# UNCERTAINTY QUANTIFICATION IN THE NET BASIN SUPPLY OF LAKE ERIE AND LAKE MICHIGAN

## *Quantification de l'incertitude en l' approvisionnement net du bassin du Lac Erie et du Lac Michigan*

Carlo, DeMarchi[1]

Department of Geological Sciences, Case Western Reserve University, Cleveland, Ohio 44106-7216, USA
carlo.demarchi@case.edu


Qiang, Dai

Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, Michigan 48109-2125, USA
qiangdai@umich.edu


Mary, E. Mello

Cooperative Institute for Limnology & Ecosystems Research, University of Michigan, Ann Arbor, Michigan 48109-1041, USA
mary.e.mello@gmail.com


Timothy, S., Hunter

NOAA Great Lakes Environmental Research Laboratory, Ann Arbor, Michigan 48105-2945, USA
tim.hunter@noaa.gov

## KEY WORDS

River runoff, Lake evaporation, Over-lake precipitation, Monte Carlo analysis, Great Lakes.

## ABSTRACT

*This paper analyzes the uncertainty in the widely used NOAA Great Lakes Environmental Research Laboratory (GLERL) estimates of two key components of the Great Lakes' net basin supply (NBS): over-lake precipitation and tributary flow. GLERL estimates over-lake precipitation using Thiessen Polygon interpolation of gage data, which is affected by the lack of gages on the lakes themselves and the scarcity of gages in parts of the draining basin. Recently, estimates based on merging precipitation radar data and rain gage data proved to be a better alternative to the traditional estimate methods for some of the lakes. We compared traditional over-lake precipitation models with the radar-based estimates for 2002-2007, computed the relative error distribution, and used it for modeling the uncertainty in historical over-lake precipitation estimates, indicating that uncertainty is around ±30% for Lake Erie to around ±25% for Lake Michigan.*

*Tributary runoff should be the least uncertain component in the Great Lakes NBS estimates, but the partial monitoring of most lake basins requires its consideration. We analyzed the algorithm used by GLERL to produce the tributary flow data. We modeled the uncertainty in each single sub-basin runoff estimate and evaluated the overall uncertainty in tributary runoff to a lake using a Monte Carlo technique. Results indicate that it is around 23% for both lakes. The uncertainty of the overall NBS is also presented complementing these results with literature values for the uncertainty on evaporation.*

---

[1] Corresponding author

**RESUME**

*Quantification de l'incertitude en l' approvisionnement net du bassin du Lac Erie et du Lac Michigan*
*Cet article analyse l'incertitude dans les estimations par le NOAA Great Lakes Environmental Research Laboratory (GLERL) de deux éléments clés en l'approvisionnement net du bassin des Grands Lacs (NBS), la précipitation au-dessus du lac et la contribution des rivières. GLERL estime la précipitation au-dessus du lac à l'aide de interpolation des données de jauge ave les polygones de Thiessen. Cette méthode est affectée par le manque de jauges sur les lacs eux-mêmes et la rareté des mesureurs dans certaines parties du bassin du drainage. Récemment, les estimations basées sur la fusion des données de radar météorologique et des données de jauge s'est avéré pour être une meilleure alternative aux méthodes traditionnelles d'estimation pour certains des Grands Lacs. Nous comparée les modèles traditionnelles de la précipitation au-dessus du lac avec les estimations basées sur le radar pour 2002-2007, calculée la distribution d'erreur relative et utilisé il pour la modélisation de l'incertitude dans les estimations historique de la précipitation au-dessus du lac, indiquant que l'incertitude varient autour de ±20 % pour le lac Érié à autour de ± 23 % pour le lac Michigan.*
*Le ruissellement affluent doit être le composant moins incertain dans les estimations de la NBS des Grands Lacs, mais la surveillance seulement partielle de la plupart des bassins exige l'examen. Nous avons analysé l'algorithme utilisé par GLERL pour évaluer le flux affluent, modélisé l'incertitude dans l'estimation de ruissellementde chaque sous-bassin hydrographique et évalué l'incertitude globale dans les eaux de ruissellement affluent à un lac à l'aide d'une technique de Monte Carlo. Les résultats indiquent que l'incertitude est environ 15 % pour les deux lacs. Enfin, nous avons utilisé ces résultats et les estimations sur l'incertitude d'évaporation rapporté dans la littérature pour calculer l'incertitude globale dans la NBS.*

## 1. INTRODUCTION

The Laurentian Great Lakes in North America (Figure 1) are the greatest single freshwater resource on earth, accounting for about 20% of the world's total freshwater and providing drinking water to 40 million U.S. and Canadian citizens, water for hydro and thermal power generation, and support for important shipping, fishing, and recreation industries.
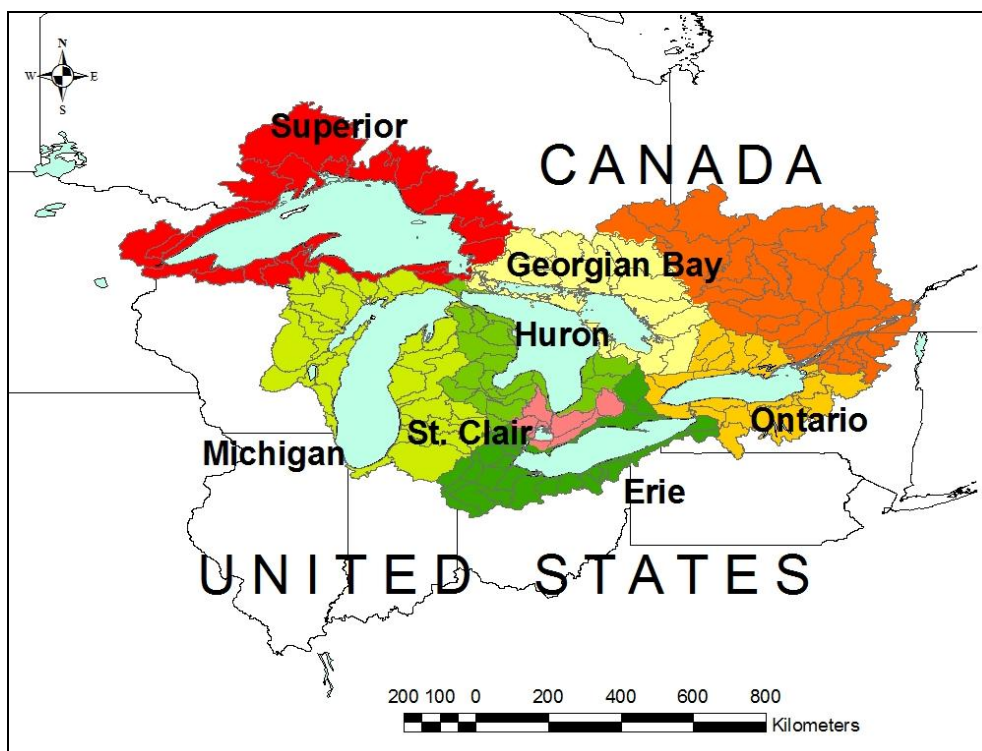


**Figure 1**: The Great Lakes drainage basins.

It is therefore understandable that considerable effort has been invested in studying and managing the Great Lakes, starting with their water balance. However, the huge sizes of the overall basin which spans 1,440 km in the East-West direction and 920 km in the North-South direction for a total surface of 755,000

km$^2$ (Figure 1), and of the lakes themselves (e.g., 82,400 km$^2$ for Superior) make even this apparently trivial task quite problematic ([1]).

An essential element for a correct assessment of the Great Lakes water-balance is quantifying the net basin supply (NBS) of each single lake, which is the net amount of water entering each lake, not counting the supply of water from upstream lakes and the loss to downstream lakes. NBS are used for calibrating hydrologic models, generating operational forecasts, and detecting long term hydroclimatic trends. Therefore, improving NBS estimates and understanding their uncertainty is critical for the management of the Great Lakes. NBS can be computed in two ways: by the residual method, which uses measured lake levels and channel input and outputs, or by the component method, which considers the actual inputs (over-lake precipitation, runoff from the drainage basin, and net groundwater flux) and outputs (evaporation) to/from each lake, but not the connecting-channel flows or change in storage ([1]), resulting in:

$$NBS = P + R - E + G \tag{1}$$

where $P$ is the over-lake precipitation, $R$ is the river and stream runoff to a Great Lake, $E$ is the evaporation from the lake surface, and $G$ is the net groundwater flux into a Great Lake.
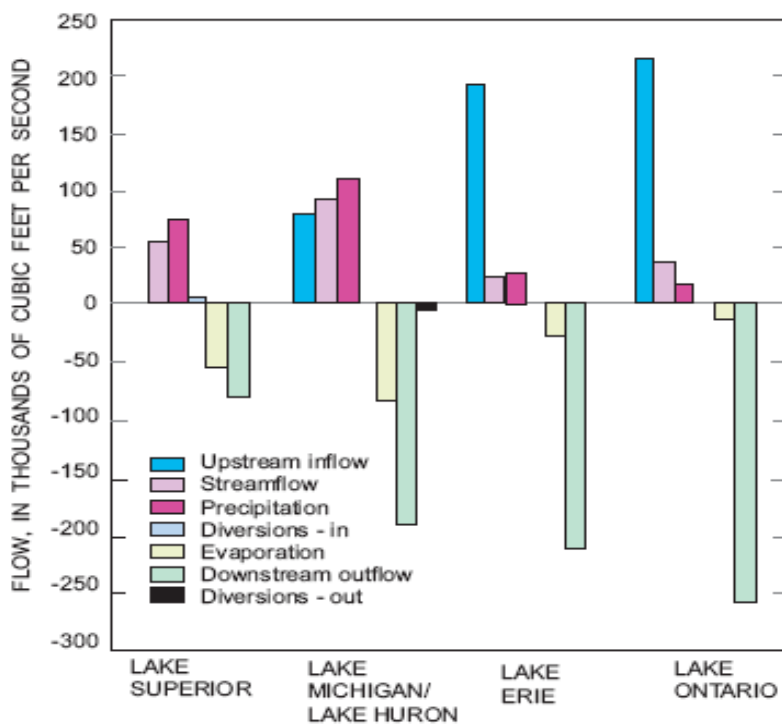


**Figure 2**: Components of Great Lakes inflows and outflows. Positive values denote inflows, and negative values denote outflows for each lake ([1]).

Net groundwater inflow is generally considered negligible ([1]; N.G. Grannemann, personal communication), while the other three components are similar in magnitude because the lake surface is a relevant part of the Great Lake's basin (Figure 2),

Of these three components of the Great Lakes NBS, river runoff is potentially the least uncertain, since river discharge measurements can be done in places easily accessible and are quite accurate. For example, river discharge data collected by the United States Geological Survey (USGS) in cooperation with local public and private agencies estimates are rated as "excellent," "good," or "fair." These terms signify confidence that 95 percent of the reported daily streamflows are within 5 percent, 10 percent, or 15 percent of their true values, respectively. Sometimes a "poor" rating is given, indicating that 95 percent of the reported daily streamflows are believed to be within an unspecified distance larger than 15 percent from their true value ([1], [2]). River discharge measurements in Canada are performed by the Water Survey of Canada (WSC) and likely have similar levels of accuracy. However, because their scarce population does not justify considerable expenditures for flood protection or because their small size, American and Canadian agencies do not maintain river discharge gages in large parts of the basin (Table 1).

| Superior | Michigan | Huron | Georgian Bay | St.Clair | Erie | Ontario |
|----------|----------|-------|--------------|----------|------|---------|
| 34% | 24% | 34% | 29% | 50% | 22% | 25% |

**Table 1**: Percentage of the Great Lakes Basin area that is ungaged (modified from [3])

Runoff from the ungaged areas needs to be estimated, thus increasing the overall uncertainty of the river runoff estimates. The magnitude of such uncertainty is dependent on the extent of the ungaged area, the season, and the method used to carry out the estimate. Table 1 refers to the situation in the late 1980s-early 1990s, but during earlier periods, the ungaged fraction of the watershed was normally larger (e.g., in 1960, just 36% of Georgian Bay basin was monitored). In their 2004 review, Neff and Nicholas states that an exhaustive analysis of uncertainty in river runoff is impossible because it involves hundreds of gages with highly variable flows and accuracy, and the reliability of current methods used to estimate runoff in ungaged areas is unknown. However, based on their experience they evaluate the uncertainty in monthly runoff estimates to be between 15 and 35% (Table 2). This figure is close to the estimated uncertainty in the other NBS components, which were also based on expert opinion and not on a formal analysis.

| Runoff | Overlake precipitation | Lake evaporation | Connecting channels |
|--------|------------------------|------------------|---------------------|
| 15 – 35% | 15 – 45% | 10 – 35% | 5 – 15% |

**Table 2**: Uncertainty ranges for net basin supply ( [1])

Precipitation falling directly on the surface of a water body is a negligible component in the water balance of most lakes, because most catchment basins are several times larger than the lake surface. Thus, drainage basins collect much more precipitation than lake surfaces and deliver it to the lake in form of tributary runoff. However, the Great Lakes cover an area of approximately 244,000 km$^2$, about one-third of the total area of their drainage basin. Consequently, as shown in Figure 2, the amount of water that falls directly on the surface of the Great Lakes is comparable in magnitude to the water that enters the lakes through runoff and to the evaporation losses.

Quantification of over-lake precipitation suffers of two uncertainty components: instrumentation error and monitoring network's insufficiency and spatial unhomogeneity. Precipitation gages tend to underestimate precipitation largely because of wind-induced turbulence near the opening of the gage, adhesion of moisture to the internal walls of the gage, tipping bucket undercatchment during high-intensity events, evaporative losses in heated gages ([4]), reduced sensitivity of weighting gages ([5]), and occasional mechanical and electrical failures. Bias is considered normally of less than 5% ([5]), but can reach up to 40 percent for single months ([4]). Since, accuracy is generally inversely proportional to temperature and wind velocities, and rain is more accurately gaged than snowfall, Legates and DeLiberty ([6]) estimate that in U.S. part of the Great Lakes Basin gages underestimates precipitation in average by 6-7% during summer and 20-28 % during winter. Another relevant source of uncertainty is represented by the inhomogeneous distribution of rain gages over the Great Lakes basin and especially by the lack of stations on peninsulas or on islands, which would represent over-lake conditions better than gages on the mainland (Figure 3).

Because of this, estimates by traditional techniques, such as Thiessen Polygons and inverse square distance weighting, are not very reliable, mainly because rain and snow vary strongly in time and space, while some parts of the lake surface 70-80 km away from the nearest rain gage. Further, on-shore rain gages are not always representative of lake precipitation, since precipitation mechanisms over the surface of large lakes are often different from those over land. The most obvious of such distortions is the lake-effect snow, which takes place in late fall and winter, when cold air masses moving across the Great Lakes from north and west entrain the moist and relatively warm air floating over the lake surface. Once these air masses reach the colder southern and eastern lake shores, moisture condensates and precipitates in larger quantities than over the lake surface ([1]). Differences in the precipitation mechanisms are present also during spring and summer (e.g., [7]; [8] for Lake Victoria). Such factors make over-lake precipitation estimates so uncertain that

GLERL prefers to use the average precipitation over the draining watershed as a surrogate for the real over-lake precipitation ([9]).
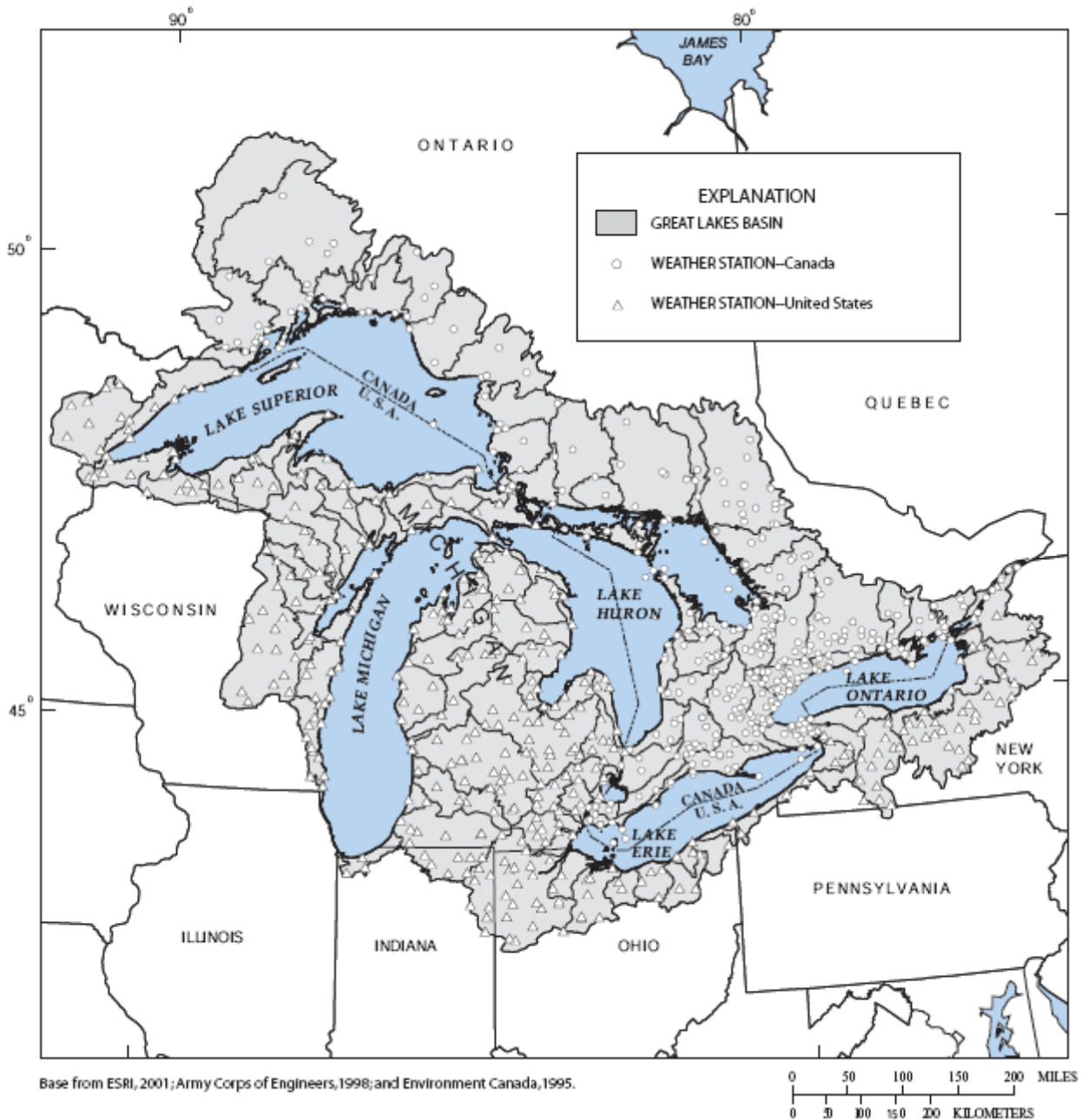


**Figure 3**: Weather stations in the Great Lakes Basin. After [1].

Given the lack of off-shore gages, it is difficult to find extensive studies on over-lake precipitation uncertainty. Based on talks with experts in the Great Lakes, Neff and Nicholas [1] estimate that uncertainty is "likely to be more than 15 percent and could be as high as 45 percent during some winters" (Table 2).

In this paper we analyze the uncertainty in the over-lake precipitation and runoff components of the NBS estimates computed by the National Oceanic and Atmospheric Administration's Great Lakes Environmental Research Laboratory (GLERL), which are widely used by scientists and managers working on the Great Lakes hydrology. Such study was motivated by the necessity of understanding the uncertainty in the NBS estimates to assess whether the conveyance of the St. Clair River, connecting Lake Huron to Lake St. Clair, and of the Detroit River, connecting the latter to Lake Erie had changed following the 1960 dredging ([10]). Further, new methodology to better estimates NBS components are likely to be based on new sensors and, therefore, cannot be applied to the past. Thus, understanding the uncertainty in GLERL NBS will still be important for future long term studies. Section 2 of this paper describes the methodology

employed to assess the uncertainty in GLERL's NBS estimates. Section 3 presents and discusses the results obtained. Finally, Section 4 outlines the conclusions that can be drawn from this paper.

## 2. METHODS

### 2.1 GLERL approach to estimating Monthly River Runoff

GLERL has subdivided the Great Lakes basin into 121 sub-watersheds (Figure 1) grouped into seven lake basins (Superior, Michigan, Huron, Georgian Bay, St. Claire, Erie, and Ontario). Some of these watersheds actually coincide with a single river system: for example Huron_10 coincides with the Saginaw River and Erie_6 with the Maumee River. In other cases, a GLERL watershed encompasses two or more physically separated river systems: for example, Huron_8 merges together the Rifle and the AuGres rivers; Erie_15 includes the Conneaut and Girard rivers. This was done to facilitate the modeling of the smaller river basins.

Watershed runoff is estimated by using streamflow records from major rivers, available from the USGS for U.S. streams and the WSC for Canadian streams. Daily runoff values provided by these agencies are summed for each watershed within a lake basin. Daily watershed runoff estimates are computed by summing all daily non-overlapping streamflow gage values in the watershed, dividing it by the sum of the drainage areas, and finally multiplying this discharge intensity for the total area of the watershed, to extrapolate for residual ungaged areas ([9]). For example, in the hypothetic watershed represented in Figure 4, the runoff is computed according to equation (2) when gage *a* has valid data, and according to equation (3) when gage *a* does not have valid data, but gage *b* has valid data.
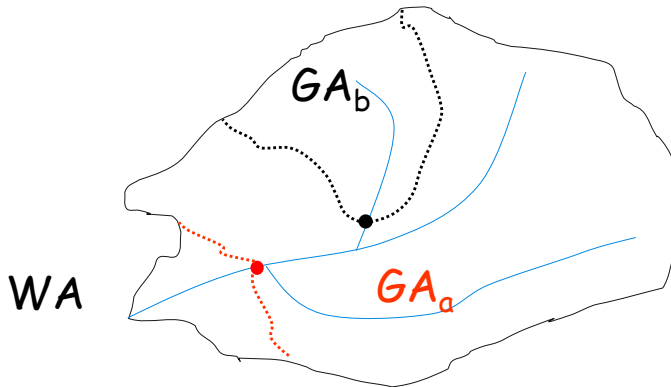


**Figure 4**: Example of GLERL gaged watershed.

$$R_A{}' = \frac{WA}{GA_a} R_a \qquad (2)$$

$$R_A{}'' = \frac{WA}{GA_b} R_b \qquad (3)$$

where, $R_A$ is Runoff from watershed A, $R_a$ is the runoff measured at gage *a*, $R_b$ is the runoff measured at gage *b*, $GA_a$ is the drainage area at gage *a*, *and* $GA_b$ is the drainage area at gage *b*.

Daily lake basin runoff estimates are computed in a similar manner. First, all outputs of $N_G$ watersheds for which there are available observations are estimated with the method described earlier. They are then summed up ($R_{Gaged}$) and divided by the total area of the gaged watersheds ($WA_{Gaged}$). Then, this "discharge per unit area" is multiplied for the total area of the ungaged watersheds to obtain the total flow from the ungaged watersheds ($R_{Ungaged}$) as shown in (4)

$$R_{Ungaged} = (WA_{u1} + WA_{u2} + ... + WA_{uN}) \frac{R_{Gaged}}{WA_{Gaged}} \qquad (4)$$

where $R_{Ungaged}$ is the runoff from all ungaged watersheds, $R_{Gaged}$ is the runoff from all ungaged watersheds, $WA_{u1}, WA_{u2}, ..., WA_{uN}$ is the area of the ungaged watersheds *1*, 2, ..., *N*, and $WA_{Gaged}$ is the total area of gaged watersheds.

Daily flows from the gaged and ungaged watersheds are summed up to obtain the daily flow to the lake ( $R$ ).

$$R = R_{Gaged} + R_{Ungaged}$$

$$R_{Gaged} = \frac{WA_1}{GA_1} R_1 + \frac{WA_2}{GA_2} R_2 + \ldots + \frac{WA_{NG}}{GA_{NG}} R_{NG}$$

$$R_{Ungaged} = (WA_{u1} + WA_{u2} + \ldots + WA_{uN}) \frac{R_{Gaged}}{WA_{Gagged}} \tag{5}$$

$$R = \frac{WA_1}{GA_1} R_1 + \frac{WA_2}{GA_2} R_2 + \ldots + \frac{WA_{NG}}{GA_{NG}} R_{NG} + \frac{\frac{WA_1}{GA_1} R_1 + \frac{WA_2}{GA_2} R_2 + \ldots + \frac{WA_{NG}}{GA_{NG}} R_{NG}}{WA_1 + WA_2 + \ldots + WA_{NG}} (WA_{u1} + WA_{u2} + \ldots + WA_{uN})$$

Monthly basin runoff is computed by simply summing the daily basin runoff estimates for all days in each month and dividing it by the lake surface ([9]). Complete years of historical daily runoff data begin in 1908 (Superior), 1910 (Michigan), 1915 (Huron), 1935 (St. Clair), 1914 (Erie), and 1916 (Ontario), while the ungaged fraction is summarized in Table 1 ([3]).

## 2.2 Evaluating the Uncertainty in GLERL Estimates of Monthly River Runoff

Uncertainty in the GLERL estimates of river runoff has three sources:
1)    Errors in the USGS and WSC estimates of discharge at gage locations (*Measurement Errors*);
2)    Errors caused by extending the discharge per unit area measured at the most downstream gages to an entire watershed (*Same-basin Errors*); and
3)    Errors caused by applying the basin-wide average discharge per unit area in the gaged watersheds to the ungaged watershed in the lake basin (*Inter-basin Errors*).
It is practically impossible to evaluate the total uncertainty in river runoff estimates in analytical form because of the large number of the error sources (7 to 27 watersheds per lake basin) and the fact that error distributions may be different for different sources of error. Thus, this research adopts a Monte Carlo analysis approach, where the uncertainty of each source in (7) is simulated by randomly generating an ensemble of alternative equiprobable discharges.

### 2.2.1   Measurement Errors

As mentioned in Section 1, USGS rates its streamflow gages as "excellent," "good," or "fair", signifying confidence that 95 percent of the reported daily streamflows are within 5 percent, 10 percent, or 15 percent of their true values, respectively. A "poor" rating is given to gages where 95 percent of the reported daily streamflows are believed to be within a bound that is more than 15 percent from their true ([1],[2]). The rating of a stream gage can change from year to year, reflecting changes in gaging devices, rating curves, and stream conditions, and also from season to season, reflecting changes in ice, flow, and wind conditions. USGS does not give further information on the statistical distribution of the measurement error, but it is likely that large errors have less probability than small errors, and, since there is no mention of bias in gage measurements, it is likely that the mean of the error distribution is zero.

In this research we used the middle USGS category of "good" for all American and Canadian gages (it was assumed MCS gages had a similar level of accuracy). Further, it is supposed that the relative measurement error $\varepsilon$ is described by a Normal distribution with zero mean and standard deviation equal to 0.05105.

### 2.2.2   Same Basin Errors

In the majority of Great Lakes watersheds, USGS and WSC river gages are located somewhat upstream of the watershed's mouth. Thus, (3) or (4) are just an approximation of the actual discharge, the goodness of which is a function of several factors, including the fraction of the watershed that is gaged, climatic characteristics of the watershed, season, soil structure, presence of reservoirs and abstractions, and land use.

To correctly represent the same-basin error, we would need a large number of measurements taken at watershed mouths and compare them with the estimates obtained by prorating upstream gage data. Unfortunately, the number of gages close to watershed mouths is small (only 29 watersheds out of 105 have a

gaged fraction above 85%), not allowing a good representation of the same-basin error distribution. The strategy adopted in this research is computing the relative difference $\phi$ between the discharge per unit area measured at several gages in a basin and the discharge per unit area obtained by prorating the measurements at upstream gages. Using the hypothetical example of Figure 4, that is:

$$\phi = \frac{\dfrac{R_b}{GA_b} - \dfrac{R_a}{GA_a}}{\dfrac{R_a}{GA_a}} \tag{6}$$

where $R_a$ is the runoff measured at gage $a$, $GA_a$ is the drainage area at gage $a$, $R_b$ is the runoff measured at gage $b$, $GA_b$ is the drainage area at gage $b$.

Basins where reservoir operations dictate river discharge (such as the Fox-Wolf River basin in Wisconsin) were not included, since this is not a situation normally found at the mouths of watersheds discharging in the Great Lakes and causes abnormally high errors. We eliminated from the pool also gages that presented errors much larger than those typically affecting basins of similar size and gaged fraction, and gages constantly underestimating or overestimating the flow, obtaining the gage distribution of Figure 5. Overall, this approach allowed using a large number of samples for generating the same-basin error distributions, thus accounting for the influence of several factors.
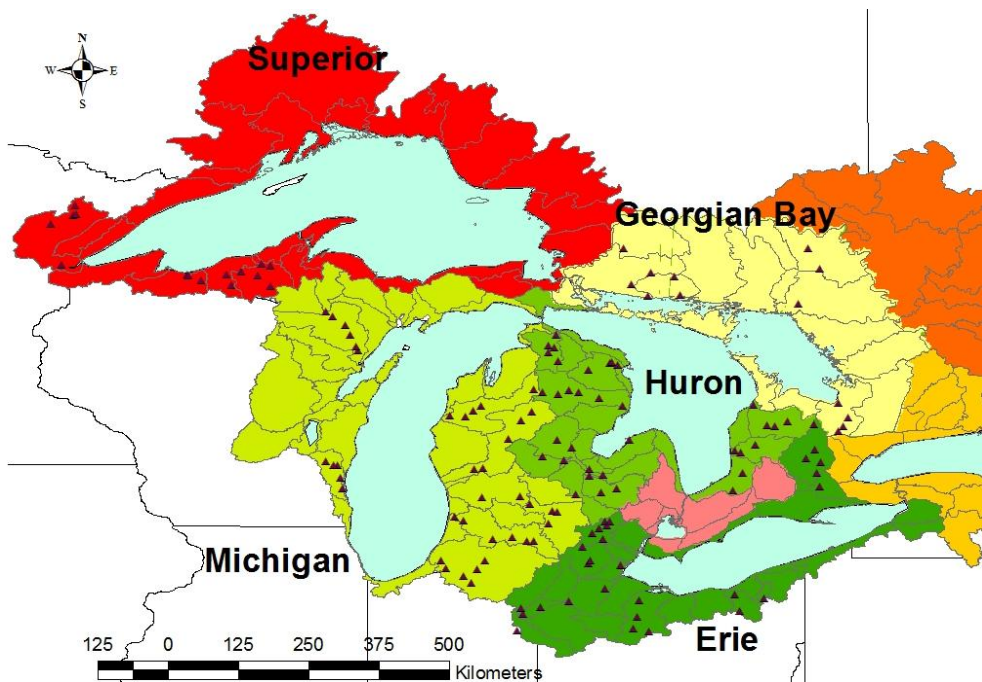


**Figure 5**: Gages used for same-basin error estimation.

We first subdivided the distribution of relative errors by Great Lake, to partially account for similarity in land use and climate. It is clear, however, that because the Great Lakes' size, watersheds in the same Great Lake basin may still present large differences in land use (e.g., the agricultural Maumee River vs. the more urbanized Cuyahoga River) and climate (e.g., watersheds in the eastern sides of the lakes are affected by lake-effect snow more than watersheds in the western sides of the lakes). We then subdivided the distribution of relative errors for each lake in six classes, according to the gaged fraction of the watershed (0—20%, 20—40%, 40—60%, 60—80%, 80—95%, and 95—100%) and assigned a different function to each class. In the example of Figure 4, the gaged fraction of the watershed is given by $GA_b/GA_a$. Because there were no coupled gages in the 95—100% class that could supply experimental data, and because the error in the 80—95% was so small, we assumed that the error in this class is limited to the measurement error. Finally, we included the seasonal effect on prorating the measured discharge per unit area to the entire watershed by creating a different error function for each month.

Working with sample distributions is quite difficult, because it requires keeping in memory thousands of data that need to be interpolated to create the cumulative probability of each error class in the Monte Carlo ensemble generation ([8], [11]). Further, this approach is more susceptible to the effect of the sample randomness and does not allow a good analysis of how error distribution changes with gaged watershed fraction and season. So the approach taken here was to fit the sample distributions obtained for different lakes, months, and gaged fraction classes with a set of probability density functions using the program Palisade Corporation Risk 4.5.3 for Excel. The probability density functions tested in this research are the following:

1) Log-normal, with lower boundary set to -1;
2) Log-logistic, with lower boundary set to -1;
3) Pearson 5, with lower boundary set to -1;
4) Inverse Gaussian, with lower boundary set to -1;
5) Gamma, with lower boundary set to -1;
6) Weibull, with lower boundary set to -1;
7) Normal;
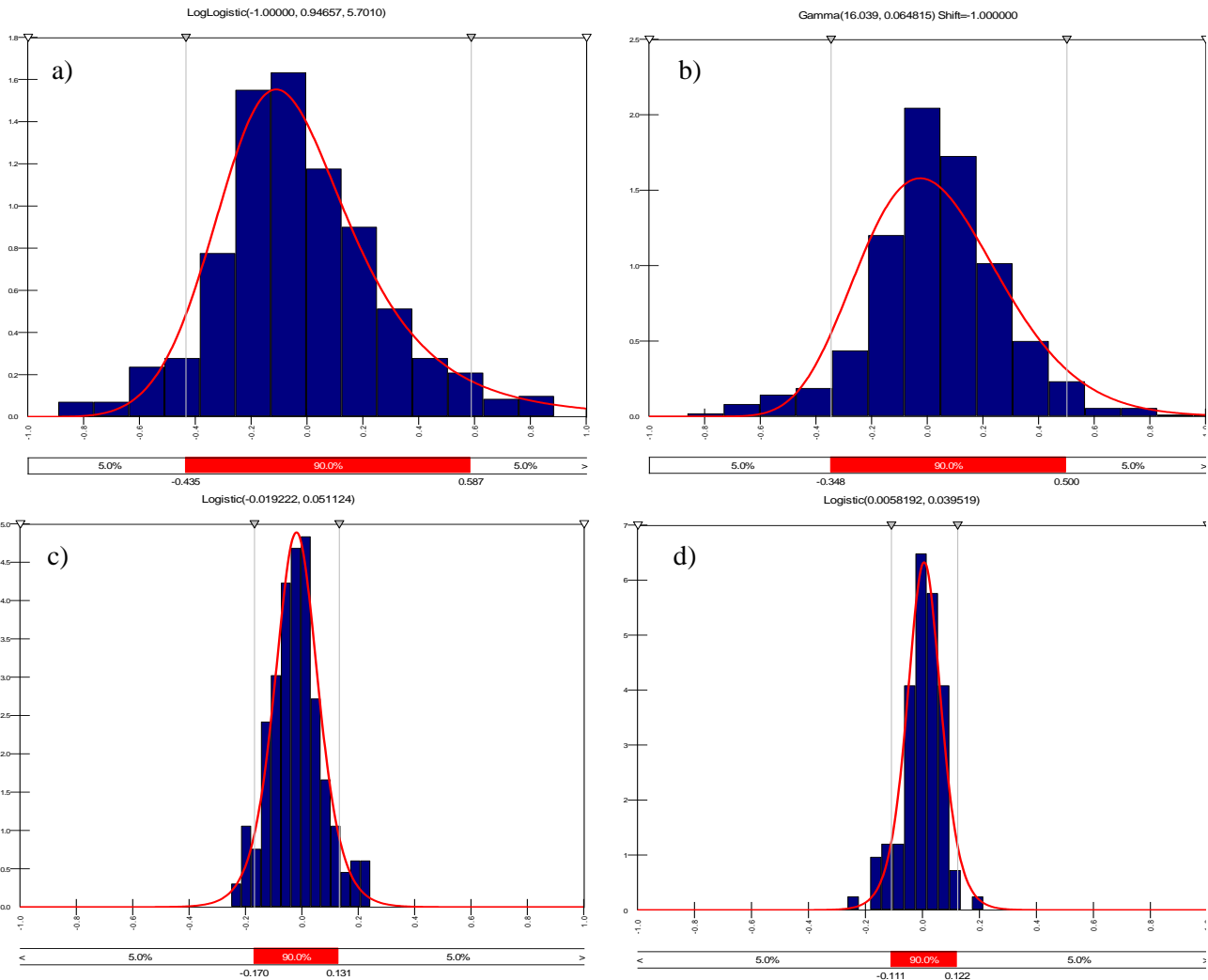8) Logistic;
9) Extreme value;



**Figure 6**: Same-basin error sample distribution and analytical fit for Lake Michigan in March as function of gaged watershed fraction: a) 0—20%; b) 20—40%; c) 60—80%; and d) 80—95%.

Probability distributions 1) through 6) were chosen from the consideration that the relative error by definition has a lower boundary equal to -1 (see Figure 6 and equation 8). Distributions 7) to 9) were chosed because the error associated with watershed largely gaged has very narrow sample distributions around 0 (Figure 6-c, and Figure 6-d). After comparing the goodness of fit for the error distributions in the six lakes, we found that

the Log-logistic distribution best fitted the sample distributions for 0—20% gaged fraction, followed by the Gamma distribution; the Gamma distribution was the best fit for the 20—40% gaged fraction, followed by the Log-logistic; while for the 40—60%, 60—80%, and 80—95% gaged fractions, the best fit was given by the Logistic distribution followed by the Normal (Figure 6).

*2.2.3 Inter-basin Errors*

To estimate the error introduced by the GLERL computation of discharge from ungaged basins, we compared the discharge per unit area from each gaged watershed with the average discharge per unit area computed with the GLERL method using only the remaining gaged watersheds in the same lake basin (Jackknife method). That is:

$$\varphi_k = \frac{\dfrac{\dfrac{R_1}{GA_1}WA_1 + \dfrac{R_2}{GA_2}WA_2 + \ldots + \dfrac{R_{k-1}}{GA_{k-1}}WA_{k-1} + \dfrac{R_{k+1}}{GA_{k+1}}WA_{k+1} + \ldots + \dfrac{R_{NG}}{GA_{NG}}WA_{NG}}{WA_1 + WA_2 + \ldots + WA_{k-1} + WA_{k+1} + \ldots + WA_{NG}} - \dfrac{R_k}{GA_k}}{\dfrac{R_k}{GA_k}} \qquad (7)$$

where $\varphi_k$ is the Inter-basin error for basin $k$, $R_i$ is the unoff measured at the gage used for watershed $i$, $GA_i$ is the drainage area at gage used for watershed $i$, $WA_i$ is the area of watershed i, and $NG$ is the number of gaged watersheds in the lake basin.

Figure 7 shows that the inter-basin errors easily reach 200-250% in Lake Michigan and even 600-800% in Lake Erie. The latter result is due to the fact that in the Lake Erie basin there are a single very large basin, the Maumee River (17,540 km$^2$), two medium-large basins, the Sandusky (5,000 km$^2$) and Grand River –Ontario (7,100 km$^2$), and a large number of small watersheds. This creates an imbalance in discharge dynamics between these watersheds. For example, during summer many small watersheds often become almost dry, while the Maumee River always preserves at least a small flow. Similarly, small watersheds can respond very quickly to the isolated convective thunderstorms characterizing summer precipitation, while the larger Maumee would not be altered significantly. Further, river discharge in most Lake Erie watersheds is dominated by runoff because of the poor soil permeability and intense tile-drainage application, thus featuring very "spiky" and inhomogeneous hydrographs which exacerbate the situation ([12], [13]).
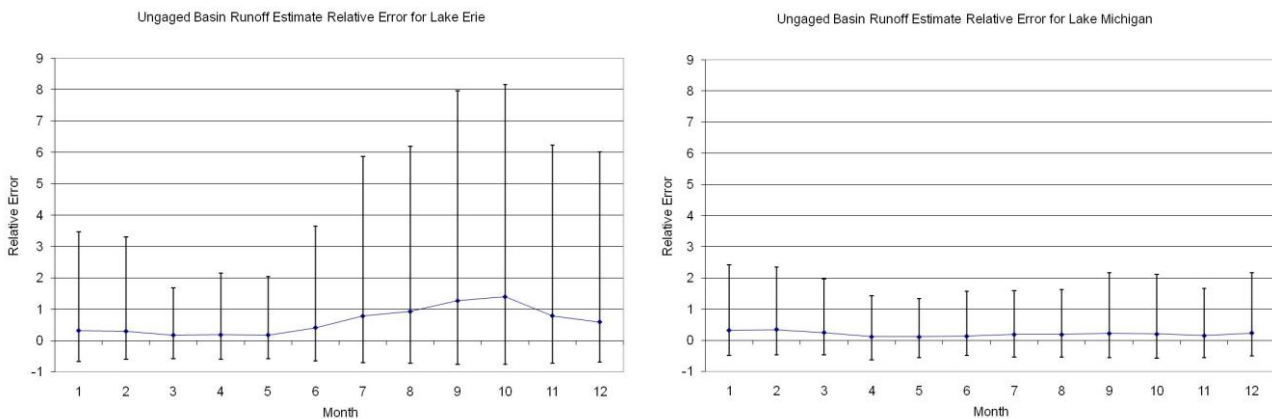


**Figure 7**: Inter-basin error distributions for the Upper Great Lakes. Average of the distribution (marker), 2.5%-tile (lower bracket), and 97.5%-tile (upper bracket).

*2.2.3 Total Uncertainty in the River Runoff to a Lake*

Total uncertainty of runoff estimates to a lake is estimated using a Monte Carlo simulation ([14]), which computes an ensemble of alternative equiprobable total monthly runoff to a lake estimates using the GLERL nominal estimates and the same-basin and inter-basin relative error distributions. That, is:

$$R'_j = \frac{WA_1}{GA_1} R_1 \frac{1}{1+\phi_{1,j}} + \frac{WA_2}{GA_2} R_2 \frac{1}{1+\phi_{2,j}} + ... + \frac{WA_{NG}}{GA_{NG}} R_{NG} \frac{1}{1+\phi_{NG,j}} +$$

$$\frac{\frac{WA_1}{GA_1} R_1 + \frac{WA_2}{GA_2} R_2 + ... + \frac{WA_{NG}}{GA_{NG}} R_{NG}}{WA_1 + WA_2 + ... + WA_{NG}} (WA_{u1} \frac{1}{1+\varphi_{1,j}} + WA_{u2} \frac{1}{1+\varphi_{2,j}} + ... + WA_{uN} \frac{1}{1+\varphi_{uN,j}}) \qquad (11)$$

where $R'_j$ is the total runoff to a lake estimate, alternative $j$, $R_k$ is the runoff measured at the gage used for watershed $k$, $GA_k$ is the drainage area at the gage used for watershed $k$, $WA_k$ is the total area of watershed $k$, $NG$ is the number of gaged watersheds in the lake basin, $\phi_{k,j}$ is the same-basin error for basin $k$, alternative $j$, $\varphi_{k,j}$ is the inter-basin error for basin $k$, alternative $j$, and $uN$ is the number of ungaged basins.

Monte Carlo simulation was performed by a Mathworks Matlab program using the statistical toolbox and an ensemble size of 50,000 elements.

## 2.3 GLERL approach to estimating Monthly Over-lake Precipitation

GLERL subdivides each lake basin into a regular grid of 1 km x 1 km cells. Precipitation is computed for each pixel using Thiessen polygon interpolation of all available daily data from stations in the basin or within approximately 0 - 30 km from the basin, depending upon station density near the edge of the basin. Over-land precipitation is computed considering only precipitation falling over land pixels, including islands and smaller inner lakes. Over-lake precipitation is computed considering only open-water pixels. Finally, over-basin precipitation is computed considering all pixels ([15], [16]). Mean daily areal precipitation is cumulated for each lake and basin to obtain monthly values ([9]). Because of the small number of off-shore gages, over-lake precipitation is dominated by the few gages along lakeshores, which are probably not representative of the over-lake precipitation (e.g., lake-snow effect). Over-land precipitation, on the other hand, is the result of the contributions of many more stations, thus probably better reflecting the climatological precipitation. Consequently, GLERL estimates the direct precipitation component of the NBS estimates by multiplying the average over-land precipitation times the lake area.

## 2.4 Evaluating the Uncertainty in GLERL Estimates of Monthly Over-lake Precipitation

The lack of offshore rain gages makes a direct evaluation of over-lake precipitation estimate uncertainty quite challenging. The strategy adopted for this work is improving estimates of precipitation for a limited period (2002-2007) by integrating rain gages and additional information, such as weather radar data. The comparison between these better products and coincident GLERL estimates would be used to assess reliability and uncertainty band affecting the latter over the entire 1948-2008 period.

*2.4.1 Merging NOAA NCEP Multi-sensor Precipitation Estimate data and Gage Data*

Radar based precipitation estimates provide precipitation data at high spatial and temporal resolution over extended areas, making them very useful in storm and flood forecasting. However, their application for quantitative hydrology is more questionable due to several limitations including hardware calibration, uncertain reflectivity-rainfall (Z-R) relationships, ground clutter, bright band contamination, mountain blockage, anomalous propagation, range-dependent and seasonal biases ([17]). The NOAA National Centers for Environmental Prediction's Multisensor Precipitation Estimates Stage IV ([18]) optimally merge hourly radar or satellite information and hourly rain gage data to produce real-time hourly multisensor estimates at 4-km resolution. Estimates are first visually quality controlled by the U.S. National Weather Service at the regional level and then mosaicked into a national product. The overall procedure is designed to overcome most of the limitations described above.

Studies of the MPE data in the Great Lakes region ([19]-[21]) have revealed that MPE underestimates gage precipitation by about 10%, probably because (1) the number of gages used to de-bias the radar estimates is limited; and (2) the gages used to de-bias the radar data are hourly gages, which are known for underestimating cumulative precipitation ([22]). Further, correlation between gage data and MPE at the monthly level is generally strong over the American side of the watershed, and over Lake Erie, Lake Ontario, Lake St.Clair, and southern portions of Lake Huron and Lake Superior ([20], [21]).

Therefore, [20] and [21] showed that MPE can correctly identify areas of high precipitation and low precipitation for most of Lake Ontario, Lake Erie, Lake Saint Clair, Lake Michigan, and part of Lake Huron. However, before using it for long term quantitative purposes, its bias must be removed by merging it with daily gage data. [20] considered several ways to merge monthly gage and MPE data and showed that the best performances were obtained by a universal kriging algorithm using single-month MPE as external drift. However, even the much simpler MPE*, which is obtained at each pixel by multiplying the MPE value times the ratio of the average of gage data and the average of coincident MPE pixels at neighboring gages, outperforms gage-only methods such as inverse square distance weighting (IDW) and ordinary kriging (OK). We adopted the simpler MPE* to estimate monthly over-lake precipitation ([21]). However, differently from [20], who worked only on the smaller Lake Erie, we first needed to screen out all pixels for which MPE was not reliable, which were identified as the pixels having mean 2002-2007 summer precipitation below 1.8 mm/day and regions showing precipitation rates significantly lower than neighboring pixels and borders falling along straight lines or shorelines or coinciding with radar locations.

After this operation, the adjusted MPE* was computed for each valid MPE pixel and precipitation over the invalid MPE pixels was obtained by inverse square distance weighting interpolation of MPE* over valid pixels and gage data. Because the number of MPE pixels is much larger than the number of gages (80,500 vs. 1,030), a simple interpolation of the nearest pixel or gage would give a disproportionate weight to the MPE pixels. Thus, the precipitation over each invalid pixel, that we name IMPE*, is computed according to the following algorithm (Figure 8):

1) The region is subdivided into 4 quadrants centered on the pixel;
2) The five nearest valid MPE pixels or rain gages in each quadrant are identified; and
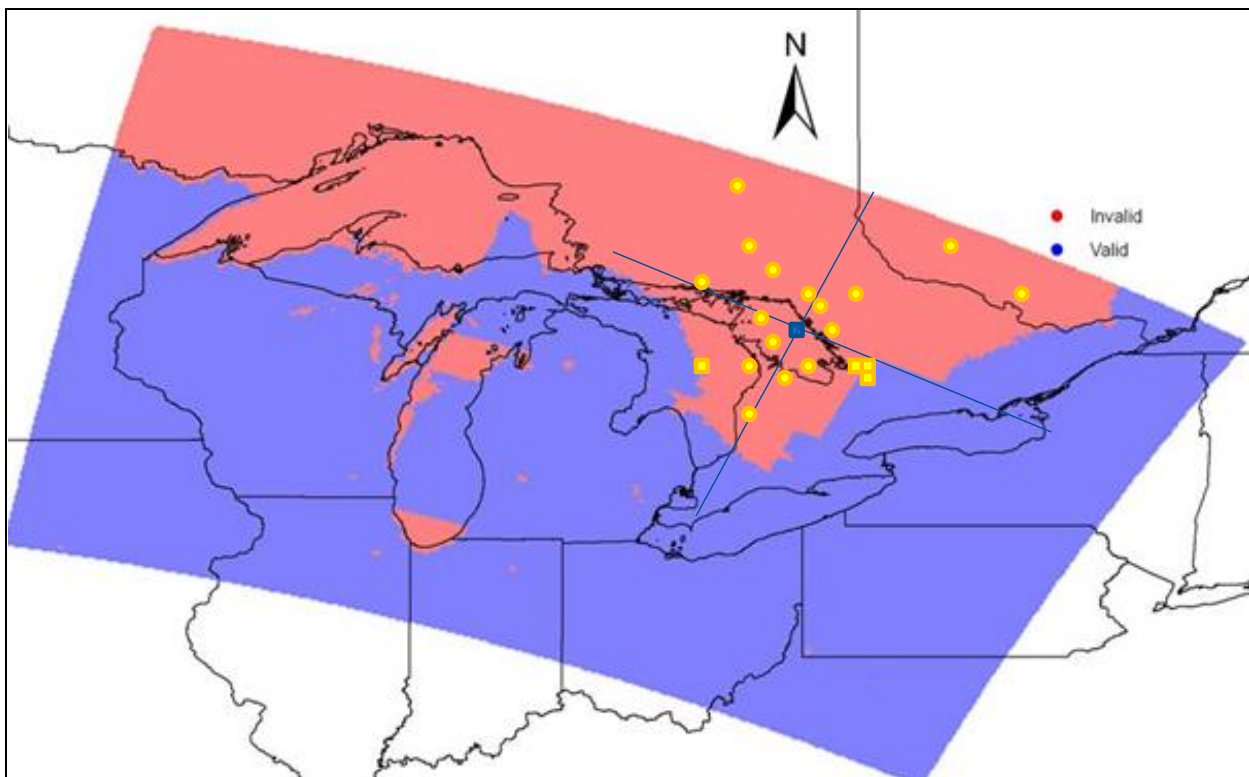3) Inverse square distance weighting interpolation of the selected pixels and gages is performed.



**Figure 8**: MPE* pixel interpolation. Precipitation over an invalid pixel (blue square) is determined with the following steps: 1) The region is subdivided into 4 quadrants centered on the pixel; 2) The five nearest valid MPE pixels (yellow squares) or rain gages (yellow circles) in each quadrant are identified; and 3) Inverse square distance weighting interpolation of the selected pixels and gages is performed.

By comparing IMPE* and GLERL Thiessen Polygon estimates, we showed that the latter estimates of overlake precipitation based on the interpolation of coastal rain gages over the lake (TP-Lake in Figure 9) are better than those obtained using the areal precipitation over the draining watershed (TP-Land in Figure 9) in all lakes ([20], [21]). Thus, TP-Lake was used as base for computing the over-lake precipitation in our uncertainty analysis.
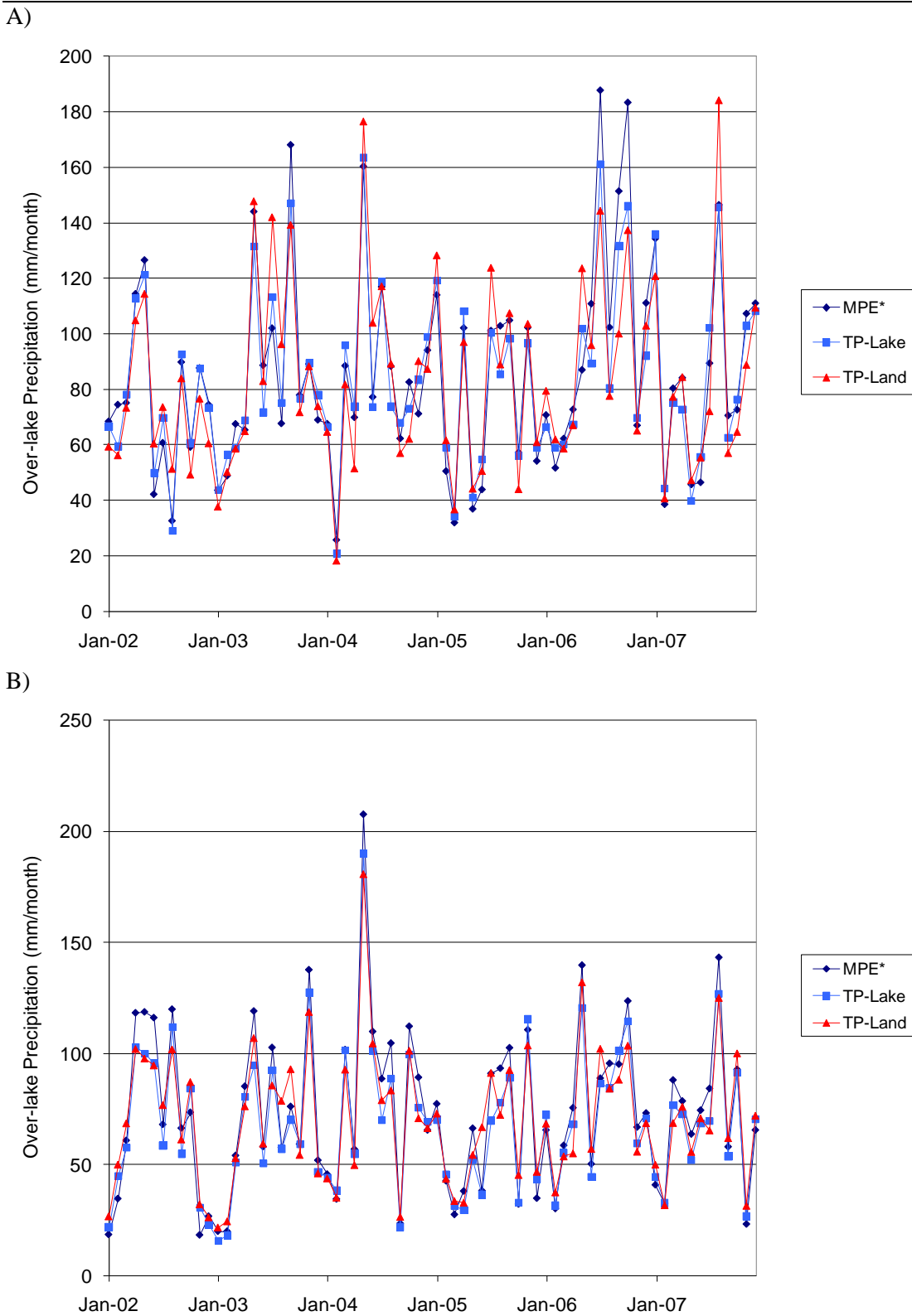
A)



B)



**Figure 9**: IMPE*, GLERL Thiessen Polygon, CaPA estimates. A) Lake Erie; B) Lake Michigan.

## 2.4.2 Computing Confidence Interval for GLERL Precipitation Estimates

The approach taken here is assuming that IMPE* is very close to the real precipitation and that the distribution of differences between this and TP-Lake can be used to model the uncertainty affecting TP-Lake for the period before 2002. That is:

$$\varepsilon_i = \frac{TP_{Lake_i} - IMPE^*_i}{IMPE^*_i} \tag{12}$$

where $TP_{Lake_i}$ is the GLERL TP-Lake for month i, $IMPE^*_i$ is the IMPE* for month i.

As mentioned earlier, the different regimes of precipitation (mostly frontal during fall and winter and mostly convective during spring and summer) and the different lake-shore interplay, likely make the error a function of the season. Thus, the computation of the $\varepsilon$ distribution has been separated into Summer (April-September) and Winter (October-March) months. Unfortunately, the limited number of elements (72) did not allow a partition of the year into a higher number of seasons [21].

As for the runoff, it was deemed more practical for the Monte Carlo simulation to model the error with analytical probability distributions instead of using the error sample distribution. Given the generally narrow distribution of the error, it was decided to test the following PDFs:

1) Normal;

2) Logistic;

3) Extreme value;

Although all three functions fitted the sample error distributions well (Logistic and Extreme value had a $\chi^2$ P Value above 0.25 for all eighteen distributions, while the Normal had a fit above 0.25 for fifteen), the Logistic performed slightly better (average P Value = 0.76 vs 0.70 of the Extreme value and 0.66 of the Normal) and was adopted for the Monte Carlo generation of the precipitation ensemble.
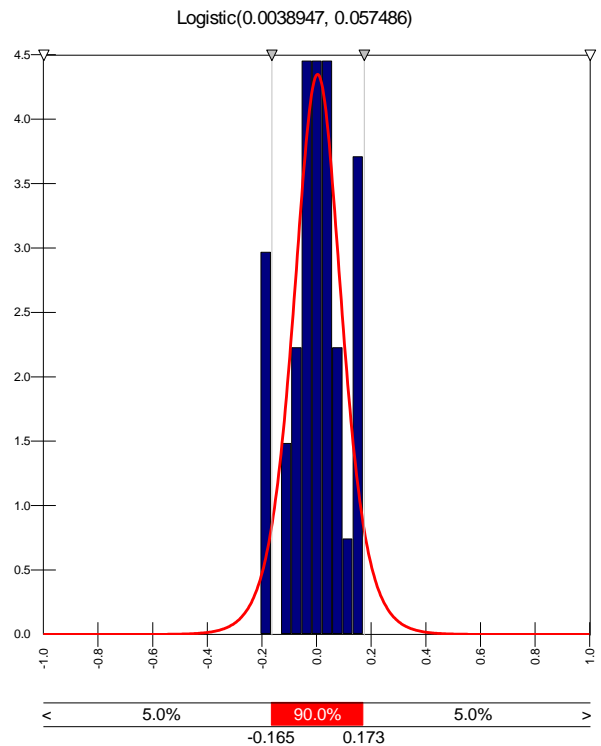
A)
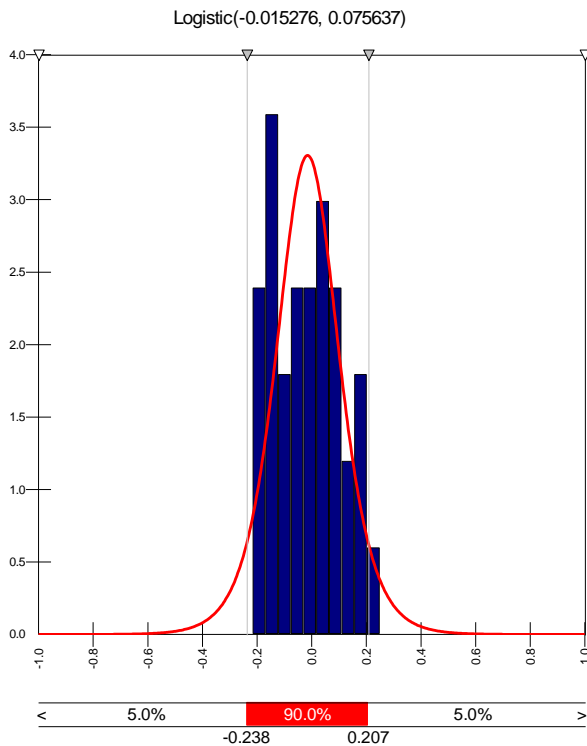
B)



**Figure 10**: Distribution of TP-Lake relative error for Lake Erie: A) April-September; B) October-March.

## 2.5 Evaluating the Uncertainty in GLERL Estimates of Lake Evaporation

Assuming that the net exchange lake-groundwater is negligible, the only missing component of the NBS is evaporation. Unfortunately, quantifying the uncertainty in evaporation estimates is quite difficult, since there are no measures of evaporation with which comparing the estimates. A field study is currently being carried out by [23] who directly measures evaporation over Lake Superior using an eddy covariance system.

Meteorological data collection began on June 1st, 2008 at Stannard Rock lighthouse, 24 miles southeast of Manitou Island. A second system was installed at Spectacle Reef, Lake Huron, in the summer 2009. Until this project produces a number of elements sufficient to compare them with GLERL estimates, the generation of the evaporation ensemble will be based on the estimates by [1], who characterize evaporation uncertainty as between 10% and 35%. Thus, the error distribution used in this work is a symmetric distribution (no bias was mentioned in [1] such that the P(Error<-35%)=2.5%; P(Error>35%)=2.5%. Since the Logistic error distribution looked better suited to model runoff and precipitation errors than the normal distribution, we used it for the evaporation as well. The evaporation uncertainty range was generated from GLERL data using a 50,000-element ensemble.

## 3. RESULTS AND DISCUSSION

### 3.1 Uncertainty in River Runoff to Lake Erie and Lake Michigan

Table 3 indicates that the Monte Carlo simulation monthly runoff is slightly higher than GLERL estimates. The smallest difference (<3%) is for Lake Erie, which is the best monitored basin, and highest for Lake Michigan (5.7%), despite being almost as well monitored. As mentioned earlier, the this result may be due to the fact that many tributaries of Lake Michigan have a strong groundwater component that reappears in the downstream reaches and to the lake-effect snow, which affects the eastern coastal areas. On the other hand, the good monitoring of the watershed reflects on a relatively small uncertainty in monthly runoff with 95% of estimates falling within 25% from the mean for both lakes.

| Lake | GLERL Average (mm/month) | Monte Carlo Average (mm/month) | Bias [mm/month] (%) | Correl. | Monte Carlo NBS 95% Conf. Interv. [mm/month] (%) |
|------|------|------|------|------|------|
| Erie | 70.4 | 72.12 | [-1.7] (-2.4%) | 0.9997 | [-6.3, +8.27]] (-9%,+22%) |
| Michigan | 54.72 | 58.03 | [-3.3] (-5.7%) | 0.9988 | [-5.2, +6.8] (-9%,+23%) |

**Table 3**: Differences between Monte Carlo Simulation and GLERL runoff estimates (1986-2005).

### 3.2 Uncertainty in Over-Lake Precipitation in Lake Erie and Lake Michigan

Table 4 indicates that the Monte Carlo simulation monthly over-lake is slightly higher than GLERL estimates. The difference (<3%) is smaller for Lake Erie, which is less wide, making the GLERL estimates more effectives, and higher for Lake Michigan (8.2%), which is wider. However, the uncertainty in monthly precipitation is higher for Lake Erie than for Lake Michigan (31% vs 25%). This uncertainty estimation address just the effect of the unequal distribution of rain gages, not the effect of rain gage undercatchment. Therefore, actual uncertainty is likely larger.

| Lake | GLERL Average (mm/month) | Monte Carlo Average (mm/month) | Bias [mm/month] (%) | Correl. | Monte Carlo NBS 95% Conf. Interv. [mm/month] (%) |
|------|------|------|------|------|------|
| Erie | 78.0 | 79.8 | [-1.8] (-2.3%) | 0.999 | [-16.9, +25.1]] (-21%,+31%) |
| Michigan | 68.9 | 75.0 | [-6.1] (-8.2%) | 0.992 | [-13.4, +19.1] (-18%,+25%) |

**Table 4**: Differences between Monte Carlo simulation and GLERL overlake precipitation estimates (1986-2005).

### 3.3 Uncertainty in Lake Erie and Lake Michigan's NBS

In order to estimates the NBS uncertainty, the errors in the NBS three components (evaporation, precipitation, and runoff) were assumed independent from each other. This allowed generating an ensemble

of 50,000 equiprobable alternative NBS estimates simply by combining each element of the Precipitation, Evaporation, and Runoff ensembles. Figure 11 shows two examples of such analysis. Lake Erie NBS estimates work quite well with no bias and relatively narrow uncertainty bands, because both TP-Lake and runoff had small bias and small uncertainty. On the other hand, Lake Superior NBS is affected by large uncertainty since precipitation and runoff uncertainty were higher and had larger biases, especially at the seasonal level.

Table 5 shows that, in the hypothesis that evaporation estimates are not biased, GLERL NBS is lower than Monte Carlo NBS between 3% and 14%, due mostly to runoff underestimation.

Monte Carlo analysis indicates that, assuming evaporation uncertainty is at 35%, the comprehensive NBS uncertainty varies between 30% and 60%. Such uncertainty is actually higher than the uncertainty on each single component and is magnified by the fact that evaporation is subtracted from precipitation and runoff, decreasing the value of the combination and increasing the impact of uncertainty. Indeed, despite being the most well monitored watershed, Lake Erie has one of the highest uncertainties, probably due to the fact it is the only lake for which evaporation is higher than both over-lake precipitation and runoff.

| Lake | GLERL Average NBS (mm/month) | Monte Carlo Average NBS (mm/month) | Bias [mm/month] (%) | Correl. | Monte Carlo NBS 95% Conf. Interv. [mm/month] (%) |
|---|---|---|---|---|---|
| Erie | 66.48 | 70.87 | [-4.4] (-6.2%) | 0.9999 | [-37, +43] (-52%,+60%) |
| Michigan | 63.02 | 72.93 | [-9.9] (-13.6%) | 0.9979 | [-28, +31] (-38%,+42%) |

**Table 5**: Average monthly GLERL NBS-Lake, Monte Carlo NBS best estimate, and 95% confidence intervals for 1988-2005.

## 4. CONCLUSIONS

This paper analyzed the algorithm employed by GLERL to estimate river runoff and over-lake precipitation to the Great Lakes to evaluate the uncertainty affecting GLERL NBS, which are widely used for analyzing the Great Lakes water balance. GLERL estimates total runoff discharge to a lake by extending the average discharge per unit area in the monitored portion of a watershed to the unmonitored portion of the watershed. The analysis of this component used data from 131 gages to compare downstream gage records and estimates based on upstream gage data and indicated that GLERL's approach is acceptable only when the gaged portion of the watershed is larger than 60%. For smaller fractions, uncertainty increases and relevant biases may appear. In watersheds with a strong groundwater component (like Lake Michigan) or with strong lake-effect snow (like eastern Lake Erie), simple extension of discharge intensity at upstream gages to the entire watersheds can lead to underestimating the input to the lakes. Even larger errors are committed when the average discharge per unit area from the gaged portion of a basin is used to estimate the discharge of completely ungaged watersheds. Although, this is a rare case after the 1970s (Table 6), it was much more frequent earlier. Given that 16% of the basin has a monitoring rate between 40-60% (Table 6), a relocation of the gages in this category further downstream or an expansion of the number of gages at downstream locations could serve the purpose of decreasing runoff uncertainty at a reasonable cost. Even less expensive could be improving the methodologies used for estimating discharge in partially gaged basins by taking advantage of distributed hydrologic model outputs and data from neighboring watersheds that are better monitored.

The approach taken for evaluating over-lake precipitation uncertainty was to integrate rain gage data and the NOAA NCEP MPE radar-based precipitation estimates to improve over-lake precipitation estimates for 2002-2007 ([20], [21]). We then analyzed the differences between this product and GLERL over-lake precipitation estimates obtained by Thiessen interpolation of shore gage data to determine the differences between the two. Results show that the bias is small for Lake Erie, but more relevant for Lake Michigan, and the uncertainty in monthly precipitation is limited to around 25-30% for both lakes. The addition of Canadian radars to MPE should improve over-lake precipitation estimates over Lake Huron and Georgian Bay. Until the end of 2007, the improvement for Lake Superior was minimal. Apparently, this is because the NWS radars in Duluth and Marquette continue to not adjusting or wrongly adjusting Radar retrievals over lake

surfaces. Further improvements to IMPE* will include an extension of MPE* to the area covered by the Canadian radars and a better interpolation of missing MPE pixels.
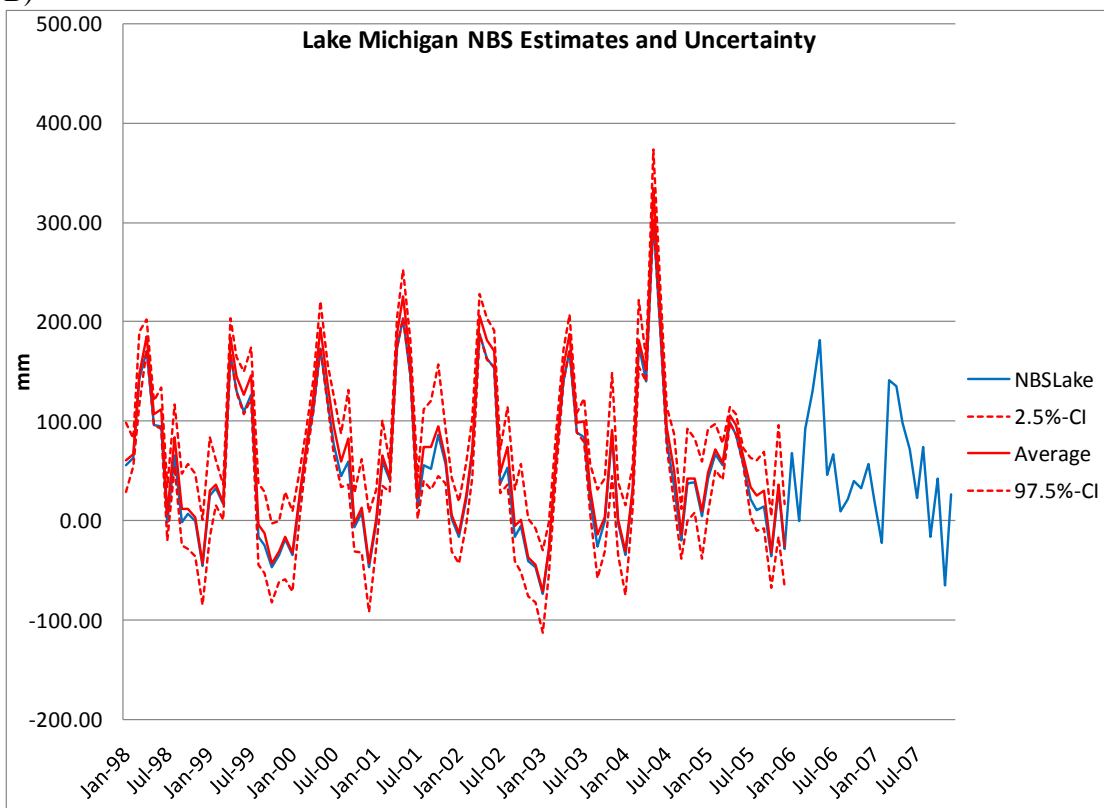
A)
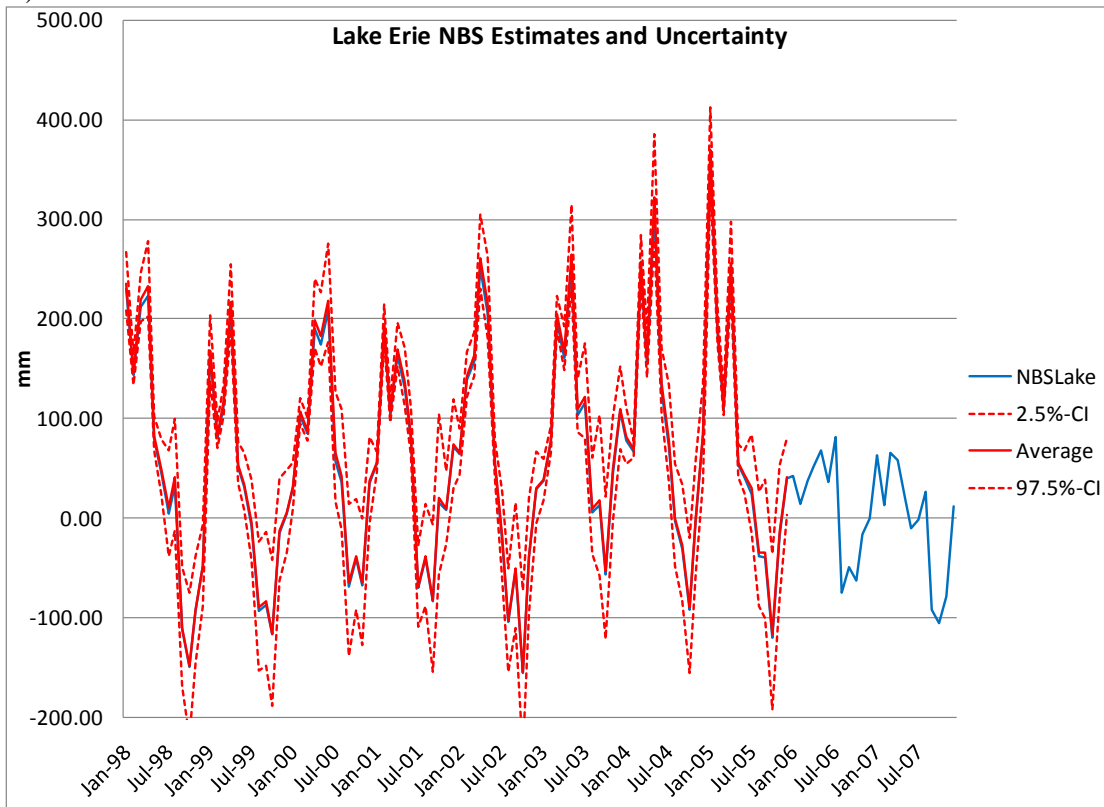


B)



**Figure 11**: GLERL NBS-Lake, Monte Carlo NBS best estimate and 95% confidence interval using A) IMEP*Lake Erie; B) Lake Michigan.

Overall, in the hypothesis that evaporation estimates are unbiased and that their uncertainty is around 35%, GLERL NBS estimates would underestimate the real NBS by 6.2% for Lake Erie and 13.6% for Lake Michigan. However, overall uncertainty in monthly NBS would be in the60% range for Lake Erie and in the 40% range for Lake Michigan.

| | Erie | St.Clair | Huron | Georgian Bay | Michigan | Superior | UGL |
|---|---|---|---|---|---|---|---|
| 0 | 1.4 | 4.2 | 10.1 | 6.4 | 2.2 | 0.7 | 3.3 |
| 0-20% | 4.7 | 0.0 | 10.9 | 0.0 | 8.1 | 10.1 | 6.8 |
| 20-40% | 3.9 | 27.3 | 15.8 | 23.5 | 3.7 | 34.1 | 18.1 |
| 40-60% | 20.6 | 0.0 | 6.3 | 15.3 | 12.9 | 23.6 | 16.2 |
| 60-80% | 17.6 | 56.3 | 0.0 | 5.4 | 21.2 | 16.8 | 15.6 |
| 80-95% | 15.4 | 0.0 | 43.3 | 30.5 | 35.7 | 10.3 | 24.4 |
| 95-100% | 36.5 | 12.2 | 13.6 | 19.0 | 16.1 | 4.5 | 15.6 |

**Table 6**: Fractions of the basin that in 1988-2005 belong to different gaged watershed portion categories.

## ACKNOWLEDGEMENTS

## REFERENCES AND CITATIONS

[1] Neff, B.P., & Nicholas, J.R. (2004). *Uncertainty in the Great Lakes Water Balance*. Scientific Investigations Report 2004-5100, United States Geological Survey, Denver, Colorado.

[2] United States Geological Suervey (2008). *Water Resources of the United States—2008 Annual Data Report. Documentation*. (http://wdr.water.usgs.gov/current/documentation.html#stage). Accessed Jan. 2009.

[3] Lee D.H. (1992). Computation of net basin supplies—A comparison of two methods, in *International Joint Commission levels reference study, Final Report Subtask 19.1.2a*, Great Lakes Environmental Research Laboratory, Ann Arbor, Michigan.

[4] Groisman, P.Y., & Legates, D.R. (1994). The accuracy of United States precipitation data. *Bulletin of the American Meteorological Society*, **75**(3), 215-217.

[5] Winter, T.C. (1981). Uncertainties in estimating the water balance of lakes. *Water Resources Bulletin* **17**, 82-115.

[6] Legates, D.R., & DeLiberty, T.L. (1993). Estimates of biases in precipitation gage measurements—An example using the United States raingage network, in Groisman, P.Y., and Legates, D.R., eds., 1994, *Proceedings from the eighth symposium on meteorological observations and instrumentation*: Anaheim, California, American Meteorological Society. J48-J51, figure 1, p. 219.

[7] Miner, T.J., & Fritsch, J.M. (1997). Lake-effect rain events, *Mon Wea Rev.*, **125**, 3231-3248.

[8] DeMarchi, C. (2006). *Probabilistic estimation of precipitation combining geostationary and TRMM satellite data*. Doctoral Dissertation, Georgia Institute of Technology, Atlanta, Georgia.

[9] Croley, T.E. II, & Hunter, T.S. (1994*). Great Lakes monthly hydrologic data*: NOAA Technical Report #TM-083. NOAA .Great Lakes Environmental Research Laboratory, Ann Arbor, Michigan.

[10] International Upper Great Lakes Study (2009). *Impacts on Upper Great Lakes Water Levels: St. Clair River. Final Report to the International Joint Commission*. International Joint Commission, Ottawa, Ontario; Washington, D.C., pp. 244.

[11] DeMarchi, C., Georgakakos, A., & Peters-Lidard, C. (2007). Uncertainty characterization in a combined IR/Microwave scheme for remote sensing of precipitation. In *Quantification and Reduction of Predictive Uncertainty for Sustainable Water Resources Management (Proceedings of Symposium HS2004 at IUGG2007, Perugia, July 2007), IAHS Publ. 313*. Boegh, E., Kunstmann, H., Wagener, T., Hall, A., Bastidas, L., Franks, S., Gupta, H., Rosbjerg, D., & Schaake, J. (Eds.). IAHS Press, Centre for Ecology and Hydrology, Wallingford, United Kingdom, 70-77.

[12] DeMarchi, C., Croley, T.S. II, He, C., & Hunter, T.S. (2009). Application of a distributed watershed hydrology and water quality model in the Great Lakes basin. In *Proceedings of the 7th International Symposium on Ecohydraulics 2009*, Concepcion, Chile, January 12-16, 2009.

[13] DeMarchi, C., Xing, F., Croley,T.S. II, He, C., & Wang, Y. (2009). Application of a Distributed Large Basin Runoff Model to Lake Erie: Model calibration and analysis of parameter spatial variation. *Journal of Hydrological Engineering (In review)*.

[14] Hammersley, J. M., & Handscomb, D.C. (1975). *Monte Carlo Methods*. Methuen, London, United Kingdom.

[15] Croley, T.E. II, & Hartmann, H.C. (1985).Resolving Thiessen polygons, *Journal of Hydrology*, **76**, 363-379.

[16] Croley, T.E. II (2003). Hydrological scenario development, in Lavender, B., Smith, J.V., Koshida, G., and Mortsch, L.D., eds., *Binational Great Lakes-St. Lawrence basin climate change and hydrologic scenarios report*, Environment Canada Environmental Adaptation Research Group, Downsview, Ontario, pp 45-88.

[17] Smith J.A., Seo, D.J., Baeck, M.L., & Hudlow, M.D.(1996). An intercomparison study of NEXRAD precipitation estimates, *Water Resources Research*, **32**, 2035–2045.

[18] Lin, Y., & Mitchell, K.E. (2005). The NCEP Stage II/IV hourly precipitation analyses: development and applications, *Proc., 19th Conf. on Hydrology, American Meteorological Society*, San Diego, CA, Paper 1.2.

[19] Watkins, D.W. Jr., Li, H., & Cowden, J.R. (2007).Adjustment of Radar-Based Precipitation Estimates for Great Lakes Hydrologic Modeling, *Journal of Hydrologic Engineering* **12**(3), 298-305.

[20] Chatterjee, A., DeMarchi, C., & Michalak, A.M. (2009). Estimating over-lake precipitation in the Great Lakes combining Radar and rain gages. In *Proceedings of the 7th International Symposium on Ecohydraulics 2009*, Concepcion, Chile, January 12-16, 2009.

[21] DeMarchi, C., Dai, Q., Mello, M.E., and Hunter, T.S. (2009). *Estimation of Overlake Precipitation and Basin Runoff Uncertainty*. Prepared for the International Upper Great Lakes Study Case Western Reserve University, Cleveland, Ohio. 64 pp.

[22] Hollinger, S.E., Angel, J.R., & Palecki, M.A. (2002). *Spatial Distribution, Variation and Trends in Storm Precipitation Characteristics Associated with Soil Erosion in the United States*, Contract report CR 2002-08, Illinois State Water Survey. Champaign, Illinois.

[23] Spence, C., Blanken, P.D., Hedstrom, N., Mackay, M.D., and Hunter, T.S. (2009). *A Comparison of Modelled and Observed Evaporation from Lake Superior*. Hydroclimate Technical Working Group Task 3.2.3, International Upper Great Lakes Study. Saskatoon, Saskatchewan.