7.7 Cloud Detection with the Use of Ground-Based Full-Sky Imaging Polarimetry

In many meteorological stations the accurate determination of sky conditions, especially the detection of clouds, is a desirable yet rarely attainable goal. Traditionally, sky conditions are reported by human observers with considerable discrepancies between individual and subjective reports. In practice, employing human observers is not always feasible due to budgetary constraints. Human observers can be replaced by automatic full-sky imager systems, like the Scripps-produced Whole Sky Imager, or the TSI-880 Total Sky Imager produced by the Yankee Environmental Systems, Inc. (YES 2001). These systems provide real-time processing and display of daytime sky conditions using common image processing algorithms, which detect the clouds radiometrically by filtering the colour picture of the sky so that the approximate value of the cloud cover fraction can be calculated.

Using the additional information obtained by evaluating both the degree and angle of polarization patterns of cloudy skies measured by full-sky imaging polarimetry in the red (650 nm), green (550 nm) and blue (450 nm) spectral ranges, the algorithms of radiometric cloud detection (e.g. Saunders 1986; Saunders and Kriebel 1988; Derrien et al. 1993; YES 2001) can be significantly improved. In this chapter we show an efficient combined radiometric and polarimetric algorithm developed by Horváth et al. (2002a), which performs the detection of clouds more efficiently and reliably than an exclusively radiometric cloud detection algorithm. In the future, this or similar improved polarimetric algorithms can accomplish cloud detection with ground-based automatic instruments, which could be a new generation of the presently existing ground-based automated total sky imagers using exclusively radiometric algorithms for cloud detection.

7.7.1 Algorithmic Cloud Detection

Using full-sky imaging polarimetry, one obtains the values of nine optical variables for every pixel of the sky image: I_r, I_g, I_b, p_r, p_g, p_b, α_r, α_g, α_b, that is, radiance I, degree of linear polarization p and angle of polarization α measured in the red (r), green (g) and blue (b) spectral ranges. The essence of the cloud detection algorithm of Horváth et al. (2002a) is that for every pixel of the sky picture seven decisions are made:

- (1) Analysing the values of I_r, I_g and I_b, the colour of the pixel is determined, and it is decided if the pixel may belong to a colourless cloud or to a blue sky region.
- (2)-(7) Using the values of p_r, p_g, p_b, α_r, α_g or α_b, it is again decided if the pixel may be part of a cloud or a clear sky region.

Every decision is the outcome of its specific subalgorithm, called "detector". Detector (1) is symbolized by *IRGB* since it uses the *I*-values measured in the red (*R*), green (*G*) and blue (*B*) spectral ranges. Detectors (2)-(7) are symbolized by *PR, PG, PB,* αR, αG and αB, because they use the measured values of p_r, p_g, p_b, α_r, α_g or α_b, respectively. If detector *IRGB* identifies a pixel as "cloud", the pixel qualification is weighted by 3, because the decision relies on the use of three input data. The total weight of a pixel qualification is *i*, if it is identified as "cloud" by *i* detectors among detectors *PR, PG, PB,* αR, αG, αB (these six detectors make their decisions using one input datum each). The partial weight is 0 in every case when the pixel is identified as "clear sky" by a given detector.

If the investigated pixel is under- or overexposed at least in one of the *R, G, B* spectral ranges, detector *IRGB* is inactive resulting in a 0 partial weight value. Similarly, any other detector is inactive, if the pixel is under- or overexposed in the corresponding spectral range. Let the number of active detectors be *m*. The partial weights are summed up, thus finally the investigated pixel has a total weight *n* ranging from 0 to 9. *n* tells how many times the pixel was identified as "cloud"; *n* is called the "number of cloud identification". At a given *m*-value, *n* is proportional to the likelihood of cloud: the higher is *n*, the greater the probability that the pixel belongs to a cloud in the picture. The authenticity (or reliability) of *n* is proportional to the number *m* of active detectors. The distributions of the *n*- and *m*-values in the sky can be represented by colour-coded maps (Fig. 7.7.5).

In the case of "radiometric cloud detection" only detector *IRGB* is used. "polarimetric cloud detection" uses only detectors *PR, PG, PB,* αR, αG and αB. We speak about "combined (radiometric and polarimetric) cloud detection" if all seven detectors are used. As we see below, the combined cloud detection algorithm has seven control parameters: c, $p_0(q)$, $\Delta\alpha(q)$, where $q = R, G, B$. Setting appropriately their values, certain types of clouds can be reliably detected.

The optimal values of these control parameters can be empirically determined in the following way: In the digitized colour picture of a given cloudy sky the clouds are visually identified by inspection with the nacked eye and each pixel is marked accordingly. The resulting cloud pattern serves as a "control pattern". Changing the value of the control parameter of a given detector, the visually detected control clouds are compared with the clouds recognized by the detector: The numbers of pixels are counted, where (i) there is cloud in the control pattern but the pixel is identified erroneously as clear sky; (ii) there is clear sky in the control pattern but the pixel is identified erroneously as cloud; (iii) the pixel is identified as cloud; (iv) the pixel is under- or overexposed. Dividing these numbers by the number of pixels of the entire sky picture, we obtain

- the proportion *PCDS* of clouds detected (erroneously) as (clear) sky,
- the proportion *PSDC* of (clear) sky detected (erroneously) as cloud,
- the proportion *PCC* of cloud cover,
- the proportion *PUO* of under- and/or overexposed pixels,
- the proportion $PED = PCDS + PSDC$ of erroneous detection.

That value of the control parameter of a given detector is considered as optimal $[c^*, p_0^*(q), \Delta\alpha^*(q)]$, ($q = R, G, B$) at which the *PED*-value is minimal (*PED**), that is, where the correlation between the pixels of the algorithmically and visually detected clouds and clear sky regions is maximal (Fig. 7.7.4).

7.7.2 Radiometric Detection of Colourless Clouds

Detector *IRGB* functions in the following way: Apart from the reddish orange clouds illuminated by the sunset or sunrise glow, or from the very high altitude bluish cirrus clouds, the clouds are generally colourless, ranging from dark grey to bright white, independently of their radiance and position in the sky (Können 1985; Coulson 1988). The pixels of such "grey" clouds on the sky picture possess approximately the same radiances in all three (*r, g, b*) spectral ranges. Thus, if the differences $\Delta I_{b-r} = |I_b - I_r|$ and $\Delta I_{b-g} = |I_b - I_g|$ are less than $\varepsilon = c \cdot I_b$, where c is a control parameter to be appropriately chosen as described above, then detector *IRGB* assumes that the given pixel belongs to a colourless cloud, else to the blue sky. ε is proportional to I_b due to the blueness of scattered skylight.

7.7.3 Radiometric Detection of Overexposed and Underexposed Parts of the Sky Image

Around the sun image disc the photoemulsion inevitably becomes overexposed. If in a given spectral range (*r, g, b*) the digitized brightness value I of a pixel reaches 254 (maximum is 255), the pixel is considered as overexposed in that part of the spectrum. At about 90^o from the sun, where the clear blue sky is the darkest, the photoemulsion may be underexposed, especially in the red, where the skylight intensity is the lowest. Landmarks and vegetation (usually near the horizon) may also show up in a sky picture, and they too, are generally underexposed on the photoemulsion, as are the sun occulter. These underexposed areas can be detected by the following algorithm: If the values I_r, I_g, I_b of a pixel are smaller than a given threshold t, it is assumed that the given pixel is underexposed. Setting appropriately the value of t, the underexposed regions of the sky as well as the underexposed landmarks and vegetation can be reliably identified.

7.7.4 Polarimetric Detection of Clouds on the Basis of the Degree and Angle of Linear Polarization

Detectors *PR, PG* and *PB* function as follows: Apart from 145° from the sun in the sky with water clouds (where rainbow scattering dominates with high p) independently of the wavelength, in a given celestial point p of cloud pixels is lower than that of clear sky pixels due to multiple scattering (Können 1985; Coulson 1988). As a first approximation the change of p of skylight versus the angular distance γ from the sun can be described by the single-scattering Rayleigh model, in which $p(\gamma) = p_{max} \cdot \sin^2\gamma/(1+\cos^2\gamma)$, $\cos\gamma = \sin\theta_s \sin\theta \cos\varphi + \cos\theta_s \cos\theta$, where θ_s is the solar zenith angle, θ and φ are the angular distances of the observed celestial point from the zenith and the solar meridian, respectively. Detector *Pq (q=R,G,B)* assumes that the given pixel positioned at angular distance γ from the sun belongs to a cloud if p is lower than the threshold $p_{threshold} = p_0(q) \cdot \sin^2\gamma/(1+\cos^2\gamma)$, where $p_0(q)$ is a control parameter to be appropriately chosen as described above.

Under certain meteorological conditions (if parts of the clouds and the air-space between the clouds and the surface of the earth are not directly lit by the sun) in a given celestial point the angle of polarization α of cloud pixels differs considerably from $\alpha_{clear\ sky}$ of clear sky pixels independently of the wavelength (Können 1985; Pomozi et al. 2001b). Detector αq *(q=R,G,B)* assumes that the given pixel belongs to a cloud if the difference $|\alpha - \alpha_{clear\ sky}|$ is larger than the threshold $\Delta\alpha(q)$, which is a control parameter to be appropriately chosen as described above. For these decisions the α-patterns of the clear sky (Fig. 7.7.2) measured in the *R, G, B* spectral ranges can be used as control with the same solar zenith angle as that of the investigated cloudy sky (Fig. 7.7.1). Prior to these decisions the α-patterns of both the corresponding clear and cloudy skies are smoothed by convolution with an appropriate two-dimensional rotation-symmetric Gaussian function to eliminate the inevitable small noise of higher spatial frequencies.

7.7.5 Detection of Clouds by Radiometric, Polarimetric and Combined Algorithms

Figure 7.7.1 shows the p-and α-patterns of a sky with fast moving cumuli measured by the three-lens three-camera full-sky imaging polarimeter of Horváth et al. (2002a) in the red (650 nm), green (550 nm) and blue (450 nm) spectral ranges. In Fig. 7.7.2 the corresponding patterns of a clear, cloudless sky are seen with the same solar position as in Fig. 7.7.1.

Figure 7.7.3A shows the photograph of the partially cloudy sky, the polarizational characteristics of which are represented in Fig. 7.7.1. To test the cloud detection performance of any algorithm control patterns would be needed with known, well-defined clouds. Unfortunately, such control cloud patterns are generally not available. Thus, the first step of cloud detection is to construct a

relatively good approximation of such control patterns: Placing trust in the excellent pattern recognition and good brightness and colour discrimination ability of the human visual system, one can detect and recognize visually the clouds in the colour picture of the sky (Fig. 7.7.3A), which is displayed on the screen of a computer, and those celestial regions can be shaded by a mouse-guided "paint brush" where clouds are seen. Figure 7.7.3B, serving as a control cloud pattern, shows the visually detected clouds in white and the clear sky regions in black. The percentage of cloud cover in this control pattern is $PCC = 56.1\%$.

Figure 7.7.3C represents the clouds detected radiometrically with the use of the algorithm (detector *IRGB*) described above. In Fig. 7.7.3C regions of the sky are checkered, if their pixels are under- or overexposed at least in one of the three (red, green, blue) spectral ranges in which measurements were performed. In these checkered regions detector *IRGB* is inactive. Figures 7.7.3D,E,F and 7.7.3G,H,I show the clouds detected polarimetrically at 650, 550 and 450 nm using the p-patterns in Fig. 7.7.1 and the α-patterns in Figs. 7.7.1 and 7.7.2, respectively.

The cloud detection performance of every detector is determined by a control parameter, the value of which is optimal if the proportion $PED = PCDS + PSDC$ of erroneous detection is minimal. Figure 7.7.4 shows the change of *PED* as a function of the corresponding control parameters c, or $p_0(q)$, or $\Delta\alpha(q)$ for detectors *IRGB, PR, PG, PB,* αR, αG and αB. We can see that the graphs $PED(c)$, $PED[p_0(q)]$, $PED[\Delta\alpha(q)]$ generally possess a definite minimum. The positions of these minima are chosen as the optimal values c^*, $p_0^*(q)$, $\Delta\alpha^*(q)$ of the control parameters.

A detector is inactive at those pixels of the picture of the sky where under- or overexposure occurs. This is the case in the checkered regions in Figs. 7.7.3C-I, where there is no information about the real sky conditions. The radiometric detector *IRGB* actually involves three detectors (*IR, IG* and *IB*), which can function only together. Since detector *IRGB* is inactive if under- or overexposure occurs in at least one of the three (red, green, blue) spectral ranges, the number m of active (neither underexposed nor overexposed) detectors can be 0, 2, 4 or 9, when the investigated celestial point is under- or overexposed in 3, 2, 1 or 0 spectral ranges, respectively. Thus, the number n of "cloud identification" can be 0,1,...,8,9. Figures 7.7.5A and 7.7.5B show the colour-coded celestial maps of n and m calculated for the partially cloudy sky in Fig. 7.7.3A, the optical characteristics of which are shown in Fig. 7.7.1. m is proportional to the authenticity (or reliability) of the (cloud or clear sky) detection. Figure 7.7.5C shows the map which combines maps A and B. At a given m-value, $n/n_{max}(m)$ is the likelihood of cloud, while $1-n/n_{max}(m)$ is the likelihood of clear sky.

Figure 7.7.5D shows the cloudy and clear sky regions detected by the combined (radiometric and polarimetric) algorithm such that the pixels with larger or smaller $n(m)$-values than $n(m)^*$ were considered to belong to clouds or clear sky regions, respectively. For $n(m)^*$ the proportion of erroneous detection $PED\ (= PCDS + PSDC)$ is minimal (PED^*). A similar procedure is applied in the case of the polarimetric cloud detection, when only the p- and α-patterns are used. Figure 7.7.5C shows the threshold values $n(m)^*$.

In Table 7.7.1 the lower and upper limits of the proportion of cloud cover determined by the radiometric, polarimetric and combined (radiometric and polarimetric) cloud detection algorithms are compared. PCC_{det} is the value of PCC determined by the radiometric (Fig. 7.7.3C), polarimetric (Figs. 7.7.3C-I) and combined (Fig. 7.7.5D) algorithms. The lower and upper limit of PCC is $PCC_{min} = PCC_{det} - PSDC$ and $PCC_{max} = PCC_{det} + PCDS + PUO$, respectively. The real value PCC_r of the proportion of cloud cover is somewhere between PCC_{min} and PCC_{max}. As an approximate value of PCC_r we obtained 56.1% by visual cloud detection (Fig. 7.7.3B). The reliability of a cloud detection algorithm is characterized by the difference $\Delta PCC = PCC_{max} - PCC_{min}$: the smaller is ΔPCC, the higher is the reliability. We can see in Table 7.7.1 that ΔPCC is largest (33.1%) for the radiometric, smaller (20.8%) for the polarimetric and smallest (14.7%) for the combined cloud detection. In the case of the combined cloud detection the interval, in which PCC_r can be, is about the half of that obtained for the radiometric cloud detection. This demonstrates well that the combined algorithm can detect clouds more reliably than the exclusively radiometric or the purely polarimetric algorithm alone. The power of the combined cloud detection algorithm in comparison with the radiometric one is ensured by the fact that the information contributed by the radiometric subalgorithm (subdetector *IRGB*) to the final decision (cloud or clear sky) is only one third by weight. The polarimetric algorithm is more reliable than the radiometric one, because the former is based on the use of twice as many detectors as the latter.

Table

Table 7.7.1. The results of cloud detection with the use of Figs. 7.7.1-7.7.5. *IRGB*: radiometric detector. *POL*: polarimetric detector involving detectors *PR, PG, PB, αR, αG* and *αB*. *COM*: combined detector combining the *IRGB* and *POL* detectors. *PUO* = *P*roportion of *U*nder- and/or *O*verexposure, *PED* = *P*roportion of *E*rroneous *D*etection, *PCDS* = *P*roportion of *C*louds *D*etected (erroneously) as (clear) *S*ky, *PSDC* = *P*roportion of (clear) *S*ky *D*etected (erroneously) as *C*loud, *PED* = *PCDS* + *PSDC*, *PCC* = *P*roportion of *C*loud *C*over, PCC_r = "real" value of *PCC* detected visually, PCC_{det} = value of *PCC* detected by a given algorithm, PCC_{min} = lower limit of *PCC*, PCC_{max} = upper limit of *PCC*. In the first row of the second (widest) column of the table the inequality $PCC_{min} \le PCC_r \le PCC_{max}$ is seen, where $PCC_{min} = PCC_{det} - PSDC$ and $PCC_{max} = PCC_{det} + PCDS + PUO$. The percentage values of these terms are given in the table for the different types of detector. Difference $\Delta PCC = PCC_{max} - PCC_{min}$ gives the uncertainty or error of PCC_r: the value of PCC_r is somewhere between PCC_{min} and PCC_{max}. The number of pixels of the entire sky is N = 346207, to which all percentage values are related. (After Table 3 of Horváth et al. 2002a, p. 555).

detection type	$PCC_{det} - PSDC =$		$PCC_{min} \le$	$PCC_r \le$	$PCC_{max} =$	$PCC_{det} + PCDS + PUO$			$\Delta PCC = PCC_{max} - PCC_{min}$
	(%)	*(%)*	*(%)*	*(%)*	*(%)*	*(%)*	*(%)*	*(%)*	*(%)*
PR	45,9	9,7	36,2	56,1	80,8	45,9	18,7	16,2	44,6
PG	38,2	6,4	31,8	56,1	62,9	38,2	22,3	2,4	31,1
PB	49,9	8,5	41,4	56,1	68,3	49,9	11,2	7,2	26,9
αR	59,7	13,8	45,9	56,1	84,8	59,7	8,9	16,2	38,9
αG	55,1	16,1	39,0	56,1	72,6	55,1	15,1	2,4	33,6
αB	51,6	16,0	35,6	56,1	75,8	51,6	17,0	7,2	40,2
IRGB	50,7	4,5	46,2	56,1	79,3	50,7	6,0	22,6	33,1
POL	59,4	12,1	47,3	56,1	68,1	59,4	7,9	0,8	20,8
COM	53,5	6,1	47,4	56,1	62,1	53,5	7,8	0,8	14,7

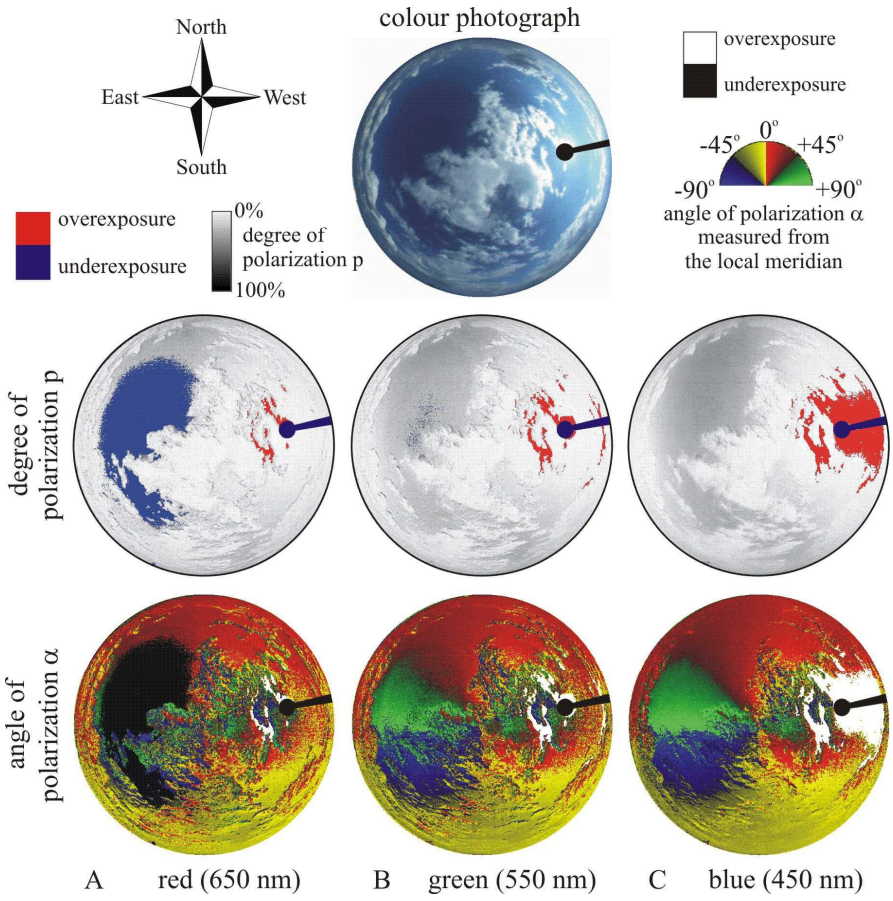

Fig. 7.7.1. Photograph and the patterns of the degree of linear polarization p and angle of polarization α of a sky with fast moving cumuli measured by the three-lens three-camera full-sky imaging polarimeter of Horváth et al. (2002a) in the red (650 nm), green (550 nm) and blue (450 nm) spectral ranges at Kunfehértó (46°23'N, 19°24'E, Hungary) on 15 August 2000 at 17:00 (local summer time = UTC+2). (After Fig. 7 of Horváth et al. 2002a, pp. 547-548).

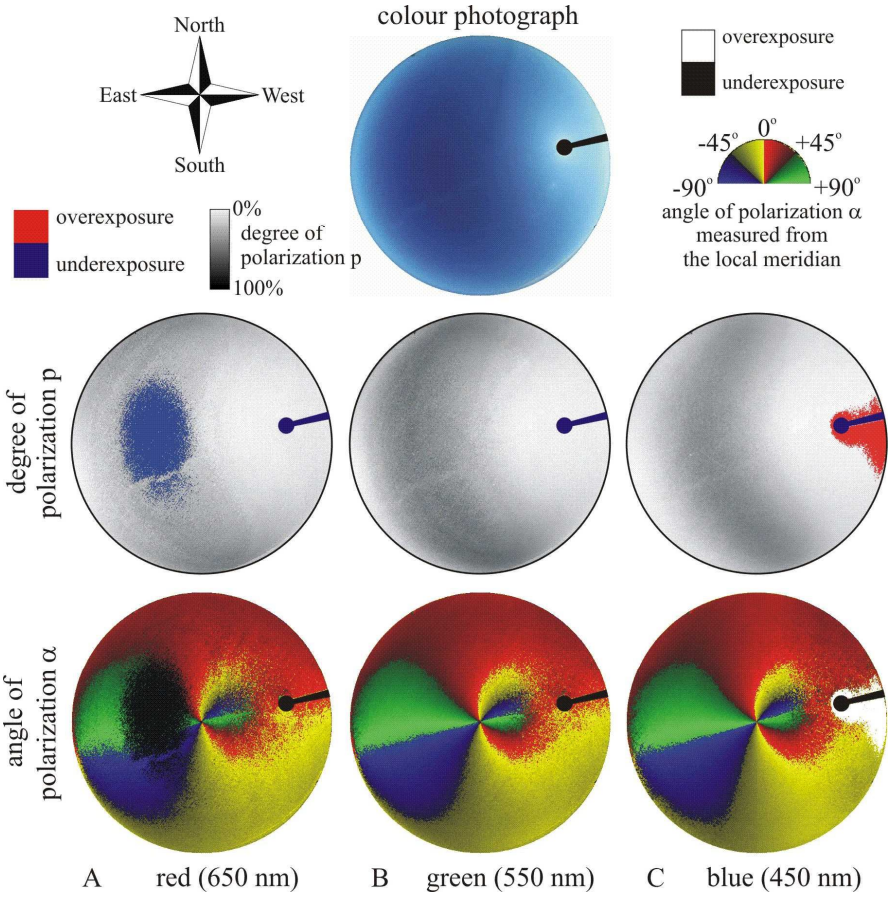

Fig. 7.7.2. As Fig. 7.7.1 for a clear, cloudless sky (with the same solar position) measured on 17 August 2000 at 17:00. (After Fig. 8 of Horváth et al. 2002a, pp. 548-549).

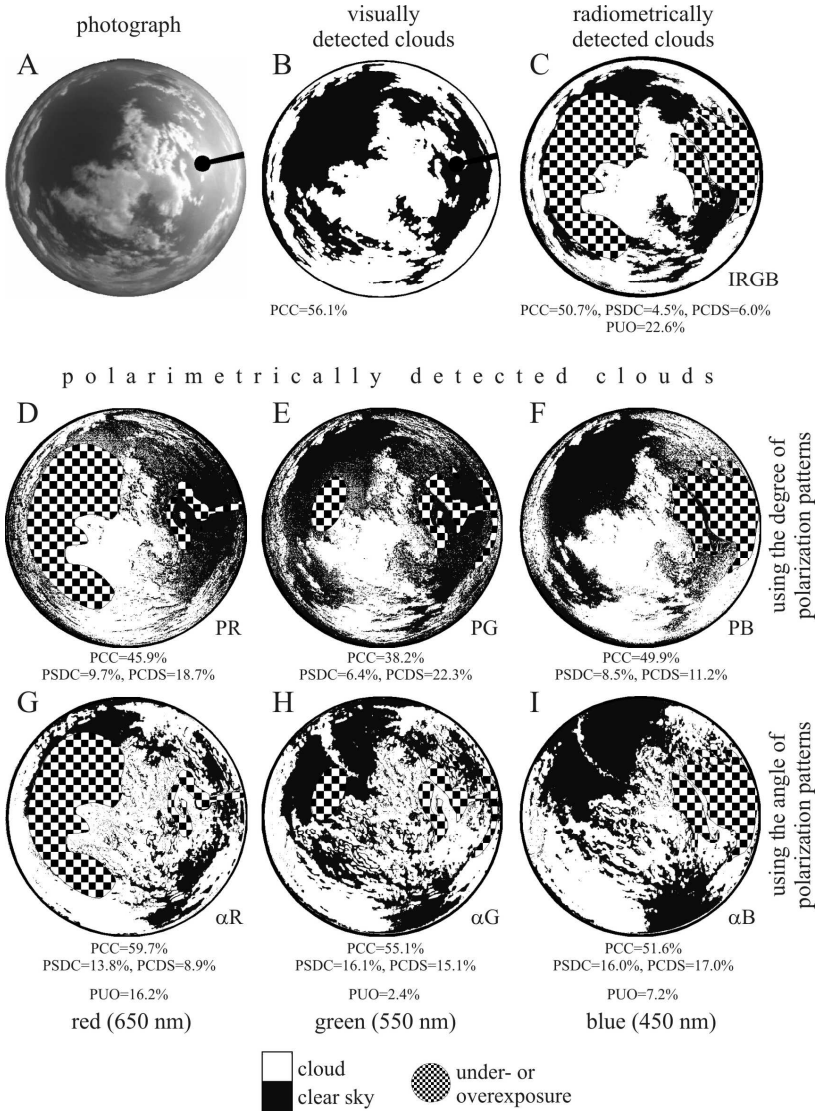

Fig. 7.7.3. A: Photograph (identical with that in the first row of Fig. 7.7.1) of the partially cloudy sky, the polarizational characteristics of which are shown in Fig. 7.7.1. B: Cloudy (white) and clear (black) sky regions detected visually by the nacked eye in picture A. C: Clouds detected radiometrically, where the under- or overexposed celestial areas are checkered. D-I: Clouds detected polarimetrically at 650, 550 and 450 nm using the *p*- or α-patterns in Figs. 7.7.1 and 7.7.2. *PCC* = *P*roportion of *C*loud *C*over determined by the different detectors *IRGB, PR, PG, PB, αR, αG* and *αB*. *PSDC* = *P*roportion of (clear) *S*ky *D*etected (erroneously) as *C*loud, *PCDS* = *P*roportion of *C*louds *D*etected (erroneously) as (clear) *S*ky, *PUO* = *P*roportion of *U*nder- and/or *O*verexposure. (After Figs. 9 and 10 of Horváth et al. 2002a, pp. 550-551).

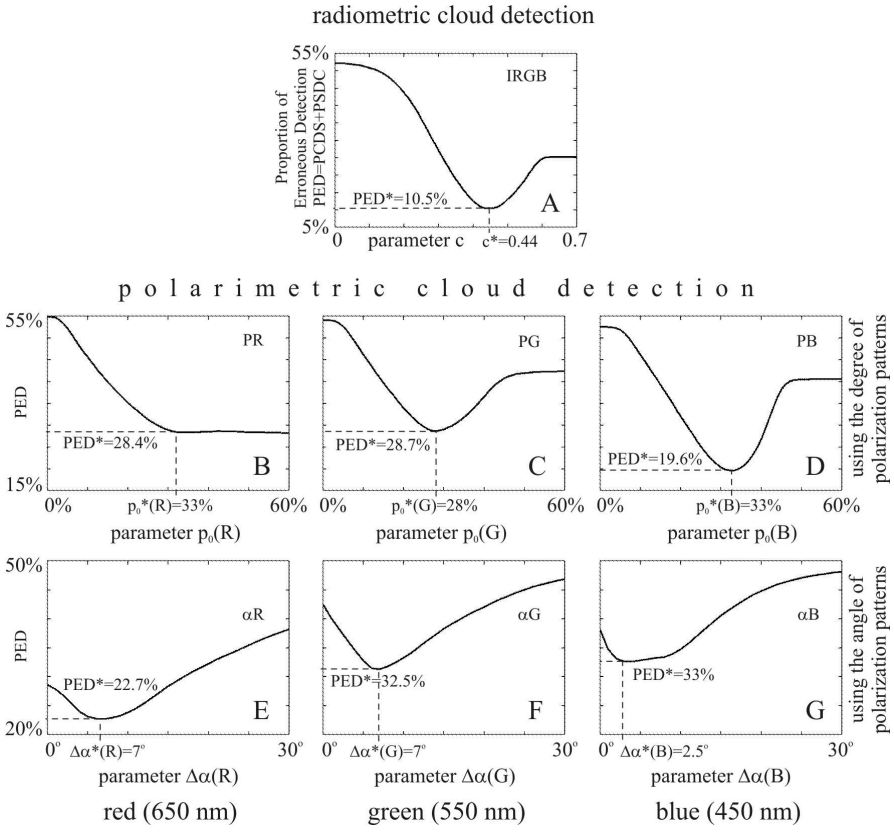

Fig. 7.7.4. Proportion *PED = PCDS + PSDC* of pixels of the sky detected erroneously by the different detectors *IRGB, PR, PG, PB, αR, αG* and *αB* as a function of parameters *z* [*z = c*, or $p_0(q=R,G,B)$, or $\Delta\alpha(q=R,G,B)$] in the red (650 nm), green (550 nm) and blue (450 nm) spectral ranges. The positions *z** of the minima *PED** of the graphs are marked with dashed straight lines. A: Radiometric cloud detection with control parameter *c*. B, C, D: Polarimetric cloud detection with control parameter $p_0(q=R,G,B)$ using the *p*-patterns in Fig. 7.7.1. E, F, G: Polarimetric cloud detection with control parameter $\Delta\alpha(q=R,G,B)$ using the α-patterns in Figs. 7.7.1 and 7.7.2.

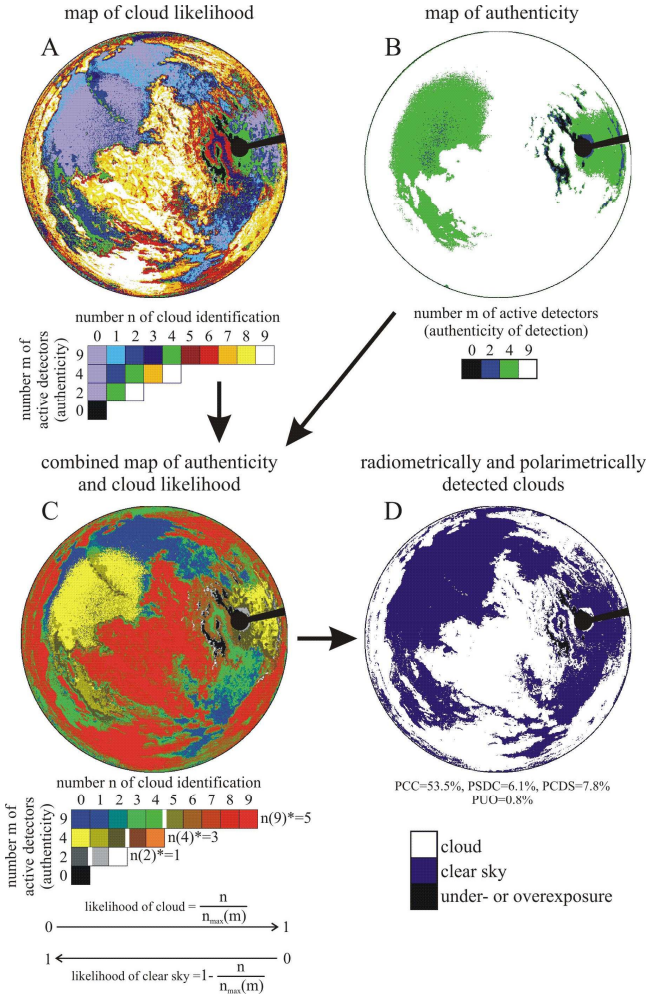

Fig. 7.7.5. A: Colour-coded map of the number *n* of cloud identification calculated for the partially cloudy sky in Fig. 7.7.3A, the optical characteristics of which are shown in Fig. 7.7.1. B: Colour-coded map of the number *m* of active (neither underexposed nor overexposed) detectors calculated for the partially cloudy sky in Fig. 7.7.3A. *m* is proportional to the authenticity of the (cloud or clear sky) detection. C: Map combining maps A and B. At a given *m*-value, $n/n_{max}(m)$ is the likelihood of cloud, while $1-n/n_{max}(m)$ is the likelihood of clear sky. D: Cloudy (white) and clear (blue) sky regions are detected by the combined (radiometric and polarimetric) algorithm such that pixels with larger or smaller $n(m)$ than $n(m)^*$ were considered to belong to clouds or clear sky regions, respectively. For $n(2)^* = 1$, $n(4)^* = 3$ and $n(9)^* = 5$ (the positions of which are indicated by white vertical bars in the colour palette) the proportion of erroneous detection $PED = PCDS + PSDC$ is minimal (PED^* in Fig. 7.7.4). The under- or overexposed sky regions ($m = 0$) in the maps are shaded with black. (After Figs. 11 and 12 of Horváth et al. 2002a, pp. 552-553).