NATIONAL WEATHER SERVICE
OFFICE of HYDROLOGIC DEVELOPMENT

REQUIREMENTS

Hydrologic Ensemble Forecast Service

REVISION HISTORY

Date	Version	Description	Authors
12/28/2010	V1.0	First Draft	HEFS Requirements Team
2/25/2011	V2.0	Revised Draft for HEFS Planning Team Review	HEFS Requirements Team
5/23/2011	V3.0	Final Document	HEFS Requirements Team
6/10/2011	V4.0	Final Document incorporating additional review	HEFS Requirements Team

HEFS REQUIREMENTS REPORT CONTRIBUTORS

This report reflects the substantial contributions of the following people. The HEFS Requirements Team also acknowledges helpful insights provided by others from the NWS Office of Hydrologic Development (OHD) and River Forecast Centers (RFCs).

Project Manager Donna Page, NWS Office of Hydrologic Development (OHD)

Team Members Ernie Wells (Team Leader), NWS Office of Climate, Water, and Weather Services (OCWWS)

Dr. Andy Wood, NWS Colorado Basin River Forecast Center

Eric Jones, NWS Arkansas-Red Basin River Forecast Center

Joseph Ostrowski, NWS Mid-Atlantic River Forecast Center

Dr. Kevin (Minxue) He, NWS OHD

Other Significant Contributors / Reviewers

Dr. Julie Demargne, NWS OHD Hydrologic Ensemble Prediction (HEP) group

Dr. Pedro Restrepo, NWS OHD

Mark Fresch, NWS OHD Hydrologic Software Engineering Branch

Dr. John C. Schaake, Consultant

Rob Hartmann, NWS California-Nevada River Forecast Center

HEP group members:

Drs. Satish Regonda, James Brown, Limin Wu, Haksu Lee

TABLE OF CONTENTS

Page

REVISION HISTORY .... **I**

HEFS REQUIREMENTS REPORT TEAM .... **I**

TABLE OF CONTENTS .... **II**

LIST OF FIGURES .... **III**

LIST OF TABLES .... **III**

EXECUTIVE SUMMARY **4**

1. INTRODUCTION **9**

1.1 PURPOSE 9
1.2 METHODOLOGY 9

2. HEFS OVERVIEW AND PRIORITY SUMMARY **11**

2.1 BACKGROUND 11
2.2 HEFS OBJECTIVES 11
2.3 HEFS CAPABILITIES AND COMPONENTS 12
2.4 HEFS CAPABILITY PRIORITY LEVELS AND FUNCTIONAL MODES 15
2.5 HEFS COMPONENT PRIORITY LEVELS 18
2.6 HEFS PROTOTYPE COMPONENT GAPS 20

3. HEFS SYSTEM AND COMPONENT REQUIREMENTS **22**

3.1 SYSTEM LEVEL REQUIREMENTS 22
3.1.1 Concept of Operations (Conops) summary *22*
3.1.2 Requirement Summary *26*
3.1.3 Possible Future or Deferred Requirements *28*
3.2 METEOROLOGICAL ENSEMBLE FORECAST PROCESSING (MEFP) 29
3.2.1 Conops summary *29*
3.2.2 Requirement Summary *31*
3.2.3 Prototype Readiness Levels *33*
3.3 DATA ASSIMILATOR 35
3.3.2 Requirement Summary *37*
3.3.3 Prototype Readiness Levels *38*
3.4 ENSEMBLE POST-PROCESSOR 39
3.4.1 Conops summary *41*
3.4.2 Requirement Summary *42*
3.4.3 Prototype Readiness Levels *42*
3.5 ENSEMBLE VERIFICATION SYSTEM (EVS) 43
3.5.1 Conops summary *43*
3.5.2 Requirement Summary *44*
3.5.3 Prototype Readiness Levels *45*
3.6 GRAPHICS GENERATOR (GRAPHGEN) 46
3.6.1 Conops Summary *46*
3.6.2 Requirement Summary *47*
3.6.3 Prototype Readiness Levels *48*
3.7 RELATIONSHIP OF HEFS REQUIREMENTS TO NEW YORK CITY DEPARTMENT OF ENVIRONMENTAL PROTECTION (NYCDEP) PROJECT REQUIREMENTS 48
3.8 PRODUCT DISSEMINATION / DATA SERVICES 50

REFERENCES 50

APPENDIX 52

RESPONSE TO MAJOR REVIEW COMMENTS 52
OVERVIEW OF TESTING OF PROTOTYPE COMPONENTS 54

LIST OF FIGURES

Page

Figure 1 Basic capabilities of an ensemble hydrologic forecast system.................................... 14
Figure 2 Schematic of components discussed in this document for potential inclusion in the HEFS. Blue shading indicates existing components of the NWS deterministic forecasting system. Orange shading indicates new ensemble forecasting system capabilities and components. Dashed lines indicate relatively undeveloped prototype components. 17

LIST OF TABLES

Page

Table 1 Ensemble forecast system capabilities, associated components, readiness levels and overall priority for HEFS version 1. 5
Table 2 Prototype or developmental components potentially supporting ensemble forecast system capabilities. 15
Table 3 Overall priority level for the functional modes of HEFS Version 1 capabilities:. ***Required*** – must have in HEFS; ***High*** – the capability is highly desired; ***Medium*** – desired capability but HEFS could function without it; ***Low*** – HEFS can succeed without this component, or component requires prohibitive R&D. For each of the capability's functional modes, the table indicates the necessity of function within CHPS. 17
Table 4. Component development priorities for HEFS version 1... 19
Table A.1 Testing of XEFS prototypes ....**Error! Bookmark not defined.**

EXECUTIVE SUMMARY

Providing uncertainty ranges for hydrologic forecasts at all time scales – from hours to beyond a year – is one of the most pressing needs of operational hydrologic forecasting. This need has been validated many times over the years [NRC report (2006), CFI survey (2008), multiple NWS service assessments]. Although there are a variety of approaches for providing uncertainty bounds on hydrologic predictions, the preferred strategy in weather, climate and hydrologic prediction centers on the production of ensemble predictions.

Our objective is to outline the requirements for an initial NWS ensemble forecasting service that can be developed and implemented by 30 Sep 2013 (approximately three years). We begin with a primary ***service-level*** objective: ***to produce and deliver ensemble streamflow forecasts and associated products/ data with a number of specific characteristics.*** These output forecast ensembles must:

1. Span lead times from one hour to one year or more (defaulting to climatology) with seamless transitions between lead time regimes (e.g., weather to climate, short to medium to seasonal range).
2. Be calibrated from a probabilistic standpoint for relevant forecast periods.
3. Be spatially and temporally consistent, thus linkable (routable) across RFC domains.
4. Effectively capture the information available from current operational weather to climate forecast systems by utilizing meteorological ensemble forecasts (e.g., precipitation and temperature) that are calibrated from a probabilistic standpoint for relevant forecast periods.
5. Be consistent – i.e., using similar data and methods – with retrospective forecast ensembles that are used for verification and training/optimization of user decision support tools.
6. Be verified via a comprehensive verification system that can generate products qualifying the expected performance of the output streamflow ensembles.

We also describe a broad-brush concept of operations (CONOPS) that may apply to the generation of ensemble forecasts in operations and hindcasting. We further recognize that a parallel effort is required to develop effective approaches for presentation of ensemble forecasts and dissemination of data/ products to NWS' internal partners (e.g., WFOs) and external users. At least for the initial implementation of HEFS (version 1), the product dissemination will likely utilize existing methods and formats to a great degree. Although there could be enhancement to the existing dissemination methods, a more consistent and integrated data service for these ensemble products is strongly desired. The successful implementation of HEFS version 1 will also require training and user outreach on techniques for ensemble forecasting and forecast verification, as well as forecast and verification products.

An assessment of the relative priority and readiness level for each component of the HEFS is provided in **Table 1** below. The capability priority is a reflection of our assessment of the criticality of each capability to meet the six stated service objectives. The component readiness is assessed via two categories - applied science readiness and software readiness. Section 1.2 describes these categories in more detail.

Table 1 Ensemble forecast system capabilities, associated components, readiness levels and overall priority for HEFS version 1.

Ensemble Forecast System Capability	Capa-bility Priority	Prototype Component	Applied Science Readiness	Software Readiness	Ability to meet HEFS Objective	Component Version 1 Priority
Meteorological ensemble forecast processing	Required	**EPP3**	Medium	Medium	Medium	**Required**
		MMEFS	Low	High	Low	Medium
Hydrologic Processing	Required	**Hydrologic Processor**	High	High	High	**Required**
Automatic data assimilation	High	1D-VAR	Medium	High	Low	Medium
		2D-VAR	Medium	Medium	Low	Medium
		EnKF	Low	Medium	Medium	Low
Hydrologic ensemble forecast post-processing	High	**EnsPost**	Medium	High	Medium	**High**
		HMOS	Medium	High	Low	Low
		MSCM	Low	Low	Low	Low
		TBD(Augmented Gaussian regression) [1]	--	--	--	Medium
Ensemble forecast verification	Required	**EVS**	High	High	High	**Required**
Product generation	Required	**GraphGen**	High	High	High	**Required**
Product Dissemination / Data Services	Required	**Ad-hoc, existing (AHPS web, RFC web, AWIPS WAN, etc)**	Medium	Medium	Medium	**Required**
		Integrated Data Services	--	--	--	**--**

[1] There exists a broad array of post-processing techniques that may be developable within the scope of the project. Some of these will be discussed at the June 2011 HEPEX meeting that NWS plans to attend.

The final column in Table 1, "Component Version 1 Priority", provides our assessed priority for each potential component as it relates to the most essential needs for an initial operational implementation of HEFS. In addition to the prioritized need for a specific capability, it also reflects the realities of the current prototype (e.g. state of development/testing, software engineering maturity, ability of individual component to support service-level objectives) and the relative likelihood that specific components can be ready for operational implementation within the planned schedule. In essence, "Component Version 1 Priority" is a composite assessment combining all the elements/data in the previous columns.

The readiness and priority levels included in Table 1 each reflect the balancing of several considerations, as noted earlier. More detail on the factors determining these levels is given in the component-focused subsections of Section 3. Also, to address the perspectives of reviewers that suggested different levels for some components, additional discussion is included in the Appendix section.

Key Findings / Recommendations

1) The XEFS ("eXperimental Ensemble Forecast System") Design and Gap Analysis (NWS, 2007) outlined a set of forecasting system components that are loosely used to define a pre-existing component set labeled "XEFS". Although gaps exist in the ability of these XEFS components to fully satisfy the complete set of service objectives (see #5 below) and the testing/evaluation of individual XEFS components and outputs is variable and in some cases limited (see #4 below), we nonetheless view a subset of current XEFS components as a reasonable basis to build a version 1 operational implementation of HEFS.

2) Pursuing an initial implementation of HEFS focused on the capabilities of **five required** components (EPP3, Hydrologic Processor (hindcasting), EVS, GraphGen, and some augmented version of our current ad-hoc product dissemination tools) and one **high** priority component in the hydrologic ensemble forecast post-processing category provides the best opportunity to satisfy the service objectives above within the current project schedule and anticipated resources.

3) From a component view, the meteorological ensemble forecast processing component (EPP3) will require the most resources and effort. The Hydrologic Software Engineering Branch (HSEB) of Office of Hydrologic Development (OHD) has already begun to delineate potential software development efforts for EPP3, including the following:
    a. Increased understanding and documentation of EPP3 algorithms and code (transition knowledge from John Schaake broadly to OHD personnel)
    b. Enhance code to incorporate/use new data from CFSv2 hindcasts and forecasts
    c. Enhance code to incorporate/use new data from GEFS
    d. Enhance code to mitigate limitations associated with output data in datacard format.
    e. Provide some capability to perform real-time control (MODs) within EPP3

   f. Generally enhance code to allow for improved supportability, maintainability, usability, efficiency (e.g., utilizing standard XML format inputs, improved logging and error handling, utilizing CHPS database for parameter storage, etc)
   g. Enhance code to provide diagnostic feedback to forecasters

4) Individual XEFS components have been tested/evaluated to varying degrees at OHD and at individual RFCs. In some cases that testing and evaluation, especially with regards to the variety of conditions (basins) and the extent of both objective and qualitative assessment of output products, is not sufficient to gage the prototype's likely performance in a real-time operations process.

5) Capability gaps exist in the areas of ensemble forecast post-processing and automated data assimilation. Operational data assimilation is a challenging area in which it is likely that continued research beyond the scope of the HEFS project will be needed to provide desired functionalities (i.e., substantially reducing use of MODs). In contrast, ensemble forecast post-processing is an area in which a range of developable solutions may exist, particularly for lead time longer than those addressed by the current EnsPost. Some of these solutions are adaptations of algorithms from HMOS, EnsPost, or other less mature techniques currently being explored by OHD. We recommend defining and undertaking a focused science development/testing effort toward improved ensemble post-processing in parallel with the initial stages of software development /system engineering on the required components of HEFS. Inclusion of the high priority ensemble post-processing component in the HEFS version 1 (HEFS V1) software implementation would be contingent on successful testing/evaluation and the overall project implementation schedule.

6) From a systems perspective, CHPS should provide a ready framework for implementing the new components/functionality that will comprise HEFS. But:
   a. Even the deterministic forecast processes at RFCs are still evolving with the CHPS implementation, as variations on standard workflows become possible, and we haven't had any significant end-to-end system testing of HEFS in the CHPS environment.
   b. From a user CONOPS view, there are still unknowns (with regards to user interaction, performance, utilization of MODs, etc) that make it difficult to specify/develop a fully-integrated, CHPS-based forecast environment that integrates tools for ensemble forecasting with those used for daily operational forecasting.
   c. A model calibration system within CHPS for deterministic forecasting is under development (with an uncertain completion schedule); meanwhile the HEFS version 1 will require some basic calibration/configuration capability for its different components. For HEFS V1, this calibration/configuration capability is likely to be included in the individual component software.
   d. An archive system for CHPS is under development (with an uncertain completion schedule); meanwhile the HEFS V1 will require some archive capability (e.g., flat files, database) to store all ensemble products (real-time forecasts and reforecasts) and verification products. The feasibility and

advisability of different alternatives for implementing such an archive that will interface well with CHPS needs to be explored further.

7) Implementing a rapid system prototype for end-to-end semi-operational testing at an RFC (using CHPS), with OHD assistance, would likely provide valuable insight as the software development process advances. This would build on the existing XEFS effort but link all the appropriate components in an operational context. Something akin to the CHPS Acceleration Team (CAT) concept (e.g. a HAT) could allow for this accelerated system testing at select RFCs and provide feedback supporting one or more iterations of spiral development before the initial HEFS V1 operational implementation. This focused system testing could occur concurrently with known software development efforts (like those described in #3 above to enhance EPP3). One goal of this testing and feedback loop would be to maximize system usability, including minimizing manual steps required both within and outside of the IFD.

8) Training for RFC forecasters (and others) to enhance their knowledge/understanding of ensemble forecasting, the value and utility of output ensembles, etc., is critical and will require significant effort and resources.

9) Effective communication and utilization of the information in the output ensembles will require significant outreach to internal (WFOs) as well as external users and partners. Among other needs, protocol and tools for ensemble forecast usage by WFOs in their product generation and dissemination role must be discussed and developed.

1. INTRODUCTION

1.1 Purpose

The main objective of this report is to outline the requirements for an initial NWS ensemble forecasting service that can be developed and implemented by 30 Sep 2013 (approximately two years). Specifically, this report will attempt to identify a minimum set of requirements to provide an initial ensemble forecasting capability, as well as additional, lower priority requirements and their consequences – that is, the effort needed to meet those requirements given the readiness levels of existing system prototypes. The report also assesses current prototype readiness levels. As context for the requirements discussion, the report briefly sketches the concept of operations (defined below) that applies to the ensemble Forecast Service components. Lastly, the report discusses the requirements in the context of the New York City Department of Environmental Protection (NYCDEP) Operations Support Tool (OST) development project.

1.2 Methodology

The requirements for HEFS were developed from a review of existing documentation and publications related to the EXperimental Ensemble Forecast System (XEFS) and its components and from a series of discussions and interviews between the HEFS team and other NWS personnel, including River Forecast Center (RFC) leadership (e.g., HIC Rob Hartmann) and Office of Hydrologic Development (OHD) leadership and science and software team experts. The project manager for the ongoing NYCDEP forecasting and decision support project (Jim Porter) was also interviewed, and the HEFS requirements as they pertain to that project are discussed in Section 3.7.

The HEFS team also discussed the concept of assigning Technological Readiness Levels (TRLs; see http://en.wikipedia.org/wiki/Technology_readiness_level) to the system and the components. TRLs define discrete stages (currently nine in the NASA and DOD schemas) with which to measure the status of a technology as it proceeds from exploratory research to proof of concept, then to prototype development and testing, and finally to operational implementation. The application of TRLs to the HEFS prototype components is complicated, however, because in some instances software has been engineered sufficiently to enable testing and implementation in the operational environment (a higher TRL), despite the fact that the research-level proof of concept and value has not been thoroughly demonstrated (a lower TRL). To be able to recognize both aspects of development, a single TRL rating for component capabilities is not provided in this document. Rather, two readiness evaluation categories are adopted:

- ***applied science readiness*** –the extent to which the prototype has been tested in applied case studies that span the range of situations in which it is likely to be applied. For streamflow forecasting, this range includes catchments from headwaters to regulated points, lead times from short to seasonal range, and hydroclimatic variability.

- ***software readiness*** – the extent to which software to implement the prototype is ready for operational implementation (based on usability, maintainability, supportability, computing efficiency, adaptability to function within CHPS, etc.).

Within these two categories, qualitative ratings of **low**, **medium** and **high** are given. These categories correspond respectively to significant, moderate, and minimal effort required to advance the prototype to a level required not only for operational implementation (in the case of software readiness), but also operational acceptance (in the case of applied science readiness).

In addition, we also evaluate the extent to which the current prototype(s) includes capability to meet required component-level functions that support the overall service level objectives of HEFS.

Our assessment begins by delineating service-level objectives of HEFS and mapping those to basic capabilities and assigning priorities to each capability. We then map individual prototype components to each basic capability and provide an assessment of their readiness to meet the service-level objectives. These component readiness levels are used in conjunction with the capability priority assignments to define our final component priority assignments for HEFS V1.

2. HEFS Overview and Priority Summary

This section describes the basic output objectives of an ensemble forecast service, required component capabilities, and gives a summary of the assessed priority level associated with each capability.

2.1 Background

RFCs currently issue two primary types of streamflow forecasts: (1) deterministic streamflow forecasts for lead times from hours out to 10 days, updated one or more times daily; and (2) ensemble streamflow forecasts for lead times of a month out to approximately one year. Variations on these general categories also exist, such as the NWRFC Single-Trace Procedure 120 day deterministic forecasts (see http://www.nwrfc.noaa.gov/stp/stp_des.cgi) and the recently developed experimental Met-Model Ensemble Forecast System (MMEFS) ensemble short-range forecasts (see http://products.weather.gov/PDD/MMEFS_PDD_final_11162010.pdf for more details).

Providing uncertainty ranges for hydrologic forecasts at all time scales – including not only the current longer lead water management products but also short-range flood forecasts – is one of the most pressing needs of operational hydrologic forecasting [NRC report (2006), CFI survey (2008), multiple NWS service assessments]. Although there are a variety of approaches for providing uncertainty bounds on hydrologic predictions, the preferred strategy – and the only practical alternative – for the simulation and optimization of water resources systems in weather, climate and hydrologic prediction centers, is based on the production of ensemble-based predictions.

2.2 HEFS Objectives

The HEFS ***service-level*** objective is to produce ensemble hydrologic, hydraulic and water management system forecasts with a number of specific characteristics. These output forecast ensembles must:

1. Span lead times from one hour to one year or more (defaulting to climatology) with seamless transitions between lead time regimes (e.g., weather to climate, short to medium to seasonal range).
2. Be calibrated from a probabilistic standpoint for relevant forecast periods.
3. Be spatially and temporally consistent, thus linkable (routable) across RFC domains.
4. Effectively capture the information available from current operational weather to climate forecast systems by utilizing meteorological ensemble forecasts (e.g., precipitation and temperature) that are calibrated from a probabilistic standpoint for relevant forecast periods.

5. Be consistent – i.e., using similar data and methods – with retrospective forecast ensembles that are used for verification and training/optimization of user decision support tools.
6. Be verified via a comprehensive verification system that can generate products qualifying the expected performance of the output streamflow ensembles.

2.3 HEFS Capabilities and Components

An operational ensemble prediction service centers on a ***hydrologic, hydraulic and water management simulation system*** (i.e., the river forecast models run by the RFCs) that can be run in either deterministic or ensemble mode. For convenience and consistency with previous XEFS documentation, this river system simulation capability will be referred to as the ***Hydrologic Processor***[1]. To meet the objectives of operational ensemble prediction, the HEFS should include the following five capabilities:

- ***Meteorological ensemble forecast processing (MEFP)*** – the ability to construct calibrated (i.e., reliable and as narrowly-distributed, or *sharp*, as possible) ensemble meteorological forecasts at time and space scales suitable for input to the hydrologic forecasting system.
- ***Hydrologic Processing*** **–** the ability to propagate the uncertainty in the meteorological input ensembles into the hydrologic output ensembles.
- ***Hydrologic ensemble forecast post-processing*** – the ability to adjust forecast outputs to eliminate systematic bias and spread deficiencies and achieve a calibrated hydrologic (or hydraulic or water management system) ensemble forecast.
- ***Ensemble forecast verification*** – the ability to measure the skill and statistical characteristics of the ensemble forecasts.
- ***Product generation*** – the ability to translate outputs from the forecast system into formats to support partner decision processes.
- ***Product Dissemination/Data Services*** – the ability to deliver basic products and data to internal/external users and partners.

Additionally, the ensemble prediction system would ideally include:

- ***Automatic data assimilation*** – the ability to use observations of stream flow or stage, snowpack and other hydrologic state variables in an automated procedure to improve state variables and/or initial conditions for operational hydrologic models and correct errors in the hydrologic, hydraulic or water management simulations.

A simple schematic depicting these basic capabilities for HEFS is shown below in Figure 1.

---

[1] The Community Hydrologic Prediction System (CHPS), currently being implemented at the RFCs, encompasses this capability and will serve as the core framework for the HEFS.

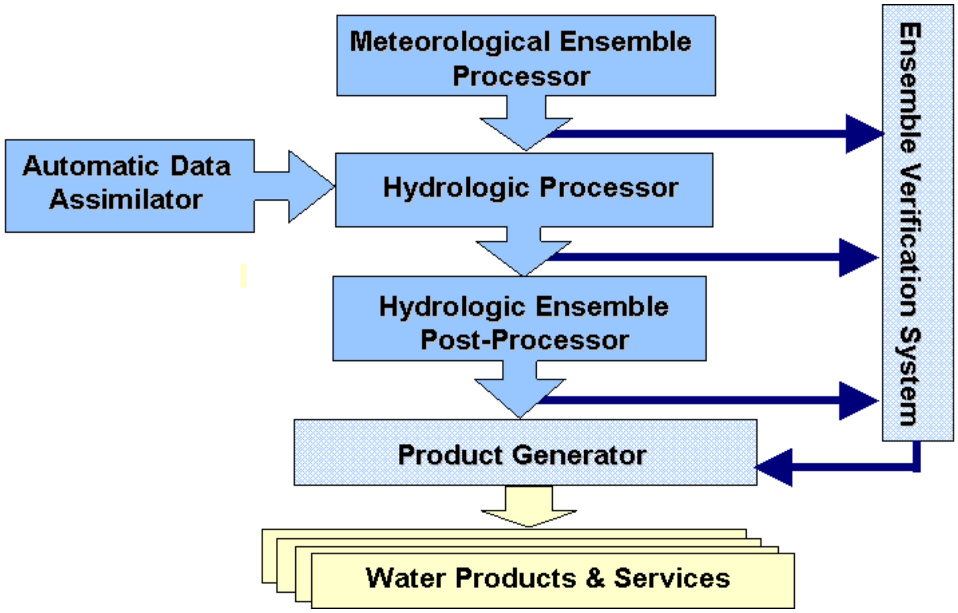

Figure 1. Basic capabilities of an ensemble hydrologic forecast system

To expedite development and testing of the basic capabilities outlined above, the OHD has been developing an EXperimental Ensemble Forecast Service System (XEFS), a prototype hydrologic ensemble prediction system for testing and use at RFCs. The XEFS currently includes the following major prototype components:

- **Meteorological Ensemble Pre-processor (EPP3)** produces meteorological ensemble forecasts for lead times from six hours to one year utilizing combinations of the RFC operational single-valued forecasts, the NCEP Climate Forecast System (CFS) ensemble forecast means, and/or the NCEP/ESRL Global Forecast System (GFS) ensemble means.
- **Data Assimilator (DA)** consists of several methods for data assimilation: primarily 1D-VAR for adjustment of 3-parameter Muskingum routing model parameters as well as background flow (i.e., flow at the downstream location of a river reach at the beginning of the assimilation window before the assimilation), but also 2D-VAR and EnKF for adjustment of state variables of the SAC-SMA model and SNOW17 model, respectively.
- **Hydrologic Ensemble Post-processor (EnsPost)** provides forecast trace-specific adjustments using the latest streamflow observation and a Gaussian regression model developed with past flow simulations and observations to correct bias and spread in the hydrologic forecast ensemble for lead times up to approximately five days (the regression parameter is constant for all lead times and therefore limit the EnsPost application to short term).
- **Hydrologic Model Output Statistic (HMOS)** provides uncertainty bounds for short-term deterministic ("single-value") streamflow forecasts using the latest streamflow observation and a Gaussian regression model for each individual lead

time developed using past flow forecast–observation pairs, but could apply forecast trace-specific corrections for lead times up to approximately five days.

- **Ensemble Verification System (EVS)** analyzes the quality of retrospective and real-time forecast ensembles (meteorological inputs and hydrologic outputs) and calculates a broad range of forecast verification metrics.
- **Graphics Generator (GraphGen)** emulates the Ensemble Streamflow Predication Analysis and Display Program (ESPADP) software in generating graphical products from forecast system ensemble outputs.

Additional background information is given in the XEFS Design and Gap Analysis report (http://www.weather.gov/oh/rfcdev/docs/XEFS_design_gap_analysis_report_final.pdf) and at the OHD XEFS one-stop web page (http://www.weather.gov/oh/XEFS).

Several ongoing development efforts outside XEFS have produced prototype procedures that could also serve the capabilities of HEFS. ***MMEFS*** has advanced to experimental operational status at Eastern Region RFCs (see http://www.erh.noaa.gov/mmefs/index.php). MMEFS forms ensemble meteorological forecasts by interpolating raw NWP ensemble outputs from several operational sources (SREF, NAEFS, GEFS) to suitable time-space formats for input to the RFC hydrologic models. Despite significant limitations in its current implementation (e.g., no statistical calibration of the raw meteorological ensemble forecasts), it can be viewed as an alternative approach for meteorological ensemble forecast generation to the XEFS EPP3 component. Another potential component prototype supports the hydrologic ensemble forecast post-processing capability: the ***multi-scale CDF matching (MSCM)*** approach for bias-correction of forecast traces for multiple temporal scales has been explored to a limited extent by the OHD HEP group to extend hydrologic post-processing to medium and long term.

Several other potential options (e.g., Indicator Co-Kriging non-parametric processor, General Linear Model – Post-Processor (GLM-PP), OpenDA, etc.) are being explored and pushed to varying levels of development at OHD . These have yet to be comprehensively tested by OHD and the NWS, but are among those currently being evaluated using the HEPEX post-processing testbed dataset (initial results will be presented at the HEPEX workshop on June 7-9, 2011, but were not available at the time of writing of this report). Some of these potential techniques may be discussed in the specific capability sections later in this report, as they may be included in future development versions of HEFS.

The prototype efforts described above are depicted as a potential ensemble forecast system via the schematic shown in **Figure 2**.

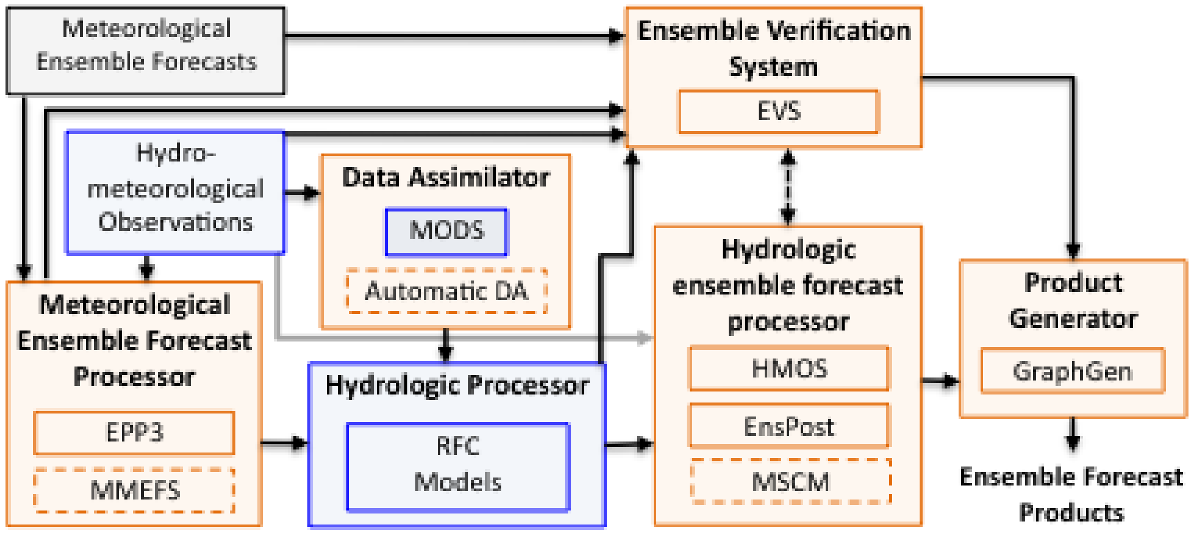

Figure .2 *Schematic of components discussed in this document for potential inclusion in the HEFS. Blue shading indicates existing components of the NWS deterministic forecasting system. Orange shading indicates new ensemble forecasting system capabilities and components. Dashed lines indicate relatively undeveloped prototype components.*

2.4 HEFS Capability Priority Levels and Functional Modes

HEFS capabilities have three general modes of functionality, all of which are desired within the CHPS environment, but some of which may be acceptable, at least initially, outside of CHPS. The first mode is ***configuration***, the development of settings or parameters that govern the application of components in real-time or hindcasting. Configuration has also been referred to as *calibration*, but the term configuration is adopted here to avoid confusion with other uses of the term calibration (e.g., traditional model calibration, and also the ensemble forecast statistical calibration that is accomplished by running the calibration programs within EPP3 and EnsPost). The second mode is application in ***real-time operations*** within CHPS. The third mode is ***hindcasting***, which is the application of the component (and system) to produce forecasts for past (retrospective) initialization dates. Hindcasts are used to evaluate the performance of real-time forecasts, and are required as input to support the ensemble forecast post-processing capability, ensemble forecast verification, and decision support needs for users who develop decision rules based on past system or forecast performance (e.g., the Operation Support Tool developed by NYCDEP).

Note that the priority levels here represent, to some degree, three considerations: the importance of having a capability to achieve HEFS objectives; the level of applied science readiness; and the level of software readiness. In determining capability priority levels, we also consider that configuration and hindcasting activities are performed infrequently and can generate output that can either be imported into CHPS or analyzed outside of CHPS, while still providing benefit. For each of the capability's functional modes (Operations, Configuration, and Hindcasts), we indicate the necessity of that function occurring within CHPS. The priority levels also account for NYCDEP project requirements.

As **Table 2** indicates, from a capability perspective, ***a baseline (initial) ensemble forecast service*** must include:

- Calibrated meteorological ensemble forecast processing, configurable either within or outside of CHPS, to produce calibrated meteorological ensembles at the basin scale.
- Hydrologic processing to produce raw hydrologic ensembles by propagating the uncertainty in the meteorological input ensembles into the hydrologic output ensembles.
- Hindcasting capability to retroactively generate calibrated meteorological ensembles and streamflow ensemble forecasts for multiple past years using a fixed automated forecasting system, to the extent possible.
- A comprehensive verification capability.
- Product generation to provide a suite of products from the ensemble forecasts, ensemble reforecasts, and verification outputs.
- Product delivery to disseminate the ensemble and verification products to users.

These required capabilities generally emulate our current deterministic forecasting capabilities that also do not currently include automatic data assimilation or post-processing capabilities. *These two capabilities are nonetheless recognized as being essential to an objective forecasting system, and are recommended for continued parallel research and development efforts designed to enable their inclusion in future versions of HEFS.* An initial HEFS will need to include/add a robust hindcasting capability, which is required for comprehensive verification and to satisfy specific requirements of the NYCDEP project. Verification is deemed necessary primarily for organizational acceptance of ensemble forecasting and for due diligence related to the NYCDEP project, but it also has an obvious benefit in helping RFCs understand the strengths and weaknesses of the services they provide.

Table 2 Overall priority level for the functional modes of HEFS Version 1 capabilities:. ***Required*** **–** must have in HEFS; ***High*** **–** the capability is highly desired; ***Medium*** **–** desired capability but HEFS could function without it; ***Low*** **–** HEFS can succeed without this component, or component requires prohibitive R&D. For each of the capability's functional modes, the table indicates the necessity of function within CHPS.

Capability	Purpose	Capability Priority	Mode must function within CHPS		
			Operations	Configuration	Hindcasts
Meteorological Forecast Ensemble Processing	Provide calibrated ensemble meteorological forcings as input to the hydrologic processor	**Required**	Yes	No	No
Automatic Data Assimilation	Use real-time observations of streamflow, snowpack and other variables to adjust model parameters, states, input or output	High	Yes	Yes	Yes
Hydrologic Processing	Propagate the uncertainty in the meteorological input ensembles into the hydrologic output ensembles	**Required**	Yes	Yes	Yes
Hydrologic Ensemble Forecast Post-processing	Calibrate hydrologic ensemble output to reduce systematic biases in forecast mean and spread	High	Yes	Yes	Yes
Ensemble Forecast Verification	Calculate a range of verification metrics for meteorological input and hydrologic output forecasts	**Required**	No	No	No
Product Generation	Produce graphical and analysis products based on forecast ensemble traces	**Required**	Yes	Yes	Yes
Product Dissemination / Data Services	Deliver basic products and data to internal/external users and partners	**Required**	No	No	No

The hydrologic ensemble forecast post-processing (ensemble calibration) capability is highly desirable, in that it would improve the quality and usability of the forecasting system. Although prototypes which offer a limited version of this capability exist, further R&D effort is needed to attain applied science readiness (i.e., comprehensive proof of concept at required time and space scales). As noted earlier, an automatic data assimilation capability would also be beneficial to an ensemble forecast service, but such a capability in a comprehensive form is not available from the existing prototypes.

2.5 HEFS Component Priority Levels

The XEFS and additional prototype components (**Table 1**) have attained varying levels of development readiness and been tested/evaluated to varying degrees at OHD and at individual RFCs. Details of our assessment of each component are found in Section 3. Further details of the testing and verification efforts associated with each component can also be found there. Table A1 of the Appendix contains a high level compilation of some basic attributes of the testing regime for each of the current XEFS components.

Attempting to include all of these components in the initial operational implementation of HEFS would require effort that likely exceeds the limits of the three-year HEFS development and implementation plan. We therefore assign priorities to the further development and incorporation of each system component (**Table 3**) as input to aid allocation resource decisions for future HEFS R&D and software implementation. In particular, we focus on component prioritization for the initial implementation version of HEFS.

The **capability priority** is a reflection of our assessment of the criticality of each capability to meet the six stated service objectives. The component readiness is assessed via two categories - **applied science readiness** and **software readiness** (see section 1.2 for details). The final column "**Component Version 1 Priority**" provides our assessed priority for each potential component as it relates to the most essential needs for an initial operational implementation of HEFS. In addition to the prioritized need for a specific capability, it also reflects the realities of the current prototype (e.g. state of development/testing, software engineering maturity, ability of individual component to support service-level objectives) and the relative likelihood that specific components can be ready for operational implementation within the planned schedule. In essence, "**Component Version 1 Priority**" is a composite assessment combining all the elements/data in the previous columns

In our assessment, an initial implementation of HEFS focusing on the capabilities of **five required** components (EPP3, Hydrologic Processor (hindcasting), EVS, GraphGen, and some augmented version of our current ad-hoc product dissemination tools) and **one high** priority component in the hydrologic ensemble forecast post-processing category provides the best opportunity to satisfy the service objectives above within the current project schedule and anticipated resources.

Table 4. Component development priorities for HEFS version 1.

Ensemble Forecast System Capability	Capability Priority	Prototype Component	Applied Science Readiness	Software Readiness	Ability to meet HEFS Objective	Component Version 1 Priority
Meteorological ensemble forecast processing	Required	**EPP3**	Medium	Medium	Medium	**Required**
		MMEFS	Low	High	Low	Medium
Hydrologic Processing	Required	**Hydrologic Processor**	High	High	High	**Required**
Automatic data assimilation	High	1D-VAR	Medium	High	Low	Medium
		2D-VAR	Medium	Medium	Low	Medium
		EnKF	Low	Low	Medium	Low
		4D-VAR	Low	Low	Medium	Low
Hydrologic ensemble forecast post-processing	High	**EnsPost**	Medium	High	Medium	**High**
		HMOS	Medium	High	Low	Low
		MSCM	Low	Low	Low	Low
		TBD(Augmented Gaussian regression) [1]	--	--	--	Medium
Ensemble forecast verification	Required	**EVS**	High	High	High	**Required**
Product generation	Required	**GraphGen**	High	High	High	**Required**
Product Dissemination / Data Services	Required	**Ad-hoc (AHPS web, RFC web, AWIPS WAN, etc)**	Medium	Medium	Medium	**Required**
		Consistent, Integrated Data Service[2]	--	--	--	**High**

[1]There exists a broad array of post-processing techniques that may be developable within the scope of the project. Some of these will be discussed at the June 2011 HEPEX meeting that NWS plans to attend.

[2]There is a need for an integrated data service and associated policies to connect ensemble forecasts to decision support systems of users (including WFOs), but prototypes for this service were not addressed in detail in this report.

2.6 HEFS Prototype Component Gaps

It should be noted that the initial implementation almost certainly will not be able to completely satisfy all of the stated HEFS service-level objectives (see Section 2.2). Gaps or "breakpoints" in the ability of existing XEFS components to support specific HEFS service-level objectives include the following:

- DA algorithms as currently developed and tested apply only to Muskingum model routing parameters (1D-VAR), hourly timestep headwater basins (2D-VAR), or very limited test basins (EnKF). Thus most of the desired capability to replace or significantly reduce the use of many common MODs (related to SNOW17/ SAC-SMA model states, parameters and forcing adjustment, etc.) at all RFCs remains well beyond the current DA capabilities. This limitation undermines objective 5.
- The current EnsPost approach (using only the latest streamflow observation and a Gaussian regression model developed with past flow simulations and observations) is designed primarily for short-range ensembles with lead times of 1-5 days, thus cannot support the calibration of forecast ensembles beyond the short range (the regression parameter is constant for all lead times). This limitation undermines objectives 1 and 2.
- The current EnsPost is calibrated using observed and simulated flows, and applied to each ensemble trace with the assumption that the hydrologic forecast ensemble/ensemble trace exhibit the error structure of the retrospective simulations. However, real-time simulations and forecasts are not entirely consistent with retrospective ones (due to MODs). Hence EnsPost may not yield optimal benefits (i.e., may not adjust forecast ensemble for biases). This limitation, which exists for many post-processing approaches, may undermine HEFS objectives 2 and 5 (i.e., limitations in real-time will also exist in hindcasting).
- The HMOS approach (as tested to date[2]) applies to deterministic forecasts and requires the existence of consistent retrospective forecasts (or archives of operational forecasts to capture all MODs. In all existing implementations (ie, those evaluated by OHD), it has been developed for short range forecasts only (2-3 days) because of the rapid decay of forecast skill for fixed lead times. Additionally, the routing of HMOS uncertainty from segment to segment has not been explored. HMOS could possibly be recoded to expand its capabilities (e.g., operating on variable length future periods or applied to output ensemble means), but because

[2] As noted elsewhere in the report (such as the component sections of Section 3), deficiencies described in the current post-processing approaches (HMOS, EnsPost and MSCM) could be reduced by further research, development and testing (e.g. of variations in implementation), as well as by augmentation with other approaches. This report focuses on components in the state for which sufficient test results were available by which to judge their potential for application in version 1 of HEFS.

such adaptations have not been tested, we consider the current implementation. The limitations described above undermine objectives 1, 2 and 3.

- The MSCM approach is applicable across all time scales, but only performs a bias-correction (i.e., it corresponds to a specific case of EnsPost for which the regression parameter is equal to 1) and does not ensure reliable (i.e., probabilistically calibrated) forecasts. This limitation undermines objective 2 and 5.
- The current MMEFS approach uses raw interpolated NWP output that is known to be biased and unreliable, and does not currently include any blending of the ensembles from the different NWP models. (i.e., flow ensemble outputs from each individual NWP model can be produced and routed downstream). This limitation undermines objectives 2 and 4, and to some degree objective 3 (as no blending is performed between the different NWP model outputs).
- No attempts have been made to develop a fully-integrated, CHPS-based ensemble forecast environment which incorporates and integrates ensemble forecasting tools and outputs into the workflow for daily operational forecasting.
- Although multiple RFCs have provided prototypes for web presentation of ensemble forecast information, the feedback still is being gathered and has not been fully evaluated to help determine presentation requirements. Efforts to determine standard verification products have been initiated with the NWS Hydrologic Forecast Verification Team and will require additional concerted efforts, as well as further experience within the RFCs. It is possible that a uniform product application may not be a realistic or desirable objective.

As detailed above, capability gaps exist in the areas of ensemble forecast post-processing and automated data assimilation. Operational data assimilation is a challenging area in which it is likely that continued research beyond the scope of the HEFS project will be needed to provide desired functionalities (i.e., substantially reducing use of MODs). In contrast, ensemble forecast post-processing is an area in which a range of more readily developable solutions may exist, particularly for lead time longer than those addressed by the current prototype component versions (e.g., EnsPost and HMOS) that have been tested to date. We recommend defining and undertaking a focused science development/testing effort toward improved ensemble post-processing in parallel with the initial stages of software development /system engineering on the required components of HEFS. Inclusion of the “high priority” ensemble post-processing component in HEFS V1 would be contingent on successful testing/evaluation and the overall project implementation schedule. Individual component strengths and weakness are discussed in greater detail in **Section 3.**

3. HEFS SYSTEM AND COMPONENT REQUIREMENTS

3.1 System Level Requirements

This section defines requirements that apply to the HEFS system as a whole, including both specific functionality and also general design principles.

3.1.1 CONCEPT OF OPERATIONS (CONOPS) SUMMARY

Real-time operations

Ensemble forecasts may be run in bulk (fully automated), interactively down to the segment level, or in a process that is a mixture of batch processing and interactive, manual effort. Thus operational and hindcast mode activation of ensemble forecast components must be scriptable as well as offer interactive controls through the CHPS IFD.

The Hydrologic Processor capability, or CHPS deterministic core, must also be extended in minor ways. In particular, CHPS must have the ability to impose a subset of MODs on each ensemble member to represent as closely as possible some deterministic aspects of the river system. For example, SETQ mods that are used to incorporate known reservoir controls and/or diversions in the short range must be applicable within the ensemble streamflow forecast context. This capability is available within CHPS, but is not easily accomplished within NWSRFS (e.g., for ESPs), or in any of the current experimental ensemble implementations.

Running ensembles as a batch process, the forecaster should be able to configure which preset MODs (perhaps resulting from a deterministic CHPS run) will be applied during the run, as well as all necessary control inputs (start and end times, warm states, etc.) and configuration settings for all CHPS modules used in the deterministic runs. Not all MODs used in real-time operations can be handled in this fashion. Examples of those that can include blends and CHANGETS or SETQ time-series, such as for diversions or consumptive uses. Those that cannot realistically be handled this way include SACCO and MFC MODs. Ultimately an automated data assimilation capability and/or expert systems approaches may be optimal to address these latter mods, but this issue reflects the long-standing difficulty of incorporating objective techniques into a real-time process that remains extensively dependent on the use of manual modifications to the forecasting system.

Running ensembles as an interactive process, the forecaster should be able to launch and control all aspects of the hydrologic ensemble generation process within a CHPS workflow at the segment level. Hydrologic ensembles could be influenced by any input data or state adjustments made by the forecaster; that is, the forecaster may make MODs (such as a WECHNG) and then generate new ensemble outputs that reflect the MODs. Desirable controls within each component are detailed in component specific sections.

The fully batch and hybrid batch-interactive configuration are the most likely configurations for ensemble hydrologic forecasting. The fully batch mode is exemplified by the way MMEFS is currently run: i.e., as a fully automated process without any manual intervention or MODs. Due to the time involved in generating ensemble meteorological forecast forcings, a hybrid run is likely to involve generating these in batch before the interactive IFD run is begun. The interactive run would then allow the forecasters to make MODs and trigger the RFS models to run for a segment using the ensembles (as well as the deterministic forecast forcings). The forecaster may evaluate and further modify the ensemble forecasts, possibly re-running them, before proceeding downstream to the next segment. As in the deterministic run, evaluation is typically a subjective process leveraging the forecaster's knowledge of the basin and recognition of upstream and downstream simulation and forecast results, but it may also include objective measures or ancillary information. Ensemble forecast graphics and suitable diagnostics will be generated in IFD to inform the forecaster decisions. Additional higher quality graphics and products may be generated at the end of the interactive run.

Real-time MODs are a regular and widespread feature of short range river forecasting operations at all RFCs. Some MODs apply only to observed or simulated variables in the past (i.e., the model spinup period, prior to the forecast period). These MODs do not directly operate on variables during the forecast period, but indirectly affect the forecast period outcomes by changing the model states used to initialize the forecast. Examples of such MODs include SACCO (modifies model soil water tank contents on a given date) and WECHNG (scales model snow water equivalent on a given date). Other MODs can operate on model variables both in the spinup period and the future period. Examples of these MODs include MFC (changes model melt behavior for a specified period), CHGBLEND (modifies the connection of observed and future flows), and SETQ (defines flows in the basin water balance, e.g, reservoir outflows or diversions).

In OFS, MODs applied for flood forecasting are generally NOT included in long range (seasonal) ESP forecasts, except for those that have affected the initial forecast states. In CHPS, it will be possible to include some or all of these flood forecast MODs in future ensemble simulations (including seasonal ESPs) by pre-configuring the ensemble simulation workflows. As with the short-range forecasts, forecasters will need to be able to see which MODs are in effect (such as a MFC that extends into the future) and to alter them depending on the simulation results. Thus the real-time controls that are available now to forecasters in CHPS for short range deterministic forecasting, should be available to forecasters for ensemble runs, regardless of their length. As noted earlier, the application of MODs in the ensemble context (or in any other objective context) is a vexing problem for the river forecasting centers. We speculate here as to how forecasters may try to reconcile MOD use with HEFS component use, but recognize that this is an area that poses significant unknowns for RFC operations.

Additional real-time controls (effectively MODs) that alter the application of HEFS component techniques (such as EPP3, EnsPost or Automatic DA) are likely to be desirable. These would also need to be available to a forecaster, possibly through the same modifier dialogue that contains the existing MODs. Many should also be available at the group level. For example, a real-time control (MOD) of whether to use 2 days of RFC forecast versus 5 in creating an EPP3 ensemble should be applicable at the group level, rather than need to be set within every segment.

It is important to note that OHD and the RFCs have limited end-to-end experience using the HEFS components within the operational environment (especially the CHPS operational environment) or with the integration of the current manual and deterministic and probabilistic forecast processes and information. The use of HEFS may vary depending on the variation in forecasting demands: e.g., during flooding, forecasters may pay little attention to the longer term effects of MODs or real-time controls, whereas during non-flooding times, forecasters may spend more time to evaluate the effects of model state changes not only in the week ahead, but for water supply range forecasts. The overarching requirements recognize the need for flexibility, extensibility, and modularity in the design of the HEFS components. The speculative nature of the CONOPS for HEFS underscores this need.

HEFS component configuration

The components of HEFS have different input requirements to support configuration. Note, configuration refers to off-line (not real time) efforts to determine settings and parameters of the HEFS components, and not real-time control or modification of settings within ranges that are allowed for each component. For example:

- Meteorological ensemble forecast generation requires a small number of date or scalar settings (e.g., number of days of forecasts to use from one NWP source) as well as access to significant volumes of retrospective forecasts (meteorological hindcasts) and corresponding meteorological observations.
- Hydrologic ensemble forecast post-processing configuration also requires simple settings as well as access to retrospective hydrologic simulations and observations, and ideally also to significant volumes of hydrologic hindcasts made using hydrologic models and the meteorological hindcasts (for alternative hydrologic post-processing techniques that use the ensemble forecast information).
- Configuring the data assimilation component (using VAR techniques but not ensemble techniques) requires access to retrospective model simulation state and output variables as well as corresponding hydro-meteorological observations, but not to large hindcast datasets (which may be required for ensemble techniques).

Thus all three of these components must train their configuration settings based on retrospective (historical) simulations and/or hindcast datasets of hydrometeorological variables. The ensemble verification and product generation components, in contrast, operate on these datasets, but do not require them for configuration: configuration entails specifying the types of statistics or plots to be generated and dimensions of hindcasts to expect, and other simple, mostly scalar quantities. Configuration of these components involves specifying a manageable number of settings.

Data management and access is an important consideration in designing the requirements for configuration (parameter setting) of the first three components mentioned above. It is currently possible to generate meteorological ensemble forecast generation parameters (at least for EPP3) outside of CHPS, and would require less of a development effort to import the parameters rather than to incorporate the parameter configuration process within CHPS and afford it access to meteorological hindcasts. The advantages of maintaining parameter configuration for data assimilation and post-processing outside of CHPS are less clear, however – especially because the hydrologic hindcasts that may be needed for post-processing configuration would be generated from within CHPS and archived in a CHPS dataset.

A concept of the configuration workflow that leaves meteorological ensemble forecast processor (MEFP) configuration (i.e., specifying settings and parameters) outside of CHPS includes:

- Forecasters access most simple (dates, scalars, etc.) configuration settings within CHPS IFD (for example, from a configuration button included in the toolbar at the top left with the other major functional control groups). This button launches an array of tabbed widgets/pages/dialogues corresponding to each component's configuration inputs.
- CHPS imports timeseries-based configuration parameters for the meteorological ensemble forecast processor, with other selected settings available for modification through an IFD-based configuration dialogue.
- Assuming that hydrologic hindcasts have been generated, the hydrologic ensemble forecast processor could be configured through an IFD-based configuration dialogue.
- EVS configuration can occur outside of CHPS, as it mostly involves setting up a batch input file that determines the statistical analyses and verification computations that will be performed on each segment.
- Data assimilation can be configured through its own IFD-based configuration dialogue; and
- Product generation can be configured through its own IFD-based configuration dialogue.

Alternatively, if meteorological ensemble forecast processor parameters must also be configured within CHPS, the workflow also includes the following step.

- CHPS imports the full historical archive of meteorological hindcasts (e.g., gridded sub-daily timestep time-series) as needed to configure the meteorological ensemble forecast processor, and either discards them after generating parameters and diagnostic plots/results, or stores them in a way that does not slow operations.

All configuration parameters would receive a version-identifying tag and are made available via those tags to the components for use in real-time operation and hindcast modes. Multiple versions of configuration parameters can exist, or at least be archived and retrieved.

Hindcasting

The production of streamflow hindcasts of sufficient length for forecast performance evaluation (approximately 30 years minimum) is a requirement for the NYCDEP project, for which the hindcasts will be used to develop optimal water management operations guidelines. Hindcasts are also required to support forecast verification capabilities and would expand the range of methods available for post-processing the hydrologic ensemble forecasts by using all the ensemble hindcast information (not just the hydrologic simulations).

Hindcast production is an intensive process, and therefore has never been a routine activity in association with operational forecasting within NWS. Currently, NWSRFS affords the capability to run streamflow hindcasts (i.e., RFS models are run in a loop using retrospective hydrometeorological observations and forecasts, if available) for most operations (but known limitations e.g., for reservoir modeling), which prevents it to be run for all forecast points. As long as NWSRFS remains implemented with current operational configurations of models – that is, those being used in HEFS – it could be used to provide hindcasts for HEFS for most forecast points. This situation will not be viable indefinitely, particularly with the full-scale operational adoption of CHPS and phase-out of NWSRFS. Thus HEFS components or CHPS will be required to support hindcasting. In addition, the model configurations and data feeds must remain identical to those in CHPS for the hindcasts to be valid – this is a significant obstacle, if not impossible.

Hindcasts will ideally be run each time that major model or system configuration changes occur for those portions of the system affected by the changes. Because they are labor/CPU intensive, they are unlikely in practice to be run more than once or twice each year. Hindcasts will certainly be batch processes without forecaster intervention (i.e., manual MODs).

Alternatives for hindcasting practices depend on whether HEFS ingests raw meteorological ensemble forecasts and calibrates them (generating MAP and MAT ensembles) or calibration occurs externally to CHPS and HEFS ingests calibrated ensemble forecasts. Either way, the forecaster will first run external programs to generate the calibrated meteorological hindcast inputs (outside CHPS), and then use CHPS/HEFS to first generate carryover states (initial conditions) and then run and process the hydrologic hindcasts, involving hydrologic simulation, and potentially the ingest of prescribed MODs (e.g., SETQs, TSCHNG), hydrologic post-processing to generate calibrated hydrologic ensembles, verification, and analysis/graphics generation.

3.1.2 Requirement Summary

Overarching requirements applicable to all HEFS component software include:

- User friendliness, providing: (a) access to guidance, help, tips, etc., -- at all forecaster decision points; (b) default settings for all inputs required to execute HEFS components, or as appropriate; (c) comprehensive error messaging and generation of diagnostics, where appropriate.
- Flexibility, modularity and extensibility. Because much of the exact functionality proposed for CHPS/HEFS to date has not yet been run operationally, it is very likely that the requirements discussed in this document will evolve. As experience in running and evaluating the performance of components is gained, the components of the system will also need to evolve (sometime rapidly), thus flexibility is a critical principle of the software design. If components are modular, upgrades can be targeted and unintended consequences due to dependencies can be controlled and minimized.

- Components must be configurable for operations and hindcasting in accordance with the CONOPS selected for this process (i.e., internal or external to CHPS, in some cases – for example, if EPP3 parameter configuration remains outside of CHPS for HEFS v.1). The configuration tools (or tool with subcomponents, or component of CHPS) would organize the various programs for configuration that pertain to each component, and provide a diagnostic "smart" tool to evaluate the configuration and guide configuration improvements. Because this process demands more knowledge and expertise of statistical concepts and methods than anything RFCs do today, and the interface will need to teach forecasters to some extent during the process. These tools should streamline the configuration process (e.g. push the "config EPP3" button and return some indication of the quality of the configuration.
- Controls for all components need to be implemented in a consistent XML schema that supports user-defined input or parameter validation (e.g., can detect if a year variable is a valid year).
- CHPS-HEFS components must provide functionality for an end-to-end hindcaster capability that will support verification and possibly additional forecast products, such as analog information[3] from past historical events using the current forecasting system, especially for rare events. Some sophisticated users (NYCDEP) will need hindcast datasets to calibrate their own decision support system. Hindcasts must be supported by the MEFP (EPP3), DA, hydrologic processors, and EnsPost. Hindcast products could be generated by the Graphics Generator (GraphGen) components.
- Training modules must be prepared for all components and for the system as a whole. The training modules must include both science-oriented components and "knobology" (mechanics of operating component controls). To the extent possible, help information should be available at every stage of component implementation within CHPS (configuration, operation, hindcasting and verification).
- Documentation of all HEFS components is required. This documentation should summarize the concept behind algorithms in use, give examples to illustrate performance, instruct forecasters in component configuration, and reference existing publications related to the components. It should also provide a brief background on the hydrologic topic (e.g., data assimilation) to which the component relates.
- Archiving of inputs and outputs (to the extent possible), as well as auxiliary information that could be used to select analog forecasts, and specifications or metadata associated with each forecast run (i.e., versioning info).
- All ensemble processes that take significant time (~20 minutes or greater, e.g., ensemble generation or ensemble forecast runs, and especially hindcasts) should be cancellable and exit gracefully. This may also apply to configuration activities.

---

[3] An analog assessment capability will require the archiving and retrieval of sufficient information to compare current events with similar events in the past using various criteria (storm type, climate indices, etc.). The analog assessment capability is considered a future priority.

- Publishing and storing ensemble input or output or run information should be separate from deterministic run information. If a forecaster decides not to proceed with an ensemble process, it should not affect the ability of the forecaster to proceed and store the deterministic process inputs, ancillary information (e.g., MODs) or outputs.
- Hindcasting is a CPU-intensive processing task that generates large volumes of output and consumes significant CPU and bandwidth. The software needs to be able to assess current hardware resources (storage, possibly network bandwidth, CPUs) and determine if a hindcast run is feasible or advisable, and provide feedback to the forecaster.
- Hindcast runs may need to leverage a central computing facility for the most CPU or storage intensive tasks. The topic of a central facility for archiving hindcasts has been raised in discussions during the preparation of this report, but is not addressed in any detail in this document.
- Components' ensemble activities must be able to be automated (e.g., piService, cron, etc.).
- All components must be able to evaluate inputs and outputs not only for numerical consistency with expected data requirements, but for quality, and provide feedback to RFCs to indicate when results may be compromised and for what possible reasons.
- The configuration tool should limit the degrees of freedom (i.e., the knobs exposed to the forecasters) to the most sensitive ones, initially. A tradeoff exists between affording extensive control over the functioning of a component and the danger of overwhelming a forecaster. Configuration parameter should minimize the need for ad hoc judgment applied on case by case basis, and must be accompanied by default settings to initiate calibration programs.
- A functioning but possibly "version 1" level archive system: especially critical for storing ensemble reforecast datasets of forcing input ensembles and hydrologic output ensembles, and supporting multi-year forecast verification.

3.1.3 Possible Future or Deferred Requirements

Some requirements may be wholly or partially deferred until a "Version 2" stage of HEFS. Deferring HEFS build activities, functionality development and extension may be advisable to allow experienced gained during the use of "Version 1" to dictate specifications for the future version. For the system as a whole, possible deferred activities include:

- Building a comprehensive archiving system that encompasses inputs/outputs of HEFS components, as well as auxiliary information for real-time verification. Because ensemble forecasts using all components of HEFS have not yet been produced on a routine basis operationally, it is still unclear what practices and requirements will arise for archiving related outputs.

- A comprehensive, unified component configuration tool. A desire for such a tool was expressed by various contributors to the report, but designing the configuration tool design will also require a range of experience with the components that is not available at the time of this report.
- A unified verification tool that can process both deterministic and ensemble forecasts. To some extent, this capability can be configured within CHPS using the GraphGen and EVS components, but further conceptualization of how this capability is provided in the CHPS may be warranted.
- An analog identification, analysis, and display tool for operations. The idea of accessing and evaluating analogs to current events from prior forecasting situations (i.e., hindcasts) was raised frequently by contributors to the report. This practice does not exist in operations yet, thus an exploration and development effort is recommended.

3.2 Meteorological Ensemble Forecast Processing (MEFP)

Meteorological ensemble forecast processing (also called “Ensemble Pre-Processing” or “EPP” in related NWS documents)[4] is the essential input to any ensemble hydrologic forecasting system. The HEFS component that accomplishes MEFP transforms raw gridded meteorological forecast ensembles that, in practice, always suffer from bias and spread deficiencies into meteorological forecast ensembles that are calibrated (unbiased, with reliable spread) to the input characteristics (temporal and spatial scale, geographic unit, location) of the hydrologic model. The concept of statistical calibration is the defining feature of the MEFP component.

3.2.1 CONOPS SUMMARY

Real-time operations

Several cron operations to download and process real-time ensemble forcings would run automatically. Currently, with EPP3, the GFS reforecast and CFS forecast ensembles are processed into raw (uncalibrated) MAP and MAT ensembles for each segment. These have the same format as the calibration forcing data. RFC deterministic forecasts would also be processed (as in the current CNRFC application), which, rather than downloading from an external site, as is the case of the GFS and CFS forecasts, involves the extraction of the RFC deterministic forecast from the database. In the future, other sources of single-valued

[4] It may be worth reconsidering some terms that have been used in prior specifications for HEFS. E*nsemble pre-processor* (EPP) is imprecise, as it does not specify the type of ensemble (hydrologic/meteorological) and pre-processing in the hydrologic context is post-processing in the NWP context. The terms should be clear and specific across disciplines, particularly where scientists and practitioners from those disciplines may interact. Similarly, “*ensemble post-processing*” could be changed to “ensemble streamflow forecast calibration [or hydrologic ensemble post-processing]”. Note that post-processing in hydrology would be considered pre-processing for water management and decision support.

forecasts will also be used (e.g., GEFS ensemble means, CFSv2 ensemble means from lagged ensembles).

The MEFP could then generate ensemble forcings via a batch process that runs in the CHPS background, once all meteorological inputs (RFC forecasts, model forecasts) are available. These would be stored in the CHPS datastore to be accessed either interactively during a CHPS deterministic model forecast run, or during a batch run that follows the deterministic run. It is likely that MEFP ensemble outputs would be archived to some extent (ie, for at least a subset of dates) to facilitate verification analyses. In the interactive case, the forecaster would activate segment operations using the ensemble forcings and assess their quality, perhaps deciding to modify the segment or change the deterministic run, before re-running the batch-generated ensemble forcings through the models and proceeding to the next segment.

The forecaster may also want to change the blend[5] between inputs of the MEFP, and regenerate them for a particular segment, in which case MEFP would be run interactively at the segment level. MODs would only be allowed to MEFP to the extent supported by the MEFP configuration parameters.

No capability has as yet been explored to allow additional controls on MEFP. These might include the ability to scale the spread of the ensemble for specified time periods, or to adjust the central tendency of the ensemble toward or away from a deterministic forcing (MAP and/or MAT). Such controls would produce an altered forcing ensemble that could undermine space-time coherence resulting from the Schaake Shuffle (Clark et al., 2004) if the rank structure in time and space is not preserved, but it may be a desirable MOD nonetheless. It is possible that such MODs could be based on objective criteria, such as the behavior of analogues, but it is likely that other considerations will also influence the decision to modify an ensemble forecast – e.g, the outlooks provided by alternative meteorological forecasts, such as from a WFO.

The forecasters would not typically alter the MEFP configuration settings, but may wish to change the application of the MEFP through MODs that are a separate set of real time *control* files. For example, the default "policy" may be to use RFC forcings for the first 5 days, then switch to GFS through lead day 14. The configuration parameters would have been developed to support this application, as well as others, including, perhaps, the use of RFC forcings for only 2 days followed by GFS. MODs could thus regulate some aspects of the ensemble processing and/or application to the hydrologic simulation.

Component Configuration

Configuration files contain the parameters that govern MEFP, and it is not anticipated that these parameters will be changed often. They will probably be established by the DOH (or whoever is managing the ensemble forecasts) using configuration tools, and may change seasonally, as model forcing archives are improved or altered, or for other reasons.

---

[5] For example, EPP3 could be configured for 14 days from the GFS reforecast and 240 days from the CFS reforecast, but the forecaster could reduce these time periods in operations.

MEFP current configuration requires organizing the proper archives of past forecasts and observations, then running several configuration programs that generate the parameters used by MEFP ensemble forcing generator programs for precipitation and temperature. Several programs for analyzing and plotting the results of the configuration exist, and these and/or other locally developed analyses would be run to ascertain the quality of the configuration. Metrics of interest may pertain to either the quality of the meteorological forecasts or the quality of hydrologic forecasts derived from them.

Configuration could either take place within CHPS, requiring that all hindcasts be available through the CHPS interface, or outside of CHPS using the existing EPP3 software. The key parameters would then be imported to CHPS before use in real-time or hindcasting operations. Either option is possible and will enable a functional HEFS system, though the organization, format and storage of MEFP files will differ depending which option is adopted. The choice of option may depend on the resources/time available to HSEB during the HEFS v1 development period.

Hindcasting

Hindcasting use of EPP3 (the only component that currently supports MEFP in a hindcasting context) could be executed separately from performing model runs. In the first case, the RFC would generate 25-30 years worth of calibrated retrospective forecast MAP and MAT ensembles, which should be stored in the CHPS datastore. These should be analyzed for skill and error characteristics as a central part of the evaluation of the EPP3 configuration. The EPP3 hindcasts can also be generated in sequence with model-runs, streamflow post-processing and other HEFS components, in which case the outputs would be archived in the CHPS datastore or imported from the disk to support the model execution. Hindcasts would likely be performed by the RFC approximately once per year or when major data/model change were implemented; and they would be scheduled not to conflict with significant real-time operations. Currently, hindcasts can be performed outside of CHPS, but it may be possible to implement CHPS drivers for hindcasting that is controlled by a CHPS IFD dialogue.

3.2.2 Requirement Summary

This section is intended to give readable outline of major requirements.

HEFS version 1

The MEFP component's overarching functional requirement include the following:

- It must generate short- to medium-range (2 week) ensemble forecasts of mean areal watershed-scale (segment) precipitation and temperature to serve as forcings for operational hydrologic models that will generate streamflow ensembles; and
- The ensemble forecasts must be calibrated, which means that the probability distributions at varying time-scales are reliable while still being as sharp (narrow) as possible. Reliability means that the forecast probabilities verify with observed frequencies over a sufficiently long period of time to generate robust statistics.

- The ensemble forecasting technique must support hindcasting using consistent data sources and methods, to the extent possible. Hindcasting and verification need to be performed on all segments to validate the MEFP configuration and to describe expected performance to support forecasters and users' decision processes.

Subsidiary functional/technical requirements include:

- Forecast leads must ultimately span from 1 hour from present to 1 year from present. Initially, the application must support 6-hourly timestep forecasting, with the 1-hour timestep capability being developed as second priority.[6] A two-year lead time is also an objective in the western US but is a lower priority.
- MEFP ensembles must be able to be generated on local, UTC, or other desired time clock.
- MEFP must be able to be run as a pre-process (in batch mode) for more efficient forecast operations.
- Configuration of parameters necessary for MEFP approaches may occur within or outside of CHPS, but CHPS will facilitate the storage, maintenance, querying and analysis of those parameters. Configuration programs will provide sufficient diagnostic information to support evaluation of the configurations if desired. If configuration occurs outside of CHPS, the import of configuration settings will be supported.
- A training module must be developed to instruct forecasters in configuration as necessary.
- A graphical interface to assist in running configuration programs to optimize component settings will be developed (perhaps patterned after the Python-based Lhotak GUI described in the EPP3 documentation). This interface may or may not be within CHPS, depending on the priority assigned to embedding the capability in CHPS. Configuration parameters should be evaluated in terms of their importance to achieving successful implementation of the MEFP approach, and training/GUI effort prioritized accordingly.
- MEFP approaches will provide diagnostic information to support monitoring of meteorological ensemble forecast performance by forecasters.
- For real-time and hindcasting use, forecasters must be able to control the time period of application of forecasting information from different sources (e.g., specify 2 days from the RFC with 5 days of GFS) at the segment level (interactively) or in batch mode. Such implementation changes are possible without reconfiguring

[6] At CBRFC, for example, all segments in the southern half of the forecast domain, including key segments that flow into Lake Mead, are run at a 1-hour timestep. Although simulation of Lake Mead does not require 1 hour timestep inputs, adjustments to the RFC models such as the ones the upper Colorado River basin 1-hour segments will be necessitated if 1 hour MEFP is not available.

MEFP parameters, within limits (e.g., if EPP3 is configured for use of GFS out to 15 day lead times, it can be applied for use of *any* period of GFS up to 15 days).

- MEFP ensemble members must be archivable with a unique identifier which indicates the source and time periods of component forecasts, among other details; or else metadata for the ensemble must be associated with the members.
- MEFP data ingest components must be flexible enough to upgrade in a timely fashion when it is necessary to make use of new versions of climate forecast sources: e.g., the new GEFS reforecast, CFSv2, NAEFS, etc.

Future Requirements / Recognized Needs

We recognize that the application of HEFS components may lead to discontinuities at spatial and temporal boundaries between application areas, and future work should be considered to reduce these discontinuities

EPP3's use of the current Schaake Shuffle technique limits the number of ensemble members to the number of years in the historical calibration data set. For example, given a 50-year historical data set, one could generate nearly 50 ensembles (losing a few at the edges). Running EPP3 for upstream and downstream basins for which the historical dataset does not cover the same historical period will limit the ensemble size. Enhancements to the Schaake Shuffle technique could allow the generation of more ensemble members.

The current version of MEFP may consist of the EPP3 using the three forecast sources noted above. This capability should be flexible enough to use single-valued forecasts from other forecast sources. In the future, MEFP should be extensible to include functionality that merges (via model averaging or combination) meteorological forecasts from multiple sources, provided they satisfy the hindcasting requirement.

3.2.3 PROTOTYPE READINESS LEVELS

This section discusses the readiness levels of the prototypes currently in use or development within NWS. Two credible pathways exist, EPP3 and MMEFS. A third pathway, "QPF Surfaces"[7], is not considered a credible approach and will not be discussed further in this document.

Of the two pathways, MMEFS & EPP3, only one pathway, EPP3, is capable of meeting all of the overarching requirements. EPP3 processes 6 hourly forecast ensemble means from two climate model sources (GFS, 0-14 days) and CFS, 0-240 days), from deterministic RFC

[7] "QPF Surfaces" describes the practice of forcing hydrologic models with MAP and MAT timeseries constructed of extreme probability QPF surfaces from HPC (e.g., 5% and 95% exceedence probabilities). The assigned probabilities of these surfaces correspond to approximately 6-hourly timesteps for a nominal spatial resolution of 32 km, thus the joint likelihood of these forcings (and the streamflow produced by them) becomes progressively more extreme as they are aggregated across space (multiple catchments) and time (multiple timesteps) and integrated by the hydrologic process. The QPF surfaces cannot in theory produce reliable uncertainty bounds when used in this way.

QPF forecasts, and climatological ESP forcings (CLIM) and generates ensemble forecasts that seamlessly combine the RFC, GFS, CFS and CLIM information. It has separate sets of modules for precipitation and temperature forecasts, including parameter generation, ensemble generation and analysis. Currently, EPP3 ensemble generation component can be executed within the CHPS environment through an EPP3 model adapter. A number of ancillary programs also exist to extract and reformat RFC, GFS and CFS forecasts from their native formats. EPP is presently limited in generating ensemble members only for the number of calibration forcing years available at RFCs (which must match between RFCs sharing a river basin, e.g., the Mississippi R. basin).

EPP3 has been evaluated and found to produce reliable meteorological forecasts for a range of individual river segments in more than one RFC. Aggregate performance over spatially nested segments has not been comprehensively tested, although the evaluation, which is nonetheless promising, has focused on a range of performance characteristics. The EPP ensemble forecasts have been evaluated for the following conditions: seasonal verification, verification for different thresholds, verification at 6-h and 24-h time steps, and verification with spatial aggregation (up to 10 segments). Hydrologic forecast evaluation was done using the RFC single-valued forecasts up to 5 lead days into the future, and using the GFS ensemble means up to 14 lead days in the future. The EPP-ESP (and EPP-ESP-EnsPost) streamflow ensembles have been verified at 24-hr time steps (as mean daily flows) up to 14 lead days into the future and for different flow thresholds. The EPP3 software is now running operationally, though in experimental form, in several RFCs and resulting streamflow forecasts are considered plausible.

The EPP3 code suite is not thoroughly understood within OHD except by the creator, although this situation is rapidly changing as members of the OHD HEP and HSEB groups work with the software.[8] **For the EPP3 application as a whole, the applied science readiness level is therefore MEDIUM, and the software readiness level is also MEDIUM.**

A second pathway, MMEFS, processes 6-hourly forecasts from output precipitation and temperature grids of the NAEFS (combining GEFS and CMCE) and also from SREF. These forecast grids are currently interpolated to create MAPs and MATs, and thus have no bias-correction or calibration step. This approach could also make use of gridded inputs from the new GEFS reforecast (expected to be available in summer 2011). MMEFS cannot support hindcasting using its present data sources due to lack of raw hindcasts available.

MMEFS has been implemented operationally in four eastern US RFCs, but little comprehensive evaluation of the forcings or flows has been performed. The aforementioned calibration step is lacking and it is widely known that the direct interpolation of current weather or climate model-scale NWP outputs for forcing a hydrologic model is inappropriate. Developing a calibrating and/or downscaling strategy

[8] John Schaake retains a comprehensive knowledge of the code and can provide valuable and efficient guidance in working with or upgrading EPP3 code. To avoid delays in the development of an enhanced and operational EPP (upgraded to process the new GFS and CFS reforecasts), it may be worth considering retaining some of his time as a consultant for a limited number of hours during version 1 development.

(correcting bias and spread) is necessary in order for this approach to meet service-level objectives. Thus, significant science development and testing effort remains, despite the existing experimental operational implementation. **For the MMEFS application as a whole, the applied science readiness level is therefore LOW, and the software readiness level is also LOW (given missing components).**

3.3 Data Assimilator

The data assimilation capability is a natural extension of the hydrologic (and hydraulic and water management) system and is not specific to the ensemble forecast context. Currently, a de facto data assimilation is accomplished at all RFCs by a manual real-time modification (MOD) process. The inclusion of a data assimilation capability in the ensemble Forecast Service has two primary motivations: (1) some types of manual MODs will not be possible for ensemble simulations, i.e., forecasters simply do not have time to modify multiple forecast traces; (2) the MOD process results in non-systematic, ad hoc alternations of the simulation system and destroys the consistency between real-time and retrospective procedures – a consistency that is a basic assumption of the post-processing and verification components of the hydrologic ensemble forecasting system; and (3) the use of MODs makes it extremely difficult, if not impossible, to do hindasting. A comprehensive data assimilation capability, if possible, would replace or reduce the use of MODs.

The XEFS Data Assimilator (DA) is designed to generate optimal states for operational hydrologic models to produce improved ensemble snow and flow forecasts. In theory, it uses either deterministic or ensemble data assimilation algorithms to account for uncertainty in hydrologic models (e.g., model structure, parameters, and states) and observational data (e.g., model forcing and output). In operations, the data assimilator can provide objective guidance for the manual modifications (run-time MODs) made by human forecasters and may have the potential to automate a subset of these modifications by using the current observations in real time together with a quantitative assessment of various sources of uncertainty.

The DA prototypes being developed under the XEFS R&D activities include the one-dimensional variational data assimilator (1D-VAR) for a hydrologic river routing model, the two-dimensional variational data assimilator (2D-VAR) for the lumped SAC-SMA model, the four-dimensional variational data assimilator (4D-VAR) for the gridded SAC-SMA model, and the Ensemble Kalman Filter (EnKF)-based data assimilator for the lumped SNOW17 model. Another ensemble-based data assimilator based on the Maximum Likelihood Ensemble Filter (MLEF) is still under development and will not be discussed further in this document.

The 1D-VAR prototype assimilates river flow observations at both upstream and downstream locations of a river reach into the 3-parameter Muskingum routing model to adjust model parameters as well as downstream flow at the beginning of the assimilation time window to bring the modeled river discharge at the downstream location (a combination of routed upstream flow and lateral inflow to the river reach) into better agreement with flow observations.

The 2D-VAR prototype assimilates streamflow observations into the SAC-SMA, and adjusts mean field biases in precipitation and potential evapotranspiration (PE) data to improve streamflow analysis and forecasts at the outlet of a basin. The 4D-VAR assimilates streamflow at the outlet as well as interior locations, and, if available, in-situ soil moisture observations into the distributed SAC-SMA, and adjust mean field biases in the gridded precipitation and PE to improve streamflow analysis and forecasts at the outlet as well as interior locations in a basin. Note that the distributed SAC-SMA model is not widely used in RFC operations.

The current EnKF prototype assimilates snow water equivalent data into the lumped SNOW17 and generates ensemble snowmelt plus rainfall estimates, which serve as an input to the lump SAC-SMA. The EnKF has the capability to simultaneously consider uncertainty in model initial states, parameters, and forecasted model forcing (i.e., FMAP and FMAT).

3.3.1 Conops summary

Real-time operations

The 1D-VAR is applied on a segment by segment basis. In real-time operations, the 1D-VAR requires the most recent streamflow observations at both upstream and downstream locations of a river reach as input. It operates on a single-valued streamflow simulation and forecast at the downstream location of a river reach. It can also take ensemble river flow simulations at both locations from the hydrology process (e.g. ESP) configured with ensemble forcing. In such a case, the 1D-VAR provides ensemble streamflow predictions at the downstream location. The above procedure can be run as an automated part of the segment workflow. Forecasters may wish to exert real-time control over the application of 1D-VAR by specifying the assimilation time window, and possibly by rejecting the use of the 1D-VAR (i.e., turning 1D-VAR "off") to assess the accuracy of flow estimates with and without 1DVAR.

The 2D-VAR has been applied in a hands-off, automated mode in real-time experimental operations, and this would be the default mode of operations. It would likely be an automated step in a segment workflow. Forecasters would likely require an option to turn off the 2D-VAR, however, in order to allow the use of traditional MODs in the cases of rainfall-runoff events where simulated flows after assimilation show poor agreement with observed flows.

The 4D-VAR can be run in real-time operations at a user-specified frequency as a cron job and produces updated model states for the distributed SAC-SMA model at the forecast time. If the forecaster accepts the 4D-VAR updated states, the states are forwarded to generate streamflow forecasts. If the forecaster chooses to override all or part of the 4D-VAR updated states, all or part of the old state variables are reinstated via MODs. The 4D-VAR is run again until the forecaster satisfies with the results. All 4D-VAR updated states and streamflow forecasts are automatically archived.

Component Configuration

The integration of 1D-VAR into the CHPS is based on the Open source Data Assimilation software (OpenDA) with collaborative development efforts between the Deltares and the OHD. Configuration files include the OpenDA wrapper for the 1D-VAR prototype, the CHPS model adapter for OpenDA, the CHPS forecast workflows, and parameter definition files. The 1D-VAR-OpenDA has now been successfully tested within CHPS for an ABRFC test location.

The parameter definition file for 1D-VAR contains the calibrated parameters for the 3-parameter Muskingum routing model. These parameters need to be calibrated off-line by human forecasters using the calibration tool developed at the OHD, or some other methods developed locally. Calibration experiments indicated about seven minutes to calibrate model parameters with 10-yr data set from one river reach. Other parameters that require forecasters' control/adjustment include the length of the assimilation time window and the observational error variances for river flow at upstream and downstream locations of a river reach. These parameters are pre-specified by the forecasters in the definition file.

The integration of 2D-VAR, 4D-VAR, and EnKF into CHPS is similar as that of the 1D-VAR, except that a different OpenDA wrapper should be developed for each prototype individually. Ultimately, configuration would take place within CHPS, though template support (including default parameters) would be desirable. The OpenDA wrapper for the 2D-VAR prototype is planned to be developed by FY12. As such, the 2D-VAR can only be run in off-line mode instead of in the CHPS environment at the current stage. In contrast, the EnKF algorithm is already available in the OpenDA. The work remaining is to build an interface connecting the SNOW17 and SAC-SMA to the OpenDA. This work is projected to be finished in two years. Once 2D-VAR and EnKF are integrated into CHPS, the integration of 4D-VAR into CHPS will be decided depending on the progresses of distributed modeling work and RFCs' preference on the use of the distributed SAC-SMA.

Hindcasting

Hindcasting use of 1D-VAR, 2D-VAR, 4D-VAR, and EnKF could be conducted in the CHPS environment or off- line. In either case, the 1D-VAR hindcasts can be carried out in sequence with other HEFS components (e.g., EPP3, ESP, EnsPost, and EVS) or in a stand-alone form. Hindcasts can be performed as needed to find optimal estimates of observational error variance values being used in real-time operations.

3.3.2 REQUIREMENT SUMMARY

HEFS version 1

The data assimilator component's functional requirements include:

- DA components must be included as an automated part of a workflow, rather than be applicable as a MOD.
- DA components must be able to be run in a hindcast mode.

- DA components must support forecasting across a range of forecast lead times (short- to long-term).
- There must be an interface that allows human forecasters to turn on or off different data assimilators in HEFS operations, thereby accepting or rejecting the results from a certain data assimilator. This control must be possible at the segment level, and be 'sticky' (i.e., the last setting must be able to persist from one run to the next) or user-selectable, applying for an arbitrary length of time.
- For real-time and hindcasting operations, forecasters must have control on the length of the assimilation and forecasting time windows. These controls will require a partial or total rerun of the operations for a given segment and downstream dependents.
- DA trace output must be archived with a meaningful identifier, or metadata describing the use of any DA in the forecast process for a given run must be stored.
- DA components must provide diagnostic information and messaging if DA algorithm limitations or assumptions are violated.

Future Requirements / Recognized Needs

The 1D-VAR is limited by the fact that the tested prototype has only been applied to the assimilation of streamflow data to adjust parameters of the three-parameter Muskingum routing model. The 2D-VAR and 4D-VAR are only applicable to headwater basins at the current stage by assimilating streamflow, precipitation, and PET data. In the future, DA's capabilities should be extended to adjust initial states of operational hydrologic models (e.g., SNOW17 and SAC-SMA) by assimilating additional data including primarily snow and (in the few places where they exist) soil moisture observations. In addition, the DA functionality should be enhanced to incorporate ensemble assimilation techniques (e.g., MLEF, EnKF), which still require extensive exploration.

The OpenDA wrapper for the 2D-VAR -lumped SAC-SMA prototype is under development and is expected to be completed in one year. As such, the 2D-VAR can only be run in off-line mode instead of in the CHPS environment at the current stage. In contrast, the EnKF algorithm is already incorporated to the OpenDA. The work remaining is to build an interface connecting the SNOW17 model to the OpenDA. This work is projected to be finished in two years. Developing the OpenDA wrapper for 4D-VAR and its integration into CHPS will be contingent on the progress of distributed modeling work and the overall utility and priority of using distributed SAC-SMA at the RFCs.

3.3.3 Prototype Readiness Levels

Five different data assimilators have been explored with varying levels of effort and development at OHD: 1D-VAR, 2D-VAR, 4D-VAR, MLEF, and EnKF. The most mature data assimilation implementations from a testing and application standpoint use the 1D-VAR and 2D-VAR approaches. They have reached a sufficiently advanced readiness level to be considered for implementation in the first version of HEFS; however, because the benefits/utility of the specific applications tested are deemed limited, their overall priority for inclusion in HEFS v1 is "medium". The 1D-VAR approach is currently the only data assimilator that has been tested in CHPS environment (it can also be executed off-line).

Testing has included the following:

- The 1D-VAR evaluation was carried out for four river reaches in the WGRFC service area, which include LOLT2-CRKT2, RSRT2-TDDT2, BRVT2-BWRT2, and BWRT2-DWYT2. The evaluation was carried out with 11-year data set and the 1D-VAR operates on an hourly time step.
- The 2D-VAR (for the SAC-SMA model) has been evaluated at three ABRFC basins on event-based scenarios. It has also been evaluated at 23 WGRFC basins during a 10-year period at an hourly time step. The 4D-VAR (for the gridded SAC-SMA model) was evaluated for four basins in the ABRFC as well as five basins in the WGRFC on an hourly time step. These basins have various data period ranging from two to eleven years. Results show improvements in streamflow forecasts relative to raw model forecasts up to approximately 12 hours lead time, in general.
- The EnKF has been applied at OHD to assimilate SNOTEL SWE data into the SNOW17 model at two basins operating at a daily time step. The updated rain plus snowmelt output was input into the SAC-SMA model to generate streamflow predictions. Preliminary unpublished results from these exploratory experiments show promise to support improvements in streamflow forecasts at lead times out to a week or longer.

The 1D-VAR code is easy to understand, but its integration into CHPS via the OpenDA interface was not straightforward. However, the integration has been completed and deemed successful, yet the current efforts only focused on limited test locations. For the 1D-VAR application, **the applied science readiness level is therefore MEDIUM**, and **the software readiness level is HIGH**. The other data assimilation approaches have readiness levels reflecting lower stages of maturity in OHD testing.

3.4 Ensemble Post-processor

Hydrologic ensemble post-processing is a statistical application that corrects systematic errors in the mean and spread of raw model ensemble forecast output – a process that is analogous to the calibration of meteorological forcing inputs. As noted above, several streamflow post-processing algorithms have been developed by OHD – HMOS, EnsPost, and MSCM – and several others exist at lesser stages of development (e.g., Indicator Co-Kriging non parametric processor, General Linear Model – Post-Processor) [9].

HMOS directly models the total uncertainty associated with the operationally-produced single-valued streamflow forecast, and generates ensemble streamflow forecasts for any lead-time based on conditional simulation. HMOS performance chiefly depends on the skill

[9] These post-processing techniques, along with HMOS, are currently being tested and evaluated with the HEPEX post-processing testbed dataset. Initial results will be presented at the HEPEX workshop on June 7-9, 2011.

in the single-valued operational forecast, and to some extent on other factors such as QPF skill, basin drainage area, length of the archive and classification criterion as well. Therefore, HMOS performance is location specific. HMOS application in the ABRFC study region suggests successful application of HMOS (i.e., reliable ensembles) for the first 2-3 days of forecast lead time, after which the relationships between the single-valued forecast and observations degrade, consequently HMOS performance as well. However, the association between operational forecasts and observed flows is preserved in the ensemble forecast in the mean sense for the entire forecast horizon. It has not been applied with different length forecast events (e.g., 6 hour forecast periods for short leads, 2 day forecast periods for longer leads, etc.), which could extend the applicability of the technique. The algorithms in HMOS are applicable to longer lead forecast correction, but testing or application in that context has has only recently been undertaken (results were not available at the writing of this document).

As stated above, the HMOS approach in the form that has been developed and tested by OHD is dependent on the existence of consistent retrospective forecasts (or archives of operational forecasts to capture all MODs). The HMOS evaluation on ABRFC test basins was complicated by inconsistencies in these archived forecasts. These inconsistencies, not unique to ABRFC, include:

- Some years of archived forecasts were generated using a 24-hr QPF, while other forecasts were generated using a 12-hr QPF.
- Data set consists of forecasts that are issued at different times.
- Forecast MODs over the period of the record are inherently variable (e.g., general forecaster skill/experience improves over time, mods are made by different forecasters and different mods can be made even given similar conditions/events).

EnsPost calibrates uncertainty in an ensemble forecast using simulations, which are devoid of input uncertainty (neglecting the uncertainty in the observed forcing inputs); therefore, EnsPost is applied on forecast ensembles assuming the input uncertainty has already been accounted for (e.g., by the MEFP). EnsPost operates on single traces from an ensemble forecast. EnsPost addresses biases in the ensemble mean and spread if the raw forecast ensemble has biases similar to simulations (used for EnsPost configuration). EnsPost algorithm is a combination of the widely-used probability (or CDF) matching technique and linear regression in the Gaussian space, and models residual error using autoregressive model. Currently benefits are limited to short-range lead time, i.e., approximately 5-forecast lead days.

The HMOS and EnsPost approaches are not unrelated, in that they are both regressive and could be implemented in ways that reduce the distinction between them. It may be appropriate to regard them as two forms of a similar approach to ensemble calibration. Each can be extended in various ways that reduce their deficiencies in satisfying HEFS objectives. For example, HMOS parameters can be calculated for variable future periods to optimize the skill captured by the technique, and it can be applied to ensemble means from GFS instead of deterministic operational RFC forecasts (which may be in limited supply). These adaptations have not been tested as yet, but they join other options for expanding the hydrologic ensemble post-processing capability that can be incorporated into HEFS in the future.

MSCM performs a bias-correction on ensemble traces, rather than a full statistical calibration, but is applicable to unlimited forecast lead times. It consists of applying the probability (or CDF) matching technique at multiple temporal scales. Its algorithm could be combined with the EnsPost component to perform bias-correction beyond the short term, even if such approach would be an incomplete solution. For example, long-lead streamflow forecast post-processing via the CDF-matching approach was shown to be less skillful than a calibration based on the EPP3 temperature algorithm in Wood and Schaake (2008).

In summary, and to reiterate points made earlier, EnsPost, HMOS and MSCM, as well as other techniques under consideration for hydrologic ensemble post-processing exhibit different strengths and weaknesses relating to meeting HEFS objectives. But none of these techniques alone completely meets the desired objectives. Each can be extended in various ways to address current limitations, and some effort is currently underway to do so. When this document was prepared, however, these adaptation efforts were not mature, thus the assessed readiness of the prototypes was based on their current rather than possible future form.

3.4.1 CONOPS SUMMARY

Real- time operations

The ensemble post-processing component will be run as an automated part of the segment workflow, and forecasters will see post-processed flow results as the ensemble forecast completes. Forecasters may wish to alter the application of the post-processor, either by turning it off, or by modifying such parameters as are allowed without reconfiguring/retraining the algorithm. Forecasters may wish to see a display of both adjusted and unadjusted results from the raw flow ensembles. Metadata from the application should be storable, perhaps as part of a run summary file.

Component Configuration

Configuration could take place outside of CHPS, provided key parameters and settings could be imported into CHPS to support real-time and hindcasting runs. Configuration could also take place within CHPS, provided all diagnostic feedbacks and interactive controls used in configuration were made available via a CHPS dialogue page. In the latter case, substantial help and guidance information would be accessible via the same dialogue page.

Hindcasting

The hindcasting application of ensemble post-processing would also be an automated part of the segment workflow. Post-processing real-time control settings, and on/off decision, should be available at the segment or group level to apply during an entire hindcast. Metadata from the application of the post-processor on a given run should be storable, perhaps as part of a run summary file.

3.4.2 Requirement Summary

HEFS version 1

The ensemble post-processing component must:

- correct or reduce systematic bias and spread deficiencies across multiple lead times (out to seasons), and particularly in the short to medium ranges that may be considered in the flood forecasting process.
- Strive to recognize differences in flow regime, if possible (e.g., rising & falling limb error categories), in training and application. Because this is challenging, initial versions of the component may fall short of this objective.
- As noted in the System Requirements section, be modular and extensible to allow for a variety of post-processing approaches to be incorporated as they are developed.
- Provide diagnostics on algorithm performance, and particularly notify forecasters if algorithm assumptions are violated.
- Be applicable to flow segments downstream of reservoirs, particularly where local or intervening flows contribute significant uncertainty.
- Be configurable by forecasters (not just OHD), via a step-wise or otherwise straightforward process.
- Calibrate the forecast ensemble such that the results may be routed to downstream segments.
- Be applicable in a hindcasting mode.

In addition, sufficient training and documentation must exist that the approach is not a black box to forecasters.

Future Requirements / Recognized Needs

Currently, no prototype component at OHD provides a streamflow calibration capability that extends to the desired time range of the HEFS products (i.e., beyond 7 days). This gap may be rectifiable within the scope of HEFS version 1. Ongoing work to support the HEPEX post-processing workshop slated for June 2011 may indicate techniques for development and incoporation in HEFS. If not, and post-processing applies just to one temporal portion of the ensemble output, those outputs must be sequenced into adjoining periods with a smoothing capability for edge effects. That is, breaks in flow continuity that appear unnatural, and are artifacts of post-processing application, should be avoidable.

3.4.3 Prototype Readiness Levels

As documented in Seo et al (2006), the EnsPost approach is configurable via straightforward data analysis and optimization of a single parameter per site (i.e. relatively parsimonious). Seo et al. (2006) reported modest reductions in forecast RMSE during the first five days of up to about 10 percent, and maintenance of reliable and unbiased forecast distributions. Thus the algorithm appears to be moderately effective for short range forecasting, at least in the portions of a river basin that are minimally affected by impairments to flow (regulation, diversion, etc.). Note that significant portions of all RFC forecasting domains are affected by flow regulation or impairment, representing a hurdle for implementation of any form of objective post-processing. In addition, EnsPost has not been widely tested in an operational RFC setting. Based on these results and interviews of the HEP group, **the applied science readiness level is MEDIUM.**

EnsPost can be also executed in the CHPS environment through an EnsPost model adapter. The EnsPost code is likely not maintainable indefinitely in this configuration, as it has problematic dependencies on RFS/OFS. But it can be reengineered to be more consistent with the FEWS/CHPS language conventions. Based on these observations, **the software readiness level is MEDIUM.**

3.5 Ensemble Verification System (EVS)

The EVS is designed for verifying ensemble forecasts of hydrologic and hydrometeorological variables, such as temperature, precipitation, streamflow and stage, issued at discrete locations (points or areas) for any lead time. It is capable of producing a comprehensive suite of verification metrics for ensemble forecasts.

3.5.1 CONOPS SUMMARY

Real- time operations

Verification is not a central element of routine RFC operations – meaning that most verification analyses related to long series of past forecasts will not generally be performed during operational forecast production. Verification results are relevant to forecast production, however, and it may be desirable for some verification graphics or statistics from EVS to be available or accessible to the forecaster during operations. For example, a forecaster facing an extreme event may wish to see statistics or a scatter plot describing the performance of the forecast system for events (perhaps including analogues) of the current magnitude and seasonality. If analysis times permit and supporting hindcast data were efficiently staged, it's conceivable that an interface in IFD to perform this type of analysis dynamically using EVS functionality would be desirable. Such a capability may already exist via the CHPS-EVS Model Adapter. Alternatively, such analysis could be pre-generated and stages for simple display within IFD. In this latter usage, EVS would be utilized in non-operational staff time to support operations, but not within the operational process.

Because EVS can calculate a wide range of metrics, it may also be adaptable for application within a forecast workflow to generate statistics or graphical analyses to provide performance feedback for the real-time forecast process. For example, plots or statistics describing running metrics of forecasts for a trailing 30 day period might be of interest. Such a usage would presumably be driven by batch process as an automated part of the workflow, and generate data that could be ingested and displayed in CHPS IFD.

Component Configuration

EVS is itself a configuration tool, in the sense that it can generate its own batch file of settings that control the verification metrics and graphics that will be generated. RFCs may configure EVS through interactive exploration of its capabilities followed by exporting a batch file (or part of one that will serve as a template for expansion to the remainder of their segments). Alternatively, general templates providing a subset of verification metrics may be developed for the RFCs, together with scripts to tailor them to the particular RFCs model and forecast configuration. EVS may be run in the CHPS environment as part of a workflow using its CHPS-EVS Model Adapter or outside CHPS, accessing input forecast and observed flat files that are stored outside the CHPS local database.

Hindcasting

In general, hindcasts from the other HEFS components are required by the EVS, although archived operational forecasts can also be verified, in principle. EVS itself is not used during hindcasting, but operates on the results of hindcasts. Hindcasts of other HEFS components are therefore a pre-condition to the use of EVS within HEFS. Verification analyses are applied to an adequate sample of past forecasts that are considered methodologically consistent with real-time forecasts (from which it follows that techniques such as DA and EnsPost and MEFP must be applicable in hindcast mode, if used in operations). The assessment of the quality of ensemble forecasts, especially for extreme events, requires the use of multiple years of forecasts. Conditional verification should also be performed to describe the ensemble performance from low to large events and could include additional criteria of conditioning (e.g., flow ensembles corresponding to temperature ensemble mean below freezing level).

3.5.2 REQUIREMENT SUMMARY

HEFS version 1

EVS in its current form provides nearly all of the operational/technical capabilities that are required within the first version of HEFS. Others that may need further attention include:

- Configuration support – that is, the design and provision of a **default** set of verification analyses that can be applied in batch to a set of hindcasts. This may involve also specifying the required hindcast experiments and framework, which should be designed in collaborating with RFC personnel. Scripts to tailor the default batch files to RFC model configuration will also be helpful.
- The development of workflow templates to provide real-time feedback verifications for forecasters (including display capabilities)
- The development of comprehensive training and tutorial materials for forecasters and possibly external partners, including examples of forecast verification applications.
- Automated expert advice to forecasters to help guide against improper generation of statistics, e.g., for samples with insufficient sample size, and/or to suggest interpretation of results.

Future Requirements / Recognized Needs

EVS may be expanded to focus on statistics or combinations of statistics that support real-time feedback and control of the forecast systems – for example, monitoring for deviations in system performance from long term norms. EVS may also expand to include greater capability for analogue identification and analysis (though an initial capability for conditioning on multiple variables – e.g., precipitation, temperature, streamflow, and stage - and multiple statistics – mean, total, minimum, maximum for different temporal aggregation windows - is already available). Many analog selection techniques have been demonstrated in the literature and utilized in operational centers. Initial efforts with the NWS Hydrologic Forecast Verification Team have helped determined criteria for analog selection, potential analog displays, and current CHPS capability for such displays. However such effort will require a robust archive database with auxiliary information to support analog selection via an appropriate query tool. There are also several scientific challenges for verifying ensemble forecasts of hydrometeorological and hydrologic variables, particularly within an operational context. It is envisaged that new verification methods and metrics will be developed and implemented in the EVS over time. For example, recent work within the HEP group evaluated the timing or 'phase' errors in hydrologic forecasts separately from the errors in amplitude. There are also major scientific challenges for verifying extreme hydrologic events and for other applications where sample data are limited (e.g. monthly volumes for water supply forecasting).

Depending on the level of integration of EVS with CHPS, it may be convenient at some point to source the calculation of various statistics required in applying other components, e.g., GraphGen, EnsPost, DA and MEFP, to EVS subroutines or functions. EVS would essentially become a statistical engine with an API that supports the other components, as well as its own verification objectives. Furthermore, it may be desirable to allow the EVS verification results to be processed by GraphGen, allowing custom plotting tailored to operational forecasting needs. Currently, the graphics produced by the EVS are somewhat limited in their ability to be configured. But the numerical outputs associated with these graphics are also written by EVS in XML format and could potentially be read by GraphGen.

3.5.3 PROTOTYPE READINESS LEVELS

EVS has received an extensive level of testing in various operational modes relevant to the envisioned functionality for CHPS, including as a standalone interactive, an interactive session within the CHPS environment via a CHPS-EVS model adapter, or a batch process being called from the system or from a CHPS workflow. The metrics calculated by EVS are unlikely to contain algorithm related deficiencies, and are sufficient to verify the forecasts anticipated for HEFS. The EVS code is written in Java and is relatively consistent with code conventions/philosophy adopted in CHPS/FEWS. Based on these observations, **both the applied science and software readiness levels are HIGH.**

3.6 Graphics Generator (GraphGen)

GraphGen is designed to construct graphical and associated tabular products and metadata based on observations and deterministic and ensemble forecasts contained within the CHPS operational database. The flexibility of GraphGen expands the CHPS and FEWS display capabilities beyond current ESPADP functionality, and should allow forecasters to configure ensemble and deterministic forecast displays as needed to satisfy the goals of the ensemble forecasting endeavor.

3.6.1 CONOPS SUMMARY

Real- time operations

The primary task of GraphGen will be to generate graphical products, ASCII tabular products and XML products. A graphical product will consist of one or more chart series calculated from user-selected input time series or ensembles that have been aggregated according to user specification. In real-time operations, the use of GraphGen will mostly be as an automated part of a forecast workflow, although forecasters may choose to detail plots further, possibly on a temporary basis, as shift time permits. The forecaster may choose to modify the appearance of the chart series, as well as the plot title, axis labels, axis limits and tick spacing, thresholds, and many other features of the chart. Finally, the parameters to generate associated graphics products can be specified. Products can be generic, capable of being applied to any segment, or segment-specific. GraphGen will support building templates that can be combined to form a single product. For example, a template in the ensemble forecast context might be for a chart that displays the observed time series, the deterministic forecast time series, and the ensemble-based uncertainty bounds. In addition, template products can be modified as needed and applied to multiple (or all) segments.

Four CHPS plug-in components comprising the GraphGen user interface will facilitate building and viewing products:

1. a panel displaying a selectable list of all products for all segments;
2. a panel displaying thumbnails for all products associated with the currently active forecast segment in the CHPS interface, allowing for the forecaster to view selected products during the standard forecasting process;
3. a panel providing the ability to design and build a product;
4. a panel displaying a single specific product.

Component Configuration

GraphGen plot configuration can take place within CHPS, or by modification of CHPS configuration templates. Default CHPS configuration templates for common products of GraphGen will therefore be desirable, along with system-level scripts to tailor them to individual RFC model implementations.

Hindcasting

GraphGen may be utilized within a CHPS-based hindcasting activity if it generates individual plots or data series for particular hindcasts within the hindcast series. It could also be applied for generating summary plots and extracted data sets that illustrate the overall performance of sets of hindcasts, a functionality that somewhat overlaps with EVS. Depending on the desired level of integration between the EVS and CHPS, the numerical outputs from the EVS could be read and processed by GraphGen. Given the greater level of configurability of GraphGen graphics versus EVS graphics, this enhancement may be desirable in the short-to-medium term. At present, whether a graphic is sought from EVS or GraphGen will depend on the plot features available and the difficulty of specifying them. Forecasters will adapt template configurations for these graphics to their model implementations, or develop them interactively and (hopefully) share them in a common repository for other RFCs and partners to use.

3.6.2 REQUIREMENT SUMMARY

HEFS version 1

The GraphGen component must:

- Provide interactive dialogues which enable the user to build and display deterministic and ensemble products.
- Replicate, at a minimum, all the current functionality of ESPADP.
- Have the capability to accept input data from other HEFS components.
- Present data against local or UTC time coordinates

User Interface Requirements:

- The system shall allow the forecaster to interface with the data by two methods – an interactive graphical user interface and a workflow module configuration (non-interactive).
- The user interface shall have a graphical presentation that allows the user to specify input conditions (i.e., product settings)
- The user interface shall have the capability of displaying graphical output
- The system output display/charts shall have the look and feel of the charts displayed in CHPS
- The system shall allow the user to interface with the data non-interactively via a workflow module configuration.
- Previously generated chart parameters/settings shall be stored (locally) in the CHPS database and shall be accessible from the CHPS database (specified as “local data settings storage” within the requirements sub-sections)
- The non-interactive workflows shall create a graphical output file, or a display within CHPS, or both
- The system shall have the capability of creating an output file as an image (e.g. .PNG), as ASCII text in a table format, or any widely used product format that is expected to continue usage in the future.
- The GraphGen shall have the capability to generate deterministic forecast products

In addition, sufficient training and documentation must exist that the design and specification of plots is straightforward for forecasters. A GraphGen user manual shall be provided with command references, 'knobology', examples of common graphics and templates, and instructions for constructing and changing the graphics and templates.

Future Requirements / Recognized Needs

Enhancements and software bugs have been identified by RFC users of the GraphGen software. As testing continues, additional user enhancements will likely be identified. To meet these developing needs GraphGen will follow a plug-in architecture allowing for easy expansion of functionality. Initially, the GraphGen will be able to gather time series from the CHPS database and Pi-timeseries compliant (schema defined by Deltares) XML files. However, the sources from which GraphGen gathers input time series will be easily extensible by adding input series plug-ins. Furthermore, the calculations performed by the GraphGen will be easily extensible, having a plug-in architecture for both aggregation computations and chart series calculations. A plug-in architecture will also be employed to allow for quick extension of the appearance modification capabilities (e.g., chart labels, axis limits, etc) and adding further types of outputs to be generated.

An API will be needed for GraphGen to enable it to interoperate with CHPS and maintain consistency with FEWS through upgrade cycles.

And effort to coordinate GraphGen functionality with the EVS software to ingest and plot or analyze verification results to go along with forecast products. Work should continue with all RFCs to identify the most meaningful forecast and verification products for forecasters and diverse end users (internal and external), perhaps including analogue-oriented products.

3.6.3 Prototype Readiness Levels

The GraphGen consists of a graphical user interface that is seamlessly integrated within the CHPS interface, as well as a model adapter that allows for products to be generated within CHPS workflows. It possesses the same look-and-feel as existing components of the CHPS interface. It will be easily extensible from its present form, allowing for quick additions of new source of input time series, new types of aggregations, new types of calculations, new ways in which the appearance of displayed charts can be modified, and new types of output products. GraphGen has been relatively well-tested within OHD as development has proceeded, but is not yet widely implemented in the RFCs. Nonetheless, **the applied science readiness is considered HIGH**, especially since little science is involved in its operation. The **software readiness also appears to be HIGH**, given the CHPS compatibility of the prototypes.

3.7 Relationship of HEFS Requirements to New York City Department of Environmental Protection (NYCDEP) Project Requirements

NYCDEP has awarded a contract to the engineering firm Hazen and Sawyer to develop the Operations Support Tool (OST). The OST is a state-of-the-art decision support system for the NYC water supply system. The OST will integrate multiple sources of critical near real-time operations data – streamflow data, in-reservoir water quality data, etc., - into an advanced version of NYCDEP's existing OASIS-W2 water supply-water quality model. The OST will combine this current data with inflow forecast data and system operating rules to project the likely range of reservoir levels and water quality over the coming weeks and months.

Based on interviews with NYCDEP and Hazen and Sawyer representatives, the requirements for the NYCDEP project strongly align with NWS requirements for HEFS. These are summarized in the remainder of this section. There is, however, one stated desire that cannot be met by the HEFS package: that is, a complete consistency between hindcast methodology and real-time forecast methodology. Inconsistencies between real-time and retrospective forecasting arise from many sources (e.g., meteorological analyses, MODs, reservoir operations, river gage errors, data discovery during operations, etc.). NWS River forecasting operations will require significant cultural, procedural, scientific and technological changes to fully eliminate these inconsistencies.

Aside from this issue, **primary** requirements of the NYCDEP project include the following:

- Daily time-steps streamflow forecasts, updated daily with lead times of two weeks.
- Forecasts are for unregulated inflows to projects and side/intervening flows downstream of reservoirs
- Best possible forecast NWS can provide, with no restrictions on what NWS decides is best (e.g., using GFS or other forecasts)
- Non-demanding latency requirements (could receive forecasts well after typical release times from RFCs, e.g. mid-day
- At least 30 ensemble members
- Strong but possibly not complete consistency in methods between real-time forecasts and retrospective hindcasts. The number of years were not specified, but OHD indicates interest in hindcasts of ~30 years. NYCDEP could accept mods in real-time practice, based on John C. Schaake's assertion that MODs tend to improve forecast quality.
- A number of forecast sites on the order of about 3 dozen, but this estimate may have grown since the discussion.

Strongly desired extensions of these requirements include:

- Forecast lead times out to one year (or at least to June 1 of each year)
- During flood events, more frequently updating forecasts with up to 3-hour timesteps.

Less strongly desired extensions to the requirements include:

- At least 50 ensemble members per forecast
- *Exact* consistency between real-time and retrospective forecasts

- A set of other forecast variables corresponding to streamflow ensembles to apply for water quality and storm timing and intensity analyses: precipitation, temperature, wind speed, humidity, solar radiation and possibly others (to be verified through further discussion with NYCDEP)

3.8 Product Dissemination / Data Services

A significant effort is required to develop effective approaches for presentation of ensemble forecasts and dissemination of data/ products to NWS' internal partners (e.g., WFOs) and external users. At least for the initial implementation of HEFS (version 1), the product dissemination will likely utilize existing methods and formats to a great degree. **These include AHPS web pages, RFC web pages, AWIPS WAN, etc.** Although there could be enhancement to the existing dissemination methods, a more consistent and integrated data service for these ensemble products is strongly desired.

REFERENCES AND RELEVANT PUBLICATIONS BY TOPIC

General:

NRC (National Research Council), 2006: Completing the Forecast: Characterizing and Communicating Uncertainty for Better Decisions Using Weather and Climate Forecasts, Committee on Estimating and Communicating Uncertainty in Weather and Climate Forecasts, National Research Council ISBN: 978-0-309-10255-1, 124 pages.

Meteorological Ensemble Forecast Processing:

Schaake, J., Demargne, J., Hartman, R., Mullusky, M., Welles, E., Wu, L., Herr, H., Fan, X., and Seo, D.-J. 2007a: Precipitation and temperature ensemble forecasts from single-value forecasts. Hydrology and Earth Systems Sciences Discussions, 4, 655-717.

Wu, L., Seo, D.-J., Demargne, J., Brown, J., Cong, S., and Schaake, J., 2011. Generation of ensemble precipitation forecast from single-valued quantitative precipitation forecast for hydrologic ensemble prediction via meta-Gaussian distribution models. Journal of Hydrology, 399, 281-298.

Hydrologic Forecast Post-Processing:

Brown, J.D. and Seo, D-J., 2010. A nonparametric post-processor for bias correcting ensemble forecasts of hydrometeorological and hydrologic variables. *Journal of Hydrometeorology*. 11(3), 642-665.

Clarke M, Gangopadhyay S, Hay L, Rajagopalan B, Wilby R. 2004. The Schaake Shuffle: a method for reconstructing space-time variability in forecasted precipitation and temperature fields. *Journal of Hydrometeorology* 5: 243–262.

Regonda, S., D-J., Seo and B. Lawrence: Short-term Ensemble Streamflow Forecasting Using Operationally Produced Single-valued Streamflow Forecasts – A Hydrologic Model

Output Statistics (HMOS) Approach, Journal of Hydrology (in preparation).

Seo, D.-J., Herr, H., and Schaake, J. (2006): A statistical post-processor for accounting of hydrologic uncertainty in short-range ensemble streamflow prediction. *Hydrology and Earth Systems Sciences Discussions*Hydrol. Earth Syst. Sci. Discuss., 3, pp 1987-2035.

Wood, Andrew W. and J.C. Schaake, 2008: Correcting Errors in Streamflow Forecast Ensemble Mean and Spread. J. Hydrometeor, 9, 132–148. doi: 10.1175/2007JHM862.1

Data Assimilation:

Lee H., Seo D.-J. and Koren V., 2010. Assimilation of streamflow and in-situ soil moisture data into operational distributed hydrologic models: Effects of uncertainties in the data and initial model soil moisture states. *Advances in Water Resources* (under review).

Seo, D. J., V. Koren, and N. Cajina (2003), Real-time variational assimilation of 18 hydrologic and hydrometeorological data into operational hydrologic forecasting, *Journal of Hydrometeorology*J. Hydrometeorol., 4(3), 627-641.

Seo, D.-J., L. Cajina, R. Corby and T. Howieson, 2009. Automatic State Updating for Operational Streamflow Forecasting via Variational Data Assimilation, 367, *Journal of Hydrology*, 367, 255-275.

Verification:

Brown, J.D., Demargne, J., Seo, D-J. and Liu, Y., 2010. The Ensemble Verification System (EVS): a software tool for verifying ensemble forecasts of hydrometeorological and hydrologic variables at discrete locations. Environmental Modelling and Software. , 25(7), 854-872.

Demargne, J., Brown, J.D., Liu, Y., Seo, D-J., Wu, l., Toth, Z. and Zhu, Y., 2010. Diagnostic verification of hydrometeorological and hydrologic ensembles. Atmospheric Science Letters, 11 (2), 114-122.

Demargne J, Wu L, Seo DJ, Schaake J. 2007. Experimental hydrometeorological and hydrologic ensemble forecasts and their verification in the U.S. National Weather Service. Quantification and Reduction of Predictive Uncertainty for Sustainable Water Resources Management (Proceedings of Symposium HS2004 at IUGG2007, Perugia, July 2007). IAHS Publication 313: 177-187.

APPENDIX

Response to Major Review Comments

A number of conflicting perspectives arose in the course of the report preparation and review. Although the report does not adopt these perspectives, they are recorded below with a brief response from the HEFS Requirements Team.

Comment: Hydrologic ensemble post-processing should be a **required** component, not just rated as "high priority" in Table 1. Most forecast users will need calibrated flow ensembles since they will not develop their own post-processing technique. Some of the hydrologic biases could be removed in real-time by the forecasters by applying MODs. However, the flow ensemble reforecasts cannot include any MODs and therefore the hydro biases can only be removed with the hydro post-processor.

Response: We felt strongly that, in order to provide input for decisions to efficiently allocate limited resources over a constrained development schedule, we needed to provide as much differentiation as possible between even highly desired components. The "Component Version 1 Priority" in Table 1 provides our assessed priority for each potential component as it relates to the most essential needs for an initial operational implementation of HEFS. In addition to the prioritized need for a specific capability, it also reflects the realities of the current prototype (e.g. state of development/testing, software engineering maturity, ability of individual component to support service-level objectives) and the relative likelihood that specific components can be ready for operational implementation within the planned schedule.

The "high priority" assessment (versus "required") for ensemble post-processing is based on the following considerations:

- NYCDEP does not require the NWS post-processed flow ensembles because they are developing their own post-processing technique for their own water supply decision support system. Other users may also choose this path, or, as with MMEFS, interpret them qualitatively in initial stages of use.
- The benefits of recent enhancements to EnsPost/HMOS hydro post-processors are being demonstrated only recently (e.g., HEPEX workshop, which Report Team members could not attend). Despite these activities within OHD, the current level of maturity and degree of testing of the enhancements is low.
- The report recognizes the importance and priority of ensemble post-processing, and strongly recommends that parallel R&D efforts continue in support of this capability. If development and testing of post-processing techniques prove successful and timely, we strongly advocate including a post-processing component in HEFS v1 if the development schedule permits. If not, a more broadly considered strategy for post-processing, developed in time for HEFS version 2, may be advisable.
- Real-time mods and particularly ADJUSTQ blends can help compensate for a lack of post-processing capability, though this baseline in comparison to post-processing has never been evaluated or recognized. This is the only strategy currently available, in fact, for many

regulated forecast locations.

- Unlike required/essential capabilities (e.g., MEFP, hydrologic processing, product generation, etc.), it is clearly possible to provide a useful ensemble forecasting service without a streamflow forecast calibration capability. RFCs, and most other forecasting enterprises in the world, have long produced deterministic forecasts (and more recently ensemble forecasts) lacking post processing in which users still find benefit. Uncalibrated ensemble forecasts are not ideal, but are arguably not an element without which the first version of HEFS cannot exist in usable form. As noted above, the priority levels are intended to guide allocation of development effort to achieve a working HEFS version 1 within a fixed time period. *After* required elements are secured, hydrologic ensemble post-processing (and other non-required elements) are highly desired inclusions.

Comment: Why is EPP3 applied science rating medium?

Response: EPP3 has been evaluated more thoroughly than most other science components of described in the report. Nonetheless, there are many situations and aspects for which it has not been evaluated – for example, the aggregate performance over multiple segments within a river basin has not been tested, nor is their a complete understanding of the effects of various EPP3 configuration permutations.

Comment: Why can't there be a full consistency between hindcast methodology and real-time forecast methodology?

Response: Inconsistencies between real-time and retrospective forecasting arise from many sources, a partial list of which includes differences in retrospective versus real-time meteorological analyses, routine MODs, reservoir operations, river gage errors, data discovery during operations (e.g., real-time specification of unknown diversions or consumptive uses to achieve a river balance). Any one of these elements alone creates a signficant inconsistency, and together they remove any real hope of exact consistency. River forecasting operations at NWS will require significant cultural, procedural, scientific and technological changes to fully eliminate these inconsistencies.

Comment: Why is MMEFS ability to meet HEFS objectives "low"?

Response: MMEFS as currently implemented cannot support hindcasting (unless limited to GFS and CFS reforecast sources) or calibration. Note, MMEFS's use of full ensemble information from a range of models, rather than generated ensembles (from ensemble means, as in EPP3), is a strength, and a primary reason that MMEFS has been supported as a potential component.

Comment: The automatic data assimilation capability is rated as "high" (not required): How are you going to do ensemble forecasting without automatic data assimilation (ADA)?

Response: The capability priority represents science readiness level as well as potential benefit from techniques developable in the near future. There are no operational ADA pilots as of now in the RFCs and most other forecasting agencies (though some do exist), which indicates the degree to which a gap exists between research-level applications and operational realities. Though data assimilation science is advanced and offers real opportunities for model state error reduction, many DA approaches do not address the need to update in an integrated fashion a physically coherent vector of models states (e.g, soil moistures and snow variables across multiple segments or elevation zones), and therefore cannot comprehensively replace MODs that are used routinely in RFC operations. For most forecast locations in the RFCs, it is not practical or possible to implement existing automated data assimilation techniques, due to imperfect data collection platforms and manual data specification practices. The lack of ADA will continue to compromise our ability to do consistent ensemble hindcasting except in specialized locations for which ADA data requirements can be satisfied. Rather than wait for automated data assimilation to meet all needs, ensemble prediction must advance imperfectly while the science and applicability of ADA advances, and we strive to incorporate as many partial solutions ADA can support as possible .

Overview of Testing of Prototype Components

The individual XEFS components have been tested to varying degrees at OHD and at individual RFCs. Further details of the testing and verification efforts associated with each component are discussed in greater detail in Section 3. **Table A.1** below contains a high level compilation of some basic attributes of the component testing to date.

Note: The table reflects testing data collected from interviews with HEP group personnel and a survey of RFCs.

Table A.1 Testing of HEFS prototype components by OHD and the RFCs

(testing categories: **LIMITED**: for a small fraction of forecast points for a short period of time or just a few events; **MEDIUM**: for a representative range of points over a moderate range of conditions; **EXTENSIVE**: for a large fraction of points and a wide range of conditions or long period)

Prototype Component	**Tested as individual component** (LIMited, MEDium, EXTensive)	**Number of forecast points tested per RFC (respectively)**	**Tested w/ other HEFS components** (yes, no)	**Multi-year hindcasting verification** (number of years)	**Assessment for specific events** (LIMited, MEDium, EXTensive)
EPP3 and/or EPP2	OHD: MED (MA, AB, CN, NW) RFC: CB LIM	OHD: 10, 9, 5+, 4 RFC: CB 219	OHD: ESP, ESP-EnsPost RFC: CB no	OHD: 27 years (GFS); 8 years (RFC) RFC: CB no	OHD: no RFC: CB LIM
MMEFS	OHD: LIM (w/ SREF; AB, MA, CN, NW) RFC:	OHD: 16, 10, 12, 11 RFC:	OHD: no RFC:	OHD: 4.5 years RFC:	OHD: no RFC:
EnsPost	OHD: MED (CN, CB, MA) RFC:	OHD: 8, 13, 5; 5 w/ EPP-ESP RFC:	OHD: EPP-ESP RFC:	OHD: 27 years (w/ GFS) RFC:	OHD: no RFC:
Multi-scale CDF matching	OHD: LIM (CB, CN, MA, WG) RFC:	OHD: 12, 6, 6, 14 RFC:	OHD: no RFC:	OHD: RFC:	OHD: no RFC:
HMOS	OHD: MED (AB, CN, MOPEX basins) RFC:	OHD: 6, 5; 12 (MOPEX) RFC:	OHD: no RFC:	OHD: ~10 years; 36 years (MOPEX) RFC:	OHD: LIM RFC:
1DVAR	OHD: LIM (WG) RFC:	OHD: 4 RFC:	OHD: no RFC:	OHD: 10 years RFC:	OHD: LIM RFC:

2DVAR	OHD: LIM (AB, WG) RFC:	OHD: 3, 23 RFC:	OHD: no RFC:	OHD: 10 years RFC:	OHD: LIM RFC:
4DVAR	OHD: LIM (AB, WG) RFC:	OHD: 4, 5 RFC:	OHD: no RFC:	OHD: 3-13 yrs RFC:	OHD: LIM RFC:
2DMLEF	OHD: LIM (WG) RFC:	OHD: 3 RFC:	OHD: no RFC:	OHD: 5 yrs RFC:	OHD: LIM RFC:
EnKF	OHD: LIM (NW) RFC: no	OHD: 1 RFC:	OHD: no RFC:	OHD: 20 years RFC:	OHD: LIM RFC:
EVS	OHD: MED (5 RFCs) RFC: CB LIM	OHD: ~15 points/RFC RFC: CB 8	OHD: yes RFC: CB no	OHD: 27 years RFC: CB no	OHD: LIM RFC: CB LIM
GraphGen	OHD: RFC:	OHD: RFC:	OHD: RFC:	OHD: RFC:	OHD: RFC:

Note: The EPP2 and EPP3 ensemble forecasts have been evaluated for the following conditions: seasonal verification, verification for different thresholds, verification at 6-h and 24-h time steps, and verification with spatial aggregation (up to 10 segments). The evaluation was done using 1) the HPC/RFC single-valued forecasts up to 5 lead days into the future (Schaake et al, 2007; Wu et al, 2011) and 2) the ensemble mean of the GFS forecasts up to 14 lead days into the future. The EPP-ESP and EPP-ESP-EnsPost streamflow ensembles have been verified at 24-hr time steps (as mean daily flows) up to 14 lead days into the future and for different thresholds (Demargne et al, 2007; Demargne et al, 2010).

Note: The SREF precipitation ensembles, which have been archived between April 2006 and August 2010 using all available cycles, have been verified for 3 RFCs (AB-, MA-, NW-, and CN-RFCs) using 10 to 20 basins per RFC (Brown et al., 2011). Conditional verification results are given by forecast lead time, amount of precipitation, season, forecast valid time, and accumulation period. OHD started to look into the potential benefits of merging the SREF ensembles with the GEFS ensembles for precipitation using a Bayesian Model Averaging approach. No verification has been done at OHD for temperature and streamflow.

Note: The HMOS processor, the Indicator Co-Kriging non-parametric processor, and the General Linear Model – Post-Processor are currently being tested as part of the HEPEX hydrologic post-processing and verification testbed. Initial results will be presented at the HEPEX workshop in June 7-9, 2011 in the Netherlands. The testbed datasets include: 1) flow simulations for 12 MOPEX basins with 8 different hydrologic models; 2) flow ensemble reforecasts for 2 CNRFC test basins based on EPP3-generated GFS-based and climatology-based ensembles.

Note: The verification has been performed with the Ensemble Verification Program (up to ~2008) and the Ensemble Verification System and included various metrics for the ensemble mean forecasts and the ensemble forecasts to describe the different forecast attributes (e.g., correlation and bias for the ensemble means, skill compared to reference forecast, reliability, resolution, and event discrimination).

Note: 1DVAR evaluation was carried out for four river reaches in Texas. Currently, 1DVAR application to ABRFC river reaches is underway. 2DVAR evaluation for 3 basins in the ABRFC region was carried out for significant events only (Seo et al., 2003). 2DVAR has also been evaluated for 23 basins in the WGRFC region through the experimental operation at the WGRFC since 2003 (Seo et al., 2009). The performance of 2DMLEF on streamflow prediction for 23 WGRFC basins has been under evaluation. 4DVAR evaluation was performed with the hydrologic and hydrometeorologic dataset from four basins in the ABRFC and five basins in the WGRFC region. Evaluation of 1DVAR, 2DVAR, 2DMLEF, and 4DVAR were performed on an hourly time step and for streamflow prediction up to 72 lead hours.

RFC Notes:

CBRFC – has run EPP3 outside of CHPS on an automated daily basis for five months for the upper Colorado R. basin, using both CFS and GFS to drive flow forecasts, and is making a qualitative and informal assessment of the performance but has not run EPP3 hindcasts. CBRFC has experimented with EVS for peak flow forecast assessment for approximately 8 points.

CNRFC – <details>

NWRFC – <details>

APRFC – <details>

MBRFC – <details>

NCRFC – <details>

NERFC – <details>

MARFC – <details>

OHRFC – <details>

SERFC – <details>

WGRFC – <details>

LMRFC – <details>

ABRFC – <details>