

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15

Hierarchical color similarity metrics for step-wise application on sky monitoring surface cameras

S. L. Mantelli Neto ^{1,2}, A. C. Sobieranski ², E. B. Pereira ¹, A. von Wangenheim ²

¹INPE Brazilian National Institute for Space Research
Av. dos Astronautas 1758 São José dos Campos SP Brazil 12227-010
²UFSC Federal University of Santa Catarina
Campus Universitário Trindade, Florianópolis SC Brazil 88040-900

Key Points:

- Most methods existent on literature for surface camera cloud assessments uses dichotomic results and discards one dimension of colour space
- Those methods are adopted because domain monitored is logarithmic and cameras linear devices causing distortions and saturation to fit range
- We proposed a new method based on atmospheric scattering dealing with pixel distortions and full color space use reducing analysing errors

Corresponding author: Sylvio Luiz Mantelli Neto, sylvio@lepten.ufsc.br

16 **Abstract**

17 Digital cameras on the surface are frequently used for monitoring atmospheric conditions.
 18 Several methods were developed to use the images for synoptic observations, cloud as-
 19 sements, short term forecasting and so on. However, there are some restrictions not
 20 considered by these methods, especially when a linear camera is used to observe loga-
 21 rithmic ranges of atmospheric luminance. Cameras accommodate the scene to a linear
 22 scale causing distortions on pattern distributions by pixel value saturation (PVS) and
 23 drifts from its original hues. This brings on some simplifying practices commonly found
 24 in the literature to overcome these problems. But those practices result in loss of data,
 25 misinterpretation of valid pixels and restriction on the use of computer vision algorithms.
 26 The present work begins by illustrating these problems performing supervised learning
 27 for two reasons: all observation systems seek out automation of human synoptic obser-
 28 vation in order to provide a sound mathematical modeling of the observed patterns. A
 29 new modeling paradigm is proposed to map the sky patterns to represent the existent
 30 physical atmospheric phenomena not considered by the literature. We validate the pro-
 31 posed method, and compared the results using 1630 images against two well-established
 32 methods. A hypothesis test showed that results are compatible with currently used bi-
 33 nary approach with advantages. Differences were due to PVS and other restrictions not
 34 considered by the methods existent on literature. Finally, the present work concludes
 35 that the new paradigm presents more meaningful results of sky patterns interpretation,
 36 allows extended daylight observation periods and uses a higher dimensional space.

37 **1 Introduction**

38 The observation of current atmospheric conditions from the surface is an impor-
 39 tant feature to be monitored, especially in order to assess cloud coverage, amount and
 40 category. These parameters are especially important in the climate research area (Kasten
 41 & Czeplak, 1980), (Marty & Philipona, 2000), (Bojanowski et al., 2013), atmospheric
 42 physical models (Harrison et al., 2008), (Nardino & Georgiadis, 2003), (Yamanouchi &
 43 Charlock, 1993), (Cess et al., 1995) and validation of satellite-based resources (Martins
 44 et al., 2007), (Martins et al., 2003). Clouds are also a major source of uncertainty in the
 45 assessment of solar energy (Hu & Stamnes, 2000). In particular, a considerable effort has
 46 been spent on computer-based methods able to assess *nowcasting* conditions.

47 Synoptic observation (SO) is one activity always present on monitoring stations.
 48 SO evaluation of clouds is usually performed by humans and is highly subjective and vari-
 49 able (WMO, 2008, chap. 15). For these reasons, several research groups have been aim-
 50 ing at replacing a highly human-dependent activity through cameras and computer-based
 51 methods. The World Meteorological Organization (WMO) calls continuous sky moni-
 52 toring equipment Synoptic Observation Systems (SOS) (WMO, 2008). SOSs usually em-
 53 ploy all-sky digital cameras, algorithms, and methods for continuous monitoring (Bradley
 54 et al., 2010). Several commercial types of equipment are being used for this purpose at
 55 considerable costs, such as: Total Sky Imager-(TSI)
 56 (<http://www.yesinc.com/products/data/tsi880/index.html>), Whole Sky Imager-(WSI)
 57 (<http://www.arm.gov/instruments/wsi>), MOONGLOW
 58 (<http://www.allskycam.com/index.php>) and so on. Some research groups are trying to
 59 find more affordable equipment and better surface image-based methods for cloud de-
 60 tection.

61 Surface based sky pattern analysis, however, has been restricted when automated
 62 systems are used to reproduce the human qualitative analysis of the environment. It is
 63 important to evaluate what could actually be classified with present methods due to sys-
 64 tem limitations.

65 When focusing on cloud detection and quantification, the most common outcome
 66 expected from automated image analysis approaches found in the literature is the clas-

67 sification of image into cloud or sky patterns. This kind of pixel value-based classifier
 68 that classifies pixel values in either one or other category by thresholding is a *dichotomizer*
 69 as shown by (Duda et al., 2001, sec.2.4.2). Authors noticed by observation the presence
 70 of more than two patterns present on color space representing other physical *phenom-*
 71 *ena* (i.e. red, yellow, etc) that are not cloud and sky patterns (white and blue) (S. Man-
 72 telli, 2001; Naylor, 2002). These additional *phenomena* are misclassified by dichotomous
 73 non-hierarchical methods. This moved the authors to find a more adequate approach to
 74 deal with systemic technological restrictions.

75 The present work focuses on cloud detection and quantification. If no proper cloud
 76 detection method is used, there is no meaning going on more complex cloud classifica-
 77 tion. The objective will be on the evaluation of computer-based methods using cameras
 78 to quantify the clouds like: computer vision, digital image processing and machine learn-
 79 ing (ML) algorithms. An extended analysis using additional sensors and methods for cloud
 80 assessments could be observed on the review made by Tapakis and Charalambides (2013).

81 In our approach the sky observation will be replaced by a computational model cen-
 82 tered on a new hierarchical color similarity measure that matches SOS's functions. We
 83 propose tackling the problem of developing a more reliable SOS through a broader and
 84 more systemic analysis, considering not only the computer algorithms to quantify the
 85 clouds, but also the monitoring sensor, its capabilities to observe the environment, and
 86 its possible outcomes.

87 2 Objectives

88 The objectives of this work are

- 89 • to evaluate present surface camera-based methods for the quantification of clouds
 90 that employ computer vision and related approaches for digital image processing,
 91 and cloud coverage analysis and
- 92 • to propose a new hierarchical partially non-isotropic color space model that over-
 93 comes some of the shortcomings of those models.

94 In this context, our paper offers two main contributions:

- 95 • we demonstrate the existence of patterns representing physical phenomena reg-
 96 istered by surface cameras by performing an exploratory data analysis (EDA) on
 97 customized color spaces based upon pattern occurrences;
- 98 • we propose a novel hierarchical image analysis method to characterize these pat-
 99 terns that takes into consideration atmospheric, optics and surface camera lim-
 100 itations.

101 The validation of our approach was performed through comparison to methods de-
 102 scribed in the literature, which we implemented and compared to our results. This com-
 103 parison was possible through reducing the color space dimension of the results obtained
 104 by our approach when comparing our results to the results achieved through traditional
 105 methods for nearly 1630 images.

106 In the next paragraphs we will describe the role of luminance in the system, the
 107 experimental set-up and propose a new set of patterns that could be perceived from im-
 108 ages acquired through state-of-the-practice equipment. We will base our model on the
 109 atmospheric optical physics theory (Naylor, 2002), abandoning a simple dichotomized
 110 approach.

111 **3 State of the Art**

112 A commonly used cloud pattern classification method is value thresholding (VT).
 113 It is based on pixel values or their combination in order to determine whether a given
 114 pixel represents a cloud or clear sky area. VT was used by (Souza-Echer et al., 2006) with
 115 classification criterion based on the saturation (S) dimensions only out of Hue, Satur-
 116 ation and Lightness (HSL) color space. Hue and Lightness pixel dimensions were discarded,
 117 restricting the cloud and sky patterns variability on the task at hand (Newell & Simon,
 118 1971) to only one dimension (1-D). A variation of VT was developed by (Kazantzidis et
 119 al., 2012) using the difference $R-B$ between red (R) and blue (B) pixel dimensions on
 120 Red, Green and Blue (RGB) color space to classify pixels into cloud or sky patterns. Only
 121 the green dimension was discarded and the variability of the cloud/sky patterns were re-
 122 stricted to two dimensions (2D). If pixel dimensions are reduced to the (R,B) plane, their
 123 correspondent color range is also reduced to a range varying from black to magenta, in-
 124 stead of black to white (Gonzalez & Woods, 2007, sec 6.2.1) in the color space. A clas-
 125 sification method based on *linear thresholding* (LT) was developed in (H. W. S. J. John-
 126 son R., 1989.). It used normalized Red (R) and Green (G) pixel dimensions R/B to clas-
 127 sify pixels into cloud or sky patterns. The limitations of these methods are the same al-
 128 ready described by 2D usage of color space. But VT and LT are simpler, easily repro-
 129 ducible and widely used for data comparison between methods.

130 More elaborated methods based on machine learning have also been proposed to
 131 discriminate between cloud and clear sky patterns. An approach employing *neural net-*
 132 *works* (NNs) trained upon training sets obtained from previously classified images con-
 133 taining *RGB* values normalized by clear sky models, as described by (Iqbal, 1983), was
 134 developed in order to classify sky images (S. Mantelli, 2001; S. L. Mantelli et al., 2005).
 135 The work presented in (Cazorla et al., 2008) also used NNs with optimized parameters
 136 by means of *genetic algorithms* (GA) to identify the same two classes, *sky* and *cloud*.
 137 The parameters were obtained from a variance matrix 9×9 of R and average values of
 138 R and B dimensions only. But NN and GA methods depend heavily on a large set of
 139 implementation conditions, not easily replicated, unless these parameters (training sets,
 140 NN configuration, etc.) are supplied in detail by the original authors. NN is a power-
 141 ful computational resource when correctly modeled, and is can be executed in parallel
 142 due to the nature of the technique. On the other side, the use of NN as a linear or non-
 143 linear mapper, could be methodologically misleading due to its black-box working sim-
 144 ply like a linear regression method. Some authors reported difficulties on the adequate
 145 parametric representation of the problem caused by the complex structure of NNs
 146 (Johannet et al., 2007; Qiu & Jensen, 2004; Setiono et al., 2000), while others indicate
 147 semantic errors during the development of specific applications (Jain et al., 2000; Zhang,
 148 2007). Improper or restricted modeling of task environment and inadequate size of train-
 149 ing sets, together with lack of sound mathematical base can lead NN to useless results.
 150 One problem with dichotomizing is the two class problem solution. Any pixel value ex-
 151 istent in the observation domain that does not belong to one or either class, is randomly
 152 misplaced in some of them. (Cazorla et al., 2015) also developed an adaptive method
 153 to classify sky images. (Yang et al., 2016) used background subtraction together with
 154 an adaptive thresholding to evaluate current conditions to detect clouds. Parameters for
 155 the adaptive thresholding were obtained from a set of clear sky reference images. An-
 156 other problem is that, according to (Reinhard et al., 2005; Mitsunaga & Nayar, 1999)
 157 sky luminance range spans nearly 5 orders of magnitude, from 10^1 to 10^5 cd/m^2 . If we
 158 consider that typical cameras are capable of monitoring luminance values of approximately
 159 2 orders of magnitude, from 0 to 10^2 , it becomes clear that, in order for an adaptive method
 160 using cameras to span the entire scale of logarithmic luminance in discrete steps (multi-
 161 exposure approach), approximately $s = \frac{10^5}{10^2} = 10^3$ steps will be necessary.

162 The luminance variation in the sky is intense, non-uniform, and dependent on an-
 163 gular and atmospheric conditions (Perez et al., 1993). All the methods mentioned above

164 present difficulties and even errors (Sabburg & Wong, 1999), leading to a deteriorating
 165 SOS performance in classifying patterns, especially near the higher intensity regions. The
 166 authors additionally understand that the reduction of dimensionality produced by 1D
 167 and 2D color-spaces also aggravates these problems, restricting analysis range and prun-
 168 ing data variability, preventing a more detailed analysis. Some authors prefer to avoid
 169 this kind of situation removing higher intensity regions from image analysis by “*man-*
 170 *ual cropping*” (Qingyong et al., 2011, sec. 2.a). Other authors used the same “*manually*
 171 *cropped*” method and extended it to all-sky images (Marquez & Coimbra, 2013), ignor-
 172 ing that such approach does not configure an *all-weather* method. In that case, the anal-
 173 ysis, performed by a dichotomizer, is optimized to work only on a subset of the environ-
 174 ment captured by all-sky systems. Ceiling all-sky images by cropping higher intensity
 175 regions removes a significant amount of information from the image analysis process. This
 176 phenomenon will later be illustrated on figure 1 and these pixels will be defined as per-
 177 taining to two distinct color-subspaces: *Diffusion of Non-Specific Scattering (DNSS)*
 178 and *Diffusion of Rayleigh Scattering (DRAY)* patterns (Grossberg & Nayar, 2003). An-
 179 other aspect described in (Qingyong et al., 2011) is that, when a part of the images used
 180 for the analysis are taken manually by different operators under distinct camera expo-
 181 sition adjusts, an adaptive method seems to be appropriate to compensate the variable
 182 image acquisition exposition adjusted manually.

183 Some digital image processing methods like background subtraction (Piccardi, 2004)
 184 and spatial geometric *locus* (S. L. Mantelli et al., 2010) were also used to detect clouds.
 185 Background subtraction is a computer vision technique commonly used on image detec-
 186 tion of moving objects using a static background. An automated cloud detection method
 187 based on the green channel or 1D of total-sky visible images (Yang et al., 2015, 2016)
 188 mentioned a better performance of thin clouds detection. Although the article recognized
 189 effects of Rayleigh and Mie scattering on image acquisition, no formal treatment is men-
 190 tioned on regions with saturated pixel values. Spatial geometric locus of patterns on color
 191 space deals with the full dimensional range of color space and do not use dichotomizing.
 192 The classification results also show patterns in luminance gradients, but they are also
 193 restricted by a mapping into three separate classes.

194 Other methods assisted by physical models like (Mejia et al., 2016) and (Kurtz et
 195 al., 2017) employs supervised learning and analysis. Learning process is implemented from
 196 a Radiative Transfer Model (RTM) and simulated images of various cloud optical depth
 197 (τ_c). Authors also considered different solar (\mathcal{V}_s) and pixel (\mathcal{V}_z) position related to cam-
 198 era zenith angle (θ_0). But although the method pointed out some new and important
 199 features to be used on image analysis and not used in our work, it still uses only a 2D
 200 color space based upon *Radiance Red Blue Ratio* (RRBR), losing information. Another
 201 important consideration used by the method is the adjust of saturated pixels to 1, mean-
 202 ing that every saturated pixel is supposed to be a cloud (Mejia et al., 2016, sec. 5.1). Ac-
 203 cording to our observations, pixels representing clear sky pixels also saturate and drift
 204 from their original hues at the end of the color scale, also presenting PVS.

205 Additional methods, employing multispectral approaches combining infrared cam-
 206 eras, polarimeters and Longwave (LW) have also been used to support surface cameras
 207 (Feister et al., 2000; Feister & Shields, 2005; Schade et al., 2008; Kreuter et al., 2009).
 208 Nevertheless, the image classification methods used employ the same approach of pix-
 209 elwise value-based segmentation using linear thresholding mentioned previously. These
 210 methods focus on different discriminating functions but keep the same dichotomizing ap-
 211 proach.

212 However, our experience has demonstrated that false positives, false negatives, non-
 213 classifiable patterns and more than two simultaneous patterns are always present in the
 214 task environment, and are not adequately handled by a dichotomizer. Dichotomizers also
 215 do not allow the definition of uncertainty and errors by parametric analysis nor handle
 216 multi-category patterns. Another limitation of these approaches is the reduction of di-

217 dimensionality, underestimating the resources existent on the task environment causing
 218 loss of information, performance reduction, and increased error (Jain et al., 2000).

219 **3.1 The role of sky luminance of images taken from the surface**

220 The independent variables analyzed by automated systems using cameras are the
 221 image pixel values in the *RGB* color space (Gonzalez & Woods, 2002). In a color im-
 222 age each pixel represents a 3-dimensional data unit of 24 bits/pixel with 8 bits or 255
 223 digital number values (DN) for every color dimension, representing a total of 256^3 dif-
 224 ferent colors. However, only 256 and 256^2 colors are available from respectively
 225 1-dimensional and 2-dimensional counterpart classification methods. We understand that
 226 a higher-dimensional pixel representation, preserving acquired data, can better describe
 227 the variability of these data and improve the quality of pattern classification in the color
 228 space. The luminance axis of *DN* values, for example, scale relatively to the main di-
 229 agonal of color space. Luminance is limited to a range of $\sqrt{256^2 + 256^2 + 256^2} = 443.40$
 230 for the *RGB* cube in color images with 8 bits per channel (S. L. Mantelli Neto, 2010).
 231 For a 2-dimensional model employing typical 8 bit per channel color codification, on the
 232 other side, only a projection of the *RB* diagonal $\sqrt{256^2 + 256^2} = 362.03$ is available.
 233 An increase of pixel resolution using a finer camera will improve the image quality, but
 234 will not affect the luminance scale of the system due to the limited span of the color do-
 235 main. This restriction is valid for any image resolution. Algorithms that could be used
 236 to determine cloud height and type would also have its performance downgraded by di-
 237 mensional reduction.

238 Cameras and displays are not able to sense and reproduce the luminance scale ex-
 239 istent on natural scenes (Tsin et al., 2001; Koslof, 2006; Inanici & Navvab, 2006), reduc-
 240 ing the overall amount of useful data. That restriction causes pixel value saturation (PVS)
 241 ¹, pattern distribution distortions and difficulties when applying digital image process-
 242 ing algorithms, ML methods (i.e. neural networks, fuzzy logic, genetic algorithms and
 243 so on) or statistical analysis. These restrictions also apply to file structures registering
 244 the information representing the images; i.e. JPG, PNG, BMP. Table 1 illustrates some
 245 light conditions existent in natural scenes that could be perceived by the human eye.

Table 1. Levels of luminance found in some external environment. Source: (Reinhard et al., 2005, chap. 1 tab. 1.1).

Condition	luminance cd/m^2
Starlight	10^{-3}
Moonlight	10^{-1}
Internal Illumination	10^2
Sunlight	10^5
Maximum intensity of CRT monitors	10^2

246 The table shows that natural scenes span luminances up to eight orders of mag-
 247 nitude, ranging from nearly 10^{-3} to 10^5 cd/m^2 . Images obtained by SOS camera equip-
 248 ment will operate on a 10^2 cd/m^2 scale, causing saturation and pattern distortions. An
 249 example of this fact is illustrated on figures 1 and 2. Figure 1 illustrates a considerable
 250 amount of saturated pixel values and their regions on RGB and HSL color spaces for a

¹ It is important to notice the difference between HSL color saturation (S), from of pixel value saturation (PVS) The former, is the name of the color space dimension. The latter is a distortion caused by a pattern that spans above the end of color scale.

251 blue sky pattern. Figure 2 show the saturation points indicated by saturation of blue
 252 SB, saturation of green SG and total saturation ST labels. It is important to notice that
 253 on the saturated region SAT, is not possible to discriminate between a cloud, a saturated
 254 cloud a saturated blue sky pixel, because they have the same end of scale value. The present
 255 technologies are slowly overcoming these limitations with high contrast monitors, new
 256 file formats and *High Dynamic Range Imaging* (HDRI) (Inanici & Navvab, 2006; Rein-
 257 hard et al., 2005; Debevec & Malik, 1997; Moeck & Anaokar, 2006). However, they are
 258 not presently available at everyday meteorologic and photo-voltaic settings and their use-
 259 age is left as a suggestion for future research work. Currently, the state-of-the-practice
 260 is limited to adapt a logarithmic luminance task environment to a linear camera sensor.
 261 As a consequence patterns are distorted or trimmed on high-intensity regions.

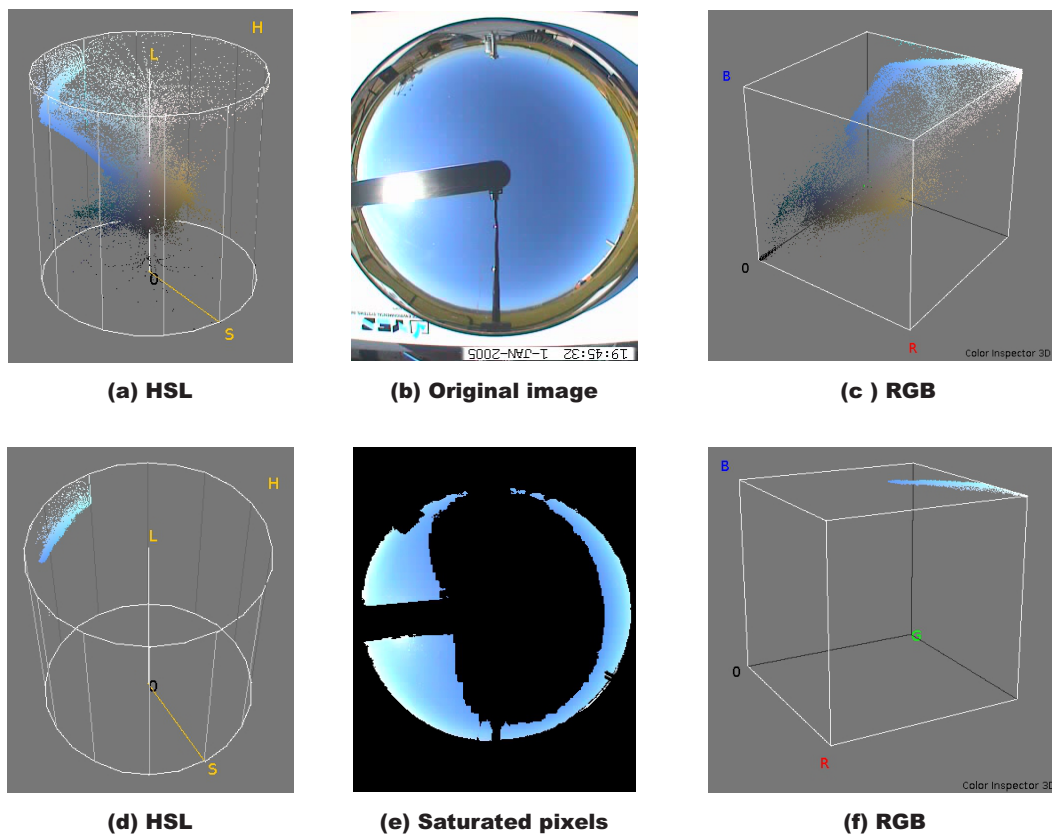


Figure 1. All sky images taken on Jan 1st 2005 at 19:45 GMT illustrating saturation problems on image. On first row, the original image and its respective pixel distribution on HSL and RGB color spaces. On the second row only saturated pixels are illustrated on image and their respective pixel distributions on HSL and RGB color spaces. Equipment self image, shading band, camera support, surrounding obstructions and non saturated pixels data were masked to black.

4 Material and methods

In this section we will describe our experimental set up, our dataset and data acquisition parameters and the methodology we followed in order to develop our approach.

265 **4.1 Experimental set up**

266 Our experiment was deployed at the Brazilian Space Research Institute (INPE) South-
 267 ern Observatory station (SMS) (<http://sonda.ccst.inpe.br/basedados/saomartinho.html>)
 268 located in São Martinho da Serra City, Rio Grande do Sul State, Brazil LAT.: 29° 26'
 269 34" S (−29, 4428°), LONG.: 53° 49' 23" W (−53, 8231°), ALT.: 489m. The sky imager
 270 site is also co-located with a solar sun photometer (AERONET) ,
 271 (<http://aeronet.gsfc.nasa.gov/>) , a BSRN-compatible station (<http://www.bsrn.awi.de/>),
 272 a Brewer Spectrophotometer, UV sensors, etc. A detailed description of the environment
 273 and the data can be found in (S. L. Mantelli Neto et al., 2014). The dataset is freely avail-
 274 able at (S. Mantelli Neto & von Wangenheim, 2019).

275 The equipment used to obtain the images was a TSI-440A manufactured by YAN-
 276 KEE Environmental (<http://www.yesinc.com>). It belongs to the SONDA project
 277 (<http://sonda.ccst.inpe.br/index.html>). TSI acquires an image obtained from a reflec-
 278 tor with an observation angle of 160°. The resident software system allows images to be
 279 obtained automatically at selected intervals, only when the Sun above 5° of elevation.
 280 The resident program was changed in order to obtain images at lower elevation angles
 281 too. Image resolution of TSI is 352x288 or 1013x766 pixels per image. Nearly 50 % of the
 282 image pixels generated by the sky imager was not useful because they record horizon ob-
 283 structions (poles, buildings, etc.), equipment self-image and a mirror shading band. A
 284 total of 1630 images were analyzed, starting January 2005. Images were acquired in JPG
 285 file format during daylight every 15 minutes at Greenwich Mean Time (GMT).

286 We made the whole dataset we acquired for this work publicly available at our site.
 287 The dataset includes sky imager data and image masks and is available at
 288 (<http://www.lapix.ufsc.br/sky-monitoring-surface-cameras>). The dataset is composed
 289 of two .zip files: *Images* and *Masks*.

290 **4.2 Parametrization of task environment**

291 Atmospheric patterns have their volumetric distributions on color space distorted
 292 due to saturation of camera scale. Figure 2 illustrates the distribution of most common
 293 sky patterns on two different color spaces HSL and RGB. Arrows are labeled and used
 294 to indicate similar regions on image and both color spaces. Figure 2 (a) illustrates a typ-
 295 ical blue sky, or *Rayleigh Scattering Pattern* (Lillesand & Kiefer, 1994, sec. 1.3), (Naylor,
 296 2002, sec. 1.2) indicated on the figure as **RAYL**. **RAYL** was parametrized by a Bayesian
 297 method using supervised learning guided by exploratory data analysis (EDA) and mul-
 298 tivariate statistical analysis (MSA). Three clear sky images sampled from different days,
 299 showing different luminance conditions were used as a reference sample (S. L. Mantelli Neto
 300 et al., 2014, sec. 2.6.6). Non-sky pixels representing equipment self-image, shading band,
 301 poles, building obstructions and surrounding horizon also indicated on figure 2, were con-
 302 sidered outliers and manually removed by masking.

303 We defined the **RAYL** pattern was as a ground truth (GT) by an average, a co-
 304 variance and an error matrices indicated on equation 1.

$$\mathbf{RAYL} = \begin{vmatrix} 88.65 \\ 128.48 \\ 191.90 \end{vmatrix} + \begin{vmatrix} 147.90 & 190.11 & 235.32 \\ 190.11 & 275.95 & 342.42 \\ 235.32 & 342.42 & 456.11 \end{vmatrix} + \begin{vmatrix} 0.035 \\ 0.049 \\ 0.062 \end{vmatrix} \quad (1)$$

305 After parameterization of **RAYL** GT patterns, unknown pixels could be classified
 306 using the linear Mahalanobis distance (MD) (Mahalanobis, 1936), also known as statis-
 307 tical distance (A. R. Johnson & Wichern, 2007). MD has F-distribution and could be
 308 calculated according to the equation 2 recommended by Mahalanobis

$$D^2 = n(\mathbf{x} - \mu)^T \cdot \Sigma^{-1} \cdot (\mathbf{x} - \mu) \quad (2)$$

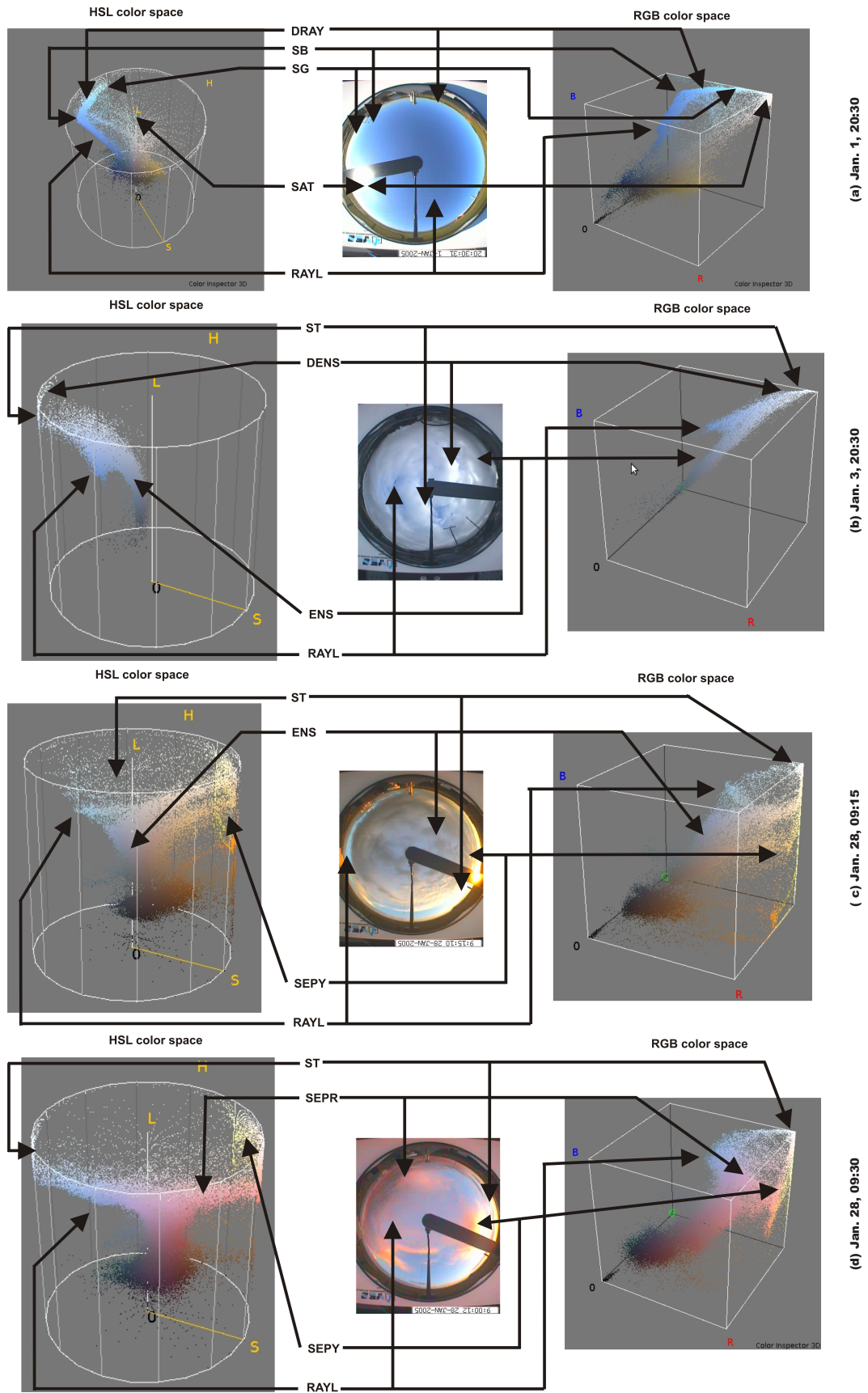


Figure 2. Sample images showing RGB and HSL pattern *loci*. A clear sky on (a), a cloudy sky on (b), a partly cloudy sky yellow colors on (c) and a partly cloudy sky red colors on (d).

309 where, D^2 : is the pixel squared MD from the GT pattern being classified, \mathbf{x} (r,g,b):
 310 is the pixel vector to be classified, represented by its color dimensions, μ : is the average
 311 GT vector, T : is the transpose matrix operation, $^{-1}$: is the inversion matrix operation,
 312 n : is the number of pixels used to determine the GT, Σ : is the GT covariance matrix.
 313 The Mahalanobis distance, when used to generate a customized color-space, is based upon
 314 a space generated from the covariance matrix of a set of reference pixel-values that rep-
 315 resent typical values for a given phenomenon. It has been employed successfully in the
 316 past by the authors in several different environments (Sobieranski et al., 2009) (Sobieranski
 317 et al., 2011) (de Carvalho et al., n.d.).

318 We established a discrimination threshold based on pixel values using F-scores on
 319 the same way as the traditional hypothesis testing. Statistical tables are easily found on
 320 related literature according to degrees of freedom on the formula suggested by (A. R. John-
 321 son & Wichern, 2007, chap. 5) and illustrated in the equation 3;

$$D^2 \leq \frac{(n-1)p}{(n-p)} F_{p,n-p,(\alpha)} \quad (3)$$

322 where, p : is the degrees of freedom corresponds to the number of color space di-
 323 mensions ($p = 3$), $n-p$: is the degree of freedom of the GT sampled population. If $(n-p)$
 324 ≥ 120 the statistics tables consider the degree of freedom to be infinite (∞), $\alpha =$
 325 $0,05$: is the level of confidence established for the evaluation test, $F_{p,n-p,(\alpha)}$: is the dis-
 326 crimination threshold from GT. This value is obtained from an percentage point from
 327 a F-distribution table. $F_{3,\infty,(0,1)} = 3,78$; $F_{3,\infty,(0,05)} = 2,61$; $F_{3,\infty,(0,01)} = 2,08$.

328 For practical reasons and large population cases, the distance and threshold val-
 329 ues needed to be adjusted according to the application otherwise this criteria could disc-
 330 card many pixels. The threshold values were tuned up to $D_{RAYL}^2 = 22.68$.

331 Figure 2 (a) shows a discontinuity point **SB** caused by the saturation of the blue
 332 dimension and a second discontinuity point **SG** caused by the saturation of the green
 333 dimension on the **RAYL** pattern. In this case the multivariate model is not valid for that
 334 region, due to a discontinuity. Mathematically the saturated portion of **RAYL** could
 335 be defined as a function in a different interval. This situation is also indicated on figure
 336 2 (a) as *Diffusion of Rayleigh Scattering* pattern or simply **DRAY**. A different criterion
 337 is necessary to classify **DRAY** because its occurrence *locus* is along the end of the scale
 338 planes. Most classification methods existent in literature have difficulties to classify **DRAY**
 339 and ignore or remove it.

340 A possible physical meaning of this specific saturation is the forward scattering caused
 341 by aerosols or water vapor, with a higher optical density near the solar disk or at low
 342 solar elevation angles (near the surface horizon) (Long et al., 2009), as illustrated in fig-
 343 ures 4(a)-4(d). Although the pattern transitions on camera image from **RAYL** to **DRAY**
 344 seems to be "smooth" due to pattern whitening; in the color space this transition is "abrupt".
 345 We believe that this discontinuity in the color space imposes a certain difficulty on clas-
 346 sification methods that erroneously consider a continuous distribution as given in the anal-
 347 ysis. Some approaches prefer to discard **DRAY** pixel data from its analysis (Qingyong
 348 et al., 2011). **DRAY**, however, still means mostly clear sky and represents a significant
 349 amount of image pixels that need to be classified. Pixels in the last region in the figure
 350 are near their full saturation value, indicated in the figure as **SAT**. For clarity, the **SAT**
 351 point on first column of figure 2 (a) is coincident to the center of HSL top circle. Clas-
 352 sification of **DRAY** and **SAT** is made by separating saturated blue dimension of pixel
 353 values obtained after determination of the Euclidean Geometric Distance (EGD) *locus*
 354 (S. L. Mantelli Neto et al., 2014, sec. 2.6.2), according to described by the equation 4.
 355 *EGD* is used because it considers the geometric location of pixels.

$$\mathbf{DRAY} = [(B \geq 255) \text{ AND } (EGD \geq 52.5)] \tag{4}$$

356 A typical distribution of patterns could also be noticed on a partially covered sky
 357 as illustrated on figure 2 (b). White clouds or *Non-Selective Scattering* pattern **NSS**
 358 (Lillesand & Kiefer, 1994, sec. 1.3) also have a typical distribution in the color space.
 359 Heavy gray clouds occur continuously in white cloud patterns inside the color space due
 360 to cloud thickening, but with smaller values of luminance (S. L. Mantelli et al., 2010).
 361 The same multivariate Bayesian method and *criteria* were used to parameterize **NSS**
 362 (S. L. Mantelli Neto et al., 2014, sec. 2.6.5). A *GT* pattern was defined as a ground truth
 363 (GT) by the average, covariance and an error matrices indicated on equation 5.

$$\mathbf{NSS} = \begin{vmatrix} 141.61 \\ 160.07 \\ 174.01 \end{vmatrix} + \begin{vmatrix} 889.144 & 944.645 & 1039.041 \\ 944.645 & 1092.758 & 1214.737 \\ 1039.041 & 1214.737 & 1390.940 \end{vmatrix} + \begin{vmatrix} 0.076 \\ 0.085 \\ 0.096 \end{vmatrix} \tag{5}$$

364 The pattern tune-up for classification of **NSS** using the Mahalanobis distance, in-
 365 dicated an statistical threshold value of $D_{NSS}^2 = 29.1$.

366 A discontinuity region was also observed in the distribution of **NSS** and is indi-
 367 cated as **ST** on figure 2 (b). This saturation was observed on several images and is de-
 368 noted in the present work as *Diffusion of Non-Selective Scattering* pattern **DNSS**. Clas-
 369 sification of **DNSS** occurrences is performed by separating the *B* blue dimension sat-
 370 urated pixel values obtained after determination of the Euclidean Geometric Distance
 371 (EGD) *locus* (S. L. Mantelli Neto et al., 2014, sec. 2.6.1), according to described by logic
 372 and arithmetic equation 6.

$$\mathbf{DNSS} = (B \geq 255) \text{ AND } (EGD \leq 52.5) \tag{6}$$

373 Additionally, extended observation intervals during daylight using the sky imager
 374 has shown the presence of new patterns, not previously reported by any reviewed method.
 375 Although they were obviously present, as illustrated in figures 2 (c) and 2 (d). By sun-
 376 set and sunrise, two color patterns in yellow and red could be noticed on sky scenes some-
 377 times simultaneously to white and gray clouds. Those patterns occur when the hues of
 378 Sunlight after sunset (or before sunrise) is reflected by clouds (Naylor, 2002, sec. 4.3),
 379 (Richards, 1995). Those two patterns occurs at different *locus* in color space, meaning
 380 clouds and were defined as distinct ones.

381 A yellow pattern is noticed when the sun is a bit higher above horizon indicating
 382 a *Selective Scattering Pattern in Yellow* (Naylor, 2002, sec. 1.2), in the present research
 383 denoted as **SEPY** (S. L. Mantelli Neto et al., 2014, sec. 2.6.3). The figure 2 (c) illus-
 384 trates *locus* occurrences of **SEPY** patterns in the HSL and RGB linear color spaces.

385 For the definition of **SEPY** patterns, a typical image was selected possessing a clear
 386 evidence of its presence, as illustrated in figure 2 (c). All the other patterns were cleared
 387 out by masking them from the image, and only **SEPY** was left (S. L. Mantelli Neto et
 388 al., 2014, sec.2.6.3). After the analysis of pattern occurrence, it was noticed that the best
 389 way to classify **SEPY** was by pixel hue value *H* interval discrimination of the HSL color
 390 space. Typical values were extracted and refined from image samples and defined by the
 391 logic and arithmetic equation 7.

$$\mathbf{SEPY} = \{\forall H \in [0, 1] \mid (H > 0.0833) \text{ AND } (H \leq 0.1667)\} \tag{7}$$

392 Red pattern is noticed when the sun is a bit lower when compared to yellow, in-
 393 dicated a *Selective Scattering Pattern in Red*, in the present research indicated as **SEPR**.

394 The figure 2 (d) illustrates a *locus* occurrence of **SEPR** pattern in the HSL and RGB
 395 linear color spaces. Saturation was also noticed in the **SEPY** and **SEPR** pixel values.
 396 However, the discriminating method used based on Hue angle, allowed a precise separa-
 397 tion even in the presence of saturated values. Hue angle discrimination was not used
 398 to separate **RAYL** and **NSS** patterns, because they occur in coincident hue angles and
 399 are very difficult to discriminate.

400 For the definition of the **SEPR** pattern, a typical image possessing a clear evidence
 401 of its presence was also selected. All other patterns were cleared out by masking and only
 402 **SEPR** was left (S. L. Mantelli Neto et al., 2014, sec. 2.6.4). After analyzing the image
 403 and the pattern occurrence, we noticed that the best way to classify **SEPR** was by hue
 404 H pixel value discrimination of the HSL color space. The **SEPR locus** is very distinct
 405 from the other patterns and can be easily discriminated employing this method. Typ-
 406 ical values were extracted and refined from image samples and defined by the logic and
 407 arithmetic equation 8.

$$\mathbf{SEPR} = \{\forall H \in [0, 1] \mid (H \leq 0.0833)\} \tag{8}$$

408 We chose a hierarchic order to be used in the classification of individual pixels, with
 409 a stepwise masking-out of the already classified pixels, in order to analyze saturated re-
 410 gions with higher pixel values prior to non-saturated ones. Otherwise, the algorithm will
 411 not properly classify the patterns. Table 2 describes the resume of the principal crite-
 412 ria used for classification in the present work. After the definition of the criteria used
 413 for pattern classification, they were implemented using a software prototype to gener-
 414 ate the results. Figure 3 illustrates the stepwise hierarchical application of the proposed
 415 color-metric.

Table 2. Principal classification criteria proposed in the present research; h is the hierarchical order, the Pattern attributed and the meaning.

h	Criterion of classification	Pattern	Meaning
1	$\{\forall R, G, B \in [0, 255] \mid ((B \geq 255) \text{ AND } (EGD \leq 52, 5))\}$	DNSS	cover
2	$\{\forall R, G, B \in [0, 255] \mid ((B \geq 255) \text{ AND } (EGD \geq 52, 5))\}$	DRAY	clear
3	$\{\forall H \in [0, 1] \mid (H > 0.0833) \text{ AND } (H \leq 0.1667)\}$	SEPY	cover
4	$\{\forall H \in [0, 1] \mid (H \leq 0.0833)\}$	SEPR	cover
5	$D_{NSS}^2 \leq F_{p,n-p,(\alpha)} = 29, 01$	NSS	cover
6	$D_{RAY}^2 \leq F_{p,n-p,(\alpha)} = 22, 68$	RAY	clear
7	Non classifiable on above cases	NC	undetermined

416 5 Results

417 Current method was applied on 1630 surface images. The detailed method devel-
 418 oped with parameters, results and comparison with related literature on nearly 7000 im-
 419 ages, figures, tables, and charts are too massive to be included in the present document.
 420 They are made available as a 511 page technical report (S. L. Mantelli Neto et al., 2014).
 421 The representative set of images on figure 4 shows original and analyzed images of mixed,
 422 clear, cloudy conditions illustrating the proposed classification patterns. Figure 5 indi-
 423 cates the segmentation color codes used on next figures.

424 Figures 4(a)-4(d) highlight the significant amount of saturated pixels classified as
 425 **DRAY** in light blue. Some methods found on related literature have more difficulty to
 426 classify higher intensity pixels specially when the sun is at lower solar elevation angles

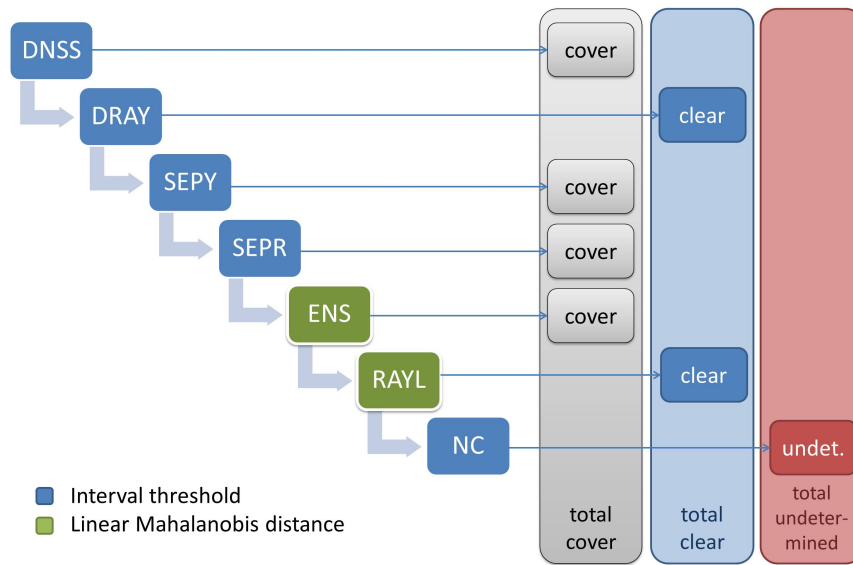


Figure 3. Hierarchical step-wise application of our color similarity metric. The classes *cover*, *clear* and *undetermined* are additive.

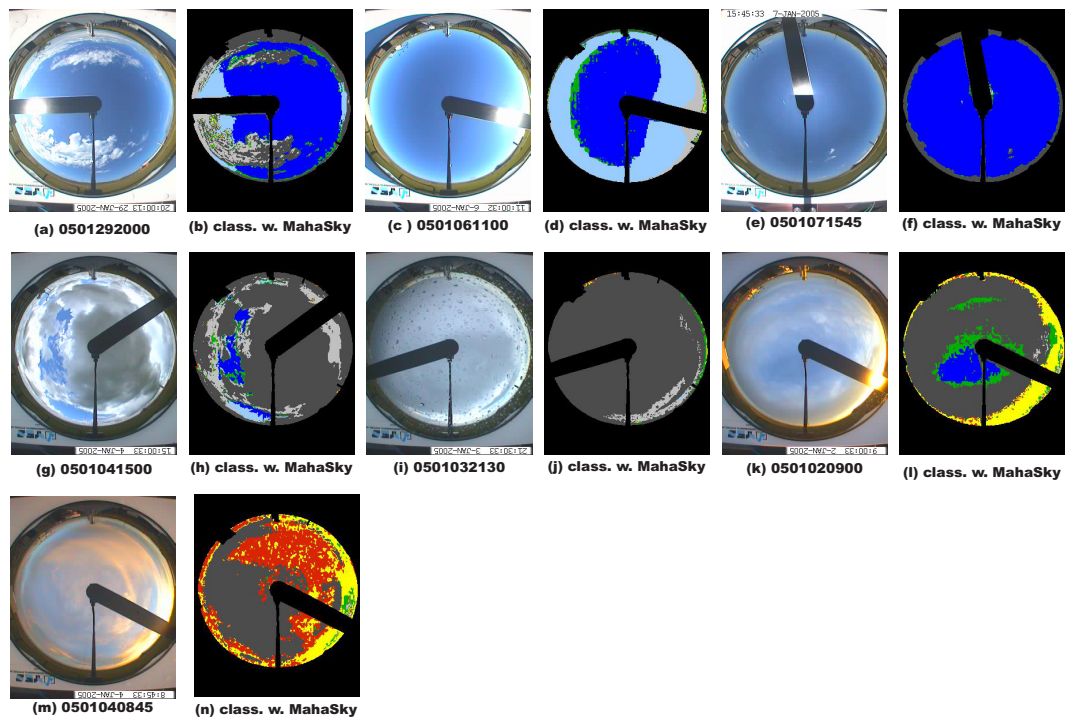


Figure 4. Original images and analysis results for various sky, ranging from clear through covered and presenting lower solar elevation.

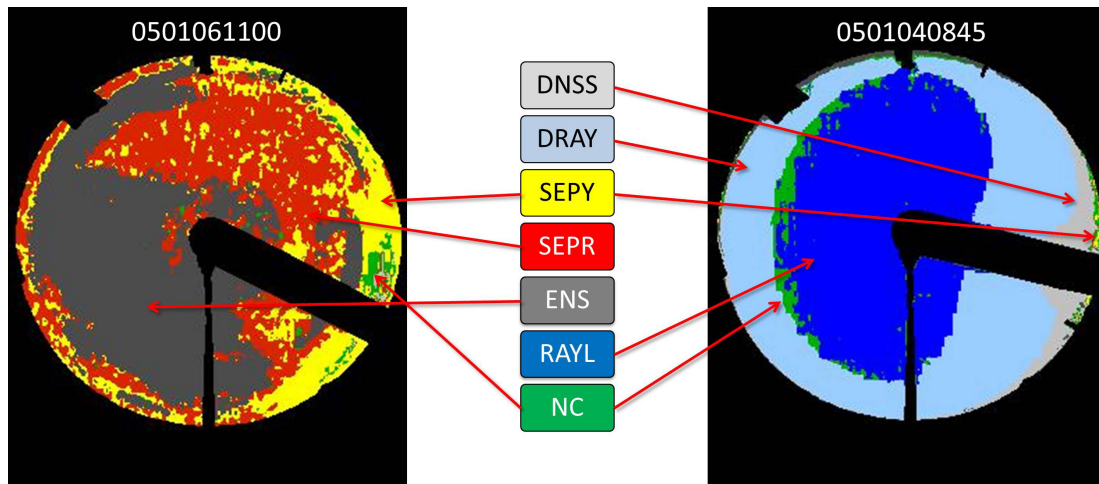


Figure 5. Detail of the results for two different images showing the color-code also employed in figure 4, 7 and 8.

427 or longer slant optical path. Additional evidence of this fact could be observed when we
 428 plot all occurrences of **DRAY** or saturated pixels obtained from all analyzed images as
 429 described on figure 6. The occurrence of other patterns like **DNSS**, **SEPY**, **SEPR** also
 430 increased with longer slant optical path, although in small proportions. Figures 4(e)-4(f)
 431 illustrates a (**RAYL**) in darker blue on a clear sky at higher solar elevations where in gen-
 432 eral, the classification methods perform without difficulty. Figures 4(g)-4(j) illustrates
 433 a covered sky where could be noticed saturated pixels or **DNSS** pattern in light gray.
 434 Figures 4(k)-4(l) illustrates **SEPY** pattern and Figures 4(m)-4(n) illustrates **SEPR** pat-
 435 tern respectively in yellow and red colors. It is important to notice that **SEPY** and **SEPR**
 436 are not possible to be determined by 2-D classification approaches, because the green
 437 dimension is not considered. Finally, the **NC** pattern in green occurs mostly at the tran-
 438 sition between **RAYL** and **NSS** where thresholding could be improved by refining. Few
 439 methods deal with **NC** patterns that could reach up to 60% at very low solar elevations
 440 with extremely dim light conditions. This **NC** amount could explain two effects: (i) the
 441 reason there is a small bias among different methods and (ii) why some classification meth-
 442 ods avoid analyze images at low solar elevations ($< 5^\circ$) under little light conditions. A
 443 full month of all sky image analysis where all the above-mentioned situations were clas-
 444 sified could be observed in figure 7. This kind of data could be latter on used on the as-
 445 sements, comparisons and validation of solar energy reaching the surface.

446 Figure 8 shows the detailed analysis of a mixed cloud condition day with occur-
 447 rence of all proposed patterns and compares the results with a dicotomizer method.

448 5.1 Validation

449 Some aspects have to be considered concerning the methods to be compared for
 450 this validation, since they have to be subject to the *same sources of variability*
 451 (Montgomery, 2005, ch. 1). Although the present method was implemented in the
 452 3-dimensional domain, it is being compared with a 2-dimensional domain method. We
 453 employed a look-up table method, using the *clear/cover* classification obtained from ta-
 454 ble 2 to convert the results to binary all-sky dichotomizer approach for the sake of our
 455 comparison. It is important to notice that there is no correspondence of **NC** pattern on
 456 dichotomizer. This will cause small aleatory difference on results.

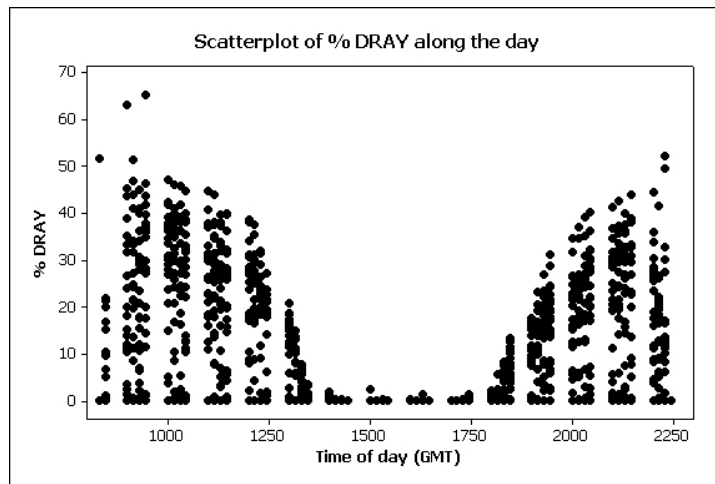


Figure 6. Percentage of **DRAY** daily occurrences for January 2005.

457 After conversion, the proposed method (MahaSky) was statistically paired compared to Long (Long et al., 2006) and EGD (S. L. Mantelli Neto et al., 2014) methods,
 458 and their differences are illustrated on figure 9.
 459

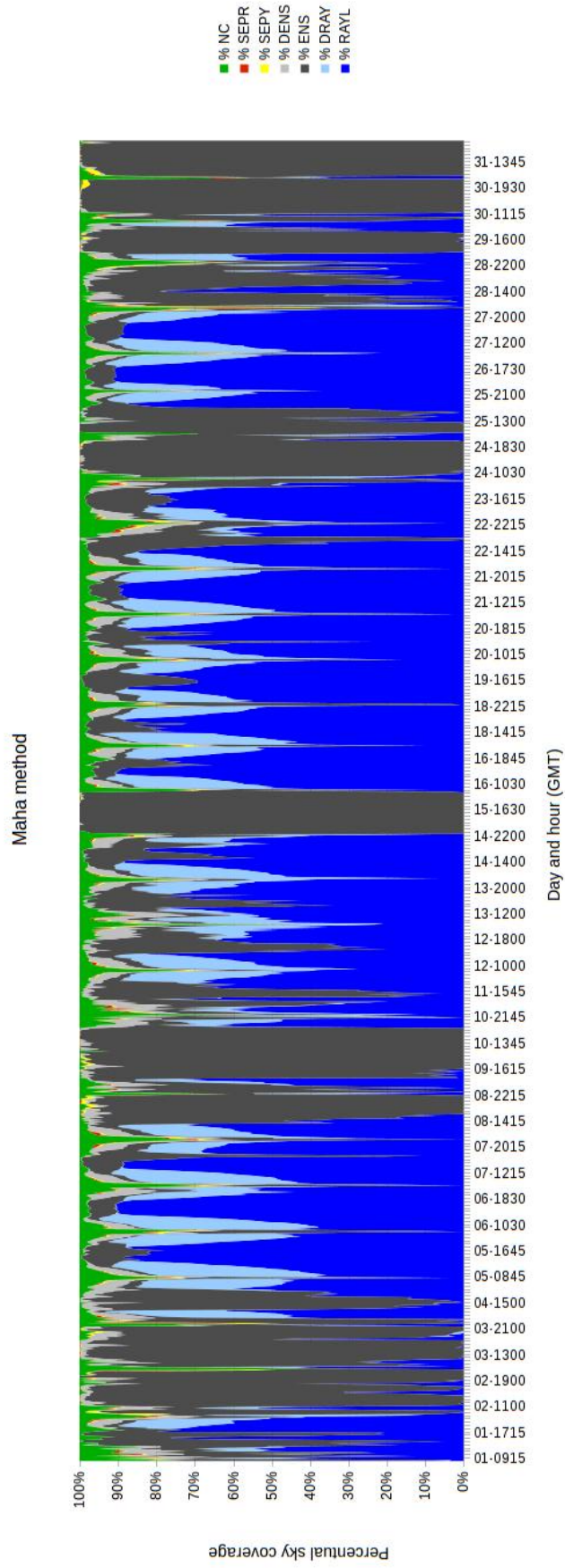


Figure 7. Cloud coverage estimation using the Maha method for January 2005.

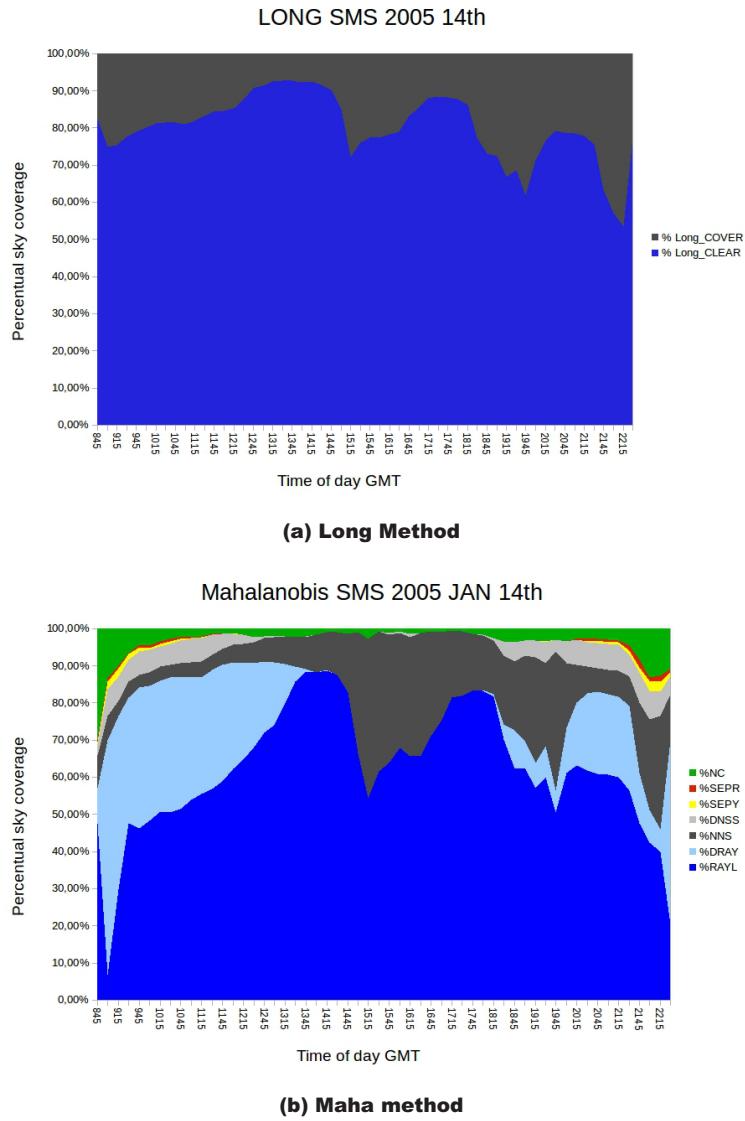


Figure 8. One day data comparison between a dichotomizer and Maha methods for January 14th 2005.

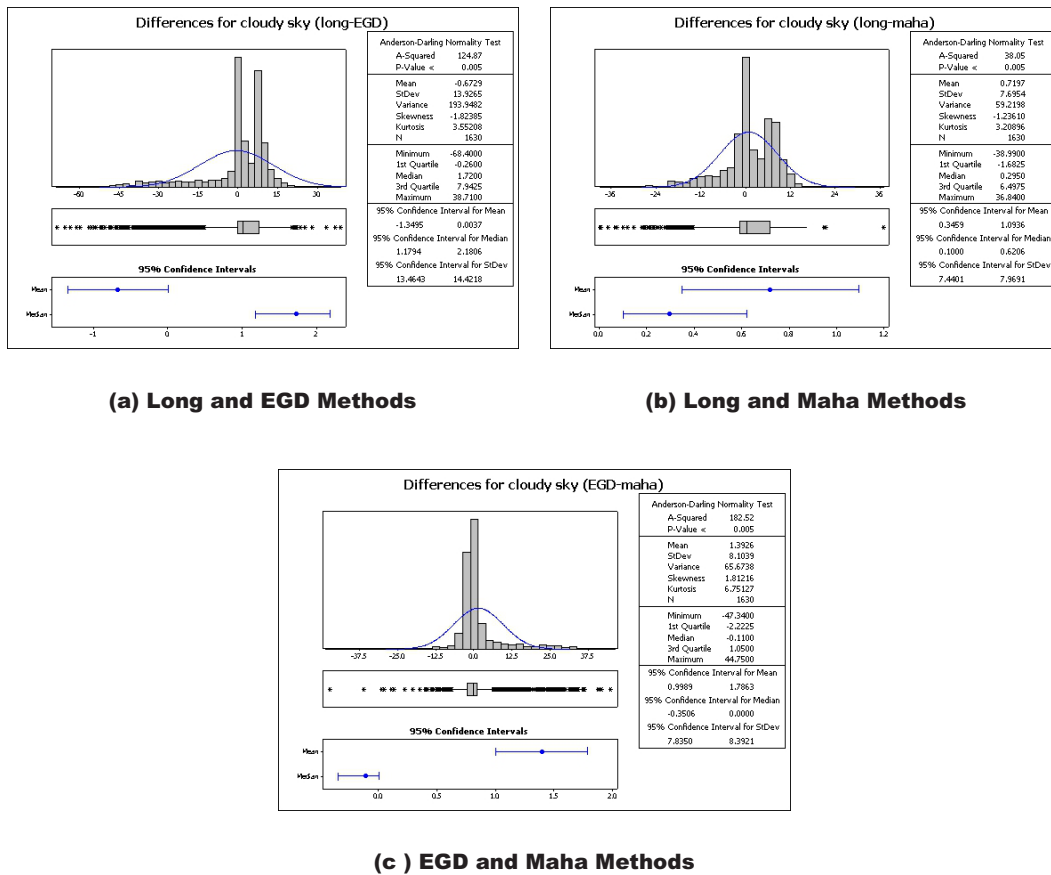


Figure 9. Paired comparison of differences in cloud coverage estimation among methods for January 2005.

460 The paired differences on 1630 images among the proposed downgraded method
 461 and two other ones existent on literature, were checked by hypothesis testing. Analy-
 462 sis indicated that z-scores differences existent were bellow the critical values, indicating
 463 that methods are similar, and aleatory factors caused the differences. Differences could
 464 also be due to the establishment of a different task domain dimension used during anal-
 465 ysis. Another difference noticed was the different training data set used to define the pa-
 466 rameters, which are subject to distinct meteorological and atmospheric conditions. The
 467 **NC** pattern not used on (Long et al., 2006) method could introduce a bias because it
 468 does not consider the **NC** class on pixel for analysis.

469 **6 Conclusions**

470 Current methods presented in the literature, allied to camera restrictions, limit the
 471 overall performance and, consequently, the quality of the results in monitoring the sky
 472 for cloud coverage estimation. This indicates that the more features are available for cloud
 473 classification, the better will be the conditions to classify cloud types, species, and va-
 474 rieties.

475 Additionally, we presented a new methodology for classification of sky patterns us-
 476 ing surface cameras. A key feature of the method is that it does not focus only on the
 477 development of a new pixel classification algorithm but considers several aspects of Syn-
 478 optoc Observation Systems (SOS) as a whole. Common practices used by related works
 479 during classification, like dichotomizers and reduction of dimensionality were not employed
 480 in this work. Dichotomizer methods were not used for two reasons. Dichotomizer does
 481 not handle properly false positives and negatives because they are assigned to either one
 482 or other pattern. The dichotomizer does not handle multi-category patterns, necessary
 483 for the implementation of the proposed method. Reduction of dimensionality causes loss
 484 of information to be classified.

485 3-dimensional analysis on RGB and HSL color spaces allowed the proposed method
 486 to obtain more information from sky data. One example is the better usage of the lu-
 487 minance diagonal on the whole RGB color space, which spans the interval [0,442], against
 488 [0,361] for 2-D and [0,255] for 1-D counterparts . Exploratory Data analysis indicated
 489 that patterns, existent on logarithmic luminance scale domain are distorted due to sig-
 490 nal saturation by linear systems used to monitor and store data. Until the present tech-
 491 nology does not develop systems that match the human perceived luminance scale, ad-
 492 ditional patterns were presented in a more appropriate approach, according to atmospheric
 493 physics by considering them discontinued functions intervals as follows. Rayleigh scat-
 494 tering in blue (RAYL), non-specific scattering indicated by clouds (NSS), selective scat-
 495 tering in red (SEPR) and yellow (SEPY) colors, saturation of Rayleigh scattering caused
 496 by forward scattering (DRAYL) and saturation of non-specific scattering indicated by
 497 clouds (DNSS). New proposed patterns allowed an appropriate analysis of the images
 498 bellow 5° of solar elevation, allowing the extension of the daylight observation period.
 499 The image dataset employed in this work is freely available at (S. Mantelli Neto & von
 500 Wangenheim, 2019).

501 **6.1 Limitations and Future Work**

502 It is possible that the model may require some adjustment to take into account sea-
 503 sonal or local variability. A better classification could be achieved if the full pattern dis-
 504 tribution was not trimmed by saturation and distortions were not limited by linear sen-
 505 sor response. A combination of techniques to detect different features could also be used
 506 to produce a more meaningful result.

507 Another limitation of this approach is that it, as do all the other models we inves-
 508 tigated, does not take into consideration spatial context information of the pixels in the

509 image. It is a method that operates exclusively in the domain of pixel values, classify-
 510 ing each pixel independently and without taking into consideration any information that
 511 could be gained from its surroundings. Observing figures 4 and 5 it is possible to observe
 512 that *NC* patterns sometimes appear in boundaries between regions, as in fig.4(d) and
 513 (h), where *NC* forms boundaries between *DRAY* and *RAYL* and between *NSS* and
 514 *RAYL* regions respectively, and sometimes *NC* patterns appear exclusively inside other
 515 patterns, as in fig.4(l), where the *NC* pattern is contained inside a *SEPY* region. One
 516 could postulate that, when *NC* patterns occur inside homogeneous regions or in bound-
 517 aries between regions that have the same *meaning*, as in *DRAY* and *RAYL* in fig.4(d),
 518 they can be computed to the total area of that specific *meaning*. These issues are left
 519 as suggestions for future investigation and implementation.

520 7 Acknowledgements

521 The authors would like to thank INPE operational teams of INPE/CPTEC/LIM
 522 (<http://lim1.cptec.inpe.br/lim/in>), and INPE/CRS/OES
 523 (<http://www.inpe.br/crs/institucional.php>) and INPE/CCST/LABREN
 524 (<http://www.ccst.inpe.br/>) and the research team of UFSC-INCOD
 525 (<http://www.incod.ufsc.br/?lang=en>), specially to Luís Gustavo Lorgus Decker for help-
 526 ing with image masking. We also like to thank the grants from FINEP by SONDA project
 527 (22.01.0569.00), PETROBRAS (0050.0029348.07.4) and CNPQ by HIDROCAM project
 528 (478694/2011-3).

529 References

- 530 Bojanowski, J. S., Donatelli, M., Skidmore, A. K., & A., V. (2013, Novem-
 531 ber). An auto-calibration procedure for empirical solar radiation mod-
 532 els. *Environmental Modelling & Software*, *49*, 118-128. Retrieved from
 533 <http://www.sciencedirect.com/science/article/pii/S1364815213001801>
 534 doi: 10.1016/j.envsoft.2013.08.002
- 535 Bradley, E., Roberts, D., & Still, C. (2010, January 2010). Design of an image anal-
 536 ysis website for phenological and meteorological monitoring. *Environmental*
 537 *Modelling & Software*, *25*(1), 107-116. doi: 10.1016/j.envsoft.2009.07.006
- 538 Cazorla, A., Husillos, C., Antón, M., & Alados-Arboledas, L. (2015). Multi-
 539 exposure adaptive threshold technique for cloud detection with sky im-
 540 agers. *solar energy*, *114*, 268-277. Retrieved from [http://ac.elsa-](http://ac.elsa-cdn.com/S0038092X1500064X/1-s2.0-S0038092X1500064X-main.pdf?_tid=de4a79a6-704a-11e6-88aa-00000aacb362&acdnat=1472737934_8a2de6b8f6a1d0dd864291ab50d6598c)
 541 [-cdn.com/S0038092X1500064X/1-s2.0-S0038092X1500064X-main.pdf](http://ac.elsa-cdn.com/S0038092X1500064X/1-s2.0-S0038092X1500064X-main.pdf?_tid=de4a79a6-704a-11e6-88aa-00000aacb362&acdnat=1472737934_8a2de6b8f6a1d0dd864291ab50d6598c)
 542 [?_tid=de4a79a6-704a-11e6-88aa-00000aacb362&acdnat=1472737934](http://ac.elsa-cdn.com/S0038092X1500064X/1-s2.0-S0038092X1500064X-main.pdf?_tid=de4a79a6-704a-11e6-88aa-00000aacb362&acdnat=1472737934_8a2de6b8f6a1d0dd864291ab50d6598c)
 543 [_8a2de6b8f6a1d0dd864291ab50d6598c](http://ac.elsa-cdn.com/S0038092X1500064X/1-s2.0-S0038092X1500064X-main.pdf?_tid=de4a79a6-704a-11e6-88aa-00000aacb362&acdnat=1472737934_8a2de6b8f6a1d0dd864291ab50d6598c) doi: [http://dx.doi.org/10.1016/](http://dx.doi.org/10.1016/j.solener.2015.02.006)
 544 [j.solener.2015.02.006](http://dx.doi.org/10.1016/j.solener.2015.02.006)
- 545 Cazorla, A., Olmo, F. J., & Alados-Arboledas, L. (2008). Development of a sky
 546 imager for cloud cover assessment. *Journal of the Optical Society of America*,
 547 *25*(1), 29–39. doi: <http://dx.doi.org/10.1364/JOSAA.25.000029>
- 548 Cess, R. D., Zhang, M. H., Minnis, P., Corsetti, L., Dutton, E., Forgan, B., ...
 549 Zhou, Y. (1995). Absorption of solar radiation by clouds: Observations versus
 550 models. *Science*, *267*(5197), 496-499. doi: 10.1126/science.267.5197.496
- 551 Debevec, P. E., & Malik, J. (1997). *Recovering high dynamic range radiance maps*
 552 *from photographs* (Tech. Rep.). University of California at Berkeley. Retrieved
 553 from <http://www.cs.berkeley.edu/~malik/papers/debevec-malik97.pdf>
- 554 de Carvalho, L. E., Neto, S. M., Sobieranski, A., Comunello, E., & von Wangen-
 555 heim, A. (n.d.). Improving graph-based image segmentation using nonlinear
 556 color similarity metrics. *International Journal of Image and Graphics*, *15*(4),
 557 1550018.
- 558 Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Patterns classification* (2nd ed.).
 559 John Wiley & Sons. Retrieved from <http://www.wiley.com/WileyCDA/>

560 WileyTitle/productCd-111858600X.html

561 Feister, U., & Shields, J. (2005, 10). Cloud and radiance measurements with the
562 vis/nir daylight whole sky imager at lindenber (germany). *Meteorologische*
563 *Zeitschrift*, 14(5), 627-639. Retrieved from [http://dx.doi.org/10.1127/](http://dx.doi.org/10.1127/0941-2948/2005/0066)
564 [0941-2948/2005/0066](http://dx.doi.org/10.1127/0941-2948/2005/0066) doi: 10.1127/0941-2948/2005/0066

565 Feister, U., Shields, J., Karr, M., Johnson, R., Dehne, K., Woldt, M., ... Potsdam,
566 M. O. (2000). *Ground-based cloud images and sky radiances in the visible*
567 *and near infrared region from whole sky imager measurements*. Retrieved from
568 [http://jshields.ucsd.edu/publications/pdfs/22%20Feister%202000](http://jshields.ucsd.edu/publications/pdfs/22%20Feister%202000.pdf)
569 [.pdf](http://jshields.ucsd.edu/publications/pdfs/22%20Feister%202000.pdf)

570 Gonzalez, R. C., & Woods, R. E. (2002). *Digital image processing* (2nd ed.). Pren-
571 tice Hall.

572 Gonzalez, R. C., & Woods, R. E. (2007). *Digital image processing* (3rd ed.).
573 Prentice Hall. Retrieved from [http://folk.uio.no/ainard/Folder2/](http://folk.uio.no/ainard/Folder2/Digital%20Image%20Processing%203rd%20ed.%20-%20R.%20Gonzalez,%20R.%20Woods.pdf)
574 [Digital%20Image%20Processing%203rd%20ed.%20-%20R.%20Gonzalez,%](http://folk.uio.no/ainard/Folder2/Digital%20Image%20Processing%203rd%20ed.%20-%20R.%20Gonzalez,%20R.%20Woods.pdf)
575 [20R.%20Woods.pdf](http://folk.uio.no/ainard/Folder2/Digital%20Image%20Processing%203rd%20ed.%20-%20R.%20Gonzalez,%20R.%20Woods.pdf)

576 Grossberg, M. D., & Nayar, S. K. (2003). Determining the camera response from
577 images: What is knowable? *IEEE TRANSACTIONS ON PATTERN ANAL-*
578 *YSIS AND MACHINE INTELLIGENCE*, 25(11), 1455-1467. Retrieved from
579 <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=1240119> doi:
580 10.1109/TPAMI.2003.1240119

581 Harrison, R. G., Chalmers, N., & Hogan, R. J. (2008). Retrospective cloud determi-
582 nations from surface solar radiation measurements. *Atmospheric Research*, 90,
583 54-62. doi: 10.1016/j.atmosres.2008.04.001

584 Hu, Y., & Stamnes, K. (2000). Climate sensitivity to cloud optical properties. *Tel-*
585 *lus*, 52, 81 - 93. doi: 10.1034/j.1600-0889.2000.00993.x

586 Inanici, M. N., & Navvab, M. (2006). The virtual lighting laboratory: Per-pixel lu-
587 minance data analysis. *LEUKOS*, 3,2, 89?104. doi: 10.1582/LEUKOS.2006.03
588 .02.001

589 Iqbal, M. (1983). *An introduction to solar radiation*. New York Academic Press.

590 Jain, A., Duin, R., & Mao, J. (2000, Jan). Statistical pattern recognition: a review.
591 *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(1),
592 4-37. doi: 10.1109/34.824819

593 Johannet, A., Vayssade, B., & Bertin, D. (2007). Neural networks: From black box
594 towards transparent box application to evapotranspiration modeling. *World*
595 *Academy of Science, Engineering and Technology*, 30, 162-169. Retrieved from
596 <http://www.waset.org/journals/waset/v30.php>

597 Johnson, A. R., & Wichern, D. W. (2007). *Applied multivariate statistical anal-*
598 *ysis* (6th ed.). Pearson Education International. Retrieved from [https://](https://www.pearsonhighered.com/program/Johnson-Applied-Multivariate-Statistical-Analysis-6th-Edition/PGM274834.html)
599 [www.pearsonhighered.com/program/Johnson-Applied-Multivariate](https://www.pearsonhighered.com/program/Johnson-Applied-Multivariate-Statistical-Analysis-6th-Edition/PGM274834.html)
600 [-Statistical-Analysis-6th-Edition/PGM274834.html](https://www.pearsonhighered.com/program/Johnson-Applied-Multivariate-Statistical-Analysis-6th-Edition/PGM274834.html)

601 Johnson, H. W. S. J., R. (1989.). *Automated visibility and cloud cover mea-*
602 *surements with a solid-state imaging system*. (Tech. Rep. Nos. 89-7, GL-
603 TR-89-0061). University of California, San Diego, Scripps Institution of
604 Oceanography, Marine Physical Laboratory, SIO Ref. Retrieved from
605 <file:///home/sylvio/Downloads/ADA216906.pdf> (128 pp.)

606 Kasten, F., & Czeplak, G. (1980). Solar and terrestrial radiation dependent on the
607 amount and type of cloud. *Solar Energy*, 24(2), 177 - 189. Retrieved from
608 <http://www.sciencedirect.com/science/article/pii/0038092X80903916>
609 doi: [https://doi.org/10.1016/0038-092X\(80\)90391-6](https://doi.org/10.1016/0038-092X(80)90391-6)

610 Kazantzidis, A., Tzoumanikas, P., Bais, A., Fotopoulos, S., & Economou, G.
611 (2012). Cloud detection and classification with the use of whole-sky ground-
612 based images. *Atmospheric Research*, 113(0), 80 - 88. Retrieved from
613 <http://www.sciencedirect.com/science/article/pii/S0169809512001342>
614 doi: <http://dx.doi.org/10.1016/j.atmosres.2012.05.005>

- 615 Koslof, T. (2006). *Visualizing high dynamic range images* (Tech. Rep.). University
616 of California at Berkeley. Retrieved from [http://vis.berkeley.edu/courses/
617 cs294-10-sp06/wiki/images/5/52/ToddWriteup.pdf](http://vis.berkeley.edu/courses/cs294-10-sp06/wiki/images/5/52/ToddWriteup.pdf)
- 618 Kreuter, A., Zangerl, M., Schwarzmann, M., & Blumthaler, M. (2009). All-sky
619 imaging: a simple, versatile system for atmospheric research. *Applied Op-
620 tics*, 48, 1091 - 1097. Retrieved from [http://www.opticsinfobase.org/
621 DirectPDFAccess/503B44E9-BDB9-137E-C455BA62A9D969C4_176587.pdf?da=
622 1&id=176587&seq=0](http://www.opticsinfobase.org/DirectPDFAccess/503B44E9-BDB9-137E-C455BA62A9D969C4_176587.pdf?da=1&id=176587&seq=0) doi: 10.1364/AO.48.001091
- 623 Kurtz, B., Mejia, F., & Kleissl, J. (2017, December). A virtual sky imager
624 testbed for solar energy forecasting. *Solar Energy*, 158, 753-759. Re-
625 trieved from [https://www.sciencedirect.com/science/article/pii/
626 S0038092X1730899X](https://www.sciencedirect.com/science/article/pii/S0038092X1730899X) doi: <https://doi.org/10.1016/j.solener.2017.10.036>
- 627 Lillesand, T. M., & Kiefer, R. W. (1994). *Remote sensing and image interpretation*.
628 John Wiley & Sons.
- 629 Long, C. N., Dutton, E. G., Augustine, J. A., Wiscombe, W., Wild, M., McFar-
630 lane, S. A., & Flynn, C. J. (2009). Significant decadal brightening of
631 downwelling shortwave in the continental united states. *Journal of Geo-
632 physical Research: Atmospheres*, 114(D10), n/a-n/a. Retrieved from
633 <http://dx.doi.org/10.1029/2008JD011263> doi: 10.1029/2008JD011263
- 634 Long, C. N., Sabburg, J. M., Calbo, J., & Page, J. D. (2006, may). Retrieving
635 cloud characteristics from ground-based daytime color all-sky im-
636 ages. *Journal of Atmospheric and Oceanic Technology*, 23, 633-652. doi:
637 <http://dx.doi.org/10.1175/JTECH1875.1>
- 638 Mahalanobis, P. C. (1936). On the generalized distance in statistics. *Proceedings Na-
639 tional Institute of Science. India*. Retrieved from [http://ir.isical.ac.in/
640 dspace/handle/1/1268](http://ir.isical.ac.in/dspace/handle/1/1268)
- 641 Mantelli, S. (2001). *Desenvolvimento de uma nova metodologia para a estima-
642 tiva da cobertura de nuvens usando uma câmera de superfície e comparando
643 com imagens de satélite*. (Master's thesis, Universidade Federal de Santa
644 Catarina, Departamento de Informática e Estatística). Retrieved from
645 <http://repositorio.ufsc.br/xmlui/handle/123456789/83106>
- 646 Mantelli, S. L., v. Wangenheim, A., & Pereira, E. B. (2005). Modelo prelim-
647 inar de estimativa de cobertura de nuvens, no espaço de cores rgb obti-
648 das a partir de imageador automático. In *Proc. xii symp. brasileiro de
649 sensoriamento remoto, goiania brazil*. (p. 4123-4131). Retrieved from
650 <http://urlib.net/ltid.inpe.br/sbsr/2004/11.16.17.23>
- 651 Mantelli, S. L., v. Wangenheim, A., Pereira, E. B., & Comunello, E. (2010). The use
652 of euclidean geometric distance on rgb color space for classification of sky and
653 cloud patterns. *Journal of Atmospheric and Oceanic Technology*, 27(9), 1504 -
654 1517. doi: 10.1175/2010JTECHA1353.1
- 655 Mantelli Neto, S., & von Wangenheim, A. (2019). *Sky monitoring surface cam-
656 era dataset from são martinho da serra, rs, southern brazil*. [http://www.lapix
657 .ufsc.br/sky-monitoring-surface-cameras](http://www.lapix.ufsc.br/sky-monitoring-surface-cameras). LAPIX/UFSC.
- 658 Mantelli Neto, S. L. (2010). *Desenvolvimento de metodologia para a estimativa da
659 cobertura de nuvens usando uma de câmera de superfície e comparando com
660 as imagens de satélite* (Doctoral dissertation, Universidade Federal de Santa
661 Catarina). Retrieved from [http://repositorio.ufsc.br/xmlui/handle/
662 123456789/83106](http://repositorio.ufsc.br/xmlui/handle/123456789/83106)
- 663 Mantelli Neto, S. L., Pereira, E. B., Thomaz Junior, J. C., Wangenheim, A. v.,
664 Decker, L. G. L., & Coser, L. (2014). *Atmospheric pattern studies from a
665 surface imager during january 2005 at inpe southern regional center, são mar-
666 tinho da serra rs brazil* (Tech. Rep.). INPE (Instituto Nacional de Pesquisas
667 Espaciais INCOD (Instituto Nacional de Convergência Digital). Retrieved from
668 <http://urlib.net/sid.inpe.br/mtc-m21b/2014/08.19.17.27>
- 669 Marquez, R., & Coimbra, C. F. M. (2013). Intra-hour dni forecasting based on

- 670 cloud tracking image analysis. *Solar Energy*, *91*, 327-336. doi: <http://www>
 671 [.sciencedirect.com/science/article/pii/S0038092X1200343X](http://www.sciencedirect.com/science/article/pii/S0038092X1200343X)
- 672 Martins, F., Pereira, E., & Abreu, S. (2007). Satellite-derived solar resource
 673 maps for brazil under {SWERA} project. *Solar Energy*, *81*(4), 517 - 528.
 674 Retrieved from [http://www.sciencedirect.com/science/article/pii/](http://www.sciencedirect.com/science/article/pii/S0038092X0600199X)
 675 [S0038092X0600199X](http://www.sciencedirect.com/science/article/pii/S0038092X0600199X) doi: <http://dx.doi.org/10.1016/j.solener.2006.07.009>
- 676 Martins, F., Souza, M., & Pereira, E. (2003). Comparative study of satellite and
 677 ground techniques for cloud cover determination. *Adv. Space Res.*, *32*(11),
 678 2275-2280. doi: DOI:10.1016/S0273-1177(03)90554-0
- 679 Marty, C., & Philipona, R. (2000). The clear-sky index to separate clear-sky from
 680 cloudy-sky situations in climate research. *Geophysical Research Letters*, *27*,
 681 2649 - 2652. Retrieved from [ftp://ftp.pmodwrc.ch/pub/publications/](ftp://ftp.pmodwrc.ch/pub/publications/gr1\%20csi.pdf)
 682 [gr1\%20csi.pdf](ftp://ftp.pmodwrc.ch/pub/publications/gr1\%20csi.pdf)
- 683 Mejia, F. A., Kurtz, B., Murray, K., Hinkelman, L. M., Sengupta, M., Xie, Y., &
 684 Kleissl, J. (2016). Coupling sky images with radiative transfer models: a new
 685 method to estimate cloud optical depth. *Atmospheric Measurement Tech-*
 686 *niques*, *9*(8), 4151-4165. Retrieved from [https://www.atmos-meas-tech.net/](https://www.atmos-meas-tech.net/9/4151/2016/)
 687 [9/4151/2016/](https://www.atmos-meas-tech.net/9/4151/2016/) doi: 10.5194/amt-9-4151-2016
- 688 Mitsunaga, T., & Nayar, S. K. (1999, June). Radiometric self calibration. *Proc. CS*
 689 *Conf. Computer Vision and Pattern Recognition*, *1*, 374-380. Retrieved from
 690 http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=786966&tag=1
 691 doi: 10.1109/CVPR.1999.786966
- 692 Moeck, M., & Anaokar, S. (2006). Illuminance analysis from high dynamic range im-
 693 ages. *Leukos*, *2*(3), 211-228. doi: DOI:10.1582/LEUKOS.2006.02.03.005
- 694 Montgomery, D. C. (2005). *Design and analysis of experiments*. John Wiley and
 695 Sons Inc.
- 696 Nardino, M., & Georgiadis, T. (2003). Cloud type and cloud cover effects on the sur-
 697 face radiative balance at several polar stations. *Theoretical and Applied Clima-*
 698 *tology*, *74*, 203 - 215. Retrieved from [http://www.dvgu.ru/meteo/library/](http://www.dvgu.ru/meteo/library/30740203.pdf)
 699 [30740203.pdf](http://www.dvgu.ru/meteo/library/30740203.pdf) doi: 10.1007/s00704-002-0708-2
- 700 Naylor, J. (2002). *Out of the blue*. Cambridge University Press. Retrieved from
 701 <http://dx.doi.org/10.1017/CB09780511536595>
- 702 Newell, A., & Simon, H. A. (1971). Human problem solving: The state of the the-
 703 ory in 1970. *American Psychologist*, *26*(2), 145-159. Retrieved from [http://](http://psycnet.apa.org/journals/amp/26/2/145/)
 704 psycnet.apa.org/journals/amp/26/2/145/ doi: 10.1037/h0030806
- 705 Perez, R., Seals, R., & J., M. (1993). All-weather model for sky luminance dis-
 706 tribution - preliminary configuration and validation. *Solar Energy*, *50*(3), 235-
 707 245. Retrieved from [http://www.sciencedirect.com/science/article/pii/](http://www.sciencedirect.com/science/article/pii/S0038092X93900171)
 708 [0038092X93900171](http://www.sciencedirect.com/science/article/pii/S0038092X93900171)
- 709 Piccardi, M. (2004, October 07). Background subtraction techniques: a review. In
 710 *Systems, man and cybernetics, 2004 IEEE international conference on* (Vol. 4,
 711 pp. 3099-3104 vol.4). IEEE. Retrieved from [http://dx.doi.org/10.1109/](http://dx.doi.org/10.1109/icsmc.2004.1400815)
 712 [icsmc.2004.1400815](http://dx.doi.org/10.1109/icsmc.2004.1400815) doi: 10.1109/icsmc.2004.1400815
- 713 Qingyong, J., Lu, W., & Yang, J. (2011). A hybrid thresholding algorithm for cloud
 714 detection on ground-based color images. *J. Atmos. Oceanic Technol.*, *28*,
 715 1286?1296. doi: <http://dx.doi.org/10.1175/JTECH-D-11-00009.1>
- 716 Qiu, F., & Jensen, J. R. (2004). Opening the black box of neural networks for
 717 remote sensing image classification. *International Journal of Remote Sens-*
 718 *ing*, *25*(9), 1749-1768. Retrieved from [http://www.tandfonline.com/](http://www.tandfonline.com/doi/abs/10.1080/01431160310001618798?journalCode=tres20)
 719 [doi/abs/10.1080/01431160310001618798?journalCode=tres20](http://www.tandfonline.com/doi/abs/10.1080/01431160310001618798?journalCode=tres20) doi:
 720 [10.1080/01431160310001618798](http://www.tandfonline.com/doi/abs/10.1080/01431160310001618798?journalCode=tres20)
- 721 Reinhard, E., Ward, G., Pattanaik, S., & Debevec, P. (2005). *High dynamic range*
 722 *imaging: Acquisition, display, and image-based lighting (the morgan kaufmann*
 723 *series in computer graphics)* (1st ed.). San Francisco, CA, USA: Morgan
 724 Kaufmann Publishers Inc.

- 725 Richards, J. A. (1995). *Remote sensing digital image analysis* (2nd ed.). Springer-
 726 Verlag.
- 727 Sabburg, J., & Wong, J. (1999). Evaluation of ground-based sky camera sys-
 728 tem for use in surface irradiance measurements. *Journal of Atmospheric*
 729 *and Oceanic Technology*, 16, 752-759. doi: [http://dx.doi.org/10.1175/
 730 1520-0426\(1999\)016<0752:EOAGBS>2.0.CO;2](http://dx.doi.org/10.1175/1520-0426(1999)016<0752:EOAGBS>2.0.CO;2)
- 731 Schade, N. H., Macke, A., Sandmann, H., & Stick, C. (2008). Total and par-
 732 tial cloud amount detection during summer 2005 at westerland (sylvt, ger-
 733 many). *Atmospheric Chemistry and Physics*, 8, 13479-13505. Retrieved from
 734 <http://www.atmos-chem-phys.org/9/1143/2009/acp-9-1143-2009.pdf>
 735 doi: 10.5194/acpd-8-13479-2008
- 736 Setiono, R., Leow, W. K., & Thong, J. Y. L. (2000). Opening the neural network
 737 black box: an algorithm for extracting rules from function approximating
 738 artificial neural networks. In *Proceedings of the twenty first international*
 739 *conference on information systems* (pp. 176-186). Atlanta, GA, USA: Asso-
 740 ciation for Information Systems. Retrieved from [http://portal.acm.org/
 741 citation.cfm?id=359640.359738](http://portal.acm.org/citation.cfm?id=359640.359738)
- 742 Sobieranski, A. C., Abdala, D. D., Comunello, E., & von Wangenheim, A.
 743 (2009). Learning a color distance metric for region-based image segmen-
 744 tation. *Pattern Recognition Letters*, 30(16), 1496 - 1506. Retrieved from
 745 <http://www.sciencedirect.com/science/article/pii/S0167865509002098>
 746 doi: <https://doi.org/10.1016/j.patrec.2009.08.002>
- 747 Sobieranski, A. C., Comunello, E., & von Wangenheim, A. (2011). Learning a
 748 nonlinear distance metric for supervised region-merging image segmentation.
 749 *Computer Vision and Image Understanding*, 115(2), 127 - 139. Retrieved from
 750 <http://www.sciencedirect.com/science/article/pii/S1077314210002006>
 751 doi: <https://doi.org/10.1016/j.cviu.2010.09.006>
- 752 Souza-Echer, M. P., Pereira, E. B., Bins, L., & Andrade, M. A. R. (2006). A simple
 753 method for the assessment of the cloud cover state in high-latitude regions by
 754 a ground-based digital camera. *Journal of Atmospheric and Oceanic Technol-*
 755 *ogy*, 23(3), 437-447. doi: <http://dx.doi.org/10.1175/JTECH1833.1>
- 756 Tapakis, R., & Charalambides, A. (2013). Equipment and methodologies for
 757 cloud detection and classification: A review. *Solar Energy*, 95, 392 - 430.
 758 Retrieved from [http://www.sciencedirect.com/science/article/pii/
 759 S0038092X12004069](http://www.sciencedirect.com/science/article/pii/S0038092X12004069) doi: <http://dx.doi.org/10.1016/j.solener.2012.11.015>
- 760 Tsin, Y., Ramesh, V., & Kanade, T. (2001). Statistical calibration of the ccd
 761 imaging process. In *Proc. int.conf. computer vision* (Vol. 1, p. 480-487). Re-
 762 trieved from <http://ieeexplore.ieee.org/document/937555/?part=1> doi:
 763 10.1109/ICCV.2001.937555
- 764 WMO. (2008). Guide to meteorological instruments and methods of observations.
 765 (7th ed.) [Computer software manual]. World Meteorological Organization,
 766 7bis, avenue de la Paix, Case postale 2300, CH-1211 Geneva 2, Switzer-
 767 land. Retrieved from [http://www.wmo.int/pages/prog/gcos/documents/
 768 gruanmanuals/CIMO/CIMO_Guide-7th_Edition-2008.pdf](http://www.wmo.int/pages/prog/gcos/documents/gruanmanuals/CIMO/CIMO_Guide-7th_Edition-2008.pdf) (WMO-No. 8 I.15-1
 769 - I.15-11)
- 770 Yamanouchi, T., & Charlock, T. P. (1993). Radiative effects of clouds icesheet
 771 and sea ice in the antarctic. *Proceedings of Yokohama Symposia J2 and J5,*
 772 *223*. Retrieved from [http://www.cig.ensmp.fr/~iahs/redbooks/a223/
 773 iahs_223_0029.pdf](http://www.cig.ensmp.fr/~iahs/redbooks/a223/iahs_223_0029.pdf) doi: 10.1029/96JD02866
- 774 Yang, J., Min, Q., Lu, W., Ma, Y., Yao, W., Lu, T., ... Liu, G. (2016). A to-
 775 tal sky cloud detection method using real clear sky background. *Atmo-*
 776 *spheric Measurement Techniques*, 9, 587 - 597. Retrieved from [http://
 777 www.atmos-meas-tech.net/9/587/2016/amt-9-587-2016.pdf](http://www.atmos-meas-tech.net/9/587/2016/amt-9-587-2016.pdf) doi:
 778 10.5194/amt-9-587-2016
- 779 Yang, J., Min, Q., Lu, W., Yao, W., Ma, Y., Du, J., ... Liu, G. (2015). An au-

780 tomated cloud detection method based on the green channel of total-sky
781 visible images. *Atmospheric Measurement Techniques*, 8(11), 4671–4679.
782 Retrieved from <http://www.atmos-meas-tech.net/8/4671/2015/> doi:
783 10.5194/amt-8-4671-2015
784 Zhang, P. G. (2007, January). Avoiding pitfalls in neural network research.
785 *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS*
786 *PART C APPLICATIONS AND REVIEWS.*, 37, 3-16. doi: 10.1109/
787 TSMCC.2006.876059