ANALYSIS OF CLOUD INFORMATION FROM SURFACE WEATHER REPORTS

Stephen G. Warren[1], Carole J. Hahn[2] and Julius London[3]

[1]Department of Atmospheric Sciences, University of Washington, Seattle, Washington, USA

[2]Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, Colorado, USA

[3]Department of Astrophysical, Planetary and Atmospheric Sciences, University of Colorado, Boulder, Colorado, USA

[This is the same article published in *Clouds in Climate II* (proceedings of WCRP workshop at Columbia, Maryland, 1987, pp. 75-80), slightly revised.]

1. INTRODUCTION

Ground-based observations of clouds can be a useful complement to satellite observations. The satellite observations give much better spatial and temporal coverage, expecially over the oceans, but the ground observers are better able to distinguish cloud types because they are closer to the clouds and can resolve individual clouds within their field of view. Multiple cloud layers often occur together, so the views from above and below are complementary; the cloud climatology from ground-based observations is more accurate for low clouds than for high clouds. Ground-based observations also offer the potential for investigating changes in cloud amounts over many years; however, the apparent changes sometimes turn out to be due to changes in observing procedure or to sampling biases.

2. CLOUD INFORMATION IN WEATHER REPORTS

Cloud descriptions are coded into weather reports from stations on land and from ships in the ocean. The WMO "synoptic code" is used by both. Some land stations also (or instead) report observations in the "airways code", in which the clouds are described in a different manner. In order to produce a consistent cloud climatology for the whole globe, we have restricted our analysis to the "synoptic" observations.

The information on clouds in the synoptic weather reports consists of total cloud cover (N), low-or-middle cloud amount (N_h), low cloud type (C_L), middle cloud type (C_M), high cloud type (C_H), present weather (ww), and base height of the lowest cloud (h). If information was lacking in a particular category, a slash (/) was recorded by the observer.

N and N_h are integers from 0 to 8, signifying eighths of sky-cover, rounded to the nearest eighth, except that N=0 means completely clear sky and N=8 means completely unbroken overcast. N=9 means "sky obscured", often due to fog, rain or snow. In cases of N=9 we consult the ww code to determine the cloud type, if any. C_L, C_M, and C_H can obtain values 1-9, signifying one of 27 defined types (9 for each level) or 0, meaning no clouds at that level. N_h is the amount of all low clouds present, but if C_L=0 then N_h is the amount of middle clouds. The base height of the lowest cloud present, whether low or middle, is coded in h as a number from 0 to 9. This is the height of the lowest part of the lowest cloud present, even if the predominant low cloud layer is higher.

About 99% of the synoptic weather observations from land stations report ww, 98% report N, 95% report N_h and C_L, 80% report C_M and 63% report C_H. Ship observations are comparable. The number of reports of C_M and C_H is lower than that of C_L, because observations of low overcast lack information about C_M and C_H.

3. <u>DATA COVERAGE</u>

The land observations are normally made every three hours, with somewhat more at GMT hours divisible by 6. Typically about 4100 stations appear in the archive at 0, 6, 12, 18 GMT, of which about 3000 stations also report at 3, 9, 15, 21 GMT. About 20% of the stations do not make observations at night.

Most ships make weather observations only four times per day. In a recent year an average of 1150 ships appeared in the archive at 0, 6, 12, 18 GMT and 160 ships at 3, 9, 15, 21 GMT. There is little tendency for fewer observations at night, but the nighttime observations may not be transmitted promptly by radio so it is important to have the complete logbook records.

In many parts of the ocean the accuracy of the computed mean cloud amounts is limited by the random sampling error because of the scarcity of observations. This is not the case on land, where the random error due to inadequate temporal sampling is very small so that bias errors become relatively more important.

We have analyzed 116 million observations from land stations for the years 1971-1981 and 52 million observations from ships in the ocean for the years 1930-1981. However, our climatology of cloud *types* for the ocean is restricted to the period 1952-1981 because although the synoptic code was defined in 1930, it was not until after 1950 that the reporting procedures for 1930 were uniformly adopted. We used only data that had already been archived on magnetic tape; such data are available for the land, with global coverage, only since January 1971, so our land analysis covers only eleven years.

We are now analyzing the years 1982-present.

4. QUALITY CONTROL

Some of the observations are internally inconsistent so that they had to be corrected if possible, or else discarded. Each observation was put through a series of tests which were possible because of redundancy in the synoptic code. For example, we thus were able to eliminate observations from automatic weather stations which were incorrectly coded (they should not contain any cloud information). When N_h=8, a common mistake is to code C_M=C_H=0. We change these observations to C_M=C_H=/ prior to our analysis, because the higher cloud types are not normally observable through a lower overcast. Observations from China prior to 1980 used an incorrect definition of N_h; we were also able to detect these.

The quality-control procedures we have used are intentionally rather minimal, because quality control can sometimes do more harm than good, by introducing biases into the analysis. In our preliminary work we carried out a pilot study of total cloud cover over the ocean and obtained spurious trends of cloud cover increasing with time because we rejected all observations that had been flagged as erroneous in the data set we were using at that time. These trends appeared because (1) the flagged data had greater average cloud cover than the unflagged data, and (2) the fraction of data that were flagged decreased in time from 40% in 1954 to 0% in 1976. The trends in zonal average ocean cloud cover which were based on those analyses (Figure 3 of Warren *et al.*, 1981) are therefore suspect.

5. METHOD OF ANALYSIS

The latitude-longitude grid for most of our analyses is $5^{o}x5^{o}$. To resolve the annual cycle we obtain average cloud amounts for each of four seasons or (for total cloud cover) twelve months. The diurnal cycle is resolved by obtaining average cloud amounts every three hours where possible, otherwise every six hours. The 27 cloud types are not all studied individually; they are grouped into six major types: cirrus + cirrostratus + cirrocumulus (Ci/Cs/Cc), altostratus + altocumulus (As/Ac), nimbostratus (Ns), stratus + stratocumulus + fog (St/Sc/fog), cumulonimbus (Cb), and cumulus (Cu). In our current work for the 1980s we retain separate statistics for Ci, Cs, Cc, As, Ac, St, Sc, and fog, so we now have eleven types instead of six.

5.1 Frequency-of-occurrence and amount-when-present

The amount of a cloud type is defined as the fraction of the sky covered by that type. The time-averaged amount can be obtained as the product of frequency-of-occurrence (fraction of weather observations in which a cloud of this type is present, whether visible or not) and amount-when-present (the average fraction of the sky which is covered by this cloud type when it is present, whether visible or not). For example, if cumulus is present in 30% of the weather observations from a station, and if it covers an average of 40% of the sky when it is present, then the average amount of cumulus at that station is 12%.

Our method is designed to obtain the true average amounts of each type of cloud, not the non-overlapped amounts seen from above or below. To obtain amount-when-present (awp) of a middle or high cloud it uses the random-overlap assumption when just two layers of cloud are present, and assumes the awp is the same on average in observations where it cannot be calculated as in observations where it can. The frequency of occurrence (f) of a cirrus-type cloud is assumed to be the same when the high level cannot be seen (because of lower overcast) as when it can be seen in the presence of lower clouds. (If this were not done, the amounts of upper clouds would be underestimated because the ground observer is often unable to see them when they are present.) The frequency of occurrence of As/Ac clouds hidden behind a low overcast is assumed to be the same as f(As/Ac) when low cloud is present in amount 3/8 or greater, because As/Ac is positively correlated with St and negatively correlated with Cu.

5.2 Averaging procedure

The true mean cloud cover, or frequency of occurrence or awp of a cloud type, may differ from the mean of a finite number of reports, which is what we compute. Inadequate sampling can lead to both random errors and biases.

A series of tests was performed to estimate the expected error in average total cloud cover as a function of the number of observations used to compute the average. The average of all the observations in the period of record from a land station or an ocean weathership was assumed to be the true average cloud cover for that period. Averages formed from subsets of varying size drawn randomly and repeatedly from the complete set differ from the "true" average, giving an estimate of the expected sampling error. The expected error turns out to be smaller for the ocean areas than on land, because of the smaller day-to-day variability in cloud cover over ocean. To reduce the expected error in seasonal mean cloud cover below 3%, only 100 observations are needed, as opposed to 200 over land.

Various sequences of steps in the averaging procedure can be applied to eliminate sampling biases due to diurnal variations, geographical variations and trends. For the land cloud analysis we judged the diurnal sampling bias to be the most severe (because some cloud types undergo a systematic diurnal cycle, and some $5^o x 5^o$ grid boxes contain stations which report observations only in daytime), so we chose a procedure to eliminate that bias: all observations for a particular reporting hour were first averaged together, irrespective of year, then the eight means were averaged together.

On the ocean (unlike the land) the clouds in a $5^o x 5^o$ grid box are not reported regularly. Some days are sampled by several ships; other days by only one ship. In the ocean we therefore average together all observations made during a single three-hour period on a particular date in a particular $5^o x 5^o$ grid box, to form a "compressed observation". Subsequent analysis of average cloud amounts is based on compressed observations.

6. BIASES

The use of compressed observations reduces the fair-weather bias (the tendency for more ships to enter a $5^o x 5^o$ box on days of fair weather), the foul-weather bias (the tendency of ships to oversample stormy or foggy weather because they are traveling more slowly), the day-night bias (somewhat more reports are transmitted by ships during the day than at night), and the trend bias (a box may be sampled by more ships each synoptic hour in later years than in earlier years). Furthermore, the presence of duplicate observations in the dataset does not damage our analysis if they are preaveraged by us to form a compressed observation.

The most serious bias, both on land and ocean, is the "night-detection bias": observers are often unable to detect the upper clouds Ci/Cs/Cc and As/Ac at night when they are present. For this reason we used only daytime observtions to obtain average amounts of Ci/Cs/Cc and As/Ac. We are therefore unable to investigate the diurnal cycle of these cloud types.

A bias which causes our cumulonimbus amounts to be excessive is that the instructions for the synoptic code require the low cloud type to be coded as Cb if any Cb is present, even if other low clouds cover more of the sky.

Biases that apply to ships, but not to land stations, were investigated by comparing the reports of stationary weatherships to those of transient ships passing nearby. The weatherships reported a seasonal average cloud cover 2.0% larger than that of the transients; 1.8% if compressed observations are used. This is the sum of two biases: 0.4% is a fair-weather bias and 1.4% is an "observer bias"; the better-trained observers on the weatherships estimate greater cloud cover (for certain cloud types) than do the observers on transient ships.

7. ERRONEOUS REPORTS

Twenty land stations were omitted from our analysis because they routinely transmit erroneous reports. Individual ships transmitting erroneous reports, however, are more difficult to detect because no one ship dominates the climatology of a grid box. The most damage is done by reports in which latitude or longitude is incorrectly transmitted (at least 1% of all reports). These reports can contaminate the cloud analysis in sparsely-sampled regions of the southern hemisphere.

8. EARTH-COVER VS. SKY-COVER

The fractional cloud amounts reported are fractional "sky cover", which is the fraction of the celestial hemisphgere covered by clouds. This is in general larger than the "earth cover", which is the fraction of earth covered by clouds when the clouds are projected vertically. There is of course no difference between the two quantities if the cloud cover is 0% or 100%. The difference depends on the ratio of vertical to horizontal dimension of the cloud, so is most significant for Cu and Cb. Ground observers are instructed to report sky cover, but some observers perform a mental compensation so as to report earth cover rather than sky cover.

The difference between earth cover and sky cover is probably partly responsible for some of the difference between cloud cover estimates from satellite observations and those from surface observations; however, satellites also do not observe earth cover unless they are viewing to the nadir.

9. RESULTS

The results of our analyses are presented as maps in four atlases (Hahn *et al.*, 1982, 1984; Warren *et al.*, 1986, 1988), along with considerable discussion about the method of analysis. The results are also archived on a single magnetic tape (containing both land and ocean analyses) available from NCAR (Boulder, Colorado) or the Carbon Dioxide Information Analysis Center (Oak Ridge, Tennessee).

10. REFERENCES

Hahn, C.J., S.G. Warren, J. London, R.M. Chervin, and R. Jenne, 1982: *Atlas of Simultaneous Occurrence of Different Cloud Types over the Ocean*, NCAR Technical Note TN-201+STR, Boulder, CO, 212 pp. (NTIS number PB83-152074).

Hahn, C.J., S.G. Warren, J. London, R.M. Chervin, and R. Jenne, 1984: *Atlas of Simultaneous Occurrence of Different Cloud Types over Land*, NCAR Technical Note TN-241+STR, Boulder, CO, 216 pp.

Warren, S.G., C. Hahn and J. London, 1981: Ground-based observations of cloudiness for cross-validation of satellite observations. In *Clouds in Climate* (W. Rossow, ed.), NASA-GISS, New York, 174-179.

Warren, S.G., C.J. Hahn, J. London, R.M. Chervin and R.L. Jenne, 1986: *Global Distribution of Total Cloud Cover and Cloud Type Amounts over Land.* NCAR Technical Note TN-273+STR, Boulder, CO, 29 pp. + 200 maps (NTIS number DE87-006903).

Warren, S.G., C.J. Hahn, J. London, R.M. Chervin and R.L. Jenne, 1988: *Global Distribution of Total Cloud Cover and Cloud Type Amounts over the Ocean.* NCAR Technical Note TN-317+STR, Boulder, CO, 42 pp. + 170 maps.