

Evaluation of NOAA Climate Outlooks in Extended Great Lakes Water Levels Forecasts

Thomas E. Croley II¹

Abstract

The Great Lakes Environmental Research Laboratory (GLERL) simulates time series of daily Great Lakes hydrology by first estimating initial hydrological conditions and then using a daily meteorology time series (scenario) taken from the historical record. They do this to make a deterministic hydrology “forecast” (including lake levels) from a representative meteorology scenario. GLERL repeats this for other segments of the historical record in an “operational hydrology” or “ensemble” approach in their Advanced Hydrologic Prediction System (AHPS). The resulting set of lake level scenarios serves as a statistical “sample” for inferring probabilistic lake levels outlooks that properly consider antecedent hydrological conditions. GLERL does this every day to produce six-month outlooks. Meanwhile, the National Oceanic and Atmospheric Administration’s Climate Prediction Center publishes each month multiple long-lead probabilistic meteorology outlooks. GLERL transforms these meteorological outlooks into equations for sample weights and solves them simultaneously. Their AHPS methodology weights their samples of six-month lake levels scenarios each day to include the effects of these meteorology predictions. GLERL simulated both deterministic and probabilistic lake level outlooks over 1995-2000 without the use of antecedent conditions or meteorology predictions and then added them into the simulation to assess the value of each in making the forecast.

Introduction

The Great Lakes Environmental Research Laboratory (GLERL) built the Advanced Hydrologic Prediction System (AHPS) for forecasting Laurentian Great Lakes water level probabilities six months into the future. They evaluated AHPS, US, and Canadian lake level outlooks in 1997, and again more recently, as they accumulated operation experience. They wanted to quickly determine the value of considering antecedent conditions and (separately) available meteorological outlooks in making hydrological outlooks. Also considered was the suitability of GLERL’s AHPS forecasts relative to the methods actually used officially by US and Canadian authorities responsible for issuing joint Great Lakes levels outlooks. The following sections briefly describe GLERL’s AHPS for the Great Lakes, the study design to organize the comparisons, and the deterministic comparisons for assessing the value of antecedent conditions and meteorological outlooks in GLERL’s AHPS. Comparisons with other agency outlooks are then summarized only.

Great Lakes AHPS

GLERL developed, calibrated, and verified conceptual model-based techniques for simulating hydrological processes in the Laurentian Great Lakes (Superior, Michigan, Huron, Georgian

¹Research Hydrologist, Great Lakes Environmental Research Laboratory, National Oceanic and Atmospheric Administration, US Department of Commerce, 2205 Commonwealth Blvd., Ann Arbor, Michigan 48105-2945; phone 734-741-2238; croley@glerl.noaa.gov

Bay, St. Clair, Erie, and Ontario). GLERL integrated the models into a system to estimate water and energy balances, whole-lake heat storage, and lake levels. Croley et al. (1996) conveniently summarizes details of the integrated models. The modeling system is modularly built, allowing model upgrades to be added as they develop. The modeling system is coupled with near real-time data acquisition and reduction to enable representation of current system states. Inputs are daily meteorology (air temperature, dew point temperature, precipitation, wind speed, and cloud cover) for all available stations. Optional inputs are snow water equivalent, soil moisture, lake water temperature, and lake levels. Daily provisional point data is converted into areal averages for each watershed and lake surface (Croley and Hartmann, 1985) for GLERL's runoff and lake thermodynamics models to estimate basin moisture and lake heat as antecedent (initial) conditions to a forecast. A deterministic "forecast" of all hydrologic variables, including lake supply, is made then by simulating the hydrology from the point of estimated initial conditions forward with a meteorological scenario (taken from the historical record). The resulting lake supply scenarios, one for each lake, then are used with connecting-channel routing and lake regulation models to determine a lake levels scenario. This is repeated for alternate meteorological scenarios (other segments of the historical record) in an "operational hydrology" or "ensemble" approach. The resulting set of scenarios serves as a statistical sample for inferring probabilities associated with both meteorology and hydrology. Probabilistic hydrologic outlooks can be made directly from this sample for each variable of interest. The resulting probabilistic hydrologic outlooks would properly consider antecedent hydrological conditions, but not available meteorological predictions.

Multiple long-lead probabilistic meteorological outlooks are now available to the water resource engineer or hydrologist. They are defined over different time periods at different time lags; they forecast either event probabilities or most-probable events. Considered here are outlooks from the National Oceanic and Atmospheric Administration's (NOAA's) Climate Prediction Center (CPC), which changed in January 1995 from issuing a few relatively short-term outlooks of meteorological probabilities to a new multiple long-lead outlook. Each outlook is issued mid-month and consists of a 1-month forecast for the next (full) month and thirteen 3-month forecasts, going into the future in overlapping fashion in 1-month steps (14 forecasts). Each forecast predicts probabilities of average air temperature and total precipitation falling within upper-third and lower-third intervals (4 equations). Thus, each outlook yields 56 probability equations each month for any location in the continental United States. The probability equations are transformed into equations involving sample weights and are solved simultaneously for physically relevant values of the weights (Croley, 1996, 1997, 2000, 2001). The solution may involve an optimization, when there is more than one set of weights possible, and therefore may require an objective function to select between various solutions. Since all 56 equations may not be satisfied simultaneously, they are ordered in priority and as many as possible are used (the lowest-priority equations are discarded). More weight is given to those sample scenarios whose corresponding historical meteorological record segments contain events appropriate to the meteorological forecasts. For example, more weight is given to those six-month lake level scenarios corresponding to monthly air temperatures in the upper third of their range when the meteorological outlook calls for above-normal monthly air temperatures; the value of the weight depends on the probabilities specified in the meteorological outlook. GLERL's AHPS methodology biases the sample of six-month lake level scenarios each month by weighting them to agree with these equations and then estimates outlook probabilities from the weighted sample by using the Weibull estimator (Croley, 2001).

Study design

GLERL began in September 1997 to evaluate their Great Lakes AHPS; the data at hand included an earlier 1982-1988 comparison of net basin supply (NBS) forecasts (Croley, 1993; Croley and Lee, 1993). (Net basin supply consists of over-lake precipitation and basin runoff less lake evaporation.) GLERL assembled beginning-of-month lake levels of record, diversions of record, lake outflows of record, actual monthly and quarter-monthly average lake levels, NBS (computed as the daily algebraic sum of over-lake precipitation, estimated lake evaporation, and basin runoff to each lake for 1954-1997), and all NOAA CPC meteorological outlooks through August 1997. GLERL also reduced all meteorological data to daily spatial averages over each of the watersheds and water surfaces of the Great Lakes; for 1948-1995, data consisted of final quality-controlled values (as reported by the collecting agencies) and for 1996-August 1997, data were provisional (as received in near-real time and unchecked for quality).

GLERL simulated probabilistic lake level outlooks for 1995-August 1997 with alternate “operational hydrology” methods. Firstly, GLERL assembled all six-month NBS time series from the historical record (1954-1997) that started the same month as each month of the period 1995-August 1997 into a sample for that month, from which to estimate a six-month forecast beginning that month. Only the period of record preceding each month of 1995-August 1997 was used to assemble the sample for that month, simulating operation in real time. For example, for the first month of the period, January 1995, GLERL used all six-month NBS time series beginning in January, prior to 1995, to build a sample; for the thirtieth month of the period, June 1997, GLERL used all six-month NBS time series beginning in June, prior to 1997, to build a sample. GLERL then converted each sample of six-month NBS time series into a sample of six-month lake level scenarios with appropriate routing and regulation models; they used the resulting samples to infer a six-month probabilistic outlook of lake levels beginning each month of the period with the Weibull estimator. This operational hydrology method represents forecasts without consideration of antecedent conditions and meteorological outlooks.

Secondly, GLERL simulated six-month lake level scenarios with their AHPS (which uses estimates of antecedent moisture and heat storage conditions with six-month pieces of the historical meteorological record). They did this for each month of the period 1995-August 1997 and assembled the six-month lake level scenarios into a sample for that month from which to estimate a six-month forecast beginning that month. Again, only historical meteorology preceding each forecast was used, simulating data availability in real time. Only provisional data were used, as they would have been available in near real time. Since no weightings were used, this represents forecasts, for each month of the period, that consider antecedent conditions but do not use meteorological outlooks.

Thirdly, GLERL simulated six-month lake level forecasts with their AHPS, using both antecedent conditions and NOAA’s 1- and 3-month meteorological outlooks, for each month of the period. GLERL used ten methods for considering these meteorological outlooks in their hydrological outlooks. The first five methods used a mixture of simultaneous meteorological outlooks over seven lake basins, ordered as indicated in Table 1. They used different objective functions, however, to select among competing sets of weights: a) minimization of the sum of squared differences between each weight and unity while using the most meteorological outlooks (Croley, 1996, 1997, 2000), b) minimization of the sum of squared differences between each weight and unity while forcing all weights non-zero (use all hydrological scenarios), c) maximization of probability of mid-third (normal) values for the first six-month air temperature and precipitation over all basins (Croley, 2000, 2001), d) maximization of probability of first six-month air tem-

Table 1. Ordered meteorological outlooks for the “All-Lakes” outlooks.

Order	Basin ^a	Meteorology ^b	Range ^c	Order	Basin ^a	Meteorology ^b	Range ^c
1	SUP	1-mo T	Lower Third	29	HUR	3-mo T	Lower Third
2	SUP	1-mo T	Upper Third	30	HUR	3-mo T	Upper Third
3	MIC	1-mo T	Lower Third	31	GEO	3-mo T	Lower Third
4	MIC	1-mo T	Upper Third	32	GEO	3-mo T	Upper Third
5	HUR	1-mo T	Lower Third	33	ERI	3-mo T	Lower Third
6	HUR	1-mo T	Upper Third	34	ERI	3-mo T	Upper Third
7	GEO	1-mo T	Lower Third	35	ONT	3-mo T	Lower Third
8	GEO	1-mo T	Upper Third	36	ONT	3-mo T	Upper Third
9	ERI	1-mo T	Lower Third	37	SUP	3-mo P	Lower Third
10	ERI	1-mo T	Upper Third	38	SUP	3-mo P	Upper Third
11	ONT	1-mo T	Lower Third	39	MIC	3-mo P	Lower Third
12	ONT	1-mo T	Upper Third	40	MIC	3-mo P	Upper Third
13	SUP	1-mo P	Lower Third	41	HUR	3-mo P	Lower Third
14	SUP	1-mo P	Upper Third	42	HUR	3-mo P	Upper Third
15	MIC	1-mo P	Lower Third	43	GEO	3-mo P	Lower Third
16	MIC	1-mo P	Upper Third	44	GEO	3-mo P	Upper Third
17	HUR	1-mo P	Lower Third	45	ERI	3-mo P	Lower Third
18	HUR	1-mo P	Upper Third	46	ERI	3-mo P	Upper Third
19	GEO	1-mo P	Lower Third	47	ONT	3-mo P	Lower Third
20	GEO	1-mo P	Upper Third	48	ONT	3-mo P	Upper Third
21	ERI	1-mo P	Lower Third	49	STC	1-mo T	Lower Third
22	ERI	1-mo P	Upper Third	50	STC	1-mo T	Upper Third
23	ONT	1-mo P	Lower Third	51	STC	1-mo P	Lower Third
24	ONT	1-mo P	Upper Third	52	STC	1-mo P	Upper Third
25	SUP	3-mo T	Lower Third	53	STC	3-mo T	Lower Third
26	SUP	3-mo T	Upper Third	54	STC	3-mo T	Upper Third
27	MIC	3-mo T	Lower Third	55	STC	3-mo P	Upper Third
28	MIC	3-mo T	Upper Third	56	STC	3-mo P	Lower Third

^aGreat Lakes basin: SUP (Lake Superior), MIC (Lake Michigan), HUR (Lake Huron), GEO (Georgian Bay), STC (Lake St. Clair), ERI (Lake Erie), and ONT (Lake Ontario).

^bMeteorological variable including first month or first three month forecast designation.

^cRange of meteorological variable over which probability is forecast.

perature and precipitation in one-third ranges as suggested by extended meteorological outlook over all basins, and e) no objective.

The second five methods used meteorological outlooks over each basin individually, ordered as indicated in Table 2, and the same objective functions as above but defined only over each basin. (Of course, for lake level forecasting simultaneously on all lakes, only the simultaneous consideration of meteorological outlooks over all basins, as in the first five methods, is appropriate. However, consideration of meteorological outlooks over each basin independent of the other basins, in the second five methods, was attempted to discern possible improvements in single-lake forecasts of hydrological variables other than simultaneous lake levels on all lakes.) Inspection of results revealed the best method for forecasting lake levels to be one that used a mixture of simultaneous meteorological outlooks over the seven lake basins ordered as in Table 1 and using a minimization of the sum of squared differences between each weight and unity while using all non-zero weights. Actually, as revealed shortly, the meteorological outlooks added very little to the hydrological outlooks, and the manner in which the meteorological outlooks were

Table 2. Ordered meteorological outlooks for each “Individual Lake” outlook.

Order	Meteorology ^a	Range ^b	Order	Meteorology ^a	Range ^b
1	1 st 1-mo T	Lower Third	29	7 th 3-mo T	Lower Third
2	1 st 1-mo T	Upper Third	30	7 th 3-mo T	Upper Third
3	1 st 1-mo P	Lower Third	31	7 th 3-mo P	Lower Third
4	1 st 1-mo P	Upper Third	32	7 th 3-mo P	Upper Third
5	1 st 3-mo T	Lower Third	33	8 th 3-mo T	Lower Third
6	1 st 3-mo T	Upper Third	34	8 th 3-mo T	Upper Third
7	1 st 3-mo P	Lower Third	35	8 th 3-mo P	Lower Third
8	1 st 3-mo P	Upper Third	36	8 th 3-mo P	Upper Third
9	2 nd 3-mo T	Lower Third	37	9 th 3-mo T	Lower Third
10	2 nd 3-mo T	Upper Third	38	9 th 3-mo T	Upper Third
11	2 nd 3-mo P	Lower Third	39	9 th 3-mo P	Lower Third
12	2 nd 3-mo P	Upper Third	40	9 th 3-mo P	Upper Third
13	3 rd 3-mo T	Lower Third	41	10 th 3-mo T	Lower Third
14	3 rd 3-mo T	Upper Third	42	10 th 3-mo T	Upper Third
15	3 rd 3-mo P	Lower Third	43	10 th 3-mo P	Lower Third
16	3 rd 3-mo P	Upper Third	44	10 th 3-mo P	Upper Third
17	4 th 3-mo T	Lower Third	45	11 th 3-mo T	Lower Third
18	4 th 3-mo T	Upper Third	46	11 th 3-mo T	Upper Third
19	4 th 3-mo P	Lower Third	47	11 th 3-mo P	Lower Third
20	4 th 3-mo P	Upper Third	48	11 th 3-mo P	Upper Third
21	5 th 3-mo T	Lower Third	49	12 th 3-mo T	Lower Third
22	5 th 3-mo T	Upper Third	50	12 th 3-mo T	Upper Third
23	5 th 3-mo P	Lower Third	51	12 th 3-mo P	Lower Third
24	5 th 3-mo P	Upper Third	52	12 th 3-mo P	Upper Third
25	6 th 3-mo T	Lower Third	53	13 th 3-mo T	Lower Third
26	6 th 3-mo T	Upper Third	54	13 th 3-mo T	Upper Third
27	6 th 3-mo P	Lower Third	55	13 th 3-mo P	Upper Third
28	6 th 3-mo P	Upper Third	56	13 th 3-mo P	Lower Third

^aMeteorological variable including period-of-forecast designation.

^bRange of meteorological variable over which probability is forecast.

used was not very significant. Thus the best method, just identified, is only marginally better than most of the other methods investigated. Only its results are presented since they are representative of the other methods.

All of these methods yielded a set of six-month probabilistic lake level outlooks, which were simplified to yield a set of six-month deterministic outlooks for comparison to actual conditions. The simplifications consisted of taking the mean, the median, the mid-range between the 5% and 95% quantiles, the mid-range between the 20% and 80% quantiles, and the mode (assuming a log-Pearson Type III distribution). There were little differences between uses of the various combinations, but the mean consistently gave the better results. Only results for it are presented here since it is representative of the other combination methods. GLERL then compared each deterministic forecast with what actually occurred to find the effects of considering antecedent moisture and heat storage conditions, and considering meteorological outlooks.

More recently, GLERL updated all data sets and repeated all of these calculations in a second evaluation, for the period 1996-2000, to take advantage of data available since August 1997 and to extend the observations made in the first evaluation. While archived meteorological data were used in the simulated forecasts, all data from 1996 onward was actually provisional data received in near real time, as GLERL made their actual forecasts, amended with later corrections as they

were received. Thus, the provisional data set archived at GLERL and used in the 1996-2000 evaluation contains some corrections not available at the time of the actual forecasts. Therefore, simulating forecasts with this data set is not exactly equivalent to forecasting in near real time; however, the evaluated goodness of the forecast can be regarded as the “potential” possible with the present near-real-time data delivery system if no errors (recognized after the fact) occur. Results are presented herein for the first time for all of the deterministic evaluations.

Deterministic Comparisons

GLERL compared first-month forecast monthly means and actual monthly average levels by using root mean square error (RMSE), bias, and sample correlation, $\hat{\rho}$.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (f_i - a_i)^2} \quad (1)$$

$$Bias = -\frac{1}{n} \sum_{i=1}^n (f_i - a_i) \quad (2)$$

$$\hat{\rho} = \frac{\frac{1}{n} \sum_{i=1}^n (f_i - \bar{f})(a_i - \bar{a})}{\sqrt{\frac{1}{n} \sum_{i=1}^n (f_i - \bar{f})^2} \sqrt{\frac{1}{n} \sum_{i=1}^n (a_i - \bar{a})^2}} \quad (3)$$

where f_i = first-month forecast monthly mean value i of n first-month forecasts, a_i = corresponding first-month actual monthly average lake level i their respective sample means are:

$$\bar{f} = \frac{1}{n} \sum_{i=1}^n f_i \quad (4)$$

$$\bar{a} = \frac{1}{n} \sum_{i=1}^n a_i \quad (5)$$

Forecasts were also compared to climatology (monthly-means from the historical record), used as a reference forecast. The climatological outlooks serve as benchmarks against which more sophisticated forecast methods can be compared. A skill measure was developed that aids in determining which methods best forecast extreme events. Climatology is also used in this measure to weight differences between forecasts and actual values to emphasize extremes.

$$Skill = \frac{\frac{1}{n} \sum_{i=1}^n |f_i - a_i| \frac{|a_i - \hat{\mu}_j|}{\hat{\sigma}_j}}{\frac{1}{n} \sum_{i=1}^n |\hat{\mu}_j - a_i| \frac{|a_i - \hat{\mu}_j|}{\hat{\sigma}_j}} \quad (6)$$

where j = month of the year (1-12) corresponding to observation i , $\hat{\mu}_j$ = reference historical mean (climatological) forecast value for month j , and $\hat{\sigma}_j$ = reference historical (climatological) standard deviation for month j . GLERL calculated statistics for forecast months 2-6 as well.

RMSE and skill are measures of the absolute differences between forecast and actual values; low values of each of these measures indicate better performance. Skill is weighted to reflect the differences at extreme values more than differences near normal values. Bias is a measure of the shift between the distributions of forecast and actual values. Correlation is a measure of how well the timing of variability is captured by the forecast method.

One important note about all results on Lake St. Clair in particular, and on Lake Erie to a lesser extent, is that ice formation drastically effects levels and is not accounted for in the operational hydrology methods considered here. Therefore, Lake St. Clair results, and Lake Erie to a lesser extent, should be viewed with caution when comparing forecast methods there.

There is significant improvement when forecasting NBS directly from current antecedent conditions each month instead of using historical NBS in the straightforward operational hydrology approach. Both evaluation periods for RMSE in the first two rows of Figure 1 reveal this. This is more apparent in the correlation plots in the last two rows of Figure 1. However, there appears to be very little improvement in the forecasts by considering available meteorological outlooks as far as RMSE and correlation are concerned. In fact, there are times when the use of meteorological outlooks slightly degrades RMSE and correlation. However, skill (which measures the ability to forecast non-central levels) does show improvement (lower values) when meteorological outlooks are used, as shown for the 1995-August 1997 evaluation in the first row of Figure 2. The 1996-2000 evaluation does not show this improvement in skill. Figures 1 and 2 demonstrate that improvements in forecasting resulting from considering antecedent conditions are much greater than those associated with using meteorological outlooks.

In terms of bias, the differences are not large and, on Lakes St. Clair and Erie, considering antecedent conditions actually increases bias; see the last two rows in Figure 2. This suggests a problem in the computation of component NBS from antecedent conditions on these lakes, undoubtedly related both to ice formation and to ignored water balance groundwater terms, as well as more poorly estimated evaporation (particularly on Lake St. Clair). Also, the positive bias in the third row of Figure 2 indicates that forecasts under-predict at all lags on all lakes in the 1995-August 1997 evaluation; since the bias is almost linear with lag for the 1995-August 1997 evaluation, a near-constant bias exists in forecasting NBS for all forecast months. For the 1996-2000 evaluation in the fourth row of Figure 2, bias is much closer to zero on all lakes than in the earlier evaluation, with the same problem as already noted on Lakes St. Clair and Erie.

Summary

Consideration of antecedent conditions greatly improves Great Lakes AHPS water level forecasts except in a few cases. Considering available meteorological outlooks generally improves estimation of extremes somewhat but has little effect overall; it may have more impact on a case-by-case basis. AHPS forecasts generally have lower RMSE, higher correlation, better skill, and lower maximum error than the US, Canadian, or jointly coordinated forecasts of lake levels (shown in Figure 3 only for 1995-August 1997 RMSE and Correlation). This suggests that AHPS has generally the smallest differences with actual levels, best captures the timing of variations of lake levels, and is most-consistently best at the extremes over different periods. However, AHPS was often more biased than at least one of the other three during earlier evaluations.

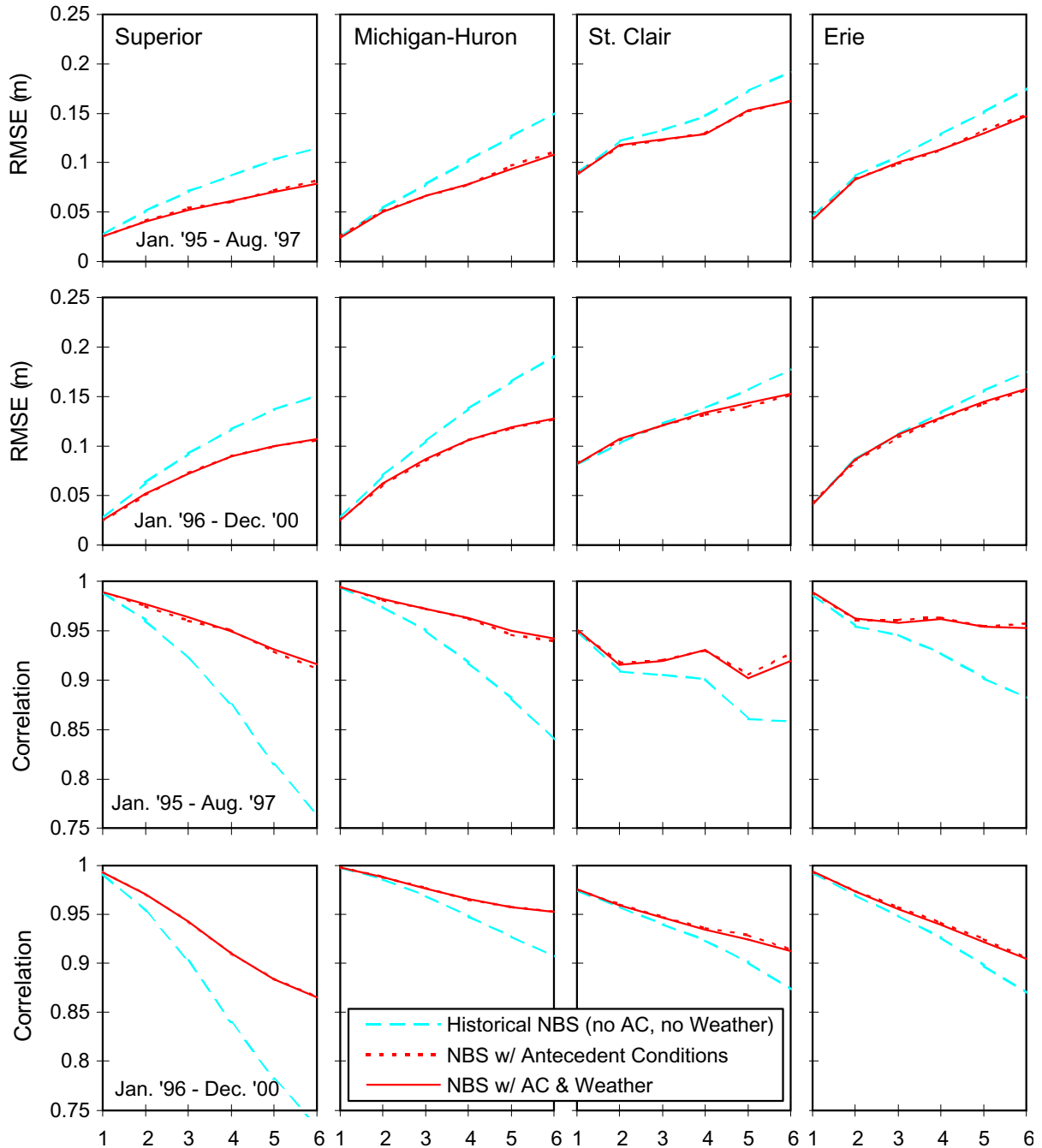


Figure 1. AHPS lake level forecasts RMSE and correlation vs. forecast month.

This suggests that it is generally under-predicting slightly during those times of high levels (both 1993-1995 and 1995-August 1997), but the other forecasts are less consistent from period to period. Results for 1996-2000 are similar.

GLERL’s probabilistic hydrologic outlooks are state-of-the-art. They a) fully and correctly utilize NOAA and others’ probabilistic long range meteorological outlooks for multiple areas simultaneously, b) explicitly account for basin soil moisture and snow pack and lake heat storage and ice cover initial conditions, c) allow daily extended outlook generation, taking advantage of

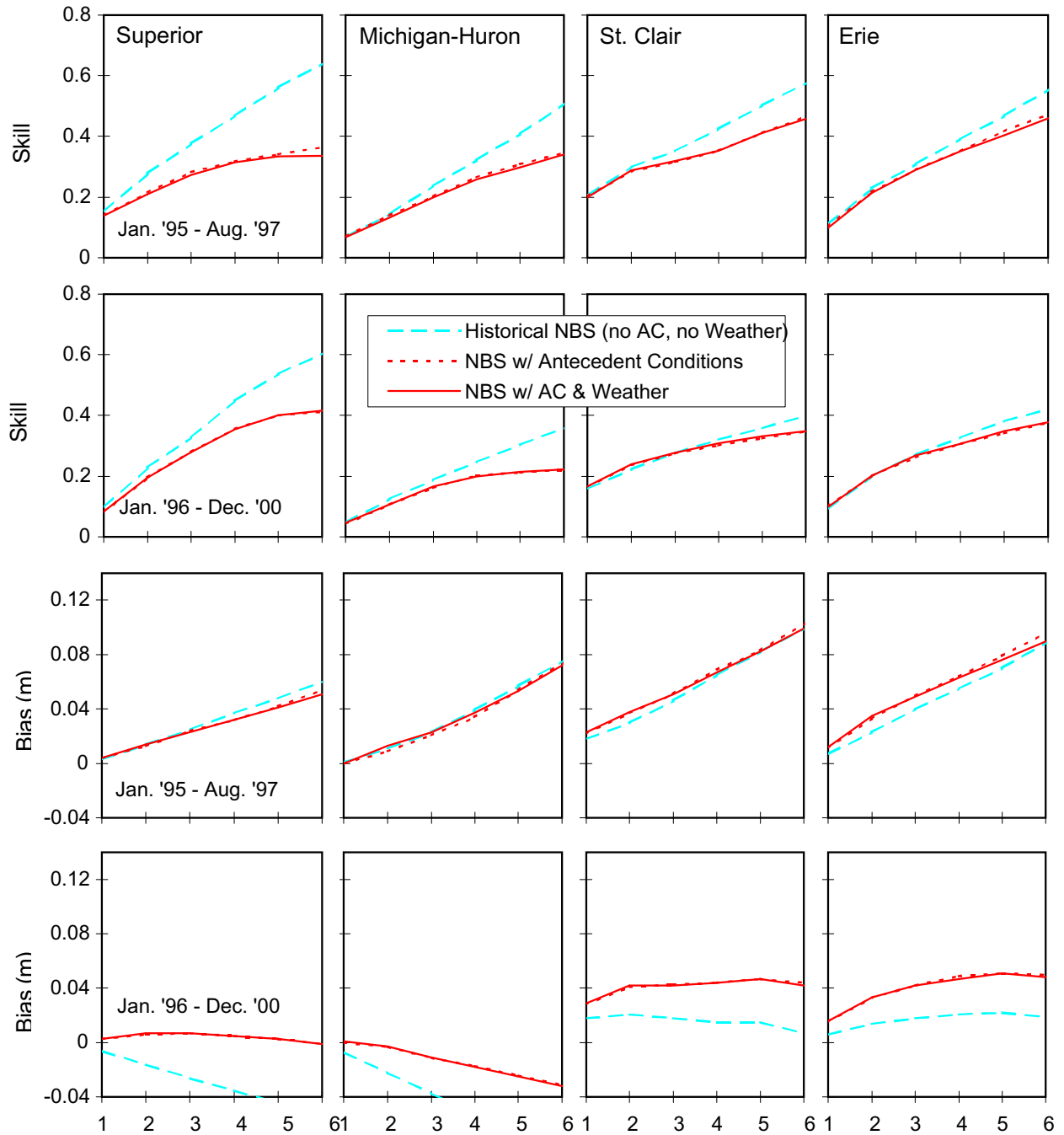


Figure 2. AHPS lake level forecast skew and bias vs. forecast month.

near-real-time data availability to offer continuously updated probabilistic outlooks, d) utilize hydrology models in a modularly-built package that allows upgrades to be “dropped in” as developed and tested, e) provide probabilistic outlooks for each lake and river watershed, capitalizing on improving weather prediction skill and hydrometeorological observations, f) properly consider the wide range of possibilities that always exist, g) incorporate some of the uncertainty inherent in forecast estimates, and h) allow decision makers to consider risk.

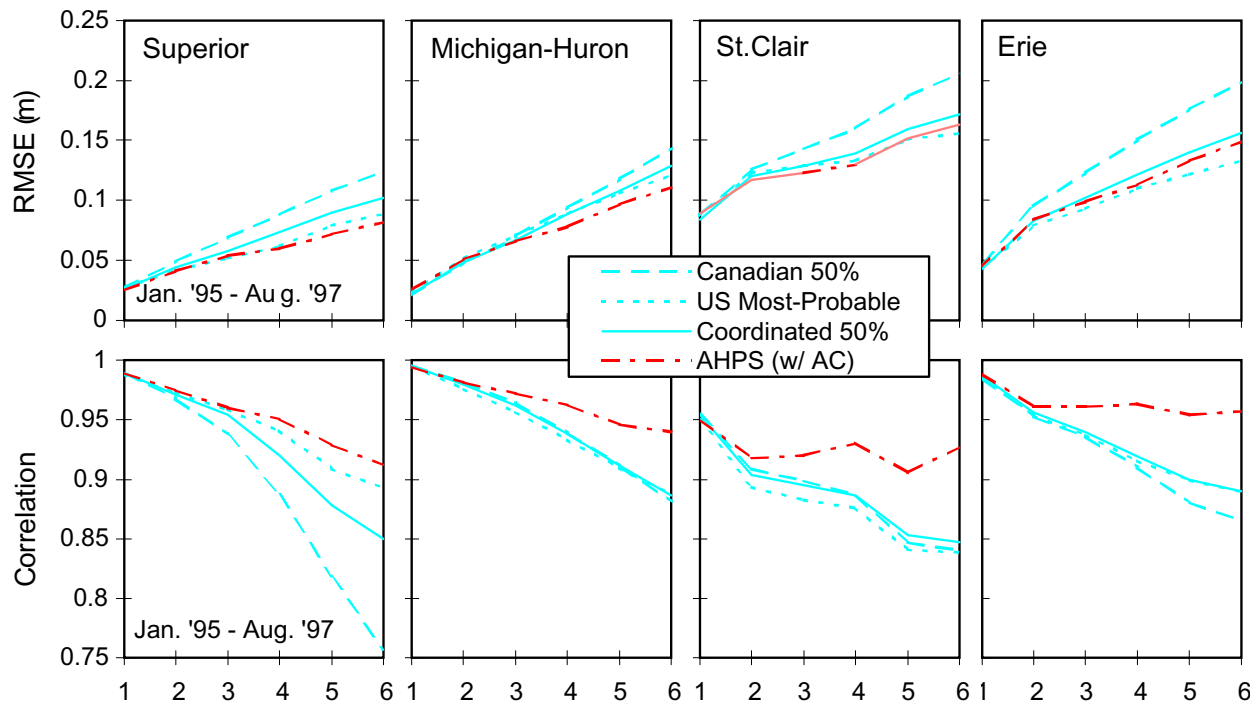


Figure 3. Selected AHPS. US. Canadian. and coordinated forecast statistics.

Acknowledgments

This is GLERL contribution no. 1236.

References

- Croley, T. E., II, 1993. Probabilistic Great Lakes hydrology outlooks, *Wat. Resour. Bull.*, AWRA, **29**(5):741-753.
- Croley, T. E., II, 1996. Using NOAA's new climate outlooks in operational hydrology, *J. Hydrol. Eng.*, ASCE, **1**(3):93-102.
- Croley, T. E., II, 1997. Mixing probabilistic meteorology outlooks in operational hydrology, *Journal of Hydrologic Engineering*, ASCE, **2**(4):161-168.
- Croley, T. E., II, 2000. *Using Meteorology Probability Forecasts in Operational Hydrology*. ASCE Press, American Society of Civil Engineers, Reston, Virginia, 214 pp.
- Croley, T.E., II, 2001. Climate-biased storm-frequency estimation. *Journal of Hydrologic Engineering*, ASCE, **6**(4):275-283.
- Croley, T. E., II, Hartmann, H. C., 1985. Resolving Thiessen polygons. *J. Hydrol.*, **76**:363-379.
- Croley, T.E., II, Lee, D. H., 1993. Evaluation of Great Lakes net basin supply forecasts. *Wat. Resour. Bull.*, **29**(2) 267-282.
- Croley, T. E., II, Quinn, F. H., Kunkel, K. E., Changnon, S. A., 1996. Climate transposition effects on the Great Lakes hydrological cycle. *NOAA Tech. Memo. ERL GLERL-89*, Great Lakes Environmental Research Laboratory, Ann Arbor, 100 pp.