WATER RESOURCES BULLETIN
VOL. 29, NO. 2 AMERICAN WATER RESOURCES ASSOCIATION APRIL 1993

EVALUATION OF GREAT LAKES NET BASIN SUPPLY FORECASTS[1]

Thomas E. Croley II and Deborah H. Lee[2]

ABSTRACT. Evaluation of the Great Lakes Environmental Research Laboratory's (GLERL's) physically-based monthly net basin supply forecast method reveals component errors and the effects of model improvements for use on the Laurentian Great Lakes. While designed for probabilistic outlooks, it is assessed for giving deterministic outlooks along with other net basin supply forecast methods of the U.S. Army Corps of Engineers and Environment Canada, and with a stochastic approach commissioned by the Corps. The methods are compared to a simple climatological forecast and to actual time series of net basin supplies. Actual net basin supplies are currently determined by estimating all components directly, instead of as water-balance residuals. This is judged more accurate and appropriate for both forecasting and simulation. GLERL's physically-based method forecasts component supplies while the other methods are based on residual supplies. These other methods should be rederived to be based on component supplies. For each of these other methods, differences between their outlooks and residual supplies are used as error estimates for the rederived methods and component supplies. The evaluations are made over a recent period of record high levels followed by a record drought. Net basin supply outlooks are better than climatology, and GLERL's physically-based method performs best with regard to either component or residual net basin supplies. Until advances are made in long-range climate outlooks, deterministic supply outlooks cannot be improved significantly.
(KEY TERMS: forecasting; hydrology; Great Lakes; water supply.)

INTRODUCTION

Six-month forecasts of Great Lakes water levels began in 1952 with the publication of a monthly bulletin of lake levels by the U.S. Lake Survey (DeCooke and Megerian, 1967). In 1975, the U.S. Army Corps of Engineers assumed this responsibility when the U.S. Lake Survey was abolished. Initially, based upon judgmental consideration of the factors that affect lake levels, the method was improved in following years with statistical relationships between these factors and monthly water supplies to the lakes (Noorbakhsh and Wilshaw, 1990). In Canada, the first six-month water-level forecast, based on probabilistic water supplies, appeared in the Canadian bulletin in 1973 (Southam and Yee, 1990). Today, the bulletins are produced by both the U.S. Army Corps of Engineers and the Canadian Hydrographic Service under the auspices of the Coordinating Committee on Great Lakes Basic Hydraulic and Hydrologic Data. Initially, commercial interests such as navigation and power entities were the major users of the bulletins. Today, about 8,000 copies of the U.S. bulletin and 2,600 copies of the Canadian bulletin are distributed monthly to a much wider audience including governmental agencies, commercial interests, recreational boaters, and riparians.

Great Lakes water-level forecasts are made by routing outlooks of water supplies through a multiple-channel flow routing model of the Great Lakes, given the initial lake levels and connecting channel flows. Although there is some uncertainty in the initial conditions at the time an outlook is made, the greatest uncertainty is in the upcoming water supplies. Efforts to improve forecasts at the Great Lakes Environmental Research Laboratory (GLERL) began in 1982 with the adaptation of their physically-based rainfall-runoff models to the estimation of net basin supplies. However, they used historical derived evaporation as forecast evaporation estimates. Subsequent evaluations (Croley and Hartmann, 1987; Hartmann and Croley, 1990) showed GLERL does an excellent job of modeling and forecasting monthly runoff; they identified the largest component errors as associated with lake evaporation estimation. They recently incorporated lake evaporation and heat storage models (Croley, 1989, 1992). The Corps also looked at improving their net basin supply forecasts by using

[1]Paper No. 93026 of the *Water Resources Bulletin*. **Discussions are open until December 1, 1993.**
[2]Research Hydrologists, Great Lakes Environmental Research Laboratory, National Oceanic and Atmospheric Administration, 2205 Commonwealth Boulevard, Ann Arbor, Michigan 48105-1593.

stochastic forecasts based on indirectly derived estimates (Buchberger, 1992).

Forecasts of Great Lakes monthly net basin supplies are evaluated here by comparisons with actual net basin supplies. The computation of "actual" net basin supplies is described first. The evaluations are over a recent time period when record high water levels were followed by a record drought; this period is described next. Then, GLERL's new forecast method is described and evaluated to determine component errors. These error estimates are used to assess forecast improvements consequent with use of the new evaporation model. Following this, other net basin supply forecast methods are described, and all methods are compared to a simple reference climatological method to measure relative skill levels between methods.

ACTUAL NET BASIN SUPPLIES

The water supplies to a lake, referred to as the net basin supplies (NBS), are defined as

$$NBS = P + R - E \quad (1)$$

where P is over-lake precipitation, R is basin runoff to the lake, and E is lake evaporation. NBS have been computed indirectly by both the Corps and Environment Canada as a residual in a water balance:

$$\Delta S = NBS + I - O + D \quad (2)$$

where ΔS is change in lake storage computed from lake-level changes, I and O are inter-basin inflow and outflow through a natural channel, respectively, and D is inter-basin diversion into the lake (negative for diversion out). Equation (2) ignores changes in storage due to thermal expansion, consumptive use, or groundwater contributions.

Historical monthly runoff data begins January 1908 (Superior), March 1901 (Michigan), April 1901 (Huron), May 1934 (St. Clair), January 1900 (Erie), and October 1915 (Ontario). Monthly precipitation data begins earlier than January 1900 for all lakes, but data are used only from that point on. Lake evaporation estimates can only be made starting in January 1948, as wind speed and humidity data are only generally available at that time. Thus, computed NBS from Equation (1) are available from January 1948. Residual NBS, from Equation (2), depend on connecting channel flows and water level data, which exist preceding January 1900; data are used only from that point on. All data exist through 1988.

Water balance errors can result in significant differences between NBS estimates computed from Equations (1) and (2). Table 1 summarizes the comparison of the two monthly NBS series over the evaluation period (August 1982 - December 1988, discussed subsequently). While correlation is good, there are large differences between the component and residual methods for all lakes, as reflected by the bias and root mean square error (RMSE) statistics. On all lakes, the RMSE is about one third of the standard deviation of component NBS. On Lakes Superior, Michigan, Huron, and Erie, the bias is about one tenth of the mean component NBS, and on Lake Ontario it is about 4 percent. Comparison plots of average seasonal characteristics of NBS from both methods (Lee, 1992a) reveal little difference in the timing of the seasonal maximum and minimum, but reveal significant differences in the monthly means during the late summer and fall months. Annual comparison plots (Lee, 1992a) show the residual NBS are consistently lower than the component NBS throughout the evaluation period used here for Lakes Superior, Michigan, Huron, and Ontario. This agrees well with the bias in Table 1.

Residual Method Errors

The differences between the two methods of computing NBS derive largely from residual errors in the

TABLE 1. Comparison of Residual with Component Monthly NBS Over the Evaluation Period (August 1982-December 1988).

	Superior	Mich-Hur	Erie	Ontario
Mean Component NBS (mm)	75	87	63	162
Mean Residual NBS (mm)	68	78	70	156
Standard Deviation Component NBS (mm)	64	60	98	108
Standard Deviation Residual NBS (mm)	63	66	102	118
RMSE (mm)	23	20	27	33
Correlation	0.94	0.96	0.97	0.96
Bias	7	9	-7	7

water balance used in Equation (2) as computed and coordinated by the U.S. Army Corps of Engineers and Environment Canada. These include errors in estimating change of storage, inter-basin inflows and outflows, and diversions, and errors in ignoring thermal volumetric changes, consumptive use, and groundwater. The change in storage on a lake is computed from the difference between the beginning-of-period and the end-of-period water levels (Lee, 1992a). Both are estimated with two-day averages of spatial means. Average estimates of instantaneous values and remnant short-term fluctuations cause errors in estimates of storage change and, ultimately, NBS. Croley (1987) evaluated errors in storage change for Lakes Superior and Erie and found the largest error in fall and winter months, when storm activity is most frequent.

Lake outflows are determined by direct measurement (Lakes Superior and Ontario), stage-discharge relationships (Lakes Michigan, Huron, and St. Clair), or a combination (Lake Erie) and are generally considered accurate within 5 percent. However, small outflow errors can result in large errors in residual NBS. Quinn and Guerra (1986) have shown that a 5 percent error in the Detroit or Niagara River flows can result in a 34 percent error in the residual NBS of Lake Erie. Errors in outflows estimated for Lakes Michigan, Huron, and St. Clair may exceed 5 percent during ice jams, since ice retardation must often be estimated. Other sources of large errors in flow estimation include flow reversals on the Detroit River during seiche activity on Lake Erie, past dredging activities in the St. Clair River, and negotiated changes in flow rating equations.

Some diversion into Lake Superior is measured before it enters the basin even though it passes through regulated Lake Nipigon before it enters Lake Superior. This is because the timing and amount of diverted water from Lake Nipigon cannot be separated from its natural flow. Likewise, diversions from Lake Michigan, in Equation (2), include actual withdrawal from the lake and redirected lake drainage (IJC, 1985), even though only lake withdrawal should be used. Both cause errors in the residual NBS estimates.

Omissions in Equation (2) further contribute to error in residual NBS estimation. Ignoring thermal expansion or contraction of the lake water body can result in 100 percent errors in monthly, residual NBS (Meredith, 1975) on each lake. This difference is largest for June through November, when the lakes' heat storage increases and then declines. Quinn and Guerra (1986) verified this in a Lake Erie water balance. Ignoring the total consumptive use [about 2,850 cfs in 1989, (GLC, 1989)] will affect residual NBS (computed for the entire Great Lakes Basin) as much as omitting the Lake Michigan diversion; they are about the same magnitude. Groundwater contributions to the Great Lakes are generally taken as small (Great Lakes Basin Commission, 1975) and conveniently ignored since they are unknown. However, any groundwater errors are directly reflected in the residual NBS estimation.

Component Method Errors

There are, of course, also errors associated with the component method of estimating NBS. These include errors in estimating precipitation, runoff, and evaporation. Over-lake precipitation estimates can be different, depending on which land-based meteorologic stations are used and how lake/land precipitation ratios are determined from short-term studies. For the Great Lakes, where lake effects on near-shore meteorology are significant and the drainage basins have relatively low relief, the use of all available meteorologic stations throughout the basin may be less biased than the use of only near-shore stations. Over-lake precipitation is estimated here from overland precipitation. It is, in turn, measured throughout the basin and Thiessen weighted. Direct measurement of over-lake precipitation may be possible in the future with the implementation of the Next Generation Radar (NEXRAD) as part of the National Weather Service modernization program.

Runoff into the Great Lakes is available from the United States Geological Survey and the Water Resources Branch of Environment Canada. Daily runoff values provided by these agencies were summed for each watershed within a lake basin. The runoff was extrapolated over ungauged areas; between 22 percent and 43 percent of the Great Lakes basins remain ungauged (Lee, 1992a) and runoff error potential exists. Drainage from the Chicago area is added to Lake Michigan and should be redirected; however, in the use of NBS in lake level routing, the entire Lake Michigan diversion (including both lake outflow and redirected Chicago drainage) is removed. In recognition of this typical use, Chicago drainage is not redirected in the computation of NBS components, as is the case with the residual NBS.

While data generally exist for basin runoff and over-lake precipitation, lake evaporation is not measurable, but now may be modeled. With the advent of improved surface thermodynamic and lake heat-storage models, it is now possible to estimate lake evaporation fairly accurately (Croley, 1989, 1992). Evaporation errors are not assessable (no lake-wide measurements exist), but the model does quite well in reproducing other observable quantities (surface water temperatures, water temperatures at depth, ice cover, dimictic turnover timing, and so forth).

Forecast Error Evaluations

Based on our review of the sources and magnitudes of errors in the computation of NBS, we believe that actual NBS are determined more accurately now by estimating all components directly, instead of as a water balance residual. Thus, the appropriate goal is taken here to be to generate forecasts of component NBS, rather than residual NBS. Likewise, the appropriate goal for forecast evaluation should be to compare methods of forecasting component supplies with actual component supplies. GLERL's physically-based method forecasts component supplies, while the other methods are derived from and forecast residual supplies. These other methods should be rederived for component supplies. For the GLERL method, comparison is made to component NBS. For each of the other methods, comparison is made to residual NBS. It seems reasonable to assume that the errors of forecasting residual supplies with these other methods would be about the same as the errors of forecasting component supplies with the rederived methods, since they are all statistical methods. Comparisons are also made between all methods and component supplies and between all methods and residual supplies, to round out the comparisons.

EVALUATION PERIOD

Recent extreme hydrometeorological events are considered as a rigorous test for the forecast techniques. Forecasts were evaluated from August 1982 through December 1988; this period was chosen because of the availability of all information upon which the forecast techniques depend. This period includes an extremely wet period from January 1984 through October 1986, and an extremely dry period from November 1986 through June 1988. For the evaluation period, supplies to the Great Lakes basin averaged 12 percent above normal. But, for the wet and dry periods, supplies averaged 131 percent and 63 percent of normal, respectively (Southam and Yee, 1990). The wet period concluded 10 years of above-normal annual precipitation. The highest annual precipitation on record (1900-1989) occurred in 1985, and the highest monthly precipitation occurred September 1986. The above-normal precipitation trend ended abruptly toward the end of 1986, and this century's worst drought occurred in the Great Lakes basin from November 1986 through June 1987. Although precipitation was above normal the second half of 1987, near-record evaporation occurred, reducing supplies. Dry weather returned in the spring of 1988 with the driest April-to-June period on record (Lee and Noorbakhsh, 1990).

GREAT LAKES NET BASIN SUPPLY FORECAST PACKAGE

Package Description

The advent of real-time hydrometeorologic data collection and transmission, and the increased accessibility to and computational power of computers, have enabled outlooks of monthly NBS with physically-based conceptual models of the hydrologic process. GLERL developed such a package for use on the Great Lakes. It incorporates near real-time data handling algorithms, a large-basin runoff model applied to all 121 watersheds draining into the Great Lakes, a lake evaporation and heat storage model applied to each of the Great Lakes, and meteorologic (climate) outlooks (Croley and Hartmann, 1987, 1990). The forecast package can generate deterministic or probabilistic outlooks of NBS several months into the future (Croley, 1993). The U.S. Army Corps of Engineers currently use it in an experimental mode; the Midwest Climate Center and the New York Power Authority use it in an operational mode.

The forecast package integrates hydrologic modeling and near real-time data reduction. The basic inputs (after calibration for a specific application) are daily values of precipitation, air temperature, humidity, wind speed, and cloud cover for all available stations throughout the lake basin. Optional input consists of measurements of snow water equivalent, soil moisture, and surface water temperature. The point data are converted to areal averages for each watershed and lake surface by Thiessen weighting. The areal averages are then utilized by the runoff and the lake evaporation and heat storage models to estimate basin moisture and lake heat as initial conditions to a forecast. Sections of the historical record then are selected, for input to the models as possible future scenarios, which best "match" the probabilistic 30-day and 90-day climate outlooks published by the U.S. National Weather Service (NWS) for each lake (Climate Analysis Center, 1990).

Historical meteorologic sequences, each beginning on the same day-of-the-year as a desired forecast start, are selected by grouping in five ways. The first grouping (group A) uses all years of the historical meteorology from 1948 to 1988. Group B uses historical meteorology sequences that match the 30-day temperature outlook. Group C uses sequences that match the 30-day temperature and precipitation outlooks in

TABLE 2. Meteorological Stations Available August 1982 for Simulated Operational Forecasting.

Lake Basin	Total	AFOS[a] (US)	AFOS[a] (Canada)	NCDC[b] (US)	AES[c] (Canada)
Over-Land Meteorology Stations[d]					
Superior	109	6	9	67	27
Michigan	161	19		142	
Huron	162	6	9	49	98
St. Clair	62	1	3	15	43
Erie	153	7	4	81	61
Ontario	234	4	4	89	137
Over-Lake Meteorology Stations[e]					
Superior	6	2	4		
Michigan	4	4			
Huron	5	1	4		
St. Clair	2	1	1		
Erie	6	4	2		
Ontario	8	2	6		

[a] "Automation Of Field Operations and Services" (AFOS) data are available immediately through connection with the National Weather Service system.

[b] National Climatic Date Center (NCDC) data are available two months after observation.

[c] Atmospheric Environment Service (AES) data are available six months after observation.

[d] Daily over-land minimum and maximum air temperatures and precipitation.

[e] Over-land stations representing daily over-lake air temperature, humidity, wind speed, and cloud cover.

the winter and spring months or the 30-day and 90-day temperature outlooks in the summer and fall months. Group D uses sequences that match the 30-day temperature and precipitation outlooks and the 90-day temperature outlook. Group E uses all climate outlook information by selecting sequences that match all: the 30-day and 90-day temperature and 30-day and 90-day precipitation outlooks. A historical meteorologic sequence is defined as "matching" (Croley, 1993) when 1) more than 30 percent of its (daily) values are less than the 30 percent quantile if so forecast by NWS, 2) more than 30 percent of it is more than the 70 percent quantile if so forecast by NWS, or 3) more than 40 percent of it is between the 30 percent and 70 percent quantiles if so forecast by NWS.

For each basin, the historical daily data sequences of each grouping are used with the models to compute corresponding groups of NBS scenarios. A probabilistic outlook then may be made (Croley, 1993) or a deterministic outlook may be made by either averaging these supply scenarios or selecting one of them in some manner. A probabilistic outlook makes use of all information in a meaningful way (Croley, 1993), but a deterministic outlook can be easily compared to outlook methodologies of other agencies. Unfortunately, a deterministic outlook averages or otherwise ignores much of the climate outlook information that is present.

Deterministic Outlook Selection

The forecast procedure was used in a simulated operational mode. For each month of the evaluation period, reception from a near real-time data acquisition network was simulated. Table 2 illustrates the data available for the first month of the simulated outlook (August 1982) and is typical of data sets used in each month of the simulated outlooks and of actual operating conditions available today. Data were used in the simulated forecasting as they would be available in real time. Thus, only 15 stations were used to estimate over-land hydrology, and 6 stations were used to estimate over-lake conditions for the Superior August 1982 outlook. However, the additional data available later for August 1982 were used (when they became available) to update the model estimates of initial conditions for later simulated outlooks.

The outlook sequences, resulting from the meteorological sequences selected in each group, were "combined" in three ways to yield a deterministic outlook

for each group: the average, the median, and the mid-interval of the range (average of the maximum and minimum). Each way of combining the scenarios to make a deterministic outlook was assessed in the following manner. Each of the outlook variables (runoff, over-lake precipitation, lake evaporation, NBS, and over-lake air temperature) was compared to actual values for each month of the outlook over the evaluation period. "Actual values" refers to measured runoff, observed precipitation, modeled actual evaporation, and component NBS [computed from the components with Equation (1)].

We computed the mean and variance of both outlook and actual data for each of the NBS components. We compared both by computing RMSE, correlation, and bias between them for each estimator. A table of comparison statistics was generated for each lake basin (Superior, Michigan, Huron, St. Clair, Erie, and Ontario), for each evaluation period (wet, dry, and full), for each group of historical meteorologic segments (groups A, B, C, D, and E), and for each method of selecting a single deterministic outlook from a group (average, median, and mid-interval). Thus, 270 tables were generated to compare each deterministic GLERL outlook method to actual data. The method of selecting a single deterministic outlook from a group seemed to make little difference. Use of the mean, as opposed to the mid-interval or median, gave marginally better results. Only results based on the mean deterministic scenario are considered further here.

FORECAST PACKAGE ERROR EVALUATION

Component Error Evaluation

Statistics of comparison with actual data, for each outlook group, are summarized in the first column of Table 3 for the one-month NBS components outlooks on Lake Superior. Column 1 of Table 3 was constructed by taking one-month (first-month) outlook statistics over the full evaluation period for average group sequences (groups A, B, C, D, and E) from the 270 tables described above. Deterministic outlooks derived from climatology [mean of measured runoff, measured precipitation, modeled evaporation, and component NBS from Equation (1) over the historical period for each month] and "perfect" forecasting (using actual meteorology) are also summarized in column 1 of Table 3.

From the first column of Table 3, each GLERL deterministic outlook (each group) appears roughly comparable to climatology. However, climatology statistics mean little for short periods. While group A outlooks have lower RMSE than climatology, bias is generally higher. On other lakes, climatology is much poorer than the various groups. All components (except precipitation) of all groups have lower RMSE than climatology for the other lakes. Not shown in Table 3 are statistics that measure how well extremes are predicted, which is when climatology suffers; this is discussed later.

The comparison statistics in the first column of Table 3 can be interpreted as forecast errors and associated with different components of the forecast package. Forecast errors can be broken into two groups: 1) model and data errors and 2) weather forecast errors. Model and data errors (type 1) are runoff model error, evaporation and heat storage model error, and data measurement errors. Weather forecast errors (type 2) are NWS forecast error, errors in matching NWS outlooks, and selection error in choosing sequences from a limited historical sample.

We can estimate type 1 errors by eliminating weather forecast errors (type 2), comparing a "perfect" outlook with actual components. A "perfect" outlook uses actual meteorology to predict NBS components. This is done at the bottom of column 1 of Table 3. Runoff model error appears to be very small in the Table 3 comparison of the perfect outlook with actual monthly values. The RMSE (8 mm) is about a third of the normal standard deviation, correlation is very high (0.94), and bias (2 mm) is about 4 percent of the normal mean. Note that the perfect precipitation outlook in column 1 has a RMSE of 0 mm, a correlation of 1, and a bias of 0 mm. This is because precipitation is not modeled and there is no corresponding model error. While forecast meteorology is transformed through models to forecast runoff or evaporation, forecast over-lake precipitation comes directly from forecast meteorology (no models). Furthermore, a consequence of comparing outlook components with component NBS, computed with Equation (1), is that evaporation model and data errors are not revealed. They appear in the column 1 perfect outlook to be almost nothing (almost zero bias and RMSE and unit correlation), since the evaporation model is used both in the outlooks and in the computation of actual NBS. The alternative, to compare outlooks with residual NBS, was objectionable since: 1) the outlooks are designed for component supplies and there are large data errors in residual supplies, and 2) resulting estimates of type-1 model and data errors would contain these large residual supply data errors. Thus, evaporation model errors are not assessable at present.

Likewise, we can estimate statistics of weather forecast errors (type 2) by comparing outlooks directly to the perfect outlook, as is done in the second column of Table 3. These statistics are derived in the same

TABLE 3. Lake Superior Comparison of One-Month Outlook Components with Actual Values[a] and with Perfect Outlooks[b].

Compared with Actual Values						Compared with Perfect Outlooks				
Error Statistic	Runoff	Precip.	Evap.	NBS	Air Temp.	Error Statistic	Runoff	Precip.	Evap.	NBS
Mean[c]	51	72	49	75	2.9	Mean (mm)	53	72	48	77
Standard Deviation[c]	22	33	41	64	11.5	Standard Deviation (mm)	20	33	41	64
GROUP A[d]						GROUP A				
RMSE[c]	10	24	12	32	5.7	RMSE (mm)	7	24	12	31
Correlation	0.89	0.72	0.96	0.88	0.87	Correlation	0.94	0.72	0.96	0.89
Bias[c]	0	4	-3	8	-0.2	Bias (mm)	2	4	-3	10
GROUP B[d]						GROUP B				
RMSE	11	25	13	35	5.5	RMSE (mm)	8	25	13	34
Correlation	0.88	0.69	0.95	0.86	0.89	Correlation	0.94	0.69	0.95	0.87
Bias	0	5	-1	7	-0.7	Bias (mm)	2	5	-2	9
GROUP C[d]						GROUP C				
RMSE	11	27	14	36	5.9	RMSE (mm)	8	27	14	36
Correlation	0.88	0.63	0.94	0.85	0.87	Correlation	0.94	0.63	0.95	0.86
Bias	0	6	-1	8	-0.7	Bias (mm)	2	6	-2	10
GROUP D[d]						GROUP D				
RMSE	12	28	15	38	6.1	RMSE (mm)	9	28	15	36
Correlation	0.86	0.56	0.94	0.83	0.86	Correlation	0.93	0.56	0.94	0.85
Bias	-1	2	0	2	-0.9	Bias (mm)	1	2	-1	4
GROUP E[d]						GROUP E				
RMSE	11	29	15	37	6.1	RMSE (mm)	8	29	15	36
Correlation	0.88	0.57	0.94	0.85	0.87	Correlation	0.94	0.57	0.94	0.87
Bias	0	5	0	5	-1.1	Bias (mm)	1	5	0	7
Climatological[e]						Climatological				
RMSE	16	23	13	33		RMSE (mm)	13	23	13	32
Correlation	0.72	0.75	0.95	0.86		Correlation	0.77	0.75	0.95	0.88
Bias	0	4	0	3		Bias (mm)	1	4	0	5
Perfect[f]										
RMSE	8	0	1	8						
Correlation	0.94	1.00	1.00	0.99						
Bias	-2	0	1	-2						

[a] "Actual" NBS is computed from component estimates and includes evaporation model error.

[b] "Perfect" outlooks use the actual meteorology.

[c] Units are mm for Runoff, Precipitation, Evaporation, and NBS, and °C for Air Temperature.

[d] Group A corresponds to use of all historical meteorologic sequences in the outlook generation; group B sequences match the National Weather Service 30-day temperature (T) outlooks; group C sequences match 30-D T and precipitation (P) (Winter & Spring forecasts) or 30-D T and 90-D T (Summer & Fall); group D sequences match 30-D T, 30-D P, and 90-D T; group E sequences match 30-D T, 30-D P, 90-D T, and 90-D P.

[e] "Climatological" outlooks are the long-term average monthly value for each month of the annual cycle.

[f] The compairson of perfect outlooks to actual values represents estimates of model and data errors in the forecast package.

manner as column 1 except that the perfect outlook is used in place of actual outcomes in the comparisons. Thus, differences between an outlook and the perfect outlook in the second column of Table 3 represent type-2 weather forecast errors.

To illustrate the relationships between the different types of errors depicted in columns 1 and 2 of Table 3, consider bias which is directly additive. For example, the Group A runoff total forecast bias (0 mm in column 1) is the sum of the Group A runoff model and data error bias (-2 mm in column 1 under "perfect" forecast) and the Group A runoff weather forecast error (2 mm in column 2), allowing for round off to the nearest millimeter. RMSE and correlation are not additive in this sense. Likewise, since there is no precipitation model (and no model error for precipitation, resulting in zero bias in column 1 under "perfect" forecast), the precipitation total forecast error statistics in column 1 are equivalent to the precipitation weather forecast errors in column 2 for each forecast method. Another relationship for bias in columns 1 and 2 is that NBS bias equals runoff bias plus precipitation bias minus evaporation bias, for every forecast method represented in those tables. This is a consequence of the NBS definition in Equation (1).

Column 1 of Table 3 reveals that the largest error component of all GLERL deterministic outlooks (all groups) on Lake Superior, is the precipitation outlook, in terms of both RMSE and bias. This is also true for both Lakes Michigan and Huron, except group C runoff bias exceeds precipitation bias. On Lakes St. Clair and Ontario, bias and RMSE are larger for runoff than for precipitation, while evaporation errors are uniformly lower. On Lake Erie, runoff and precipitation RMSE and bias are similar, and evaporation RMSE is still uniformly lower. Column 2 of Table 3 reveals that the precipitation error is all type-2 weather forecast error. Likewise, the bulk of runoff RMSE is due to type-2 weather forecast error. (Evaporation RMSE appears to be largely type-2 weather forecast error also, but remember that evaporation model and data errors cannot be estimated from these tables). We expect runoff and evaporation to have less total outlook error associated with them than with precipitation, since models allow consideration of initial (current) conditions at the beginning of an outlook. The "memory" of the system, preserved through storages of heat and moisture in the basin and lake, provides information on the first month of an outlook outside of the climate outlook. It appears now that precipitation weather forecast error is the largest component error of our NBS outlooks, and improvement of NWS climate outlooks will improve NBS outlooks more than model improvements.

GLERL Outlook Evaluations

The usefulness of the 30-day and 90-day climate outlooks was analyzed by comparing the five groups in column 1 of Table 3. The RMSE is minimum in column 1 for each component and for NBS for group A. It is important to note that use of group A corresponds to no use of NWS meteorology outlooks. However, the bias is large for group A as might be expected when ignoring climate outlooks. By using climate outlook information, the bias of both precipitation and evaporation drop low enough that the bias of the component NBS is more acceptable in group D. This is achieved at the expense of the RMSE, but it may be more important to minimize bias when forecasting extremes. The improvement (lowering) of bias when climate outlook information is used at the expense (raising) of RMSE was observed, in general, for other lakes and for other months of the outlooks (second, third, fourth, fifth, and sixth were looked at).

Group D does a much better job than group A in minimizing bias for each component and for NBS. In general, on other lakes, groups A or B minimize RMSE, and groups D or E minimize bias, with some exceptions. Group D uses almost all available climate outlook information regarded with any confidence (30- and 90-day temperature and 30-day precipitation outlooks; 90-day precipitation is not very useful since it is such a poor outlook). Bias is reduced then as the selected historical meteorological scenarios better match actual meteorology for group D than for group A. However, fewer scenarios are used in group D than in group A and variation in the mean is larger, resulting in larger outlook variations and RMSE for group D than for group A.

In terms of forecast accuracy, lower bias and higher RMSE correspond to more variable outlooks that are more accurate over the long term. Both statistics are important to consider. In general, low forecast bias is important, but, in the day-to-day operations of an agency, highly variable forecasts represent little useful information. The answer to this dilemma lies in a redesign of an agency's use of forecasts to avoid deterministic outlooks (which will always present this dilemma) and to embrace probabilistic outlooks. Although NWS climate outlooks do not greatly improve NBS components deterministic outlooks when used in the manner described above, Croley (1993) shows that they contribute useful information in the development of probabilistic outlooks.

Note also in column 1 of Table 3, there is some error compensation as components add to form NBS. This error compensation means that one cannot select the "best" (minimum-bias group say) method based on

NBS statistics alone. Indeed, the minimum-bias outlook of NBS in column 1 of Table 3 is group D, but neither runoff nor evaporation are forecast with minimum bias by group D. The additional improvement possible in both components is not recognized by basing the choice on NBS alone. Instead, an effort should be made to select the best outlook of each component.

Table 4 identifies which group outlook has minimum RMSE or minimum bias for each component (runoff, precipitation, and evaporation). NBS may be forecast as the algebraic sum of the best component outlooks in Table 4. For example, the "minimum-component-RMSE" method for Lake Superior uses group A runoff, precipitation, and evaporation, while the "minimum-component-bias" method for Lake Superior uses group B runoff, group D precipitation, and group E evaporation to compute the NBS outlook. In subsequent comparisons with other agencies, both the minimum-component-RMSE and minimum-component-bias methods are used as well as groups A through E.

TABLE 4. Minimum-Component-Error Groupings for Deterministic Outlooks.

	Runoff	Precipitation	Evaporation
Minimum-Component-RMSE Method			
Superior	A	A	A
Michigan	D	B	B
Huron	C	A	B
St. Clair	A	B	B
Erie	A	A	E
Ontario	A	A	A
Minimum-Component-BIAS Method			
Superior	B	D	E
Michigan	E	C	D
Huron	D	E	B
St. Clair	D	E	D
Erie	D	C	A
Ontario	D	C	C

OVERVIEW OF OTHER FORECAST METHODS

U.S. Army Corps of Engineers

The U.S. Army Corps of Engineers currently uses a combination of two statistical techniques to forecast monthly NBS for six months into the future: multiple linear regression and time series analysis. As described by Noorbakhsh and Wilshaw (1990), multiple linear regression is used to relate NBS to antecedent and forecasted precipitation and temperature and to antecedent NBS [computed as a residual from Equation (2)]. Time series analysis is used to extend the historical trend of NBS [computed as a residual from Equation (2)], adjusted for cyclical and seasonal variations, for up to 12 months. The forecaster considers recent actual supplies, the NWS 30-day and 90-day precipitation outlooks, and the performance of both the first-month trend and the regression forecasts made the previous month. Then, the forecaster decides whether to use the regression estimate, the first-month trend estimate, or a combination of the two. Often the regression and the first month trend values are averaged and used for the supply for the first month. NBS for the following five months are taken from the trend analysis and may be shifted to match the selected first month value, at the forecaster's prerogative.

Environment Canada

Environment Canada prepares the outlooks published by the Canadian Hydrographic Service. They base their six-month water-level outlook on a "probabilistic" outlook of NBS. They built three NBS outlook sequences for any six-month period (Dumont, 1990): the 5 percent, 50 percent, and 95 percent exceedance scenarios (sic). To derive these scenarios, they first computed the average residual NBS, for each month of the year, from the 1900-1986 historical record by Equation (2). They assumed that deviations from these averages were normally distributed and then estimated the distribution parameters. They identified the 5 percent, 50 percent and 90 percent exceedance quantiles for these deviations and added each to the monthly average for each month of the year. This gave them an annual cycle of 5 percent, 50 percent, and 95 percent exceedance sequences from which they then selected six-month sequences, one starting each month of the annual cycle, as their forecast NBS scenarios. These sequences do not change from year to year as they are independent of the current hydrologic conditions of the basin. The analysis of the Canadian forecast, presented here, is based upon the average of the 5 percent and 95 percent supply outlooks.

Univariate ARMA Model

Another NBS forecast method was developed during Phase II of the International Joint Commission's Levels Reference Study for the U.S. Army Corps of

Engineers (Buchberger, 1992). It is based on Box-Jenkins Autoregressive-Moving Average (ARMA) time series modeling. The model was calibrated with historical NBS from 1900-1986 and can generate 1- through 12-month outlooks, including confidence limits. Univariate ARMA (1,1) models were used to generate six-month outlooks of NBS for Lakes Superior, Michigan, Huron, and St. Clair, and univariate ARMA (2,0) models were used for Lakes Erie and Ontario. The only inputs required by the model are the previous month's NBS, and the previous month's forecast error.

METHOD OF FORECAST COMPARISON

As mentioned earlier, these other forecast methods should be rederived to be based on component supplies instead of on residual supplies. Since that is not possible here, we used differences between each method and residual supplies [computed by Equation (2) and coordinated between the U.S. Army Corps of Engineers and Environment Canada (U.S. Army Corps of Engineers, 1992)] as estimates of the differences that would exist between a rederived method (of forecasting component supplies) and the actual component supplies. Additionally, to counter criticism that might result from our use of inconsistent bases of comparison, we also calculated comparison statistics between GLERL's method and residual supplies to compare with the other agency methods. We also calculated comparison statistics between the other-agency methods and component supplies to further extend the comparisons. Evaluation statistics are RMSE, bias, correlation, and maximum error. These are computed between outlooks and actual values over the evaluation period.

Forecasts were also compared to climatology. For component NBS, the climatological outlook consists of the 1948-1988 monthly mean component supplies. For residual NBS, the reference outlook consists of the long-term (1900-1989) monthly mean coordinated residual supplies. The climatological outlooks serve as benchmarks against which more sophisticated forecast methods can be compared. A skill measure was developed for this comparison that aids in determining which methods best forecast extreme events. It is defined as the ratio of the weighted mean absolute difference between the actual value, X_i, and the outlook value, F_i, to that between the actual value and the reference outlook value, R_i:

$$\text{Skill} = \frac{\frac{1}{n}\sum_{i=1}^{n} P_i|X_i - F_i|}{\frac{1}{m}\sum_{i=1}^{m} P_i|X_i - R_i|} \tag{3}$$

where n and m are sizes of the outlook and reference data sets (sans missing data over the evaluation period), respectively. The weight P_i, is the absolute difference between the actual value and its long-term climatologic mean for month j (corresponding to observation i), μ_j, divided by the long-term climatologic standard deviation for month j, σ_j:

$$P_i = \frac{|X_i - \mu_j|}{\sigma_j} \tag{4}$$

Thus, a larger weight or penalty is applied to forecast errors as the actual event approaches the extremes of its range. A value of 0 indicates a perfect forecast while a value of 1 indicates that the forecast method has the same level of skill as the reference outlook. A value between 0 and 1 indicates the outlook has greater skill than the reference outlook, and a value greater than 1 indicates the outlook has less skill than the reference outlook (Lee, 1992b).

Evaluation statistics for each method were computed for each month of a six-month outlook; first-month NBS outlook comparisons are presented in Table 5 for the Corps method, the Canadian method, the ARMA method, the five GLERL groups (methods), and the two composite GLERL methods. Recall that the GLERL method statistics are compared to component supplies, and the other agency statistics are compared to residual supplies, on which they are based. Table 6 provides additional evaluation statistics for the GLERL methods based on comparisons to residual supplies. Table 7 provides additional evaluation statistics for the other agency methods based on comparisons to component supplies. Table 5 is offered as the best mechanism for comparing all methods, but Tables 6 and 7 are provided for completeness.

RESULTS AND DISCUSSION

In terms of RMSE, Table 5 reveals that NBS outlooks are most accurate for Lake Superior, followed by Lakes Michigan, Huron, Erie, Ontario, and St. Clair. This is not surprising since RMSE reflects natural variability of NBS, and they are more variable with

TABLE 5. Comparison of First-Month NBS Outlook Alternatives with Actual Supplies.

Method	Skill	RMSE	Correlation	Bias	Max. Error
Lake Superior					
Corps[a]	0.96	44	0.75	-2	132
Canadian[b]	1.00	44	0.75	3	131
ARMA[c]	0.92	42	0.77	4	132
Group A[d]	0.87	32	0.88	8	116
Group B	0.92	35	0.86	7	122
Group C	0.94	36	0.85	8	124
Group D	0.92	38	0.83	2	115
Group E	0.82	37	0.85	5	95
Min. BIAS[e]	0.81	35	0.87	3	111
Min. RMSE[f]	0.87	32	0.88	8	116
Lake Michigan-Huron					
Corps	0.94	51	0.68	2	214
Canadian	1.00	53	0.65	9	215
ARMA	0.93	51	0.69	10	224
Group A	0.81	46	0.71	14	205
Group B	0.76	43	0.74	13	192
Group C	0.76	44	0.73	10	175
Group D	0.74	45	0.71	3	170
Group E	0.83	49	0.68	-2	156
Min. BIAS	0.81	46	0.70	2	167
Min. RMSE	0.74	42	0.75	10	187
Lake St. Clair					
Corps	0.93	327	0.51	14	1052
Canadian	1.01	350	0.55	169	1125
ARMA	0.72	299	0.67	-78	821
Group A	0.77	270	0.82	80	955
Group B	0.76	274	0.81	84	889
Group C	0.78	275	0.79	63	841
Group D	0.85	312	0.72	56	1019
Group E	0.75	283	0.79	61	930
Min. BIAS	0.79	287	0.78	61	930
Min. RMSE	0.77	269	0.82	80	943
Lake Erie					
Corps	0.98	76	0.71	4	337
Canadian	1.00	76	0.72	18	326
ARMA	1.01	78	0.70	22	333
Group A	0.89	61	0.78	5	240
Group B	0.90	65	0.76	7	247
Group C	0.89	69	0.75	0	252
Group D	0.93	75	0.70	-4	279
Group E	0.71	68	0.77	-11	177
Min. BIAS	0.91	70	0.74	1	265
Min. RMSE	0.72	56	0.84	3	111
Lake Ontario					
Corps	0.92	82	0.75	3	256
Canadian	0.99	85	0.74	23	257
ARMA	0.92	81	0.76	24	255
Group A	0.83	72	0.78	19	209
Group B	0.80	72	0.78	22	217
Group C	0.78	71	0.77	13	204
Group D	0.82	79	0.68	4	216
Group E	0.94	82	0.67	3	271
Min. BIAS	0.79	73	0.73	10	208
Min. RMSE	0.83	72	0.78	19	209

[a] U.S. Army Corps of Engineers trend and regression analysis compared to residual supplies.

[b] Canadian forecasts (average of 5 percent and 95 percent probabilistic outlooks, prepared by Environment Canada) compared to residual supplies.

[c] Univariate ARMA Model of NBS commissioned by Corps compared to residual supplies.

[d] GLERL's grouping method compared to component supplies. Group A is all historical meteorological sequences; Group B sequences match the NWS 30-day temperature (T) outlooks; Group C match 30-D T and precipitation (P) (Winter & Spring forecasts) or 30-D T and 90-D T (Summer & Fall); Group D match 30-D T, 30-D P, and 90-D T; Group E match 30-D T, 30-D P, 90-D T, and 90-D P.

[e] GLERL's grouping that minimizes individual NBS component bias compared to component supplies.

[f] GLERL's grouping that minimizes individual NBS component RMSE compared to component supplies.

decreasing latitude. In fact, in terms of maximum error (largest absolute difference between an outlook and actual supplies), the variation with latitude is even more apparent. (Note it is not possible to compute residual supplies nor their comparison statistics on Lake Michigan or Lake Huron individually. Those lakes are treated as one lake hydraulically since flows between the two are unknown. Thus results are shown only for the combined Lake Michigan-Huron.) The order of increasing variability is Superior, Michigan-Huron, Ontario, Erie, and St. Clair. Lake St. Clair is anomalous since it has a small surface area with respect to its supplies; thus, supplies expressed as a depth over this lake are very large as are their natural variabilities. In terms of bias, Table 5 indicates that the picture is not so clear, largely because bias does not reflect natural variabilities in NBS.

Differences between the different forecast methods are immediately apparent in Table 5. Several gross comparisons are easily made between GLERL's seven methods (A, B, C, D, E, minimum-component-bias, and minimum-component-RMSE) and the others (Corps, Canadian, and ARMA). In terms of RMSE, GLERL's methods are uniformly better than the others with only two exceptions (group D on St. Clair and group E on Ontario). In terms of correlation, GLERL's methods are uniformly better with only five exceptions (group E on Michigan-Huron, group D on Erie, and groups D, E, and minimum-component-bias on Ontario). In terms of maximum error, GLERL's methods are better except for Lake St. Clair and group E on Ontario. On Lake St. Clair, the ARMA method has lowest maximum error, followed then by all GLERL methods. In terms of skill, GLERL's methods are better except for St. Clair, group C on Superior, and group E on Ontario. On Lake St. Clair, the ARMA method has lowest (best) skill, followed then by all GLERL methods. Finally, in terms of bias, the comparisons are more mixed; however, on Lakes Superior, Michigan-Huron, and Erie, one or more GLERL methods have the smallest bias. On Lake Ontario, the Corps method has slightly smaller bias, and on Lake St. Clair, both the Corps and the ARMA methods have smaller bias.

Thus, on Lakes Superior, Michigan-Huron, and Erie, it is possible to pick a GLERL method that is better, with respect to every statistic computed in Table 5, than the other-agency methods. These are group D on Superior, the minimum-component-bias method on Lake Michigan-Huron, and group C, the minimum-component-bias method, and the minimum-component-RMSE method on Lake Erie. With minor trade-offs, it is possible to use a GLERL method that is better in every respect but one than any of the other methods, on all of the Great Lakes.

By taking skill and bias as the most important criteria for selecting a forecast method, the following recommendations are made from Table 5. Use the minimum-component-bias method on Lake Superior; it has the best skill with low RMSE (not the lowest but better than the other agency methods) and very low bias. On Lake Michigan-Huron, either group D or the minimum-component-bias method offer combinations of good skill and bias. The group D method is recommended for both lakes since it offers very good skill and bias on all comparisons. On Lake Erie, use the minimum-component-RMSE method; it offers the best skill and very low bias while also offering the best RMSE, correlation, and maximum error of any method. On Lake Ontario, use the minimum-component-bias method; it is better than the other-agency methods in all respects except for Corps bias. It also has very good skill, RMSE, and maximum error. Its bias is not the best of the GLERL methods, but methods with better bias have unacceptably worse skill, RMSE, correlation, and maximum error.

On Lake St. Clair, while GLERL methods do well in terms of RMSE and correlation, it appears that the ARMA method offers the best skill, and the Corps method offers the best bias. However, groups C and E both offer better skill than the Corps method and better bias than the ARMA method, while giving better RMSE and correlation than all of the other-agency methods. Group C is recommended over Group E since the latter sometimes may be empty in practice (sometimes no historical sequence matches all NWS outlook parameters). To summarize, the recommended forecast methods (considering primarily skill and bias scores) are the minimum-component-bias methods for Lakes Superior and Ontario, the group D method for Lake Michigan-Huron, the group C method for Lake St. Clair, and the minimum-component-RMSE method for Lake Erie.

While the GLERL methods were derived from component NBS to forecast component supplies, it is interesting to assess their performance in the forecasting of residual supplies. Table 6 provides comparison statistics between the GLERL methods and residual supplies (refer to Table 5 for the other-agency statistics). In Table 6, we see that the recommended methods (identified above and underlined in Table 6) have better skill than the other-agency methods on all lakes. In fact, all GLERL methods still have better skill than all other-agency methods on all lakes, except for group D on Lake St. Clair. Also, the recommended methods have lower maximum error than the other-agency methods on all lakes except St. Clair. Bias is now seen to be good but not better than the best other-agency method on each lake except Ontario, where it is now best. Correlation is better for the recommended methods than for the other-agency

TABLE 6. Comparison of GLERL's First-Month NBS Outlooks with Residual Supplies.

Method	Skill	RMSE	Correlation	Bias	Max. Error
Lake Superior					
Group A[a]	0.87	41	0.79	0	139
Group B	0.89	42	0.78	-1	147
Group C	0.90	43	0.77	1	142
Group D	0.90	46	0.76	-6	127
Group E	0.77	45	0.78	-3	112
Min. BIAS[b]	0.81	43	0.79	-4	117[d]
Min. RMSE[c]	0.87	41	0.79	0	139
Lake Michigan-Huron					
Group A	0.85	48	0.70	5	175
Group B	0.81	45	0.73	3	163
Group C	0.77	45	0.74	0	145
Group D	0.78	49	0.70	-7	140
Group E	0.86	55	0.66	-12	127
Min. BIAS	0.84	51	0.69	-8	137
Min. RMSE	0.77	44	0.75	0	157
Lake St. Clair					
Group A	0.69	250	0.77	91	817
Group B	0.69	260	0.75	95	939
Group C	0.70	259	0.75	74	864
Group D	0.78	300	0.68	67	1078
Group E	0.62	262	0.77	80	813
Min. BIAS	0.66	268	0.75	80	903
Min. RMSE	0.69	249	0.77	91	814
Lake Erie					
Group A	0.92	72	0.72	11	322
Group B	0.93	75	0.70	13	329
Group C	0.93	79	0.69	7	335
Group D	0.98	84	0.64	4	362
Group E	0.71	74	0.73	-1	182
Min. BIAS	0.97	81	0.68	9	348
Min. RMSE	0.74	65	0.79	13	150
Lake Ontario					
Group A	0.81	77	0.78	12	222
Group B	0.80	79	0.77	15	220
Group C	0.76	79	0.76	6	210
Group D	0.81	89	0.65	-4	229
Group E	0.91	91	0.66	-4	284
Min. BIAS	0.79	82	0.71	2	221
Min. RMSE	0.81	77	0.78	12	222

[a] GLERL's grouping method compared to residual supplies. Group A is all historical meteorological sequences; Group B sequences match the NWS 30-day temperature (T) outlooks; Group C match 30-D T and precipitation (P) (Winter & Spring forecasts) or 30-D T and 90-D T (Summer & Fall); Group D match 30-D T, 30-D P, and 90-D T; Group E match 30-D T, 30-D P, 90-D T, and 90-D P.

[b] GLERL's grouping that minimizes individual NBS component bias compared to residual supplies.

[c] GLERL's grouping that minimizes individual NBS component RMSE compared to residual supplies.

[d] Recommended forecast methods, identified in the text from inspection of Table 6, are underlined here.

methods on all lakes except Ontario, where it is still acceptably close. The recommended methods' RMSE is better than the other-agency methods on Lakes Michigan-Huron, St. Clair, and Erie, and it is very close elsewhere. Thus, even when used as residual supply outlooks, the recommended methods are still the methods of choice.

Another interesting exercise is to consider the other-agency methods (derived from residual supplies to forecast residual supplies) as methods to forecast component supplies. Table 7 provides comparison statistics between the other-agency methods and component supplies (the recommended method statistics are repeated from Table 5 for convenience). Interestingly enough, the other-agency methods on both Lake Superior and Lake Erie have lower RMSE, higher correlation, and lower maximum error when compared to component supplies (Table 7) than when compared to residual supplies (Table 5), even though they were derived from residual supplies. However, they have poorer (larger) skill and higher bias on Lake Superior; skill is about the same and maximum error is also reduced on Lake Erie. On the other lakes, they are generally poorer when evaluated against component supplies than against residual supplies (with the exception of St. Clair bias for the Corps and Ontario maximum error). The recommended method is better with respect to every statistic than all of the

TABLE 7. Comparison of First-Month NBS Outlook Alternatives with Component Supplies.

Method	Skill	RMSE	Correlation	Bias	Max. Error
Lake Superior					
Corps[a]	1.02	39	0.81	5	125
Canadian[b]	1.07	38	0.83	10	121
ARMA[c]	0.97	36	0.85	11	118
Recommended[d]	0.81	35	0.87	3	111
Lake Michigan-Huron					
Corps	0.97	55	0.61	11	243
Canadian	0.98	56	0.61	18	245
ARMA	0.94	55	0.63	19	253
Recommended	0.74	45	0.71	3	170
Lake St. Clair					
Corps	1.04	370	0.52	4	1276
Canadian	1.12	399	0.54	185	1713
ARMA	0.81	337	0.67	-89	971
Recommended	0.78	275	0.79	63	841
Lake Erie					
Corps	1.00	70	0.75	-3	255
Canadian	1.00	67	0.76	12	244
ARMA	1.00	69	0.74	15	250
Recommended	0.72	56	0.84	3	111
Lake Ontario					
Corps	1.04	87	0.69	9	243
Canadian	1.07	86	0.71	30	244
ARMA	1.02	86	0.70	31	242
Recommended	0.79	73	0.73	10	208

[a] U.S. Army Corps of Engineers trend and regression analysis compared to component supplies.

[b] Canadian forecasts (average of 5 percent and 95 percent probabilistic outlooks, prepared by Environment Canada) compared to component supplies.

[c] Univariate ARMA Model of NBS commissioned by the Corps compared to component supplies.

[d] Recommended forecast method, identified in text from inspection of Table 6, compared to component supplies: the Minimum-Component-Bias method on Lakes Superior and Ontario, the Group D method on Lake Michigan-Huron, the Group C method on Lake St. Clair, and the Minimum-Component-RMSE method on Lake Erie.

TABLE 8. Comparison of Lake Superior Second-Month and Sixth-Month NBS Outlooks with Actual Supplies.

Method	Skill	RMSE	Correlation	Bias	Max. Error
Second-Month Outlooks					
Corps[a]	0.97	45	0.75	-2	131
Canadian[b]	1.01	45	0.75	3	131
ARMA[c]	0.94	42	0.77	5	132
Group A[d]	0.95	33	0.88	8	103
Group B	0.94	33	0.88	12	107
Group C	0.92	33	0.88	12	107
Group D	1.00	36	0.86	9	111
Group E	1.02	38	0.85	12	111
Min. BIAS[e]	1.00	36	0.85	7	110
Min. RMSE[f]	0.94	33	0.88	10	104
Sixth-Month Outlooks					
Corps	1.02	46	0.77	-10	134
Canadian	1.00	42	0.78	0	131
ARMA	0.99	42	0.78	2	135
Group A	1.04	33	0.88	4	96
Group B	1.06	34	0.87	5	102
Group C	1.09	35	0.86	3	115
Group D	1.11	38	0.85	8	121
Group E	1.15	40	0.84	6	107
Min. BIAS	1.10	34	0.87	5	101
Min. RMSE	1.05	33	0.88	4	100

[a] U.S. Army Corps of Engineers trend and regression analysis compared to residual supplies.

[b] Canadian forecasts (average of 5 percent and 95 percent probabilistic outlooks, prepared by Environment Canada) compared to residual supplies.

[c] Univariate ARMA Model of NBS commissioned by the Corps compared to residual supplies.

[d] GLERL's grouping method compared to component supplies. Group A is all historical meteorological sequences; Group B sequences match the NWS 30-day temperature (T) outlooks; Group C match 30-D T and precipitation (P) (Winter & Spring forecasts) or 30-D T and 90-D T (Summer & Fall); Group D match 30-D T, 30-D P, and 90-D T; Group E match 30-D T, 30-D P, 90-D T, and 90-D P.

[e] GLERL's grouping that minimizes individual NBS component bias compared to component supplies.

[f] GLERL's grouping that minimizes individual NBS component RMSE compared to component supplies.

other-agency methods on all lakes except for Corps bias on Lakes St. Clair and Erie (it is close on Erie).

Considering forecast performance beyond the first outlook month, the second- and sixth-month outlook evaluations are summarized for Lake Superior in Table 8. Generally, errors associated with any method are larger than for the first outlook month (compare Tables 5 and 8) and become obscured by sampling variability. However, while skills and biases are comparable for all methods for the second- through sixth-month outlooks, other statistics (RMSE, correlation, and maximum error) associated with all of the GLERL methods are better than those associated with all of the other-agency methods, respectively. It also appears that marginal skill may be observed during the second-month outlooks for the GLERL methods. Even so, the lack of improvement beyond the first outlook month is marked, and little useful information is contained in the outlooks; this suggests that the outlooks are not better than climatology for later-month outlooks after the second (and probably after the first).

CONCLUSIONS

There are significant differences between NBS computed from components or as water-balance residuals. This presents two problems: 1) selection of the best supply estimate (component or residual), and 2) comparison of other-agency forecast techniques (statistically derived from residual supplies) and GLERL's methods (based on component supplies). Regarding the first problem, there are large residual errors in residual supplies, that mask model and data errors, but we cannot estimate evaporation errors with component supplies. However, based on the sources and magnitudes of errors in NBS computations, we believe that component supply estimates are more accurate than residual estimates. Furthermore, residual supplies are an inappropriate goal of forecasting since improvements in models, data collection, or climate outlooks cannot be utilized fully as they become available in the future. Agencies, that derive NBS as a residual, should consider the component method and identify errors where there are large differences with the residual method. Computation of residual supplies should include consumptive use, thermal volumetric changes, and groundwater flows. Improvements in component supplies include ongoing studies of lake heat storage and evaporation, and expansion of the existing stream gage network in both Canada and the United States. (At a minimum, the current network should be maintained.) NEXRAD should be explored for over-lake precipitation measurement.

Regarding the second problem of comparing component and residual supply forecasts, all methods were compared with each other with regard to forecasting both. While derived to forecast component supplies, GLERL's recommended methods were chosen over other-agency methods for forecasting either component or residual supplies. We recommend that water supply forecasts be made using GLERL's method for groups C or D on every lake. The minimum-component-bias or minimum-component-RMSE methods can be adopted for Lakes Superior, Erie, and Ontario as recommended in the discussion. However, evaluation-period data were used to select those methods, and then they were evaluated against the same data. Thus, only groups A, B, C, D, and E were considered with other-agency methods in unbiased comparisons.

Precipitation forecast error now appears as the largest source of error in GLERL's NBS forecasts. Improvement of the NWS climate outlooks will improve NBS forecasts more than model improvements. This presents a dilemma to people using deterministic forecasts. How can their operations be improved when significant improvements in long-range weather forecasting do not appear imminent? The answer lies in the redesign of their use of forecasts to avoid deterministic outlooks, and to embrace probabilistic outlooks. Fuller use is made of climate outlook information, and the trade-offs of minimizing both RMSE and bias do not exist. Improved climate forecasting and effective use of forecasts in decisions are now the principal scientific challenges (Changnon, 1990).

ACKNOWLEDGMENTS

Ms. Gail Monds of the U.S. Army Corps of Engineers, Detroit District provided the NBS ARMA forecasts. This paper is GLERL contribution no. 823, supported in part by Phase II of the International Joint Commission Levels Reference Study.

LITERATURE CITED

Buchberger, S.G., 1992. Modeling and Forecasting Great Lakes Monthly Net Basin Supplies. Report prepared for U.S. Army Corps of Engineers, Detroit District, Detroit, Michigan.

Changnon, S.A., 1990. The Dilemma of Climatic and Hydrologic Forecasting for the Great Lakes. Proceedings of Great Lakes Water Level Forecasting and Statistics Symposium, Great Lakes Commission, Windsor, Ontario, May 17-18, 13-25.

Climate Analysis Center, 1990. Monthly and Seasonal Weather Outlook for May 1990. Monthly and Seasonal Weather Outlook 44(9), National Weather Service, Washington, DC.

Croley, T. E., II, 1987. Wind Set-Up Errors in Mean Lake Levels. Journal of Hydrology 92:223-243.

Croley, T. E., II, 1989. Verifiable Evaporation Modeling on the Laurentian Great Lakes. Water Resources Research 25(5):781-792.

Croley, T. E., II, 1992. Long-Term Heat Storage in the Great Lakes. Water Resources Research 28(1):69-81.

Croley, T.E., II, 1993. Probabilistic Outlooks of Great Lakes Hydrology. Water Resources Bulletin (submitted as companion paper).

Croley, T. E., II, and H.C. Hartmann, 1987. Near Real-Time Forecasting of Large Lake Supplies. Journal of Water Resources Planning and Management 113(6):810-823.

Croley, T.E., II, and H.C. Hartmann, 1990. GLERL's Near Real-Time Hydrological Outlook Package. Proceedings of Great Lakes Water Level Forecasting and Statistics Symposium, Great Lakes Commission, Windsor, Ontario, May 17-18, 63-72.

DeCooke, B.G. and E. Megerian, 1967. Forecasting the Levels of the Great Lakes. Water Resources Research 3:397-403.

Dumont, B.S., 1990. Forecasting Water Levels in the Great Lakes: A Probabilistic Method. Proceedings of Great Lakes Water Level Forecasting and Statistics Symposium, Great Lakes Commission, Windsor, Ontario, May 17-18, 41-52.

Great Lakes Basin Commission, 1975. Great Lakes Basin Framework Study - Appendix 11: Levels and Flows, Great Lakes Basin Commission, Ann Arbor, Michigan.

Great Lakes Commission, 1989. 1989 Annual Report of the Great Lakes Regional Water Use Data Base Repository, Appendix A (Draft). February, 1991.

Hartmann, H.C. and T.E. Croley II, 1990. An Evaluation of GLERL's Hydrological Outlook Package. Proceedings of Great Lakes Water Level Forecasting and Statistics Symposium, Great Lakes Commission, Windsor, Ontario, May 17-18, 177-196.

International Joint Commission, 1985. Great Lakes Diversions and Consumptive Uses Report. International Joint Commission, Windsor, Ontario.

Lee, D.H., and N.L. Noorbakhsh, 1990. An Evaluation of Great Lakes Water Level Forecasting. Proceedings of Great Lakes Water Level Forecasting and Statistics Symposium, Great Lakes Commission, Windsor, Ontario, May 17-18, 157-176.

Lee, D.H., 1992a. Computation of Net Basin Supplies: A Comparison of Two Methods. Final Report Subtask 19.1.2a (Scenarios Based Upon 1900-1989 Supplies), Task 19.1.2, Task Group 2, Working Committee 3, Phase II - International Joint Commission Levels Reference Study, Great Lakes Environmental Research Laboratory, Ann Arbor, Michigan.

Lee, D.H., 1992b. An Evaluation of Deterministic Great Lakes Water Supply Forecasts. Final Report Subtask 19.3.6a (Forecast Evaluation), Task 19.3.6, Task Group 2, Working Committee 3, Phase II - International Joint Commission Levels Reference Study, Great Lakes Environmental Research Laboratory, Ann Arbor, Michigan.

Meredith, D.M., 1975. Temperature Effects on Great Lakes Water Balance Studies. Water Resources Bulletin 11:60-68.

Noorbakhsh, N.L., and R. Wilshaw, 1990. Forecasting Water Levels in the Great Lakes Using Multiple Linear Regression and Time Series Analyses of Net Basin Supplies. Proceedings of Great Lakes Water Level Forecasting and Statistics Symposium, Great Lakes Commission, Windsor, Ontario, May 17-18, 53-62.

Quinn, F.H., and B. Guerra, 1986. Current Perspectives on the Lake Erie Water Balance. Journal Great Lakes Res. 12(2):109-116

Southam, C. F., and P.P. Yee, 1990. Review of the Probabilistic Forecasting method used by Environment Canada. Proceedings of Great Lakes Water Level Forecasting and Statistics Symposium, Great Lakes Commission, Windsor, Ontario, May 17-18, 145-156.

U.S. Army Corps of Engineers, 1992. Great Lakes Net Basin Supplies. Unpublished report, Detroit District, U.S. Army Corps of Engineers, Detroit, Michigan.