Managing the transition from human cloud observations to automatic cloud measurements at an AWS

R. Boers, W. Wauben en T. Brandsma

R&D Observations and Data Technology, KNMI, PO Box 201

De Bilt, Netherlands

1. Introduction

In the last 20 years manned weather stations have been replaced by automatic weather stations (AWS) and this process is expected to continue in the next decades. Ceilometers are most often employed as a substitute for the human observer in the measurement of fractional cloudiness although their merit is debatable (Boers et al., 2010). A ceilometer probes only a small portion of the sky, whereas a human observer scans the whole sky. As a result, time series of cloudiness have become inhomogeneous.

Here, we describe a method to homogenize the time series of cloudiness. The paper is organized as follows. First we describe the origin of the problem in more detail including the manner in which the distribution functions of fractional cloudiness are affected by the switch to automatic cloud measurements. Next, a method is introduced to correct for this break. The correction procedure, a so-called quantile-quantile correction, is applied to the time series of fractional cloudiness after the break. The method is evaluated using a two-year period during which human observer and ceilometer cloud measurements were taken side-by-side. Lastly, the method is applied to homogenize the time series of fractional cloudiness over the Netherlands.

A short discussion and conclusion section will treat the implication and usability of the correction procedure.

2. The origin of the problem

In 2002 all human observers at the AWS - sites of the Netherlands were replaced by ceilometers. A ceilometer measurement of cloudiness constitutes an entirely different method of observation than the previous practice of visually scanning the sky for clouds as done by a human observer (Wauben et al., 2006). Hence discontinuities in cloud fraction are to be expected. Even so, a user of climate data who procures his/her data from traditional climate data repositories and who is ignorant of the change in observation method might be persuaded by Figure 1 that cloud fraction as observed over the Netherlands has no discontinuity at the year 2002.

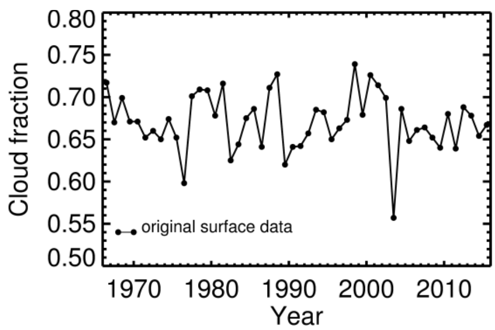

Figure 1. Annual averaged cloud fraction for all AWS in the Netherlands as a function of time, not accounting for the change of observation method in the year 2002.

Any discontinuities, should they exists, are at least partly concealed by the large negative anomaly in fractional cloudiness as a result of the hot and relatively cloudless year 2003. Yet, when cloud fraction distribution functions are plotted for the period before and after 2002 (see Figure 2) it becomes apparent that a break occurred. A closer investigation into the underlying cloudiness data is imperative, as there are substantial differences in distribution function between the period before and after the break.

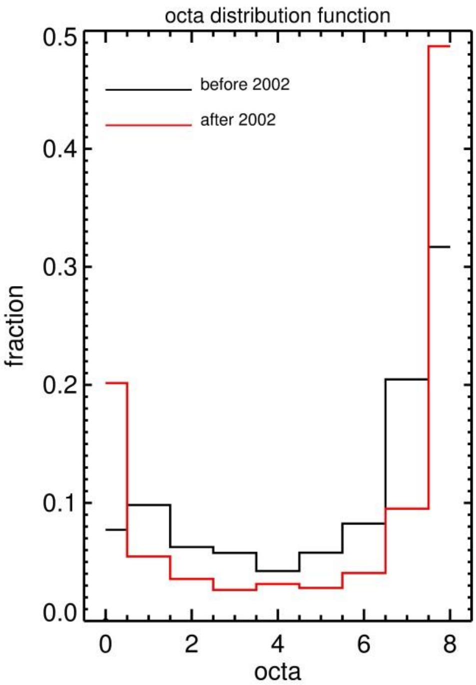

Figure 2. Cloud fraction okta distribution function from before (black) and after (red) 2002 over the Netherlands.

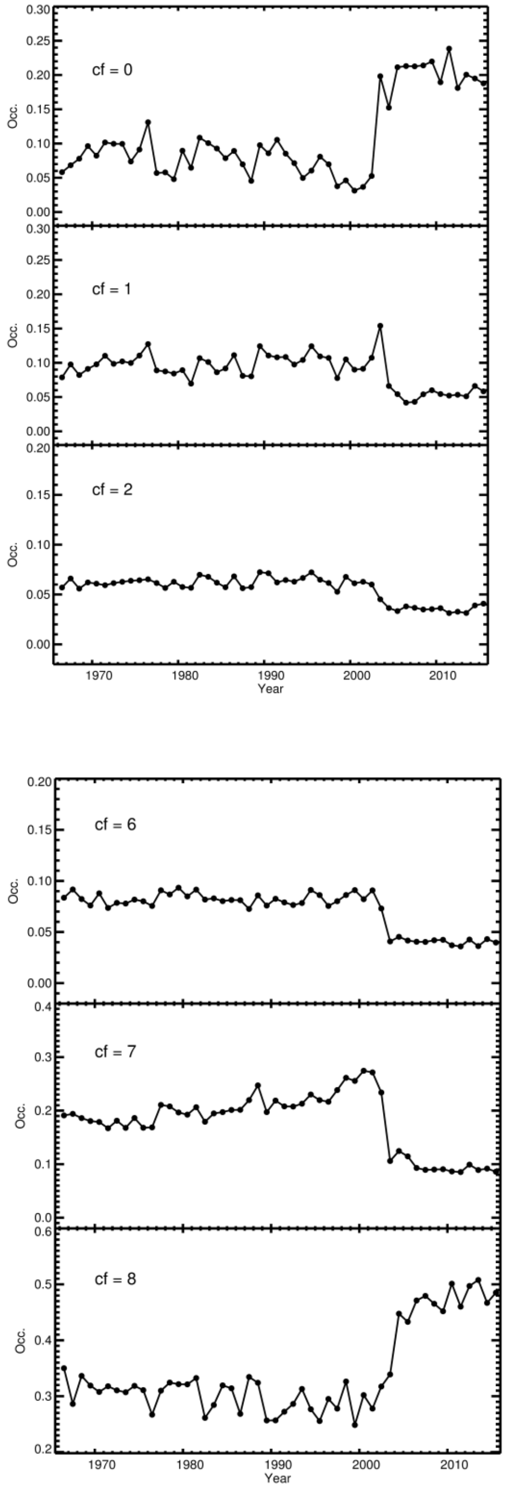

Figure 3. Relative occurrence of hourly okta values as a function of time.

As documented before (Boers et al., 2010) a ceilometer measurement using a time period of 30 minutes will more frequently assign 0 or 8 okta as a ceilometer covers only a track close to the zenith whereas an observer also reports the clouds or gaps in the cloud deck that may or may not exist elsewhere in the sky. Consequently, the 0 and 8 okta intervals are over-represented when compared to data obtained from the human observer and the middle okta regions, and particularly 1 and 7 okta, are relatively under-represented. Analysis of the relative occurrence of individual okta values clearly demonstrates that such breaks occur. Figure 3 shows a time series for okta 0,1,2 and 6,7,8. In this figure, and the remainder of the paper, hourly okta values of the ceilometer are derived from the last 30 minutes of the hour (the last 10 minutes has double weight (Wauben, 2002). The hourly okta values of the observer are instantaneous observations about 10 minutes before the hour.

Discontinuities in relative occurrence are clearly present with the largest discontinuities arising at the 0,1 and 7,8 okta values. The discontinuity in occurrences for the intermediate okta values of 3,4,5 are smaller than any of the discontinuities shown in Figure 3, but they are still clearly visible.

The conclusion is therefore justified that despite the absence of a clear discontinuity in the averaged cloud fraction in Figure 1, discontinuities are present in the relative occurrence of the okta values and must be taken into account for specific applications.

3. Quantile-quantile correction of the okta occurrence values

A quantile-quantile (q-q) correction is commonly applied to output from meteorological models to adjust the calculated distribution functions to distribution functions that are actually observed (Li et al., 2010). In a q-q correction cumulative distribution functions (CDF's) of a model / input variable and of an observation / output variable are obtained from the data at hand. Next, individual model output data points at the model CDF-value are corrected to a new value by shifting them over to a point where the observed CDF-value is identical to the model CDF-value. Here we applied this technique by using ceilometer and human observed fractional cloudiness instead of 'model / input' and 'observation / output' variables, respectively. As cloudiness data are indicated by nine discrete okta values all hourly data were first converted to fractional cloudiness using the conversion factors of Boers et al., 2010). To arrive at a smooth cloudiness distribution function, the time series of fractional cloudiness was then smoothed by a 2-point Gaussian smoother after which the q-q correction was applied. Finally, the q-q corrected fractional cloudiness data were converted back to discrete okta values.

It should be pointed out from the outset that a q-q correction can never be used to perfectly reconstruct the human observer observation time series from a ceilometer time series. Although the q-q correction changes individual ceilometer values of fractional cloudiness it will generally not make the optimal correction for each specific situation. There is simply not enough information present in a ceilometer time series for that purpose. All that the q-q correction does is to adjust individual okta values towards values that are consistent with the desired distribution functions. It can nevertheless be expected that a q-q correction should 'improve' the data so that statistical quantities such as ceilometer fractional cloudiness are more in line with the human observer fractional cloudiness. For example the q-q method and smoothing will correct for the surplus of 8 okta values by changing isolated 8 okta values or 8 okta values that are near boundaries of an overcast situation. It should also be noted that the q-q method will also adjust the overall cloud fraction, forcing the averaged ceilometer fractional cloudiness to match that of the human observation of cloudiness.

The technique described above, was applied to data obtained in the years 2000 and 2001 when human observation and ceilometer measurements of cloudiness were taken side-by-side at four AWS's in the Netherlands, namely De Bilt, De Kooy, Eelde and Maastricht. Figure 4 shows two so-called contingency matrices from De Bilt (2000 – 2001) comparing all hourly cloudiness of human observer and ceilometer individually. In a contingency matrix data points are counted in bins where each bin represents an okta value from the human observer (horizontal axis) and an okta value from the ceilometer data (vertical axis).

De_Bilt contingency matrix

ceilometer data [original]

0	647	302	303	14	2	2	0	1	6	2
1	1624	663	738	90	47	28	17	10	8	5
2	901	213	331	98	76	65	46	37	11	8
3	882	129	241	85	75	77	87	71	70	24
4	686	48	109	58	57	85	75	92	98	49
5	898	69	108	46	50	81	81	111	184	141
6	1685	87	171	55	56	82	120	164	342	577
7	4753	106	154	70	74	94	112	200	587	3253
8	5354	5	23	16	17	30	44	59	253	4760
okta		1622	2178	532	454	544	583	746	1560	8833
	okta	0	1	2	3	4	5	6	7	8

human observer data

De_Bilt contingency matrix

ceilometer data [q-q corrected]

0	647	149	376	66	20	10	5	4	6	3
1	1624	232	697	271	146	115	73	57	18	4
2	901	54	221	146	134	106	99	66	56	4
3	882	38	161	101	92	89	127	141	107	5
4	686	11	67	46	69	78	110	148	137	8
5	898	18	83	43	45	56	122	195	288	23
6	1685	20	116	75	69	70	148	308	718	134
7	4753	22	76	92	83	97	169	477	2214	1427
8	5354	0	7	10	16	13	53	118	1307	3741
okta		544	1804	850	674	634	907	1519	4876	5429
	okta	0	1	2	3	4	5	6	7	8

human observer data

Figure 4. Contingency matrixes for the Bilt hourly observations of cloudiness (in okta, 2000 and 2001). Fig. 4a compares human observer (horizontal axis) versus original ceilometer data (vertical axis). Fig 4b compares the human observer with q-q corrected ceilometer data. The comment on the horizontal axis should be interpreted to mean that within the matrix (see text) the okta values represent human observer data. Similarly, for the vertical axis the comment denotes the ceilometer data.

In Figure 4, the matrix itself is contained in the rectangle to the right and upward from the two thick solid lines. The okta values for either set of observations are denoted in the far left blue column and lowest blue line. The column and line just to the left of and downward from the matrix denote the sums of observations in the horizontal bins / vertical bins respectively.

As an example to read this plot: The okta bin 8 on the horizontal axis / okta bin 7 on the vertical axis contains the number 3253. This means that there are 3253 hourly points in the two year period for which the human observer registered a cloudiness of okta 8 whereas the ceilometer registered okta 7.

Figure 4a compares human observer with uncorrected ceilometer data, Figure 4b compares human observer with q-q corrected ceilometer data. There is an improvement between Figure 4a and 4b in so far as that in Figure 4b a larger fraction of measurements fall along the diagonal or in bins at most one okta removed from the diagonal when compared to Figure 4a. This is particularly true for the high cloudiness data points. Furthermore there is a reduction in points whereby the difference in okta indicated by the human observer and okta indicated by the ceilometer is larger or smaller than 2 oktas. Overall, the sum on the diagonals is increased by 10% from the original to the corrected data.

As expected an identity matrix whereby all adjusted observations perfectly match the original time series is not achieved by this method, but an improvement of the results is obtained.

4. Application of the q-q technique to homogenize time series of fractional cloudiness

The question remains what the impact is on the data comparison when synchronous time series are absent. The latter situation is, of course, the common one as prior to the year 2001 only measurements taken by the human observer are recorded, and from 2001 onwards only ceilometer data were taken. The investigator is allowed considerable flexibility in assigning the baseline human observed cloudiness distribution (from the years prior to 2002) and the baseline ceilometer distribution (from the years after 2002). However, the assumption must be made that the two distribution functions are representing the same basic atmospheric conditions. Therefore, it is prudent to select one or several years close to the break when constructing the distribution functions. Here we calculated the relative occurrences of okta values from distribution functions averaged over one to five years counting backwards (for the human observer measurements) or forwards (for the ceilometer measurements) from the break at 2002.

Results for the station of De Bilt are shown in Figure 5. Error bars are not shown but were, in general, less than 1%. It is clear, however, that the q-q correction is quite capable of removing most of the discontinuities in relative occurrences within a particular okta value.

Finally we show the time series of fractional cloudiness (Figure 6). Figure 6 shows that in the Netherlands, during the years 2002 – 2015 the ceilometers observed substantially less clouds than would have been measured by the human observer.

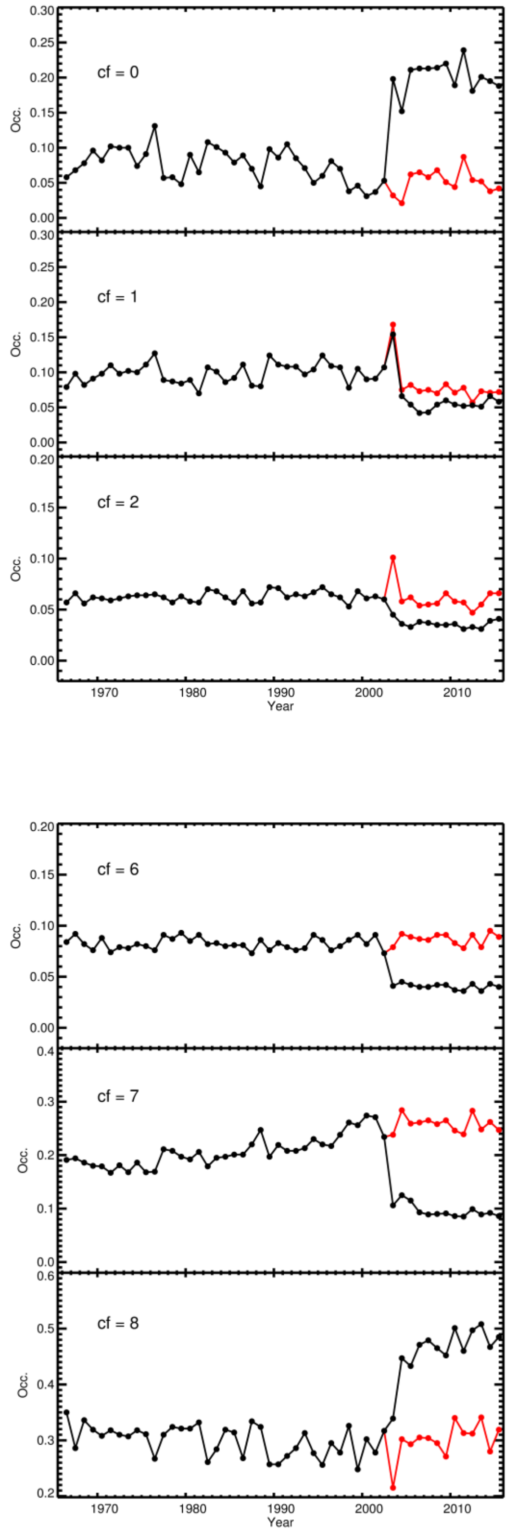

Figure 5. Break-corrected [in red] relative occurrence of cloudiness in an okta class (for cf = 0,1,2,6,7,8 okta). The original time series also appeared in Figure 3.

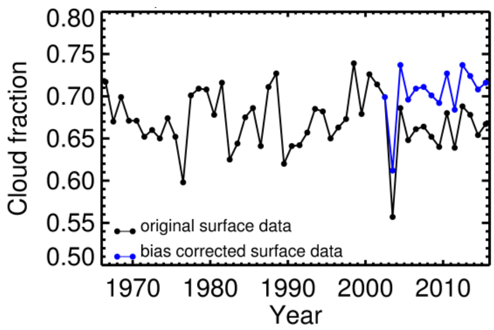

Figure 6. Break-corrected time series of fractional cloudiness for the Netherlands from 1966 – 2015. The original time series appeared also in Figure 1.

The reason is that the ceilometer used by KNMI has a poor sensitivity for high clouds. A comparison with an independent satellite-observed data set confirms this result (see Boers et al., 2017, their Figure 1).

5. Conclusions

In this paper we demonstrated a procedure to homogenize a time series of fractional cloudiness across a break in observation technique. Time series of fractional cloudiness after the break were quantile-quantile corrected using CDF's from data prior to the break and from data after the break.

Results show that fractional cloudiness in the Netherlands as observed by ceilometers is substantially smaller than recorded by the human observer, and that the q-q correction was able to remove the break.

Caution must be taken when applying this procedure to other stations:

a) The procedure is not to be taken as a means to obtain new and accurate time series of cloudiness. It should be realized that perfect time series representing a non-existent human observer can never be reproduced by any procedure correcting ceilometer observations.

b) The procedure is a statistical approach and therefore will only obtain a statistically credible result. In our case the yearly mean fractional cloudiness can be well reconstructed within acceptable margins. As such it is useful when considering trends on a time scale of several years to decades or more.

c) As longer time series of ceilometer data become available it will become more apparent how well this procedure performs under a wide range of atmospheric conditions. At all times, however, it is important to obtain a suitable independent set of observations to benchmark the results. In particular the modern satellite observations of MODIS and SEVIRI provide an adequate source of data with sufficient resolution to select pixels that match the location of surface observations.

6. Literature

Boers, R. T. Brandsma and A. P Siebesma, 2017: Impact of aerosols and clouds on decadal trends in all-sky solar radiation over the Netherlands, Atmos. Chem. Phys, 17, 1-20.

Boers, R., M. de Haij, W. Wauben, H. Klein Baltink, B. v. Ulft, M. Savenije and C. Long, 2010: Optimized fractional cloudiness determination from five ground-based remote sensing techniques, J. Geophys. Res., 115, D24116.

Li, H., J. Sheffield, E. F. Wood, 2010: Bias correction of monthly precipitation and temperature fields from Intergovernmental Panel on Climate Change AR4 models using equidistant quantile matching. J. Geophys. Res., 115, D10101.

Wauben, W., 2002: Automation of visual observation at KNMI; (ii) Comparison of automated cloud reports with routine visual observations. Annual Meeting, Amer. Meteor. Soc., Orlando, Florida, USA.

Wauben, W., H. Klein Baltink, M. de Haij, N. Maat and H. The, 2006: Status, evaluation and new developments of the automated cloud observations in the Netherlands. IOM 94 (TD 1354) WMO Technical Conference, Geneva, Switzerland.