Climatology of 24-h precipitation totals at SYNOP stations

Thomas Haiden, ECMWF

1. Introduction

For the computation of SEEPS and other scores which involve station climatology the weights of the SEEPS scoring matrix and the 1,2,..,99 percentiles have been computed at each station for the 30-yr period 1980-2009 (Rodwell et al. 2010). They have been made available to facilitate the computation and inter-comparison of quantitative precipitation forecast verification between NWP centres. The quality control procedures used in the compilation of the data are described in Section 2. The data is in ascii format, details of which are described in Section 3. Section 4 explains how SEEPS is computed using an example program which is included in the download package.

2. Quality control procedure

a. Rejection of individual observations by latitudinal filtering

24-h precipitation totals have been constructed from 6-h, 12-h, and 24-h SYNOP observations disseminated via GTS. In a first step all negative values, and all values larger than an upper limit which depends on latitude, are rejected. The latitudinal dependence of this limit is based on the envelope of the observed latitudinal distribution of the bulk of 99th percentiles, which can be approximated by

$$P_{99}(\varphi) \approx P_{99}(0)\cos(\varphi) \qquad (1)$$

where $P_{99}(0) = 200$ mm (continuous blue curve in Figure 1). The condition for rejection of individual precipitation observations has been defined as $P > 5\,P_{99}(\varphi)$, which is shown by the dotted blue curve in Figure 1. According to this limit, at the equator values larger than 1000 mm, and at 60° latitude values larger than 500 mm are rejected. The fact that expression (1) goes to zero at the poles does not pose a problem because the station closest to a pole is located at 82.5°N. (At 82.5°, the limiting amount is still 130 mm.) The set of percentiles retained after this initial filtering (but before subsequent filtering based on percentiles) is shown in the panels on the left in Figure 1.

b. Rejection of stations based on percentiles

Erroneous values which are not large enough to be individually rejected may still lead to unphysical cumulative distribution functions (CDFs), and additional stations which should be rejected can thereby be identified.

For each month and station, percentile values larger than 100 mm which occurred more than once (e.g. $P_{97} = P_{98} = P$, where $P > 100$ mm) were identified and, after manual inspection, in most cases discarded. In this process, many instances of two specific values, 420.0 mm and 819.1 mm, were found. Stations with percentiles assuming either of these two values, even if just once, were rejected (after manual inspection).
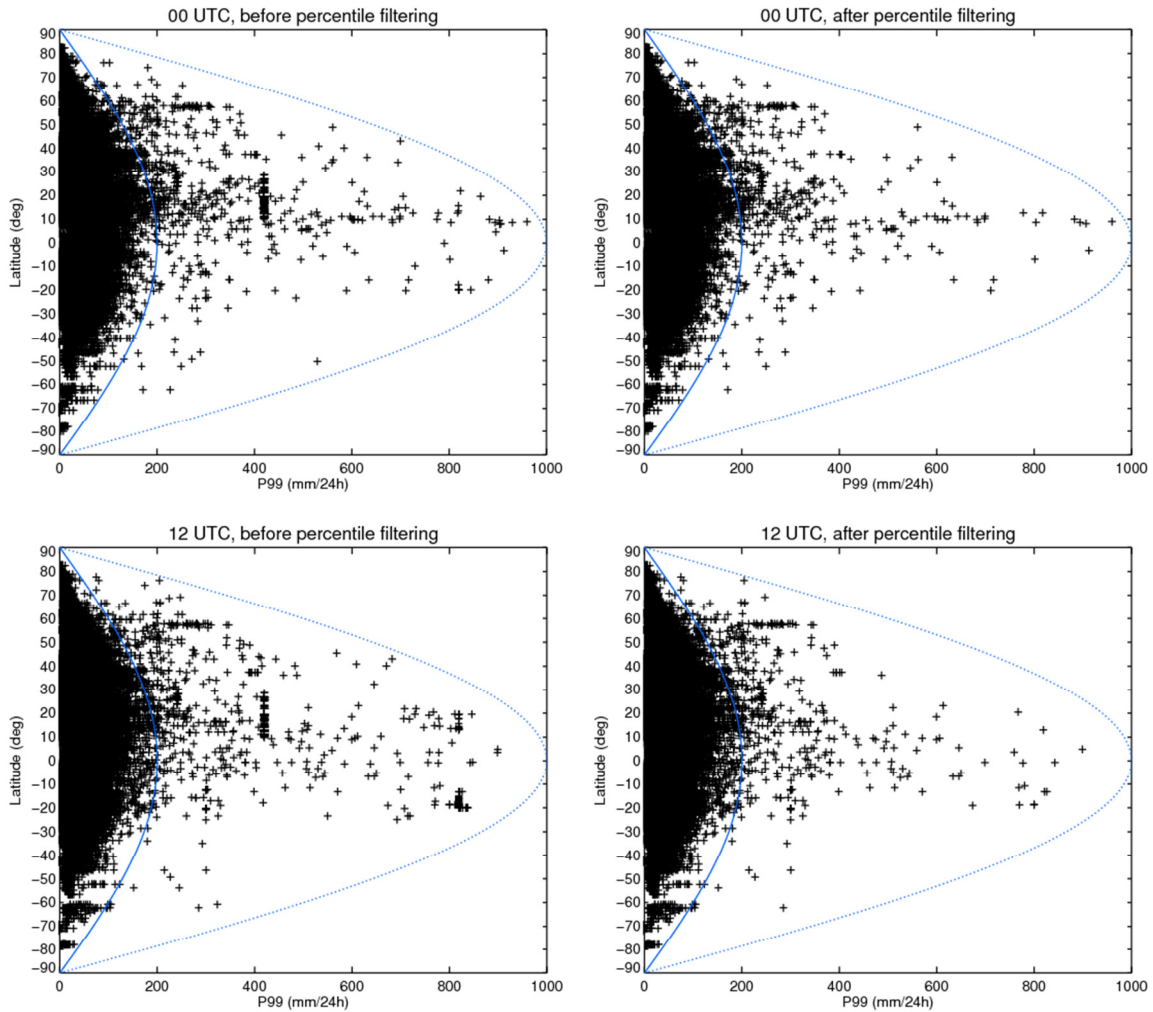
Figure 1: Latitudinal distribution of 99th percentiles of 24-h precipitation at 00 UTC (top) and 12 UTC (bottom) after latitudinal filtering but before percentile filtering (left) and after both latitudinal and percentile filtering (right). The continuous blue curve shows expression (1), which approximately envelopes the bulk of values and is used as a reference. The dotted blue curve is 5 times the envelope which is used to filter individual observations.

Finally, stations where P_{99} is more than twice the latitudinal envelope given by (1) were rejected if the ratio between consecutive percentiles near the upper limit of the distribution was larger than 10. The set of percentiles retained after the percentile filtering is shown in the panels at the right in Figure 1.

For 00 UTC, the percentile-based filtering rejects 44 stations, but most of them only for 1-2 months out of 12. The number of rejected station×months is 218, which is 0.37% of the total (59106 station×months). The majority of these stations are located in Africa, and Central and South America. For 12 UTC, the filtering rejects 69 stations and 246 station×months which is 0.38% of the total (64355 station×months).

It is very likely that there are still erroneous values in the data which were not captured by the filtering algorithm described above. Further filtering could be based on a more sophisticated approach, such as taking into account to what extent stations are affected by orographic precipitation, comparison with nearby stations, or consistency checks with other quantities

like cloudiness. However, since the SEEPS weights are not particularly sensitive to the number of the heaviest precipitation events, and insensitive to their exact values, the resulting SEEPS score is hardly affected by such quality control. The percentiles at the high end of the distribution, on the other hand, are naturally sensitive to the heavy precipitation events, and need to be used with this caveat in mind.

3. Data format

The percentiles and SEEPS weights have been written to monthly ascii files, where for each month there is a 00 UTC and a 12 UTC file. The filenames are

```
PPT24_percentiles_[mm]_[hh]
PPT24_seepsweights_[mm]_[hh]
```

where `mm=00,01,..,12`, and `hh=00,12`. After a two-line header, each line contains data for one station, where stations are ordered by decreasing latitude. In the percentile files, the following 105 columns can be found:

lat	Latitude (deg)
lon	Longitude (deg)
sid	WMO station number
yyyymmdd	Date of the latest observation used
height	Station elevation (m)
nobs	Number of observations available
1,2,..,99	Percentiles (mm)

Since the period considered here is 30 years, the maximum possible value of `nobs` is 930 (in a month with 31 days). Only those stations for which at least 150 observations were available, corresponding to ~5 years of data, were included in the files. The files containing the SEEPS weights have 18 columns and a similar form:

lat	Latitude (deg)
lon	Longitude (deg)
sid	WMO station number
height	Station elevation (m)
nobs	Number of observations available
p1	Probability of the 'dry' category
t1	Threshold between the 'dry' and 'light' categories (mm)
p2	Probability of 'dry' + probability of 'light'
t2	Threshold between the 'light' and 'heavy' categories (mm)
s11,s12,..,s33	Weights of the 3×3 SEEPS contingency table

The set of stations (and their ordering) is identical in the percentiles and in the SEEPS weights files.

4. How to compute SEEPS

The short, self-contained, standard C-program `seeps.c` demonstrates the computation of the SEEPS score for an example time period of 10 days for an example station, using the climatological information described in Section 3. The derivation and properties of SEEPS are discussed in detail in Rodwell et al. (2010). The use of SEEPS in a recent global model inter-comparison study has been documented in Haiden et al. (2012).

References

Haiden, T., M. J. Rodwell, D. S. Richardson, A. Okagaki, T. Robinson, and T. Hewson, 2012: Intercomparison of global model precipitation forecast skill in 2010/11 using the SEEPS score. *Mon. Wea. Rev.,* **140,** 2720-2733.

Rodwell, M. J., D. S. Richardson, T. D. Hewson, and T. Haiden, 2010: A new equitable score suitable for verifying precipitation in NWP. *Q. J. R. Meteorol. Soc.*, **136,** 1344-1363.

For questions, contact: thomas.haiden@ecmwf.int