4.11 OPTIMIZED MERGING OF HOURLY PRECIPITATION WITH DAILY COOP PRECIPITATION DATA

Dongsoo Kim *
NOAA National Climatic Data Center, Asheville, North Carolina

1. INTRODUCTION

Precipitation reports from NOAA's Cooperative Observer (COOP) network have been the most important data source in describing long term climate change and variability of precipitation even though some concerns of quality have been reported (e.g., Daly et. al, 2007). When it comes to understanding precipitation patterns such as frequency and intensity, daily observations limit detailed analyses. However, hourly precipitation data are greatly needed to study sub-daily scale patterns of heavy rains in response to global change scenario. The traditional hourly precipitation data are currently served by NOAA's National Climatic Data Center (NCDC Data Set 3240, referred to Hourly Precipitation Data, HPD). As Brooks and Stensrud (2000) concluded the station spacing of HPD misses most of the truly large precipitation events.

Recently, NCDC reprocessed more than 10 years of historical Hydrometeorological Automated Data System (HADS) precipitation data that are not part of HPD. The data provide an opportunity to narrow spacing of hourly stations and to assess quality of daily COOP and integrate into hourly hydro-climate database. However, the challenge remains as how to control errors which are unique to each observing system. It is a known problem that the operators of COOP rain gauges occasionally report a 24-hour total on its perceived day of occurrence rather than on the day it occurred. These so-called "shifting" errors are difficult to detect. Reprocessed hourly HADS precipitation data (Kim et. al, 2006) which have similar spatial density with daily COOP stations (Figure 1) present an opportunity to examine possible COOP observation time shifts independently. Since HADS hourly data are automated, reporting in the wrong day (even wrong hour) is less likely, while COOP daily precipitation amount is relatively more reliable than HADS precipitation. Thus, an optimal approach in the merging of hourly and daily precipitation data is necessary.

As a part of rigorous quality assurance of HADS hourly precipitation data, we developed an optimization method that minimizes the differences of daily COOP and nearby time-aggregated HADS precipitation subject to the following condition being met: both COOP and HADS must record precipitation ($\geq$ 0.01 inch/hour). This condition is termed "co-raining".

The minimization of the objective function is to estimate the hour of the day when preceding 24 hour accumulation of HADS hourly data best matches COOP daily precipitation during the month. Through this optimization, we can select hourly precipitation and daily COOP data to provide best input to Multi-sensor Precipitation Estimate (Nelson et. al, 2008).

* *Corresponding author address:* Dongsoo Kim, NOAA/NESDIS/NCDC, 151 Patton Ave, Asheville, NC 28801; e-mail: dongsoo.kim@noaa.gov

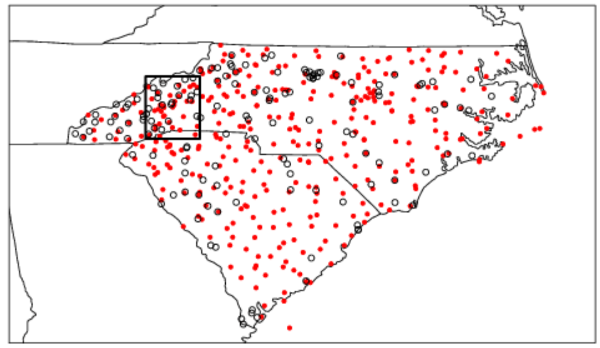

Figure 1. Distribution of hourly HADS stations (open circles) and daily COOP stations (solid dots) in September 2004 in pilot domain of North and South Carolinas. COOP stations distribution is more homogeneous than that of HADS. Box shows 1^ox1^o spatial dimension.

2. OPTIMIZATION METHOD

We formulate an objective function of daily observational hour with hour range between 0 and 35. This parameter has shifting role in the aggregation of hourly precipitation data for daily sum, hence we call it a "shifting parameter" which varies from 00 UTC to 11 UTC next day. The optimization algorithm finds the value of the shifting parameter that minimizes mean squared differences between daily COOP, P^C, and aggregated HADS data, P^H during a month. The superscripts H and C stand for HADS and COOP respectively. The objective function is expressed;

$$F(k) = \frac{1}{n}\sum_{i}^{n} \left(P_i^H(k) - P_i^C\right)^2,$$

where

$$P_i^H(k) = \sum_{j(i)+k}^{j(i)+23+k} p_{j(i)}^H.$$

The variable k is the shifting parameter in (0, 35), i is an index for day, index function $j(i)$ converts day of month to hour of month, $j(i)$ = 24*(i-1) + 1, and $p_{j(i)}^H$ is HADS hourly precipitation at the data point $j(i)$, n is number of days when either COOP or HADS reports precipitation. Both P^H and P^C are in units of inch/day, $p^H_{j(i)}$ is in inch/hour. We do not compute the objective function when both stations report no-rain.

We also compute co-raining function *m(k)* to constrain solution space such that minimizer is found near the largest value of *m*. The co-raining function is an integer function of number of days when both HADS and COOP report rains. This function is a qualitative function as it considers only raining cases regardless of how heavy they are. Sometimes, trace of precipitation in both COOP and HADS does not have a necessarily desirable solution. In case of heavy rains with short duration, the largest value in co-raining function may not yield the best result.

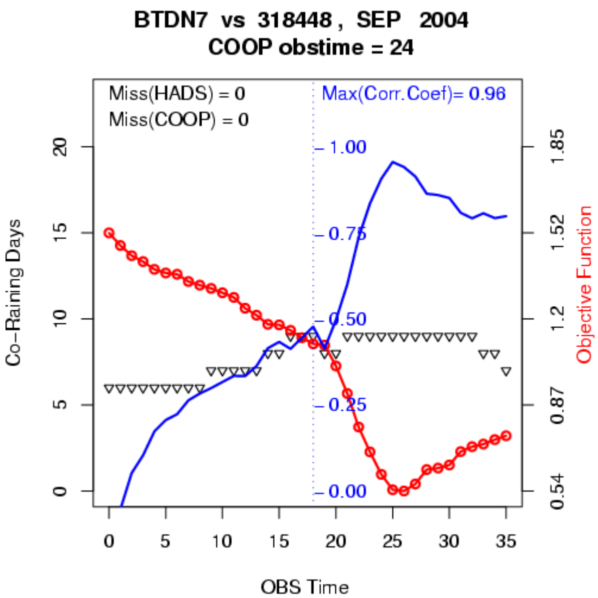

Figure 2. Objective function applied to COOP station 318448 and HADS station BTDN7 distanced by 7 km for the September, 2004 (red colored open circles). The co-raining function is represented by triangular symbols. The best match between two is at 26th hour in LST, two hours after operator's record at midnight. The blue solid line is correlation coefficient function.

The optimization process is illustrated via two examples. The first example is the pair of COOP station (318448) and HADS station (BTDN7) near Swannonoa, North Carolina. The second one is 313976 and HVLN7 near Hendersonville, North Carolina. The first example is in Figure 2 that display three relevant quantities in determining the shifting parameter which best matches the pair of locations. The objective function *F(k)* is in red circles connected with solid line. The co-raining function *m(k)* is denoted in triangles, and correlation coefficient function is in solid line. The first example shows that optimum shifting parameter is found regardless of co-raining constraint at 26^{th} which is 2AM next day. It differs 2 hours from reported COOP observation time.

The second example in Figure 3 shows sensitivity of solution to the constraint. The co-raining function *m(k)*=12 forces solution to be 8AM, while relaxed condition $m(k) \geq 11$ forces solution to be 6AM. Both estimated observation times are near reported observation time.

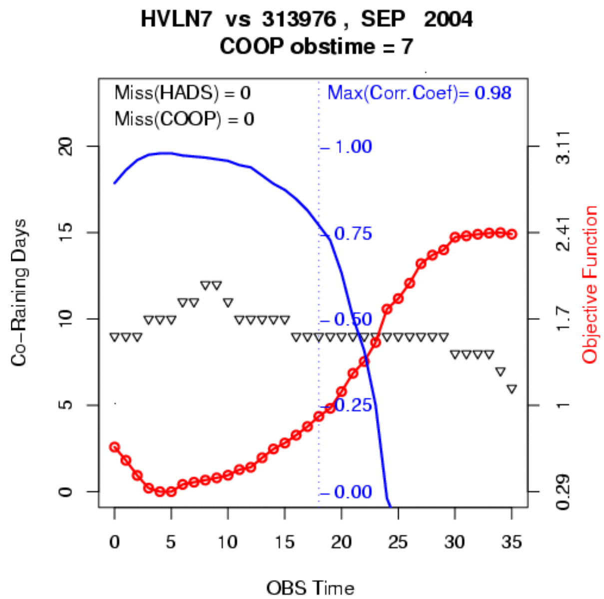

Figure 3. The same as Fig. 2 except station pair of 313976/HVLN7. The COOP station is 313976 whose observation is reportedly made at 7 AM LST, and nearby HADS station HVLN7 distanced by 12 km. The co-raining days of 11 and 12 impose a solution be found between 6AM and 10AM, so the optimum solution is 6AM.

3. DISCUSSION IN LOCAL DOMAIN PAIRS

We have applied optimization process to available pairs in the local domain, within +/- 0.5 degree from the respective COOP stations (Box in Fig.1). Then, estimated shifting parameters are compared with reported COOP observation time. Discrepancies of estimated shifting parameter from reported COOP observation time are attributed to variability of rain events during the month. Missing values in HADS data and erroneous COOP observation time also contribute discrepancies. There were 18 HADS stations available in the local domain of COOP station (318448). Figure 4 shows a spatial distribution of HADS stations relative to COOP station, and the differences of estimated shifting parameter from the COOP observation time within a 1^ox1^o local domain.

The negative values in Figure 4 indicate that precipitation occurred ahead of COOP observation time, while positive values show they occurred after the time. Two HADS stations are indicated in time difference field. The HADS station SAMN7 has 10 hours difference, and SWRN7 shows -24. Investigation revealed that both HADS stations missed observations during rain events. The spatial variation of time differences reflects variability of precipitation and predominant pattern during September 2004. This month had three remnants of Hurricanes (Frances, Ivan and Jeanne) which passed through the region (Fig. 5).

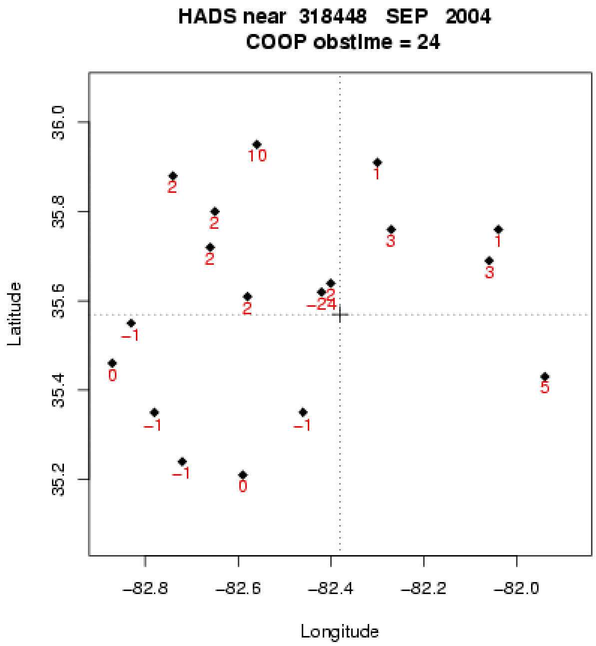

Figure 4. Spatial distribution of HADS stations relative to COOP station 318448. The numbers are time differences (estimated shifting parameter minus reported COOP time) of each pair.

In general, cases of large time differences show local minima in the objective function. In such cases, we should consider a possibility of COOP reporting time error. For example, three pairs of HADS stations with COOP (319100) in September 2004 displayed unreasonable shapes of objective function. Figure 6 is a pair between TRANS and 319100 located in Tar River Basin, coastal region of North Carolina. The official record of COOP observation time was 18 LST (6 PM), but the estimated time was at 7 AM (Fig. 6). Further investigation revealed that COOP operator had been reporting observation time at 18 LST in September and switched to 8 LST since October 2004. Therefore, it is very likely that operator actually measured daily precipitation at 8 AM but did not update observation time in September.

4. SUMMARY AND PLAN

We have developed an optimization method to estimate a parameter which is observation hour of the day. The method is demonstrated to identify possible gross errors and help users determine their usages. For example, if missing values in hourly precipitation data occur during rain events, then the objective function will not yield global minimum in parameter space. Similarly, if COOP observation time is in error, then observation time difference with respect to neighboring HADS stations will display consistent spatial pattern of time difference.

We plan to implement the optimization method to improve screening of questionable data, and thus lead to better integration of hourly and daily precipitation data. The output of optimization will be available along with reprocessed hourly precipitation data at http://eclipse.ncdc.noaa.gov:9090/hads/.

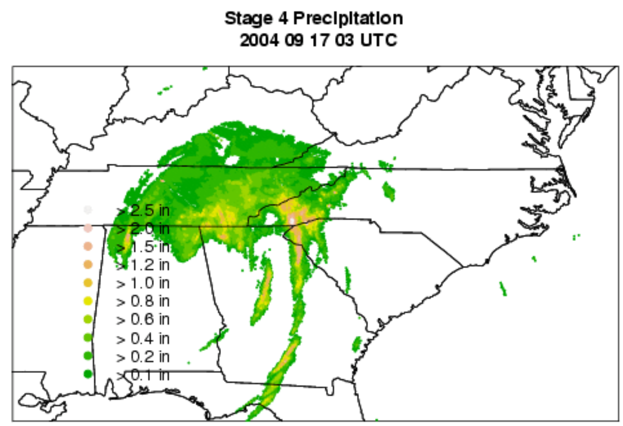

Figure 5. An hourly precipitation of Stage 4 product valid at 03 UTC 17 September 2004. Rain bands by remnant of Hurricane Ivan were moving from southwest toward northeast over Western North Carolina region.

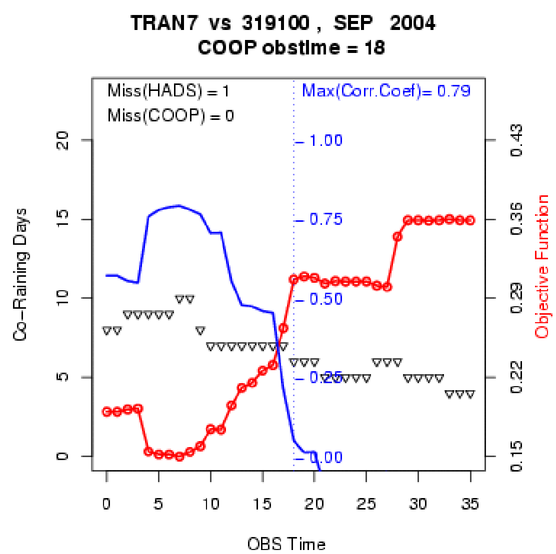

Figure 6. An objective function applied to COOP station 319100 and TRAN7, Beaufort County, North Carolina. The official COOP observation time was 18 LST, but the optimization estimate it 7 AM.

5. ACKNOWLEDGMENT

Author acknowledges careful review by Drs. Matt Menne and Sharon LeDuc and extends thanks to Drs. Seo and Nelson for continued discussions on precipitation QC issues.

6. REFERENCES

Brooks, H. E. and D. J. Stensrud, 2000: Climatology of heavy rain events in the United States from hourly precipitation observations, *Mon. Wea. Rev*., Vol 128, 1194 – 1291.

Daly, C., W. Gibson, G. Taylor, M. Doggett, and J. Smith, 2007: Observer bias in daily precipitation measurements at United States cooperative network stations. *Bull. Amer. Meteor. Soc.*, **88**, 899-912.

Kim, D., B. Nelson and L. Cedrone, 2006: Reprocessing of historic hydrometeorological automated data system (HADS) precipitation data, Preprint, AMS 20th Conf on IOAS, 29 Jan. - 2 Feb., 2006, Atlanta, GA

Kim, Dongsoo, 2007: Analyses of heavy rainfall variations in the U.S. from the reprocessed HADS precipitation data. Joint AMS 21st Conf. on Hydrology and 19th Conf. on Climate Variability and Change, 15-18 January 2007, San Antonio, Texas

NCDC, 2003: Data documentation for data set 3240 (DSI-3240), National Climatic Data Center, 151 Patton Ave, Asheville, NC 28801-5001, USA.

Nelson B. R., D. Kim, and D. J. Seo, 2008: The multi-sensor precipitation reanalysis: Science Issues, to be submitted to Journal of Hydrometeorology (in preparation)