7th ISE & 8th HIC
Chile, 2009

ESTIMATING OVER-LAKE PRECIPITATION IN THE GREAT LAKES BY COMBINING RADAR AND RAIN GAGES

ABHISHEK CHATTERJEE
Dept. of Civil and Environmental Engineering, University of Michigan, 1351 Beal Ave.
Ann Arbor, MI 48109-2125, United States

CARLO DEMARCHI
School of Natural Resources & Environment, University of Michigan, 440 Church Street
Ann Arbor, MI 48109-1041, United States

ANNA M. MICHALAK
Dept. of Civil and Environmental Engineering, and Dept. of Atmospheric, Oceanic and
Space Sciences, University of Michigan, 1351 Beal Ave.
Ann Arbor, MI 48109-2125, United States

Over-lake precipitation is a key component of the Great Lakes' water balance. Its estimation is, therefore, vital for planning and operational purposes. Yet, reliable over-lake precipitation estimates are difficult to obtain because the lack of gages on the lakes themselves and the scarcity of gages in parts of the draining basins. Traditionally, over-lake precipitation is estimated by distance-weighted interpolation methods. In spite of their wide acceptance, these methods suffer from intrinsic limitations as they fail to take into account the spatial variability of rainfall. Recently, multisensor products combining radar-based precipitation estimates and rain gage data (MPE) have provided a suitable alternative to estimates based on the sparse gage data. However, biases in the MPE data have raised serious concerns about their accuracy. A promising approach for overcoming the limitations of either of these methodologies for estimating monthly-averaged over-lake precipitation is to spatially integrate the MPE data with the gage observations in a geostatistical framework based on universal kriging. In this work, the estimates from these techniques are compared to (i) more traditional methods based on the weighted interpolation gage data only and (ii) the available MPE data. Results for Lake Erie reveal that the universal kriging setup outperforms the estimation methods based only on one of the two data types, by providing estimates with significantly lower root mean square error and lower overall bias. Overall, the results demonstrate the robustness of the proposed approach in assimilating two different information sources for providing more accurate and reliable estimates of over-lake precipitation.

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.

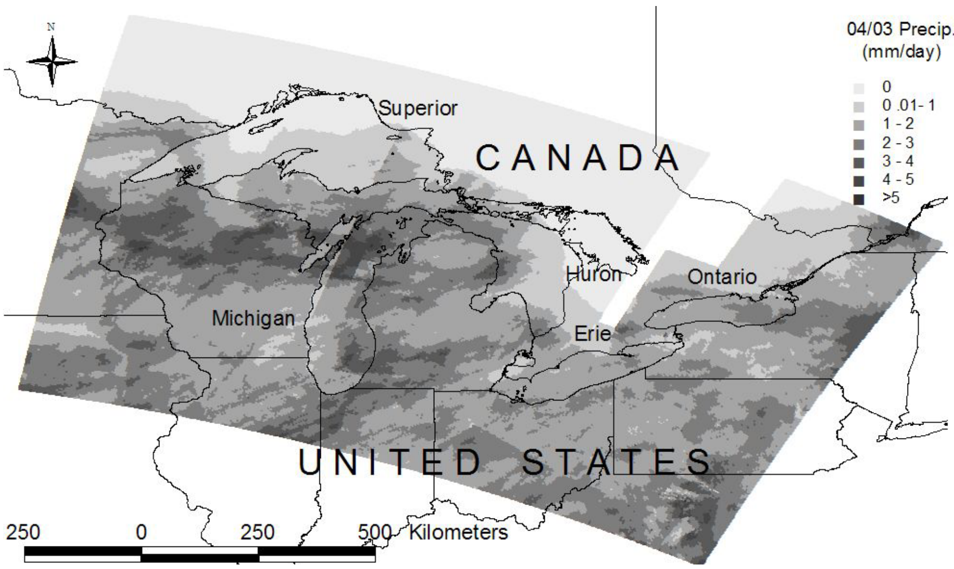

Figure 1. The Great Lakes basin and average MPE precipitation estimates for April 2003.

It is therefore reasonable that considerable effort has been invested in understanding and managing the Great Lakes, starting with their water balance. However, the huge sizes of the overall basin (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 (Neff and Nicholas [1]). Of the four components of the net basin supply (tributary runoff, over-lake precipitation, evaporation, and net groundwater inflow), only runoff is relatively well measured (although the ungaged portion of the lakes' watersheds varies between 22 and 43% − Lee [2]). The remaining three components must be quantified using some type of model. Especially problematic is the estimation of the precipitation falling directly over the lakes, which is still estimated by interpolating rain gages using the Thiessen or inverse square distance weighting methods (Croley and Hartmann [3]; Croley [4]). In this region, these techniques are not very reliable, mainly because rain and snow vary strongly in time and space, while the lack of gages on the lakes themselves and the scarcity of gages in parts of the draining basins leave 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 quantities considerably higher than over the lake surface. Differences in the precipitation mechanisms are present also during spring and summer (e.g., Miner and Fritsch [5]; DeMarchi [6] for Lake Victoria). Such factors make over-lake precipitation estimates so uncertain that the National Oceanic and Atmospheric Administration (NOAA) Great Lakes Environmental Research Laboratory (GLERL) prefers to use the average precipitation over the draining watershed as a surrogate for the real over-lake precipitation (Croley *et al.* [7]).

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 (Smith *et. al.* [8]). The NOAA National Centers for Environmental Prediction's Multisensor Precipitation Estimates Stage IV (Lin and Mitchel [9]) 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.

This paper examines the suitability of MPE data for estimating monthly over-lake precipitation for Lake Erie and presents a new methodology for improving data assimilation based on a universal kriging approach. The first section describes the area selected for the study and the datasets that were used. The second section examines the suitability of MPE for over-lake precipitation estimation. The third describes the models developed to overcome MPE's shortcomings and the statistical analyses used to evaluate their performances. Finally, the fourth section discusses model results.

STUDY AREA AND AVAILABLE DATA

Lake Erie is the fourth largest lake by surface area (25,745 km^2 with a length of 388 km and a maximum breadth of 92 km), the shallowest (with an average depth of 19 m), and the southernmost of the five Great Lakes (Figure 1). It is bounded on the North by the Canadian province of Ontario, on the South by the U.S. states of Ohio, Pennsylvania, and New York, and on the West by the state of Michigan (GLIN [10]). Meteorologically the climate of Lake Erie's shoreline is greatly affected by the lake's influence on weather systems, which usually move from the West to the East shores of the lake. To the East of the lake, the land becomes increasingly hilly. Thus, as the moist air from Lake Erie travels eastward into these hilly areas, it begins to rise, cool and precipitate its excess water, making the southwestern corner of Lake Erie one of the wettest areas in the basin (the average yearly precipitation in Buffalo, New York, is 924 mm). Such convective precipitation patterns can have a totally different spatial correlation structure than the ones inferred from the more prevalent stratiform precipitation.

Daily MPE Stage IV precipitation data for the Great Lakes region were extracted from the national mosaics for the summer months (April through September) in 2002 to 2005. Stage IV estimates are a significant improvement over the previous operational precipitation estimation algorithms (Watkins *et. al.*, [11]), because they account for beam blockage, overshooting and software errors.

Daily precipitation data from U.S. and Canadian rain gages in the Lake Erie region were compiled by NOAA GLERL. To ensure statistical significance of the MPE-gage comparison, this dataset was reduced to include only gages featuring valid data for more than 90% of the examined period, thus limiting its size to 516 gages around Lake Erie.

ASSESSING THE SUITABILITY OF MPE DATA FOR ESTIMATING OVER-LAKE PRECIPITATION IN THE GREAT LAKES BASIN

Figure 1 shows that the MPE data are available for most of the Great Lakes Basin. However, Lake Superior and Lake Huron are covered partly by areas of zero precipitation values and partly by areas of extremely low values (0.01-1 mm class) with the circular shape typical of the Radar swath. On the other hand, MPE data over Lake Ontario, Lake Erie, Lake St. Clair, and Lake Michigan show reasonable continuity with over-land precipitation. Thus it seems plausible that MPE estimates could be used at best only for these latter lakes.

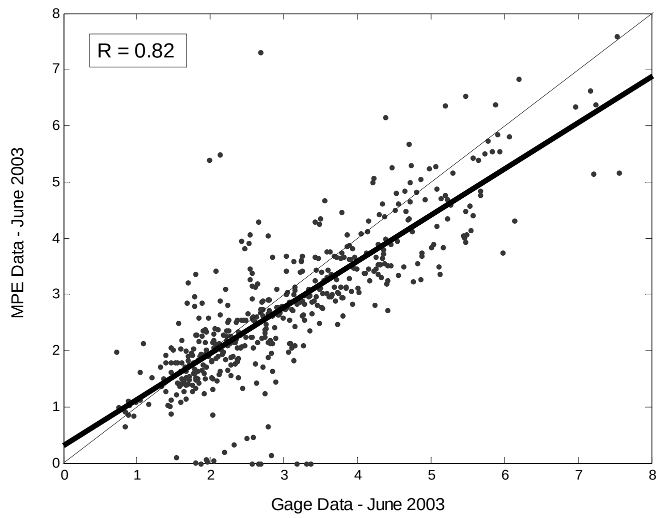

Figure 2. Linear Trend between gage and coincident MPE data – the bold line indicates the “best-fit” line while the thin line shows the 1:1 fit.

Comparisons between gage and coincident MPE estimates in the Lake Erie region (76 - 87 W, 39 - 45.5 N) indicate that the MPE underestimates the gage catch by 5-20%, but still features a strong correlation with gage data (e.g., Figure 2).

Other than April 2004, which has a correlation coefficient value of 0.44, all the other 24 months of study (April 2002-September 2005) have correlation coefficient values ranging from 0.67 to 0.94. MPE underestimates precipitation for two main reasons: (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 (Hollinger *et al.* [12]).

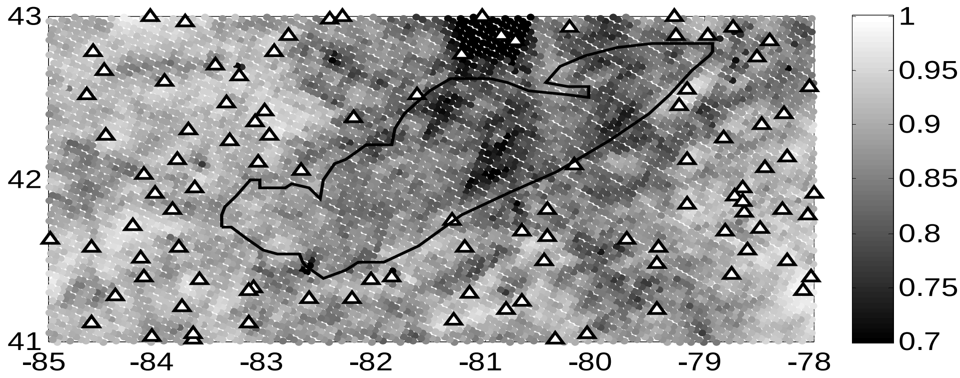

Figure 3. Correlation coefficient between MPE pixels and the most correlated gage (triangles).

Figure 3 shows the spatial variation in the correlation coefficient between an MPE pixel and the gage that is most correlated to the pixel for a portion of the domain (78-85W, 41-43N). Three facts should be noted: (1) while over-land MPE pixels have higher correlation coefficients than over-lake pixels, the latter show a smooth decrease in the western and eastern parts of the lake, likely due to the increasing distance from gages, and values comparable to similar situations over land (e.g., the area around 41.8N, 79.5W); the lower correlation at the center of the lake is probably due to occasional malfunctioning of the radar in Cleveland, Ohio. (2) Canadian gages show a good correlation with MPE despite the distance from the NWS radars. The exception is the area around 81W, again likely due to occasional malfunctioning of the radar in Cleveland, Ohio. (3) In many parts of the area, correlation coefficients are high for MPE pixels very close to gages, but decrease rapidly when moving away from the gage, even just two-three pixels (approximately 10 kilometers). This indicates that, even at monthly resolution, the precipitation field is far from smooth, thus making interpolation of rain gage data less reliable and increasing the value of using MPE data for quantifying over-lake precipitation.

These observations suggest that, if the bias in the MPE data were removed, MPE could potentially provide reliable monthly precipitation estimates over Lake Erie.

IMPROVING MPE ESTIMATES OF MONTHLY PRECIPITATION

Simple Method

Two methods were developed to improve monthly MPE estimates by merging MPE and daily rain gage data. The simplest method adjusts the MPE data in the Lake Erie region by a monthly correction factor, mathematically,

$$MPE^{*} = MPE \times \frac{G_{AVG}}{MPE_{AVG}} \tag{1}$$

where MPE* = adjusted monthly MPE;
MPE = original monthly MPE;
G_{AVG} = Average precipitation recorded by all gages in the region of application during a specific month;
MPE_{AVG} = Average MPE precipitation for pixels surrounding a gage in the region of application during the same month;

The correction factor obtained with this method is of magnitude similar to the values obtained by Watkins *et al.*, [11], though the procedures used in the two projects are quite different. Further, an inter-annual variation in the correction factor was noted. In 2002 the correction factor ranged from 1.03-1.23, and by 2005 it had dropped down to 0.95-1.12. Since the MPE data is available over a limited period, it is difficult to state whether this is due to the improving performance of MPE or an artifact of any other random factor, like more uniform precipitation in 2004 and 2005, better gage and radar operation, more consistent calibration, etc.

Universal Kriging

A shortcoming of the simple method is that the same correction is applied to the entire region of interest, neglecting spatial variations in the gage/MPE ratio. A way to overcome this limitation is by integrating gage and MPE data within a Universal Kriging (UK) framework (Matheron [13]). In this framework, the variable to be estimated Z(x) – measured at some point x_i, i=1,…,n – is considered to be composed of a deterministic component m(x) – known everywhere in the domain – and a stationary stochastic component Y(x) (the residual), that is:

$$Z(x) = m(x) + Y(x) \tag{2}$$

In essence, *m(x)* represents the "large-scale component of spatial variability" (Kitanidis [14]), while *Y(x)* represent the local deviations from this trend. For this application, the stochastic process are the features of the precipitation distribution as measured by the rain gages that are not reproduced by the MPE data, while the deterministic part is given by the MPE data, which are known at both the gage measurement and estimation locations. Since MPE data are significantly correlated with the gage data at the monthly scale (Figure 2 and 3), a linear transformation of the MPE data can be used as the trend, that is:

$$m(x) = \beta_0 + \beta_1 \cdot MPE \tag{3}$$

The stochastic component of Z(x) is modeled using an exponential generalized covariance function (GCF) derived from the variogram of the residual components of the gage precipitation.

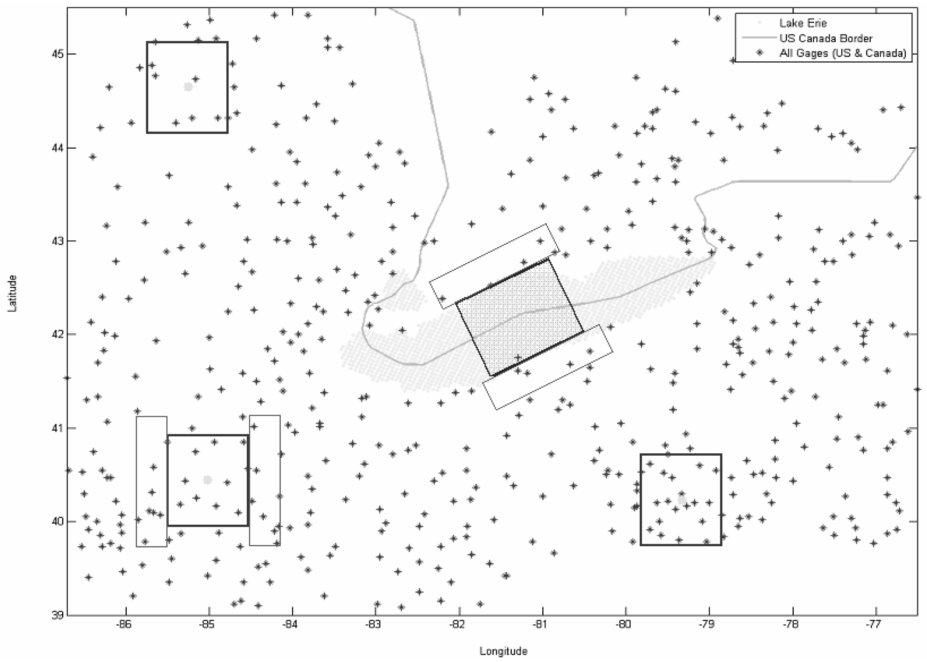

Figure 4. Rectangular areas used to emulate characteristics of the central Lake Erie – crosses represent usable gages; the shaded square corresponds to the central portion of Lake Erie; the rectangular areas at its sides contain the closest gages that are available for over-lake estimation; the hollow squares are locations with similar geometry to the central basin of Lake Erie; however, only the lower-left square has a distribution of rain gages at its borders similar to that of Lake Erie.

Evaluating MPE-based Estimates of precipitation over Lake Erie

Direct evaluation of the MPE-based estimates of over-lake precipitation and comparison with traditional methods is not possible given the absence of precipitation measurements over the lake surface itself. Thus, an indirect evaluation was performed assessing model performances for 120 rectangular areas randomly selected in the region surrounding Lake Erie, which replicate some of the conditions present in the central part of Lake Erie (Figure 4). That is, these areas are 80-km wide and have at least four valid nearby rain gages on at least two opposite sides (rectangular areas in Figure 4). These rain gages were used in the estimation schemes, while gages inside the rectangular areas provide the ground truth. The performance comparison included MPE, MPE*, UK, and two gage-only methods – inverse square distance interpolation (IDW) and ordinary kriging (OK).

RESULTS AND DISCUSSION

Table 1 shows that MPE has better correlation with the gage data and lesser RMSE than IDW and OK, but it clearly underestimates precipitation. MPE*, the simple scaling correction of MPE, has the same correlation and RMSE of MPE, but its bias is significantly reduced (it is actually better than that for IDW). Finally, the UK estimator achieves the best results with the highest correlation, lowest RMSE, and negligible bias.

Table 1. Comparison of different estimators for the period 2002-2005. Median values for 120 rectangular areas (Bold) and relative median absolute deviation (in parenthesis).

	IDW	OK	MPE	MPE*	UK
RMSE (mm/day)	**0.98** (0.09)	**1.03** (0.19)	**0.76** (0.07)	**0.72** (0.06)	**0.65** (0.06)
Correl. Coeff.	**0.83** (0.04)	**0.81** (0.06)	**0.91** (0.02)	**0.90** (0.02)	**0.92** (0.02)
Bias (%)	**-4.29** (3.5)	**-1.11** (3.78)	**-7.51** (3.19)	**-1.94** (3.45)	**-0.45** (1.72)

Users of over-lake precipitation estimates are normally more interested to the lake-wide Mean Areal Precipitation (MAP) than the precipitation at specific points in the lake. Figure 5 shows Lake Erie's MAP estimates for total precipitation for each of the 24 summer month considered in this study. In addition to UK, MPE, and IDW, Figure 5 reports UK's 95% confidence interval and NOAA's mean precipitation for the lake surface (PrecipLake) and for the draining watershed (PrecipLand) derived using Thiessen polygons (Croley *et al.* [7]). The chart reveals that, because the uncertainty of the UK estimator is quite large, all other estimation methods fall within UK's 95% confidence interval, indicating general agreement between the methods in terms of total precipitation. However, MPE most often yields the lowest estimate, confirming its likely bias. Further, PrecipLake is typically closer to UK than PrecipLand, indicating that using

PrecipLake instead of PrecipLake should produce better net basin supply estimates, at least during the summer months.

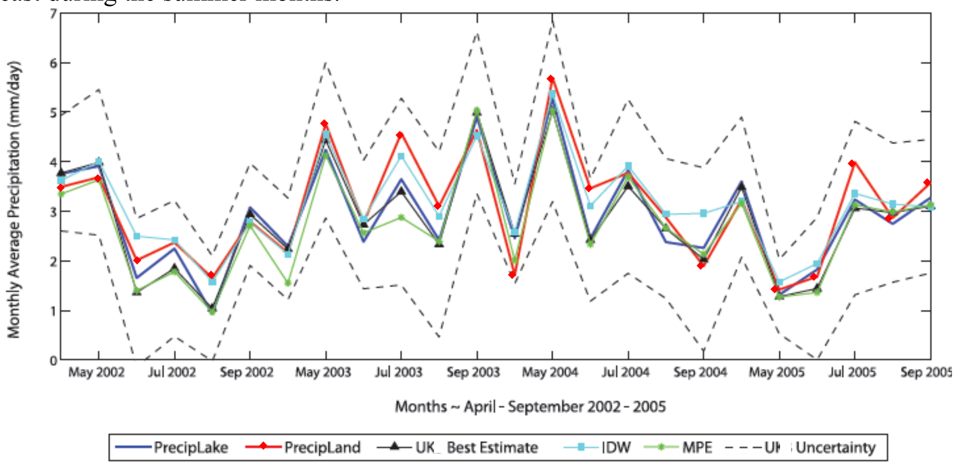

Figure 5. Lake Erie's Mean Areal Precipitation from different estimation methods.

CONCLUSIONS

Given the strong correlation with coincident rain gages in both U.S. and Canada, and the smooth variation of the correlation between over-lake MPE values and gage values along the lake shores of Lake Erie, MPE Stage IV appears to be a good candidate for estimating precipitation over Lake Erie. However, successful use of MPE for precipitation estimation first requires a correction of its bias. This can be accomplished with a simple ratio correction or by employing a more sophisticated Universal Kriging setting. In both cases, combining information from MPE data and rain gages results in better performance for areas in the Lake Erie region, relative to traditional gage-only methods. Further, lake-wide mean areal precipitation estimates are consistent with other estimation methods. Thus, merging MPE and daily gage data substantially improves over-lake precipitation for Lake Erie. Lake Ontario and Lake Michigan also show potential for similar improvements, while the application of MPE to Lake Superior and Lake Huron will require additional radar and gage measurements in the Canadian side of these lakes.

ACKNOWLEDGEMENTS

This research has been supported by NOAA Great Lakes Environmental Research Laboratory; University of Michigan's Cooperative Institute for Limnology and Ecosystem Research, College of Engineering, and Rackham Graduate School; and the International Joint Commission's International Upper Great Lakes Study.

REFERENCES

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

[2] Lee D.H., "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, MI, (1992).

[3] Croley T.E. II, and Hartmann H.C., "Resolving Thiessen polygons", *Journal of Hydrology*, Vol. 76, (1985), pp 363-379.

[4] Croley T.E. II, "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, Ont., pp 45-88.

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

[6] De Marchi C., "*Probabilistic estimation of precipitation combining geostationary and TRMM satellite data",* Doctoral dissertation, Georgia Institute of Technology, Atlanta, GA, (2006).

[7] Croley T.E. II, Hunter T.S., and Martin S.K., "*Great Lakes monthly hydrologic data*", NOAA Technical Report TM-083, NOAA GLERL, Ann Arbor, MI, (2001).

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

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

[10] Great Lakes Information Network, "Lake Erie: Facts and Figures", *http://www.great-lakes.net/lakes/ref/eriefact.html*, (2006).

[11] Watkins D.W. Jr., Li H., Cowden J.R., "Adjustment of Radar-Based Precipitation Estimates for Great Lakes Hydrologic Modeling", *Journal of Hydrologic Engineering* Vol. 12, No. 3, (2007), pp 298-305.

[12] Hollinger S.E., Angel J.R., and Palecki M.A., "*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, IL, (2002).

[13] Matheron G., "*Les Variables Régionalisées et Leur Estimation*", .Masson, Paris, France, (1965), pp 305.

[14] Kitanidis P.K., "*Introduction of Geostatistics*", University Press, Cambridge, MA, (1997).