Journal of Great Lakes Research 37 (2011) 577–583

Contents lists available at ScienceDirect

Journal of Great Lakes Research

journal homepage: www.elsevier.com/locate/jglr

An appraisal of the Great Lakes advanced hydrologic prediction system

Andrew D. Gronewold *, Anne H. Clites, Timothy S. Hunter, Craig A. Stow

NOAA, Great Lakes Environmental Research Laboratory, Ann Arbor, Michigan, 48108, USA

ARTICLE INFO

Article history:
Received 3 March 2011
Accepted 26 May 2011
Available online 20 July 2011

Communicated by Leon Boegman

Keywords:
Water levels
Great Lakes
Forecasting
Uncertainty
Probabilistic model
Model verification

ABSTRACT

Great Lakes water level forecasts are used to inform decisions ranging from personal choices of recreational activities to corporate evaluations of alternative cargo transport options. For effective decision-making it is important that these model-based forecasts include an accurate expression of the forecast uncertainty, as well as information regarding the model forecasting skill. We provide an assessment of water level forecasts from 1997 through 2009 that were made using the National Oceanic and Atmospheric Administration (NOAA) Great Lakes Environmental Research Laboratory (GLERL) Advanced Hydrologic Prediction System (AHPS). A visual comparison between observed and forecast water levels suggests that AHPS generally captures seasonal and inter-annual patterns. A more quantitative assessment based on the percentage of observations within 90% prediction intervals, however, indicates that AHPS generally underestimates the observed variability of Great Lakes water levels. This assessment provides a benchmark for forecast performance against which alternative model structures (including future evolutions of AHPS) can be tested, and a basis to identify and prioritize the implementation of those alternatives. Including a calibrated model error term into the AHPS framework, to accommodate the underestimated variability, is a priority for short-term development and research, and represents one step toward more accurately quantifying forecast uncertainty. Our results also underscore the importance of storing historical forecasts and the data from which they were derived to serve as a basis for assessing model performance and prioritizing future model improvements.

Published by Elsevier B.V. on behalf of International Association for Great Lakes Research.

Introduction

The health, livelihood, and economic security of the Great Lakes region are highly dependent on the capacity of the Lakes to support commerce, navigation, and tourism, and to serve as a water supply source. Thus, accurately forecasting water levels of the Great Lakes is an important priority for research organizations and regulatory agencies (Grima and Wilson-Hodges, 1977; IJC, 1989). As an indication of this priority, the International Joint Commission (IJC), in 1993, recommended region-wide initiatives focused on improving methodologies for lake level monitoring, modeling, and forecasting to better plan for, and in some cases mitigate, extreme flooding, erosion, shoreline damage, and detrimental impacts to shipping and hydropower infrastructure (IJC, 1993). There is a large body of research, however, both before and after the 1993 IJC report, that underscores the importance of improving Great Lakes water level forecasting. This research includes the development and application of process models (Marchand et al., 1988), statistical models (including spectral analysis, autoregressive moving average-based models, and dynamic linear models, as described in Cohn and Robinson, 1976; Irvine and Eberhardt, 1992; Lamon and Stow, 2010, respectively), as well as comparisons among different models (Meredith, 1970) and analyses of models that propagate different climate change scenarios into future Great Lakes water level variability (Angel and Kunkel, 2010).

From a regional water resources management perspective, there are several pressing questions that need to be answered in the near future about the dynamics of Great Lakes water levels, including: "For how long, and to what extent, will the levels continue to decline relative to the record high levels observed in the mid-1980s?" (Sellinger et al., 2007), and "To what extent does forecast variability affect the perceived risk of undesirable outcomes in water level-based management decisions?" (for a historical perspective on this question, see Lee et al., 1997). In addition, in light of current economic pressures and limited resources, federal and other research agencies are asking, "What is the magnitude of uncertainty in Great Lakes water level forecasts, and what improvements to model algorithms and investments in monitoring infrastructure would best address and reduce that uncertainty?" (Eberhardt and Moin, 2009).

Answers to these questions are based, at least in part, on model forecasts. However, those forecasts should not only include an explicit expression of uncertainty and variability, but their performance (or "skill", as described in Stow et al., 2009) should also be assessed in a probabilistic framework that documents how accurately the model predicts water levels *with respect to* model uncertainty. Assessing the accuracy of the model forecast uncertainty is particularly important because these forecasts are used for decision support and an accurate assessment of uncertainty provides decision makers with the

* Corresponding author. Tel.: +1 734 741 2444; fax: +1 734 741 2055.
E-mail addresses: Drew.Gronewold@noaa.gov (A.D. Gronewold), Anne.Clites@noaa.gov (A.H. Clites), Tim.Hunter@noaa.gov (T.S. Hunter), Craig.Stow@noaa.gov (C.A. Stow).

0380-1330/$ – see front matter. Published by Elsevier B.V. on behalf of International Association for Great Lakes Research.
doi:10.1016/j.jglr.2011.06.010

information needed to appropriately hedge their decisions (Reckhow, 1994). If the uncertainty bounds are too wide, the forecasts might appear accurate because all observations are encompassed within the forecast uncertainty, but unnecessarily-wide intervals (or overly-conservative intervals) would be minimally informative to decision makers. Alternatively, narrow forecast intervals that rarely capture the observations would offer decision-makers a “false sense of security”, leading them to bad decisions. We find that previous studies assessing the performance of AHPS (Croley and Lee, 1993, for example) focus primarily on deterministic metrics of model skill.

To address this research gap, we provide an assessment of the National Oceanic and Atmospheric Administration (NOAA) Great Lakes Environmental Research Laboratory (GLERL) Advanced Hydrologic Prediction System (AHPS). AHPS is a comprehensive modeling framework designed to provide seasonal probabilistic forecasts across the Laurentian Great Lakes for numerous hydrometeorological variables including net basin supplies and water levels. We focus here on an assessment of water level forecasts because, in contrast to net basin supplies, water levels can be measured directly thereby providing a benchmark to compare against model forecasts.

It is important to note that NOAA, through its National Weather Service (NWS), maintains and operates a similarly-named Advanced Hydrologic Prediction Service (also referred to as AHPS), that provides stream flow forecasts at thousands of locations throughout the United States to support flood forecasting and the development of flood maps (for details, see McEnery et al., 2005). For the remainder of this paper, we will use the acronym AHPS to refer to the version of AHPS developed by GLERL, however developing and implementing a single integrated operating platform encompassing both systems is a reasonable and practical goal for future research, particularly in light of ongoing implementation of NOAA's Community Hydrologic Prediction System (commonly referred to as CHPS, for details see Schaake et al., 2006). Currently, (GLERL's) AHPS is used routinely by governmental agencies, hydropower companies, and other groups in both the United States and Canada to support official Great Lakes water level forecasts, forecast-based management decisions, and as a tool for understanding potential impacts of climate change (Croley, 1990; Croley et al., 1998).

An extensive body of literature documents the development of the individual components of AHPS and their integration into a single modular modeling framework along with analyses of those components and of the relative skill of AHPS in forecasting net basin supplies (Croley, 2006; Croley and Lee, 1993). Here, we provide (in the following section) a brief description of AHPS to support further discussion and analysis of AHPS forecasts, how they are generated, and what changes to AHPS might lead to improved forecasting skill. Our goal is to provide an assessment that serves as both a benchmark for water level model forecasting performance against which alternative model structures (including future evolutions of AHPS) can be tested, and as a basis for identifying and prioritizing the implementation of those alternatives. Future modifications may include, but are not limited to, alternative climate and meteorological data sources (Miller, 2009), models for compiling those data, interconnecting channel flow routing algorithms, and innovative approaches to predicting flows in ungauged basins (Kokkonen et al., 2003; Krueger et al., 2010).

Overview of the Great Lakes advanced hydrologic prediction system (AHPS)

AHPS is a Windows-based graphical user interface (GUI) application that is used daily both at GLERL and other public and private agencies in the Great Lakes region for forecasting Great Lakes water levels and a range of other hydrometeorological variables. It begins (Fig. 1) with a data mining procedure that extracts daily meteorological information from NOAA's National Climatic Data Center (NCDC). Subbasin and over-lake averages are calculated for each climatological variable using an automated Thiessen polygon-based weighting algorithm (Croley and Hartmann, 1985) and subsequently input to both a Lake Thermodynamic Model (Croley, 1992) and a conceptual rainfall–runoff model which, within AHPS, is referred to as the Large Basin Runoff Model (or LBRM, as described in Croley, 2002).

The Lake Thermodynamic Model (LTM) uses adjusted over-land (i.e. subbasin) data from available weather stations to estimate over-water meteorology based on air temperature, humidity, wind speed, and cloud cover (Croley, 1989). LTM-derived evaporation estimates are then combined with precipitation estimates from the Thiessen polygon algorithm and runoff estimates from the LBRM to estimate the net water supply into each lake (i.e. “net basin supply” or NBS). NBS estimates serve as input to the Large Lakes Routing and Regulation model which encodes the current Lake Superior regulation plan and simulates water levels for Lakes Superior, Michigan, Huron, Georgian Bay, Lake St. Clair, and Lake Erie. Water level forecasts based on historical climatology from a given period of record (for details, see Croley, 1997) are then combined in a weighted ensemble framework.

Outflows and water levels for Lake Ontario are not calculated in AHPS, in part because standard protocol in the Lake Ontario regulation plan is frequently adjusted to accommodate small changes in the St. Lawrence River flows. NBS values for Lake Ontario are produced and are publicly available (along with NBS and lake level calculations for the other Great Lakes) through GLERL.

Methods

To assess the forecasting skill of the current version of AHPS, we compared 13 years (from 1997 through 2009) of 3 and 6-month average monthly water level forecasts (i.e. forecasts for each month made 3 and 6 months prior) to the corresponding lake-wide monthly average of observed water levels for Lakes Superior, Michigan–Huron, St. Clair, and Erie. For example, monthly average observed water levels for July 2008 are compared to the 3-month (ahead) forecast made on the first day of May 2008, and to the 6-month (ahead) forecast made on the first day of February 2008. Lake-wide average monthly “observed” levels are based on gauging station data from both NOAA's National Ocean Service (NOS) Center for Operational Oceanographic Products and Services (CO-OPS), and the Canadian Hydrographic Service (CHS). Following protocol established by the U.S. Army Corps of Engineers (USACE) and Environment Canada (EC), and implemented through the Coordinating Committee on Great Lakes Basic Hydraulic and Hydrologic Data, we use the average of a subset of NOS CO-OPS and CHS measurements (Table 1) which are believed to provide the most reliable and accurate representation of the average monthly water levels in each lake (for details, see U.S. Army Corps of Engineers, 2008).

We note that these average water levels are also uncertain, and standard error-based uncertainty bounds (on average water level observations) could be included in our assessment, resulting in a comparison analogous to an ANOVA (a statistical technique for comparing means for multiple independent variables). However, at the monthly scale, with multiple stations and frequent observations on each lake, this uncertainty is small relative to the forecast uncertainty, so we did not incorporate average water level uncertainty into our analysis. More importantly, including observation uncertainty would not affect the pattern of any discrepancies that are revealed in our investigation, and identifying systematic deviations is one of our main goals.

AHPS generates probabilistic forecasts (Croley, 2003) of monthly average water levels in each of the Great Lakes for the next one to ten months based on current basin moisture conditions, NWS one and three-month climatic outlooks for precipitation and temperature, and historical weather and water level data. Water level (and other

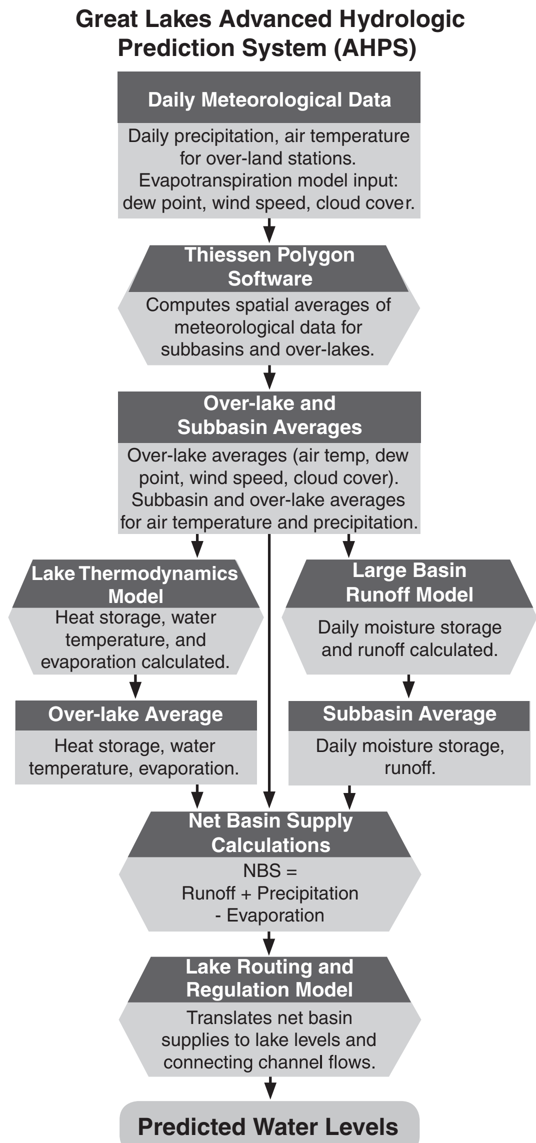

Fig. 1. Schematic representation of the Great Lakes Advanced Hydrologic Prediction System for forecasting water levels based on a single set of historical climatology (i.e. for a given period of record). Methodologies (and detailed process schematics) for combining individual water level forecasts as members of a complete water level ensemble forecast are described in Croley, 2003. In the schematic presented here, nodes represent data sources, data processing tools, and process-based models, and interconnecting arrows represent the flow of information between those nodes.

hydrometeorological data) forecasts are available to the general public through GLERL's web site however, until recently, forecasts were not archived.

To retrospectively forecast (or "hindcast") water levels, we used data from 1997 to 2009 and followed standard AHPS forecasting protocol with two modifications. First, outlook weights (Croley, 2003) used to generate the ensemble forecasts were recalculated using historical climatic outlooks available from the NWS National Center for Environmental Prediction (NCEP). Second, when "rebuilding" historical meteorological data, we used the most up-to-date station data because it would be excessively time consuming, if not impossible, to retrospectively determine *exactly* which data sources were available at the time of each monthly forecast. More specifically, over the period 1997 through 2009, additional data streams and quality assurance/quality control (QA/QC) procedures have been implemented, or become available, since the original forecasts were generated. These additional data sources and quality assurance measures are included in our "hindcast".

The importance of storing historical forecasts and the data used to generate them cannot be overemphasized, now that digital storage is not the challenge it was 20 years ago. As modeling methods improve and data uncertainty changes, ready access to past forecasts will allow continual monitoring of model performance, leading to continual improvement in the tools available.

For each probabilistic forecast we (following Croley, 2003) estimated 5%, 10%, 20%, 30%, 50%, 70%, 80%, 90%, and 95% quantiles of the water level probability distribution, and subsequently calculated the number and fraction of monthly average observed water level values between the 5 and 95% quantiles (also referred to as the inner 90% prediction interval), as well as the prediction bias (measured as the average difference between the median forecast water level and the monthly average observed water level).

Table 1
Gauging stations used to calculate Great Lakes monthly average water levels. The Fairport gauge was used (for this study) for the period between 1997 and 2007, and the Cleveland gauge was used from 2008 through 2010.

Lake	State or province	Station or locality name
Superior	Minnesota	Duluth
	Michigan	Marquette C.G.
	Michigan	Pt Iroquois
	Ontario	Michipicoten
	Ontario	Thunder Bay
Michigan–Huron	Michigan	Harbor Beach
	Michigan	Mackinaw City
	Michigan	Ludington
	Wisconsin	Milwaukee
	Ontario	Thessalon
	Ontario	Tobermory
St.Clair	Michigan	St. Clair Shores
	Ontario	Belle River
Erie	Ohio	Toledo
	Ohio	Fairport (1997–2007)
	Ohio	Cleveland (2008–2010)
	Ontario	Port Stanley
	Ontario	Port Colborne

Results

A visual assessment of the comparison between observed monthly average water levels and corresponding AHPS forecasts (Fig. 2) suggests that AHPS generally captures seasonal and interannual trends. Furthermore, the results in Fig. 2 emphasize the significant role that climate forecast uncertainty plays in determining future lake levels because, at present, it is the only source of uncertainty currently explicitly acknowledged in AHPS Great Lakes water level forecasts.

The results (Fig. 2) also suggest potential variations in forecasting skill at different times of the year, at different periods over the past 13 years, and between the 3- and 6-month forecasts. These differences are directly related to changes in the forecast bias and the relative spread of the water level predictive distribution, and warrant a more robust probabilistic assessment of forecast uncertainty which, given our goal of establishing a performance benchmark for Great Lakes water level forecasting skill, is beyond the scope of this paper (but is certainly a goal for future research). For example, the 3-month forecast for Lake Erie between 1998 and 2000 (Fig. 3) is less biased than the corresponding 6-month forecast for the same time period, however the bias for both 3- and 6-month Lake Erie water level forecasts and the efficiency of the 90% prediction intervals are comparable across the entire time period assessed in this study.

Qualitative, visual model performance evaluations, however, are highly subjective (Allen et al., 2007); a more quantitative assessment of AHPS skill, based on the fraction of observations within 90% prediction intervals (Table 2), indicates that AHPS generally underestimates the observed variability in Great Lakes water levels from 1997 through 2009. For example, the 3-month and 6-month 90% prediction intervals in Lake Superior included 64% and 69% (respectively) of the observed monthly average water levels. In general, the 90% prediction intervals, across all lakes and across both 3- and 6-month forecasts, capture between 64 and 74% of observed water levels. Our assessment of forecasting bias (Table 2) indicates that AHPS, over the 1997 to 2009 time period, tends on average to slightly underestimate water levels in Lakes St. Clair and Erie, and to overestimate water levels in Lakes Michigan–Huron.

Discussion

Rigorous, comprehensive model skill assessment is relatively rare in the refereed literature (Arhonditsis and Brett, 2004; Stow et al., 2009), while excuses to forego uncertainty analysis are common

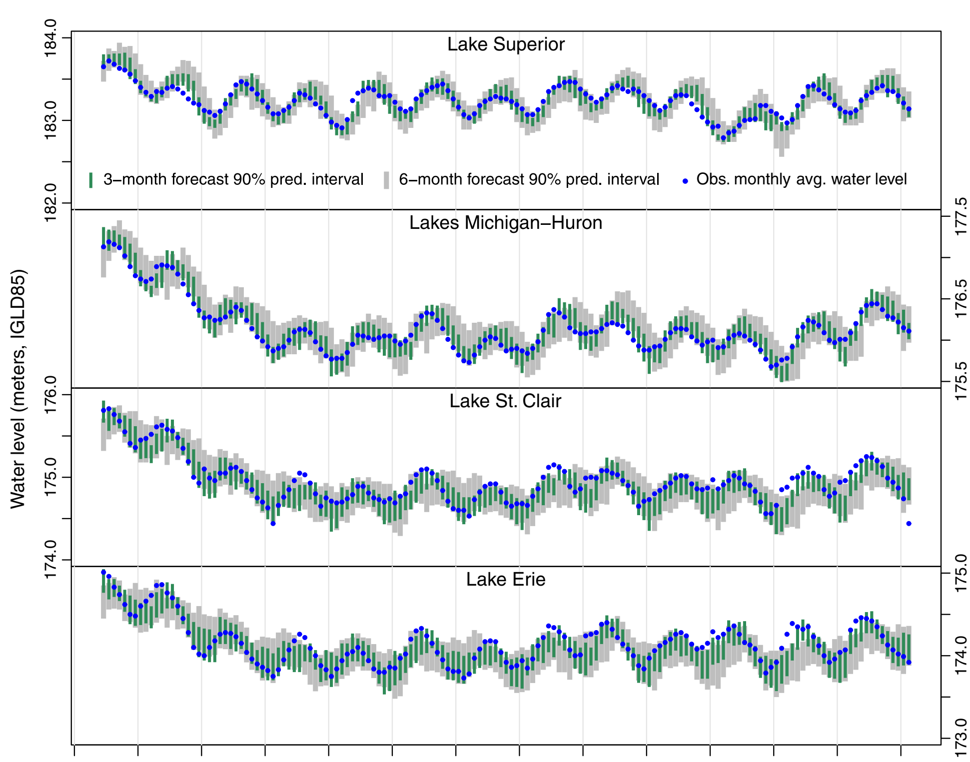

Fig. 2. Comparison between observed Great Lakes monthly average water levels between 1997 and early 2010, and corresponding AHPS 3- and 6-month forecasts. Note that the y-axis limits in each panel cover the same range (2.0 m) to facilitate cross-lake comparisons.

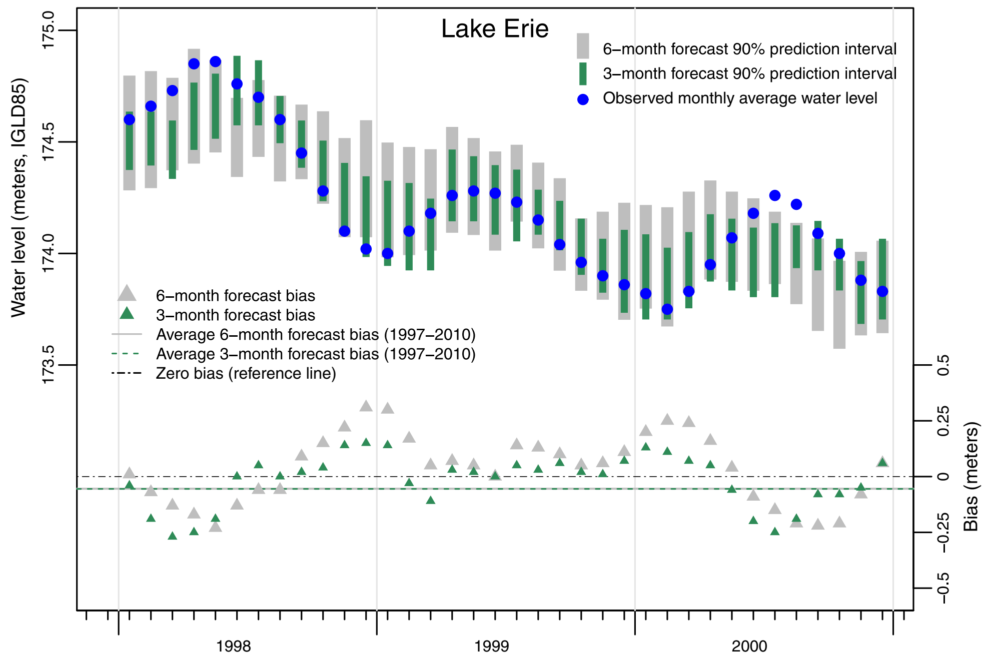

Fig. 3. Comparison between observed and forecast Lake Erie monthly water levels between 1998 and 2000, including the bias of each monthly forecast and the average monthly bias for the data record used in this study (1997–2010).

(Pappenberger and Beven, 2006). Without a careful evaluation of model skill, decision makers may assume either too much or too little credence in the model forecasts resulting in poor decisions and inappropriate actions (Gronewold et al., 2009). Additionally, a meticulous skill assessment is essential to reveal model weaknesses and identify areas where further effort should be directed to improve model performance and also result in a better understanding of the system being modeled. With these facts in mind, we have presented an initial analysis of the predictive performance of NOAA's Great Lakes AHPS, a comprehensive modular hydrological modeling system designed to forecast net basin supplies and water levels for the Laurentian Great Lakes within a probabilistic framework.

We have found that AHPS captures the main water level dynamics in four lake systems (i.e. Superior, Michigan–Huron, St. Clair and Erie) both 3 and 6 months into the future. However, of perhaps equal or greater importance is that our assessment accommodates AHPS' probabilistic forecasts by evaluating skill in light of its expression of uncertainty, and suggests that there are sources of bias and uncertainty that should be addressed in future evolutions of the AHPS modeling framework (for a related discussion of quantifying uncertainty in hydrological modeling, see Young, 2003).

In addition, we underscore the fact that many previous AHPS forecasting skill assessments are based on AHPS NBS calculations, and on comparisons between AHPS NBS calculations and NBS calculations derived by the USACE and EC (Croley, 2006).While these comparisons provide insight into the relative performance of each modeling approach, we argue that NBS forecasts do not necessarily provide a clear basis for assessing model forecasting skill because NBS values are not directly observable and because they can be derived through different methods (i.e. residual or component methods, as described in Croley and Hunter, 1994) each with different intrinsic sources of uncertainty, bias, and variability.

One general conclusion from previous NBS-based studies, however, is that AHPS is negatively biased during periods with relatively high water levels (Croley, 2002). The results we present in this paper, however, which are based on an assessment during relatively low water levels, suggest that the average bias is relatively small. Previous studies also highlight the importance of incorporating antecedent conditions in overall improvements to AHPS forecasting skill, particularly when compared with climatic outlooks (Croley, 2006). In light of our analysis here, and these previous studies, we identify in the following paragraphs several critical components of AHPS which we believe readily can and should be improved through ongoing and future research.

To begin, we recognize that the current version of AHPS does not include a residual error term to quantify differences between historical Great Lakes water level forecasts and observations (Chapra, 2003; DiToro, 1984; Gronewold et al., 2009). This error term, conditioned on data (and forecasts) over a particular calibration period, would then be used to express uncertainty in future forecasts above and beyond the uncertainty explicitly represented in model inputs and model parameters (for further discussion of quantifying residual error terms in the context of model forecasting, see Gelman and Hill, 2007; Gronewold et al., 2011). At present, AHPS expresses forecast uncertainty (as described in the Methods section) solely by using a weighted ensemble of water level forecasts using different periods of record from historical climatology as model input. Consequently, including a calibrated model residual error term into the AHPS forecasting framework is a priority for short-term AHPS development and research, and represents a first step toward more explicitly quantifying uncertainty throughout the individual components of AHPS.

We also recognize that the water level “observations” used for assessing AHPS forecasting skill in this paper are, in fact, derived from a model (although in this case the model is statistical, rather than process-based) that is also intrinsically uncertain. Uncertainty and variability in monthly “observed” water level estimates depend in part, for example, on the number of gauges (and their spatial relationship) used in each estimate (Table 1). Furthermore, there are alternative model algorithms (i.e. alternative statistical models, such as the geometric mean) and probabilistic calibration routines that might better address the uncertainty in water level observations and computation of a representative “average” value while also accounting for uncertainties associated with a relatively limited

Table 2
Summary of AHPS forecasting skill based on the percentage of average monthly water levels observed from 1997 through 2010 within AHPS 3-month and 6-month forecast 90% prediction intervals, and corresponding median water level forecast bias.

Lake(s)	Percentage (%) of observations within forecast 90% prediction interval		Bias in median water level forecast(centimeters)	
	3-month forecast	6-month forecast	3-month forecast	6-month forecast
Superior	64	69	−0.4	0.9
Michigan–Huron	68	71	1.5	6.1
St.Clair	72	73	−5.5	−3.7
Erie	74	69	−5.5	−5.4

sampling size, variability over different spatial and temporal scales and errors (perhaps minor) in the individual water level measurements themselves.

There are also, of course, sources of uncertainty arising from the models and algorithms within AHPS that individually and collectively (Fig. 1) represent the major components of the hydrologic cycle and water level regulation plan-based interventions. While some of these sources of uncertainty are acknowledged and described in previous studies, we revisit them here to establish short term research priorities and to more closely align efforts to reduce those uncertainties with ongoing regional and national research priorities, including the planning, design, and siting of climate and hydrometeorological data collection infrastructure. For example, the current AHPS methodology for gathering and compiling meteorological data, while robust and efficient, does not fully utilize the full range of real-time data available from Environment Canada (EC) or NOAA including, for example, the Meteorological Assimilation Data Ingest System, or MADIS (Miller, 2009). As mentioned in the Introduction, the NWS CHPS system represents the type of innovative and comprehensive tool which could potentially address this need. Exploring how CHPS (and similar comprehensive modeling frameworks) could be utilized to improve Great Lakes water level forecasting (through, among other potential benefits, more efficient acquisition of real-time data at higher spatial and temporal resolutions) is an exciting and important area for future research.

Similarly, as with other process-based hydrological models in the Great Lakes drainage basin, AHPS could potentially be improved through a more explicit quantification of uncertainty in precipitation, evaporation, and runoff estimates (Croley, 1989; Fortin et al., 2006; St-Hilaire et al., 2003). At present, AHPS employs a Thiessen polygon-based averaging scheme for translating daily rainfall observations at individual gauging stations into subbasin averages (Croley and Hartmann, 1985), a methodology which has been applied and evaluated rather extensively in hydrological modeling (Tabios and Salas, 1985; Teegavarapu and Chandramouli, 2005). As currently encoded in AHPS, however, precipitation estimates do not account for measurement uncertainty or within-subbasin variability. Precipitation measurement uncertainty and spatial variability, regardless of how it is ultimately quantified, would almost certainly lead to changes in water level forecasts (and the expression of variability in those forecasts), not only because precipitation measurements are believed to be a significant source of uncertainty in AHPS model inputs, but also because the use of modified precipitation estimates would require recalibration of other components of AHPS as well, including the rainfall–runoff model.

Indeed, the version of the rainfall–runoff model (the LBRM) currently encoded in AHPS has not been calibrated recently, and a more recent calibration should be incorporated in the near future to reflect recent changes in land use patterns, climatology, and in the relationship between subbasin attributes and hydrological response (Wagener and Wheater, 2006). Recalibration of the LBRM, however, should be conducted in light of availability of other rainfall–runoff model alternatives and schemes for predicting runoff in subbasins with limited or no flow observations (Kokkonen et al., 2003; Krueger et al., 2010). Such model alternatives include (but are certainly not limited to) the Distributed Large Basin Runoff Model (or DLBRM, as described in Croley et al., 2005) and the general class of data-based mechanistic models (Jakeman et al., 1990), including their implementation in readily-available GUI packages (see, for example Croke et al., 2006). Some of these alternatives (particularly the DLBRM) represent approaches that attempt to quantify model parameters at a higher spatial resolution (i.e. over a finer spatial grid) than the current AHPS algorithms. There is an ongoing debate, however, regarding the relative benefits of lumped versus distributed modeling (see Beven, 2001, for further discussion), and we view a more focused comparison of these alternatives as an area for future research. We also acknowledge that existing NOAA comprehensive hydrological modeling frameworks, including CHPS, are explicitly designed to facilitate this type of comparison through integration of multiple model algorithms, a feature which we believe has significant potential for advancing the state-of-the-art of Great Lakes water level forecasting research.

We also recognize that effectively propagating any NBS sequence into water level forecasts depends on both the selected water level regulation plan (and how it is encoded within the overall modeling framework) as well as algorithms representing the formation, aggregation, and disaggregation of ice in Great Lakes interconnecting channels. To address this need, the current regulation and routing model algorithm in AHPS should continue to be updated and revised to match the current regulation plans and state-of-the-art algorithms encoding those plans, and to include information from the most recent flow–discharge relationship studies. In addition, an ice module should be added along with a subsequent assessment of potential improvements in water level forecasting skill. We suspect that an ice module, if implemented, might help AHPS better explain some of the more extreme lake level phenomena resulting from ice jams (for example) such as the one which occurred upstream of Lake St. Clair in early 2010 (Fig. 2). For further discussion of regulation plans within the context of propagating probabilistic forecasts into lake level management decisions, see Lee et al., (1997).

Finally, this effort underscores the importance of storing historical forecasts and the data from which they were derived to serve as a basis for model performance and prioritizing future model improvements. As methods for computing uncertainty change and data becomes more accurate, having access to a wide range of forecasts will serve to improve our model's skill. By maintaining this archive of data, we hope to increase the value of AHPS to the research and operation groups in the Great Lakes community. Given that data storage is not nearly the limitation it was when AHPS was assembled, this should be relatively easy to overcome in the future.

Conclusions

The results of our study suggest that the current version of AHPS captures the main temporal dynamics of Great Lake water levels both 3 and 6 months into the future, and the portion of water level variability explained by AHPS is relatively consistent across both 3- and 6-month forecasts. Multiple improvements, however, can be made to the current version of AHPS; one of the most critical is incorporation of residual error terms. Initially, an error term should be applied directly to water level forecasts. Ideally, each model component within AHPS would include a residual error term designed to account for model uncertainty and to propagate that uncertainty into water level forecasts in a hierarchical framework (Gelman and Hill, 2007; Gronewold and Borsuk, 2010), although observations for supporting estimates of residual error may not be available for all model components.

In addition, a series of alternative process models and model calibration routines can be employed within the existing modular framework of AHPS to explore potential improvements in water level forecasting skill. Combining these efforts with the development of similar comprehensive hydrological modeling frameworks (such as NOAA's CHPS framework) could set a new standard for state-of-the-art Great Lakes water level forecasting research. While the assessment in this paper acknowledges the probabilistic nature of AHPS current forecasting system, future assessments could employ more rigorous skill assessments including the assessment of rank histograms (Elmore, 2005) and an analysis of posterior predictive p-values (Gronewold et al., 2009).

The analysis and results presented in this paper serve as a platform for launching new research initiatives focused on improving AHPS and, more generally, the forecasting of net basin supplies and water levels within the Great Lakes. We hope these and future assessments will help provide guidance for prudent investments in monitoring infrastructure networks expected to reduce uncertainty and improve model forecasting skill.

Acknowledgments

This paper is GLERL contribution number 1594. The authors thank Tom Croley, Rob Caldwell, and Jim Lewis, as well as three anonymous reviewers for providing helpful comments which improved the clarity of the paper. We also thank Cathy Darnell for providing graphics and editorial support.

References

Allen, J., Holt, J., Blackford, J., Proctor, R., 2007. Error quantification of a high-resolution coupled hydrodynamic-ecosystem coastal-ocean model: part 2. Chlorophyll-a, nutrients and SPM. J. Mar. Syst. 68 (3–4), 381–404.

Angel, J., Kunkel, K., 2010. The response of Great Lakes water levels to future climate scenarios with an emphasis on Lake Michigan–Huron. J. Great Lakes Res. 36, 51–58.

Arhonditsis, G., Brett, M., 2004. Evaluation of the current state of mechanistic aquatic biogeochemical modeling. Mar. Ecol. Prog. Ser. 271, 13–26.

Beven, K., 2001. How far can we go in distributed hydrological modelling? Hydrol. Earth Syst. Sci. 5 (1), 1–12.

Chapra, S.C., 2003. Engineering water quality models and TMDLs. J. Water Resour. Plann. Manage.-ASCE 129 (4), 247–256.

Cohn, B.P., Robinson, J.E., 1976. A forecast model for Great Lakes water levels. J. Geol. 84 (4), 455–465.

Croke, B.F., Andrews, F., Jakeman, A.J., Cuddy, S., Luddy, A., 2006. IHACRES Classic Plus: a redesign of the IHACRES rainfall-runoff model. Environ. Model. Softw. 21 (3), 426–427.

Croley, T., 1989. Verifiable evaporation modeling on the Laurentian Great Lakes. Water Resour. Res. 25 (5), 781–792.

Croley, T., 1990. Laurentian Great Lakes double-CO_2 climate change hydrological impacts. Clim. Chang. 17 (1), 27–47.

Croley, T., 1992. Long-term heat storage in the Great Lakes. Water Resour. Res. 28 (1), 69–81.

Croley, T., 1997. Mixing probabilistic meteorology outlooks in operational hydrology. J. Hydrol. Eng. 2, 161–168.

Croley, T., 2002. Large basin runoff model. Chapter 17. In: Singh, V., Frevert, D., Meyer, S. (Eds.), Mathematical Models of Large Watershed Hydrology, pp. 717–770.

Croley, T., 2003. Weighted-climate parametric hydrologic forecasting. J. Hydrol. Eng. 8, 171.

Croley, T., 2006. Using climate predictions in Great Lakes hydrologic forecasts. In: Garbrecht, J., Piechota, T. (Eds.), ASCE Task Committee Report on Climatic Variations, Climate Change, and Water Resources Engineering. American Society of Civil Engineers, Arlington, VA, pp. 164–185.

Croley, T.E., Hartmann, H.C., 1985. Resolving Thiessen polygons. J. Hydrol. 76 (3–4), 363–379.

Croley, T., Hunter, T., 1994. Great Lakes Monthly Hydrologic Data: Technical Memorandum 083. US Dept. of Commerce, National Oceanic and Atmospheric Administration, Great Lakes Environmental Research Laboratory.

Croley, T., Lee, D., 1993. Evaluation of Great Lakes net basin supply forecasts. J. Am. Water Resour. Assoc. 29 (2), 267–282.

Croley, T., Quinn, F., Kunkel, K., Changnon, S., 1998. Great Lakes hydrology under transposed climates. Clim. Change 38, 405–433.

Croley, T., He, C., Lee, D., 2005. Distributed-parameter large basin runoff model. II: application. J. Hydrol. Eng. 10, 182.

DiToro, D.M., 1984. Probability model of stream quality due to runoff. J. Eng. Manage. ASCE 110 (3), 607–628.

Eberhardt, A., Moin, S., 2009. Determining the causes of declining upper great levels—phase I of the International Upper Great Study by the St. Clair River Task Team. Proceedings of World Environmental and Water Congress 2009, vol. 342. American Society of Civil Engineers, Reston, VA, USA, pp. 1–7.

Elmore, K., 2005. Alternatives to the chi-square test for evaluating rank histograms from ensemble forecasts. Weather. Forecast. 20, 789–795.

Fortin, V., Favre, A., Said, M., 2006. Probabilistic forecasting from ensemble prediction systems: improving upon the best-member method by using a different weight and dressing kernel for each member. Q. J. R. Meteorol. Soc. 132 (617), 1349–1369.

Gelman, A., Hill, J., 2007. Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge University Press, New York, NY.

Grima, A., Wilson-Hodges, C., 1977. Regulation of Great Lakes water levels: the public speaks out. J. Great Lakes Res. 3 (3–4), 240–257.

Gronewold, A., Borsuk, M., 2010. Improving water quality assessments through a hierarchical Bayesian analysis of variability. Environ. Sci. Technol. 44, 7858–7864.

Gronewold, A.D., Qian, S.S., Wolpert, R.L., Reckhow, K.H., 2009. Calibrating and validating bacterial water quality models: a Bayesian approach. Water Res. 42 (13), 2688–2698.

Gronewold, A., Myers, L., Swall, J., Noble, R., 2011. Addressing uncertainty in fecal indicator bacteria dark inactivation rates. Water Res. 45 (2), 652–664.

IJC, 1989. Living with the lakes: challenges and opportunities. Water Levels Reference Study.

IJC, 1993. Levels Reference Study — Great Lakes-St. Lawrence River Basin, Executive Summary.

Irvine, K., Eberhardt, A., 1992. Multiplicative, seasonal ARIMA models for Lake Erie and Lake Ontario water levels. Water Resour. Bull. 28 (2), 385–396.

Jakeman, A.J., Littlewood, I., Whitehead, P., 1990. Computation of the instantaneous unit hydrograph and identifiable component flows with application to two small upland catchments. J. Hydrol. 117 (1), 275–300.

Kokkonen, T., Jakeman, A., Young, P., Koivusalo, H., 2003. Predicting daily flows in ungauged catchments: model regionalization from catchment descriptors at the Coweeta Hydrologic Laboratory, North Carolina. Hydrol. Process. 17 (11), 2219–2238.

Krueger, T., Freer, J., Quinton, J., Macleod, C., Bilotta, G., Brazier, R., Butler, P., Haygarth, P., 2010. Ensemble evaluation of hydrological model hypotheses. Water Resour. Res. 46 (7), W07516.

Lamon, E.C., Stow, C.A., 2010. Lake Superior water level fluctuation and climatic factors: a dynamic linear model analysis. J. Great Lakes Res. 36 (1), 172–178.

Lee, D.H., Clites, A.H., Keillor, J.P., 1997. Assessing risk in operational decisions using Great Lakes probabilistic water level forecasts. Environ. Manag. 21 (1), 43–58.

Marchand, D., Sanderson, M., Howe, D., Alpaugh, C., 1988. Climatic change and Great Lakes levels the impact on shipping. Clim. Chang. 12 (2), 107–133.

McEnery, J., Ingram, J., Duan, Q., Adams, T., Anderson, L., 2005. NOAA's advanced hydrologic prediction service. Bull. Amer. Meteor. Soc 86, 375–885.

Meredith, D.D., 1970. Modeling of the Great Lakes water system. J. Am. Water Resour. Assoc. 6 (1), 55–64.

Miller, P., 2009. A 2009 update on the NOAA Meteorological Assimilation Data Ingest System (MADIS). 25th Conference on International Interactive Information and Processing Systems.

Pappenberger, F., Beven, K., 2006. Ignorance is bliss: or seven reasons not to use uncertainty analysis. Water Resour. Res. 42 (5), W05302.

Reckhow, K.H., 1994. Water-quality simulation modeling and uncertainty analysis for risk assessment and decision-making. Ecol. Model. 72 (1), 1–20.

Schaake, J., Franz, K., Bradley, A., Buizza, R., 2006. The hydrologic ensemble prediction experiment (HEPEX). Hydrol. Earth Syst. Sci. Discuss. 3 (5), 3321–3332.

Sellinger, C.E., Stow, C.A., Lamon, E.C., Qian, S.S., 2007. Recent water level declines in the Lake Michigan–Huron system. Environ. Sci. Technol. 42 (2), 367–373.

St-Hilaire, A., Ouarda, T.B., Lachance, M., Bobée, B., Gaudet, J., Gignac, C., 2003. Assessment of the impact of meteorological network density on the estimation of basin precipitation and runoff: a case study. Hydrol. Process. 17 (18), 3561–3580.

Stow, C.A., Jolliff Jr., J., D. J. M., Doney, S.C., Allen, J.I., Friedrichs, M.A., Rose, K.A., Wallhead, P., 2009. Skill assessment for coupled biological/physical models of marine systems. J. Mar. Syst. 76 (1–2), 4–15.

Tabios, G.Q., Salas, J.D., 1985. A comparative analysis of techniques for spatial interpolation of precipitation. J. Am. Water Resour. Assoc. 21 (3), 365–380.

Teegavarapu, R.S., Chandramouli, V., 2005. Improved weighting methods, deterministic and stochastic data-driven models for estimation of missing precipitation records. J. Hydrol. 312 (1–4), 191–206.

U.S. Army Corps of Engineers, 2008. Monthly (April) Bulletin of Lake Levels for the Great Lakes.

Wagener, T., Wheater, H.S., 2006. Parameter estimation and regionalization for continuous rainfall–runoff models including uncertainty. J. Hydrol. 320 (1–2), 132–154.

Young, P., 2003. Top-down and data-based mechanistic modelling of rainfall-flow dynamics at the catchment scale. Hydrol. Process. 17 (11), 2195–2217.