Per the request made by the Hydrology Subcommittee, the following metadata was assembled to describe the historical monthly precipitation estimates made by GLERL. The specifics of the request, as reported to me by Tom Croley in the notes he took at the meeting on about May 6, 2008, are:

1. Description of methodology (or methodologies).
2. History of methodology changes.
3. When were values recomputed / replaced?
4. They want a short report to be posted on the Coordinating Committee website – go through NOS co-op (Tom Landon and Cary Wong)
5. Update this metadata whenever there are updates/changes to the data set.

The monthly precipitation estimates that are published by GLERL and adopted by the Coordinating Committee as the official estimate were originally developed for use in a publication entitled “Great Lakes Monthly Hydrologic Data” (Croley and Hunter, 1994). The following excerpt from that report describes the method used for computing the values presented in that report.

> Due to the quality of the available station data, estimates were derived in three ways, corresponding to the available periods of record for different segments of the historical time series. Data prior to 1930 (1918 for Lake Superior) were computed by the Lake Survey District of the U.S. Army Corps of Engineers and by the National Ocean Survey. Data for 1930-1947 (1918-1947 for Lake Superior) were computed at the Great Lakes Environmental Research Laboratory (GLERL) from monthly station values. Data for 1948-1990 were computed at GLERL using daily station data. The methodology for each period is briefly described below.
>
> The Lake Survey District of the U.S. Army Corps of Engineers, and later the National Ocean Survey, computed monthly precipitation estimates for data earlier than 1930 with an areally weighted "district" approach (Quinn and Norton, 1982). Monthly over-land precipitation data begin in 1882 (Superior), 1883 (Michigan), 1883 (Huron), 1900 (St. Clair), 1882 (Erie), and 1883 (Ontario). Districts (large areas) were designated and divided into sub-districts (smaller areas). Arithmetic means for each sub-district were computed from all stations chosen for that subdistrict. The sub-district values were areally weighted to compute district precipitation. Over-land values were computed by areally averaging district values. Over-lake values came from the use of nearshore stations. Monthly over-lake precipitation data begins in 1900 for all lakes.
>
> Quinn and Norton (1982) computed 1930-1947 monthly precipitation by using a modified Thiessen weighting approach. They areally weighted monthly station data on a 5-km grid to calculate each monthly value. Each grid square, belonging to the basin or lake, was evaluated to find its nearest station; the relative counts for each station within the basin or lake were used as Thiessen weights to areally combine station values for the basin or lake. All available stations within 25 kilometers of the basin were used, and the weights were recomputed for each month as necessary.
>
> We computed 1948-1990 monthly precipitation from all available daily data from stations in the basin or within approximately 0 - 30 km of the basin, depending upon station density near the edge of the basin. The distance was chosen to assure that we obtain the same non-zero Thiessen weights as if no stations were eliminated. Station data for the U. S. were obtained from the National Climatic Data Center (1987), and station data for Canada were obtained from the Atmospheric Environment Service (1981). We then used Thiessen weighting (Croley and Hartmann, 1985) similar to the above, but defined for a 1-km grid and recomputed every day for all watersheds within the basin and the lake. The basin watersheds were defined by the U. S. Geological Survey and the Water Survey of Canada. We constructed daily over-land basin precipitation by areally weighting the areally-averaged daily precipitation for each watershed and summing over the days in each month. We also constructed total monthly over-lake precipitation by summing the areally-averaged daily values for the lake surface.

The values that are currently provided by GLERL are similar, in that the data prior to 1948 are still from the same sources identified in the report. The method used for the 1948-present data is also fairly similar, though some fine points of the processing have changed a few times since the report. These minor ad hoc methodology changes have resulted in some variation in the values that were published. We did not regard the changes as significant, as the methodology employed was consistent with the description given within the report, and we did not realize that the values were being regarded as the official estimates. Based on interaction with personnel from NOS in the late 1980s and early 1990s, we believed that NOS was independently producing the official estimates. Not until approximately 2004 or 2005 did we become aware that this was not the case, and that NOS was actually using our values as their estimate. At this time we took steps to develop a strictly standardized procedure that can be duplicated and rigorously described. The following timeline is presented to give an approximate history of the changes that we made in our methodology. Please be aware that this is constructed based on best recollections in 2008. Therefore, some dates and details may be inaccurate.

1980s: Stations used to generate areal estimates are chosen from a static "master list" of allowed stations. This list was built in the early to middle 1980s, and was not changed until the early to middle 1990s. The exact selection criteria used when building this original list is somewhat vague, but seems to have followed the general rule of 30 km from the basin boundary. Updates to the data set were performed on an irregular basis, generally only about every 2 or 3 years due to the cost of procuring the data and the labor cost in processing it (format changes were fairly common). This irregular update schedule persisted until about 2006.

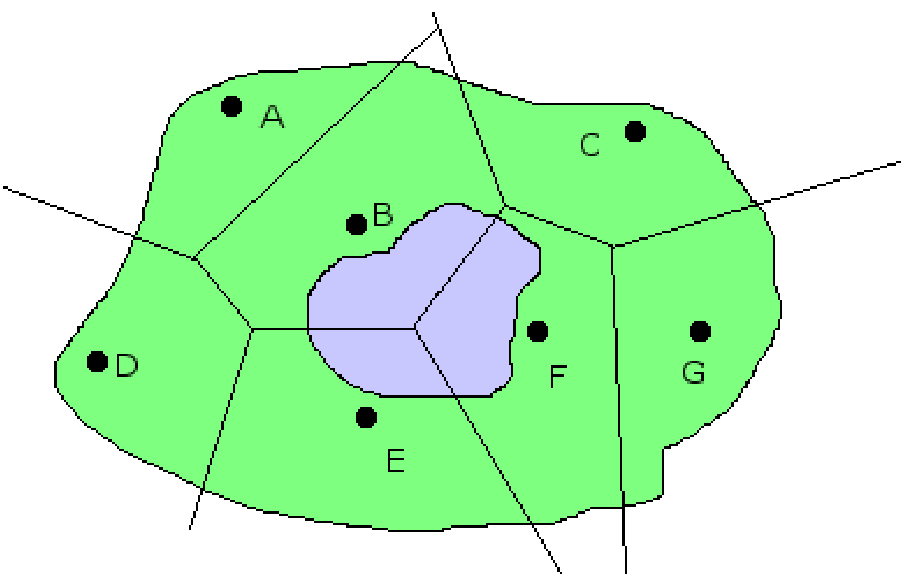
1994: The static "master list" method was still employed, but we rebuilt the list by plotting the station locations on a map of the lake and watersheds, then choosing the applicable stations in a somewhat arbitrary "eyeball" method. This allowed the use of new stations that had been added within the previous 10 years or so. The goal was to select sufficient stations outside the basin boundary so that every station that might be used with the Thiessen technique would be included. However, we also tried to minimize the total number of station files (to reduce processing time) and refrain from choosing stations from too far away (that might not be representative). This meant that the distance from the basin boundary was not a fixed value, but rather was a very subjective choice based on station density. As such, it was not a clearly reproducible process, even for the same operator. Once the master list was produced, processing proceeded as described in the report.

1997: We recognized the arbitrary nature of the station selection and desired to get something that was a little easier to describe and reproduce. Therefore, sometime in the late 1990s, we switched to a simple box selection method for stations to be included in the processing. A rectangular box (defined by latitude and longitude coordinates) was selected around each lake basin and all stations within the box were included in the processing. The advantage was easy specification of the selection criteria so that others could duplicate our results, if desired. The downsides were two-fold. First was that the selection of the box was still an arbitrary decision. The second was that some stations were now being included, particularly in the corners of the box, which were a significant distance from the lake basin watershed.

2004: After we were made aware that our estimates were being regarded as the official estimate, we realized that we needed a better station selection method; one that would be both efficient and easily repeated. So we developed software that would allow the user to choose all of the stations within a specified distance from the basin boundary. There is now just one arbitrary component (the distance) in the process. Given that information, the process is easily and completely repeatable by others. We have chosen 50 kilometers as our specified distance for each of the lakes except for Lake Superior. Due to the relatively sparser station network, particularly in the northern reaches of the basin, we chose to use 60 km as our distance for Lake Superior.

2006: We began to do annual updates of the precipitation estimates. These updates require that we procure the station data from the U.S. and Canadian data agencies. Because there is a certain amount of lag time between the collection of the data and the time that the agency completes their quality assurance checks to certify it as final, we do this update in the late summer or early fall. Under the recent processing schedules, this has allowed us to get fully QA station data from the U.S. for the entire previous year (ending 8 months or so prior to the date of our processing). The Canadian data has still been something of an issue, as their backlog for completing the QA is longer. We receive a number of final quality stations that extend through the end of the prior year, but many of the stations have only been finalized through the middle of the prior year. Thus, the monthly precipitation estimates that we produce for the most recent year (e.g. 2006 if we are processing in summer/fall of 2007) are subject to change when we do the annual update the following year. In addition, we intend to reprocess the ENTIRE 1948-present period each time. This will allow us to incorporate the results of any additional quality issues that the agencies may happen upon in their historical databases. This also means that it is possible (though not expected now) for older data values (e.g. in 1970) to be changed when we do the update.

One other question that is frequently raised with respect to these precipitation estimates is a clearer explanation of the difference in calculating overland and overlake precipitation. Consider the figure below that depicts a simple lake basin with seven precipitation stations. The applicable Thiessen polygons (approximated by freehand drawing) are also shown. The overlake value that we compute is the value over the blue area (lake surface) resulting from the weighted average of the intersection of the lake surface with the polygons. Thus, only stations B, E, F contribute to the calculated overlake value. The overland value is calculated from all 7 stations.



Latest revision: May 16, 2008