than 27% were obtained throughout almost all the year.

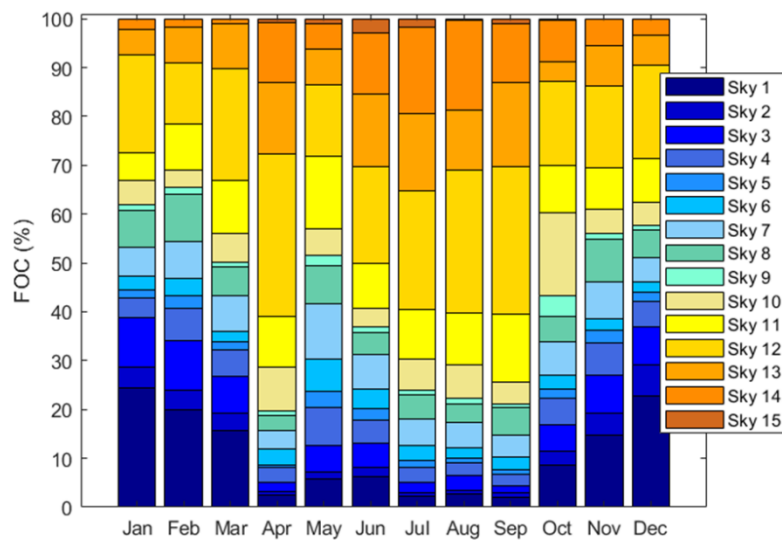


Figure 8. Monthly FOC of the CIE standard sky type in Burgos, Spain, between September 2016 and January 2019.

6. Conclusions

VSC is a useful index of daylighting availability for acceptable urban design in modern cities. There are very few studies on this parameter and its temporal distribution due, mainly, to the scarce experimental data available around the world. In this study, different methodologies for obtaining VSC were compared and an extensive campaign of measures was carried out over more than two years. Taking the VSC value calculated with luxmeter measures as a reference, the use of the sky luminance distribution from the CIE standard classification for homogeneous skies, and the one measured from a sky scanner have shown comparable RMSD values, lower than 30%. This fact shows the equivalence of both approaches for calculating VSC for surfaces facing a given orientation. The use of the CIE standard classification presents the advantage that it can be obtained from different procedures, different from the use of sky scanners, as various studies have shown.

Values of diffuse illuminance on horizontal and vertical cardinal-oriented surfaces calculated from the sky scanner device used in this work (model MS-321LR) implied an intrinsic RSMD, in comparison with the measured data from the luxmeters that ranged between 21% and 31%. The main discrepancies between the VSC values calculated from the luxmeter measures and those obtained from the sky illuminance distribution were due to the technical specifications of the sky scanner device. Therefore, the use of more accurate devices could decrease the error made by equivalent procedures to the same extent. In the study, neither climatic nor geographic variables were used, so the results are perfectly extrapolated to other locations.

The VSC in Burgos, Spain, was calculated in a measurement campaign lasting 29 months, between September 2016 and January 2019. During that period, VSC values lower than 20% in the four cardinal orientations were practically nonexistent and values between 20–40% were of greater probability. Values between 40% to 60% had FOC of almost 30% in all cardinal orientations and both the south- and the east-facing orientations presented FOC in the interval 60–100% of 20% and 10%, respectively. It means very good conditions for daylighting availability in accordance with the most common European Building Standards.

VSC values of >27%, considered by different standards as representative value for acceptable daylight levels, present FOC values ranging from 87% to 99%, so efficient city and building design could lead to significant energy savings for lighting, as well positive effects for occupant health and well-being. VSC values higher than 100% were found in all cardinal orientations, more probably on east- (FOC 30%) and south- (FOC 20%) facing vertical surfaces, showing the high energetic and luminic

potential of those orientations without underestimating the potential of the north and west facades, usually less considered for an energy-efficient design of buildings.

Author Contributions: Conceptualization and methodology, M.D.-M., M.I.D.-V., and D.G.-L.; software, D.G.-L. and M.I.D.-V.; validation, D.G.-L., and A.S.-G.; investigation, D.G.-L. and A.S.-G.; resources, M.D.-M. and C.A.-T.; data curation, D.G.-L. and A.S.-G.; writing—original draft preparation, M.D.-M. and C.A.-T.; writing—review and editing, C.A.-T.; supervision, C.A.-T., M.D.-M., and A.S.-G.; project administration, M.D.-M. and C.A.-T.; funding acquisition, M.D.-M. and C.A.-T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Regional Government of Castilla y León under the “Support Program for Recognized Research Groups of Public Universities of Castilla y León” (ORDEN EDU/667/2019) and the Spanish Ministry of Science, Innovation & Universities under the I+D+i state program “Challenges Research Projects” (Ref. RTI2018-098900-B-I00). Diego Granados López also thankfully acknowledges the economic support from the Junta de Castilla-León (ORDEN EDU/556/2019).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Aries, M.B.C.; Aarts, M.P.J.; Van Hoof, J. Daylight and Health: A Review of the Evidence and Consequences for the Built Environment. *Lighting Res. Technol.* **2015**, *47*, 6–27. [[CrossRef](#)]
2. Torrington, J.M.; Tregenza, P.R.; Noell-Waggoner, L.C. Lighting for People with Dementia. *Lighting Res. Technol.* **2017**, *39*, 81–97. [[CrossRef](#)]
3. Webb, A.R. Considerations for Lighting in the Built Environment: Non-Visual Effects of Light. *Energy Build.* **2006**, *38*, 721–727. [[CrossRef](#)]
4. Ghodrati, N.; Samari, M.; Mohd Shafiei, M.W. Green Buildings Impacts on Occupants’ Health and Productivity. *Res. J. Appl. Sci.* **2012**, *8*, 4235–4241.
5. Edwards, L.; Torcellini, P. *Literature Review of the Effects of Natural Light on Building Occupants*; National Renewable Energy Lab: Golden, CO, USA, 2002.
6. Dubois, M.C.; Gentile, N.; Amorim, C.N.D.; Osterhaus, W.; Stoffer, S.; Jakobiak, R.; Geisler-Moroder, D.; Matusiak, B.; Onarheim, F.M.; Tetri, E. Performance Evaluation of Lighting and Daylighting Retrofits: Results from Iea Shc Task 50. *Energy Procedia* **2016**, *91*, 926–937. [[CrossRef](#)]
7. Ferreira, C.; Soares, C.P.; Rocha, P. Research on energy saving potential of daylighting in tropical climates: a case study of the building IBOPE, Brazil. In Proceedings of the 12th Conference of International Building Performance Simulation Association, Sydney, Australia, 14–16 November 2011.
8. Li, D.H.W. A Review of Daylight Illuminance Determinations and Energy Implications. *Appl. Energy.* **2010**, *7*, 2109–2118. [[CrossRef](#)]
9. Tregenza, P.; Mardaljevic, J. Daylighting Buildings: Standards and the Needs of the Designer. *Lighting Res. Technol.* **2018**, *50*, 63–79. [[CrossRef](#)]
10. Darula, S.; Christoffersen, J.; Malikova, M. Sunlight and Insolation of Building Interiors. *Energy Procedia* **2015**, *78*, 1245–1250. [[CrossRef](#)]
11. Kómar, L.; Kocifaj, M. Uncertainty of Daylight Illuminance on Vertical Building Façades When Determined from Sky Scanner Data: A Numerical Study. *J. Sol. Energy* **2014**, *110*, 15–21. [[CrossRef](#)]
12. Chen, W.; Li, D.H.W.; Lou, S. Estimation of Irregular Obstructed Vertical Sky Components under Various Cie Skies. *Energy Procedia* **2019**, *158*, 309–314. [[CrossRef](#)]
13. Li, D.H.W.; Lou, S.; Lam, J.C.; Wu, R.H.T. Wu. Determining Solar Irradiance on Inclined Planes from Classified Cie (International Commission on Illumination) Standard Skies. *Energy* **2016**, *101*, 462–470. [[CrossRef](#)]
14. Littlefair, P. *Site Layout Planning for Daylight and Sunlight: A Guide to Good Practice*; Building Research Establishment: Watfold, UK, 2011.
15. Ng, E.; Cheng, V.; Gadi, A.; Mu, J.; Lee, M.; Gadi, A. Defining Standard Skies for Hong Kong. *Build. Environ.* **2007**, *42*, 866–876. [[CrossRef](#)]
16. Alshaibani, K. The Use of Sky Luminance and Illuminance to Classify the Cie Standard General Skies. *Lighting Res. Technol.* **2015**, *47*, 243–247. [[CrossRef](#)]
17. Li, D.H.W.; Cheung, K.L.; Tang, H.L.; Cheng, C.C.K. Identifying Cie Standard Skies Using Vertical Sky Component. *J. Atmos. Sol.-Terr. Phys.* **2011**, *73*, 1861–1867. [[CrossRef](#)]

18. Li, D.; Lam, T.; Wu, T. Estimation of Average Daylight Factor under Obstructed CIE Standard General Skies. *Lighting Res. Technol.* **2014**, *46*, 187–197. [[CrossRef](#)]
19. Li, D.H.W.; Lou, S.; Ghaffarianhoseini, A.; Alshaibani, K.; Lam, J.C. A Review of Calculating Procedures on Daylight Factor Based Metrics under Various CIE Standard Skies and Obstructed Environments. *Build. Environ.* **2017**, *112*, 29–44. [[CrossRef](#)]
20. Darula, S.; Kittler, R.; Kómar, L. Sky Type Determination Using Vertical Illuminance. *Przełąd Elektrotechniczny*. **2013**, *89*, 315–319.
21. Kittler, R.; Darula, S. The Method of Aperture Meridians: A Simple Calculation Tool for Applying the ISO/CIE Standard General Sky. *Lighting Res. Technol.* **2006**, *38*, 109–119. [[CrossRef](#)]
22. Uetani, Y.; Aydinli, S.; Joukoff, A.; Kendrick, J.D.; Kittler, R.; Koga, Y. *BS ISO 15469:2004. Spatial Distribution of Daylight-CIE Standard General Sky*; CIE: Vienna, Austria, 2003.
23. ISO. *ISO-15469:2004 (E). Spatial Distribution of Daylight-CIE Standard General Sky*; CIE: Geneva, Switzerland, 2004.
24. Markou, M.T.; Kambezidis, H.D.; Bartzokas, A.; Katsoulis, B.D.; Muneer, T. Sky Type Classification in Central England During Winter. *Energy* **2005**, *30*, 1667–1674. [[CrossRef](#)]
25. Markou, M.T.; Kambezidis, H.D.; Katsoulis, B.D.; Muneer, T.; Bartzokas, A. Sky Type Classification in South England During the Winter Period. *Build Res. J.* **2004**, *52*, 19–30.
26. Suárez-García, A.; Granados-López, D.; González-Peña, D.; Díez-Mediavilla, M.; Alonso-Tristán, C. Seasonal Characterization of CIE Standard Sky Types above Burgos, Northwestern Spain. *J. Sol. Energy* **2018**, *169*, 24–33. [[CrossRef](#)]
27. Tregenza, P.R. Standard Skies for Maritime Climates. *Light. Res. Technol.* **1999**, *31*, 97–106. [[CrossRef](#)]
28. Torres, J.L.; de Blas, M.; García, A.; Gracia, A.; de Francisco, A. Sky Luminance Distribution in the North of Iberian Peninsula During Winter. *J. Atmos. Sol.-Terr. Phy.* **2010**, *72*, 1147–1154. [[CrossRef](#)]
29. Torres, J.L.; de Blas, M.; García, A.; Gracia, A.; de Francisco, A. Sky Luminance Distribution in Pamplona (Spain) During the Summer Period. *J. Atmos. Sol.-Terr. Phy.* **2010**, *72*, 382–388. [[CrossRef](#)]
30. Chaiwivatworakul, P.; Chirarattananon, S. Distribution of Sky Luminance in Tropical Climate. In Proceedings of the Joint International Conference on Sustainable Energy and Environment, Hua Hin, Thailand, 1–3 December 2004; pp. 530–537.
31. Li, D.H.W.; Tang, H.L. Standard Skies Classification in Hong Kong. *J. Atmos. Sol.-Terr. Phy.* **2008**, *70*, 1222–1230. [[CrossRef](#)]
32. Li, D.H.W.; Chau, C.; Wan, K.K.W. A Review of the CIE General Sky Classification Approaches. *Renew. Sust. Energy Rev.* **2014**, *31*, 563–574. [[CrossRef](#)]
33. Umemiya, N.; Kanou, T. Classification of Sky Conditions by the Ranges of Insolation Indices Considering CIE Standard for General Sky. *J. Light Vis. Environ.* **2008**, *32*, 14–19. [[CrossRef](#)]
34. Wong, S.L.; Wan, K.K.W.; Li, D.H.W.; Lam, J.C. Generation of Typical Weather Years with Identified Standard Skies for Hong Kong. *Build. Environ.* **2012**, *56*, 321–328. [[CrossRef](#)]
35. Lou, S.; Li, D.H.W.; Lam, J.C. CIE Standard Sky Classification by Accessible Climatic Indices. *Renew. Energy* **2017**, *113*, 347–356. [[CrossRef](#)]
36. Li, D.H.W.; Lau, C.C.S.; Lam, J.C. Standard Skies Classification Using Common Climatic Parameters. *J. Sol. Energy Eng.* **2004**, *126*, 957–964. [[CrossRef](#)]
37. Alshaibani, K. The Use of Horizontal Sky Illuminance to Classify the CIE Standard General Skies. *Lighting Res. Technol.* **2016**, *48*, 1034–1041. [[CrossRef](#)]
38. Alshaibani, K. Classification Standard Skies: The Use of Horizontal Sky Illuminance. *Renew. Sust. Energy Rev.* **2017**, *73*, 387–392. [[CrossRef](#)]
39. Janjai, S.; Tohsing, K.; Nunez, M.; Laksanaboonsong, J. A technique for mapping global illuminance from satellite data. *J. Sol. Energy* **2008**, *82*, 543–555. [[CrossRef](#)]
40. de-Simón-Martín, M. Characterisation of Solar Diffuse Irradiance on Vertical Surfaces. Ph.D. Thesis, University of Burgos, Burgos, Spain, 2015.
41. Alshaibani, K. Finding Frequency Distributions of CIE Standard General Skies from Sky Illuminance or Irradiance. *Light. Res. Technol.* **2011**, *43*, 487–495. [[CrossRef](#)]
42. Li, D.H.W.; Li, C.; Lou, S.W.; Tsang, E.K.W.; Lam, J.C. Analysis of vertical sky components under various CIE standard general skies. *Indoor Built. Environ.* **2016**, *25*, 703–711. [[CrossRef](#)]
43. CIBSE. *Desktop Guide to Daylighting—Good Practice Guide 245*; CIBSE: London, UK, 1998.

44. Mansfield, K.P. *British Standard BS 8206-2 (2008) Lighting for Buildings-Part 2: Code of Practice for Daylighting*; British Standards Institution: London, UK, 2008.
45. Mayor of London. Daylight, Sunlight and Overshadowing. *Environ. Statement Main Rep.* **2012**, *2*, 1–16.
46. Olina, A.; Zaimi, N. Daylight Prediction Based on the Vsc-Df Relation. Master Thesis in Energy-efficient and Environmental; Buildings Faculty of Engineering, Lund University, Lund, Sweden, 2018.
47. Littlefair, P. Passive Solar Urban Design: Ensuring the Penetration of Solar Energy into the City. *Renew. Sustain. Energy Rev.* **1998**, *2*, 303–326. [[CrossRef](#)]

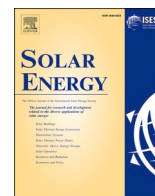


© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Paper III

Granados-López, D., Suárez-García, A., Díez-Mediavilla, M. and Alonso-Tristán, C., 2021. *Feature selection for CIE standard sky classification*. *Sol. Energy* 218, 95-107. DOI: 10.1016/j.solener.2019.11.060

Abstract: There are several compilations of sky classifications that refer to Meteorological Indices (MIs) (variables usually recorded at meteorological ground stations), due to the scarcity of sky scanner devices that can supply the experimental data needed to apply the CIE standard sky classification. The use of one rather than another MI is never justified, because there is no standardized criterion for their selection. In this study, forty-three MIs, traditionally used to define different sky conditions, are reviewed. Feature Selection (FS) is a key step in the design of a sky-classification algorithm using MIs as an alternative to data from sky scanners. Four procedural methods for FS -Pearson, Permutation Importance, Recursive Feature Elimination, and Boruta- are applied to an extensive data set of MIs that includes CIE standard sky classification data, which was used as a reference. The use of FS procedures significantly reduced the original set of MIs, permitting the construction of different classification trees with high performance for the sky classification. In the case of the Pearson FS method, the classification tree only used two MIs. The advantage of the Pearson FS method is that it functions independently from the machine-learning algorithm used latter for the sky classification.



Feature selection for CIE standard sky classification

Diego Granados-López^a, Andrés Suárez-García^{a,b}, Montserrat Díez-Mediavilla^a,
Cristina Alonso-Tristán^{a,*}

^a Research Group Solar and Wind Feasibility Technologies (SWIFT), Electromechanical Engineering Department, Universidad de Burgos, 09006 Burgos, Spain

^b Centro Universitario de la Defensa. Escuela Naval Militar de Marín, 36920 Marín, Pontevedra, Spain

ARTICLE INFO

Keywords:

CIE standard sky classification
Feature selection
Meteorological indices
Machine learning

ABSTRACT

There are several compilations of sky classifications that refer to Meteorological Indices (MIs) (variables usually recorded at meteorological ground stations), due to the scarcity of sky scanner devices that can supply the experimental data needed to apply the CIE standard sky classification. The use of one rather than another MI is never justified, because there is no standardized criterion for their selection. In this study, forty-three MIs, traditionally used to define different sky conditions, are reviewed. Feature Selection (FS) is a key step in the design of a sky-classification algorithm using MIs as an alternative to data from sky scanners. Four procedural methods for FS -Pearson, Permutation Importance, Recursive Feature Elimination, and Boruta- are applied to an extensive data set of MIs that includes CIE standard sky classification data, which was used as a reference. The use of FS procedures significantly reduced the original set of MIs, permitting the construction of different classification trees with high performance for the sky classification. In the case of the Pearson FS method, the classification tree only used two MIs. The advantage of the Pearson FS method is that it functions independently from the machine-learning algorithm used latter for the sky classification.

1. Introduction

Daylight, as part of the solar irradiance is an essential natural resource even for human health. In building design, projection of daylight can increase the energy efficiency of buildings (Dubois et al., 2016; Fouquart et al., 1990; Li, 2010) and will have positive effects on the well-being of occupants (Aries et al., 2015; Edwards and Torcellini, 2002). Natural lighting availability is highly dependent on luminance levels and sky conditions. In 2003, the Commission International de L'Eclairage (CIE) (Uetani et al., 2003) and the International Organization for Standardization (ISO) (ISO, 2004) both adopted 15 Standard Sky Luminance Distributions (SSLD), five clear, five overcast and five partly cloudy sky conditions. They provide the most versatile definition of skylight at various localities and daylight climate regions, making it possible to simulate an annual daylight profile at any point on earth in absolute units, based on typical luminance sky patterns.

The sky scanner is the standard instrument for measuring sky luminance distribution (Li, 2010). Despite the high interest in those measurements, very few studies at only a handful of European (Markou et al., 2005; Markou et al., 2004; Suárez-García et al., 2018; Torres et al., 2010a, b; Tregenza, 1999) and Asian (Chaiwiwatworakul and

Chirarattananon, 2004; Li and Tang, 2008; Ng et al., 2007; Zi et al., 2020) locations have been conducted to characterize the sky under the CIE standard, mainly due to the scarcity of sky scanner devices available to gather sky luminance data.

Different alternatives to the use of sky scanners have been proposed for classifying the skies (Li et al., 2014b), including the use of different climatic parameters or meteorological indices (Li et al., 2004; Lou et al., 2017; Umemiya and Kanou, 2008; Wong et al., 2012), vertical (Chen et al., 2019; Darula et al., 2013) and horizontal illuminance data (Alshaibani, 2016b; Alshaibani, 2017), as well as satellite data (Janjai et al., 2008). Added difficulties for sky classification (Allard et al., 2015) include the variability of sky conditions and their sensitivity to many stochastic variables.

Sky classification in various studies relies on Meteorological Indices (MIs), factors usually recorded at meteorological ground stations that, to a greater or lesser extent, affect the luminance and radiance distributions: sun position, cloud coverage, turbidity, and weather conditions, among others. Such climatic indices within certain ranges will lead to sky luminance and radiance distributions with similar features, and a straightforward approach is to describe those distributions by sky conditions (Lou et al., 2017). The selection of the MIs depends on their

* Corresponding author.

E-mail addresses: catristan@ubu.es, cristinaalonso.tristan@gmail.com (C. Alonso-Tristán).

<https://doi.org/10.1016/j.solener.2021.02.039>

Received 15 September 2020; Received in revised form 24 November 2020; Accepted 15 February 2021

Available online 8 March 2021

0038-092X/© 2021 International Solar Energy Society. Published by Elsevier Ltd. All rights reserved.

availability. The number of MIs used and the conditions that define each sky type vary in each study, complicating the task of defining a taxonomy that could unequivocally describe the specific characteristics of each sky type (Dieste-Velasco et al., 2019; Perez et al., 1990a), even in a reduced classification with only three sky types: clear, partly cloudy, and overcast conditions.

In recent years, machine learning (ML) tools, such as Artificial Neural Networks (ANN's) (Li et al., 2010) and classification trees have, among others, been applied to sky classification. Supervised machine learning tools permit the identification of patterns and relationships between inputs and outputs, as long as the algorithm has sufficient examples to train recognition. In this paper, a set of sky type samples labeled as CIE Standard Sky Classification will be used as the training set for sky classification purposes and a test set of available MIs will be used as input for the algorithm.

The work flow of a supervised Machine-Learning (ML) tool is shown in Fig. 1. The first step for every ML tool is to filter and to analyze the input data so as to categorize it and to control its quality. The second step is the Feature Selection (FS) procedure: the identification of related features within a set of data and the removal of irrelevant or less important features that contribute little or nothing to the definition of the target variable, so as to achieve models of greater accuracy. FS is one of the core concepts of ML that will impact on the performance of the developed model, improving its precision and reducing its complexity and overfitting as well as its runtime.

Following acceptable FS, the algorithm is trained using part of the input data set (training set), typically 80% of the total, using the remaining 20% for validation tests. Training set data and test set data are exchanged as many times as necessary, until the algorithm is considered validated.

In this study, a total of forty-three MIs describing sky conditions are borrowed from past studies for their use as variables to define sky types. The use of one rather than another MI is not justified, because there is no standardized criterion for selecting MIs. It is necessary to compare the information that each of them offers, removing those that offer redundant or insignificant information for the definition of sky types (Yang and Pedersen, 1997). Therefore, FS is a key step in the design of a sky classification algorithm using MIs as an alternative to data from sky scanners. The main objective of this study is to determine, through a FS procedure, the most suitable MIs and their precise number for the optimization of the sky classification algorithms. Forty-three MIs were included in the study, calculated from half-hourly experimental data records collected at Burgos, Spain, between September 2016 and December 2019. The following FS criteria were selected: Pearson (Biesiada and Duch, 2007), Permutation Importance (Gregorutti et al., 2017), Recursive Feature Elimination, and Boruta (Degenhardt et al., 2019).

This study reports an extensive review of the MIs that define different sky conditions and features that are suitable for sky classification.

Structured and rigorous FS procedures can determine the usefulness of the information in these indices, with a high degree of success, for the problem of sky classification, the informative equivalence between some of the MIs, and the number of MIs that may be needed for sky classification in line with the CIE standard. It was proven that the Pearson FS procedure performed accurate sky classification into three sky conditions (clear, partly cloudy and overcast conditions), in accordance with the CIE Standard Classification, requiring only two MIs. The FS results, processed in a classification tree to test their validity, confirmed that the intervals of definition of the MIs for each sky type were close to the intervals that were established in each study for the individual use of each MI.

The structure of this paper will be as follows. Following the Introduction in Section 1, the methodology will be explained in Section 2, where the experimental facility and the data processing needed to calculate the MIs and the experimental campaign is introduced in Section 2.1. In Section 2.2, the CIE Standard Sky classification of Burgos, Spain, gathered during the experimental campaign will be described, as reference data for sky classification. The MIs with their data on sky conditions that were available for the FS procedures will be reviewed in Section 2.3. Then, the different FS procedures used in this work and the results of their application to the experimental MIs will be described in Section 2.4. In Section 2.5, the classification trees will be introduced, together with the machine learning algorithm used to test the performance of the FS procedure; and in Section 2.6 the metrics used to test FS performance will be presented. Finally, the main results and the conclusions of the study will be summarized in Sections 3 and 4.

2. Methodology

The present work was developed in four steps: data collection, Feature Selection (FS), classification trees, and classification metrics. Several meteorological variables were collected between 21 September 2016 and 31 January 2020. The dataset contained over eight-thousand samples that were used for the evaluation of 43 MIs. The size of the data set lent support for the conclusions of this work. Following the calculation of the MIs, the classification tree was employed in conjunction with the FS procedure to classify sky cloudiness (clear, partial or overcast) following the established CIE patterns. The ML classification tree algorithm was selected, because it can process and extract the rules for sky labelling. FS, for maximum simplification of the classification trees, was applied, in an effort to reduce the number of MIs serving as ML algorithm inputs to a minimum. Finally, the outputs of the classification tree algorithm were analyzed using several metrics.

2.1. The experimental facility

The experimental campaign during which the meteorological data were recorded for the processing of each MI in this study was performed

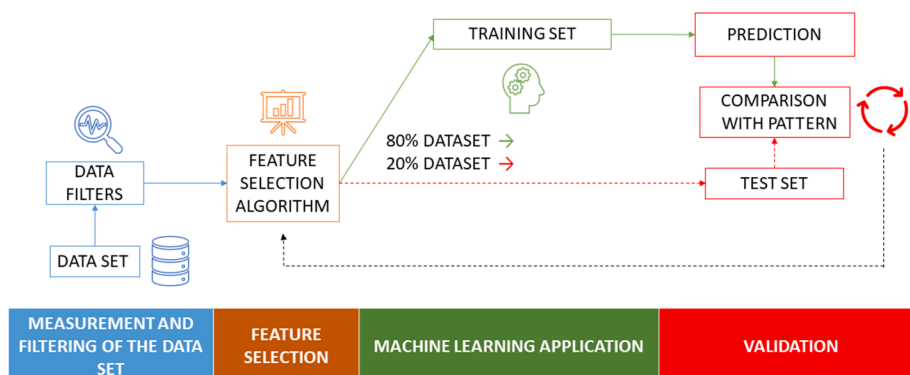


Fig. 1. Workflow of Supervised Machine-Learning tool.

in Burgos, Spain. Data collection took place at a meteorological facility located at the Higher Polytechnic School of Burgos University (LON. 42°21'04"N, LAT. 3°41'20"W, 856 m above mean sea level). A commercial sky scanner from Eko instruments, model MS-321LR, was used for CIE Standard classification. Its technical specifications are shown in Table 1. Measurements from 4-minute scans were taken every 15 min, from September 2016 to December 2017. From January 2018, the scans were taken every 10 min. The device was adjusted on a monthly basis to measure from sunrise to sunset. First and last daily records were discarded, to avoid measurements with solar altitudes equal to or lower than 7.5°. Data higher than 50 kcd/m^2 and lower than 0.1 kcd/m^2 were also discarded, following the technical specifications of the sky scanner. Seven lux sensors, EKO, model ML-020S-O, technical specification listed in Table 2, were also used: four of them recorded vertical global illuminance in the four cardinal orientations and three lux sensors recorded horizontal, global, beam, and diffuse illuminance. Horizontal global, diffuse and beam irradiance were measured using Hukseflux pyranometers, model SR11 and a Hukseflux pyrhemliometer, model DR01. The technical specifications of the pyranometers and the pyrhemliometer are shown in Table 3. The beam illuminance and irradiance sensors were installed on a sun tracker, model Sun-Tracker 3000, from Geónica. The diffuse illuminance and irradiance sensors were obscured from direct sunlight by a shadow hat. Illuminance and irradiance data were recorded every 10 min (averaging recorded scans of 30 s). Fig. 2 shows the experimental equipment.

CIE quality criteria (Comission Internationale de, L.E, 1995) were used for analyze and filter illuminance data while irradiance data were analyzed and then filtered using conventional quality criteria (Gueymard and Ruiz-Arias, 2016). To match simultaneous records of illuminance and irradiance data, half-hourly and hourly sky scanner measurements were used in this study, from September 2016 to December 2017, and from January 2018, ten minutes records. If the illuminance and irradiance data failed to pass the quality criteria, then all the simultaneous data sets were rejected.

The measurement campaign extended between 21 September 2016, and 31, January 2020. Following their analysis and the filtering process, the experimental data amounted to 8829 items.

2.2. CIE standard classification of Burgos skies

Supervised Machine Learning needs examples for training the classifier algorithm. In this work, the CIE standard classification served as a benchmark for estimating the performance of the supervised machine learning algorithm and for testing the FS procedure. Several works have reported that the CIE Standard sky classification provides a good overall framework for representing the actual conditions for homogeneous skies (Li et al., 2011b; Li et al., 2004; Li et al., 2010; Markou et al., 2005). Tregenza (Tregenza, 2004) gave a detailed description of the CIE standard classification procedure following a discrete integration methodology, the same method that was used for sky classification in Burgos. The labelling of CIE sky types was as follows: I.1 to III.1: cloudy; III.2 to IV.3: partially cloudy; and, IV.4 to VI.6: clear skies. More information on the classification method can be obtained from a previous work (Suárez-García et al., 2018). In the experimental campaign between 21 September 2016, and 31 January 2020, clear skies predominated in

Table 1
Sky scanner specifications.

Model	MS-321LR Sky Scanner
Dimensions (W × D × H)	430 mm × 380 mm × 440 mm
Mass	12.5 kg
FOV	11°
Luminance	0 to 50 kcd/m^2
Radiance	0 to 300 W/m^2
A/D Convertor	16 bits
Calibration Error	2%

Table 2
Luxmeter technical specifications.

Model	ML-020S-O
Illuminance Range	0 to 150,000 lx
Output	0 to 30,000 μV
Impedance	280 Ω
Operating temperature range	-10 °C to 50 °C
Temperature response	0.4%

Table 3
Pyranometers and Pyrhemliometer technical specifications.

Model	SR11	DR01
Measurement Range	0–3000 W/m^2	0–4000 W/m^2
Calibration uncertainty	<1.8% (k = 2)	<1.2% (k = 2)
Spectral Range	285–3000 $\times 10^{-9} m$	200–4000 $\times 10^{-9} m$
Sensitivity (nominal)	$15 \times 10^{-6} V/(W/m^2)$	$10 \times 10^{-6} V/(W/m^2)$
Operating temperature range	-40 °C to 80 °C	-40 °C to 80 °C
Temperature response	< $\pm 2%$ (-10 °C to 40 °C)	< $\pm 1%$ (-10 to +40 °C)

Burgos (52%) while overcast skies were present in 15% and partially cloudy skies in 33% of cases, as shown in Fig. 3.

2.3. Meteorological indices

Skies of the same category are assumed to share identical well-defined sky luminance patterns (Darula and Kittler, 2002), which is the straightforward approach for sky classification. Once the skies have been identified, the daylight on any surface can be estimated, by integrating the luminance distribution of the sky dome over each surface (Granados-López et al., 2020). Therefore, any climatic parameter based on lighting measurements can potentially identify a given sky condition. Table 4 describes the 43 MIs reviewed in this work.

The US National Bureau of Standards (NBS) recommends the use of the horizontal diffuse fraction, k_d , for sky classification (Fakra et al., 2011): low k_d values indicate clear sky conditions and high values are usually present, but not exclusively so, in overcast conditions (Li et al., 2015). Alternatively, high values of the horizontal direct fraction, k_b , are representative of clear skies, due to the high values of the solar irradiation beam component (Ferraro et al., 2010) while low k_b values predominate on cloudy days.

Perez's Sky clearness, ϵ_p , maybe one of the most widely used MIs for sky characterization, was originally proposed to define the ratio of illuminance and irradiance, known as luminous efficacy, K . The sky's brightness index, Δ , is often used with the clearness index, ϵ_p , for sky classification (Li et al., 2004; Perez et al., 1990a).

Luminous efficacy, K , can be modeled through different parameters such as the solar zenith angle, Z_s ; Perez's sky clearness index, ϵ_p ; the sky's brightness index, Δ , and, the atmospheric precipitable water content (Perez et al., 1990a).

Relative heaviness, Ω , (Chung, 1992) is proportional to the amount of solar radiation entering into clouds. Cloud cover, CC , is often used as an indicator of sky conditions (Muneer et al., 2007): 0 oktas is the CC value for clear skies and 8 oktas is assigned in overcast conditions. A sky classification based on MIs was performed by Igawa et al. (Igawa et al., 2004) using Igawa's sky index, S_i , the clear sky index, k_c , and the cloudless index, Cle .

A CIE-based standard classification of skies using global horizontal illuminance, $LxGH$, and Kittler's index, k_t , was proposed by Lou et al. (2019). Kittler's index, k_t , is widely used for illumination studies, due to the high information content that it provides when only global irradiation data are available. However, k_t is only available when the zenith sun angle is under 80°, $Z_s < 80^\circ$. An alternative and globally valid definition was proposed by Perez et al. (1990b), k_{t2} , that used cloud cover, CC , together with relative humidity, RH , among other factors,

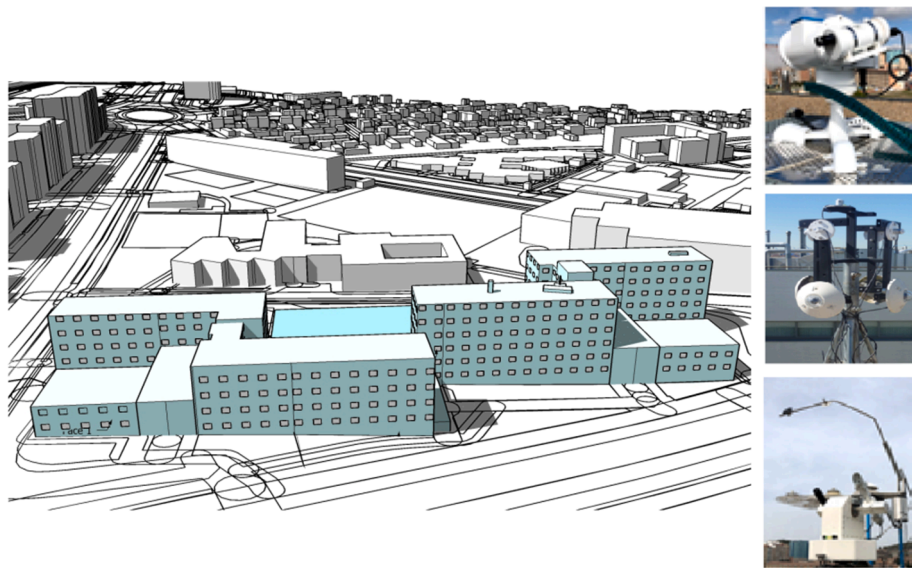


Fig. 2. Experimental equipment on the roof of the Higher Polytechnic School of Burgos University, Spain.

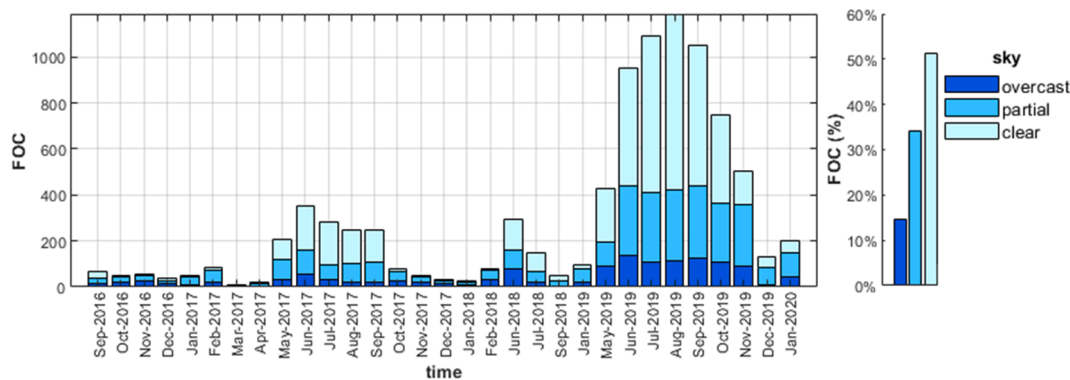


Fig. 3. Monthly distribution of the Frequency of Occurrence (FOC) and total FOC (%) of clear, partly cloudy, and overcast sky conditions in Burgos, Spain (from September 21st, 2016 to January 31st, 2020).

effectively contributing to better definition of the atmospheric conditions.

The cloud ratio on irradiance, C_e (Rahim et al., 2004) originally defined as the proportion of diffuse to global irradiance, was used in the estimation of solar radiation. Umemiya and Kanou (2008) introduced a new definition in terms of illumination, C_v , and used it for sky tree classification. The cloud ratio is 1 in overcast sky conditions and 0 for clear skies, and it will vary quickly and with some frequency when the sky is partly cloudy. The cloudless index, Cle , is often defined in terms of the standard cloud ratio, C_{es} , and the cloud ratio, C_e . C_{es} is defined as a polynomial fit of the lower limit of C_e .

Perraudeau’s nebulosity index, OPF , introduced by Perraudeau in 1989 (Kambezis et al., 1998), classifies the skies into five categories (Kambezis, 2018). This index has since been modified by other authors (Fakra et al., 2011) and is defined in this work as FP . The clearness function, F , was compared to the MIs Δ , ϵ_p , and k_t for sky classification (Muneer, 2007). Low values of F , indicate overcast sky conditions and values near to 1 are obtained under clear sky conditions.

The Klucher index, FK (Klucher, 1979) depending only on k_d , has also been used for sky classification. Markou et al. (2005) prepared a simple sky classification by modelling direct solar irradiance data, P_e , characteristic of each sky type. This proposal suggested the use of experimental MIs for sky classification: horizontal global irradiance, $RaGH$; horizontal diffuse irradiance, $RaDH$; horizontal beam irradiance,

$RaBH$; and south-facing global vertical irradiance, $RaGVS$.

Umemiya and Kanou (2008) proposed the turbidity index, $TURV$, permeability, $PERM$, Unemiyas’s Cloud Ratio, $CLDV$, and global, diffuse, and beam illuminance, $EVGM$, $EVDM$, and $EVSM$, normalized to the optical mass, M_v , as effective sky condition sorters. They produced a sky classification with 7 types of skies that used a classification tree based on the turbidity index, $TURV$; Kittler’s index, k_t ; sky brightness, Δ ; and, normalized global illuminance, $EVGM$. A similar proposal was introduced by Lou et al. (2017) using solar altitude, α_s ; Kittler’s index, k_t ; the turbidity index, $TURV$; air temperature, T ; and, relative humidity, RH . Other variables, used for meteorological forecasting have been proposed among which MIs for sky classification such as wind speed, WS ; relative humidity, RH ; cloud cover, CC ; and air temperature, T , among others (Inman et al., 2013).

Li et al. [12] proposed a group of MIs that obtained a very accurate sky classification. They used a ratio of zenith illuminance, L_z , and horizontal diffuse illuminance, L_{xDH} , named $LERT$, as a measure of sky brightness (Li et al., 2006; Markou et al., 2005). The luminous turbidity index, t_v , refers to the attenuation of solar radiation in the atmosphere, due to the molecules contained into the air (water, dust or aerosols) (Li et al., 2016; Pasero and Mesi, 2010). In overcast sky conditions, t_v , is very high, because there is no direct solar-irradiation component. Under clear or partly cloudy sky conditions, t_v is a very interesting parameter, due to its high sensitivity to ambient pollution (Lou et al., 2017). It is related to CIE standard sky types VI.6, VI.5, and IV.4 (Kocifaj, 2011).

Table 4

Definition of the 43 MIs reviewed as candidates for sky classification. L_0 is the Luminous solar constant (133.8 kLux) and I_{SC} is the standard global irradiance ($1361.1 \frac{W}{m^2}$) (Gueymard, 2018).

Ratio Zenith Illuminance to horizontal diffuse Illuminance	Ratio Global Illuminance	Ratio Diffuse Illuminance	Luminous Turbidity index	Vertical Sky Component
$LERT = \frac{L_z}{L_{xDH}}$	$C1 = \frac{L_{xGH}}{L_{Oh}}$	$C2 = \frac{L_{xDH}}{L_{Oh}}$	$t_v = \frac{Ln(\frac{L_{Oh}}{L_{xBH}})}{A_v M_v}$	$VSC = \frac{RaDH}{RaDV}$
Normalized Global Illuminance	Normalized Beam Illuminance	Normalized Diffuse Illuminance	Cloudless Index	Igawa's Sky Index
$EVGM = M_v \frac{L_{xGH}}{L_0}$	$EVSM = M_v \frac{L_{xBH}}{L_0}$	$EVDM = M_v \frac{L_{xDH}}{L_0}$	$Cle = \frac{1 - k_d}{1 - Ces(M)}$	$S_i = \frac{RaGH}{0.84 \frac{I_{SC}}{M_v} e^{-0.0675M_v}} + \sqrt{Cle}$
Direct Fraction	Cloud Cover	Illuminance Cloud Ratio	Irradiance Cloud Ratio	Standard Cloud Ratio
$k_b = \frac{RaBH}{RaGH}$	$CC(\%Clouds)$	$C_v = \frac{L_{xDH}}{L_{xDH} + L_{xBH}}$	$C_e = \frac{RaDH}{RaDH + RaBH}$	$Ces = 0.01299 + 0.07698M_v - 0.003857M_v^2 + 0.0001054M_v^3 - 0.000001031M_v^4$
Umeyiya's Cloud Ratio	Relative Heaviness	Clear Sky Index	Clearness Index	Zenith Angle Independent Clearness Index
$CLDV = \frac{L_{xDH}}{L_{xGH}}$	$\Omega = \frac{L_{xGH}}{Sin\alpha_s}$	$k_c = \frac{L_{xGH}}{0.84 \frac{I_{SC}}{M_v} e^{-0.0675M_v}}$	$k_t = \frac{RaGH}{I_0 Sin\alpha_s}$	$k_{r2} = \frac{k_r}{1.4}$
Luminous Efficacy	Brightness Index	Perez's Clear sky index	Original Perraudau's Nebulosity Index	Perraudau's Nebulosity Index
$K = \frac{L_{xGH}}{RaGH}$	$\Delta = \frac{RaDH M_v}{I_{sc} \epsilon_0 \sin\alpha_s}$	$\epsilon_p = \frac{(RaDH + RaBH) + 1.04Z^3}{1 + 1.04Z^3}$	$OFFP = \frac{1 - k_D}{1 - \frac{E_{clear}}{E_{clear}} + RaGH}$	$FP = \frac{1 - k_D}{1 - 0.12037(SinZ_s)^{-0.82}}$
Klucher's Clearness Index	RaBH, RaDH, RaGHRaGVS, LxGH	Optical Mass	Scattering Angle Ref	Turbidity
$FK = 1 - k_D^2$	Direct, Diffuse, Global (Horizontal and Vertical South oriented) Irradiance global horizontal Illuminance	$M_v = (\sin\alpha_s + 0.50572(\alpha_s + 6.07995))^{-1.6364}$	$\chi = \arccos(\cos Z_s \cos Z_p + \sin Z_s \sin Z_p \cos \phi_p - \phi_s)$	$TURV = \frac{1 + 0.0045M_v}{0.1M_v} Ln(\frac{L_0}{L_{xBH}})$
T, RH, WS, L_z, α_s	Diffuse fraction	Clearness Function	Modeled direct solar irradiance	Permeability
Temperature, Relative humidity, Wind speed, Zenith luminance, solar altitude.	$k_d = \frac{RaDH}{RaGH}$	$F = \frac{RaGH - RaBH}{I_{sc} \epsilon_0 \sin\alpha_s}$	$P_e = \frac{RaGH - RaDH}{\sin(\alpha_s)}$	$PERV = M_v \sqrt{\frac{L_{xBH}}{L_0}}$

C1, defined as the ratio of horizontal global illuminance, L_{xGH} , and horizontal extraterrestrial illuminance, L_{Oh} , evaluates the ambient clarity. Low values of C1 are characteristic of the passage of a cloud on a clear day while a high C1 value can reflect a cloud opening zone on a completely overcast day (Alshabani, 2016a; Kittler and Danda, 2000). C2 is defined as the ratio of horizontal diffuse illuminance, L_{xDH} , and horizontal extraterrestrial illuminance, L_{Oh} , so high C2 values are characteristic of partly cloudy skies, while low C2 values are characteristic of cloudy or completely clear skies (Li et al., 2006; Li et al., 2010; Markou et al., 2005).

The vertical sky component, VSC, was also proposed as an MI for sky classification (Li et al., 2011b). Defined as the ratio of the vertical diffuse illuminance and horizontal diffuse illuminance, it can easily be obtained experimentally. Littlefair established an international standard for the indoor daylight evaluation of buildings (Littlefair, 2012) based on VSC, which is highly dependent on the solar altitude, α_s , and the scattering angle, χ (Alshabani, 2011; Li et al., 2014a).

2.4. Feature selection

43 MIs were selected (Table 1) for the study. Each one represents certain characteristics of the sky that are suitable for sky classification. The final objective of the present work is to distinguish the most representative MIs for sky classification according to the CIE taxonomy.

The most simple and demanding methodology is the full combinatorial method. It proposes to test all the possible combinations of all MIs: at first, only one MI would be considered for the CIE classification; then, all combinations of two MIs would be used for the task and so on (Visa et al., 2011). Li et al. (2011a) followed this path to evaluate the performance of several MIs in neural networks for weather data classification. It was feasible because only five MIs were considered which

meant a total of 30 sets of MIs for testing. In the present study, an analysis of the 43 variables implied over a trillion combinations, which was not feasible. FS was therefore essential to solve this task.

Several FS techniques are used widely in the ML field to find the most important variables or to reject the most redundant ones. There are different types of FS algorithms. On the one hand, no clustering algorithms are used with the Filter Methods that base their decision on a statistical index that evaluates the dependence between the MIs. On the other hand, the Wrapper Methods evaluate the information provided by each MI using clustering algorithms, which implies a higher computational cost (Solorio-Fernández et al., 2019).

The FS Filter methods are used to study the similarity of MIs through a statistical parameter, a mathematical expression that serves to eliminate redundant or non-informative indices. These methods are independent from the ML algorithm used later on (Mitra et al., 2002; Yu and Liu, 2003). Hence, their results may be used as an input of any ML algorithm. They are an efficient procedure, in so far as they reduce the input dimensionality of the ML algorithm and prevent overfitting. In this study, a widely used statistical parameter will be used: the Pearson correlation coefficient.

The FS Wrapper methods perform a global evaluation of the entire set of variables that creates a ranking of relevance. The ML algorithm executes the ranking and, consequently, the score is not universal and they cannot be applied to any other ML algorithm (Wald et al., 2014). In other words, the Wrapper methods will produce different rankings for different ML algorithms. Wrapper FS approaches are commonly used in the field of renewable energy applications due to their higher performance (Salcedo-Sanz et al., 2018). Permutation Importance, Recursive Feature Elimination and Boruta methods are all included within this category. In this work, the FS algorithms used for simplifying the classification trees are: Pearson correlation coefficient, Permutation

Importance, Recursive Feature Elimination and Boruta. The following paragraphs describe them and their use in other fields of the ML.

2.4.1. Pearson correlation coefficient criterion (*P*)

The Pearson criterion is based on the Pearson correlation coefficient, r . If two datasets X and X' are strongly correlated, the Pearson coefficient is 1 (direct correlation) or -1 (inverse correlation). However, a Pearson coefficient near 0 implies a weak or null correlation.

In this work, the Pearson criterion was applied in two steps: firstly for selecting the MIs with a strong correlation to the CIE cloudiness classification. Only the MIs with Pearson correlation coefficients above a certain threshold were selected and used in the next step for detecting the MIs with high correlations between them and for selecting the most important ones. After both steps, only the most important independent MIs for the classification were selected.

2.4.1.1. Permutation Importance (PI). Permutation Importance (PI) or the Mean Decrease in Accuracy (MDA) (Nembrini, 2019) algorithm is used to analyze how the score of the prediction model decreases when the data of a single variable is randomly permuted, generating random noise. Permutation feature importance is defined as decreasing in a model score when a single feature value is randomly shuffled (Bommert et al., 2020). A PI index of 0% means null relevance of this feature for the classification. Usually, a threshold of 5% is employed, considering only MIs with a permutation importance above 5% as important and discarding any others (Altmann et al., 2010).

2.4.1.2. Recursive feature Elimination (RFE). The Recursive Feature Elimination (RFE) method fits a model, so as to remove the weakest features until a specified number of variables is reached. A great number of ML classification algorithms such as Decision Trees, Support Vector Machines (Weston et al., 2001), and Random Forests (Diaz-Uriarte and Alvarez de Andres, 2006), among others, attach a weight to each input for the classification. The features are ranked in each loop and a few features per loop are removed, in an attempt to lower their interdependencies and collinearity. Also, the final size of the feature set cannot be initially specified and the number is established when there is no global improvement in the accuracy of the model. This method has been widely used with high-dimensional data sets (Escanilla et al., 2018; Paul et al., 2015). Fields where the algorithm has successfully been applied include genetics (Darst et al., 2018), materials science (Sharp et al., 2018), cancer studies (Duan et al., 2005), sports (Paul et al., 2015) and solar and wind forecasting (Benamrou et al., 2020; Feng et al., 2017).

2.4.1.3. Boruta (BOR). The Boruta (BOR) method, rather than comparing features between each other, competes with a randomized version of so-called “shadow features”. In each iteration, the importance given by the classification algorithm to each original feature is compared with the highest feature importance recorded among the shadow features. Each time the importance of a feature is higher than this threshold, it is called a “hit”. A feature is considered useful, if it performs better than the best randomized feature. Counting the number of hits, the selection of a feature is decided after a number of trials. In the same way as RFE, the BOR method performs a top-down search for relevant features, progressively eliminating irrelevant ones (Kursa and Rudnicki, 2010). RFE and Boruta have been compared on many occasions and in scientific fields such as genetics (Kursa, 2014) and spectroscopy (Poona et al., 2016). Permutation Importance is highly sensitive and effective (Gregorutti et al., 2017) when applied to biological data (Degenhardt et al., 2019).

2.5. Classification trees

A classification tree is an algorithm that classifies datasets into

certain outcome categories by using a sequence of “partitions”, or “splits”. However, the more complex the category analysis, the larger the sequence of splits that may be needed. Since the first implementation of Breiman et al. in 1984 (Breiman, 1984), classification trees have been used in a very large variety of disciplines, such as meteorology, medicine, and statistics, among others, and likewise CIE Standard skies classification (Umemiya and Kanou, 2008).

The structure of the classification tree can be implemented by several criteria. The one chosen for the sky cloudiness classification is the Classification and Regression Tree (CART) (Breiman, 1984). It looks for successive binary splits that chooses the partitions, in order to obtain the highest performance. Both the Gini (D'Ambrosio and Tutore, 2011) and the Entropy (Witten et al., 2016) indices were considered to fit the classification tree. The Gini Index points to how often a randomly chosen element from the set would be incorrectly labelled. The entropy index considers the disorder of a grouping by the target variable. Both of them are performance measures of the classification tree.

The classification tree algorithm was selected above other ML classification algorithms, due to the transparency of the results it can obtain. The classification tree produces a diagram that can be more easily understood than those produced by other ML techniques such as, Support Vector Machines, Neural Networks, Random Forest and Gradient Boosting, traditionally known as “black boxes”.

2.6. Classification metrics

Confusion matrices are a useful tool for the performance characterization of an classification algorithm. Four possible cases can be obtained in a classification procedure: true positive (TP), true negative (TN), false positive (FP) and false negative (FN). The number of each one in the confusion matrix summarizes the performance of a dichotomic classification, as shown in Fig. 4.

From the confusion matrix Precision, Pr , and Recall, Re , indices are defined in Eqs. (1) and (2), respectively, in order to measure the performance of the classification algorithm. Pr is the probability that a positive prediction was correct, while Re is the percentage of correctly detected positive cases. Both indices are independent from each other and could be explained by a very precise and not a very sensitive algorithm. Both indices are grouped in the $f1$ factor, defined in Eq. (3) as the harmonic average of precision and recall.

$$Pr = \frac{TP}{TP + FP} \quad (1)$$

$$Re = \frac{TP}{TP + FN} \quad (2)$$

$$f1 = \frac{2}{\frac{1}{Pr} + \frac{1}{Re}} = \frac{2TP}{2TP + FN + FP} \quad (3)$$

Sky classification is a multiclass classification. It is therefore necessary to sum up the indices for each category, which yields a global result for the classification algorithm performance, as shown in Fig. 5. Pr , Re , and $F1$ indices were calculated for each CIE Standard sky condition (Clear, Partial and Overcast).

Two different procedures can be used to obtain the global values: the macro-average and the weighted-average. The macro-average calculates the global value for each index as the average of the index for each category, regardless of the size of the category within the sample. Therefore, a low performance of the classification algorithm in one of the categories may affect overall performance, despite performing well in the other categories. This problem is as common in imbalanced class distribution as it is for the case of sky classification (see Fig. 2, where the FOC of clear, partial, and overcast conditions differ). The weighted-average yields the global value, by adding the results for each category and the weighting that represents the category size over the total number of cases. Therefore, the weighted aggregation used in this work

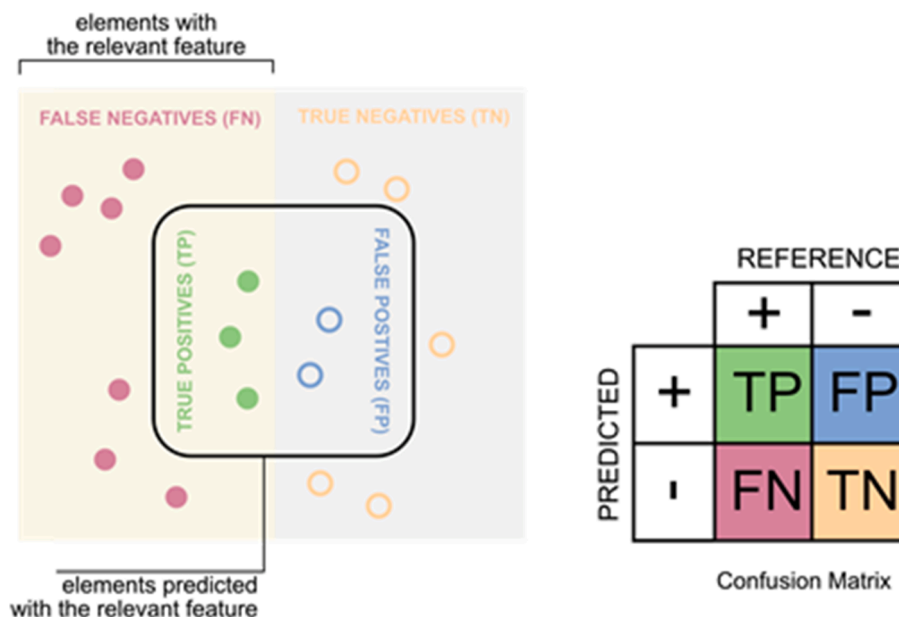


Fig. 4. Confusion matrix: possible cases in the comparison of the prediction with the actual data.

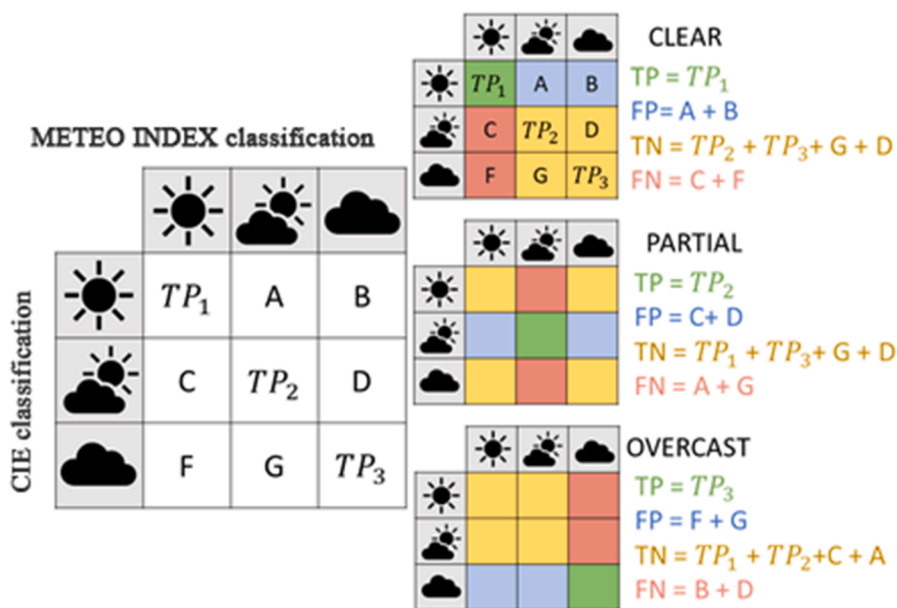


Fig. 5. Confusion Matrix for multi-class sky classification.

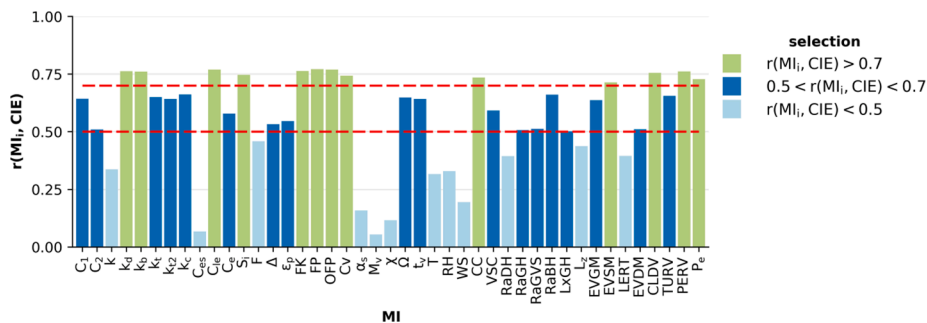


Fig. 6. Absolute value of the Pearson correlation between the MIs under consideration and the CIE Standard Sky classification, $r(MI_i, CIE)$.

will reward those algorithms with good performance in the most numerous classes and will have a lesser effect on those with poor performance, in the classes with fewer samples.

3. Results

3.1. Feature selection

Figs. 6 and 7 show the results of the FS using the Pearson correlation coefficient. Fig. 6 shows the absolute value of the correlation between each MI and the CIE classification, denoted as $r(MI_i, CIE)$. Following Thumb's rule (Mukaka, 2012), three r intervals were considered: high ($0.9 \geq r \geq 0.7$), moderate ($0.7 \geq r \geq 0.5$), and negligible ($r < 0.5$) correlations. The MIs with $r(MI_i, CIE) \geq 0.7$ were moved to the second stage of the Pearson correlation coefficient criterion. In the second step, an effort was made to discard the redundant MIs. Here, two MIs, MI_i and MI_k , are redundant if $r(MI_i, MI_k) > 0.9$. Fig. 7 represents the correlation between the MIs. The MIs for which $r(MI_i, MI_k) \geq 0.9$ are shaded in blue. The redundant groups of the MIs were formed by grouping the MIs with very high correlation represented as blue squares, for each column of the matrix of Fig. 7. Each MI was only included in one group, as shown in Table 5. In this Table 5, all MIs in the same group were considered to have the same information and only one of them, the one with the highest value of $r(MI_i, CIE)$, was needed to reflect the information of the rest. CC was related to the others MIs and was therefore included in the selection process. As can be seen, the original set of 43 MIs was reduced to two. 13 features (MIs) were selected from the Permutation Index (PI) results as necessary for CIE Standard Sky classification, as shown in Fig. 8. All of them caused a 5% decrease in the performance of the classification algorithm when they were randomly shuffled. The red line in the figure represents the aforementioned threshold. All the scores above the line, represent an impact higher than 5%. Other thresholds could be considered. However, the optimal threshold for each algorithm is a matter for further research.

Thirteen MIs were selected using the Recursive Feature Elimination (RFE) FS procedure that sets a minimum number of MIs needed for accurate CIE Standard Sky classification at 13 MIs. Fig. 9 shows the curve produced by recursive reduction of the number of MIs. The curve maintains an excellent $f1$ above 13 features when the most suitable variables that the algorithm selected were $CC, LxGH, VSC, WS, OFP, C_{le}, k_t, K, LERT, C_e, RH, L_z$. When fewer variables were in the classification tree,

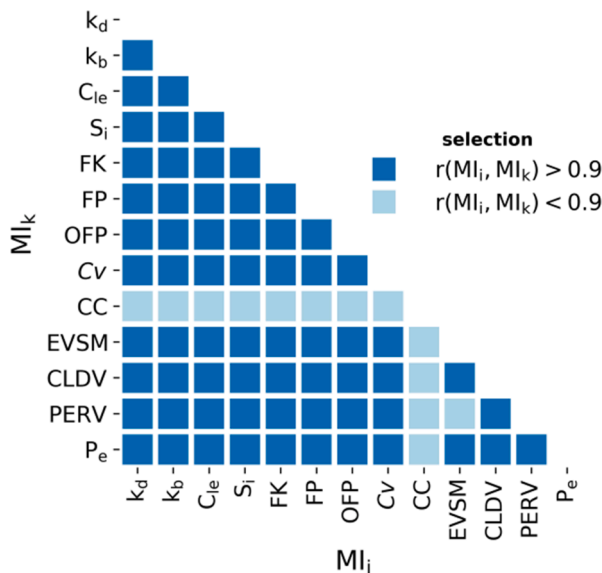


Fig. 7. Pearson correlation absolute value between MIs, $r(MI_i, MI_k)$, for MIs with $r(MI_i, CIE) \geq 0.7$.

the performance of the classification algorithm drastically decreased, because the most informative features were removed from the model. Conversely, redundant information was included, whenever additional variables were added.

One hundred trial tests of the Boruta (BOR) FS methodology were completed. Fig. 10 shows the hits of each one of the MI. All MIs got a hit and the maximum number of hits was below ten. The MI with a number of hits higher than one was selected.

The MIs selected by each FS procedure are summarized in Table 6. With the exception of the Boruta method, the FS procedures reduced the original set of MIs to a little less than 75%, selecting different MIs. The reduction in the number of variables required for the classification process, reflects the usefulness of the FS. Fewer variables to be measured and/or calculated implies less instrumentation and data storage, and simplifies the classification algorithm. Simpler models reduced the necessary computing power and, for example, made its implementation easier for lighting control systems.

The results of different feature methods selection, show the relationship existing between the variables, which in some cases can be directly deduced from the definition thereof, shown in Table 1 while in other cases does not appear so clearly reflected. The MIs selected by the FS algorithms can be classified into three types: variables related to the cloud conditions, others related to daylighting, brightness or clearness conditions of the skies and geometrical variables. While the Pearson FS method eliminates those variables that are most related to each other, in order not to include redundant information, the Boruta method does not eliminate a priori, highly related variables that can add distinctive nuances useful for classification. PI and RFE methods reach a compromise between information and complexity.

3.2. Classification trees

The classification trees for CIE Standard Sky Classification from the MIs selected by Pearson, Permutation Importance, RFE and Boruta FS procedures are shown in Figs. 11–14. Starting in the main left node, if the condition is met, the path of the upper branch is followed and, if not, the path followed is the one indicated by the lower branch. Evaluating each node consecutively, the sky conditions would be obtained. The number inside the nodes represents the number of samples inside each partition. The number of binary partitions or levels of the classification tree is a previously set parameter. In this work all the classification trees have four levels. An increased number of levels might increase the precision of the classification algorithm in the same way as complexity. The starting MI and the number of levels of the classification tree were selected following the Gini and the Entropy criteria, previously introduced.

FP and CC are MIs selected by the Pearson FS method for the classification tree. Both MIs are related to the cloud conditions, through the diffuse horizontal fraction (ratio diffuse horizontal irradiation to global horizontal irradiation) and the percentage of sky covered by clouds, respectively. The CIE Standard decision tree obtained from the variables selected by the Pearson FS method identifies the clear sky type by one of these cases: a) $FP > 0.51$, and $CC \geq 0.53$; b) $FP > 0.78$; c) $FP \leq 0.51$ and $CC \leq 0.66$.

Although the PI FS methods selected 11 MIs for the CIE sky classification, only three were used in the four-level classification tree: FP , CC , and VSC . The Vertical Sky Component, VSC , linked the classification to the daylighting.

The classification tree obtained with the MIs selected by the RFE FS method started with the CC , a variable which directly classified the skies as clear if $<61.7\%$. On the second and third levels of the classification tree, original Perraudeau's Index, OFP , and VSC , were evaluated. At the last level, the MI selected to fit the classification was luminous efficacy, K . Again, two of the MIs were related to daylighting (VSC, K) and CC were related to cloud coverage.

Boruta FS methods selected 34 MIs for the CIE standard sky classi-

Table 5
Results of Pearson FS method.

Group 1	k_d	k_b	C_{le}	S_i	FK	FP	$EVSM$	$CLDV$	$PERV$	P_e	OPF	C_v
$r(MI_i, CIE)$	0.762	0.761	0.770	0.747	0.764	0.771	0.714	0.755	0.762	0.729	0.769	0.740

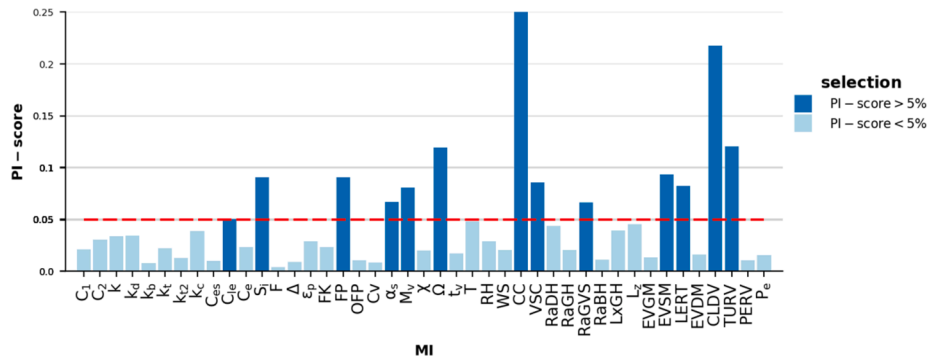


Fig. 8. Permutation Index (PI) results in feature selection of MIs for CIE Standard Sky classification.

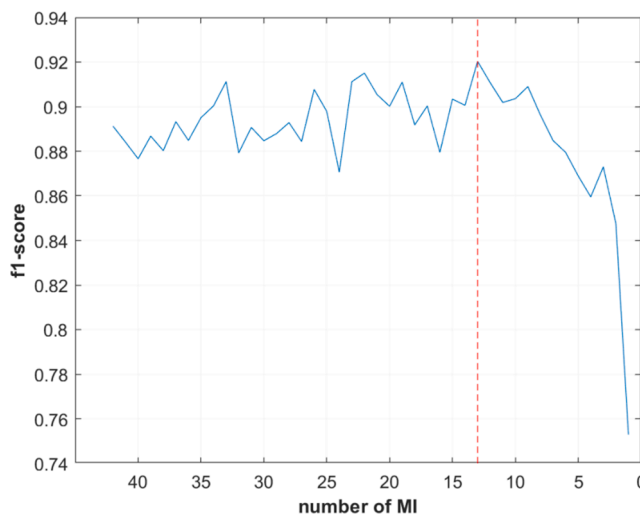


Fig. 9. Recursive Feature Elimination (RFE) results in FS of MIs for CIE Standard Sky classification.

fication, but four were necessary to build the four-level classification tree. The sky classification started by evaluating Perradeau’s nebulosity index, FP (Kambežidis et al., 1998). At the second level, cloud cover, CC , and the vertical sky component, VSC were investigated.

Finally, the scattering angle, χ , a geometrical variable, was investigated.

As regards the intervals established by the classification trees for each partition with respect to the one established by the authors in their original works, it is important to remark that the number of sky categories is different for some MIs (for example FP establishes 5 sky categories, instead of three). However, the original intervals and those obtained were in consonance.

3.3. Analysis of the classification trees using different metrics

Fig. 15 shows the results of the Pr , Re , and $f1$ metrics obtained for the classification trees calculated from the MIs selected by each FS

Table 6
Summary of the features (MIs) selected by each one of the FS algorithms.

Feature Selection	MI selected	Number
Pearson correlation coefficient criterion (P)	FP, CC	2
Permutation Importance (PI)	$C_{le}, S_i, FP, \alpha_s, M_v, \Omega, CC, VSC, RaGVS, EVSM, LERT, CLDV, TURV$	13
Recursive Feature Elimination (RFE)	$K, k_t, C_{le}, C_e, OPF, \chi, RH, WS, CC, VSC, LxGH, L_z, LERT$	13
Boruta (BOR)	$C_1, C_2, CC, CLDV, C_e, \chi, C_{le}, C_v, \epsilon_p, EVDM, EVGM, EVSMF, FK, FP, LERT, L_z, OPF, \Omega, PERV, P_e, RaBH, RaDHRaGVS, \Delta, S_i, TURV, VSC, k_b, k_c, k_d, k_t, k_{t2}, t_v$	34

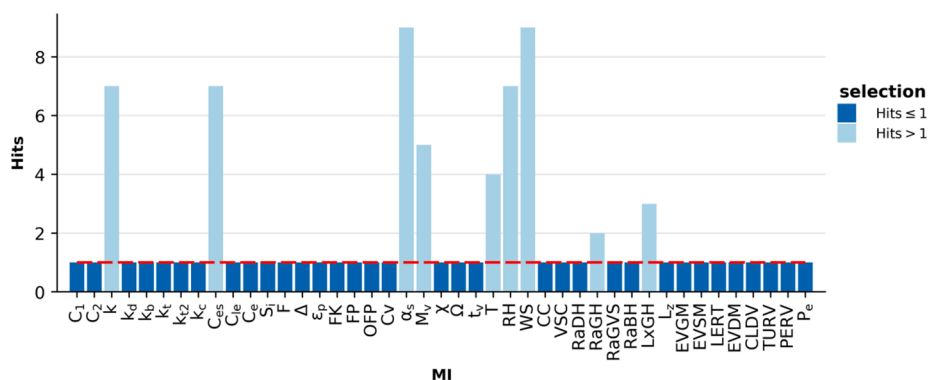


Fig. 10. Results of the Boruta FS methodology for CIE Standard Sky Classification after 100 trial tests.

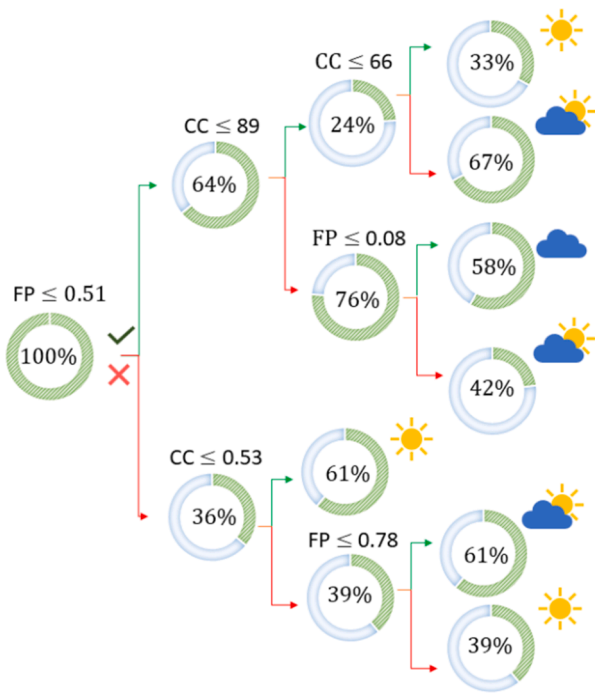


Fig. 11. CIE standard sky classification tree (MIs selected with the Pearson FS method).

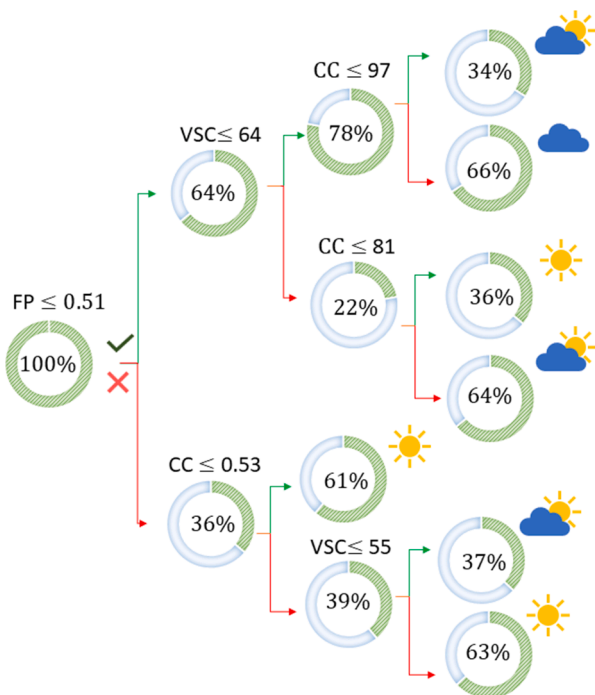


Fig. 12. CIE standard sky classification tree (MIs selected with the PI FS method).

procedure and for the different sky conditions (clear, partial, and overcast). The best results for all metrics were obtained in overcast sky conditions, with P_r and Re_{above} 85% and reaching 90% in the case of the classification tree that applied the four MIs selected for the RFE method (CC , OFP , VSC , K). The classification trees that used the MIs selected by both PI (FP , VSC , CC), and BORUTA (FP , VSC , CC , χ) obtained the same results. The simplest classification tree, from the two MIs selected by the P method, (CC , FP), yielded worse f_1 metrics.

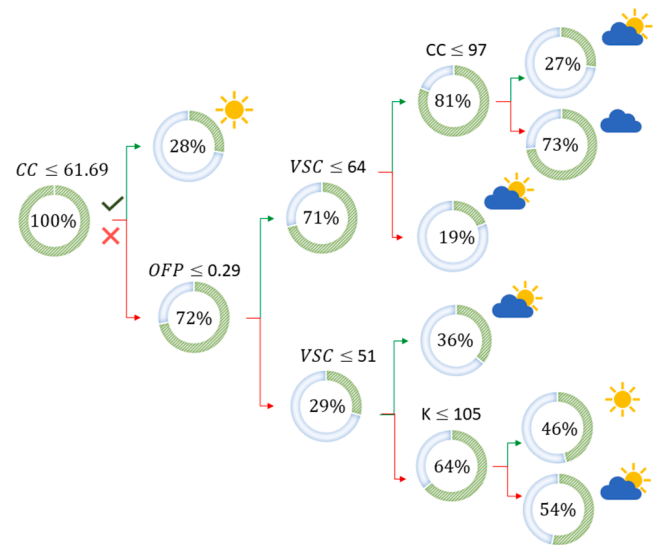


Fig. 13. CIE standard sky classification tree (MIs selected with the RFE FS method).

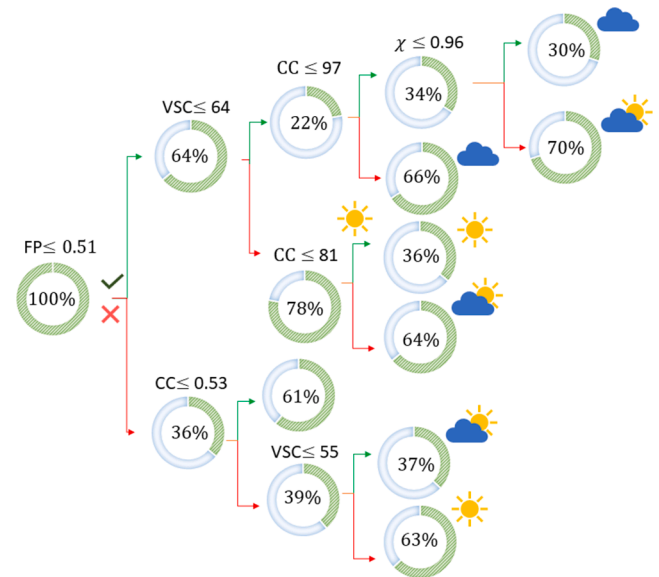


Fig. 14. CIE standard sky classification tree. (MIs selected with the Boruta FS method).

In the identification of clear sky conditions, all classification trees presented more dispersion of the metrics value: f_1 ranged from 65% (P method) to 87% (BORUTA method), but all indices exceeded 65%. The identification of partially covered skies was worse, lowering the values of all indices by between 55% and 70%. The irregularity of partially covered sky conditions, the high variability of the MIs for these conditions and the dependency of cloud cover with respect to the Sun might explain this fact. In every case, the RFE FS method yielded the closest values of the three metrics.

The weighted-averaged global f_1 is shown in Fig. 16. As can be seen, all the classification trees yielded results between 74% and 77%, the highest value of which was produced by the RFE FS procedure, very close to the BOR and PI FS procedure with the same value of f_1 . The Pearson method, also the simplest classification tree, showed the lowest f_1 value. Taking into account the number of MIs used by each classification tree, perhaps the RFE FS methods offered the best performance with no high complexity, but the results highlighted no significant

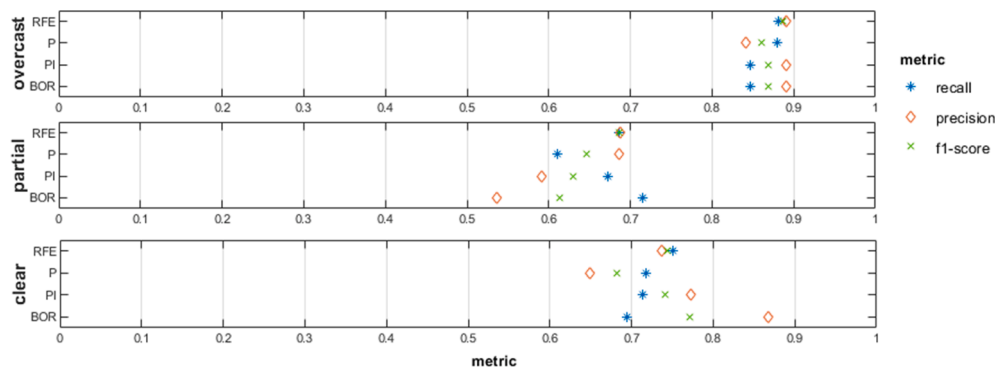


Fig. 15. Precision, Pr , Recall, Re and $f1$ indices calculated for each CIE Standard sky type for each classification tree based on the different FS procedures.

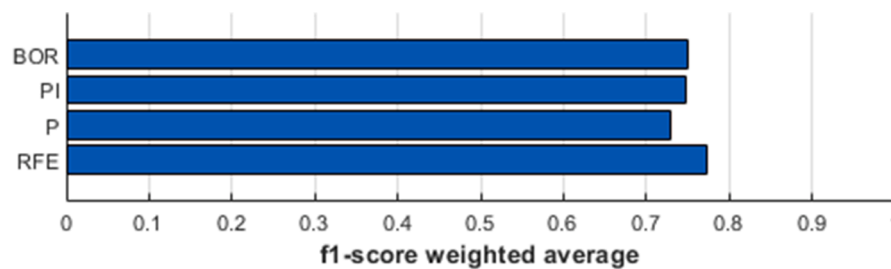


Fig. 16. Weighted-averaged global $f1$.

advantages between the classification algorithms constructed from the different feature selection procedures used in this work.

4. Conclusions

This study has highlighted the usefulness of the FS procedure for adequate determination of MIs for sky classification in accordance with the CIE Standard classification, as an alternative to the use of sky-scanner devices. The maximum number of MIs can be identified with FS for use as an input for the ML algorithm, avoiding the introduction of redundant and useless information. Four FS (filter and wrapper) methods have been reviewed and applied. The initial set of 43 MIs was drastically reduced by three of the FS algorithms (Pearson, PI and RFE), although a less significant reduction was achieved with the Boruta FS method. The main advantage of the Pearson FS procedure over and above all the other methods that were tested was its independence from the ML algorithm used after the FS procedure, with the consequent saving of time when it was necessary to verify the operation of different ML algorithms.

All the classification trees yielded performances that were similar to the CIE standard sky classification in terms of the Pr , Re and $f1$ metrics. The worse results were shown for the identification of partially cloudy conditions, while the overcast and clear sky conditions were identified with high success rates. No significant differences in the performance of the classification algorithms constructed from the MIs selected by the different FS methods have been pointed out, and the use of one or another FS method could be at the discretion of the researcher.

The MIs selected by the FS algorithms can be classified into three types: variables related to the cloud conditions, (FP , CC , $OFFP$), others related to daylighting, brightness or clearness conditions of the skies (VSC , K), and geometrical variables, such as α_s and χ .

Both the intervals established by the classification trees for each partition and those established by the authors in their original works were in consonance. However, the classification tree might be a good alternative, in order to set up these intervals independently from local climatic and meteorological conditions.

Supplementary Materials

Experimental data and Python code in this paper are available with copyleft licenses for the data and the code in the following link: <http://hdl.handle.net/10259/5563> (DOI: 10.36443/10259/5563).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors gratefully acknowledge the financial support provided by the Regional Government of Castilla y León under the “Support Program for Recognized Research Groups of Public Universities of Castilla y León” (BU021G19) and the Spanish Ministry of Science & Innovation under the I + D + i state program “Challenges Research Projects” (Ref. RTI2018-098900-B-I00). Diego Granados López express his thanks to Junta de Castilla-León for economic support (PIRTU Program, ORDEN EDU/556/2019).

References

- Alshabani, K., 2011. Finding frequency distributions of CIE Standard General Skies from sky illuminance or irradiance. *Light. Res. Technol.* 43 (4), 487–495. <https://doi.org/10.1177/1477153511404999>.
- Alshabani, K., 2016a. Average daylight factor for the ISO/CIE Standard General Sky. *Light. Res. Technol.* 48 (6), 742–754. <https://doi.org/10.1177/1477153515572939>.
- Alshabani, K., 2016b. The use of horizontal sky illuminance to classify the CIE Standard General Skies. *Light. Res. Technol.* 48 (8), 1034–1041. <https://doi.org/10.1177/1477153515624485>.
- Alshabani, K., 2017. Classification Standard Skies: the use of horizontal sky illuminance. *Renew. Sust. Energ. Rev.* 73, 387–392. <https://doi.org/10.1016/j.rser.2017.01.116>.
- Altmann, A., Tolosi, L., Sander, O., Lengauer, T., 2010. Permutation importance: a corrected feature importance measure. *Bioinform.* 26 (10), 1340–1347. <https://doi.org/10.1093/bioinformatics/btq134>.
- Allard, D., Ailliot, P., Monbet, V., Naveau, P., 2015. Stochastic weather generators: an overview of weather type models. *Journal de la Société Française de Statistique* 156 (1), 101–113. <https://hal.inrae.fr/hal-02641587>.

- Aries, M.B.C., Aarts, M.P.J., Van Hoof, J., 2015. Daylight and health: a review of the evidence and consequences for the built environment. *Light. Res. Technol.* 47 (1), 6–27. <https://doi.org/10.1177/1477153513509258>.
- Benamrou, B., Ouardouz, M., Allaouzi, I., Ben Ahmed, M., 2020. A proposed model to forecast hourly global solar irradiation based on satellite derived data, deep learning and machine learning approaches. *J. Ecol. Eng.* 21 (4).
- Biesiada, J., Duch, W., 2007. Feature selection for high-dimensional data — a Pearson redundancy based filter. In: Kurzynski M., P.E., Wozniak M., Zolnierek A. (Ed.) *Computer Recognition Systems 2. Advances in Soft Computing*. Springer, Berlin, Heidelberg, pp. 242–249.
- Bommert, A., Sun, X., Bischl, B., Rahnenführer, J., Lang, M., 2020. Benchmark for filter methods for feature selection in high-dimensional classification data. *Comput. Stat. Data An.* 143, 106839. <https://doi.org/10.1016/j.csda.2019.106839>.
- Breiman, L., 1984. *Classification and Regression Trees*. Routledge, New York.
- Comission Internationale de, L.E., 1995. Guide to recommended practice of daylight measurement, CIE 108-1994. *Color Res. Appl.* 20(1), 80–80. <https://doi.org/10.1002/col.5080200118>.
- Chaiwivatworakul, P., Chirattananon, S., 2004. Distribution of sky luminance in tropical climate. In: Proceedings of the Joint International Conference on Sustainable Energy and Environment, Thailand, 1–3 December. pp. 530–537.
- Chen, W., Li, D.H.W., Lou, S., 2019. Estimation of irregular obstructed vertical sky components under various CIE skies. *Energy Procedia* 158, 309–314. <https://doi.org/10.1016/j.egypro.2019.01.094>.
- Chung, T.M., 1992. A study of luminous efficacy of daylight in Hong Kong. *Energy Build.* 19 (1), 45–50. [https://doi.org/10.1016/0378-7788\(92\)90034-E](https://doi.org/10.1016/0378-7788(92)90034-E).
- D'Ambrosio, A., Tutore, V.A., 2011. Conditional classification trees by weighting the gini impurity measure. In: 7th Biannual Meeting of the Classification and Data Analysis Group, CLADAG 2009. Catania, pp. 273–280. https://doi.org/10.1007/978-3-642-11363-5_31.
- Darst, B.F., Malecki, K.C., Engelman, C.D., 2018. Using recursive feature elimination in random forest to account for correlated variables in high dimensional data. *BMC Genet.* 19 (Suppl 1), 65. <https://doi.org/10.1186/s12863-018-0633-8>.
- Darula, S., Kittler, R., 2002. CIE general sky standard defining luminance distributions. *Proc. eSim* 11–13.
- Darula, S., Kittler, R., Kómar, L., 2013. Sky type determination using vertical illuminance. *Przełąd Elektrotechniczny* 89 (6), 315–319.
- Degenhardt, F., Seifert, S., Szymczak, S., 2019. Evaluation of variable selection methods for random forests and omics data sets. *Brief. Bioinform.* 20 (2), 492–503. <https://doi.org/10.1093/bib/bbx124>.
- Diaz-Urriarte, R., Alvarez de Andres, S., 2006. Gene selection and classification of microarray data using random forest. *BMC Bioinform.* 7, 3. <https://doi.org/10.1186/1471-2105-7-3>.
- Dieste-Velasco, M.I., Díez-Mediavilla, M., Granados-López, D., González-Peña, D., Alonso-Tristán, C., 2019. Performance of global luminous efficacy models and proposal of a new model for daylighting in Burgos, Spain. *Renew. Energy* 133, 1000–1010. <https://doi.org/10.1016/j.renene.2018.10.085>.
- Duan, K.-B., Rajapakse, J.C., Wang, H., Azuaje, F., 2005. Multiple SVM-RFE for gene selection in cancer classification with expression data. *IEEE Trans. Nanobiosci.* 4 (3), 228–234.
- Dubois, M.C., Gentile, N., Amorim, C.N.D., Osterhaus, W., Stoffer, S., Jakobiak, R., Geisler-Moroder, D., Matusiak, B., Onarheim, F.M., Tetri, E., 2016. Performance evaluation of lighting and daylighting retrofits: results from IEA SHC task 50. *Energy Procedia* 91, 926–937. <https://doi.org/10.1016/j.egypro.2016.06.259>.
- Edwards, L., Torcellini, P., 2002. Literature review of the effects of natural light on building occupants. National Renewable Energy Lab., Golden, CO.(US).
- Escanilla, N.S., Hellerstein, L., Kleiman, R., Kuang, Z., Shull, J.D., Page, D., 2018. Recursive Feature Elimination by Sensitivity Testing. In: Proceedings of the ... International Conference on Machine Learning and Applications. International Conference on Machine Learning and Applications 2018, pp. 40–47. <https://doi.org/10.1109/ICMLA.2018.00014>.
- Fakra, A.H., Boyer, H., Miranville, F., Bigot, D., 2011. A simple evaluation of global and diffuse luminous efficacy for all sky conditions in tropical and humid climate. *Renew. Energy* 36 (1), 298–306. <https://doi.org/10.1016/j.renene.2010.06.042>.
- Feng, C., Cui, M., Hodge, B.-M., Zhang, J., 2017. A data-driven multi-model methodology with deep feature selection for short-term wind forecasting. *Appl. Energy* 190, 1245–1257.
- Ferraro, V., Igawa, N., Marinelli, V., 2010. INLUX-DBR – A calculation code to calculate indoor natural illuminance inside buildings under various sky conditions. *Energy* 35 (9), 3722–3730. <https://doi.org/10.1016/j.energy.2010.05.021>.
- Fouquart, Y., Buriez, J.C., Herman, M., Kandel, R.S., 1990. The influence of clouds on radiation: a climate-modeling perspective. *Rev. Geophys.* 28 (2), 145–166. <https://doi.org/10.1029/RG028i002p0145>.
- Granados-López, D., Díez-Mediavilla, M., Dieste-Velasco, M.I., Suárez-García, A., Alonso-Tristán, C., 2020. Evaluation of the vertical sky component without obstructions for daylighting in Burgos, Spain. *Appl. Sci.* 10 (9), 3095.
- Gregorutti, B., Michel, B., Saint-Pierre, P., 2017. Correlation and variable importance in random forests. *Stat. and Comput.* 27 (3), 659–678. <https://doi.org/10.1007/s11222-016-9646-1>.
- Gueymard, C.A., 2018. A reevaluation of the solar constant based on a 42-year total solar irradiance time series and a reconciliation of spaceborne observations. *Sol. Energy* 168, 2–9. <https://doi.org/10.1016/j.solener.2018.04.001>.
- Gueymard, C.A., Ruiz-Arias, J.A., 2016. Extensive worldwide validation and climate sensitivity analysis of direct irradiance predictions from 1-min global irradiance. *Sol. Energy* 128, 1–30. <https://doi.org/10.1016/j.solener.2015.10.010>.
- Igawa, N., Koga, Y., Matsuzawa, T., Nakamura, H., 2004. Models of sky radiance distribution and sky luminance distribution. *Sol. Energy* 77 (2), 137–157. <https://doi.org/10.1016/j.solener.2004.04.016>.
- Inman, R.H., Pedro, H.T.C., Coimbra, C.F.M., 2013. Solar forecasting methods for renewable energy integration. *Prog. Energy Combust. Sci.* 39 (6), 535–576. <https://doi.org/10.1016/j.peccs.2013.06.002>.
- ISO, 2004. ISO-15469:2004 (E). Spatial distribution of daylight-CIE standard general sky. Geneva, Switzerland.
- Janjai, S., Tohsing, K., Nunez, M., Laksanaboonsong, J., 2008. A technique for mapping global illuminance from satellite data. *Sol. Energy* 82 (6), 543–555. <https://doi.org/10.1016/j.solener.2007.11.003>.
- Kambezdids, H.D., 2018. The solar radiation climate of Athens: variations and tendencies in the period 1992–2017, the brightening era. *Sol. Energy* 173, 328–347. <https://doi.org/10.1016/j.solener.2018.07.076>.
- Kambezdids, H.D., Muneer, T., Tzortzis, M., Arvanitaki, S., 1998. Global and diffuse horizontal solar illuminance: month-hour distribution for Athens, Greece in 1992. *Light. Res. Technol.* 30 (2), 69–74. <https://doi.org/10.1177/096032719803000203>.
- Kittler, R., Danda, S., 2000. Determination of sky types from global illuminance. *Int. J. Light. Res. Technol.* 32 (4), 187–193.
- Klucher, T.M., 1979. Evaluation of models to predict insolation on tilted surfaces. *Sol. Energy* 23 (2), 111–114.
- Kocifaj, M., 2011. CIE standard sky model with reduced number of scaling parameters. *Sol. Energy* 85 (3), 553–559. <https://doi.org/10.1016/j.solener.2010.12.024>.
- Kursa, M.B., 2014. Robustness of Random Forest-based gene selection methods. *BMC Bioinform.* 15 (1), 8.
- Kursa, M.B., Rudnicki, W.R., 2010. Feature selection with the Boruta package. *J. Stat. Softw.* 36 (11), 1–13.
- Li, D., Lam, T., Wu, T., 2014a. Estimation of average daylight factor under obstructed CIE Standard General Skies. *Light. Res. Technol.* 46 (2), 187–197. <https://doi.org/10.1177/1477153512453578>.
- Li, D.H.W., 2010. A review of daylight illuminance determinations and energy implications. *Appl. Energy* 87 (7), 2109–2118. <https://doi.org/10.1016/j.apenergy.2010.03.004>.
- Li, D.H.W., Chau, T.C., Wan, K.K.W., 2014b. A review of the CIE general sky classification approaches. *Renew. Sust. Energy. Rev.* 31, 563–574. <https://doi.org/10.1016/j.rser.2013.12.018>.
- Li, D.H.W., Cheung, G.H.W., Cheung, K.L., 2006. Evaluation of simplified procedure for indoor daylight illuminance determination against data in scale model measurements. *Indoor Built Environ.* 15 (3), 213–223. <https://doi.org/10.1177/1420326x06066300>.
- Li, D.H.W., Cheung, K.L., Tang, H.L., Cheng, C.C.K., 2011a. Identifying CIE standard skies using vertical sky component. *J. Atmos. Sol.-Terres. Phys.* 73 (13), 1861–1867. <https://doi.org/10.1016/j.jastp.2011.04.015>.
- Li, D.H.W., Lau, C.C.S., Lam, J.C., 2004. Standard skies classification using common climatic parameters. *J. Sol. Energy Eng.* 126 (3), 957–964. <https://doi.org/10.1115/1.1740776>.
- Li, D.H.W., Lou, S., Lam, J.C., Wu, R.H.T., 2016. Determining solar irradiance on inclined planes from classified CIE (International Commission on Illumination) standard skies. *Energy* 101, 462–470. <https://doi.org/10.1016/j.energy.2016.02.054>.
- Li, D.H.W., Lou, S.W., Lam, J.C., 2015. An analysis of global, direct and diffuse solar radiation. *Energy Procedia* 75, 388–393. <https://doi.org/10.1016/j.egypro.2015.07.399>.
- Li, D.H.W., Tang, H.L., 2008. Standard skies classification in Hong Kong. *J. Atmos. Sol.-Terr. Phys.* 70 (8), 1222–1230. <https://doi.org/10.1016/j.jastp.2008.03.004>.
- Li, D.H.W., Tang, H.L., Cheung, K.L., Lee, E.W.M., Cheng, C.C.K., 2011b. Sensitivity analysis of climatic parameters for sky classification. *Theor. Appl. Climatol.* 105 (3–4), 297–309. <https://doi.org/10.1007/s00704-010-0392-6>.
- Li, D.H.W., Tang, H.L., Lee, E.W.M., Muneer, T., 2010. Classification of CIE standard skies using probabilistic neural networks. *Int. J. Climatol.* 30, 305–315. <https://doi.org/10.1002/joc.1891>.
- Littlefair, P.J., 2012. Building Research Establishment, Site layout planning for daylight.
- Lou, S., Li, D.H.W., Lam, J.C., 2017. CIE Standard Sky classification by accessible climatic indices. *Renew. Energy* 113, 347–356. <https://doi.org/10.1016/j.renene.2017.06.013>.
- Lou, S., Li, D.H.W., Chen, W., 2019. A study of overcast, partly cloudy and clear skies by global illuminance and its variation features. *IOP Conf. Ser.: Mater. Sci. Eng.* 556, 012015. <https://doi.org/10.1088/1757-899x/556/1/012015>.
- Markou, M.T., Kambezidis, H.D., Bartzokas, A., Katsoulis, B.D., Muneer, T., 2005. Sky type classification in Central England during winter. *Energy* 30 (9), 1667–1674. <https://doi.org/10.1016/j.energy.2004.05.002>.
- Markou, M.T., Kambezidis, H.D., Katsoulis, B.D., Muneer, T., Bartzokas, A., 2004. Sky type classification in South England during the winter period. *Build Res. J.* 52 (1), 19–30.
- Mitra, P., Murthy, C., Pal, S.K., 2002. Unsupervised feature selection using feature similarity. *IEEE Trans. Pattern Anal. Mach. Intell.* 24 (3), 301–312.
- Mukaka, M., 2012. *Statistics Corner: a guide to appropriate use of Correlation coefficient in medical research Malawi. Med. J.*
- Muneer, T., 2007. *Solar Radiation and Daylight Models*. Routledge, New York.
- Muneer, T., Younes, S., Munawwar, S., 2007. Discourses on solar radiation modeling. *Renew. Sust. Energy. Rev.* 11 (4), 551–602. <https://doi.org/10.1016/j.rser.2005.05.006>.
- Nembrini, S., 2019. On the behaviour of permutation-based variable importance measures in random forest clustering. *J. Chemomet.* 33 (8) <https://doi.org/10.1002/cem.3135>.

- Ng, E., Cheng, V., Gadi, A., Mu, J., Lee, M., Gadi, A., 2007. Defining standard skies for Hong Kong. *Handbook of Environmental Chemistry, Volume 5: Water Pollution* 42 (2), 866–876. <https://doi.org/10.1016/j.buildenv.2005.10.005>.
- Pasero, E., Mesi, L., 2010. *Artificial Neural Networks for Pollution Forecast, Air Pollution*. IntechOpen.
- Paul, J., D'Ambrosio, R., Dupont, P., 2015. Kernel methods for heterogeneous feature selection. *Neurocomputing* 169, 187–195. <https://doi.org/10.1016/j.neucom.2014.12.098>.
- Perez, R., Ineichen, P., Seals, R., Michalsky, J., Stewart, R., 1990a. Modeling daylight availability and irradiance components from direct and global irradiance. *Sol. Energy* 44 (5), 271–289. [https://doi.org/10.1016/0038-092X\(90\)90055-H](https://doi.org/10.1016/0038-092X(90)90055-H).
- Perez, R., Ineichen, P., Seals, R., Zelenka, A., 1990b. Making full use of the clearness index for parameterizing hourly insolation conditions. *Sol. Energy* 45 (2), 111–114. [https://doi.org/10.1016/0038-092X\(90\)90036-C](https://doi.org/10.1016/0038-092X(90)90036-C).
- Poona, N.K., van Niekerk, A., Nadel, R.L., Ismail, R., 2016. Random Forest (RF) wrappers for waveband selection and classification of hyperspectral data. *Appl. Spectrosc.* 70 (2), 322–333. <https://doi.org/10.1177/0003702815620545>.
- 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(7), 660–666. <https://doi.org/10.1016/j.enbuild.2004.01.012>.
- Salcedo-Sanz, S., Cornejo-Bueno, L., Prieto, L., Paredes, D., García-Herrera, R., 2018. Feature selection in machine learning prediction systems for renewable energy applications. *Renewable Sustainable Energy Rev.* 90, 728–741. <https://doi.org/10.1016/j.rser.2018.04.008>.
- Sharp, T.A., Thomas, S.L., Cubuk, E.D., Schoenholz, S.S., Srolovitz, D.J., Liu, A.J., 2018. Machine learning determination of atomic dynamics at grain boundaries. *Proc. Natl. Acad. Sci.* 115 (43), 10943–10947.
- Solorio-Fernández, S., Carrasco-Ochoa, J.A., Martínez-Trinidad, J.F., 2019. A review of unsupervised feature selection methods. *Artif. Intell. Rev.* 53 (2), 907–948. <https://doi.org/10.1007/s10462-019-09682-y>.
- Suárez-García, A., Granados-López, D., González-Peña, D., Díez-Mediavilla, M., Alonso-Tristán, C., 2018. Seasonal characterization of CIE standard sky types above Burgos, northwestern Spain. *Sol. Energy* 169, 24–33. <https://doi.org/10.1016/j.solener.2018.04.028>.
- Torres, J.L., de Blas, M., García, A., Gracia, A., de Francisco, A., 2010a. Sky luminance distribution in Pamplona (Spain) during the summer period. *J. Atmos. Sol.-Terres. Phys.* 72 (5–6), 382–388. <https://doi.org/10.1016/j.jastp.2009.12.005>.
- Torres, J.L., de Blas, M., García, A., Gracia, A., de Francisco, A., 2010b. Sky luminance distribution in the North of Iberian Peninsula during winter. *J. Atmos. Sol.-Terres. Phys.* 72 (16), 1147–1154. <https://doi.org/10.1016/j.jastp.2010.07.001>.
- Trezenza, P.R., 1999. Standard skies for maritime climates. *Light. Res. Technol.* 31 (3), 97–106. <https://doi.org/10.1177/096032719903100304>.
- Trezenza, P.R., 2004. Analysing sky luminance scans to obtain frequency distributions of CIE Standard General Skies. *Lighting Res. Technol.* 36 (4), 271–279. <https://doi.org/10.1191/1477153504lii1170a>.
- Uetani, Y., Aydinli, S., Joukoff, A., Kendrick, J.D., Kittler, R., Koga, Y., 2003. BS ISO 15469:2004. Spatial distribution of daylight-CIE standard general sky. Vienna, Austria.
- Umemiya, N., Kanou, T., 2008. Classification of sky conditions by the ranges of insolation indices considering CIE standard for general sky. *J. Light Vis. Environ.* 32 (1), 14–19. <https://doi.org/10.2150/jlve.32.14>.
- Visa, S., Ramsay, B., Ralescu, A.L., Van Der Knaap, E., 2011. Confusion matrix-based feature selection. *MAICS 710*, 120–127.
- Wald, R., Khoshgoftaar, T.M., Napolitano, A., 2014. Optimizing wrapper-based feature selection for use on bioinformatics data. *The Twenty-Seventh International Flairs Conference*.
- Weston, J., Mukherjee, S., Chapelle, O., Pontil, M., Poggio, T., Vapnik, V., 2001. Feature selection for SVMs. *Adv. Neural Inf. Process. Syst.* 668–674.
- Witten, I.H., Frank, E., Hall, M.A., Pal, C.J., 2016. *Data mining: practical machine learning tools and techniques*.
- Wong, S.L., Wan, K.K.W., Li, D.H.W., Lam, J.C., 2012. Generation of typical weather years with identified standard skies for Hong Kong. *Build. Environ.* 56, 321–328. <https://doi.org/10.1016/j.buildenv.2012.04.003>.
- Yang, Y., Pedersen, J.O., 1997. *A comparative study on feature selection in text categorization*. *Icml*, Nashville, TN, USA, p. 35.
- Yu, L., Liu, H., 2003. Feature selection for high-dimensional data: A fast correlation-based filter solution. In: *Proceedings of the 20th international conference on machine learning (ICML-03)*. pp. 856–863.
- Zi, Y., Sun, C., Han, Y., 2020. Sky type classification in Harbin during winter. *J. Asian Archit. Build. Eng.* 1–12. <https://doi.org/10.1080/13467581.2020.1752217>.

Paper IV

Granados-López, D., García-Rodríguez, A., García-Rodríguez, S., Suárez-García, A., Díez-Mediavilla, M., Alonso-Tristán, C., 2021. ***Pixel-based image processing for CIE standard sky classification through ANN.*** Complexity 2021, 1-15. DOI: <https://doi.org/10.1155/2021/2636157>

Abstract: Digital sky images are studied for the definition of sky conditions in accordance with the CIE Standard General Sky Guide. Likewise, adequate image-processing methods are analyzed that highlight key image information, prior to the application of Artificial Neural Network classification algorithms. Twenty-two image-processing methods are reviewed and applied to a broad and unbiased dataset of 1500 sky images recorded in Burgos, Spain, over an extensive experimental campaign. The dataset comprises one hundred images of each CIE standard sky type, previously classified from simultaneous sky scanner data. Color spaces, spectral features, and texture filters image-processing methods are applied. While the use of the traditional RGB color space for image-processing yielded good results (ANN accuracy equal to 86.6%), other color spaces, such as Hue Saturation Value (HSV), which may be more appropriate, increased the accuracy of their global classifications. The use of either the green or the blue monochromatic channels improved sky classification, both for the fifteen CIE standard sky types and for simpler classification into clear, partial, and overcast conditions. The main conclusion was that specific image-processing methods could improve ANN-algorithm accuracy, depending on the image information required for the classification problem.

Research Article

Pixel-Based Image Processing for CIE Standard Sky Classification through ANN

D. Granados-López ¹, **A. García-Rodríguez** ¹, **S. García-Rodríguez** ¹,
A. Suárez-García ^{1,2}, **M. Díez-Mediavilla** ¹ and **C. Alonso-Tristán** ¹

¹Research Group Solar and Wind Feasibility Technologies (SWIFT),

Electromechanical Engineering Department, Escuela Politécnica Superior, 09006 Burgos, Spain

²Centro Universitario de la Defensa, Escuela Naval Militar de Marín, Pontevedra 36920, Marín, Spain

Correspondence should be addressed to C. Alonso-Tristán; catristan@ubu.es

Received 9 September 2021; Accepted 26 November 2021; Published 20 December 2021

Academic Editor: Lingzhong Guo

Copyright © 2021 D. Granados-López et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Digital sky images are studied for the definition of sky conditions in accordance with the CIE Standard General Sky Guide. Likewise, adequate image-processing methods are analyzed that highlight key image information, prior to the application of Artificial Neural Network classification algorithms. Twenty-two image-processing methods are reviewed and applied to a broad and unbiased dataset of 1500 sky images recorded in Burgos, Spain, over an extensive experimental campaign. The dataset comprises one hundred images of each CIE standard sky type, previously classified from simultaneous sky scanner data. Color spaces, spectral features, and texture filters image-processing methods are applied. While the use of the traditional RGB color space for image-processing yielded good results (ANN accuracy equal to 86.6%), other color spaces, such as Hue Saturation Value (HSV), which may be more appropriate, increased the accuracy of their global classifications. The use of either the green or the blue monochromatic channels improved sky classification, both for the fifteen CIE standard sky types and for simpler classification into clear, partial, and overcast conditions. The main conclusion was that specific image-processing methods could improve ANN-algorithm accuracy, depending on the image information required for the classification problem.

1. Introduction

Sky conditions are crucial factors when assessing daylighting levels and solar-energy output. The sky is generally classified on the basis of cloud presence into three categories: cloudless, partially cloudy, and overcast. Many models for the calculation of global, direct, and diffuse irradiation and illumination were defined for different sky types based on the values of several climatic parameters [1]. In 2003, the Commission Internationale de L'Éclairage (CIE) adopted the set of 15 standard sky classifications proposed by Kittler et al., in 1998, categorized under 3 sky types, clear, partial, and overcast, each of five grades [2]. These CIE standard skies that classify a general spectrum of homogeneous skies throughout the world were standardized in ISO 15469:2004(E)/CIE S 011/E:2003 [3] for the purpose of evaluating

indoor visual comfort within buildings [4], solar irradiance calculations [5], and energy efficiency improvements to lighting [6], among other applications.

The CIE standard sky classification is based on taking luminance measurements [7] of diffuse luminance angular distribution in the sky vault. Skies within a CIE category have approximately the same well-defined sky luminance and solar radiance patterns.

Devices called sky scanners are used to measure sky luminance patterns. According to the CIE Guide [8], a reliable commercial sky scanner measures luminance from 145 patches of sky hemisphere. However, various alternative procedures have been developed for CIE standard sky classification [9], due to the scarcity of sky scanners available to gather sky luminance data at ground meteorological stations. In this task, Supervised Machine Learning (SML)

procedures are proposed as effective tools for sky classification, based on accessible meteorological indices [10] such as decision trees (DTs) [11], Support Vector Machines (SVMs) [12], and Artificial Neural Networks (ANNs) [13–15].

Over recent years, interest has been expressed in calibrated sky luminance maps for sky classification and cloud detection [16–19]. A digital camera equipped with a fisheye lens can map at a higher resolution than commercial sky scanners and High Dynamic Range (HDR) images can capture the full sky luminance range [20].

There are also novel image-processing methods that can help to overcome misclassification due to cloud cover. While some studies have had their focus placed on color space, the focus of others has been on the modification and combination of the original monochromatic channels, known as the spectral features. A third alternative, texture filters, adjusts the gray pixel image patterns [21].

The RGB (red, green, and blue) chromaticity color model, a basic standard for computer images, has spectral features that may be adapted to cloud detection (CD) [22]. Shorter sunlight spectrum wavelengths will scatter due to atmospheric particles, giving the sky background a blue appearance [23] where the chromaticity component is mainly blue rather than red. Clouds appear white due to the uniform scattering of visible-light wavelengths, indicating similar amounts of red and blue components. Other models successfully applied to CD include Removal Atmospheric Scattering (RAS) [24], Red-Blue Ratio (RBR) [21], Red-Blue Difference (RBD) [25], and Normalized Red-Blue Ratio (NRBR) [17].

Some strategies have been aimed at adapting the image to the color perception of the human eye. Hue Saturation Value (HSV) [17], Red-difference Chroma (YCbCr) [18], and Intensity Hue Saturation (HIS) [26], among other color spaces, have recently demonstrated their efficacy for CD.

In addition to color space and spectral features, texture procedures use the gray distribution of pixels and their spatial neighborhood to identify objects and regions. These procedures have been shown to be very effective for cloud detection [27], medical images classification [28], and traffic analysis [29]. Gray Level Cooccurrence Matrix (GLCM), Local Range (LR), local Standard Deviation (STD), and local Entropy Matrix (EM) are texture filter procedures that statistically process the textures of images for their classification.

Image processing based on spectral, texture, and color spaces offers various perspectives of the same image. Their combination for image analysis can produce successful applications such as mapping [30] and aerial photographic classification [31]. In this paper, the recently proposed alternatives to the RGB color model are reviewed and compared for the improvement of image-processing methods applied to cloud detection and sky classification using Artificial Neural Network (ANN) algorithms. In some cases, preliminary image processing significantly improved the accuracy of the ANN used to classify the same image dataset. The methods that reduce misclassification will be identified from a detailed study, in which both the CIE standard sky

classification (15 types) and the reduced classification of three categories (clear, partial, and overcast sky conditions) were all considered.

The paper will be structured as follows. A complete comparison between several image-processing methods for CIE standard sky classification though ANNs will be presented in Section 2. In Section 3, the acquisition and processing of the experimental data will be described. In Section 4, the fit of the results of the ANN models with actual sky conditions will be verified. The results of the classification algorithms will be discussed in Section 5 and, finally, succinct conclusions on the most efficient image-processing methods will be presented in Section 6.

2. Review of Image-Processing Methods for Cloud Detection

Table 1 summarizes the main characteristics of twenty-two pixel image-processing methods that were reviewed and tested in this study and classified in terms of color space, spectral, and texture features. A complete description of all the image-processing methods will be completed in this section.

2.1. Color Spaces. The RGB color space uses one channel for each of the primary colors: blue, red, and green. Implemented directly in machine learning or with previous processing, this color space will yield spectral features. The primary colors, subchannels R, G, and B, build up a monochromatic image. A grayscale (GS) image is created when only pixel intensity is recorded. As previously mentioned, the HSV space is modelled on visual human perception, which classifies objects in terms of their luminous intensity (brightness or value) and chromaticity. The chromaticity has two independent parameters, hue and saturation. Hue is the pure color that varies from red to magenta (listed as red, yellow, green, cyan, blue, and magenta). The saturation describes the dilution of a pure color in white (0 = white; 1 = pure color). The hue, saturation, and value channels can also be independently used. Clouds are mostly perceived on a grayscale, due to interactions between sunlight and the atmosphere, so different cloud cover can be analyzed through the saturation channel. This color space has proved itself to be highly effective for sky classification into three categories: blue sky, cloudy sky, and sunset sky [35].

2.2. Spectral Features Based on the RGB Model. Unlike the direct implementation of the RGB model, a spectral feature describes the change of tone and color in an image. Its capability of detecting dark clouds from high and transparent cirrus clouds has been demonstrated [25]. The RAS channel was proposed to distinguish atmospheric scatter from atmospheric background light [24]. The RAS channel is obtained from a linear combination of the panchromatic channel (Y), the bright channel (L), and the dark channel (D), defined in Table 1. Channels Y, L, and D can also be independently applied.

TABLE 1: Summary of pixel image-processing methods.

Type	Name	Purpose	Formulation	Ref.
Color space	RGB	Image visualization based on primary colors	R: red channel G: green channel B: blue channel	[31]
	Subchannels R, G, and B		R: red channel; G: green channel; B: blue channel	
	GS	Grayscale intensity image (calculated from the RGB image)	rgb2gray ¹	[32]
	HSV	Image visualization based on the perceptions of the human eye	H: hue channel S: saturation channel V: value channel	[16]
	Subchannels H, S, and V		H: hue channel; S: saturation channel; V: value channel	
Spectral feature	RAS	Image processing for removing atmospheric scattering	$RAS = Y - (L - D)$ $L = \max(R, G, B)$ bright $D = \min(R, G, B)$ dark of the channel $Y = 0.299 \cdot R + 0.587 \cdot G + 0.114 \cdot B$	[21, 24]
	Subchannels Y, L, and D		Y: panchromatic channel; L: bright channel; D: dark channel	
	RBR	Background due to atmospheric scattering (Red-Blue Ratio).	$RBR = R/B$	[24]
	RBD	Difference between red and blue channels	$RBD = R - B$	[33]
	NRBR	Blueness of the sky; high robustness to noise (Normalized Red-Blue Ratio)	$NRBR = R - B/B + R$	[33]
	C1	Combination of RBR, RBD, and NRBR channels	$RBR - RBD - NRBR$	[17]
	ARGD	Reducing sunlight interference (Adjusted Red Green Difference)	$ARGD = k \cdot R - G$ $k = 1.7$ (weight of the red channel)	[22]
C2	Combination of RBR, ARGD, and NRBR channels	$RBR - RBD - ARGD$	[22]	
Texture feature	LR	For distinguishing edges and contours (Local Range)	Rangefilt ¹	
	STD	Obtained from the standard deviation in each neighborhood	Stdfilt ¹	[34]
	EM	Randomness of the image (Entropy Matrix)	Entropyfilt ¹	

¹MATLAB function ([34]).

Different combinations of red and blue channels were proposed for cloud detection. The aim of the Red-Blue Ratio (RBR), which yields small ratios for blue skies and large ratios for clouds, is to recognize thin and opaque cloud cover and clear skies [36]. Heinle et al. [25] noted several problems related to the use of the RBR channel for detecting thick clouds and difficulties with circumsolar pixels. They therefore proposed the RBD (Red-Blue Difference) channel as an alternative. Yamashita et al. [37] performed a full revision of the blue and red channel and implemented the sky index or NRBR (Normalized Red-Blue Ratio) for separating the blue sky and clouds area. These adaptations of the RGB channels have been successfully contrasted for CD. The green channel is however often overlooked in image processing. The Adjusted Red Green Difference (ARGD) [22] was introduced to correct any possible saturation of the blue component. Linear combinations of the spectral features have been proposed in other works, such as C1 [17] and C2 [22] that are listed in Table 1.

2.3. *Texture Filters.* Texture filters use the gray pixel distribution (grayscale, from 0 to 255, GS matrix) and their

spatial neighborhood to identify objects and regions. Texture filters divide the GS matrix into local neighbors, applying a mathematical operator: range for Local Range (LR), the Entropy Matrix (EM), and the local Standard Deviation for STD image processing [34]. Figure 1 shows an example of an LR texture-filtering process. GS is a monochromatic matrix whose elements are $M_{i,j}$. The size of the GS matrix, defined by its neighbors, is represented in Figure 1 as the 9×9 blue square. Its size is smaller throughout the GS boundary (elements represented as $m_{i,j}$). The filter function applies a mathematical operator in this neighborhood and the result is included in the position (i, j) of the new matrix.

2.3.1. *Local Range Texture Filter.* The purpose of LR filtering is to make the edges and contours of an image visible. The highest value is subtracted from the smallest one within the 9×9 neighborhood, as shown in Figure 1. The function saves the result in the LR matrix.

2.3.2. *EM Texture Filter.* Entropy is a measure of the image texture randomness. The Entropy Matrix (EM) calculates the local entropy of all the GS neighborhoods [34]. The EM value

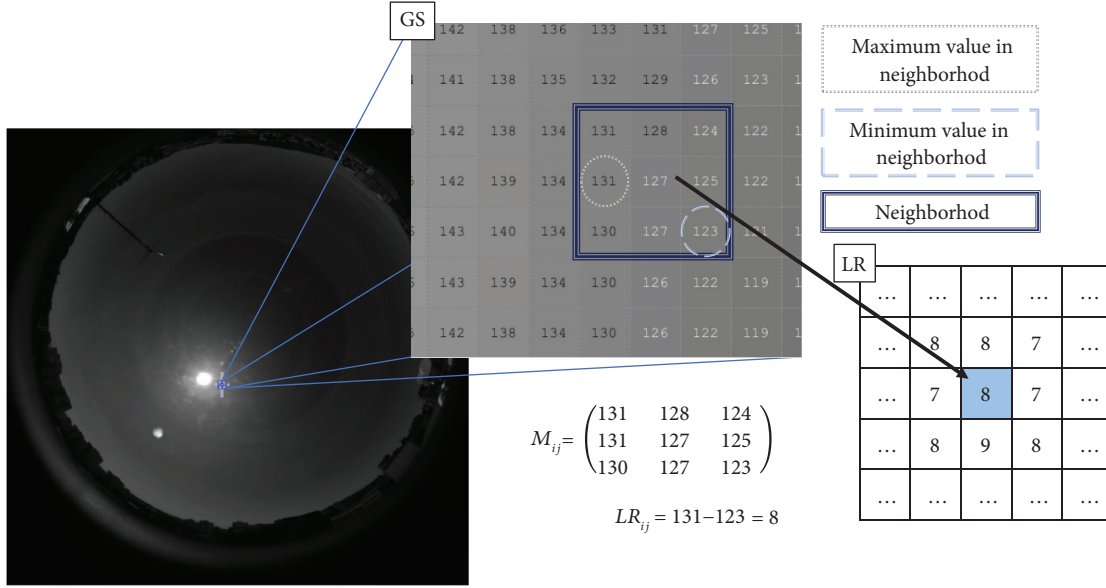


FIGURE 1: Example of LR calculation in the 9×9 neighborhood (blue area) for the element $M(i, j) = 127$.

that is directly proportional to the degree of variation of a pixel with respect to its neighbors is calculated with

$$EM_{i,j} = - \sum_{k=1}^N p_k \log_2 p_k. \quad (1)$$

In Figure 2, an image histogram with high variations is shown. p_k reflects the occurrence for the gray level p element; N is the total number of gray levels in the neighborhood.

2.3.3. Local Standard Deviation (STD) Texture Filter. The following equation is used to calculate the local Standard Deviation (STD) within each neighborhood:

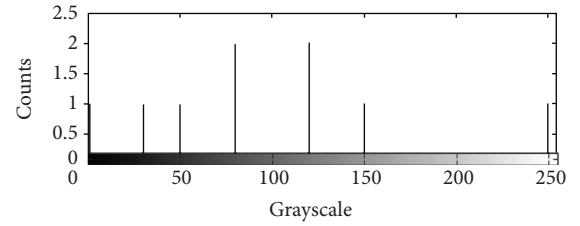


FIGURE 2: Example of the histogram for local entropy calculations.

$$M_{i,j} = \begin{pmatrix} 1 & 50 & 80 \\ 120 & 120 & 80 \\ 250 & 150 & 30 \end{pmatrix}.$$

$$STD_{i,j} = \sqrt{\frac{\sum_{k=1,0,-1}^3 \sum_{l=1,0,-1}^3 [M(i+k, j+l) - (1/N) (\sum_{k=1,0,-1}^3 \sum_{l=1,0,-1}^3 M(i+k, j+l))]^2}{N-1}}. \quad (2)$$

($N = 9$) is the number of elements in the neighborhood; k, l varies from -1 to 1 to cover the neighborhood matrix.

3. Experimental Data Acquisition and Processing

As previously stated, the main objective of this work is the analysis of image-processing algorithms for CIE standard sky classification using ANN-processed sky images. The workflow is described in Figure 3 and explained in the following sections.

3.1. Experimental Data Acquisition. The experimental data used in this work were recorded at a meteorological weather station located on the roof of the Higher Polytechnic School

building at Burgos University ($42^\circ 21' 04'' N$; $3^\circ 41' 20'' O$; 856 m above mean sea level). A complete description of the meteorological facility may be found elsewhere [1, 10, 38]. The experimental equipment is shown in Figure 4. The sky luminance distribution for characterization of sky conditions according to the CIE Standard General Sky classification was measured with a commercial MS-321LR sky scanner (EKO Instruments Europe B. V. Den Haag, The Netherlands). The sky scanner was adjusted on a monthly basis for taking measurements from sunrise to the sunset. It completed a full scan in four minutes and started a new scan every 10 minutes. The first and last measurements of the day ($\alpha_s \leq 5^\circ$) were discarded, as measurements were higher than 50 kcd/m^2 and lower than 0.1 kcd/m^2 , following the recommended specifications of the sky scanner equipment. The

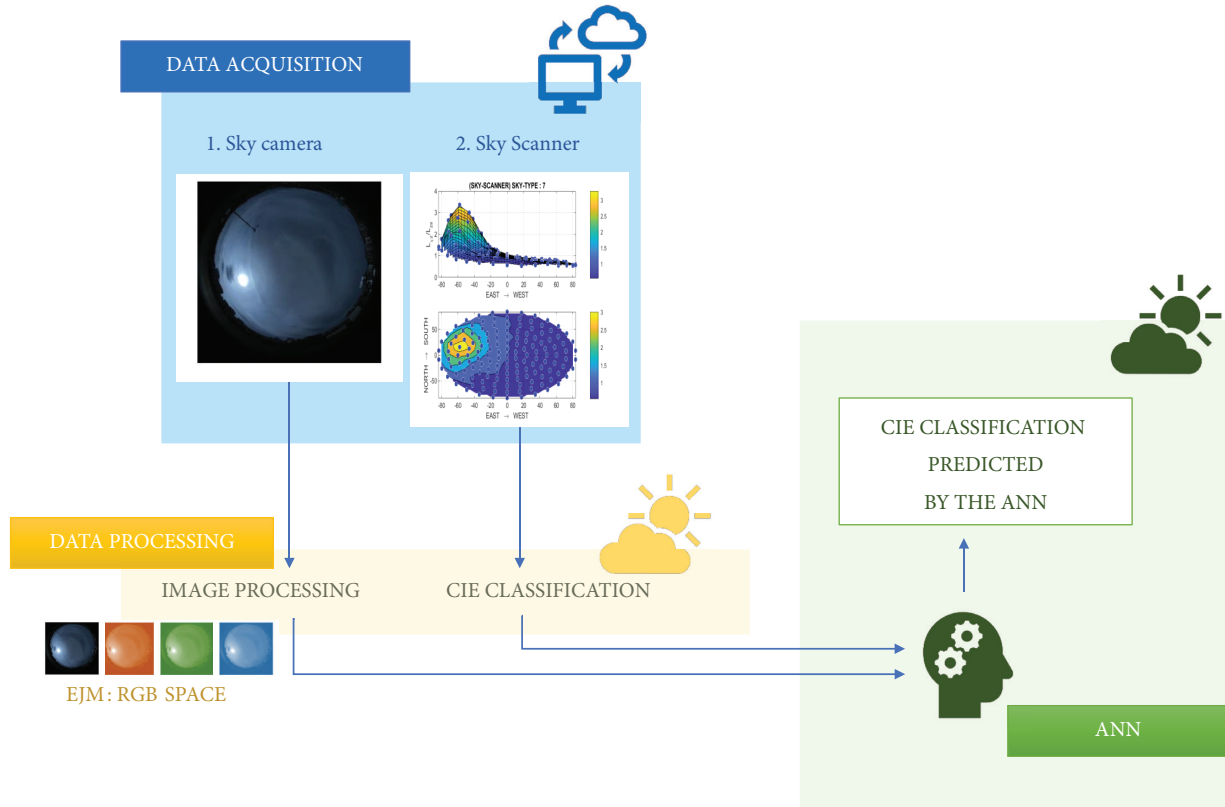


FIGURE 3: Workflow for image processing. Example with RGB space.



FIGURE 4: Location of the experimental equipment on the roof of the Higher Polytechnic School building at the University of Burgos with the SONA 201-D sky camera (top) and EKO MS-321LR (bottom) (source: Google Earth).

sky images were recorded every 15 seconds by a commercial SONA 201-D all sky camera day (Sieltec Canarias S L, Spain). The trigger frequency of the camera is one second and the image resolution is 1158×1172 pixels, recorded with the RGB color model (each pixel uses 8 bits, which have integer values from 0 to 255). Tables 2 and 3 show the technical specifications of the sky scanner and sky camera, respectively.

The experimental campaign took place between 1 November 2016 and 31 March 2020. The Normalized Luminance method (NL) proposed by Tregenza in 2004 [39], detailed in a previous paper [38], was used to determine the CIE standard sky types over Burgos during the experimental campaign. A total of 1,500 images were selected from the

TABLE 2: Sky scanner technical specifications.

Model	MS-321LR sky scanner
FOV	11°
Luminance	0 to 50 kcd/m ²
Radiance	0 to 300 W/m ²
A/D convertor	16 bits
Calibration error	2%

TABLE 3: Sky camera technical specifications.

Model	SONA 201-D
Sensor	CMOS-2.3 MP
Vision angle	<180° (fisheye lens)
Operating temperature	-40°C to 55°C
Image format	RAW

experimental dataset (more than 80,000 sky images), 100 from each CIE sky category, which were characterized by greater concordance with the CIE pattern for that category. The experimental dataset was therefore composed of one hundred sky images catalogued as CIE standard sky categories. This sky classification was used as a reference for the sky conditions.

3.2. Data Processing. The original sky images were processed using the twenty-two different channels chosen for the study and summarized in Table 1. While some of them were

directly generated from the sky images, others had to be generated through complementary channels. In Figure 5, the results of the image-processing methods applied to images of sky conditions are classified as clear, partial, and overcast, following the CIE taxonomy. As can be observed, each filter highlights different features of the images. The circumsolar area and the nearest horizon zone present the greatest difficulties for cloud detection. In Figure 5, it can be seen that the RGB image is sensitive to the circumsolar region and is capable of detecting the solar corona. However, in the RGB image, no differences can be appreciated in dark-homogeneous sky conditions. The appearance of direct day beam can be a source of errors. Although the blue channel saturated the circumsolar region, both the red and the green channels showed greater sensitivity at detecting cloudy areas. In contrast, the horizon was captured by the Y, D, L, RAS, V, STD, and EM methods, and LR mainly defined the contours. Unlike most of the other channels, the RGB model had difficulty with the directional homogeneity of the images for the detection of overcast sky conditions. The family of RAS methods (RAS, Y, D, and L) appeared to show similar levels of accuracy under all sky conditions, their main differences being near the circumsolar area.

3.3. Image Compression. The high resolution of the original sky images (1158×1172 pixels) requires their compression to reduce the dimension of the dataset, improving data storage and subsequent image processing. In this study, the original sky images were compressed to 110×110 pixels in each channel. Figure 6 shows the result of the image compression procedure to 0.89%, which facilitates ANN tuning with no loss of efficiency.

4. ANN for CIE Standard Sky Classification

Artificial Neural Networks (ANNs) are frequently used in meteorology science: CIE and cloud classification [40, 41], solar irradiance and wind speed forecasting [42–47], atmospheric pollution distribution [48, 49], and rainfall [50, 51]. ANN classification models serve to classify input information into certain categories or targets. A Supervised Machine Learning (SML) neural network is required for CIE standard sky classification where the sky types are previously known. The model works efficiently when the prediction matches the target. Modelled on the biological concept of neurons, ANN is a very powerful technique for classification problems. Figure 7 shows a conventional ANN structure, which consists of an input layer, a set of several hidden layers, and an output layer.

The information from the neurons of the input layer (X_i^0) crosses the hidden layers (one in this work), following unidirectional connections, to the output layer that has one neuron (X_i^2) per target. Each processing center or neuron is adjusted to the other neurons through an interactive process, using (3). The Scaled Conjugate Gradient method (SCG) [52] was used to fit the weights (weighting matrix, W^n) for each iteration.

$$X_i^n = W^n X^{n-1} + B, \quad (3)$$

where W^n is the weighting matrix, X^{n-1} are the input variables, and B is the bias. The neuron generates the output, X_i^n , through the activation function, $f(X_i^n)$, given by the hyperbolic tangent sigmoid transfer function in this study, as shown in [13]

$$X_i^n = f(X_i^n) = \frac{2}{1 + \exp(-2 \cdot X_i^n)} - 1. \quad (4)$$

Supervised Machine Learning requires three datasets: training, validation, and test datasets. The training group is used to determine the weighted matrix and the bias in an iterative process. The training is over when the results of the performance of the resulting model, calculated using the validation set, reach the desired quality. The test data group is used to calculate the performance of the model. Random dataset division is crucial to achieve a reliable performance. A conventional training dataset is randomly selected and consists of 70% of the total data, while the validation set and the test set each represent 15%, respectively.

The design of the ANN is adapted to the database and the process is simulated. There is no standardized procedure for establishing the most effective number of neurons and hidden layers [42], so experimentation or tuning is needed. In this study, several trials were performed in which the number of neurons (1–100) was varied, searching for the best accuracy, Acc, of the ANN, given by

$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \quad (5)$$

where TP and TN are the correct predictions of the ANN (true positives and true negatives) and FP and FN are the incorrect predictions (false positives and false negatives). Accuracy is rated by the number of correct predictions over the total number of predictions. The neural network structure (number of neurons in the hidden layer) was selected on the basis of highest accuracy. After several trials, the number of hidden layers was fixed at one.

5. Results

In Figure 8, the improved accuracy of the ANN models that used the sky images as their input is shown. Each image had previously been processed by each of the twenty-one image-processing methods summarized in Table 1, with respect to the RGB space, defined as $\Delta(\text{Acc})$ and shown in

$$\Delta(\text{Acc}) = \frac{(\text{Acc}(\text{channel } x) - \text{Acc}(\text{RGB space}))}{\text{Acc}(\text{RGB space})} \cdot 100, \quad (6)$$

where $\text{Acc}(\text{channel } x)$ and $\text{Acc}(\text{RGB space})$ are the accuracy obtained when the input of the ANN is the set of sky images processed by each method x ($x =$ each image-processing method summarized in Table 1) and RGB space, respectively. The accuracy of each ANN and the number of neurons in its hidden layer are shown in Table 4.

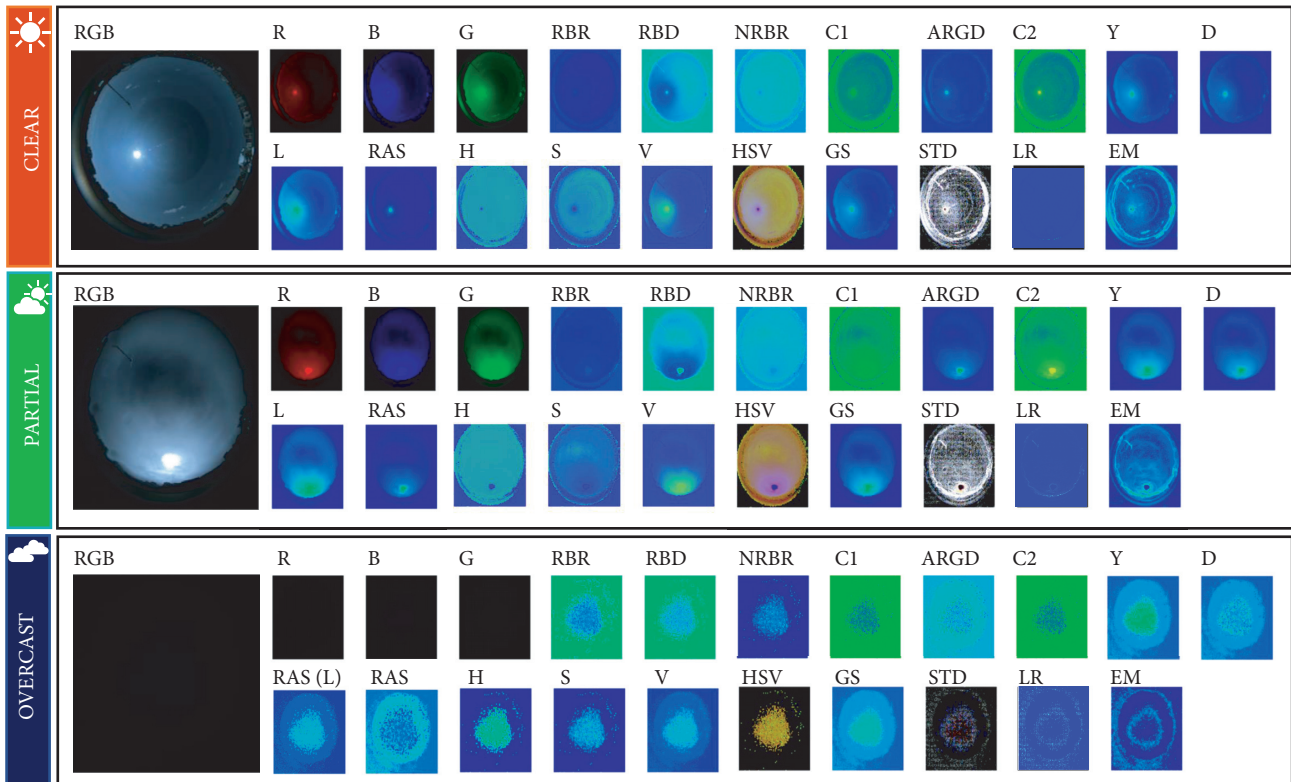


FIGURE 5: Results of the image-processing methods applied to clear, partial, and overcast CIE standard sky types.

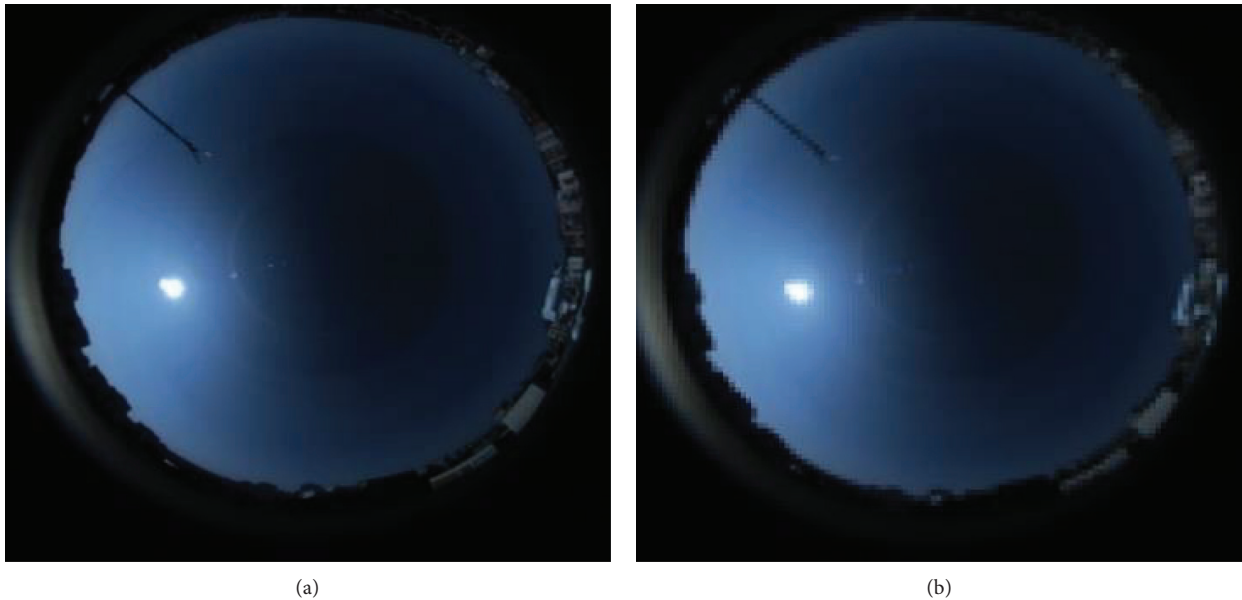


FIGURE 6: Result of the image compression procedure. (a) Actual sky image. (b) Compressed sky image.

As can be seen in Figure 8, HSV is better color space than RGB for CIE standard sky classification using images, with a small improvement in the accuracy (0.66%) with respect to RGB image processing. The GS color space and the RGB space were equally accurate. The use of the R, G, and B monochromatic channels also

improved the accuracy of the ANN for CIE standard sky classification, the G channel being the most suitable for this task. The accuracy of the ANN fitted using the individual channels, H and S, worsened over the RGB color space, while the V channel significantly improved ANN accuracy.

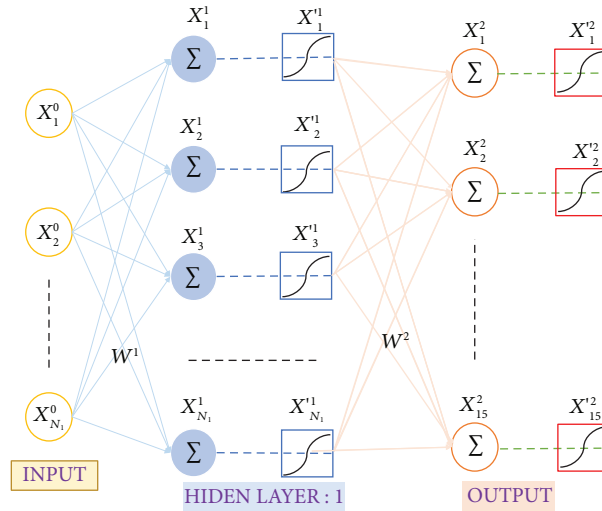


FIGURE 7: ANN system architecture.

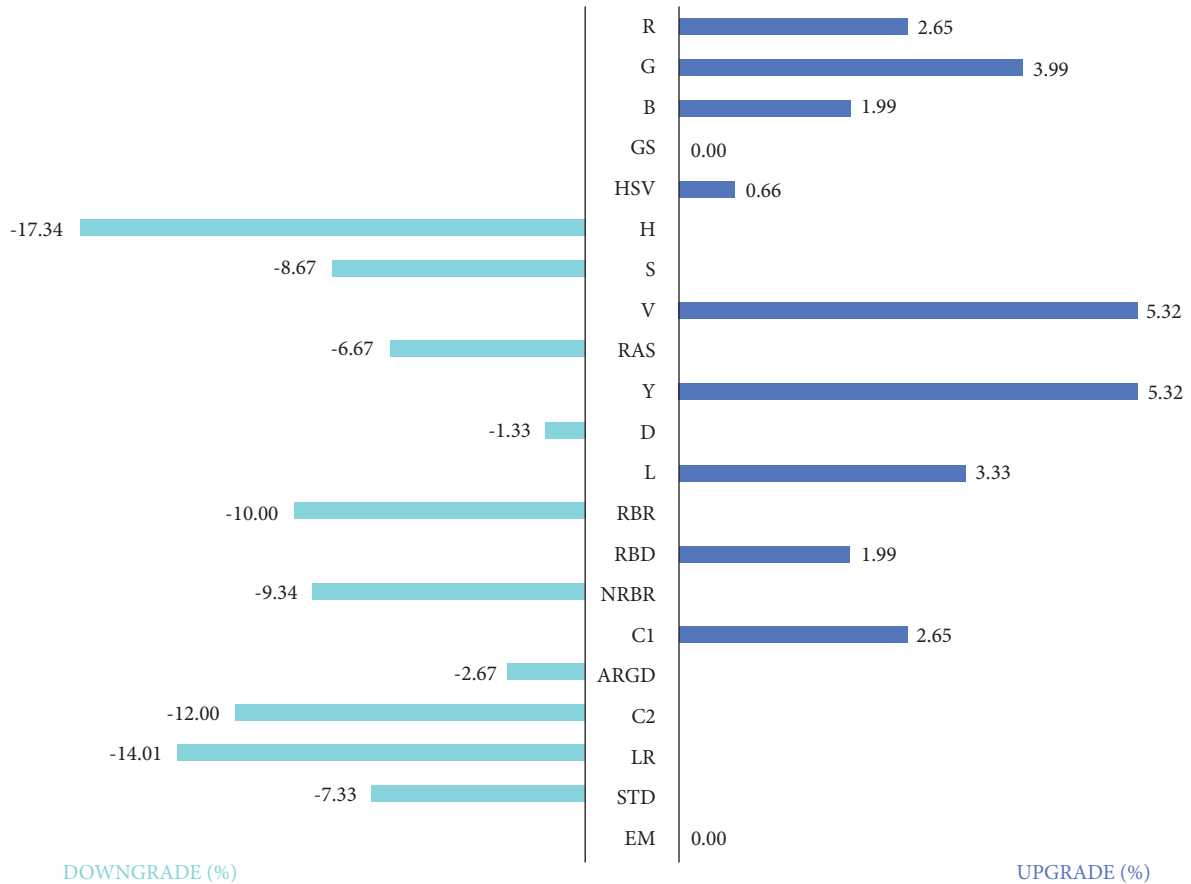


FIGURE 8: Improvement in ANN accuracy, $\Delta(\text{Acc})$, for CIE standard sky classification, using the results of each image-processing method as input, as summarized in Table 1, over ANN accuracy obtained with the original RGB images as input.

In the spectral feature category, the RAS processing method worsened the sky classification accuracy of the ANN. However, channels Y and L showed better behavior for sky classification, although they used more neurons in the hidden layer. Among the rest of spectral feature

channels, only RBD and C1 significantly improved ANN accuracy. With regard to the texture filters, EM showed little or no advantages over the use of the RGB color space and the other two filters, LR and STD, impaired the accuracy of the resulting neural network. The number of neurons in the

TABLE 4: Accuracy and structure (number of neurons in the hidden layer) of the selected ANN for each image-processing method.

Method	Acc (%)	Number of neurons
RGB	66.67	84
R	68.44	66
G	69.33	58
B	68	36
GS	66.67	58
HSV	67.11	40
H	55.11	23
S	60.89	27
V	70.22	58
RAS	62.22	25
Y	70.22	95
D	65.78	21
L	68.89	74
RBR	60	45
RBD	68	98
NRBR	60.44	27
C1	68.44	73
ARGD	64.89	17
C2	58.67	68
LR	57.33	39
STD	61.78	52
EM	66.67	58

hidden layer, shown in Table 4, never increased the accuracy of the ANN, as can be seen from the use of image-processing methods Y and V.

Figure 8 shows the results of each ANN classifying the skies into the fifteen CIE standard sky categories. A simpler classification into three categories (clear, overcast, and partial conditions) is often sufficient for many applications, such as luminous efficacy calculations [53] and lighting design in buildings [54]. The fitted results of the ANN sky classification for three categories are shown in Figure 9 and Table 5.

For CIE standard sky classification into three sky categories, lower differences in accuracy can be seen and only the G, the B, and the GS monochromatic channels and the spectral features L and C1 improved ANN accuracy. In all these classification cases, the number of neurons in the hidden layer was lower.

The accuracy index was used to group the goodness of fit of the ANN in all categories, although the fitted quality in each individual category was not processed. A confusion matrix analysis is shown in Figures 10–13. In a confusion matrix, when the Supervised Machine Learning algorithm prediction and the target match each other (TP or TN result), the corresponding diagonal boxes of the matrix are colored. When there are no matches between the prediction and the target value (FP and FN), the other boxes of the confusion matrix are filled in. The best image-processing method will have the highest number of colored boxes around the diagonal line of the matrix. The figures below represent the confusion matrices corresponding to the 15 types of CIE standard skies.

Figure 10 shows the confusion matrix of the ANN-calculated RGB-CIE sky classification for the test set (15% of

the total dataset). It can be seen that the RGB-CIE classification with machine learning misclassified cloudy and partial skies: few matches are visible in the boxes along the diagonal line. In Figure 10, the CIE standard sky classification into three categories (clear, partial, and overcast sky conditions) is also presented. Those cases classified outside the corresponding category were designated as critical, i.e., clear skies classified as either partial or overcast or vice versa.

The same information is shown in Figure 11 for the color space CIE standard sky classification, corresponding to the other color space processing methods under analysis. The red, the green, and the blue channels showed a similar behavior to the RGB color space. The red channel adequately classified CIE standard sky types 7 to 15, in other words, all clear skies and some partial sky types. The HSV color space showed a similar performance in all categories, in contrast to the RGB color space, in which the classification of clear sky types may be highlighted. Hue and saturation channels introduced too much noise, but the value channel showed good performance.

In Figure 12, the confusion matrices are shown for the spectral feature image-processing methods-CIE standard sky classification. The RBR and NRBR spectral features introduced noise, but the resultant combination, C1, reduced misclassification, improving the traditional RGB color space. It therefore appears to be an adequate alternative image-processing method for CIE standard sky classification using sky camera images. The RAS channel theoretically removed atmospheric scattering, but the confusion matrix never reflected a better performance than the RGB color space. The confusion matrix has demonstrated that it cannot distinguish the CIE sky types 1, 3, and 5. The RAS method also introduced too much noise in cloudy-to-partial sky types.

Finally, the confusion matrices are shown in Figure 13 for texture filter processing methods-CIE standard sky classification with ANN. As can be seen, all texture channels performed well, especially the EM channel, while LR largely failed for CIE standard sky classifications partial and overcast.

A detailed study for the CIE standard sky classification into three categories is presented in Figure 14, where the confusion matrices presented in Figures 10–13 were divided into four submatrices: overcast (CIE standard sky types 1 to 5), partial (CIE standard sky types 6 to 10), clear (CIE standard sky types 1 to 5), and critical that refers to cases classified out of category. The red line indicates the RGB result, taken as a baseline for accuracy improvements, $\Delta(\text{Acc})$. Some of the image-processing methods for classifying certain sky categories are highlighted in Figure 14. RBD, D, and B showed the best performance for the detection of overcast skies, increasing the performance of each respective ANN. G, S, and GS achieved better results for the detection of partial skies and clear skies were also in the same category in which the conventional RGB color space achieved its best performance.

Some channels highlighted certain sky types but drastically failed to classify other types. The blue channel saturated in clear skies, to such a point that its performance was

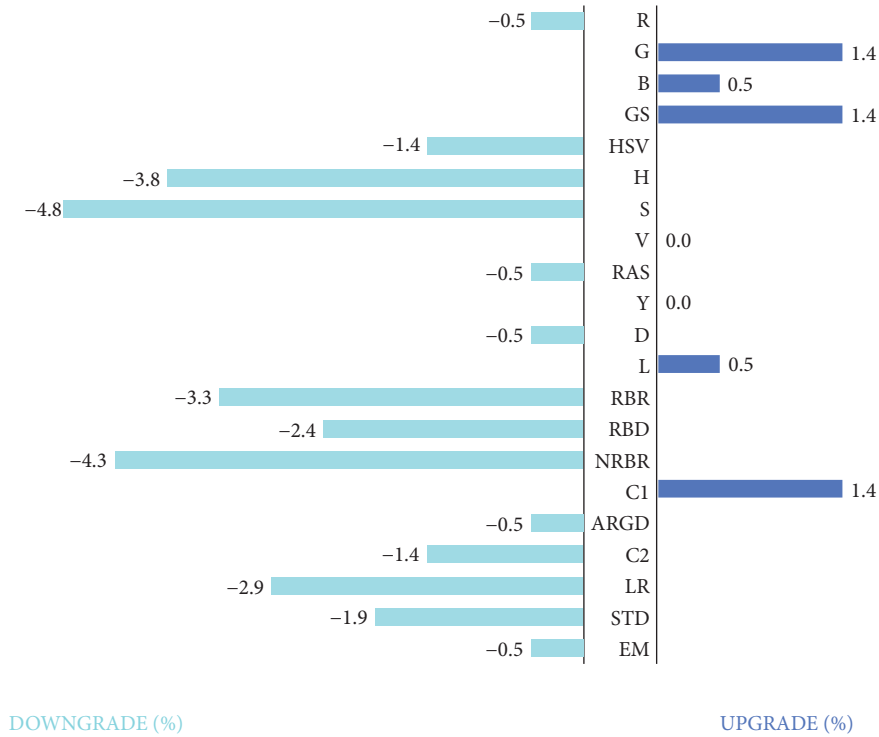


FIGURE 9: Improvement in ANN accuracy, Δ (Acc), for CIE standard sky classification in three sky categories: overcast, partial, and clear conditions, using as input the image processed by each image-processing method summarized in Table 1, over ANN accuracy obtained with the RGB images as input.

TABLE 5: Accuracy and structure (number of neurons in the hidden layer) of the ANN selected for each image-processing method.

Method	Acc (%)	Number of neurons
RGB	93.33	71
R	92.89	48
G	94.67	52
B	93.78	52
GS	94.67	52
HSV	92	11
H	89.78	23
S	88.89	73
V	93.33	15
RAS	92.89	90
Y	93.78	21
D	92.89	15
L	93.78	15
RBR	90.22	23
RBD	91.11	43
NRBR	89.33	23
C1	94.67	18
ARGD	92.89	19
C2	92	41
LR	90.67	81
STD	91.56	7
EM	92.89	68

almost the worst for clear skies detection. This behavior was also noted for the D channel.

Unfortunately, no image preprocessing method drastically improved the RGB classification in the three

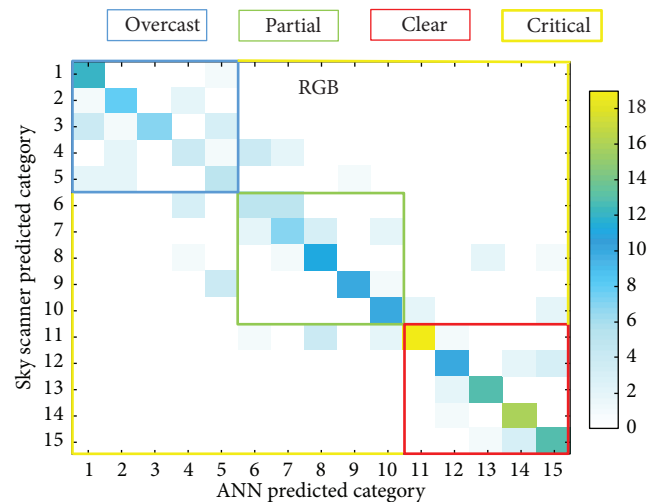


FIGURE 10: Confusion matrix for ANN-calculated RGB-CIE sky classification into fifteen categories and into three categories: overcast (CIE standard sky types 1 to 5), partial (CIE standard sky types 6 to 10), and clear (CIE standard sky types 1 to 5). Critical refers to cases classified out of category. The color scale shows the number of coincidences in each category.

subcategories (clear, partial, and overcast conditions). However, Y, green, red, RBD, V, and EM processing methods were prominent in one or two categories and their results were acceptable in all other categories, as shown in Table 6.

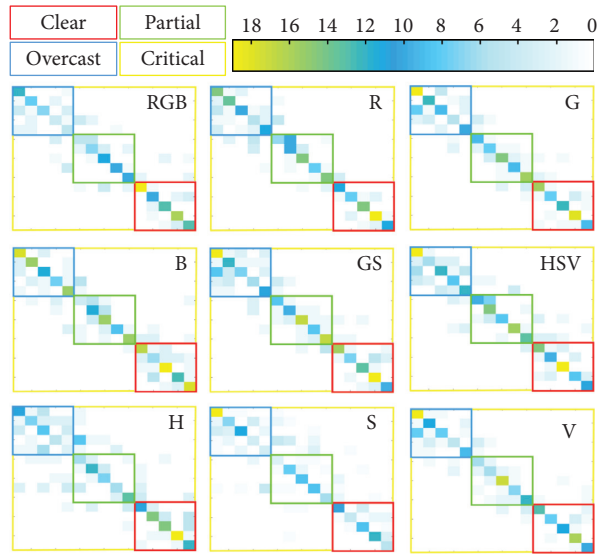


FIGURE 11: Confusion matrices color space image-processing methods-CIE standard sky classification.

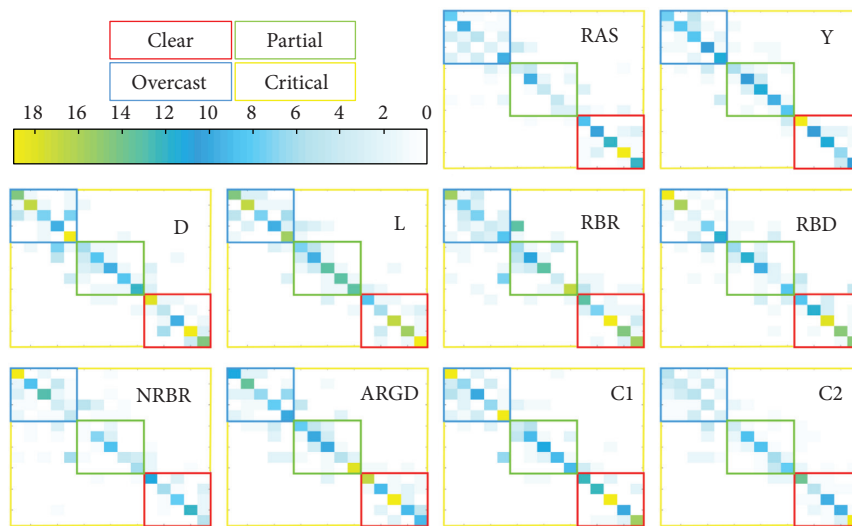


FIGURE 12: Confusion matrices spectral feature image-processing methods-CIE standard sky classification.

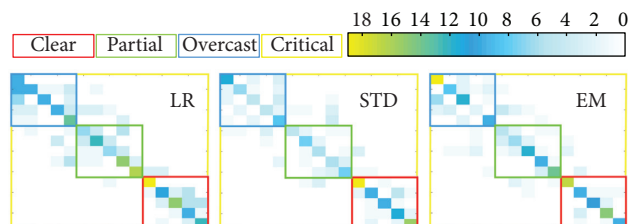


FIGURE 13: Confusion matrices for texture filter image-processing methods-CIE standard sky classification.

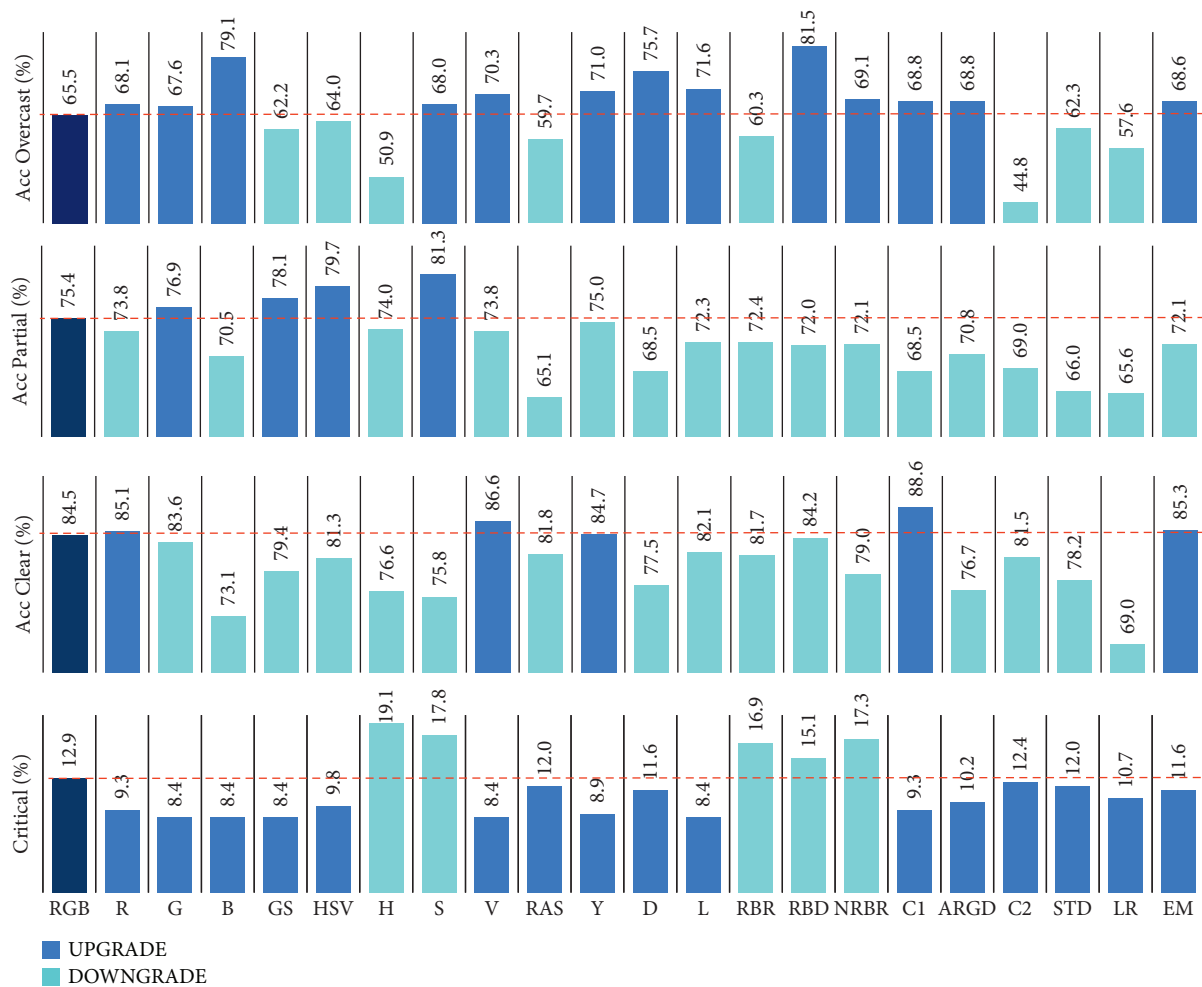


FIGURE 14: Acc for the CIE standard sky classification through sky images and ANN, using the different image-processing methods recorded in Table 1.

TABLE 6: Summary of the results.

CIE categories	One category	Two categories	Three categories
Clear	R, CI, Y, EM, and V		
Partial	G, S, HSV, and GS		
Overcast	R, G, B, RBD, NRBR, C1, ARGD, Y, D, L, S, V, and EM	R, G, CI, Y, S, V, and EM	R, G, Y, and V, EM
Critical	R, G, B, CI, C2, Y, D, L, RAS, V, HSV, GS, STD, LR, and EM		

Almost all the image-processing methods reduced critical mistakes or misclassification, which should as far as possible be avoided. Following this criterion, RBR, RBD, NRBR, ARGD, H, and S were discarded as preprocessing image methods for ANN sky classification of sky images.

6. Conclusions

Sky classification and cloud detection from sky images and machine learning can be largely improved through preliminary image processing, reducing errors in classification and simplifying algorithms. In this study, 22 sky image-processing methods have been reviewed, including the three most common categories, color spaces, spectral features, and texture filters. The CIE standard sky classification has been

selected to determine the characteristics of the sky, as it is recognized as representative of the atmospheric conditions. A very extensive unbiased dataset has been used, including 1,500 sky images and their corresponding CIE classification, calculated through the Normalized Luminance method from sky luminance distribution data. The Artificial Neural Network (ANN) was the selected machine learning algorithm.

As a first conclusion, digital cameras equipped with fisheye lens can be used as alternatives to sky scanner devices for ANN-assisted CIE standard sky classification. The accuracy of the classification algorithm can be improved with adequate preliminary image processing that highlights the sky image information and optimizes the algorithmic structure.

HSV was a better color space than RGB, as were the monochromatic channels R, G, and B, for classifying the skies on the basis of the images into the fifteen CIE standard sky types. Only the V individual channel of HSV worked better than both HSV and RGB. Spectral feature channels Y and L showed better behavior for sky classification than the RGB color space, but they used more neurons in the hidden layer. Among the rest of the spectral feature channels, only RBD and C1 significantly improved ANN accuracy. Texture filters added no significant advantages over the RGB color space.

For CIE standard sky classification as clear, partial, and overcast conditions, RGB appeared to be the best image-processing method and only the monochromatic channels G and B, GS, and the composed spectral feature C1 improved the accuracy of the RGB color space. No improvement in ANN performance was therefore noted with the use of extra channels.

In contrast to previous studies [14] which have their weakest accuracy in cloudy conditions, several channels have worked successfully, improving the accuracy of the machine learning algorithm by 10% over the RGB color space for cloudy skies. These channels were B, R, S, V, ARGD, RBD, C1, C2, Y, STD, and EM.

RGB and its primary channels, R, G, and B, were not good enough for dark cloudy conditions, due to image-processing information losses. While traditional cloud detection has usually omitted the G channel, both the G and the B channels have been shown to be equally effective. In contrast, the B channel tended to saturate on clear sky conditions.

The confusion matrices highlighted that the ANN failed to distinguish CIE sky types 1, 3, and 5.

The main conclusion is that the use of a specific image-processing method could improve the accuracy of an ANN algorithm, depending on the information required from the image for the classification problem. Future work will focus on the classification of skies according to the CIE standard using neural networks specifically designed for the classification of images such as convolutional neural networks.

Data Availability

The neural network database used to support the findings of this study has been deposited in the Institutional Repository of University of Burgos (<https://riubu.ubu.es/>).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

D. Granados-López participated in investigation, methodology, software, and visualization; A. García-Rodríguez and S. García-Rodríguez took part in investigation, methodology, formal analysis, and validation; A. Suárez-García contributed to methodology and software; M. Díez-Mediavilla was responsible for conceptualization, original

draft preparation, and supervision as well as funding acquisition and project administration; C. Alonso Tristán played a role in conceptualization, original draft preparation, and supervision (writing, reviewing, and editing).

Acknowledgments

The authors gratefully acknowledge the financial support provided by the Regional Government of Castilla y León under the “Support Program for Recognized Research Groups of Public Universities of Castilla y León” (BU021G19) and the Spanish Ministry 595 of Science and Innovation under the I+D+i state program “Challenges Research Projects” (Ref. RTI2018-098900-B-I00). Diego Granados López expresses his thanks for economic support from the Junta de Castilla-León (PIRTU Program, ORDEN EDU/556/2019).

References

- [1] A. Suárez-García, M. Díez-Mediavilla, D. Granados-López, A. González-Peña, and C. Alonso-Tristán, “Benchmarking of meteorological indices for sky cloudiness classification,” *Solar Energy*, vol. 195, pp. 499–513, 2020.
- [2] R. Kittler, S. Darula, and R. Perez, *A Set of Standard Skies Characterizing Daylight Conditions for Computer and Energy Conscious Design, Final Report*, Slovak Academy of Science, Slovakia, 1998.
- [3] Y. Uetani, S. Aydinli, A. Joukoff, and J. D. Kendrick, *Spatial Distribution of Daylight—CIE Standard General Sky*, ISOCIE, Geneva, Vienna, 2004.
- [4] D. H. W. Li, S. Li, W. Chen, and S. Lou, “Analysis of point daylight factor (PDF) average daylight factor (ADF) and vertical daylight factor (VDF) under various unobstructed CIE standard skies,” *IOP Conference Series: Materials Science and Engineering*, vol. 556, no. 1, Article ID 12044, 2019.
- [5] D. H. W. Li, S. Lou, J. C. Lam, and R. H. T. Wu, “Determining solar irradiance on inclined planes from classified CIE (International Commission on Illumination) standard skies,” *Energy*, vol. 101, pp. 462–470, 2016.
- [6] H. Xiao, Q. Kang, J. Zhao, and Y.-s. Xiao, “A dynamic sky recognition method for use in energy efficient lighting design based on CIE standard general skies,” *Building and Environment*, vol. 45, no. 5, pp. 1319–1328, 2010.
- [7] I. García, M. de Blas, and J. L. Torres, “The sky characterization according to the CIE Standard General Sky: comparative analysis of three classification methods,” *Solar Energy*, vol. 196, pp. 468–483, 2020.
- [8] “Guide to recommended practice of daylight measurement, CIE publication 108, central bureau of the CIE, Vienna, Austria 1994, softbound, 54 pp., \$59 (members)/\$88 (non-members),” *Color Research and Application*, vol. 20, no. 1, p. 80, 1995.
- [9] D. H. W. Li and S. Lou, “Review of solar irradiance and daylight illuminance modeling and sky classification,” *Renewable Energy*, vol. 126, pp. 445–453, 2018.
- [10] D. Granados-López, A. Suárez-García, D. Díez-Mediavilla, and C. Alonso-Tristán, “Feature selection for CIE standard sky classification,” *Solar Energy*, vol. 218, pp. 95–107, 2021.
- [11] S. Lou, D. H. W. Li, and J. C. Lam, “CIE Standard Sky classification by accessible climatic indices,” *Renewable Energy*, vol. 113, pp. 347–356, 2017.

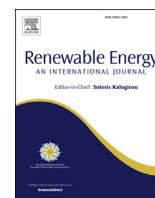
- [12] A. Taravat, F. del Frate, C. Cornaro, and S. Vergari, "Neural networks and support vector machine algorithms for automatic cloud classification of whole-sky ground-based images," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 3, pp. 666–670, 2014.
- [13] T. Kazanasmaz, M. Günaydin, and S. Binol, "Artificial neural networks to predict daylight illuminance in office buildings," *Building and Environment*, vol. 44, no. 8, pp. 1751–1757, 2009.
- [14] D. H. W. Li, H. L. Tang, E. W. M. Lee, and T. Muneer, "Classification of CIE standard skies using probabilistic neural networks," *International Journal of Climatology*, vol. 30, no. 2, pp. 305–315, 2010.
- [15] D. H. W. Li, H. L. Tang, S. L. Wong, E. K. W. Tsang, G. H. W. Cheung, and T. N. T. Lam, "Skies classification using artificial neural networks (ANN) techniques," in *Proceedings of the 6th International Conference on Indoor Air Quality, Ventilation and Energy Conservation in Buildings: Sustainable Built Environment*, IAQVEC, Sendai Japan, October 2007.
- [16] J. Alonso-Montesinos and F. J. Batlles, "The use of a sky camera for solar radiation estimation based on digital image processing," *Energy*, vol. 90, pp. 377–386, 2015.
- [17] R. Chauvin, J. Nou, S. Thil, A. Traoré, and S. Grieu, "Cloud detection methodology based on a sky-imaging system," *Energy Procedia*, vol. 69, pp. 1970–1980, 2015.
- [18] H.-Y. Cheng and C.-L. Lin, "Cloud detection in all-sky images via multi-scale neighborhood features and multiple supervised learning techniques," *Atmospheric Measurement Techniques*, vol. 10, no. 1, pp. 199–208, 2017.
- [19] M. B. Piderit, C. Cauwerts, and M. Diaz, "Definition of the CIE standard skies and application of high dynamic range imaging technique to characterize the spatial distribution of daylight in Chile," *Revista de la construcción*, vol. 13, no. 2, pp. 22–30, 2014.
- [20] D. Granados-López, I. Suárez-García, M. Díez-Mediavilla, and C. Alonso-Tristán, "Evaluation of two procedures for selecting the CIE standard sky type using high dynamic range images," in *Proceedings of the ISES Solar World Congress 2019, SWC 2019 and IEA SHC International Conference on Solar Heating and Cooling for Buildings and Industry 2019*, SHC, Santiago, Chile, January 2019.
- [21] X. Wan and J. Du, "Cloud classification for ground-based sky image using random forest," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLIII-B3-2020, pp. 835–842, 2020.
- [22] X. Li, Z. Lu, Q. Zhou, and Z. Xu, "A cloud detection algorithm with reduction of sunlight interference in ground-based sky images," *Atmosphere*, vol. 10, no. 11, p. 640, 2019.
- [23] P. Seuntjens, M. Van Boven, and D. Sekulovski, "Effect of skylight configuration and sky type on the daylight impression of a room," *WIT Transactions on Ecology and the Environment*, vol. 165, pp. 53–61, 2012.
- [24] J. Yang, Q. Min, W. Lu, Y. Ma, W. Yao, and T. Lu, "An RGB channel operation for removal of the difference of atmospheric scattering and its application on total sky cloud detection," *Atmospheric Measurement Techniques*, vol. 10, no. 3, pp. 1191–1201, 2017.
- [25] A. Heinle, A. Macke, and A. Srivastav, "Automatic cloud classification of whole sky images," *Atmospheric Measurement Techniques*, vol. 3, no. 3, pp. 557–567, 2010.
- [26] M. P. Souza-Echer, E. B. Pereira, L. S. Bins, and M. A. R. Andrade, "A simple method for the assessment of the cloud cover state in high-latitude regions by a ground-based digital camera," *Journal of Atmospheric and Oceanic Technology*, vol. 23, no. 3, pp. 437–447, 2006.
- [27] Q. Luo, Y. Meng, L. Liu, X. Zhao, and Z. Zhou, "Cloud classification of ground-based infrared images combining manifold and texture features," *Atmospheric Measurement Techniques*, vol. 11, no. 9, pp. 5351–5361, 2018.
- [28] P. K. Mall, P. K. Singh, and D. Yadav, "GLCM based feature extraction and medical x-ray image classification using machine learning techniques," in *Proceedings of the 2019 IEEE Conference on Information and Communication Technology*, IEEE, Allahabad, India, December 2019.
- [29] L. Wei and D. Hong-ying, "Real-time road congestion detection based on image texture analysis," *Procedia engineering*, vol. 137, pp. 196–201, 2016.
- [30] P. Maillard, "Spectral-textural image classification in a semiarid environment," in *Proceedings of the ISPRS Commission VII Symposium 'Remote Sensing: From Pixels to Processes'*, Enschede Citeseer, The Netherlands, May 2006.
- [31] S. Tuominen and A. Pekkarinen, "Performance of different spectral and textural aerial photograph features in multi-source forest inventory," *Remote Sensing of Environment*, vol. 94, no. 2, pp. 256–268, 2005.
- [32] X. Zhao and H. Zheng, "Cloud detection of gray photographs by cloud region detection and optimal region grow threshold," in *Proceedings of the 2nd International Conference on Intelligent Computing and Cognitive Informatics (ICICCI 2015)*, Singapore, 2015.
- [33] J. Yang, Q. Min, W. Lu et al., "An automated cloud detection method based on the green channel of total-sky visible images," *Atmospheric Measurement Techniques*, vol. 8, no. 11, pp. 4671–4679, 2015.
- [34] "Image processing toolbox for use with MATLAB, user's Guide version 5. 2004," *The MathWorks, Inc.* vol. 5, 2004, https://es.mathworks.com/help/pdf_doc/images/images_ref.pdf.
- [35] K. F. Hussain and H. A. Sayed, "Sky detection using K-HSV descriptor," *Journal of the Institute of Industrial Applications Engineers*, vol. 2, no. 1, pp. 1–5, 2014.
- [36] T. Koehler, R. Johnson, and J. Shields, "Status of the whole sky imager database," in *Proceedings of the Cloud Impacts on DOD Operations and Systems*, pp. 77–80, Department of Defense, El Segundo, CA, USA, 1991.
- [37] M. Yamashita, M. Yoshimura, and T. Nakashizuka, "Cloud cover estimation using multitemporal hemisphere imageries. International Archives of Photogrammetry," *Remote Sensing and Spatial Information*, vol. 35, pp. 826–829, 2004.
- [38] A. Suárez-García, D. Granados-López, D. González-Peña, M. Díez-Mediavilla, and C. Alonso-Tristán, "Seasonal characterization of CIE standard sky types above Burgos, northwestern Spain," *Solar Energy*, vol. 169, pp. 24–33, 2018.
- [39] P. Tregenza, "Analysing sky luminance scans to obtain frequency distributions of CIE Standard General Skies," *Lighting Research and Technology*, vol. 36, no. 4, pp. 271–279, 2004.
- [40] D. H. W. Li, T.C. Chau, and K.W. Wan, "A review of the CIE general sky classification approaches," *Renewable and Sustainable Energy Reviews*, vol. 31, pp. 563–574, 2014.
- [41] B. Bin Tian, M. A. Shaikh, M. R. Azimi-Sadjadi, T. H. V. Haar, and D. L. Reinke, "A study of cloud classification with neural networks using spectral and textural features," *IEEE Transactions on Neural Networks*, vol. 10, no. 1, pp. 138–151, 1999.
- [42] F. J. L. Lima, F. R. Martins, E. B. Pereira, E. Lorenz, and D. Heinemann, "Forecast for surface solar irradiance at the Brazilian Northeastern region using NWP model and artificial neural networks," *Renewable Energy*, vol. 87, pp. 807–818, 2016.

- [43] H. Calik, N. Ak, and I. Guney, "Artificial NARX neural network model of wind speed: case of Istanbul-Avcilar," *Journal of Electrical Engineering and Technology*, vol. 16, no. 5, pp. 2553–2560, 2021.
- [44] Y. Chen, Y. Wang, Z. Dong et al., "2-D regional short-term wind speed forecast based on CNN-LSTM deep learning model," *Energy Conversion and Management*, vol. 244, Article ID 114451, 2021.
- [45] G. Goudarzi, P. K. Hopke, and M. Yazdani, "Forecasting PM2.5 concentration using artificial neural network and its health effects in Ahvaz, Iran," *Chemosphere*, vol. 283, 2021.
- [46] Y. Kassem, H. Gökçekuş, and W. Janbein, "Predictive model and assessment of the potential for wind and solar power in Rayak region, Lebanon," *Modeling Earth Systems and Environment*, vol. 7, no. 3, pp. 1475–1502, 2021.
- [47] B. Shboul, I. Alarfi, S. Michailos, and D. Ingham, "A new ANN model for hourly solar radiation and wind speed prediction: a case study over the north & south of the Arabian Peninsula," *Sustainable Energy Technologies and Assessments*, vol. 46, Article ID 101248, 2021.
- [48] S. Gao, Z. Bai, S. Liang et al., "Simulation of surface ozone over Hebei province, China using Kolmogorov-Zurbenko and artificial neural network (KZ-ANN) combined model," *Atmospheric Environment*, vol. 261, Article ID 118599, 2021.
- [49] G. Gualtieri, S. D. Lonardo, F. Carotenuto et al., "The role of emissions and meteorology in driving CO₂ concentrations in urban areas," *Environmental Science and Pollution Research*, vol. 28, no. 23, pp. 29908–29918, 2021.
- [50] M. Ghamariadyan and M. A. Imteaz, "Monthly rainfall forecasting using temperature and climate indices through a hybrid method in Queensland, Australia," *Journal of Hydrometeorology*, vol. 22, no. 5, pp. 1259–1273, 2021.
- [51] P. Wellyantama and S. Soekirno, "Temperature, pressure, relative humidity and rainfall sensors early error detection system for automatic weather station (AWS) with artificial neural network (ANN) backpropagation," in *Proceedings of the 10th International Conference on Theoretical and Applied Physics (ICTAP)*, Lombok, Indonesia, 2020.
- [52] M. F. Møller, "A scaled conjugate gradient algorithm for fast supervised learning," *Neural Networks*, vol. 6, no. 4, pp. 525–533, 1993.
- [53] M. I. Dieste-Velasco, M. Díez-Mediavilla, D. Granados-López, D. González-Peña, and C. Alonso-Tristán, "Performance of global luminous efficacy models and proposal of a new model for daylighting in Burgos, Spain," *Renewable Energy*, vol. 133, pp. 1000–1010, 2019.
- [54] H. Kambezidis, T. Oikonomou, and D. Zevgolis, "Daylight climatology in the Athens urban environment: guidance for building designers," *Lighting Research and Technology*, vol. 34, no. 4, pp. 297–309, 2002.

Complementary Paper - CP I

Dieste-Velasco, M.I., Díez-Mediavilla, M., Granados-López, D., González-Peña, D., Alonso-Tristán, C., 2019. *Performance of global luminous efficacy models and proposal of a new model for daylighting in Burgos, Spain. Renewable Energy* 133, 1000-1010. DOI: 10.1016/j.renene.2018.10.085

Abstract: Daylighting is recognized as an important and useful strategy in the design of energy efficient buildings. Daylight is still the best source of light for good colour rendering and visual comfort. In this study, a new model of global luminous efficacy over a horizontal surface is proposed. A comparative study of eighteen classic models is presented, to obtain global horizontal illuminance, using both, the original formulation and new formulae with local adaptations, in order to determine the most suitable models for the conditions in Burgos (Spain). With this aim in mind, the selected models consisted of six models developed for all sky conditions, five models for clear sky conditions, three for partly cloudy sky and four for modelling overcast sky conditions. These eighteen models were also compared with the proposed model using experimental global illuminance measurements for different sky conditions. It was shown that the proposed model behaved in a better way than most of the classic models selected from the literature; both for all sky conditions and for particular sky conditions (clear, partly cloudy and overcast). The proposed model was therefore generally applicable, with no need to employ a different model for each particular sky condition.



Performance of global luminous efficacy models and proposal of a new model for daylighting in Burgos, Spain



M.I. Dieste-Velasco^{*}, M. Díez-Mediavilla, D. Granados-López, D. González-Peña, C. Alonso-Tristán

Solar and Wind Feasibility Technologies Research Group (SWIFT), Electromechanical Engineering Department, University of Burgos, Avda. de Cantabria s/n, Burgos, Spain

ARTICLE INFO

Article history:

Received 8 August 2018
Received in revised form
28 September 2018
Accepted 21 October 2018
Available online 22 October 2018

Keywords:

Luminous efficacy models
Illuminance
Irradiance
Modelling

ABSTRACT

Daylighting is recognized as an important and useful strategy in the design of energy efficient buildings. Daylight is still the best source of light for good colour rendering and visual comfort. In this study, a new model of global luminous efficacy over a horizontal surface is proposed. A comparative study of eighteen classic models is presented, to obtain global horizontal illuminance, using both, the original formulation and new formulae with local adaptations, in order to determine the most suitable models for the conditions in Burgos (Spain). With this aim in mind, the selected models consisted of six models developed for all sky conditions, five models for clear sky conditions, three for partly cloudy sky and four for modelling overcast sky conditions. These eighteen models were also compared with the proposed model using experimental global illuminance measurements for different sky conditions. It was shown that the proposed model behaved in a better way than most of the classic models selected from the literature; both for all sky conditions and for particular sky conditions (clear, partly cloudy and overcast). The proposed model was therefore generally applicable, with no need to employ a different model for each particular sky condition.

© 2018 Elsevier Ltd. All rights reserved.

1. Introduction

Solar-energy-based conversion systems and daylighting schemes are recognized as an important design strategy to generate clean energy that is sustainable and environmentally friendly, thereby reducing peak electricity consumption and cooling demands and saving on the total energy consumption of the building. The availability of natural light is also recommendable for reasons of visual comfort, and the physical and mental well-being of building occupants [1]. Daylighting not only improves aesthetic values, but can also lead to savings, using appropriate controls, of up to 50% on lighting energy [2]. International recommendations of energy standards and green building rating systems strongly advise architects to incorporate daylighting strategies in their building designs [3]. Illuminance data are essentially for the incorporation of daylighting in the design of energy-efficient buildings and for

suitable dimensioning of both the cooling and the heating systems. The availability of daylight has been recognized to be site-specific, although the measurement of daylight is not so common on a long-term basis [4]. An alternative method to increase illuminance data is through the use of luminous efficacy. Once the ratio of illuminance to irradiance, i.e., the luminous efficacy, is known, then measured irradiance values can be converted to illuminance values, which can in turn be used as input for a daylight simulation tool for the calculation of available daylight. The luminous efficacy value is not a constant, but will vary with solar altitude, cloud cover, and the amounts of aerosol and water vapour in the atmosphere [5]. The luminous efficacy models based on atmospheric conditions are also strongly dependent on those local variables [6]. Hence the importance of studying models of luminous efficacy to predict the values of illuminance at any one location.

Several studies have followed that pattern, mostly studying the local behaviour of luminous efficacy and its variability. Littlefair [5] reviewed different models of luminous efficacies formulated by different authors prior to 1985 at several global locations, highlighting the strong dependency of luminous efficacy on local climatic conditions. Vartiainen [7] studied the behaviour of five

^{*} Corresponding author.

E-mail addresses: midieste@ubu.es (M.I. Dieste-Velasco), mdmr@ubu.es (M. Díez-Mediavilla), dgranados@ubu.es (D. Granados-López), davidgp@ubu.es (D. González-Peña), catristan@ubu.es (C. Alonso-Tristán).

Nomenclature

a_i, b_i, c_i, d_i	Perez coefficients	L_{gh}	horizontal global illuminance (lux)
C	cloud cover	m	relative optical air mass
D	cloud ratio or sky ratio or diffuse fraction	MBE	Mean Bias Error (%)
E_0	correction factor for the sun-earth distance	n	number of data
E_{bh}	horizontal beam irradiance (W/m^2)	p_0, p_1, p_2	coefficients of the proposed model
E_{dh}	horizontal diffuse irradiance (W/m^2)	RMSE	Root Mean Square Error (%)
E_{gh}	horizontal global irradiance (W/m^2)	t	outdoor air temperature ($^{\circ}C$)
I	normal incidence direct irradiance (W/m^2)	T_d	three-hourly surface dew point temperature ($^{\circ}C$)
I_0	extra-terrestrial irradiance (W/m^2)	W	atmospheric precipitable water (cm)
I_{sc}	solar constant	$X_{measured}$	measured variable
K_g	global luminous efficacy (lm/W)	X_{model}	predicted variable
K_t	clearness index	Z	solar zenith angle (rad)
L_{bh}	horizontal beam illuminance (lux)	α	solar altitude angle (rad)
L_{dh}	horizontal diffuse illuminance (lux)	Δ	sky brightness
		ϵ	sky clearness
		Ω	relative heaviness of overcast sky

models of luminous efficacy in Finland, showing that Perez et al. model [8] was the only one that improved the predictions of the constant luminous efficacy model. De Souza et al. [9] showed the local dependency of the luminous efficacy models, improving the results obtained when local coefficients were calculated for different models. Patil et al. [10] remarked on the good behaviour of Perez et al. model with a locally adapted coefficient for different climatic zones in India, and Azad et al. [11] proposed new global and luminous efficacy models with constant forms for New Delhi.

As previously mentioned, eighteen models of luminous efficacy are reviewed and tested in this study in the city of Burgos, Spain, using both the original form of these models proposed by their authors and their local adaption to the location under study. Traditional statistical indicators RMSE (%) and MBE (%) were used to classify the models and to determine their accuracy. One year and a half of experimental data on illuminance were used in this study. In addition, a new model to predict global horizontal illuminance is proposed. This new model is analysed for all sky conditions and for particular sky conditions (clear, partly cloudy and overcast) showing the improvement in the illuminance prediction over the eighteen previously tested models for the city of Burgos, Spain.

The structure of this paper will be as follows: the experimental meteorological facility and data used for the study will be described in Section 2. Section 3 will describe the global luminous efficacy models on horizontal surfaces that are reviewed in this work. The benchmarking results of the eighteen luminous efficacy models under review will be presented in Section 4. The new model proposed for the area under study and its comparison with the others models under review will be shown in Section 5. In Section 6, the validation of both the proposed model and the eighteen luminous efficacy models under review will be presented and, finally, the main conclusions of this study will be outlined, remarking on the goals of the work and future lines of study.

2. Daylight global illuminance and solar global irradiance measurements

The experimental data for this study were gathered at a meteorological and radiometric facility located on the roof of the Higher Polytechnic School building at Burgos University ($42^{\circ}21'04''N$; $3^{\circ}41'20''O$; 856 m above mean sea level). This five-storey building, in an area with no other buildings of comparable height, has a horizon elevation angle that is lower than 10° with regard to the surface where the radiometric station is located. The experimental equipment is shown in Fig. 1.

The following meteorological data were measured: temperature, wind velocity and direction, atmospheric pressure, humidity and rainfall. Global, beam and diffuse horizontal irradiation (E_{gh} , E_{bh} , E_{dh}) and illuminance data (L_{gh} , L_{bh} , L_{dh}) were all recorded. Class 1 Hukseflux SR11 pyranometers and an EKO ML020SO Luxmeter were used to measure irradiance and illuminance data, respectively. The facility includes a SONA201D All-Sky Camera-Day and a MS-321LR sky scanner both from EKO. The experimental data were recorded on a CAMPBELL CR3000 datalogger. Experimental data were measured with a sampling time of 30 s, with average values recorded every 10 min, from 1st October 2016 to 31st March 2018, in order to determine the luminous efficacy models. The same



Fig. 1. Experimental equipment.

experimental procedure was followed from 1st April 2018 to 31st May, in order to measure the data for testing the models. The experimental 30 s values of E_{gh} , E_{bh} , E_{dh} , and L_{gh} , L_{bh} , L_{dh} were properly analysed and filtered using traditional quality criteria [12,13]. Whenever a 30 s data item failed to match the quality criteria, the values were eliminated. Fig. 2 shows the experimental values of horizontal global illuminance, L_{gh} (lux) versus horizontal global irradiance, E_{gh} (W/m^2) measured in Burgos.

3. Global luminous efficacy models on horizontal surfaces

The global luminous efficacy values (K_g) were obtained by simultaneously measuring both illuminance and irradiance on a specified surface and then computing their ratio, as shown in Equation (1):

$$K_g = \frac{L_{gh}}{E_{gh}} (lm/W) \quad (1)$$

where, L_{gh} is the global horizontal illuminance (lux) value and E_{gh} is the horizontal global irradiance (W/m^2) value. Alternatively, both the illuminance and the irradiance of particular sky elements can be measured to calculate the luminous efficacy. It is a convenient quantity for the calculation of daylight availability and lighting energy use in buildings. It enables daylight data to be generated from the more widely measured solar irradiance data for places where measured outdoor illuminance data are not recorded.

As previously mentioned, eighteen models of global horizontal luminous efficacy that cover different sky types will be reviewed in the following sections. Luminous efficacy models can be classified according to the number and type of input variables needed for their calculation: there are models of constant luminous efficacy, while others depend exclusively on solar altitude and others depend on more climatic variables. In some cases, the luminous efficacy model has a different form depending on the characteristics of the sky (clear, partly cloudy or overcast), while other models are applied to all types of sky. In the following paragraphs, the models used in this work are described. The models under review are presented in two ways: using the original coefficients given by their authors and adapted to local conditions. The previously described experimental data were used to calculate the local coefficients of the models. The non-linear Least Squares method was

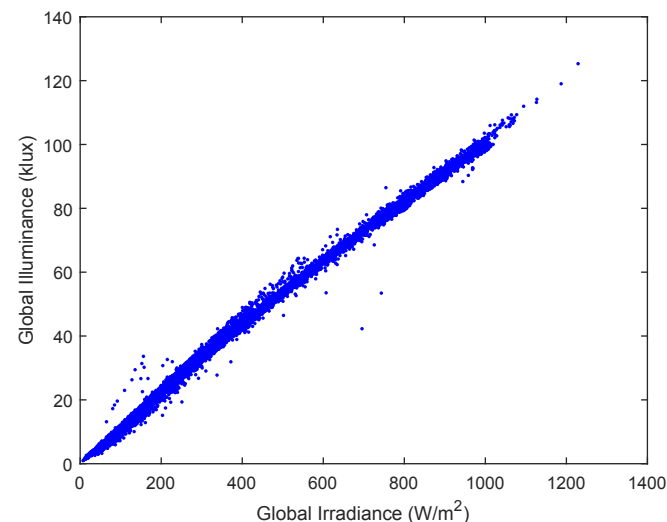


Fig. 2. Measured global illuminance vs measured global irradiance on horizontal surfaces, Burgos, (Spain).

employed using the Matlab™ 2017 fit function.

3.1. Perez et al. model (1990)

One widely used model of luminous efficacy is the Perez et al. model [8]. Applied at different locations around the world, it has consistently provided good illuminance prediction values. Diffuse, global and beam luminous efficacy can be modelled using the Perez et al. model for all kind of skies. Equation (2) allows the calculation of global luminous efficacy from radiance and the type of sky. These models were developed from illuminance data gathered at ten United States locations and three European cities covering different climatic conditions, from high altitude desert to temperate oceanic, oceanic and subtropical climates [8].

$$K_g = a_i + b_i W + c_i \cos(Z) + d_i \ln(\Delta) \quad (2)$$

where, a_i , b_i , c_i , d_i are the original coefficients of the model shown in Table 1(a). The local adaptation of these coefficients to the city of Burgos, are presented in Table 1(b). W is the atmospheric precipitable water content, defined by Equation (3). Z is the solar zenith angle, and Δ is the sky brightness as shown in Equation (4) [8].

$$W = e^{(0.07T_a - 0.075)} \quad (3)$$

$$\Delta = \frac{E_{dh} * m}{I_0} \quad (4)$$

The sky clearness parameter allows the classification of the sky, as in Equation (5) [8]:

$$\varepsilon = \left(\frac{E_{dh} + I}{E_{dh}} + kZ^3 \right) / \left(1 + kZ^3 \right) \quad (5)$$

where, $k = 1.041$ for Z in radians.

3.2. The Chung model (1992)

The Chung model [14] describes the luminous efficacy from the solar altitude, α , for the case of clear sky. This model was tested in the city of Hong Kong. In models for partly cloudy sky and overcast sky, the sky conditions are included in the models through the sky ratio of cloud ratio parameter, D , defined as the ratio of horizontal diffuse irradiance to horizontal global irradiance. The cloud ratio classifies the sky conditions as clear ($D < 0.3$), partly cloudy ($0.3 < D < 0.8$) and overcast ($D > 0.8$), and it gives different expressions for K_g calculations. Table 2 shows these expressions and the corresponding adaptation of the model to the city of Burgos.

In Table 2, $Q = E_{gh} / \sin \alpha$ shows the relative heaviness of overcast sky conditions and it represents the solar energy that passes through the cloud.

3.3. Lam and Li model (1996)

The clearness index, K_t , is obtained from Equation (6) and is defined as the ratio of the global radiation at ground level on a horizontal surface and the extra-terrestrial global solar irradiation [15]. This is the main parameter of this model, also tested in the city of Hong Kong [16].

$$K_t = \frac{E_{gh}}{I_0 \sin \alpha} \quad (6)$$

$$I_0 = I_{sc} E_0 \quad (7)$$

The clearness index classifies the sky as: clear sky ($K_t > 0.65$),

Table 1
Perez et al. model (1990).

ϵ category	Lower bound Upper bound		a) Original global luminous efficacy coefficients				b) Local global luminous efficacy coefficients for Burgos, Spain			
			a_i	b_i	c_i	d_i	a_i	b_i	c_i	d_i
1	1.000	1.065	96.63	-0.47	11.50	-9.16	109.53	0.04	-4.10	-3.14
2	1.065	1.230	107.54	0.79	1.79	-1.19	111.34	-0.63	-5.79	-2.00
3	1.230	1.500	98.73	0.70	4.40	-6.95	109.13	0.42	-5.68	-0.72
4	1.500	1.950	92.72	0.56	8.36	-8.31	103.61	0.57	2.39	1.09
5	1.950	2.800	86.73	0.98	7.10	-10.94	101.73	0.87	7.55	3.64
6	2.800	4.500	88.34	1.39	6.06	-7.60	116.20	0.61	11.61	11.03
7	4.500	6.200	78.63	1.47	4.93	-11.37	113.23	0.23	0.83	4.98
8	6.200	-	99.65	1.86	-4.46	-3.15	110.20	0.16	-17.50	-1.69

Table 2
Chung model equations for the calculation of luminous efficacy, K_g (lm/W), and for the different conditions of the sky. The original coefficients were calculated from experimental data recorded in Hong Kong. The locally adapted coefficients were calculated from the experimental data measured in Burgos, Spain.

Clear sky	Original model	$K_g = 102.2 + 0.69\alpha - 0.0059\alpha^2$
	Locally adapted model	$K_g = 95.106 + 22.493\alpha - 16.129\alpha^2$
Overcast sky	Original model	$K_g = (102.2 + 0.67\alpha - 0.0059\alpha^2) * (1.18 - 8.7 * 10^{-4} \Omega + 9.3 * 10^{-7} \Omega^2)$
	Locally adapted model	$K_g = (101.958 + 7.144\alpha - 7.387\alpha^2) * (1.135 - 2.32 * 10^{-4} \Omega + 1.77 * 10^{-7} \Omega^2)$
Partly cloudy sky	Original model	$K_g = D(135.3 - 25.7D) + (48.5 + 1.67\alpha - 0.0098\alpha^2)(1 - D)$
	Locally adapted model	$K_g = D(104.118 + 8.894D) + (83.401 + 56.696\alpha - 40.159\alpha^2)(1 - D)$

partly cloudy sky ($0.3 < K_t \leq 0.65$) and overcast sky ($0 < K_t \leq 0.3$). Following this classification, K_g is obtained through the mathematical expressions shown in Table 3. The local adaptation of the model to the city of Burgos is also presented in Table 3.

3.4. Muneer and Kinghorn model (1998)

Muneer and Kinghorn [17] proposed a model of K_g valid for all sky conditions that was tested in five different locations of UK. This polynomial model has the clearness index K_t , as an input parameter. The original expressions and the local adaptation of the model to the city of Burgos are presented in Table 4.

3.5. Robledo and Soler model (2000)

Two different models of luminous efficacy, A and B for clear sky conditions were proposed by Robledo and Soler [18] using experimental data of illuminance and irradiance measured in Madrid, Spain. The clear sky condition was determined through sky brightness ($\Delta < 0.12$) and sky clearness ($\epsilon > 5.0$). Both parameters were previously defined by Equation (4) and Equation (5). Table 5 shows the mathematical expressions of these models and their local adaptation to the city of Burgos.

3.6. Ruiz et al. model (2001)

This all sky model for luminous efficacy has the solar altitude, α , and the clearness index (K_t) as its input parameters. It was

Table 4

Muneer and Kinghorn model equations for luminous efficacy calculations, K_g (lm/W). The original coefficients were calculated with data from five different UK locations. The locally adapted coefficients were calculated with the experimental data measured in Burgos, Spain.

All sky model	Original model	$K_g = 136.6 - 74.541K_t + 57.3421K_t^2$
	Locally adapted model	$K_g = 112.952 - 5.809K_t - 9.487K_t^2$

proposed by Ruiz et al. [19] for the city of Madrid, Spain. The equation of the model and its adaptation to the local conditions of Burgos are presented in Table 6.

3.7. Robledo et al. model (2001)

Robledo et al. [20] proposed two different models of global luminous efficacy for overcast skies (Models A and B) and a third for partly cloudy skies. The sky classifications were established from the sky clearness parameter (ϵ) defined previously by Equation (5). The overcast sky condition was ($\epsilon < 1.2$) and the partly cloudy sky condition was ($1.2 < \epsilon < 5.0$). The solar altitude (α) and sky brightness (Δ), which is defined by Equation (4), were the input parameters of the models. The mathematical expressions of these models and their adaptation to the local conditions are shown in Table 7.

3.8. De Souza et al. model (2006)

De Souza et al. [9] proposed a clear sky model of luminous

Table 3
Lam and Li and model equations for luminous efficacy calculations, K_g (lm/W), and for the different conditions of the sky. The original coefficients were calculated from experimental data recorded in Hong Kong. The locally adapted coefficients were calculated from the experimental data measured in Burgos, Spain.

Clear sky	Original model	$K_g = (59.15 + 1.12\alpha - 0.0061\alpha^2)(1 - D) + 130.6D$
	Locally adapted model	$K_g = (118.752 - 0.513\alpha + 0.003\alpha^2)(1 - D) + 108.837D$
Overcast sky	Original model	$K_g = 116.2$
	Locally adapted model	$K_g = 111.744$
Partly cloudy sky	Original model	$K_g = (59.15 + 1.12\alpha - 0.0061\alpha^2)(1 - D) + (130.6 - 14.4C)D$
	Locally adapted model	$K_g = (62.240 + 2.436\alpha - 0.031\alpha^2)(1 - D) + (111.693 - 0.973C)D$

Table 5

Robledo and Soler model equations for luminous efficacy calculations, K_g (lm/W). The original coefficients were calculated from experimental data recorded in Madrid, Spain. The locally adapted coefficients were calculated from the experimental data measured in Burgos, Spain.

Clear sky model A	Original model	$K_g = 100.97 + 0.32\alpha - 0.000019\alpha^3 + 6.6257*10^{-9}\alpha^5$
	Locally adapted model	$K_g = 99.854 + 15.570\alpha - 24.505\alpha^3 + 9.459\alpha^5$
Clear sky model B	Original model	$K_g = 129.46(\sin\alpha)^{0.122}e^{-0.0029\alpha}$
	Locally adapted model	$K_g = 115.827(\sin\alpha)^{0.048}e^{-0.132\alpha}$

Table 6

Ruiz et al. model equations for luminous efficacy calculations, K_g (lm/W). The original coefficients were calculated from experimental data recorded in Madrid, Spain. The locally adapted coefficients were calculated from the experimental data measured in Burgos, Spain.

All sky model	Original model	$K_g = 104.83(\sin\alpha)^{0.026}K_t^{-0.108}$
	Locally adapted model	$K_g = 101.086(\sin\alpha)^{-0.021}K_t^{-0.060}$

efficacy for Florianopolis, Brazil. The clear sky condition was defined by ($\Delta < 0.12$ and $\varepsilon \geq 5.0$), where parameters Δ and ε are defined by Equation (4) and Equation (5), respectively. Table 8 collects the original form of the model and the form of its local adaption to the city of Burgos.

3.9. Fakra et al. model (2011)

Fakra et al. [21] established an all sky type luminous efficacy model for Saint-Pierre (Reunion Island) based on a constant form. This model and its adaptation to the local conditions of Burgos are shown in Table 9.

3.10. Mahdavi and Dervishi model (2011)

Clearness index (K_t), and the outdoor air temperature (t) are the input parameters used by Mahdavi and Dervishi [22] to calculate the global luminous efficacy in Vienna, Austria, for all sky types, as shown in Table 10, joined to the local adaptation for the city of Burgos.

3.11. Chaiwiwatworakul and Chirarattananon model (2013)

Global and diffuse luminous efficacy were evaluated by Chaiwiwatworakul and Chirarattananon [4] at Bangkok, Thailand. The sky clearness (ε) and zenith angle (Z) were used as input parameters for the all sky type model as shown in Table 11. The local adaptation of this model is also shown in Table 11.

Table 7

Robledo et al. model equations for luminous efficacy calculations, K_g (lm/W). The original coefficients were calculated from experimental data recorded in Madrid, Spain. The locally adapted coefficients were calculated from the experimental data measured in Burgos, Spain.

Overcast Sky model A	Original model	$K_g = [129.46(\sin\alpha^{0.122})e^{-0.0029\alpha}] * (1.361 - 1.091\Delta + 1.0334\Delta^2)$
	Locally adapted model	$K_g = [115.905(\sin\alpha)^{0.055}e^{-116\alpha}] * (1.128 - 0.418\Delta + 0.531\Delta^2)$
Overcast Sky model B	Original model	$K_g = 128.16(\sin\alpha)^{0.122}e^{-0.0029\alpha}\Delta^{-0.105}$
	Locally adapted model	$K_g = 117.070(\sin\alpha)^{-0.041}e^{-0.104\alpha}\Delta^{-0.022}$
Partly cloudy sky	Original model	$K_g = 120.26(\sin\alpha)^{0.077}e^{-0.0019\alpha}\Delta^{0.002}$
	Locally adapted model	$K_g = 138.173(\sin\alpha)^{0.137}e^{-0.234\alpha}\Delta^{0.022}$

Table 8

De Souza et al. model equations for luminous efficacy calculations, K_g (lm/W). The original coefficients were calculated from experimental data recorded in Florianopolis, Brazil. The locally adapted coefficients were calculated from the experimental data measured in Burgos, Spain.

Clear sky model	Original model	$K_g = 99.10 + 0.927\alpha - 0.0298\alpha^2 + 0.000422\alpha^3 - 2.2*10^{-6}\alpha^4$
	Locally adapted model	$K_g = 101.104 - 1.700\alpha + 1.362\alpha^2 - 0.294\alpha^3 - 0.241\alpha^4$

Table 9

Fakra et al. model equations for luminous efficacy calculations, K_g (lm/W). The original coefficients were calculated from experimental data recorded at Saint-Pierre, Reunion Island. The locally adapted coefficients were calculated from the experimental data measured in Burgos, Spain.

All sky model	Original model	$K_g = 121.5$
	Locally adapted model	$K_g = 103.428$

A summary of the main features of the models reviewed and the parameters used by each of them is shown in Table 12.

4. Evaluation of the global luminous efficacy models on a horizontal plane

The goodness-of-fit of the models was calculated by means of the statistical indicators MBE (%) (Mean Bias Error) and RMSE (%) (Root Mean Square Error) [21], [23]. MBE shows the trend of the model to either over-estimate or under-estimate the data. RMSE provides a measure of the deviation between the predicted values using the fitted models and the experimental measurements. Equations (8) and (9) show the statistical estimators employed in the present study.

$$MBE (\%) = 100 \frac{\sum_n (X_{model} - X_{measured})}{\sum_n X_{measured}} \tag{8}$$

$$RMSE (\%) = 100 \sqrt{\frac{\sum_n (X_{model} - X_{measured})^2}{n}} \frac{1}{\sum_n \frac{X_{measured}}{n}} \tag{9}$$

Tables 13–16 present the results obtained following the application of the statistical estimators shown by Equation (8) and (9) to the models analysed in this study. Table 13 shows the results obtained for the case of all sky conditions (six models). It can be observed that, when local coefficients were used, the model with the lowest RMSE was that of Chaiwiwatworakul and

Table 10

Mahdavi and Dervishi model equations for luminous efficacy calculations, K_g (lm/W). The original coefficients were calculated from experimental data recorded at Vienna, Austria. The locally adapted coefficients were calculated from the experimental data measured in Burgos, Spain.

<i>All sky model</i>	<i>Original model</i>	$K_g = 140.9 + 0.273t - 102K_t + 0.60t * K_t - 0.001t^2 + 77.28K_t^2$
	<i>Locally adapted model</i>	$K_g = 112.554 + 0.139t - 5.331K_t - 0.040t * K_t - 0.008t^2 - 7.140K_t^2$

Table 11

Chaiwiwatworakul and Chirarattananon model equations for luminous efficacy calculations, K_g (lm/W). The original coefficients were calculated from experimental data recorded at Bangkok, Thailand. The locally adapted coefficients were calculated from the experimental data measured in Burgos, Spain.

<i>All sky model</i>	<i>Original model</i>	$K_g = (101.65 + 13.92e^{-3.49})(\cos Z)^{(-0.18 + 0.19e^{-1.25})}$
	<i>Locally adapted model</i>	$K_g = (101.076 + 7.898e^{-2.181})(\cos Z)^{(-8.475 + 8.453e^{-0.002})}$

Table 12

Summary of the global luminous efficacy models reviewed in this work. Literature reference of the original model, year, sky type classification, input parameters used in the models and the original place of development of the model.

Ref.	Year	Authors	Sky types	Model parameters	Location
[8]	1990	Perez et al.	All	Δ, Z, W	USA and Europe
[14]	1992	Chung	Clear	α	China
			Overcast	α, Ω	
			Partly	α, D	
[16]	1996	Lam and Li	Clear	α, D	China
			Overcast	116.2 lm/W	
			Partly	α, C, D	
[17]	1998	Muneer and Kinghorn	All	K_t	UK
[18]	2000	Robledo and Soler (Model A) and Robledo and Soler (Model B)	Clear	α	Spain
[19]	2001	Ruiz et al.	All	α, K_t	Spain
[20]	2001	Robledo et al.	Partly	α, Δ	Spain
		Robledo et al. (Model A) and Robledo et al. (Model B)	Overcast		
[9]	2006	De Souza et al.	Clear	α	Brazil
[21]	2011	Fakra et al.	All	121.5 lm/W	Reunion Island
[22]	2011	Mahdavi and Dervishi	All	K_t, t	Austria
[4]	2013	Chaiwiwatworakul and Chirarattananon	All	Z, ε	Thailand

Table 13

Evaluation of the global luminous efficacy models for all skies.

Model	Original coefficients		Local coefficients	
	MBE (%)	RMSE (%)	MBE (%)	RMSE (%)
Chaiwiwatworakul and Chirarattananon	3.18	5.34	-0.41	3.61
Mahdavi and Dervishi	15.36	24.14	-0.20	3.65
Perez et al.	6.03	11.97	0.15	3.68
Ruiz et al.	4.16	6.99	-0.17	3.81
Muneer and Kinghorn	8.62	12.89	-0.28	3.86
Fakra et al.	15.11	21.24	-2.01	5.11

Chirarattananon [4] (3.61%) followed by the Mahdavi and Dervishi model [22] (3.65%) and the Perez et al. model [8] (3.68%).

Table 14 shows the results obtained for the case of a clear sky (five models). The models with the lowest RMSE values, when local coefficients were employed, were those of Robledo and Soler [18] (1.86%) followed by the model of Lam and Li [16] (2.02%).

Table 15 shows the results obtained for the case of partly cloudy

Table 14

Evaluation of the global luminous efficacy models for clear skies.

Model	Original coefficients		Local coefficients	
	MBE (%)	RMSE (%)	MBE (%)	RMSE (%)
Robledo and Soler (Model A)	0.00	2.26	-0.06	1.86
Robledo and Soler (Model B)	23.94	24.98	-0.11	1.86
Lam and Li	6.28	8.59	0.02	2.02
Chung	1.29	2.76	-0.15	2.08
De Souza et al.	-1.31	2.57	-1.00	2.34

Table 15

Evaluation of the global luminous efficacy models for partly cloudy skies.

Model	Original coefficients		Local coefficients	
	MBE (%)	RMSE (%)	MBE (%)	RMSE (%)
Robledo et al.	12.10	16.63	-0.01	3.07
Chung	-15.79	20.89	-0.28	3.38
Lam and Li	6.15	10.19	0.19	4.43

Table 16

Evaluation of the global luminous efficacy models for overcast skies.

Model	Original coefficients		Local coefficients	
	MBE (%)	RMSE (%)	MBE (%)	RMSE (%)
Robledo et al. (Model A)	22.04	29.46	-0.74	4.27
Robledo et al. (Model B)	21.48	29.09	-0.84	4.35
Chung	-6.15	10.49	-0.78	4.60
Lam and Li	3.93	8.44	-0.06	7.09

skies (three models). When local coefficients were used, the model with the lowest RMSE value was that of Robledo et al. [20] (3.07%), followed by the model of Chung [14] (3.38%).

Finally, Table 16 shows the results obtained for the case of overcast sky conditions (four models). It can be noted that when using local coefficients, the models with the lowest RMSE values were those of Robledo et al. (Model A) [20] (4.27%) and Robledo et al. (Model B) [20] (4.35%).

As was expected *a priori*, it can be affirmed from the results in Table 13–Table 16 that the models fitted with data from local measurements provided lower RMSE values than those obtained when using original coefficients.

5. Proposal of a new model to predict global luminous efficacy

In this Section, the proposed model to predict global luminous efficacy on horizontal surfaces is presented. Several models were analysed for modelling global luminous efficacy as a function of both the clearness index (K_t) and the solar altitude (α). From these results, the function that had the best fit was found to be the model shown in Equation (10). The advantage of the two parameters that this new model uses as its independent variables is that they are easily obtained. The model was firstly proposed for all sky conditions, yielding lower RMSE than any of the models shown in Table 13. Likewise, this new model also yielded lower RMSE values than the lowest ones shown in Table 14 (clear sky), Table 15 (partly cloudy sky) and Table 16 (overcast sky conditions). Therefore, as will be shown afterwards, this model can be generally applied either for all sky conditions or for other particular sky conditions.

$$K_g = p_0 * e^{p_1 * K_t * \sin(p_2 * \alpha^2)} \quad (lm/W) \quad (10)$$

Fig. 3 represents the experimental global luminous efficacy, K_g (lm/W), versus the clearness index (K_t) and Fig. 4 shows the experimental global luminous efficacy, K_g (lm/W) versus the solar altitude (α).

5.1. All sky conditions

The model fitted with experimental data measured in the city of Burgos is shown in Equation (11).

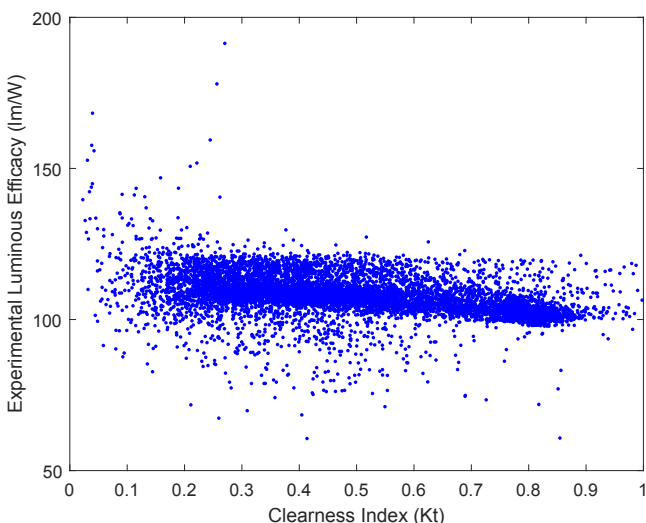


Fig. 3. Experimental luminous efficacy vs clearness index at Burgos.

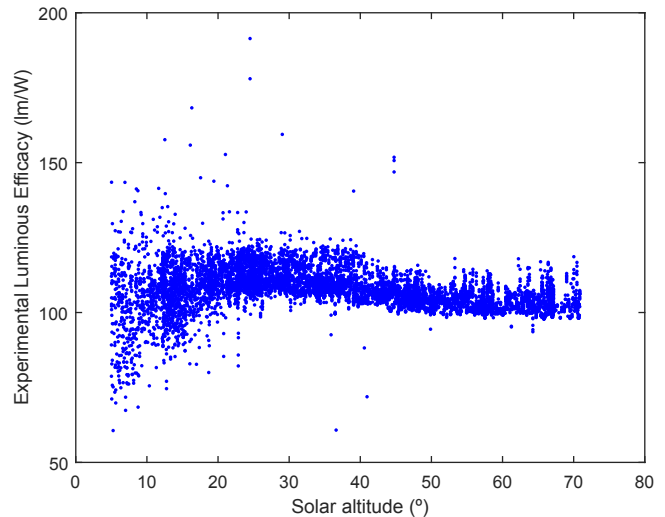


Fig. 4. Experimental luminous efficacy vs solar altitude at Burgos.

$$K_g = 111.616 * e^{-0.127 * K_t * \sin(1.232 * \alpha^2)} \quad (11)$$

The model shown in Equation (11) yielded an RMSE = 3.27% and an MBE = -0.19% for all sky conditions. This RMSE value was lower than any of the six RMSE values obtained with the models shown in Table 13. As can be observed in Table 13 and in Table 17, the lowest RMSE value was provided by Chaiwivatworakul and Chirattananon [4] (RMSE = 3.61%), higher than that obtained with the proposed model. Therefore, it can be affirmed that the proposed model was capable of predicting the global illuminance for all sky conditions more accurately than the other models analysed in Table 13, for local data measured in Burgos.

Fig. 5 shows the estimated global illuminance with the proposed model vs the measured global illuminance for all sky conditions. As can be observed in this figure, the proposed model acceptably predicted the global illuminance values for all sky conditions.

5.2. Clear sky

Equation (12) shows the proposed model, adapted for the particular case of clear sky which is defined by ($\epsilon > 5.0$ and $\Delta < 0.12$). These conditions are employed by the Robledo and Soler models [18] that have the lowest RSME values of all the models shown in Table 14. The new proposed model yielded an MBE = -0.03% and an RMSE = 1.80%. As can be observed, the RMSE was slightly lower than the one obtained with the models of Robledo and Soler [18] (1.86%). Moreover, as can be observed in Table 18, the proposed model in Equation (11), locally fitted for all sky conditions, showed a similar RMSE value to the previous ones.

$$K_g = 108.591 * e^{-0.111 * K_t * \sin(1.031 * \alpha^2)} \quad (12)$$

Fig. 6 shows the estimated global illuminance with the proposed model versus the measured global illuminance for clear sky conditions (given by $\epsilon > 5.0$ and $\Delta < 0.12$). As can be observed, the proposed model adequately predicted the global illuminance in the case of clear sky conditions.

5.3. Partly cloudy sky

Equation (13) shows the proposed model, adapted for the particular case of partly cloudy sky conditions defined by

Table 17
Comparison between the best performing model for all sky and the proposed model.

Model	Local coefficients	
	MBE (%)	RMSE (%)
Proposed model, All sky ($p_0 = 111.616$; $p_1 = -0.127$; $p_2 = 1.232$)	-0.19	3.27
Chaiwiwatworakul and Chirattananon	-0.41	3.61

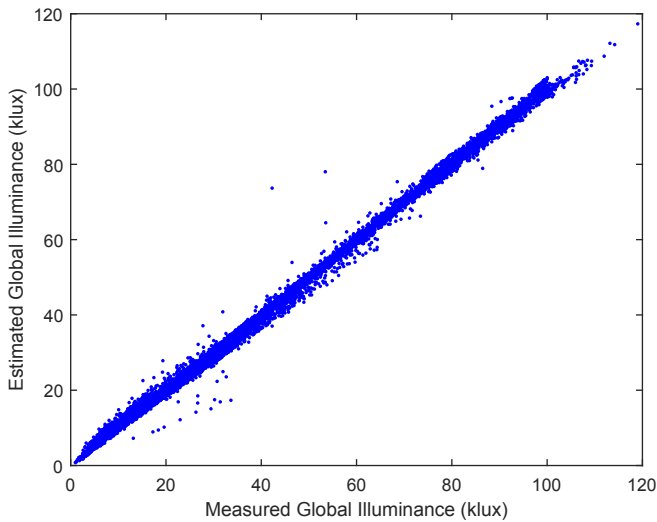


Fig. 5. Estimated global illuminance with the proposed model vs measured global illuminance for all sky conditions.

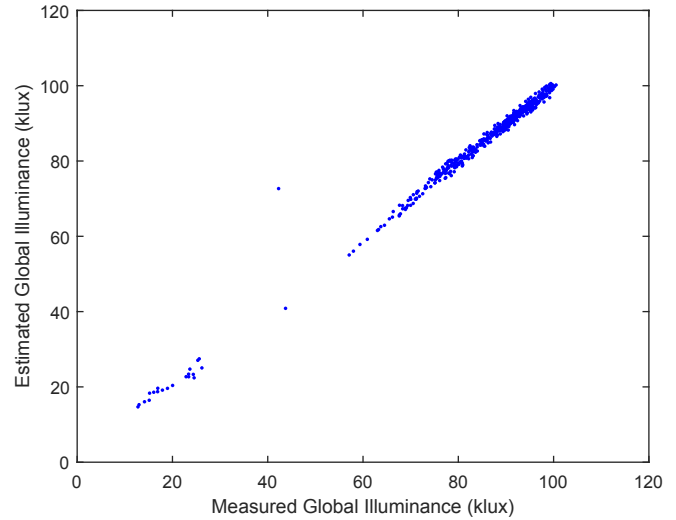


Fig. 6. Estimated global illuminance with the proposed model vs measured global illuminance for clear sky conditions given by $\epsilon > 5.0$ and $\Delta < 0.12$.

($1.20 < \epsilon < 5.0$). These conditions were employed by model of Robledo et al. [20], which is the model with the lowest RSME value of all the models shown in Table 15. The new model shown in Equation (13) yielded an RMSE of 2.89%, slightly lower than the value (3.07%) obtained with the model of Robledo et al. [20]. As observed in Table 19, the new model, locally fitted for all sky, as shown in Equation (11), yielded a lower RMSE value than the previous ones.

$$K_g = 109.152 * e^{-0.100 * K_t * \sin(1.013 * \alpha^2)} \quad (13)$$

Fig. 7 shows the estimated global illuminance with the proposed model versus measured global illuminance for partly cloudy sky conditions. As can be observed, the proposed model acceptably predicted global illuminance values for partly cloudy sky conditions defined from ($1.20 < \epsilon < 5.0$).

5.4. Overcast sky conditions

Equation (14) shows the proposed model, adapted for the particular case of overcast sky defined by ($\epsilon < 1.2$). This condition is employed by Robledo et al. (Model A) [20], which has the lowest RSME value of all the models shown in Table 16. The new model

shown in Equation (14) yielded an RMSE of 4.22%, a slightly lower value than the one obtained with Model A (Robledo et al.) [20]. On the other hand, as can be observed in Table 20, the new model, locally fitted for all sky, which is shown by Equation (11), yielded an RMSE value similar to the previous ones.

$$K_g = 111.693 * e^{-0.103 * K_t * \sin(1.241 * \alpha^2)} \quad (14)$$

Fig. 8 shows estimated global illuminance with the proposed model versus measured global illuminance for overcast sky conditions. As can be observed, the proposed model acceptably predicts the global illuminance values for overcast sky conditions.

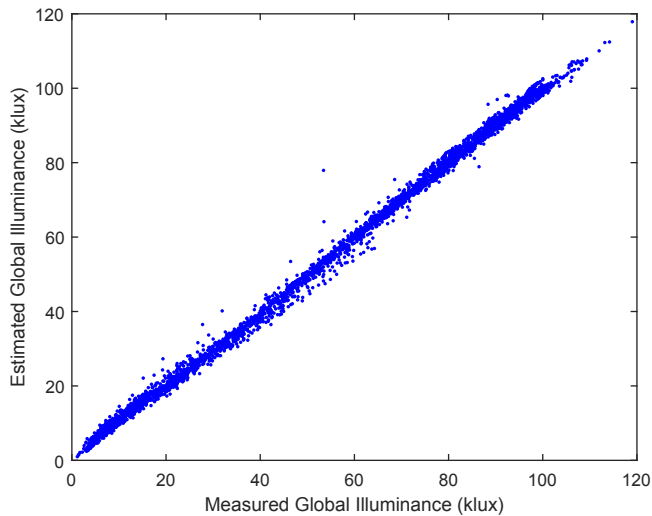
In this section, the new proposed model has been presented and analysed. It has been demonstrated that this new model yielded lower RMSE values than all the eighteen classic models considered in this study. These values have been verified both for all sky conditions and for particular (clear, partly cloudy and overcast sky) conditions. Moreover, as can be observed in Table 18-Table 20, the new model proposed for all sky conditions, which is shown in Equation (11), also provided values close to those obtained with models adapted for particular sky conditions (clear, partly cloudy and overcast).

Table 18
Comparative between the best performing model for clear sky and the proposed model, using the same sky conditions $\epsilon > 5.0$ and $\Delta < 0.12$.

Model	Local coefficients	
	MBE (%)	RMSE (%)
Proposed model, Clear sky ($p_0 = 108.591$; $p_1 = -0.111$; $p_2 = 1.031$)	-0.03	1.80
Robledo and Soler (Model A)	-0.06	1.86
Robledo and Soler (Model B)	-0.11	1.86
Proposed model, All sky ($p_0 = 111.616$; $p_1 = -0.127$; $p_2 = 1.232$)	0.88	2.01

Table 19Comparison between the best performing model for partly cloudy sky and the proposed model, using the same sky conditions ($1.20 < \epsilon < 5.0$).

Model	Local coefficients	
	MBE (%)	RMSE (%)
Proposed model, All sky ($p_0 = 111.616$; $p_1 = -0.127$; $p_2 = 1.232$)	0.31	2.84
Proposed model, Partly cloudy sky ($p_0 = 109.152$; $p_1 = -0.100$; $p_2 = 1.013$)	0.04	2.89
Robledo et al.	-0.01	3.07

**Fig. 7.** Estimated global illuminance with the proposed model vs measured global illuminance for partly cloudy sky conditions ($1.20 < \epsilon < 5.0$).

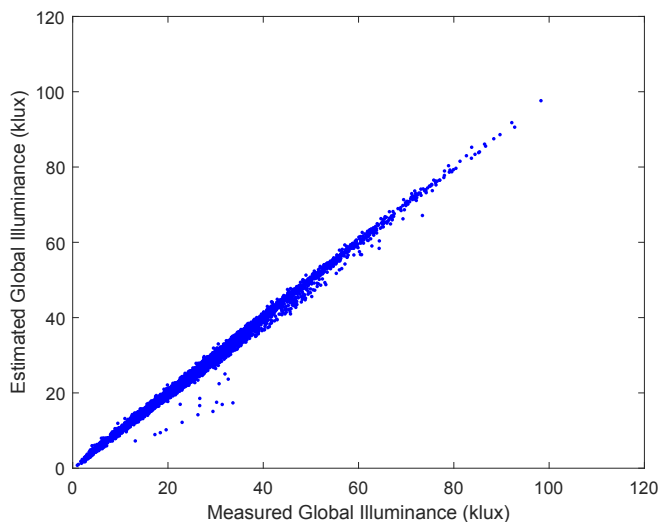
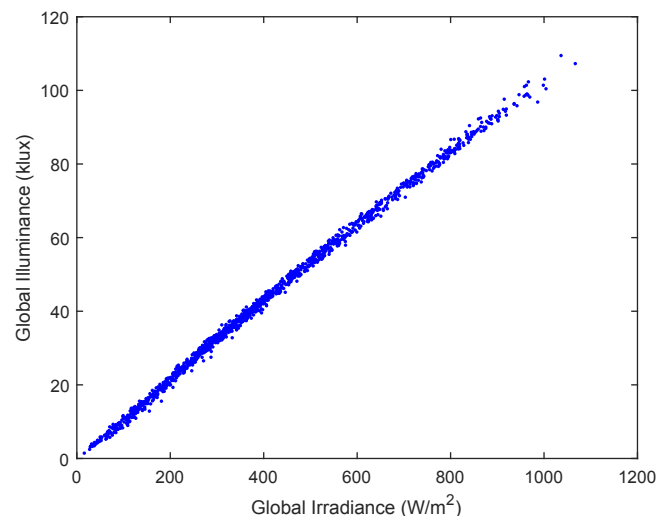
6. Validation of the global illuminance models

In Section 3, global luminous efficacy models from eighteen existing models in the literature were fitted by using local data from Burgos (Spain) and the models were then evaluated in Section 4. Moreover, the results of fitting and analysing a new model for all sky and for particular sky conditions, using the same data as the previous mentioned models, has been presented above in Section 5. In the present Section, validation of all these models is shown by employing two additional months of measurements (from 1st April 2018 to 31st May 2018). These measurements were taken, following the procedure shown in Section 2. Fig. 9 shows the experimental data employed for testing the global luminous efficacy models. This figure compares measured global illuminance versus measured global irradiance on the horizontal surface at Burgos over the test period.

Data obtained from these two additional months were used to re-evaluate both RMSE and MBE in the models that had previously been fitted with experimental data (local models). Tables 21–24 show the results obtained after evaluating the statistical estimators shown in Equation (8) and in Equation (9) taken from the

Table 20Comparison between the best performing model for overcast skies and the proposed model, using the same sky conditions (Overcast skies: $\epsilon < 1.2$).

Model	Local coefficients	
	MBE (%)	RMSE (%)
Proposed model, Overcast sky ($p_0 = 111.693$; $p_1 = -0.103$; $p_2 = 1.241$)	-0.81	4.22
Robledo et al. (Model A)	-0.74	4.27
Proposed model, All sky ($p_0 = 111.616$; $p_1 = -0.127$; $p_2 = 1.232$)	-1.37	4.40

**Fig. 8.** Estimated global illuminance with the proposed model vs measured global illuminance for overcast sky ($\epsilon < 1.2$).**Fig. 9.** Measured global illuminance vs measured global irradiance on the horizontal surface at Burgos. Test data (01/04/18–31/05/18).

luminous efficacy models that have been analysed in this study. The results obtained from the different sky conditions under study are also shown. To that end, particular sky conditions proposed by each author were applied, in order to define different sky types (clear sky, partly cloudy sky and overcast sky). The new model proposed in this study was also validated in both all sky and particular sky conditions (clear, partly cloudy and overcast). In the latter case, the conditions employed by the model with the lowest RMSE value were used in order to define the sky type.

Table 21 shows the results obtained with the testing data for the case of all sky conditions (seven models). The model of Ruiz et al. [19] (2.57%) was slightly lower than the new proposed model (2.66%). However, the MBE obtained with the proposed model (−0.01%) was ten times lower than the one obtained with the previous model (−0.1%).

The results obtained from classic clear sky models (five models) and the proposed model are shown in Table 22. In addition, the results obtained when the all sky model, given by Equation (11), was validated for this particular sky type are also compared. It is shown that the new model proposed in this study yielded the lowest RMSE values, both after validation with the all sky model coefficients (0.66%) and with the coefficients fitted with data from clear sky conditions (1.40%), followed by Robledo and Soler (Model A) [18] (1.53%).

Likewise, Table 23 shows the results obtained for classic partly cloudy sky models (three models) and these three models are also compared with the new proposed models. It can be noted that models with the lowest RMSE values are those of Robledo et al. [20] (2.43%) and the new model for partly cloudy sky (2.46%) followed by the model of Chung [14] (2.67%).

Finally, Table 24 shows the results obtained for the case of classic overcast sky models (four models) and these models are also compared with the new proposed models. It can be noted that the models proposed in this study yielded the lowest RMSE values when fitted with either overcast or with all sky conditions, followed by those of Robledo et al. (Model B) [20] (2.52%) and Robledo et al. (Model A) [20] (2.65%).

Table 21–Table 24 show the results obtained after validating the models with two additional measurements months. From these results, it can be observed that the proposed model, fitted for a specific sky condition, yield lower RMSE values for both overcast sky (Equation (14), 2.30%) and clear sky (Equation (12), 1.40%) than any of the analysed models. With regard to partly cloudy sky conditions, the RMSE obtained with the proposed model (Equation (13), 2.46%) was approximately equal to that of Robledo et al. [20] (2.43%), and the RMSE obtained with the proposed model (Equation (11), 2.66%) for all sky conditions was slightly higher than that of the model of Ruiz et al. [19] (2.57%).

It should be mentioned that the model fitted with all the data (all sky conditions), shown in Equation (11), can be generally applied for modelling particular sky types (clear, partly cloudy and overcast), because the RMSE values obtained after validating this

Table 21
Validation of the global luminous efficacy models for all skies.

Model	Local coefficients	
	MBE (%)	RMSE (%)
Ruiz et al.	−0.10	2.57
Proposed model. Equation (11)	−0.01	2.66
Chaiwiwatworakul and Chirattananon	1.23	2.81
Mahdavi and Dervishi	0.94	2.94
Perez et al.	1.31	2.98
Muneer and Kinghorn	0.36	3.22
Fakra et al.	−2.52	3.64

Table 22
Validation of the global luminous efficacy models for clear skies.

Model	Local coefficients	
	MBE (%)	RMSE (%)
Proposed model (All sky). Equation (11)	−0.26	0.66
Proposed model (Clear sky). Equation (12)	−1.21	1.40
Robledo and Soler (Model A)	−1.25	1.53
Robledo and Soler (Model B)	−1.32	1.65
Lam and Li	−0.20	2.07
Chung	−1.84	2.30
De Souza et al.	−3.07	3.43

Table 23
Validation of the global luminous efficacy models for partly cloudy skies.

Model	Local coefficients	
	MBE (%)	RMSE (%)
Robledo et al.	0.51	2.43
Proposed model (Partly sky). Equation (13)	−0.09	2.46
Chung	0.93	2.67
Proposed model (All sky). Equation (11)	0.27	2.80
Lam and Li	2.25	3.44

Table 24
Validation of the global luminous efficacy models for overcast skies.

Model	Local coefficients	
	MBE (%)	RMSE (%)
Proposed model (Overcast sky). Equation (14)	0.65	2.30
Proposed model (All sky). Equation (11)	−0.21	2.48
Robledo et al. (Model B)	0.71	2.52
Robledo et al. (Model A)	0.86	2.65
Chung	1.27	2.76
Lam and Li	2.56	4.16

model in these specific sky types were 0.66%, 2.80%, and 2.48%, respectively.

7. Conclusions

Eighteen classic global luminous efficacy models, from the existing literature, have been evaluated, both with their original coefficients and locally fitted with experimental data measured in Burgos (Spain), between 1st October 2016 and 31st March 2018. The local behaviour of the models has been noted, which leads to lower RMSE and MBE values than those obtained by using their original coefficients.

A new model to predict the global luminous efficacy on horizontal surfaces has been proposed and analysed in this study. This new model has been fitted for either all sky types or particular sky types (clear, partly cloudy and overcast). It employs the solar altitude and the clearness index (K_t) as independent variables, which have the advantage of being two easily obtained parameters.

It has been shown that with data employed for fitting the models over the period of study (01/10/16 to 31/03/18) in the city of Burgos (Spain), the new proposed model has provided lower RMSE values than any of the eighteen classic models analysed in this study, for either all sky or particular sky conditions (clear, partly cloudy and overcast). Moreover, this new model provides lower MBEs than most of the classical models analysed in this study. With regard to the results obtained with the validation data measured in the period (01/04/18 to 31/05/18), the proposed model has provided lower RMSE values for clear sky and overcast sky conditions than any of the classic models and it has provided similar RMSE

values to those obtained with the models that presented the lowest RMSE values for all sky and partly cloudy sky conditions.

It can be affirmed from these results that the model fitted for all sky conditions, shown in Equation (11), can also be applied for modelling the global illuminance in all sky types and in particular sky conditions (clear sky, partly cloudy and overcast), with no need to employ different luminous efficacy models for each specific sky type.

As future work, the proposed model could be applied to data gathered in different locations, in order to compare the results and to determine its applicability to the modelling of horizontal global illuminance.

Acknowledgements

The authors acknowledge the financial support given by the Spanish Government (Ministerio de Economía y Competitividad) (ENE2014-54601-R). David González Peña would also like to thank the Junta de Castilla-León for economic support (PIRTU Program, ORDEN EDU/310/2015).

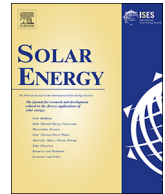
References

- [1] T. Hwang, T.K. Jeong, Effects of indoor lighting on occupants' visual comfort and eye health in a green building, *Indoor Built Environ.* 20 (1) (2011) 75–90.
- [2] P. Ihm, A. Nemri, M. Krarti, Estimation of lighting energy savings from daylighting, *Build. Environ.* 44 (3) (2009) 509–514.
- [3] Aalto University School of Science and Technology, I.E.A., Guidebook on Energy Efficient Electric Lighting for Buildings, IEA publications, Raisio, 2010.
- [4] P. Chaiwiwatworakul, S. Chirarattananon, Luminous efficacies of global and diffuse horizontal irradiances in a tropical region, *Renew. Energy* 53 (2013) 148–158.
- [5] P.J. Littlefair, The luminous efficacy of daylight: a review, *Light. Res. Technol.* 17 (4) (1985) 162–182.
- [6] P.J. Littlefair, Measurements of the luminous efficacy of daylight, *Light. Res. Technol.* 20 (4) (1988) 177–188.
- [7] E. Vartiainen, A comparison of luminous efficacy models with illuminance and irradiance measurements, *Renew. Energy* 20 (3) (2000) 265–277.
- [8] R. Perez, et al., Modeling daylight availability and irradiance components from direct and global irradiance, *Sol. Energy* 44 (5) (1990) 271–289.
- [9] R.G. de Souza, et al., Evaluation of global luminous efficacy models for Florianópolis, Brazil, *Build. Environ.* 41 (10) (2006) 1364–1371.
- [10] K.N. Patil, S.N. Garg, S.C. Kaushik, Luminous efficacy model validation and computation of solar illuminance for different climates of India, *J. Renew. Sustain. Energy* 5 (6) (2013).
- [11] A.S. Azad, D. Rakshit, K.N. Patil, Model development and evaluation of global and diffuse luminous efficacy for humid sub-tropical region, *Renew. Energy* 119 (2018) 375–387.
- [12] C.A. Gueymard, J.A. Ruiz-Arias, Extensive worldwide validation and climate sensitivity analysis of direct irradiance predictions from 1-min global irradiance, *Sol. Energy* 128 (2016) 1–30.
- [13] CIE 108-1994, Technical Report: Guide to Recommended Practice of Daylight Measurement, Commission International de l'Éclairage, 1994.
- [14] T.M. Chung, A study of luminous efficacy of daylight in Hong Kong, *Energy Build.* 19 (1) (1992) 45–50.
- [15] M. Iqbal, Introduction to Solar Radiation, Academic Press, 1983.
- [16] J.C. Lam, D.H.W. Li, Luminous efficacy of daylight under different sky conditions, *Energy Convers. Manag.* 37 (12) (1996) 1703–1711.
- [17] T. Muneer, D. Kinghorn, Luminous efficacy models - evaluation against UK data, *J. Illum. Eng. Soc.* 27 (1) (1998) 163–169.
- [18] L. Robledo, A. Soler, Luminous efficacy of global solar radiation for clear skies, *Energy Convers. Manag.* 41 (16) (2000) 1769–1779.
- [19] E. Ruiz, A. Soler, L. Robledo, Assessment of Muneer's luminous efficacy models in Madrid and a proposal for new models based on his approach, *J. Sol. Energy Eng. Trans. ASME* 123 (3) (2001) 220–224.
- [20] L. Robledo, A. Soler, E. Ruiz, Luminous efficacy of global solar radiation on a horizontal surface for overcast and intermediate skies, *Theor. Appl. Climatol.* 69 (1–2) (2001) 123–134.
- [21] A.H. Fakra, et al., A simple evaluation of global and diffuse luminous efficacy for all sky conditions in tropical and humid climate, *Renew. Energy* 36 (1) (2011) 298–306.
- [22] A. Mahdavi, S. Dervishi, A comparison of luminous efficacy models based on data from Vienna, Austria, *Build. Simulat.* 4 (3) (2011) 183–188.
- [23] G. López, C.A. Gueymard, Clear-sky solar luminous efficacy determination using artificial neural networks, *Sol. Energy* 81 (7) (2007) 929–939.

Complementary Paper - CP II

Suárez-García, A., Díez-Mediavilla, M., Granados-López, D., González-Peña, D., Alonso-Tristán, C. 2020. ***Benchmarking of meteorological indices for sky cloudiness classification***. *Sol. Energy* 195, 499–513.
DOI: 10.1016/j.solener.2019.11.060

Abstract: Sky classification is a complex problem, due in part to such abstract conceptual definitions as clear, intermediate, and overcast, as well as other intermediate ranges. The CIE (Commission Internationale de L'Éclairage) Standard classification offers a solution to this problem, although its application requires data on the luminance distribution of the whole sky that are less commonly available. A benchmarking and classification system of ten meteorological indices is introduced in this study to classify the sky types from overcast to clear. The indices can be calculated from measurements of global, diffuse, and direct irradiance that are widely available from meteorological ground stations. The classification system uses confusion matrices, a machine-learning tool that generates a visual display of the results of supervised-learning algorithms. The CIE Standard skies classification, applied to half hourly sky-scanner measurements in Burgos (Spain), over the period June 2016 - May 2017, is used in this study as a baseline reference for a comparative review of the results from the meteorological indices and their results. They are classified by four performance ratings: Accuracy, Jaccard, Cohen, and Matthews, which feature both classification similarity and the randomness of any agreement. All meteorological indices yielded a high average degree of accuracy - close to 80% - in a detailed review of their classification. Nevertheless, the results suggested that Perez's Clearness Index based on global, diffuse and direct radiation measurements offered the most precise classification of the skies, followed closely by the Klucher Clearness Index and the Perraudau Nebulosity Index.



Benchmarking of meteorological indices for sky cloudiness classification

Andrés Suárez-García^{a,b}, Montserrat Díez-Mediavilla^b, Diego Granados-López^b,
David González-Peña^b, Cristina Alonso-Tristán^{b,*}

^a Centro Universitario de la Defensa. Marín, Pontevedra, Spain

^b Research Group Solar and Wind Feasibility Technologies (SWIFT), Electromechanical Engineering Department, Universidad de Burgos, Spain

ARTICLE INFO

Keywords:

Clearness index
Sky classification
CIE standard
Cloudiness classification
Meteorological indices

ABSTRACT

Sky classification is a complex problem, due in part to such abstract conceptual definitions as clear, intermediate, and overcast, as well as other intermediate ranges. The CIE (Commission Internationale de L'Éclairage) Standard classification offers a solution to this problem, although its application requires data on the luminance distribution of the whole sky that are less commonly available. A benchmarking and classification system of ten meteorological indices is introduced in this study to classify the sky types from overcast to clear. The indices can be calculated from measurements of global, diffuse, and direct irradiance that are widely available from meteorological ground stations. The classification system uses confusion matrices, a machine-learning tool that generates a visual display of the results of supervised-learning algorithms. The CIE Standard skies classification, applied to half hourly sky-scanner measurements in Burgos (Spain), over the period June 2016 - May 2017, is used in this study as a baseline reference for a comparative review of the results from the meteorological indices and their results. They are classified by four performance ratings: Accuracy, Jaccard, Cohen, and Matthews, which feature both classification similarity and the randomness of any agreement. All meteorological indices yielded a high average degree of accuracy - close to 80% - in a detailed review of their classification. Nevertheless, the results suggested that Perez's Clearness Index based on global, diffuse and direct radiation measurements offered the most precise classification of the skies, followed closely by the Klucher Clearness Index and the Perraudau Nebulosity Index.

1. Introduction

A key aspect in the modelling of solar radiation and daylighting is sky classification. Many models for the calculation of global, direct, and diffuse irradiation and illumination (i.e. luminous efficacy) are defined for different sky types, based on the values of different climatic parameters. Searching for parameters that can quantify abstract concepts such as clear, partially cloudy, and overcast skies, as well as all possible intermediate classes, is a complex problem that researchers have addressed using different strategies. Sky conditions of the same category should have similar solar radiation and sky luminance distributions and the corresponding climatic parameters should be within certain ranges (Li and Lam, 2001).

In 2003, 15 standard sky types were defined in the CIE categorization (Uetani et al., 2003). The classification included five types of clear sky, five intermediate types, and five types of cloudy skies. Sky types of the same category have the same well-defined sky luminance pattern. Once the sky types are identified, the basic solar irradiance and daylight illuminance on the surfaces of interest can be obtained through

simple mathematical expressions (Li et al., 2013). The luminance distribution for each standard sky type can help arrive at accurate determinations of daylight illuminance (Kittler et al., 1997). Several works have reported that the CIE standard sky classification provides a good overall framework for representing the actual sky conditions and covers the whole probable spectrum of skies found in nature (Alshabani, 2011; Li and Cheung, 2006; Li et al., 2008; Li et al., 2007; Markou et al., 2005; Markou et al., 2004; Torres et al., 2010a, b; Tregenza, 2004; Wong et al., 2012). Each CIE General Standard Sky is well defined by the straightforward approach for sky classification: the sky luminance pattern. The standard instrument for measuring sky luminance distribution is a sky scanner and basic sky luminance data are available at many locations across the world.

When interpreting sky conditions, meteorological data are initially used as weighting factors to show the degree of sky clearness, but different researchers have each adopted different kinds of Meteorological Indices (MIs), each with a different range (Li et al., 2007; Lou et al., 2017; Umemiya and Kanou, 2008). Their selection depended on the availability of meteorological variables. Some previous attempts have

* Corresponding author.

E-mail address: catristan@ubu.es (C. Alonso-Tristán).

Nomenclature

Greek symbols

χ	scattering angle
$\varphi(Z)$	gradation function
γ, γ_s	angle of elevation, solar elevation
α, α_s	Azimuth angle, solar azimuth
$f(\chi)$	indicatrix function
\in	Perez's clearness index
ε_0	average value of the orbital eccentricity of the Earth
Δ	Perez's brightness index
κ	Cohen Kappa
Z, Z_s	angle from zenith, angle between sky zenith and sun
a, b, c, d, e	coefficients of the gradation and indicatrix function for CIE standard skies classification
b_p	reference number of each band
B_{sc}	extraterrestrial irradiance constant (1361.1 W/m ²)
$B(n)$	beam irradiance (W/m ²)
$B_{ext}(n)$	extraterrestrial direct solar irradiance (W/m ²)
Ces	standard cloudiness fraction
$G_{ext}(0)$	extraterrestrial global solar irradiance on a horizontal plane
Cle	cloudless index
d_n	day number of year, 1 on 1st January and 365 on 31st

	December. February has 28 days
D_V	diffuse luminance (cd/m ²)
$D(O)$	diffuse horizontal irradiance (W/m ²)
F	clearness function
FK	Klucher clearness index
FP	Perraudau nebosity index
G_{st}	standard global irradiance
$G_{clear}(0)$	clear-sky irradiance (W/m ²)
$G(O)$	global horizontal irradiance (W/m ²)
k_b	horizontal direct fraction
k_d	horizontal diffuse fraction
k_{d0}	horizontal diffuse fraction for a clear sky
k_k	Batilles clearness index
k_t	clearness index
k_{it}	auxiliary parameter for Batilles Clearness Index calculation
L_p	luminance of a sky patch measured by the Sky-scanner (cd/m ²)
L_z	Zenith luminance(cd/m ²)
MI	meteorological index
m	optical air mass
n_p	number of patches in band b
NR	luminance normalization ratio (cd/m ²)
p	reference number of a scanned sky patch
Si	Igawa index

been made to use specific MIs to classify sky conditions. The diffuse fraction, k_d , and the clearness index, k_t , were used in (Brunger and Hooper, 1993) for the classification of sky conditions. Igawa (Igawa et al., 2004) defined a clear-sky index, Si , that can be used as an index with no dependency on the solar altitude for the classification of sky conditions. (Baharuddin et al., 2010) and (Rahim et al., 2004) classified daylight data into three sky conditions – clear, intermediate, and overcast – using two methods: sunshine duration and cloud ratio methods. The sunshine duration method estimates the frequencies of the occurrence of clear, intermediate, and overcast sky from the monthly mean value of the relative sunshine duration. The cloud ratio method classifies the sky condition into three sky classes, based on the observation of the pattern of diffuse fraction graphs. (Kong and Kim, 2013) evaluated the ability of the diffuse fraction, Perez's clearness index, and the clearness index to classify the skies by means of a frequency distribution. (Umamiya and Kanou, 2008) introduced a sky classification method, by using different combinations of nine “insolation” or irradiation indices. In (Lou et al., 2017), the CIE Standard Skies identified by the luminance scan were correlated with fourteen meteorological parameters (solar altitude angle, clearness index, diffuse fraction, turbidity, air temperature, relative humidity, wet bulb temperature, and direct normal solar irradiance, among others) using the Classification Tree algorithm. (Li and Lam, 2001) investigated the prevailing sky conditions in Hong-Kong in terms of different climatic parameters (cloud cover distribution, hours of sunshine, k_t and k_d) and highlighted the merits of each one. In terms of modelling solar radiation and outdoor illuminance components, the authors concluded that k_t was the best parameter from among the four indices under study. (Li et al., 2014) analysed different alternative approaches to perform the CIE Standard skies classification, the clearness index, and the turbidity index, among others.

No previous studies have presented analyses of each MI and its sky classification capability. The frequency of occurrence of clear, intermediate, and overcast skies is determined by combining various MIs through mathematical algorithms. Likewise, no simultaneous comparison between the different approaches to sky classification has been conducted. The comparison is usually done in terms of frequency of appearance. Very few of the works just reviewed included comparisons

of their results with the CIE Standard sky classification. Nevertheless, the CIE Standard skies are internationally considered as sufficiently comprehensive to simulate skylight luminance and radiance distributions, crucial factors in passive energy efficient building designs and in active solar energy applications. Even a simple sky classification with only three categories (clear, partial, and cloudy) would allow a major improvement in lighting control systems (Li, 2010; Li et al., 2008), a precise selection of the models to estimate solar irradiance and for its prediction (Djafer et al., 2017; Ruiz-Arias and Gueymard, 2018), more accurate determination of the spectral components of solar radiation (Escobedo et al., 2009; Jacovides et al., 2007), and better prediction of photovoltaic production, to improve grid integration (Perveen et al., 2018).

In this study, a review of the different MIs used for sky classification and their benchmarking has led to the selection of those calculated with three standard variables recorded at many meteorological stations: global, beam, and diffuse irradiance. Sky conditions and characterization of CIE Standard sky types were reported in a previous publication with a full year of data recorded at Burgos, northwestern Spain, (Suárez-García et al., 2018). The sky classifications using these MIs were compared to the CIE cloudiness categories (cloudy, partial, and clear), adapting the original intervals to match the number of categories.

The main objective of this study is to assess the selected MIs presented as alternatives to the CIE classification, so as to perform sky classifications of acceptable accuracy. For sky classification, the appropriate MIs would be expected to require less expensive and more common equipment at meteorological ground stations than the sky scanner that is used to measure sky luminance and its distribution. Calculation of simple MIs is fast and allows an efficient implementation of control systems of daylighting and active solar energy applications. The use of confusion matrices as a benchmarking tool is a novelty in the study of sky conditions. In this way, both the MI classification and the CIE Standard, may be compared against the same timestamp, rather than through a frequency distribution that can only be done at the end of the data collection period.

This study is structured as follows: the CIE Standard classification will be described in Section 2. In Section 3, the MI derived from global,

beam, and diffuse irradiation will be used to define the sky types. Section 4 will examine the main characteristics of the confusion matrices classification and the indicators used for the MIs benchmarking process. The experimental facility and data used in this work, as well as the results of the study, will be described in Section 5 and in Section 6, respectively. Finally, the principal observations will be presented in Section 7 together with the main contributions of the study.

2. CIE standard classification

The CIE standard sky classification describes the luminance ratio of any given sky patch, $L_p(kcd/m^2)$, normalized by the sky's zenith luminance, $L_z(kcd/m^2)$, as the product of the relative gradation function, $\varphi(Z)/\varphi(0)$, and the relative scattering indicatrix function, $f(\chi)/f(0)$, as shown in Eq. (1):

$$\frac{L_p}{L_z} = \frac{f(\chi) \cdot \varphi(Z)}{f(0) \cdot \varphi(0)} \tag{1}$$

The gradation function (Eq. (2)) gives the luminance/radiance variation from horizon to zenith and the indicatrix function (Eq. (3)) expresses the decrease of luminance from the solar disc to sky patches far away from the sun:

$$\varphi(Z)/\varphi(0) = \frac{1 + a \cdot \exp(b/\cos Z)}{1 + a \cdot \exp b} \tag{2}$$

$$f(\chi)/f(0) = \frac{1 + c \cdot \left[\exp(d \cdot \chi) - \exp\left(d \cdot \frac{\pi}{2}\right) \right] + e \cdot \cos^2 \chi}{1 + c \cdot \left[\exp(d \cdot Z_s) - \exp\left(d \cdot \frac{\pi}{2}\right) \right] + e \cdot \cos^2 Z_s} \tag{3}$$

Z is the angle between sky zenith and the sky patch under scrutiny, (rad). Z_s is the angle between sky zenith and the Sun. Coefficients a, b, c, d, and e can be adapted to depict the 15 CIE sky conditions: five overcast, five partly cloudy, and five clear sky types, as shown in Table 1. χ is the scattering angle (rad) that represents the shortest distance from the sky patch to the solar disc and is calculated from Eq. (4):

$$\chi = \arccos(\cos Z_s \cdot \cos Z + \sin Z_s \cdot \sin Z \cdot \cos |\alpha - \alpha_s|), \tag{4}$$

where α is the azimuth angle of the sky patch and α_s is the solar azimuth.

Fig. 1 shows a sky image of three different sky conditions classified by the CIE Standard as: (a) Clear (IV.4 type); (b) partial (III.0.2); and, (c) cloudy (II.2).

The luminance distributions of individual standard skies were modelled and compared with the scanned sky luminance readings. The standard sky that was assigned had the lowest Mean-Square Error (RMSE), (Tregenza, 2004). The original criterion to define the sky type,

known as the Standard Sky Luminance Distribution method (SSLD) (Kittler et al., 1997), uses a theoretical assemblage of curves that represent the relation between the zenith luminance/diffuse luminance (L_z/D_v) ratio and the solar elevation angle. These curves converge at solar elevation values higher than 35°, making it difficult to apply this method in certain areas and at times when the solar elevation angle is higher than 35° (Li et al., 2013), which is the case at the location under study, especially in summer. Various procedures to circumvent this issue are proposed, using various methods of normalization (Li et al., 2014). In a previous paper (Suárez-García et al., 2018), the Normalization Ratio (NR) introduced by Littlefair (Littlefair and Paul, 1994; Littlefair, 1994) was used to obtain the CIE Standard sky types in Burgos, Spain.

3. Meteorological indices for sky classification

The sky classification by luminance distribution is reliable, but it has several restrictions (Li et al., 2014). The most significant problem is that luminance measurements are only available at a few sites in the world and only over short measurement periods. Alternatively, sky conditions can be evaluated using MIs calculated from meteorological variables readily accessible from most weather stations. In this study, different MIs traditionally used to classify the sky conditions were used to correlate the CIE Standard Skies determined by the luminance scan in terms of confusion matrices. The correlation can be easily interpreted and used to analyse the long-term sky conditions at sites that share similar climatic variables with the location under study. In the following paragraphs, the different MIs contemplated here are briefly described. For comparative purposes, a common classification is used in all cases, i.e., the CIE cloudiness categorization (cloudy, partial and clear, as shown in Table 1). As highlighted in Section 1, the definition of clear, partial, and cloudy sky conditions may be ambiguous, and the limits between these words are fuzzy. However, *cloudy sky* can be defined at one end of a continuous scale with *clear sky* at the other end.

The problem addressed in this work is the homogenization of the sky categories in the same number of classes without changing the original limits established by their authors. This follows a similar approach to that in (Gueymard et al., 2019) where the authors compared different MI models in an attempt to classify clear skies. Different intervals were tested, in order to mitigate the semantic effects and to maximize MI performance, keeping the class limits that the respective authors set in their original works. Fig. 2 shows the possible combinations of the limits of the intervals for the different MIs studied, keeping the ordinal relationship between cloudy, partial, and clear-sky conditions. For some of the MIs, no adaptation was needed. When more than three intervals were defined in the original work, all possible combinations for merging the intervals were tested. The adapted

Table 1
Parameters of CIE standard Sky types (Uetani et al., 2003).

	Type	a	b	c	d	e	Description
CLOUDY	I.1	4.0	-0.70	0	-1.0	0.00	Overcast with a steep gradation & azimuthal uniformity
	I.2	4.0	-0.70	2	-1.5	0.15	Overcast with a steep gradation & slight brightening toward the Sun
	II.1	1.1	-0.80	0	-1.0	0.00	Overcast with a moderate gradation & azimuthal uniformity
	II.2	1.1	-0.80	2	-1.5	0.15	Overcast with a moderate gradation & slight brightening toward the Sun
	III.1	0.0	-1.00	0	-1.0	0.00	Overcast, foggy or cloudy, with overall uniformity
PARTIAL	III.2	0.0	-1.00	2	-1.5	0.15	Partly cloudy with a uniform gradation & slight brightening toward the Sun
	III.3	0.0	-1.00	5	-2.5	0.30	Partly cloudy with a uniform gradation & a brighter circumsolar effect
	III.4	0.0	-1.00	10	-3.0	0.45	Partly cloudy, rather uniform with a clear solar corona
	IV.2	-1.0	-0.55	2	-1.5	0.15	Partly cloudy with a shaded sun position
	IV.3	-1.0	-0.55	5	-2.5	0.30	Partly cloudy with brighter circumsolar effect
CLEAR	IV.4	-1.0	-0.55	10	-3.0	0.45	White-blue sky with a clear solar corona
	V.4	-1.0	-0.32	10	-3.0	0.45	Very clear / unturbid with a clear solar corona
	V.5	-1.0	-0.32	16	-3.0	0.30	Cloudless polluted with a broader solar corona
	VI.5	-1.0	-0.15	16	-3.0	0.30	Cloudless turbid with a broader solar corona
	VI.6	-1.0	-0.15	24	-2.8	0.15	White-blue turbid sky with a wide solar corona effect

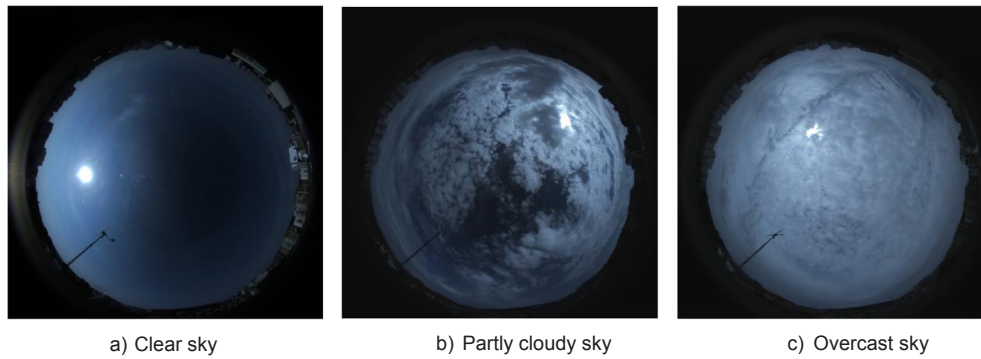


Fig. 1. Sky Image of different sky conditions classified by CIE Standard as (a) Clear (IV.4 type), (b) partial (III.2) and (c) cloudy (II.2), taken by a SONA201D All-Sky Camera-Day in Burgos, Spain, on 6/22/17 at 10:30 UTC, 3/11/2017 at 16:45 UTC and 8/23/2017 at 13:00 UTC, respectively.

interval is the one that maximized the classification metrics for each MI. The original and the adapted values of the intervals selected by each index are summarized in Table 2. The MIs are calculated using global horizontal irradiance, $G(0)$; diffuse horizontal irradiance, $D(0)$; and beam irradiance, $B(n)$.

3.1. Horizontal diffuse fraction, k_d

The horizontal diffuse fraction (or cloud ratio or cloudiness index or diffuse ratio), k_d (Eq. (5)), is defined as the ratio between the diffuse horizontal irradiance, $D(0)$, and the global horizontal irradiance, $G(0)$ (Erbs et al., 1982). k_d refers to the cloudiness of the sky and/or the turbidity of the atmosphere: the higher the proportion of the diffuse radiation in the global one, the higher the k_d (Kambezidis, 2018).

$$k_d = \frac{D(0)}{G(0)}. \tag{5}$$

Conversely, low values mean that the global radiation mainly

consists of the direct component that predominates under clear skies. The cloud ratio method classifies the sky condition into three skies based on the observation of the pattern of k_d graphs defining a sky as either clear or overcast when the value of k_d remains close to either 0 or 1, respectively, and as intermediate when k_d changes frequently and rapidly (Baharuddin et al., 2010). Different works have used the diffuse fraction for the classification of radiation data into three sky conditions. Hence, its value can be divided into three intervals from 0 (clear sky) to 1 (overcast sky). (Baharuddin et al., 2010; Kong and Kim, 2013; Li and Lam, 2001; Rahim et al., 2004). In this work, three equal intervals were applied assigning a cloudiness type to each one, which unifies the intervals used in the previously mentioned studies. Both, the original and the adapted intervals are shown in Table 2.

3.2. Horizontal direct fraction k_b

As an alternative to k_d , the horizontal direct fraction, k_b , is defined as the ratio between the direct horizontal irradiance, $B(0)$, and the

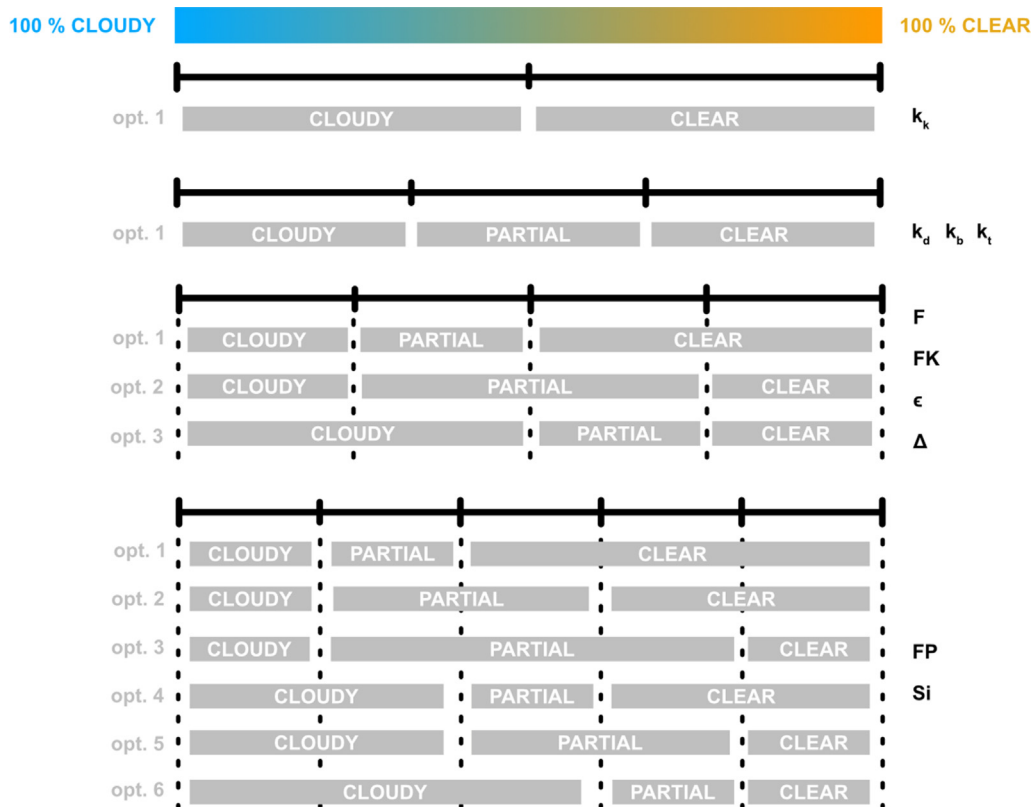


Fig. 2. Interval adaptation procedure of the original MIs.

Table 2

Summary of the MIs used to classify the skies, the original intervals used to define the clear, partial and cloudy sky conditions and the adapted ones used in this work.

Symbol	MI	Ref.	ORIGINAL	ADAPTED
k_d	Diffuse fraction	(Kong and Kim, 2013)	(0.00, 0.33] clear (0.33, 0.8) partial [0.8, 1) cloudy	(0.00, 0.33] clear (0.33, 0.8) partial [0.8, 1) cloudy
k_b	Direct fraction			[0.66, 1) clear (0.33, 0.66) partial (0, 0.33] cloudy
k_t	Clearness Index	(Wang et al., 2013)	[0.65, 1) clear (0.35, 0.65) partial (0, 0.35] cloudy	[0.65, 1) clear (0.65, 0.35) partial (0, 0.35] cloudy
F	Clearness Function	(Muneer, 2007)	[0.61, 1.00] completely clear [0.51, 0.61] clear [0.18, 0.51] partial [0.00, 0.18] completely cloudy	[0.51, 1.00] clear [0.18, 0.51] partial (0.00, 0.18) cloudy
k_k	Battles Clearness Index	(Battles et al., 2000)	$k_t > k_{kl}$ clear $k_d < k_k$ clear	$k_t > k_{kl}$ clear $k_d < k_k$ clear
FK	Klucher Clearness Index	(Klucher, 1979)	[0.61, 1.00] completely clear [0.51, 0.61] clear [0.18, 0.51] partial [0.00, 0.18] completely cloudy	[0.51, 1.00] clear [0.18, 0.51] partial (0.00, 0.18) cloudy
ϵ	Perez's clearness index	(Perez et al., 1990; Perez et al., 1987)	[6.20, ∞) completely clear [2.80, 6.20] clear [1.50, 2.80] partial [1.065, 1.50] cloudy [1.00, 1.065] completely cloudy	[2.4, ∞) clear [1.50, 2.4] partial [1.00, 1.50] cloudy
Δ	Sky brightness	(Perez et al., 1990; Perez et al., 1987)	[0.48, ∞) very bright [0.30, 0.48] bright [0.10, 0.30] partial [0.00, 0.10] very dark	[0.30, ∞) clear [0.10, 0.30] partial [0.00, 0.10] cloudy
FP	Nebulosity Index	(Kambezidis et al., 1998)	[0.90, 1.00] blue sky [0.70, 0.90] partial [0.20, 0.70] partially blue [0.05, 0.20] partially cloudy [0.00, 0.05] totally cloudy	[0.70, 1.00] clear [0.05, 0.70] partial [0.00, 0.05] cloudy
S_i	Igawa Index	(Igawa et al., 2004)	[1.70, ∞) clear [1.50, 1.70] almost clear [0.60, 1.50] partially clear [0.30, 0.60] partially cloudy [- ∞ , 0.30] totally cloudy	[1.70, ∞) clear (0.30, 1.70) partial [- ∞ , 0.30] cloudy

global horizontal irradiance, $G(0)$ as shown in Eq. (6). Hence, the direct horizontal irradiance can be calculated through the direct irradiance on a plane facing the Sun, $B(n)$, and the solar altitude angle, γ_s . As with k_b , a number of intervals can be defined between its minimum (0, overcast sky) and its maximum (1, clear sky) value. Three equal intervals were taken assigning a cloudiness type to each one of them as can be seen in Table 2. The pattern is the inverse of the one used for k_d , due to the complementarity of both indices ($k_b = 1 - k_d$).

$$k_b = \frac{B(0)}{G(0)} = \frac{B(n)\cos Z_s}{G(0)} = \frac{B(n)\sin \gamma_s}{G(0)} \tag{6}$$

3.3. Clearness index, k_t

The clearness index, k_t , (Iqbal, 1983) (Eq. (7)) is the ratio between the global horizontal irradiance, $G(0)$, and the extraterrestrial global horizontal irradiance, $G_{ext}(0)$.

$$k_t = \frac{G(0)}{G_{ext}(0)} = \frac{G(0)}{B_{sc}\epsilon_0\cos Z_s} \tag{7}$$

B_{sc} is the extraterrestrial irradiance constant (1361.1 W/m² (Gueymard, 2018; Gueymard and Ruiz-Arias, 2016)), ϵ_0 is the average value of the orbital eccentricity of the Earth, calculated from Eq. (8), and Z_s is the angle between sky zenith and sun.

$$\epsilon_0 = 1 + 0.033 \cdot \cos [2 \cdot \pi \cdot d_n / 365], \tag{8}$$

where d_n is the day the year. k_t has often been adopted to indicate the relative clearness of the atmosphere for sky categorization (Djafer et al., 2017; Escobedo et al., 2009; Kong and Kim, 2013; Wang et al., 2013)

and it indicates the percentage of solar irradiance that radiates through the atmosphere. (Mellit et al., 2008) suggested that the choice of k_t interval values would also differ from one site to another. In general, when the atmosphere is clear, a small fraction of the solar radiation is scattered, resulting in a predominance of direct sunlight yielding a high k_t reading. Under overcast skies, a large portion of the solar radiation is dispersed, so the main component is diffuse with a small k_t value. Several empirical relationships between k_d and k_t have been developed to calculate diffuse irradiation on horizontal and tilted surfaces from the global irradiation using a different time basis (Iqbal, 1983) that has been the object of numerous reviews (Khorasanizadeh and Mohammadi, 2016; Khorasanizadeh et al., 2016; Muneer et al., 2007; Tapakis et al., 2016; Torres et al., 2010a). As many categories as required may be generated for the indices k_d and k_b . The selected intervals are depicted in Table 2.

3.4. Clearness function F

The anisotropic sky-diffuse models use the horizontal diffuse fraction, k_d , and the clearness index, k_t , to describe the prevailing sky conditions. Under non-overcast conditions, the constituent components of sky-diffuse irradiance are a circumsolar (Sun's aureole) part and background diffuse irradiance. The sky clarity indices are used to relate the above-mentioned components (Muneer, 2007), the most common of which is the clearness function, F , defined by Eq. (9):

$$F = \frac{G(0) - B(0)}{G_{ext}(0)} = \frac{G(0) - B(n)\cos Z_s}{B_{sc}\epsilon_0\cos Z_s} \tag{9}$$

Muneer (2007) established four types of sky in function of the F

value that were reduced to three, which are used in this work, as shown in Table 2.

3.5. Batlles clearness index k_k

Two new indices, k_{it} and k_k , have been introduced by (Batlles et al., 2000), based on solar altitude, and are defined in Eqs. (10) and (11) as:

$$k_{it} = -0.3262 - 0.0032\gamma_s + 0.6843\log(\gamma_s), \quad (10)$$

$$k_k = 1.0827 - 0.3893\log(\gamma_s). \quad (11)$$

k_k index is restricted to clear skies, simultaneously defined by $k_t > k_{it}$ and $k_d < k_k$ (Muneer et al., 2004). This index therefore only distinguishes two categories of skies: clear and overcast. The original criteria were maintained for this study.

3.6. Klucher clearness index FK

Klucher (1979) proposed a model for estimating irradiance on a tilted surface and developed the function FK , depending on $D(0)$, and $G(0)$, as defined by Eq. (12). The original and the adapted classification of cloudiness using this index, shown in Table 2, is equal to the one established in the clearness function F .

$$FK = 1 - \left(\frac{D(0)}{G(0)} \right)^2 = 1 - k_d^2. \quad (12)$$

3.7. Perez's clearness indices ϵ and Δ

Eqs. (13) and (14) introduce the sky clearness index, ϵ , and the brightness factor, Δ , respectively, both defined by Perez in 1987 (Perez et al., 1987) and revised in 1990 (Perez et al., 1990). ϵ predicts cloud conditions using the ratio between the diffuse horizontal irradiance and the direct one on the same plane. Δ quantifies cloud thickness or aerosol loading.

$$\epsilon = \frac{\frac{D(0)+B(n)}{D(0)} + kZ_s^3}{1 + kZ_s^3}, \quad (13)$$

$$\Delta = \frac{mD(0)}{B_{ext}(n)}. \quad (14)$$

Z_s is the angle between sky zenith and sun (rad) and $k = 1.04$ (or $k = 5.53 \cdot 10^{-6}$ if Z_s is expressed in degrees). $B_{ext}(n)$ is the extraterrestrial direct irradiance ($B_{ext}(n) = B_{sc}\epsilon_0$) and m is the optical air mass calculated using the Kasten model (Kasten, 1993). Classification of the cloudiness sky types using ϵ and Δ are shown in Table 2.

3.8. Perraudau nebulosity index, FP

Derived from the original work of Perraudau and Chauvel (1986), the nebulosity index, FP , is defined by Eq. (15) (Kambeizidis et al., 1998). This index models the degree of the sky covered by clouds using the diffuse horizontal fraction, k_d , and the diffuse horizontal fraction for a clear sky, k_{d0} , given by Eq. (16).

$$FP = \frac{1 - k_d}{1 - k_{d0}}, \quad (15)$$

$$k_{d0} = \frac{G_{clear}(0)}{G_{clear}(0) + B(n)} \quad (16)$$

Here $G_{clear}(0)$, given by Eq. (17), is the clear-sky irradiance and $B(n)$ is the beam irradiance.

$$G_{clear}(0) = (0.5528 + 0.8785 \cdot \gamma - 0.01322 \cdot \gamma^2 + 0.0003434 \cdot \gamma^3) \cdot (6.9731 + 0.042496 \cdot \gamma - 8.5275 \cdot 10^{-4} \cdot \gamma^2 - 8.6088 \cdot 10^{-5} \cdot \gamma^3 + 1.984 \cdot 10^{-6} \cdot \gamma^4 - 1.6222 \cdot 10^{-8} \cdot \gamma^5 + 4.7823 \cdot 10^{-11} \cdot \gamma^6) \quad (17)$$

Depending on the FP value, five sky types can be identified that are shown in Table 2.

3.9. Igawa index Si

The Igawa Index (Igawa et al., 2004), Si , defined by Eq. (18), uses the standard global irradiance, G_{st} , the cloudless index, Cle , and the standard cloudiness fraction, Ces , calculated from Eqs. (18)–(21).

$$Si = \frac{G(0)}{G_{st}} + \sqrt{Cle}, \quad (18)$$

$$G_{st} = 0.84 \frac{B_{sc}}{m} e^{-0.0657m}, \quad (19)$$

$$Cle = \frac{1 - k_d}{1 - Ces}, \quad (20)$$

$$Ces = 0.01299 + 0.07698m - 0.003857m^2 + 0.0001054m^3 - 0.000001031m^4 \quad (21)$$

G_{st} is calculated considering the Linke turbidity factor as 2.5 in the Kasten model (Kasten, 1993) of global irradiance for a clear sky; Ces is given by a polynomial fit adjustment using the optical air mass as the independent variable and B_{sc} is the extraterrestrial irradiance constant (1361.1 W/m^2). The original criterion used by Igawa and the adapted one established following the procedure described by Fig. 2 (Section 3) are shown in Table 2.

4. Confusion matrices

In the field of machine learning, a confusion matrix is used for measuring the performance of classification algorithms. Classification is known as a supervised learning approach, because the machine is trained with selected examples of the data from which it is said to learn. After the learning process, the model that has been tuned by the data inputs will classify a new observation or sample into a given number of classes. The classification can be binary (e.g. a dichotomous classification of email as either spam or not spam) or multi-class (e.g. recognition of handwritten numbers). When the algorithm attempts to classify a sample, it processes an individual measurable property or feature. For example, in binary classification the algorithm attempts to detect the presence of a relevant feature. If it is detected, the sample is labelled as “positive” and, if otherwise, “negative”, (Fig. 3). Comparing the prediction with reality, there are four possible scenarios: the predicted positive will agree with the actual one (True Positive or TP), the predicted positive will not agree with the actual one (False Positive or FP), the predicted negative will agree with the actual one (True Negative or TN) and the predicted negative will not agree with the real one (False Negative or FN). Each row of the matrix represents the instances of the predicted class, while each column represents instances of the reference class (Powers, 2011). The matrix was so-named, because it easily visualizes whether the algorithm is confusing or mislabelling two classes.

In this present work, the reference labels are established by the CIE methodology for sky classification and the predicted labels by the MIs. Hence, the latter are the algorithms or predictive models under analysis. As mentioned above, the CIE defines fifteen types of sky that can be grouped into three types of cloudiness: clear, partial, or cloudy. The MIs distinguish between the features in an attempt to define these three classes. This task can be analyzed as a multiclass problem that can be decomposed as a multiple dichotomous classification where each

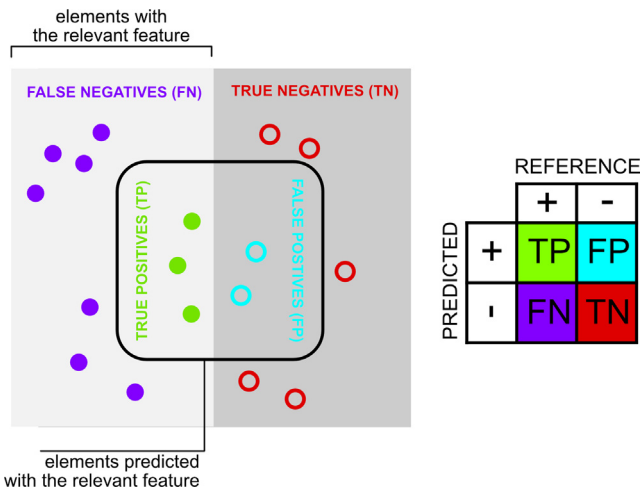


Fig. 3. Confusion matrix: possible scenarios in the comparison of the prediction to the actual data.

cloudiness categorization is predicted against the other remaining ones (Fig. 4). At the end of the process, the dichotomous classification of the MIs is evaluated using several performance ratings explained in the following sections.

4.1. Accuracy index

The Accuracy Index, also known as the Simple Matching Coefficient, represents the ratio of correct predictions, positive or negative, amongst all the cases evaluated by the algorithm expressed by Eq. (22):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{22}$$

As can be appreciated, it represents the overall portion of agreement and it is usually the starting point for analyzing the quality of a predictive model. The main weakness of the so-called Accuracy Paradox is that models of a given accuracy may have greater predictive power than others of higher accuracy (Kundel and Polansky, 2003). An illustration of this fact would be a comparison of two different algorithms designed

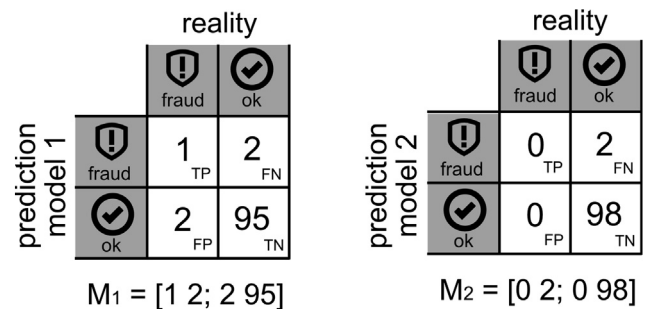


Fig. 5. Confusion matrices for a fraud detection example.

to detect insurance frauds for the same set of 100 samples, represented in Fig. 5. Given a confusion matrix (with notation [TP FP; FN TN]) for the first one of $M_1 = [12; 295]$ and for the second one of $M_2 = [02; 098]$, then the accuracy of the first and the second algorithm would be 0.95 and 0.98, respectively. In comparison with the first model, the second model would show fewer incorrect predictions and improved accuracy, globally; however, its fraud detection power would be weaker. This paradox arises in imbalanced data where there is a substantial difference in the size of the categories. The use of other metrics may therefore be advisable.

In multiclass classification performance, there are two types of averages: micro-averages and macro-averages. The micro-average sums up all the TP, TN, FP, and FN cases of all the classes, aggregating their contributions before calculating the metrics. The result would be dominated by the performance of the common categories and it would be similar to weighting it against the population size of each category. A bad performance labelling one of the classes could be masked. The macro-average - the average type used in this work - is computed by averaging the metrics after their calculation. Therefore, equal weight was given to each category, regardless of its frequency. Using the latter, the classification performance must be good in all of the (cloudy, partial and clear) classes, if it is to be among the highest positions of the metric ranking.

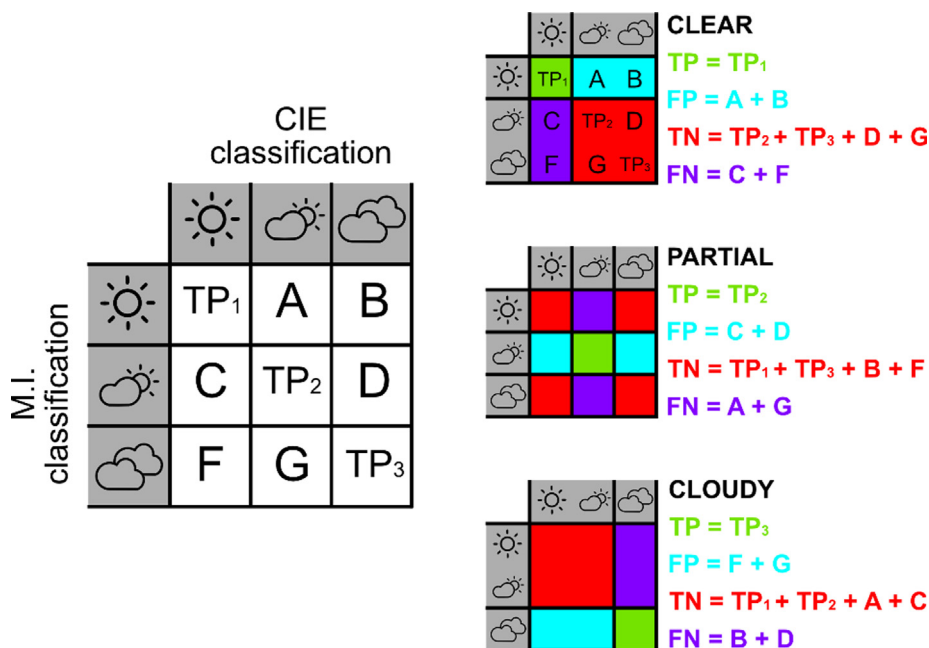


Fig. 4. Multiclass to dichotomous transformation.

4.2. Jaccard index

The Jaccard Index (Jaccard, 1912) expressed by Eq. (23), measures the similarity between the number of true positives in all the cases under evaluation. The main difference to the Accuracy Index is the omission of TN , meaning that the Jaccard Index only takes into account successful detection of the relevant attribute. In the above insurance fraud example, the Jaccard Index is 0.20 for M_1 and 0 for M_2 , showing a better reflection of the predictive power of the first algorithm.

$$Jaccard \text{ Index} = \frac{TP}{TP + FP + FN}. \tag{23}$$

The Jaccard Index is focused on the detection of the relevant feature or true positives. It makes no distinction between FP and TN. There are some scenarios where such a distinction is crucial. For example, in medical diagnostics, a model with a confusion matrix [90 10; 0 0] and another characterized by [90 0; 10 0] would have the same Jaccard Index of 0.9. However, the first model would have 10 FP cases and the second one 10 FN. The second model would have harmful consequences, so the use of the first one would be recommendable. Depending on the field of application, the use of another metric or full information would be recommendable before taking the final decision.

4.3. Cohen's Kappa

Cohen's Kappa, κ , is a measure of true agreement rather than a true prediction that reflects the possibility that the algorithm agrees with the reference by chance. It indicates the proportion of agreement beyond that expected by chance, as shown in Eq. (24):

$$\kappa = \frac{\text{observed agreement} - \text{chance agreement}}{1 - \text{chance agreement}} = \frac{p_o - p_e}{1 - p_e}, \tag{24}$$

where p_o is the overall agreement given by Eq. (25); p_e is the expected agreement by chance as calculated by Eq. (26), $1 - p_e$ is the fraction of cases on which agreement is not expected to occur by chance, and N (Eq. (27)) is the total number of cases analyzed by the algorithm.

$$p_o = \frac{TP + TN}{N}, \tag{25}$$

$$p_e = \left(\frac{TP + FP}{N}\right)\left(\frac{TP + FN}{N}\right) + \left(\frac{TN + FP}{N}\right)\left(\frac{TN + FN}{N}\right), \tag{26}$$

$$N = TP + TN + FP + FN. \tag{27}$$

Table 3 shows the strength of agreement for various ranges of κ suggested by (Landis and Koch, 1977). The choice of intervals is arbitrary, but is now in wide use (Kundel and Polansky, 2003). In the previous example, Cohen's Kappa scored 0.313 (fair strength) for M_1 and 0 (poor strength) for M_2 , avoiding the Accuracy Paradox.

Cohen's Kappa takes into account the four classes of the confusion matrix and is more informative than other confusion-matrix measures. It can produce drastically different results to Accuracy and the Jaccard Index in scenarios where the positives cases are the predominant class. For example, a classification algorithm that produces the confusion matrix $M_3 = [905; 41]$ would produce Accuracy 0.91, Jaccard Index 0.91, and κ 0.13. Nevertheless, the Jaccard Index and Accuracy will provide indicative assessments of almost perfect predictive power and κ will predict a performance near a random classifier with null predictive power, because of its incapacity to detect true negatives (Sim and Wright, 2005).

4.4. Matthews correlation coefficient

The Matthews Correlation Coefficient (Matthews, 1975) given by Eq. (28) is a measure of the quality of the binary classification:

$$Matthews \text{ Correlation} = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FN)(TP + FP)(TN + FP)(TN + FN)}}. \tag{28}$$

Essentially, it is a correlation coefficient between the reference and predicted classification returning a value between +1 (perfect prediction) and -1 (total disagreement). The 0 value is representative of a random prediction. Returning again to the Accuracy Paradox example, the Matthews Correlation is 0.31 for M_1 and 0 for M_2 . As with Cohen's Kappa, the Matthews Correlation takes into account the four classes of the confusion matrix with the same implications as those seen in the Section 4.3 (Chicco, 2017).

4.5. Combination of the confusion matrix metrics

In the preceding sections, various metrics for a classification algorithm have been explained. All of them are attempts to sum up the confusion matrix associated with the algorithm using only one number. Inevitably, the process is associated with a loss of information, because a four dimensional matrix is collapsed into one number. Each dimension attempts to highlight one aspect of interest. However, they are correlated in some way because the starting data are all the same. In fact, the ranking offered by each one is very similar.

The different rankings produced by all the metrics are combined to arrive at the best MI for classifying sky cloudiness. At the end of the process, four rankings are thus obtained, one per confusion matrix. The one designated as best model holds the best positions in all of them. So, by using all of them simultaneously, the deficiencies of the metric and its biases are avoided.

5. Experimental section

5.1. The meteorological facility

The experimental data for this study were gathered at a meteorological weather station located on the roof of the Higher Polytechnic School building at Burgos University (42°21'04"N; 3°41'20"O; 856 m above mean sea level). This five-storey building is in an area with no other buildings of comparable height, free from any external obstructions. The experimental equipment is shown in Fig. 6. The following meteorological data were measured: temperature, wind velocity and direction, atmospheric pressure, humidity, and rainfall. Global, beam, and diffuse horizontal irradiation ($G(0)$, $B(n)$, $D(0)$) were all recorded using first-class Hukseflux SR11 pyranometers and a Hukseflux DR01 pyrliometer. The facility includes a SONA201D All-Sky Camera-Day, from Sieltec Canarias and an MS-321LR sky scanner, from EKO. The experimental variables were recorded with a CAMPBELL CR3000 datalogger. Tables 4–6, respectively, show the technical specifications of the sky scanner, the pyranometers, and the pyrliometer used for this study.

5.2. Data processing

Global, beam, and diffuse irradiance data are measured from June 2016 to May 2017, for the classification and benchmarking of the

Table 3
Strength of agreement indicated with κ values.

κ	strength
$(-\infty, 0]$	poor
$[0, 0.20)$	slight
$[0.20, 0.40)$	fair
$[0.40, 0.60)$	moderate
$[0.60, 0.80)$	substantial
$[0.80, 1.00]$	almost perfect

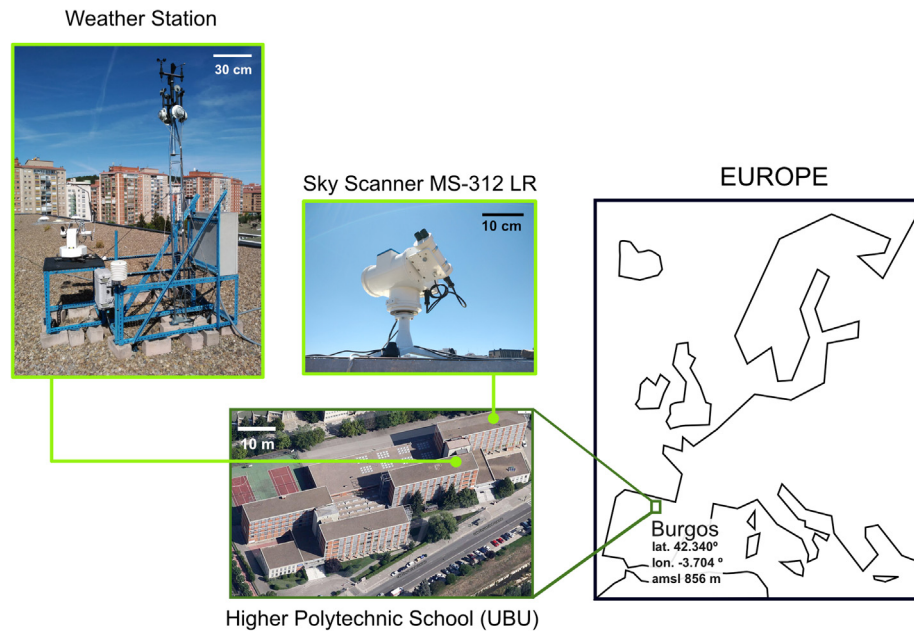


Fig. 6. Location of the meteorological station on the roof of the Higher Polytechnic School building at the University of Burgos, Spain.

Table 4
Sky Scanner specifications.

Model	MS-321LR EKO Instruments
Dimensions (W × D × H)	430 mm × 380 mm × 440 mm
Mass	12.5 kg
Aperture	11 °
Luminance	0 to 50 kcd/m ²
Radiance	0 to 300 W/m ² /sr
A/D Converter	16 bit
Calibration Error	2%

Table 5
Pyranometer specifications.

Model	SR11
ISO classification	first class
Spectral range	300 to 2800 nm
Irradiance range	0 to 2000 W/m ²
Sensitivity	15 × 10 ⁻⁶ V/(Wm ⁻²)
Calibration uncertainty	< 1.8%

Table 6
Pyrheliometer specifications.

Model	DR01
ISO classification	first class
Spectral range	200 to 4000 nm
Irradiance range	0 to 2000 W/m ²
Sensitivity	7-15 × 10 ⁻⁶ V/(Wm ⁻²)
Calibration uncertainty	< 0.3%

different MIs selected for this study. Both the pyranometers and the pyrheliometer are classified as “first class” in the ISO classification (ISO 9060:1990) with a WMO performance level that is of “good quality” (WMO, 2010). The calibration and the management of the meteorological facility is done following ISO (1992) and the WMO Guide to Meteorological Instruments and Methods of Observation (WMO, 2008. (Updated 2010)). Global, diffuse, and beam irradiation data are recorded every ten minutes (averaging recorded scans of thirty seconds). The sky-scanner completes a full scan in four minutes and starts a new

scan every fifteen minutes. So as to match simultaneous records of global, diffuse, and beam irradiation, only half-hourly and hourly sky-scanner measurements are used in this study. The irradiance data are analysed and filtered using traditional quality criteria (Gueymard and Ruiz-Arias, 2016). If irradiance data (global, diffuse, and beam irradiance data) fail to pass the quality criteria, then the three simultaneous data sets are rejected.

The number of indices, once the data from both sources had passed the quality criteria, was above 300 per month, as shown in Fig. 7. In the summer months there were fewer estimated indices than expected, because the weather station and Sky Scanner experienced several shutdowns due to servicing works. Overall, over 3600 indices were correlated with their respective sky cloudiness.

5.3. CIE standard classification of skies in Burgos, Spain

The sky scanner divides the sky into 145 patches or sectors (*p*) that cover the whole dome. The sectors are grouped into eight bands, named *b_p*, and by their solar altitude, $\gamma = (\frac{\pi}{2} - Z)$, where *Z* is the zenith angle. Fig. 8 shows the location of the sectors in the whole dome. A luminance measurement (kcd/m²) of each patch is taken four times per hour. Half-hourly and hourly measurements were recorded between June 2016 and May 2017. Continuous scanning yielded luminance data corresponding to the 145 patches (see Fig. 8) recommended for the CIE in the Guide to Daylight Measurements (Tregenza et al., 1994), which were measured and registered. Likewise, the luminance corresponding to

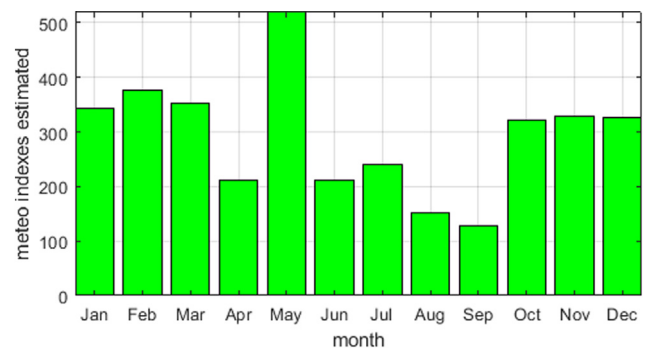


Fig. 7. Number of sets of the MIs estimated per month.

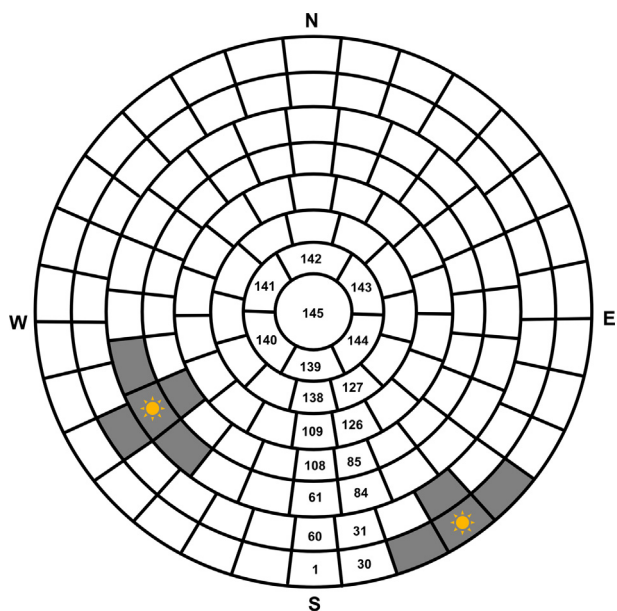


Fig. 8. Sky divided into 145 sectors (p) grouped into 8 bands (b_p). The number of patches per band (n_b) is shown in the figure. Figure shows two fictitious cases illustrating how, depending on sun position, the patches adjacent to the position of the sun are excluded for the luminance calculation.

each of the commonly considered 15 standard sky types presented in Table 1 was calculated at the same time and for the same 145 patches. The standard sky type ascribed to each record expressed the lowest RMSD (Root Mean Square Deviation) between the 145 normalized luminance values that were measured and calculated.

The sky scanner is adjusted each month to measure daylight hours from sunrise to sunset. The first and last measurement of the day (solar elevation angle equal or lower than 5°) were discarded, as were measurements higher than 50 kcd/m^2 or lower than 0.1 kcd/m^2 , following the specification of the equipment. Fig. 9 gives the frequency of occurrence (FOC) of the sky classification results by the method. One previous work (Suárez-García et al., 2018) described the details of the sky luminance measurement and classification methodology. All sky types of the CIE classification, shown in Table 1, can be found in Burgos, from overcast to very clear. The lowest frequency is for type I.2, corresponding to Overcast with the steep gradation and slight brightening toward the sun (1.7%), and the highest frequency is for type V.5. (18.3%) (Cloudless polluted with a broader solar corona). Cloudiness labelling was done with the CIE sky types: I.1 to III.1 classified as cloudy, III.2 to IV.3 as partially cloudy, and IV.4 to VI.6 as clear skies. These three categories reflect the characteristically clear skies that are predominant in Burgos (56.3%), while cloudy skies are presented in 23.7% and partially cloudy in 20% of cases, as shown in Fig. 10.

6. Results and discussion

6.1. Results of cloudiness classification from the meteorological indices

In this section, the results of the calculation of each MI used for sky classification are presented. The calculated MIs have values inside the intervals defined by their authors (Table 2). The box plots, presented in Fig. 11, show the data organized by quartiles. The high variability of their interquartile ranges (IQR) can be observed: k_t , k_b , k_d , FK , ϵ and Si . It might be thought that a wider dispersion of the values would mean a higher number of different cloudiness classifications. As will be seen later (Table 8), ϵ , FK , and k_t are ranked first, third, and fourth in the cloudiness classification. However, the second best single index (FP) had an IQR of low variance. The explanation could be a better adaption of its intervals originally established by the CIE types of cloudiness.

The results of each MIs for the cloudiness classification are summarized in Fig. 12 and presented by month. Fig. 13 shows the monthly cloudiness classification obtained from the standard CIE.

Individual analysis of sky classification with the MIs shows that practically no index is able to identify the high percentage of clear skies indicated by the CIE classification. The Klucher clearness index (FK) is the one that identifies a higher percentage of clear skies, and at the opposite extreme, the Battles clearness index identifies practically all cases as overcast. Δ , Si , FP , and F , identify most of the cases as partial cloudiness. The stacked bar graph shows an imperfect match between the different approaches. A frequency graph would be insufficient for a complete comparison: the energy contribution of the Sun depends on its elevation above the horizon, which varies according to the time of day and day of year. It is therefore necessary to compare the classification at each timestamp.

6.2. Confusion matrices

Finally, the comparison between classifications at each instant is obtained and presented in terms of the confusion matrices and the rating variables: Accuracy, κ , Jaccard, and Matthews. The overall power prediction of the MIs is shown using the data for all the months. Then, their estimations are compared with the CIE classification for all the samples that were registered. The maximum percentage of true positives in each sky type is limited by the CIE classification shown in Fig. 10. For this reason, an ideal algorithm with no errors in its classification would obtain the following classification matrices, expressed as percentages: $M_{clear} = [570; 043]$, $M_{partial} = [240; 076]$ and $M_{cloudy} = [190; 081]$. In the matrices, the values are the maximum true positives and true negatives that the algorithms can obtain. In Fig. 14 the confusion matrices for each MI used as a sky cloudiness classifier are presented. Percentages of true positives (tp), true negatives (tn), false positives (fp) and false negatives (fn) are depicted. The lowercase abbreviation is used to distinguish the percentage of the raw recount where uppercase is preferred. To give as much information as possible, the number of samples analyzed, N , is detailed in the title of each figure. For example, the confusion matrix for the k_d detecting clear skies is $M_{clear, k_d} = [1524; 4318]$ in percentage terms or $M_{clear, k_d} = [330527; 945395]$ in samples counting.

A predominant behavior can be observed in each one of the sky types. The false negatives are predominant in the clear skies (Fig. 14). There is a considerable number of cases where the MIs predict a non-clear sky (i.e. partial or cloudy) when the CIE model predicts a clear sky. Better than all of the other MIs, FK stands out from the rest with a true positive ratio near 45%, ten percentage points higher than the rest.

The behavior of the MIs slightly changes in the partly cloudy skies classification shown in Fig. 14, where the false positives and true

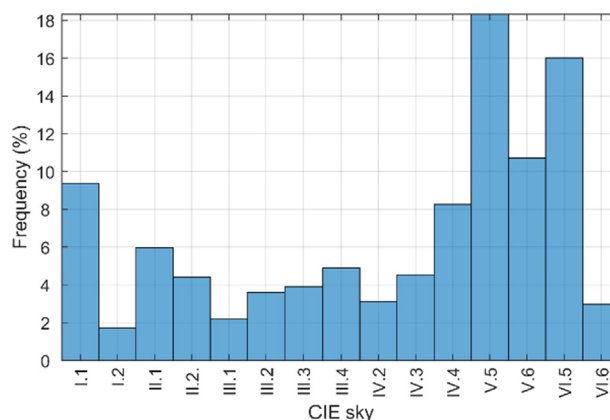


Fig. 9. Frequency of occurrence (FOC) from June 2006 to May 2017 Burgos CIE Standard Skies.

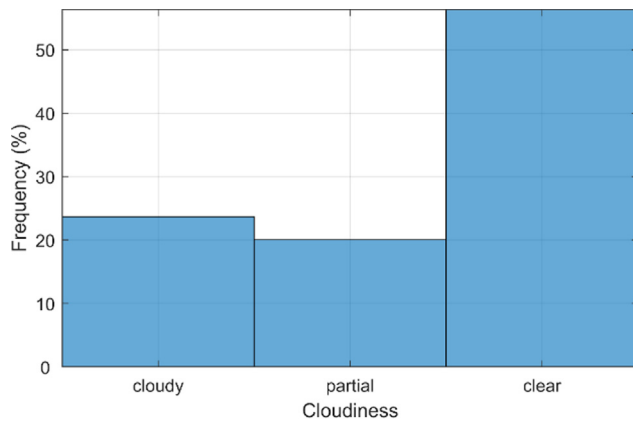


Fig. 10. Comparative study of cloudiness classifications in Burgos.

negatives are predominant. Clear and cloudy skies are clearly distinguished. However, MIs can hardly classify a partial cloudiness sky correctly. Finally, the true negatives predominate in the cloudy scenarios, with the rest of cases evenly spread out between the true positives, false positives, and false negatives. In both cloudiness scenarios, partial and cloudy skies, the true positive ratio is below 20%.

The confusion matrix metrics, shown in Table 7, summarize the performance of the MI classifications. The results are discussed in two groups: the indices that characterize similarity (Accuracy and Jaccard) and the indices that analyze randomness in the agreement (Cohen Kappa and Matthews Correlation). The results obtained by Accuracy, in some cases, show values, above 80%, such as the classification of cloudy skies by the FP and the Si indices, and the average value was above 60% in five indices (FK , FP , Si , k_i , and k_b). Nonetheless, these

high values are reached, because the true negative ratios value are included in the accuracy computation. In the present study, where there are three classes (clear, partial, and cloudy), a true negative obtained in one class implies that the sample could be in any of the other two. The Jaccard index shows an average of the meteorological classifiers below 35%. This considerable reduction is because the Jaccard Index only measures the similarity of the true positives, omitting the true negatives. The Jaccard index emphasizes the scoring capability of the algorithm to detect the desired value of the attribute under analysis. The Jaccard Index is nevertheless 45% when the Klucher Clearness classifies clear skies. Cohen’s Kappa and Matthews Correlation Coefficient both show a fair strength of agreement for most of the classifiers, as shown in Table 7. Poor performance under partly cloudy skies is also very visible from the values, below 10% obtained for both classifiers. The results obtained in the cloudy skies are the best ones, followed by the clear ones. In other words, the extreme situations are those that are detected with best accuracy.

Table 8 summarizes the average ranking obtained by the MIs. ϵ was in first place in the Jaccard Index, κ and Matthews evaluations. It was only surpassed by the FK index in the Accuracy ranking. Bearing in mind the similarity of the numerical values obtained by ϵ , FK and FP in the evaluation indices, these three MIs may be identified as the best classifiers of sky cloudiness. When studying only the capacity of the MIs to identify clear skies, FK obtained first position in all the indicators. This result was already shown in Fig. 12. None of the MIs under analysis have simultaneously shown good results in the rating variables, for the identification of either partial or cloudy skies,.

7. Conclusions

A key aspect in the modelling of solar radiation and daylighting is

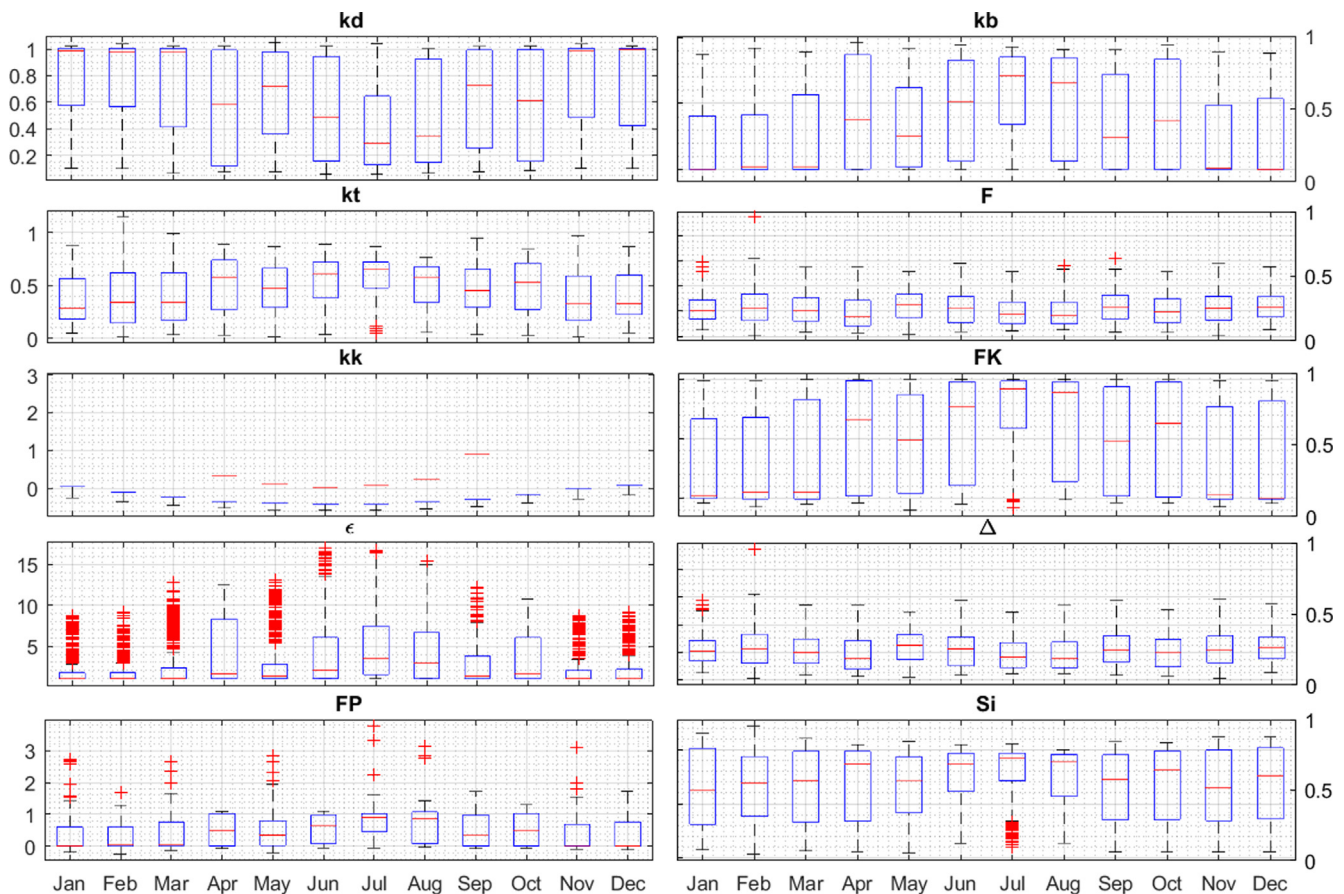


Fig. 11. MIs calculated from the experimental data of global, beam, and diffuse irradiance measured in Burgos from June 2016 to May 2017.

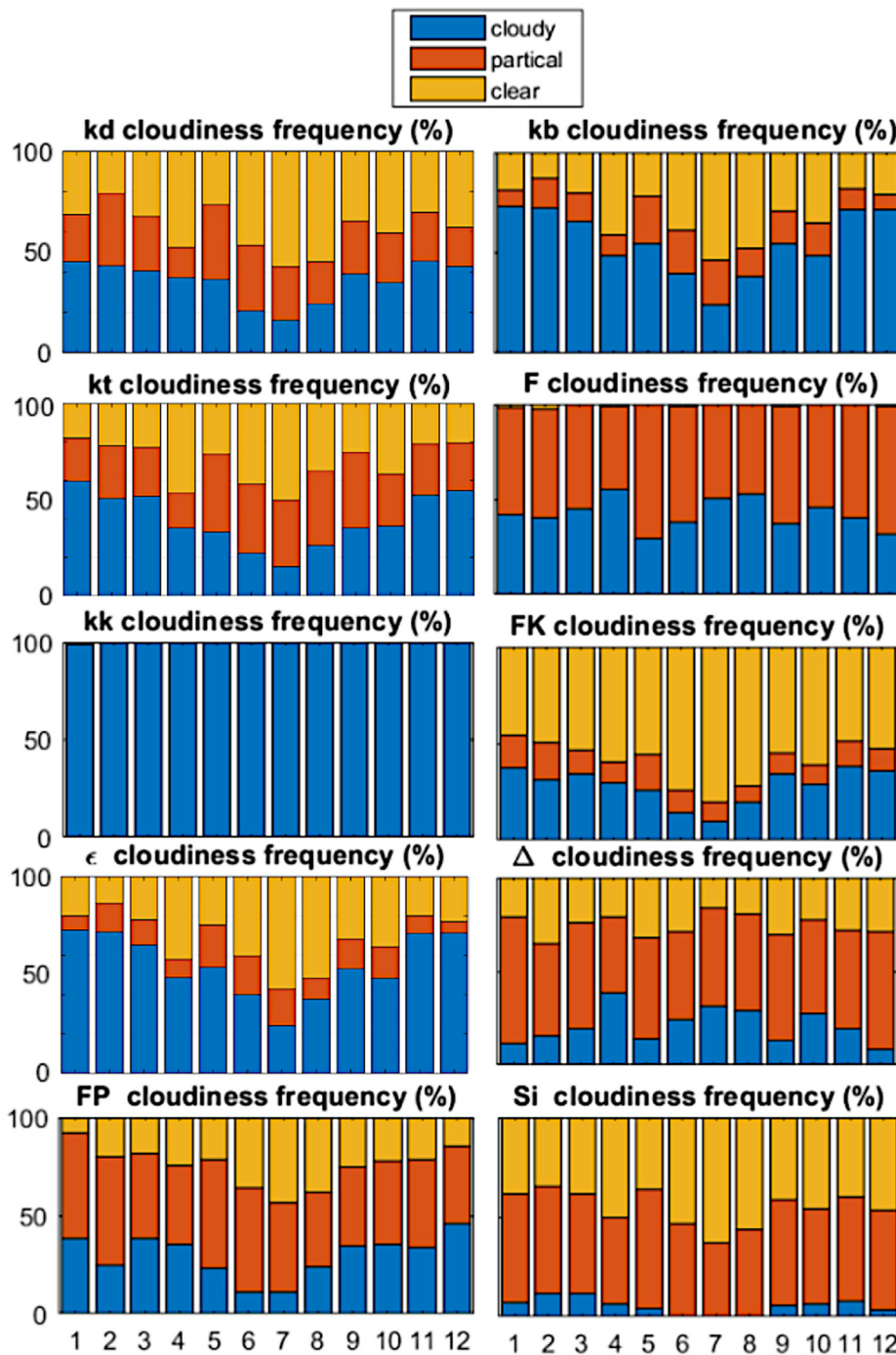


Fig. 12. Monthly cloudiness classification obtained from each meteorological index calculated from the experimental data of global, beam, and diffuse irradiance measured in Burgos from June 2016 to May 2017.

sky classification. The use of MIs as an alternative to the CIE standard classification has been investigated in this work. Various MIs for sky classification have been reviewed. Ten classical MIs obtained from global, diffuse, and/or beam irradiance measurements, recorded at weather stations have been used. The baseline reference in the one-year experimental measurement campaign was the standard CIE classification for homogenous skies. Only three cloudiness sky categories have been established: clear, partial and cloudy. This homogenization has required scaling down the number of intervals of some MIs, keeping the class limits stated in the original works. An exhaustive analysis of all the possibilities has been made to merge the intervals, always respecting the original limits and selecting the option that optimized the result of

each MI. Confusion matrices have been used as a tool for benchmarking the MIs under study and four different quality metrics have been calculated, to determine the performance of each MI as a sky classifier.

Although the classification of the skies was limited to three categories, none of the MIs under analysis distinguished the large global number of clear skies that the CIE determined in the city of Burgos; the *FK* index was the closest to this classification, in terms of frequency of occurrence of each of the sky categories. The rest of the indices, in general, indicated partially or fully covered skies as the predominant ones in the area under study. The monthly distribution over the year-long analysis of the types of sky calculated by the different indices also differed from that offered by the CIE standard, which predicted

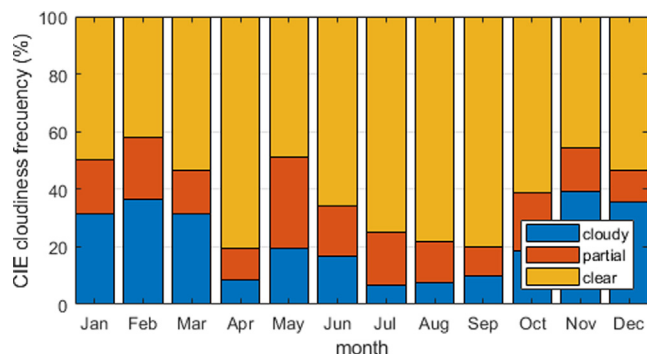


Fig. 13. CIE monthly cloudiness classification.

predominantly clear skies in summer.

The metrics for the cloudiness classification have shown that the performance of the MIs may, at best, be considered as “fair”. Classification accuracy reached values of over 70% for cloudy skies. However, this metric can lead to misunderstandings, as it accounts for the correct classification of true positives and true negatives (see the Accuracy Paradox in Section 4.1). As Table 8 shows, the best MI for the sky classification was ϵ followed closely by FK and FP indices. k_t and F showed the worst results for the classification of cloudiness of the skies into three categories. No correlation was observed between the mathematical complexity of the MIs or the number of input parameters and the classification result of the MIs. As can be observed, the original definition of the intervals is one of the main factors that justifies the difference in the performance of the MIs.

The experimental data of the present study were taken from a different location other than that used in the original studies for the definition of each MI. However, if subtle language differences are neglected, the definition of “clear” or “cloudy” sky should be independent of the location. Attending to the metrics, regardless of the accuracy, the results were quite similar. The average accuracy almost reached 70% in FK , and was above 50% in all the MIs. However, this result is

overoptimistic because of the accuracy paradox. The average value was below 33% for the Jaccard Index and below 25% for Cohen’s Kappa and Matthews correlation coefficient. In short, the capability of the MIs at classifying sky cloudiness can be considered “fair”. As demonstrated, their isolated use would not be advisable for the classification of sky cloudiness.

Several possible lines may be followed to improve on the results that have been presented here. Future work must address different combinations of these indices linked with other variables related to climatic and atmospheric conditions, which were not considered in the calculation of the indices. Another option would be to modify the original interval limits of each sky type to obtain the best classification results. The use of machine learning techniques should also be considered, to obtain models that will help to classify day cloudiness in real time. The final objective must be to reuse the big data available from meteorological stations, so as to classify cloudiness in the most reliable manner without the use of expensive sky scanner devices.

Declaration of Competing Interest

Author declares that there is no conflicts of interest.

Acknowledgments

The authors gratefully acknowledge the financial support provided by the Regional Government of Castilla y León under the “Support Program for Recognized Research Groups of Public Universities of Castilla y León” (ORDEN EDU/667/2019) and the Spanish Ministry of Science, Innovation & Universities under the I + D + i state programme “Challenges Research Projects” (Ref. RTI2018-098900-B-I00). David González Peña and Diego Granados López express their thanks to Junta de Castilla-León for economic support (PIRTU Program, ORDEN EDU/301/2015 and ORDEN EDU/556/2019, respectively). Also, Andrés Suárez García, is grateful to the International Excellence Triangular Campus-E3 for the financed stay under the programme managed centrally by the Ministry of Education, Culture & Sports (BOE

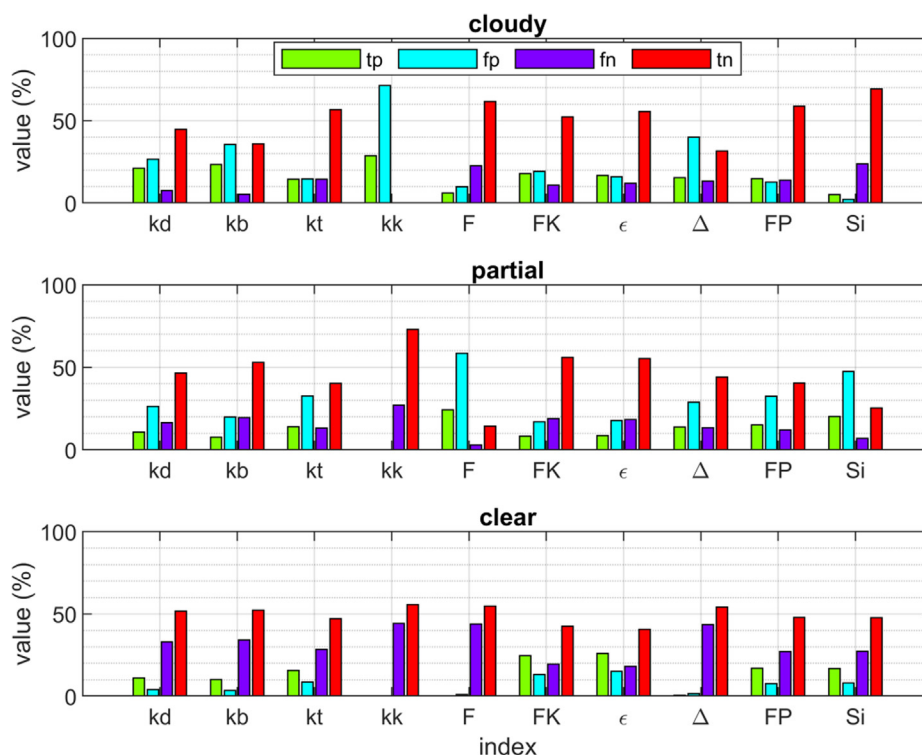


Fig. 14. Results of the Confusion matrices for cloudy skies (N = 2197), partly cloudy skies (N = 2197), and clear skies (N = 2197).

Table 7

Results of the confusion matrix indicators obtained by each of the MIs studied for the different cloudiness categories and on average by all cloudiness classification methods.

	MI	k_d	k_b	k_t	F	kk	FK	ϵ	Δ	FP	Si
ACCURACY	Av.	0.6190	0.6077	0.6267	0.5375	0.5242	0.6779	0.6758	0.5397	0.6474	0.6147
	cloudy	0.6579	0.5926	0.7105	0.6763	0.2863	0.7000	0.7211	0.7205	0.7358	<u>0.7421</u>
	partial	0.5711	0.6063	0.5421	0.3858	<u>0.7284</u>	0.6658	0.6389	0.4511	0.5553	0.4558
	clear	0.6279	0.6242	0.6274	0.5505	0.5579	<u>0.6679</u>	0.6674	0.4474	0.6511	0.6463
JACCARD	Av.	0.2701	0.2465	0.2871	0.1494	0.0954	0.3268	0.3357	0.1596	0.3143	0.2525
	cloudy	<u>0.3810</u>	0.3651	0.3309	0.1575	0.2863	0.3722	0.3735	0.0635	0.3597	0.1624
	partial	0.1994	0.1624	0.2342	<u>0.2827</u>	0.0000	0.1511	0.1948	0.1813	0.2542	0.2713
	clear	0.2298	0.2119	0.2962	0.0081	0.0000	<u>0.4570</u>	0.4387	0.2341	0.3289	0.3239
COHEN KAPPA, κ	Av.	0.1734	0.1436	0.1856	0.0423	0.0000	0.2366	0.2482	-0.0431	0.2318	0.1664
	cloudy	0.3020	0.2431	0.2940	0.0857	0.0000	0.3247	0.3442	0.0671	<u>0.3456</u>	0.1862
	partial	0.0275	0.0086	0.0551	0.0541	0.0000	0.0573	0.0796	-0.0777	<u>0.0909</u>	0.0638
	clear	0.1906	0.1792	0.2078	-0.0128	0.0000	<u>0.3278</u>	0.3207	-0.1186	0.2590	0.2493
MATTHEWS	Av.	0.2006	0.1813	0.1944	0.0812	0.0000	0.2392	0.2488	0.1120	0.2432	0.2051
	cloudy	0.3278	0.2943	0.2940	0.0919	0.0000	0.3305	0.3456	0.1283	<u>0.3457</u>	0.2532
	partial	0.0283	0.0086	0.0602	<u>0.1043</u>	0.0000	0.0593	0.0796	0.0890	0.1000	0.0892
	clear	0.2456	0.2411	0.2291	0.0473	0.0000	<u>0.3279</u>	0.3213	0.1186	0.2840	0.2728

Table 8

Ranking average summary.

	k_d	k_b	k_t	F	kk	FK	ϵ	Δ	FP	Si
ACCURACY	5	7	4	9	10	1	2	8	3	6
JACCARD	5	7	4	9	10	2	1	8	3	6
COHEN KAPPA, κ	5	7	4	8	9	2	1	10	3	6
MATTHEWS	5	7	6	9	10	3	1	8	2	4
GLOBAL	5	7	4	9	10	2	1	8	3	6

295, 10th December 2015).

References

Alshaibani, K., 2011. Finding frequency distributions of CIE Standard General Skies from sky illuminance or irradiance. *Light. Res. Technol.* 43 (4), 487–495.

Baharuddin, LauS.S., Rahim, R., 2010. Daylight availability in Hong Kong: classification into three sky conditions. *Archit. Sci. Rev.* 53 (4), 396–407.

Battles, F.J., Olmo, F.J., Tovar, J., Alados-Arboledas, L., 2000. Comparison of cloudless sky parameterizations of solar irradiance at various Spanish midlatitude locations. *Theor. Appl. Climatol.* 66 (1), 81–93.

Brunger, A.P., Hooper, F.C., 1993. Anisotropic sky radiance model based on narrow field of view measurements of shortwave radiance. *Sol. Energy* 51 (1), 53–64.

Chicco, D., 2017. Ten quick tips for machine learning in computational biology. *BioData Mining* 10 (1), 35.

Djafer, D., Irbah, A., Zaiani, M., 2017. Identification of clear days from solar irradiance observations using a new method based on the wavelet transform. *Renew. Energy* 101, 347–355.

Erbs, D.G., Klein, S.A., Duffie, J.A., 1982. Estimation of the diffuse radiation fraction for hourly, daily and monthly-average global radiation. *Sol. Energy* 28 (4), 293–302.

Escobedo, J.F., Gomes, E.N., Oliveira, A.P., Soares, J., 2009. Modeling hourly and daily fractions of UV, PAR and NIR to global solar radiation under various sky conditions at Botucatu, Brazil. *Appl. Energy* 86 (3), 299–309.

Gueymard, C.A., 2018. A reevaluation of the solar constant based on a 42-year total solar irradiance time series and a reconciliation of spaceborne observations. *Sol. Energy* 168, 2–9.

Gueymard, C.A., Bright, J.M., Lingfors, D., Habte, A., Sengupta, M., 2019. A posteriori clear-sky identification methods in solar irradiance time series: review and preliminary validation using sky imagers. *Renew. Sustain. Energy Rev.* 412–427.

Gueymard, C.A., Ruiz-Arias, J.A., 2016. Extensive worldwide validation and climate sensitivity analysis of direct irradiance predictions from 1-min global irradiance. *Sol. Energy* 128, 1–30.

Igawa, N., Koga, Y., Matsuzawa, T., Nakamura, H., 2004. Models of sky radiance distribution and sky luminance distribution. *Sol. Energy* 77 (2), 137–157.

Iqbal, M., 1983. An introduction to solar radiation. Academic Press.

Jaccard, P., 1912. The distribution of the flora in the alpine zone. 1. *New Phytol.* 11 (2), 37–50.

Jacovides, C., Tymvios, F., Assimakopoulos, V., Kaltsounides, N., 2007. The dependence of global and diffuse PAR radiation components on sky conditions at Athens, Greece. *Agric. Forest Meteorol.* 143 (3–4), 277–287.

Kambezis, H., Katevatis, E., Petrakis, M., Lykoudis, S., Assimakopoulos, D., 1998. Estimation of the Linke and Unsworth-Monteith turbidity factors in the visible spectrum: application for Athens, Greece. *Solar Energy* 62 (1), 39–50.

Kambezis, H.D., 2018. The solar radiation climate of Athens: variations and tendencies

in the period 1992–2017, the brightening era. *Sol. Energy* 173, 328–347.

Kasten, F., 1993. Discussion on the relative optical air mass. *Light. Res. Technol.* 25 (3), 129–130.

Khorasanizadeh, H., Mohammadi, K., 2016. Diffuse solar radiation on a horizontal surface: reviewing and categorizing the empirical models. *Renew. Sustain. Energy Rev.* 53, 338–362.

Khorasanizadeh, H., Mohammadi, K., Goudarzi, N., 2016. Prediction of horizontal diffuse solar radiation using clearness index based empirical models; a case study. *Int. J. Hydrogen Energy* 41 (47), 21888–21898.

Kittler, R., Perez, R., Darula, S., 1997. A new generation of sky standards. In: *Proceedings of the Eighth European lighting conference*, Amsterdam, pp. 359–373.

Klucher, T.M., 1979. Evaluation of models to predict insolation on tilted surfaces. *Sol. Energy* 23 (2), 111–114.

Kong, H.J., Kim, J.T., 2013. A Classification of real sky conditions for Yongin, Korea. *Sustain. Energy Build.* Springer 1025–1032.

Kundel, H.L., Polansky, M., 2003. Measurement of observer agreement. *Radiology* 228 (2), 303–308.

Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. *Biometrics* 159–174.

Li, D.H.W., 2010. A review of daylight illuminance determinations and energy implications. *Appl. Energy* 87 (7), 2109–2118.

Li, D.H.W., Chau, N.T.C., Wan, K.K.W., 2013. Predicting daylight illuminance and solar irradiance on vertical surfaces based on classified standard skies. *Energy* 53, 252–258.

Li, D.H.W., Chau, T.C., Wan, K.K.W., 2014. A review of the CIE general sky classification approaches. *Renew. Sustain. Energy Rev.* 31, 563–574.

Li, D.H.W., Cheung, G.H.W., 2006. Average daylight factor for the 15 CIE standard skies. *Light. Res. Technol.* 38 (2), 137–152.

Li, D.H.W., Lam, J.C., 2001. An analysis of climatic parameters and sky condition classification. *Build. Environ.* 36 (4), 435–445.

Li, D.H.W., Lam, T.N.T., Cheung, K.L., Tang, H.L., 2008. An analysis of luminous efficiencies under the CIE standard skies. *Renew. Energy* 33 (11), 2357–2365.

Li, D.H.W., Tang, H.L., Wong, S.L., Tsang, E.K.W., Cheung, G.H.W., Lam, T.N.T., 2007. Skies classification using artificial neural networks (ANN) techniques. In: *6th International Conference on Indoor Air Quality, Ventilation and Energy Conservation in Buildings: Sustainable Built Environment, IAQVEC 2007*. Sendai, pp. 61–68.

Littlefair, P.J., 1994a. A comparison of sky luminance models with measured data from Garston, United Kingdom. *Solar Energy* 53 (4), 315–322.

Littlefair, P.J., 1994b. The luminance distributions of clear and quasi-clear skies. In: *Proceedings of the CIBSE National Lighting Conference*, Cambridge, UK, pp. 267–283.

Lou, S., Li, D.H.W., Lam, J.C., 2017. CIE Standard Sky classification by accessible climatic indices. *Renew. Energy* 113, 347–356.

Markou, M.T., Kambezidis, H.D., Bartzokas, A., Katsoulis, B.D., Muneer, T., 2005. Sky type classification in Central England during winter. *Energy* 30(9 SPEC. ISS.) 1667–1674.

Markou, M.T., Kambezidis, H.D., Katsoulis, B.D., Muneer, T., Bartzokas, A., 2004. Sky type classification in South England during the winter period. *Build. Res. J.* 52 (1), 19–30.

Matthews, B.W., 1975. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim. et Biophys. Acta (BBA)-Protein Struct.* 405 (2), 442–451.

Mellit, A., Kalogirou, S.A., Shaari, S., Salhi, H., Hadj Arab, A., 2008. Methodology for predicting sequences of mean monthly clearness index and daily solar radiation data in remote areas: application for sizing a stand-alone PV system. *Renew. Energy* 33 (7), 1570–1590.

Muneer, T., 2007. *Solar radiation and daylight models*. Routledge.

Muneer, T., Gueymard, C., Kambezidis, H.D., 2004. *Solar Radiation and Daylight Models*

- (Second Edition). Butterworth Heinemann.
- Muneer, T., Younes, S., Munawwar, S., 2007. Discourses on solar radiation modeling. *Renew. Sustain. Energy Rev.* 11 (4), 551–602.
- Perez, R., Ineichen, P., Seals, R., Michalsky, J., Stewart, R., 1990. Modeling daylight availability and irradiance components from direct and global irradiance. *Sol. Energy* 44 (5), 271–289.
- Perez, R., Seals, R., Ineichen, P., Stewart, R., Menicucci, D., 1987. A new simplified version of the perez diffuse irradiance model for tilted surfaces. *Sol. Energy* 39 (3), 221–231.
- Perraudou, M., Chauvel, P., 1986. One year's measurements of luminous climate in Nantes. *Proceedings of the International Daylighting Conference*. Long Beach, CA.
- Perveen, G., Rizwan, M., Goel, N., 2018. Intelligent model for solar energy forecasting and its implementation for solar photovoltaic applications. *J. Renew. Sustain. Energy* 10 (6), 063702.
- Powers, D.M., 2011. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation.
- 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 (7), 660–666.
- Ruiz-Arias, J.A., Gueymard, C.A., 2018. Worldwide inter-comparison of clear-sky solar radiation models: consensus-based review of direct and global irradiance components simulated at the earth surface. *Sol. Energy* 168, 10–29.
- Sim, J., Wright, C.C., 2005. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys. Ther.* 85 (3), 257–268.
- Suárez-García, A., Granados-López, D., González-Peña, D., Díez-Mediavilla, M., Alonso-Tristán, C., 2018. Seasonal characterization of CIE standard sky types above Burgos, northwestern Spain. *Sol. Energy* 169, 24–33.
- Tapakis, R., Michaelides, S., Charalambides, A.G., 2016. Computations of diffuse fraction of global irradiance: part 1 – analytical modelling. *Sol. Energy* 139, 711–722.
- Torres, J.L., de Blas, M., García, A., Gracia, A., de Francisco, A., 2010a. Sky luminance distribution in Pamplona (Spain) during the summer period. *J. Atmos. Sol.-Terr. Phys.* 72 (5–6), 382–388.
- Torres, J.L., de Blas, M., García, A., Gracia, A., de Francisco, A., 2010b. Sky luminance distribution in the North of Iberian Peninsula during winter. *J. Atmos. Sol.-Terr. Phys.* 72 (16), 1147–1154.
- Tregenza, P.R., 2004. Analysing sky luminance scans to obtain frequency distributions of CIE standard general skies. *Light. Res. Technol.* 36 (4), 271–279.
- Tregenza, P.R., Perez, R., Michalsky, J., Seals, R., Molineaux, B., Ineichen, P., 1994. Guide to recommended practice of daylight measurement. *Cimmission Internationale de l'éclairage*, Vienna.
- Uetani, Y., Aydinli, S., Joukoff, A., Kendrick, J., Kittler, R., Koga, Y., 2003. Spatial distribution of daylight-CIE standard general sky. *Austria, Vienna*.
- Umemiya, N., Kanou, T., 2008. Classification of sky conditions by the ranges of insolation indices considering CIE standard for general sky. *J. Light Visual Environ.* 32 (1), 14–19.
- Wang, L., Gong, W., Ma, Y., Hu, B., Wang, W., Zhang, M., 2013. Analysis of ultraviolet radiation in Central China from observation and estimation. *Energy* 59, 764–774.
- WMO, 2008. (Updated 2010). *Guide to Meteorological Instruments and Methods of Observation*. NO 8.
- WMO, 2010. *Technical Regulations. Volume I: General Meteorological Standards and Recommended Practise*. http://library.wmo.int/pmb_ged/wmo_49-v1-2012_en.pdf.
- Wong, S.L., Wan, K.K.W., Li, D.H.W., Lam, J.C., 2012. Generation of typical weather years with identified standard skies for Hong Kong. *Build. Environ.* 56, 321–328.

Annex II:

Communications

International

- ANII- 1. Diego Granados López; David González Peña; Ana García Rodríguez; Sol García Rodríguez; Manuel García Fuente. ***Machine Learning for BIPV Production***. *EU PVSEC European Photovoltaic Solar Energy Conference and Exhibition 2021. WIP Renewable energies*. 2021. Portugal. 1529 – 1531. ISBN 3-936338-78-7.

Abstract: The increase in energy consumption in homes has never been as intense as in the last decade. The international alarm has resulted in various plans and strategies This research supports the implementation of Building Integrated Photovoltaics (BIPV) proposing a model to predict vertical PV production from vertical solar irradiation and ambient temperature data. The predictor is an Artificial Neural Network (ANN), trained using the Levenberg Marquardt algorithm. The research tested and compared several architectures of the ANN to obtain the most accurate network. The ANN chosen has a single hidden layer. Its entry has two neurons, ambient temperature, and global vertical irradiation, RaGV. This research showed that ambient Temperature and RaGV are enough to estimate BIPV production, being RaGV the most relevant variable. The introduction as input of the model of the ambient temperature increased its performance. The obtained model had good statistical results and allowed to estimate the PV production from commercial panels.

- ANII- 2. Diego Granados López; Ana García Rodríguez; Sol García Rodríguez; David Gonzalez Peña. ***Avoiding shading on photovoltaic panels in building integrated photovoltaics (BIPV)***. *ICEESEN (2020). Erciyes Üniversitesi ERÜ / KAYSERİ. 2020. Turquía. 347 – 350. ISBN: 978-625-409-147-6*

Abstract: The integration of photovoltaic (PV) panels on non-used buildings surfaces allows the electricity distributed generation near the consumption places and the electrification of remote areas avoiding the transportation cost and electrical losses. The current

study highlights the importance of carrying out a prior analysis of the nearest area to get a correct implementation of the PV panels, avoiding the shadows produced by the close obstacles. A recently refurbished public building has been analyzed: the IndUVA building of the University of Valladolid, Spain. Although the current design offers good energy production, the consideration of the roof surface in the design could have improved the electricity production up to 37%. The roof surface presents some advantages over the facade surfaces: easier and cheaper implementation of the PV panels, and also, higher production than the south-oriented surface in summer.

- ANII- 3. Diego Granados López; Montserrat Díez Mediavilla; Ana García Rodríguez; David Gonzalez Peña; Cristina Alonso Tristán. ***Towards Net-Zero Public Buildings through BIPV: a Case Study in Castilla-León(Spain).*** *37th European Photovoltaic Solar Energy Conference and Exhibition. PVSEC. 2020. Portugal. 1964 – 1967. ISBN 3-936338-73-6*

Abstract: The new European directives focused to achieve public Net-Zero Energy Buildings require for their application the knowledge of the potential of photovoltaics (PV) generation of these buildings. The available building surfaces (façades and roofs) for PV integration depend on the building's location and orientation, and relative position of the nearby buildings and elements. This work developed a technical procedure to determine the available surface for PV integration, the economic and energetic viability of the PV facilities, and their contribution of the energy consumption of the buildings through a case study. The current design has a potential benefit of 0,17 million of euros. Its greenhouse gas emissions are 0.27–0.92 Ton, ten times smaller than a production obtained through the traditional fuels (6–8.47 Ton), which are commonly used in Spain.

ANII- 4. Diego Granados López; Manuel García Fuente; David González Peña; Montserrat Díez Mediavilla; Cristina Alonso Tristán. ***Indoor Daylight Model for All Sky-Type Luminance Pattern. XII National and III International Conference on Engineering Thermodynamics. 12CNIT (2022). Universidad Carlos III de Madrid. Spain. (Accepted submission).***

Abstract: Daylight is a valuable source for illuminating indoor spaces that provides positive effects on the well-being and health of occupants, reducing stress levels, and improving mood and photobiological effects. Daylight reduces electricity consumption on lighting and increases the energy efficiency of buildings. International Net Zero Energy Building (NZEB) standards and regulations recommend the incorporation of natural lighting strategies in their design and define minimum standards. Important input data for predicting the indoor illuminance are the distributions of the sky luminance, that can be obtained from different models. In this study, DeLight of Vartiainen model was selected to predict the indoor illuminance using the CIE Standard sky luminance distribution model. Model has been tested experimentally.

ANII- 5. Diego Granados-López; Sol García-Rodríguez; Ana García-Rodríguez; Montserrat Díez-Mediavilla; Cristina Alonso-Tristán. ***Machine Learning techniques for estimation of BIPV production. XII National and III International Conference on Engineering Thermodynamics. 12CNIT (2022). Universidad Carlos III de Madrid. Spain. (Accepted submission).***

Abstract: The European Commission promotes achieving climate neutrality by 2050 . Taking advantage of renewable energy sources is crucial in the most recent European strategies, whose focus is to reduce the energy dependency of Europe and to avoid greenhouse gas emissions. The decrement of the electricity costs from utility-scale solar photovoltaic (PV) installations has been notably reduced since 2010 and nowadays is getting close to being competitive with

conventional electricity sources. Building Integrated PV facilities (BIPV) can be a leading technology in terms of cost and efficiency. PV production is highly dependent on weather conditions.

ANII- 6. Diego Granados-López; Ignacio García; Jose Luis Torres; Andrés Suárez-García; Montserrat Díez-Mediavilla; Cristina Alonso-Tristán. ***Propuesta de reagrupación de los tipos de cielo ISO/CIE mediante técnicas de aprendizaje supervisado. XVIII Congreso Ibérico y XIV Congreso Iberoamericano de Energía Solar. CIES2022 (2022). Universitat de Les Illes Balears. Spain. (Accepted submission).***

Abstract: La nubosidad del cielo determina en gran medida la radiación solar y la luz natural recibida en la superficie terrestre. En este sentido, el conocimiento de la distribución angular de la radiancia y la luminancia en el cielo, permite conocer de manera precisa la irradiancia o iluminancia recibidas en un plano de interés. Entre los modelos existentes, se encuentra el propuesto por la Comisión Internacional de Iluminación (CIE), que se consolidó en una norma ISO en 2004 (ISO 15469:2004(E)/CIE S 011/E:2003). La propuesta de la CIE establece un total de 15 tipos de cielo, cada uno asociado a un patrón específico de distribución de la luminancia en el cielo. Este patrón se puede integrar sobre un plano de interés con una orientación e inclinación determinadas para calcular la iluminancia difusa. Resulta especialmente interesante su integración sobre los planos vertical y horizontal. Concretamente, la ratio entre la iluminancia recibida sobre el plano vertical (libre de obstáculos) y el horizontal se conoce como Vertical Sky Component (VSC) y está estrechamente relacionada con la clasificación de cielos ISO/CIE.

En la Figure 36, se muestra la evolución de la VSC en función de la elevación solar y el tipo de cielo. Como se puede apreciar en la propia Figure 36, en este trabajo, se propone una agrupación simplificada de los 15 tipos de cielo de la norma ISO/CIE en cinco grupos: Grupo 1: tipos 1, 2, 3, 5; Grupo 2: tipos 4, 6, 7; Grupo 3: tipos 8, 9, 10, 11; Grupo 4: tipos 12, 13; Grupo 5: tipos 14, 15. El Grupo 1 refiere el porcentaje

de cielo más cubierto de la clasificación. El grupo 2 será también bastante oscuro, y la influencia del Sol debe empezar a ser tenida en cuenta. El grupo 3 puede mostrar una dependencia muy fuerte con la posición del Sol, y cada orientación tendrá su propia iluminancia disponible. El grupo 4 será un cielo contaminado o un poco turbio, y la VSC teórica puede ser mayor del 100% (la iluminancia difusa vertical es mayor que la horizontal). Por último, en el grupo 5 la VSC puede alcanzar valores aún más altos, disminuyendo el confort visual y aumentando la carga térmica en el interior del edificio.

La clasificación de cielos según la norma ISO/CIE tiene un gran potencial ya que permite relacionar los tipos de cielos con la distribución angular de luminancia celeste. Sin embargo, la medida de dicha distribución requiere de equipos, comúnmente conocidos como sky scanners, que raramente están presentes en las estaciones meteorológicas convencionales. Por esta razón, en la literatura científica se pueden encontrar diferentes métodos alternativos de clasificación de cielos aplicables a partir de las medidas registradas rutinariamente en las estaciones meteorológicas. Entre estas alternativas se encuentra el uso Índices Meteorológicos (IMs), que se calculan a partir de variables fácilmente accesibles desde las estaciones meteorológicas.

Este estudio propone la utilización del machine learning supervisado para clasificar el cielo a partir de IMs en los cinco tipos anteriormente establecidos según los valores de VSC. Concretamente, se han utilizado el índice de nebulosidad de Perraudau (F), el índice de brillo (Δ) y de claridad (ϵ) de Perez, la Cloud Cover (CC), el VSC, y el índice de cielo de Igawa (Si). Tomando como referencia la clasificación de cielos realizada a partir de las medidas proporcionadas por un sky scanner localizado en la estación meteorológica de la Universidad de Burgos (España), se ha comparado la clasificación ISO/CIE obtenida mediante el uso de redes neuronales artificiales (ANNs, por sus siglas en inglés) y árboles de decisión (ADs). Se ha observado que ambos

modelos presentan un buen comportamiento a la hora de clasificar los cielos a partir de IMs.

En las matrices de confusión presentadas en la Figure 37, se observa que ambos modelos ofrecen buenas predicciones a la hora de clasificar los cielos a partir de IMs, sin diferencias reseñables. Por su simplicidad, se recomienda la aplicación de ADs frente a las ANNs.

National

ANII- 7. Diego Granados López; David González Peña; Andrés Suárez García; María Isabel Dieste Velasco; María del Carmen Rodríguez Amigo; Montserrat Díez Mediavilla; Cristina Alonso Tristán. ***Obtención de la componente vertical del cielo.*** *V Jornadas de Doctorandos de la Universidad de Burgos, Servicio de Publicaciones e Imagen Institucional. 2018. España. pp 29-38. ISBN: 978-625-409-147-6 (e-book).*

Resumen: Para el estudio de la eficiencia energética en edificios, conocer los valores de irradiación e iluminación solar es de crucial importancia, sobre todo en países con frecuencias muy elevadas de cielos despejados, como es el caso de España. La estimación de la luz que llega a una determinada zona de trabajo en el interior de un edificio requiere del conocimiento de las condiciones exteriores de iluminación, en concreto de la componente difusa, que, en general, es la más importante. La cantidad de luz disponible se estima con la Componente Vertical del Cielo (Vertical Sky Component, VSC), definida como la relación entre la iluminancia difusa horizontal y la iluminancia difusa vertical en un punto de referencia. Este parámetro se puede obtener de forma teórica o experimental. En la ciudad de Burgos, donde el tipo de cielo según el estándar CIE más frecuente es el cielo despejado, se puede afirmar que, en la mayoría de las ocasiones, la iluminancia vertical es superior a la horizontal. Por lo tanto, con un diseño adecuado de los edificios en cuanto a la

distribución de ventanas y lucernarios, se podría aumentar considerablemente la eficiencia lumínica, reduciendo los costes energéticos asociados.

ANII- 8. Diego Granados López, David Gonzalez Peña, Ana García Rodríguez, Montserrat Díez Mediavilla, Cristina Alonso Tristán. ***Clasificación de cielos con árboles de clasificación. VI Jornadas De Doctorandos de la Universidad De Burgos. Servicio de Publicaciones e Imagen Institucional. Burgos 2019. ISBN: 978-84-16283-86-6 (e-book).***

Resumen: La radiación solar es considerada como uno de los recursos más valiosos que la naturaleza ofrece. Correctamente aprovechada permite aumentar la calidad de vida y desarrollar la actividad humana a la vez que se convive en armonía con el resto de los seres vivos del planeta. No obstante, presenta el inconveniente de ser altamente dependiente de la presencia y distribución de nubes en la cúpula celeste. La metodología estandarizada a nivel mundial que permite clasificar cielos fue la diseñada por la Comisión Internacional de Iluminancia (CIE), que precisa del uso de Sky-scanner, un aparato relativamente costoso raramente disponible en las estaciones meteorológicas locales. Existen multitud de modelos desarrollados para, de manera simplificada, clasificar el cielo. Estos modelos parten de la definición de variables que actúan como indicadores meteorológicos (MIs) que definen intervalos característicos. En este trabajo se propone una clasificación de la nubosidad del cielo (despejados, intermedios y nublados) basada en el Estándar Internacional de la norma CIE. La metodología propuesta tiene una gran fiabilidad, con un índice estadístico f_1 del 80%. Se ha obtenido que el uso conjunto del índice de nebulosidad de Perraudau, la cubierta nubosa y el VSC es suficiente para categorizar el cielo con un alto grado de confianza.