The Case for

EVALUATING HYDROANALYTIC INFORMATION

Sandra Knight, PhD, PE, D.WRE, D.NE
President, WaterWonks LLC

June 2021

COASTAL RESILIENCE CENTER
A U.S. Department of Homeland Security Center of Excellence

waterwonks llc

Contract Information

This informational report is based upon work supported by two agencies in two phases, the U.S. Army Engineer Research and Development Center (ERDC), Coastal and Hydraulics Laboratory (CHL) and the U.S. Department of Homeland Security. The ERDC support was funded under an existing contract with Information International Associates, Inc. and sub-contracted to WaterWonks LLC from August 2018 through February 2019. The U.S. Department of Homeland Security's Coastal Resilience Center of Excellence supported the second phase during the period from April 2019 through June 2020 through a sub-award to WaterWonks LLC using funds provided by the DHS Flood Apex Program.

CONTENT

ACKNOWLEDGEMENTS

Many individuals influenced the content of this study by graciously contributing their valuable time to share their expertise and insights with me. While the views and conclusions of the report are mine, the rich discussions and valuable resources provided by more than 70 contributors greatly influenced the outcomes. Contributing their time, expertise, feedback and references at various points in the process were the engineers, scientists and mathematicians, both researchers and practitioners, from U.S. and international agencies, private industry, laboratories and academia. These included the U.S. Army Corps of Engineers-Institute for Water Resources, the US Army Engineer Research and Development Center, the Rock Island District; the Federal Emergency Management Agency, the National Weather Service, AECOM, Atkins, M2Strategy, IIA Inc., the Iowa Institute for Hydraulic Research, the University of Georgia, the University of Maryland, the Centre for Ecology and Hydrology, UK, HR Wallingford, UK, Environment Agency, UK, Flood Forecast Center, UK and Deltares, Netherlands.

My sincere appreciation goes to the DHS S&T Coastal Resilience Center of Excellence at the University of North Carolina, Chapel Hill. The Center, and especially Tom Richardson, have supported my participation and helped to improve the content and quality of this report at each step along the way. Additionally, I am grateful to the members of the DHS S&T Flood Apex Research Review Board who provided invaluable discussions on several occasions on the content of my findings and helped shaped the recommended concepts for an evaluation framework.

A special thanks goes to ERDC Coastal and Hydraulics Laboratory, DHS S&T and FEMA and the support provided by Jeff Lillycrop (retired), Julie Rosati and Jane Smith, CHL, David Alexander, DHS S&T and Luis Rodriguez, FEMA. Their insights and eagerness to innovate and transform their programs was an inspiration for my research. Their support from the beginning to the end has made working on this project a rewarding experience.

FOREWORD

The initial purpose of this study was to explore the best strategies in evaluating numerical models and data particularly related to hydroanalytic information. Completed in two phases with two sponsors, I began to understand that while collecting best practices was interesting, it is not a satisfactory end point. The viability and credibility of the models, data, and applications needed to address the complex issues facing us in a changing climate will require a paradigm shift within our modeling community of practice. Identifying the technical requirements to assure models are valid and accurate for their intended use can be complex, but doable. Yet, as a community of practice, we have not converged on a standard set of technical processes or methodologies, much less institutionalized those methodologies. A more abstract aspect but essential element of institutionalizing best practices, fundamentally begins with leadership and an institutional culture that embraces collaboration, demands excellence, attracts talented and diverse people and values customer perspective.

As I share some of the best technical strategies and provide examples for evaluating hydroanalytic information, I also describe the importance of user engagement and institutional guidance. Models and data are sufficiently complex that the technical requirements and processes should be explicitly and consistently standardized. However, the technical aspects cannot be created in a vacuum. Acceptance or buy-in of modeling tools and data by the client or end-user requires their engagement early and often and a transparent framework for sharing. The institutional guidance - policies/governance/business processes/human resources- must establish clear and authoritative directives, be built on an organizational framework that streamlines functions and by organizations that embrace a culture of change and innovation.

In other words, to be credible developers and users of hydroanalytic models and tools, our community of practice should consider moving quickly to build a robust evaluation framework that will best serve our respective missions and the clients we support.

Sandra K. Knight, PhD, PE, D.WRE, D.NE
President, WaterWonks LLC

EXECUTIVE SUMMARY

INTRODUCTION

Revolutionary changes are occurring in data and modeling analytics. Hydroanalytics — defined broadly as any analytic data and models that involve water-related processes — will by necessity be transformed as basic applications (i.e. hydrometeorology, hydrology, hydraulics, sediment transport, coastal circulation, wave dynamics, groundwater) morph to more advanced applications that will require integration and inclusion of non-traditional data/models (i.e. remote sensing, social media, market data, ecology, behavioral sciences), a push to use leading-edge methodologies (i.e. artificial intelligence, deep learning, neural networks, coupled probabilistic and physics-based methods), the introduction of transformational technologies (i.e. blockchain, micro-processors, drones, sensor technologies, high- performance, cloud and quantum computing) and better representation of complex processes (i.e. large-scale coupling of interdisciplinary models, agent-based modeling, advances in scientific understanding of natural processes).

A robust evaluation framework is needed in the hydroanalytic community of practice to assure transferability, repeatability, reliability, usability and overall quality of data, models and information. The definition of evaluation in this report, includes any criterion, process or methodology that reviews and documents the quality and performance of software, data and models. Currently, there are dozens of models and data systems that inform critical decisions by federal, state and local governments, public sector businesses and individuals to plan for capital investments in ports, waterways, flood risk reduction and ecosystem restoration projects, improve operational activities of our infrastructure, recovery and mitigation of major hydro-climatic events, and preparation of individuals and communities to be more resilient.

To be credible developers and users of hydroanalytic models and tools, this study recommends that our community of practice should move quickly to build a robust evaluation framework that will best serve our respective missions and the clients we support. This study sought to identify best approaches and methodologies to build that framework to consistently validate and verify, quantify and/or reduce uncertainty and manage hydroanalytic information.

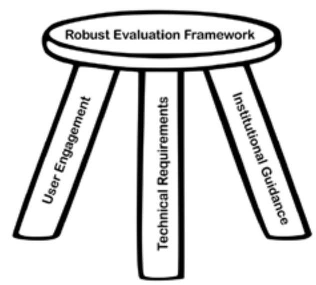
The report explains the urgency and importance to act now in creating and institutionalizing an evaluation framework. The report describes the basis for this on the desires by professionals and their organizations to:

- Reduce and communicate uncertainty
- Be relevant and/or best-of-class
- Address tough and complex problems
- Optimize and stretch diminishing resources
- Adapt to disruptors and adopt innovative science and technology
- Be transparent and collaborative.

To maintain credibility as world-class leaders in hydroanalytic modeling and solutions, organizations must be prepared to not only address the technical challenges associated with a changing planet, but also adapt their strategies to compete in a global marketplace. They must be committed to providing credible science and engineering solutions that withstand sociopolitical scrutiny while informing resilient solutions to grand challenges. To stay competitive in this rapidly evolving market with other national and international hydroanalytic laboratories, agencies and consultants, successful organizations will need to modernize their modeling and data portfolio by incorporating transformational science and technology, embracing collaborative and transdisciplinary approaches and measuring their progress through a robust evaluation framework.

LANDSCAPE REVIEW

The review captures a snapshot of an ever-evolving landscape of current practices in evaluating hydroanalytic information. While highlighting and extracting the best practices, the review uses examples to explain weaknesses or gaps. Further, best approaches to manage and standardize processes, evaluate quality and communicate uncertainties for evolving and future hydroanalytical information are synthesized into three foundational areas: user engagement, technical requirements and institutional guidance.



The best ideas and lessons learned on evaluation practices reviewed in these foundational areas were considered important to the long-term viability and sustainability of hydroanalytic tools. Acceptance or buy-in of modeling tools and data by the client or end-user requires their engagement early and often and a transparent framework for sharing. Technical requirements included software and data management, qualitative and quantitative criteria and overall quality of the information. Policies, governance, business processes and human resources are part of the building blocks needed to establish the guidance to institutionalize an evaluation framework.

User Engagement

There were five areas of user engagement explored in this study:

Know the Users
Models and software are often developed for one use or organization but expand to a broader set of users and applications. The developing agency and the user's credibility can both be tested when

their products are applied beyond the purpose for which they were originally developed without evaluating their fitness-of-use. Therefore, it is critical to understand the current and potential users and uses as modeling and data tools transfer from development to mainstream application.

Fitness of Use

Whether used for forecasting, forensics, planning, design or operations, the fitness-of-use of a model software to a specific application and the availability and quality of data should be evaluated prior to selecting the best tool for a problem. Importantly, that selection must support and be supported by the ultimate end-user.

Acceptance and Tech Transfer

Operational organizations that accept the "readiness" of a product must be prepared to defend its viability as they are financially, and often legally, responsible for its use. Developers and their users must engage to assure that product transitions are acceptable for operational use. There are many viable transition pathways for new products that can be driven by necessity but grounded in expert evaluation and review.

Building Capacity and Capability

Understanding the user also means understanding their capacity and capability to use the model outputs or to run the model. Capacity building improves the credibility of the giving and receiving organizations. Organizations reviewed in this study used numerous approaches including cooperative grants, software training and workshops, user groups and career fairs.

Credentialing

Professionals who both developed and used models expressed their concerns that the quality of modeling results, regardless of the quality of the model, were explicitly tied to the capability and proficiency of the modeler. There is a void and few examples of credentialing individuals and/or organizations that apply specifically to hydroanalytic tools. As data and software continue to move to open source and shared platforms, there is a growing need that the credibility of the model/data users be validated.

Technical Requirements

There are many good practices for technically evaluating and managing hydroanalytic information and other derivative products. The practices and methodologies reviewed included rigorous data management and standards, validating and verifying software and models, quantifying uncertainty and managing large portfolios of information. There are numerous gaps in standardization and some inconsistencies not only across the hydroanalytic community of practice, but within the agencies that develop and use hydroanalytic information and products. However, there are clearly some best practices that could provide a basis for more consistent standards and applications.

Four areas of technical requirements reviewed were:

Software and Data Management Practices

As new analytic techniques become more mainstream, organizations must commit to open source and easily accessible software and data. It is widely accepted among practitioners and is being mandated in federal practice by DoD, DHS and others where security is not compromised. Sharing codes and algorithms among advanced researchers opens the door to transparency and transdisciplinary approaches to solving the next-generation problems. While there are some pockets of data and modeling standards and management practices, they are not consistent among practitioners and/or organizations. The American Meteorological Society makes a case for a standard set of principles due to the growing complexity and increasing volume of observations and model data (Bulletin of the American Meteorological Society 2019).

Qualitative Criteria

In selecting and using models, a reliance on qualitative assessments over more quantitative assessments was a common practice. Common qualitative requirements included software and data accessibility, graphical user-interfaces and visualization, documentation and training and technical support. Hydroanalytic tools that had been peer-reviewed along with the experience of the modeler and expertise of the modeling tool were all considered acceptable qualitative criteria. Fitness-of-use (how well-suited the model was for its application) was important, but acceptable levels of accuracy were not often specified to assure the model could reproduce the processes of interest. Physics- or process-based methodologies were generally deemed better approaches more often than data-driven results because of the acceptance of the science behind the modeling approach.

Quantitative Criteria

Many good practices for using quantitative criteria and techniques, particularly when testing the validity of the software or the accuracy of the results, were found and reviewed. Benchmark testing, test beds, uncertainty quantification, statistical analysis and product scoring represent some of the more accepted practices. Benchmarking is an evidence-based process for helping compare capabilities among different models. It is an important tool for understanding existing models and assessing new ones. In general, however, there are no universal set of bench tests for various categories of hydroanalytic models. Further, test cases often do not represent the real complexity of nature nor its range of scale from large basins to near-field applications. Verification, validation and uncertainty quantification (VVUQ) takes benchmarking to a more complex level and is emerging in many fields as a desired methodology for rigorous evaluation particularly for sophisticated hydroanalytic models. VVUQ can be complimented by the use of test beds and other laboratory techniques. Common methods for quantifying uncertainty and evaluating quality that have been practiced in the meteorological community using statistics, skill scores and other forecast verification metrics are now becoming more mainstream in river and coastal forecasting.

Quality and Performance

Often a combination of qualitative and quantitative criteria, standards and expert elicitation are applied to evaluate the quality and use of individual models or a portfolio of modeling products. Identifying critical user and technical requirements (whether qualitative or quantitative) and establishing a process for rating or scoring the criteria results in a hybrid approach to evaluating hydroanalytic information. The review identified some excellent practices that could be used to assess and evaluate the technical quality of hydroanalytic information.

Institutional Guidance

Areas for improving institutional guidance for hydroanalytic evaluation include:

Policies and Standards

Good data standards and practices often are the result of good policies. As federal agencies and organizations improve and implement new data and modeling policies, they should strive to conform to the highest standards following accepted practices at international and national levels. To be useful and achieve the outcomes intended, the policies must be resourced and enforced.

Evaluation

Progressive first-class organizations embrace evaluation and employ evaluation practices such as annual reporting, external boards and special reviews to assure their organizations are keeping pace and meeting the highest level of expectations.

Governance

Organizational culture can be a leading indicator on whether an organization is world-class. Culture starts at the top with leadership, but is also driven by clear definitions of authorities, roles and responsibilities. The organization's structure can support these definitions and ease the path to success.

Business Processes

Implementation of good business process applies to hydroanalytic tools just like it does other assets. Considering and managing these assets influences how to best invest in and maintain them over their life cycle, helps in determining acquisition and technology adaptation strategies and considers how they should be marketed and resourced.

EMERGING HYDROANALYTICS

World-class problems and transformative technologies are causing revolutionary changes in data and model analytics. Integrating modeling systems, incorporating new analytic methods and the inclusion of non-traditional data will be necessary to address such areas as the consequences of disasters, the energy-water nexus, adaptation to climate change and managing aging and inadequate infrastructure. Quantifying non-traditional benefits and costs, cascading effects and compound events will add to the complexity. Hydroanalytic modeling approaches are expanding to include not just economic consequences but social behaviors using techniques such as gaming theory, agent-based modeling and multi-agent systems. Organizations around the world are turning to big data, advanced modeling approaches and hybrid or coupled analytics to model complex problems. The report highlights a few examples.

BUILDING BLOCKS TO IMPLEMENTATION

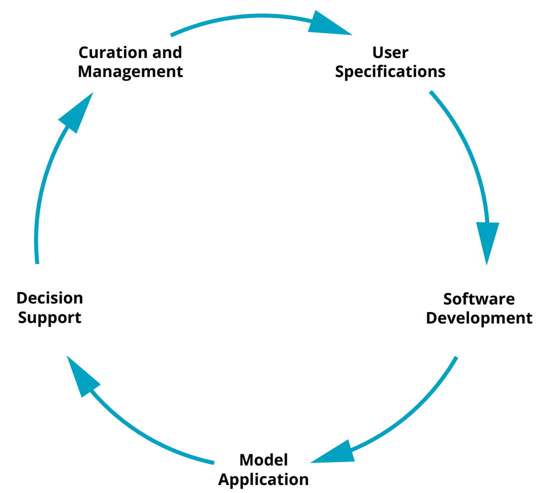
A synthesis of the methods and best practices is described as building blocks toward implementation. Each building block represents an essential consideration in the development and implementation of a robust hydroanalytic evaluation strategy. The building blocks include:

Synthesis of Best Practices

This section of the report highlights some of the best practices reviewed in each foundational area -user engagement, technical requirements, institutional guidance- described above.

The Model Evaluation Cycle

A modeling cycle is developed to help distinguish the steps from conception to development to use and curation. Within these steps, the methodologies for evaluation may be distinct, but equally critical is assigning who will execute or implement these practices. Assigning someone the responsibility and giving them the authority and resources to execute the evaluation will help assure a quality product.



Product and Portfolio Assessment

Large organizations and agencies often are responsible for a portfolio of products that develop over time and are used for multiple purposes. As part of a portfolio asset management strategy, a portfolio assessment concept is proposed that includes qualitative and quantitative data to assess the inventory of products and provides a robust way to track current conditions and identify areas of improvement. Further, in the concept presented in this study, not only are technical criteria proposed to assess the quality of hydroanalytic information, but also criteria that reflect community vulnerability and future climate and economic information. As demonstrated by the methodology developed in the U.K., and described in the report, of assigning confidence scores by rolling up the assessment data into indices, a scorecard of existing mapping and assessment products could help assess the confidence and quality in local to national products, identify areas of needed investment in hydroanalytic information, inform investment strategies and communicate requirements to the public.

Certification of Modelers and Models

To further improve quality and credibility of modelers, some options are presented to provide credentialing to individuals and/or organizations that apply hydroanalytic tools. These options could involve existing professional organizations and/or institutes or laboratories who could administer the training and evaluate the experience requirements for a certain class of models such as hydrologic, coastal circulation or hydrodynamic models.

Certification of models or analytic methods is more complicated. While it seems logical that software should meet some standards to be considered certified for hydroanalytic applications, the reality of doing that is complicated by the various requirements of end use and the lack of standards. In the absence of set standards or certifying organizations, a tiered approach is proposed to assure the selected modeling or data analytic tools meet the needs of an organization and its functional use.

NEXT STEPS

A first step is to recognize that leadership is critical for setting the right example and building the right environment for change from the top down and bottom up. It will take champions at all levels in existing professional organizations, governmental agencies, private industry and academia to encourage and implement a robust hydroanalytic framework.

The second logical step toward a robust evaluation strategy is to harmonize the evaluation standards and practices across the hydroanalytic community of practice. Because the hydroanalytic community covers a broad range of disciplines, there could be a suite of standards where lessons learned from one area could be applied and merged with others.

Third, to track progress, organizations should set goals or targets for success. Organizations should establish detailed and clearly defined metrics for managing performance in each leg of the evaluation stool and/or throughout the modeling cycle.

Fourth, the evaluation process must be institutionalized and implemented. One place to start could be done by leveraging current federal interagency mechanisms to establish a governance platform for broad implementation of hydroanalytic standards. On a smaller scale, individual organizations can require hydroanalytic evaluation as they build new and/or update or modernize their current portfolio of hydroanalytic tools.

Finally, the time seems right for the hydroanalytic community to come together and embrace the changes needed to develop and implement an evaluation framework.

Chapter 1

SETTING THE STAGE

INTRODUCTION

Revolutionary changes are occurring in data and modeling analytics. Hydroanalytics — defined broadly as any analytic data and models that involve water-related processes — will by necessity be transformed as basic applications (i.e. hydrometeorology, hydrology, hydraulics, sediment transport, coastal circulation, wave dynamics, groundwater) morph to more advanced applications that will require integration and inclusion of non-traditional data/models (i.e. remote sensing, social media, market data, ecology, behavioral sciences), a push to use leading-edge methodologies (i.e. artificial intelligence, deep learning, neural networks, coupled probabilistic and physics-based methods), the introduction of transformational technologies (i.e. blockchain, micro-processors, drones, sensor technologies, high- performance, cloud and quantum computing) and better representation of complex processes (i.e. large-scale coupling of interdisciplinary models, agent-based modeling, advances in scientific understanding of natural processes). Each additional layer of complexity adds to an already dizzying portfolio of hydroanalytic tools making it even more important to have rigorous management and evaluation strategies. Further, because of the broad interdisciplinary nature of the field, as well as the many active organizations and rapid transformation that is ongoing, collaboration with others is essential to leverage ongoing advancements and innovative technologies.

A robust evaluation framework is needed in the hydroanalytic community of practice to assure transferability, repeatability, reliability, usability and overall quality of data, models and information. The definition of evaluation in this report, includes any criterion, process or methodology that reviews and documents the quality and performance of software, data and models. Currently, there are dozens of models and data systems that inform critical decisions by federal, state and local governments, public sector businesses and individuals to plan for capital investments in ports, waterways, flood risk reduction and ecosystem restoration projects, improve operational activities of our infrastructure, recovery and mitigation of major hydro-climatic events, and preparation of individuals and communities to be more resilient. It is important that the hydroanalytic information and methodologies used to inform these decisions are consistently validated and verified, correctly applied, managed and documented, open and reproducible and help to quantify and/or reduce uncertainty.

APPROACH

The intent of the study was to gather and provide a synthesis of best ideas to evaluate the quality and performance of hydroanalytic tools. Further, the study findings are designed to inform organizations and individuals that develop and use hydroanalytic information as they apply, advance and modernize numerical products. The study findings are based on the author's interviews and discussions with experts (more than 70), review and analysis of models and model inventories and a robust collection of documents and reports. Additionally, the author took a more detailed look into the practices and policies of several large modeling organizations/agencies — Deltares (the Netherlands); the Environment Agency and HR Wallingford (United Kingdom); the U.S. Army Corps of Engineers; the National Oceanic and Atmospheric Administration-National Weather Service; and the Federal Emergency Management Agency.

The study begins by making a case for developing and implementing evaluation strategies by describing the drivers for change. A landscape overview follows presenting many of the current evaluation practices used across organizations and in literature relative to uses, users, technical and institutional practices. These practices lead to an overview of emerging hydroanalytic methodologies that are transforming and challenging more traditional hydroanalytic methods. The concepts are synthesized in Chapter 5 providing building blocks to implementing hydroanalytic information including a recap of best practices, the roles and responsibilities of evaluation throughout the model cycle, conceptual frameworks for product and portfolio quality assessment and moving the practice toward certification. The last chapter describes the need for measuring our progress toward evaluation and next steps.

DEFINITIONS

Accreditation
According to the National Institute of Standards and Technology (NIST 2017), accreditation is distinguished from certification as a designation that indicates an organization is competent and has quality management systems to perform certain activities. NIST refers to certification as the qualifications of an individual or a product. These definitions vary for other organizations.

Accuracy
Commonly defined as the ratio of the number of correct predictions over the total number of predictions.

Bias
According to National Oceanic and Atmospheric Administration (NOAA)'s glossary of forecast verification methods, bias is "the degree of correspondence between the mean of the forecasts and the mean of the observations."

Certification
NIST suggests that certification applies to professionals in certain fields and with certain skills that have been recognized from a reliable third party. In this sense, certification can also apply to model software meeting various reliability and integrity criteria established by a reputable third party.

Code
According to Demirbilek and Rosati, 2011 code is "software that implements the solution algorithms." In context of this report, the code may be developed using a variety of different programming languages dependent upon the type of problem and the capabilities of the code developer.

Community of Practice (CoP)
In its simplest form, CoPs are groups of individuals having a common interest or professional background that come together organically or deliberately. The USACE has deliberately embraced CoPs to strengthen key disciplines within the organization.

Credentials
In modeling, this can be the qualifications one possesses to use a specific software.

Curation
According to Lord 2004, curation "is the activity of managing and promoting the use of data from its point of creation, to ensure it is fit for contemporary purpose, and available for discovery and reuse."

Graphical User interface (GUI)
The GUI helps modelers and users navigate the complexities of the model using intuitive visualization tools. The GUI may aid the modeler in importing data, organize files, visualizing outputs and launching the code.

Hydroanalytic
For purposes of this paper, hydroanalytics refers to any type of analysis of information whether using simple algorithms, complex physics-based models and/or big data analytics to better understand, predict, or review anything related to water including riverine, coastal, hydrologic, near-field hydrodynamics, vessel-water interactions, waves, sediment processes and other related processes.

Model
In USACE Enterprise Standard 08101 (USACE 2011), a model is "an application or implementation of a piece of software created for a specific purpose." Models are intended to represent a real-life situation or scenario.

Open Source Software (OSS)
Defined by DoD (2009), OSS is "software for which the human-readable source code is available for use, study, reuse, modification, enhancement and redistribution by the users of that software."

Planning Model
According to EC 1105-2-412 (USACE March 31, 2011) "planning models are defined as any models and analytical tools that planners use to define water resources management problems and opportunities, to formulate potential alternatives to address the problems and take advantage of the opportunities, to evaluate potential effects of alternatives and to support decision-making. It includes all models used for planning, regardless of their scope or source."

Quantities of Interest (QOI)
QOIs are the quantities representing a physical parameter of interest. In VVUQ process, the computed estimate of a QOI, say discharge or water surface elevation, will be compared to a measured value under similar conditions (National Research Council 2012).

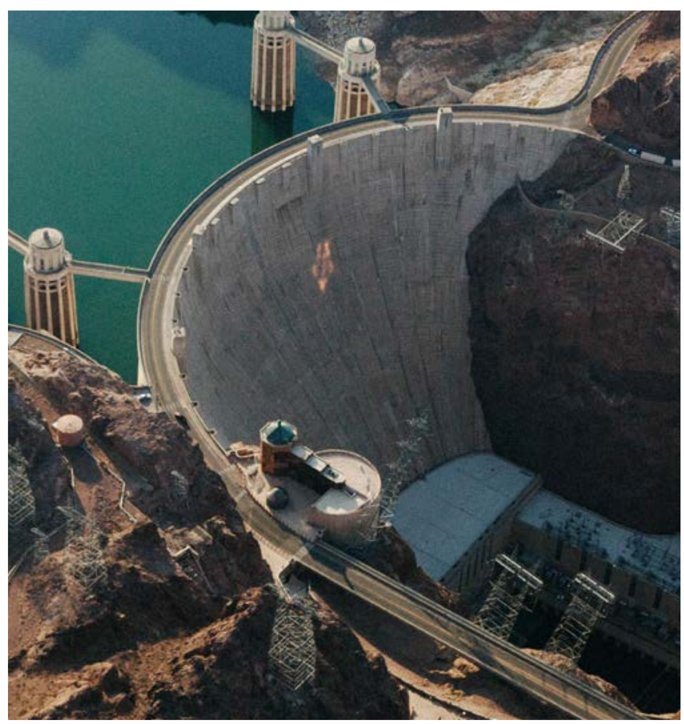
Skill Score
According to NOAA's Climate Prediction Center, skill scores "measure improvement of the forecast over the standard."

Software
In Enterprise Standard 08101 (USACE 2011), software is specifically defined as the "native code (programs) that informs a computer what to do and the associated user interface that allows the required and optional parameters to be introduced."

Uncertainty Quantification (UQ)
According to the National Academies (NRC 2012), UQ is "the process of quantifying uncertainties associated with model calculations of true, physical QOIs, with the goals of accounting for all sources of uncertainty and quantifying the contributions of specific sources to the overall uncertainty." UQ addresses the sources of error and uncertainty for QOIs.

Validation
According to the same paper (NRC 2012), validation is "the process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended uses of the model." It addresses how accurately the model represents reality for the QOIs.

Verification
And finally, the paper (NRC 2012) defines verification as "the process of determining how accurately a computer program ("code") correctly solves the equations of the mathematical model. This includes code verification (determining whether the code correctly implements the intended algorithms) and solution verification (determining the accuracy with which the algorithms solve the mathematical model's equations for specified QOIs)."



Hoover Dam, Nevada

Chapter 2

DRIVERS FOR EVALUATION

What would drive an organization to create and institutionalize a new paradigm in modeling evaluation and management? Why now? What are compelling reasons to implement and standardize evaluation strategies? Is there a gold standard for modeling practices? In review of world-class modeling organizations many faced challenges that drove them to change and/or modernize their approaches to managing their hydroanalytic portfolios. These included the desires to:

- Reduce and communicate uncertainty
- Be relevant and/or best-of-class
- Address tough and complex problems
- Optimize and stretch diminishing resources
- Adapt to disruptors and adopt innovative science and technology
- Be transparent and collaborative.

REDUCING AND COMMUNICATING UNCERTAINTY

George Box, a statistician, famously said in 1976, "All models are wrong, some are useful." Uncertainty is inherent to natural phenomena and how it is modeled. Statisticians and probabilistic specialists often describe uncertainty as either aleatory or epistemic. Aleatory, or stochastic uncertainty, relates to the natural or inherent variation of parameters or physical conditions. These perturbations can cause the outcomes of models and experiments to be different for each run. Epistemic uncertainty relates to imperfect knowledge. We don't know what we don't know. (Fox and Ukumen 2011).

Much effort is spent in the modeling community of practice to reduce these uncertainties. Considerable research and investment go toward improving how physical processes such as sediment transport are represented, or adjusting algorithms or coefficients to hit a benchmark such as water elevation. These are all important advances to the state-of-the-science. Yet, even the most sophisticated scientific-based approaches can accumulate uncertainties in the modeling chain and/or overlook a significant error source when applied. Even when modeling outputs may be certain or quantifiable, such as velocities or water levels, when coupled or integrated with other less accurate or more artful modeling processes, such as geomorphology or ecology, the outputs can be misleading and imply accuracy that can't be supported. A recent study in Nature Communications demonstrated the inherent error in quantifying the vulnerability of coastal communities from sea level rise forecasts due to the accuracy of the elevation data, not the climate science. The mean error in the elevation data exceeded the uncertainty in the predicted range of sea level rise. By improving the accuracy of the elevation data, the authors determined the global consequences of sea level rise were three-fold higher than previous estimations (Kulp and Strauss 2019)!

Unintended consequences can occur from providing data and modeling outputs without quantifying uncertainties or by communicating their meaning poorly. For instance, the National Weather Service (NWS) has an excellent reputation for communicating forecasting information, based on years of improvements and inclusion of social science. However, even the best products, such as the hurricane forecast, can leave the public confused. The cone of uncertainty used in the hurricane updates represents the statistical error in the forecasted path of the hurricane eye. Yet, the public may assume they are out of harm's way if they are located out of the boundaries, and/or if the storm category is lower at landfall. Yet, the size of the hurricane may expand impacts well beyond the cone boundaries of the cone and location of eye landfall. Additionally, surge and precipitation can dominate and/or exacerbate consequences that are not reflected in a wind-speed-only characterization, as seen in many events over the past few years. Therefore, certainty in forecasting must also be measured by how it is interpreted and understood.

A theme from numerous modeling experts was that over-investing in complex analytics may not be necessary if simpler approaches can address the question and uncertainty can be articulated. Misapplication of good products such as flood insurance rate maps and hurricane forecasts can lead to poor decisions, costly unintended consequences and even litigation against providers of these products. Progressive organizations are investing more into social and behavioral sciences to understand how to evaluate, visualize and present complex information that can improve public confidence and help in decision making. Many leading scientists and modelers are addressing technical uncertainties whether aleatory or epistemic, using methods such as Validation, Verification and Uncertainty Quantification, VVUQ, ensemble modeling and Bayesian statistics. While there will always be some inherent uncertainty in modeling, reducing it and explaining it are best practice.

BEST OF CLASS

Being a large organization, federal agency or corporation is not a condition for preeminence or excellence in a field. As we have seen in the growth of the tech industry, large corporations may dominate a market area for a time. Their success may have originated with an innovative idea in a garage start-up, but sustained success grew with further improvements of the original idea and likely with the merger and acquisitions of both large and smaller companies. Even then, the real success is evaluated by the bottom line to the shareholders and is driven by the consumer.

In discussions with many first-class hydroanalytic modelers, all were passionate believers in the software and products they were developing. In speaking with leadership from various organizations, whether academia, government or consultants, all were committed to providing the best results and striving to be pre-eminent in their respective area of expertise. While competition is real amongst model products, organizations and suppliers, there is also space for expanding capabilities, adding niche products, and considering the value of integrating and combining efforts. A history of being the primary developer and supplier of flood models, for instance, is not a satisfactory condition to stay on top.

An introspective review would ask: Who do we consider leaders in the field? How do they maintain their status? What is the gold standard in modeling products or studies? These questions really drive home the need for evaluation strategies that can identify best-of-class or status relative to others. Both successes and failures must be viewed from an objective perspective. Best of class is not driven by how an organization views itself, but how others view it.

ADDRESSING TOUGH PROBLEMS

Hydroanalytic information will need to go beyond the current state-of-practice in hydrology, hydro-met and hydrodynamic data analytics and modeling to address at least three grand challenges: aging and inadequate infrastructure, population dynamics and development, and the escalating impacts of climate change in responding to both acute and chronic hazardous events. Sociopolitical, national and/or cybersecurity issues and the acceptance of evidence-based information compound these challenges. The American Society of Civil Engineers 2017 Report Card estimates that a $4.6 trillion investment is needed by 2025 in the US to replace and repair aging infrastructure (ASCE 2017). Much of this infrastructure is indirectly impacted by water and climate (i.e. bridges, roads, rail, energy) and/or is directly designed to manage water resources (i.e. levees, dams, storm and wastewater systems, water distribution systems, ports and waterways). People moving from rural or small communities to large metropolitan areas put more stress on aging infrastructure designed and built decades ago, often for a smaller population. Meanwhile, smaller communities are left without the capital to reinvest in the most critical infrastructure needs to provide clean water and treat wastewater. People, ecosystems and infrastructure are increasingly exposed to climate change impacts such as extreme weather events, heat, drought and sea level rise. Not only are these impacts occurring simultaneously, such as riverine and coastal flooding during a major hurricane event, they can combine rapidly with other vulnerabilities to trigger cascading consequences such as melting permafrost, eroding coast lines, fires and degraded water supplies.

World class analytic solutions should strive to represent the inter-relationships between natural, sociocultural and built systems; model large regionally diverse and complex problem domains; optimize solutions to address multi-objective requirements; model interdependencies within and between systems; include non-traditional socioeconomic benefits and impacts; quantify the probabilities and impacts of compound events; evaluate socioeconomic alternatives to mitigate the rising cost of disaster; and provide ways to transparently communicate to the public.

As the problems to be modeled become more complex, costly and dangerous, a new modeling paradigm is needed. It will include non-traditional data/models (i.e. remote sensing, social media, market data, ecology, behavioral sciences), use leading-edge methodologies (i.e. artificial intelligence, deep learning, neural networks, coupled probabilistic and physics-based methods) and transformational technologies (i.e. blockchain, micro-processors, drones, sensor technologies, and high performance, cloud and quantum computing), and better represent complex processes (i.e. large-scale coupling of interdisciplinary models, agent-based modeling).

OPTIMIZING RESOURCES

Whether an operational agency or program, a research laboratory or model developer, a for-profit provider or an academic institute, optimizing resources drives investment decisions in hydroanalytic tool development and application. With fixed or fluctuating budgets, there is always a tension between investing limited resources to deliver results on time, reduce costs and/or improve quality.

Federal operational organizations are often measured by how many permits, projects or grants they can issue or complete with a fixed amount of appropriated funding in a fixed amount of time. They have statutory requirements, that may be dated or constantly changing, that may also be very prescriptive and limit their flexibility to incorporate change. For example, within FEMA, where hundreds of new flood maps are updated annually, there is a tension between meeting the statutory requirements of the National Flood Insurance Program within the authorities and appropriations provided and modernizing hydroanalytic practices. For providers under contract, implementing changes could be more costly upfront and/or slowdown production of new flood studies and maps, jeopardizing their performance.

A 2007 NASEM study indicated that Federally funded research is underfunded yet critical to transformational change (National Academy of Science, National Academy of Engineering, and Institute of Medicine 2007). Research is needed to modernize technological approaches as discussed above, and to reduce uncertainties. Yet, funding for research often is a low priority and competes with other pressing operational demands. Further, the small federal investments in research can ebb and flow depending upon subjective or unpredictable drivers including administrative priorities, Congressional authorities/appropriations, inter-relationships between funding providers and researchers, technical obsolescence, and often crisis, such as a disaster. Within that context, some federal research organizations question the trade-off between investing limited and/or unpredictable resources to research new methodologies and software, versus simply hiring the best trained professionals to provide science and engineering technical services to those that do not have capacity and/or capability. Without federal research funds, another option for agencies has been to outsource or procure new technologies. By not maintaining some baseline investment in innovation, whether as a research or operational organization, the ability to be technically relevant and a credible provider of hydroanalytic information is limited and will begin to decline.

To overcome the research investment shortfall, some organizations both in the U.S. and abroad are finding other ways to innovate. Entrepreneurial for-profit companies are pressing the state-of-the-art in hydroanalytics with minimal seed funding in hopes of finding customers that will look to them as current modeling capabilities become obsolete or unjustifiable. Some agencies, particularly those that provide hydro-met services such as the National Weather Service and UK MET, have a history of investing in research and transitioning it to operations. Importantly, world-class organizations that value sharing and collaborating contribute positively to the advancement of the community of practice and can leverage limited resources. The bottom line, as supported by the NRC study, is that commitment to long-term basic research to modernize operational systems and transform applications most certainly has a high return on investment and strengthens quality of life.

DISRUPTERS AND OPPORTUNITIES IN SCIENCE AND TECHNOLOGY

Innovations must go beyond incremental evolution of individual problem areas to better address the current challenge of integrating modeling environments, networks and systems. The hydroanalytic community of practice (CoP) must also be prepared for radical or disruptive changes in technology – small-scale and cheap processors are changing sensor technology and generating much more data to assimilate into models; satellites and drones are providing new ways to analyze large or inaccessible areas with greater detail; quantum computing, blockchain and big data analytics could also be transformational drivers to solve complex high performance computing challenges.

An area already being embraced by leading organizations is big data analytics (Artificial Intelligence (AI), Machine Learning (ML), Artificial Neural Networks (ANN), etc.). The challenges and opportunities driven by big data analytics are becoming evident and may result in disruption to the current paradigm of hydroanalytic modeling. Models are used to understand the world, make predictions, or help with responses to change. With the availability of big data and disruptive technology tools, data analytics could leapfrog current physical modeling approaches to address objectives. Yet, a rush to replace traditional modeling with data analytics without proper validation could result in poor representation of actual physical conditions and add more uncertainty to the outcomes. Because AI/ML methods are data driven, they will have difficulty addressing a highly changing system. Blended approaches, such as dynamic and ensemble modeling or combining big data analytics with physics-based models, are viable approaches and pressing their way into mainstream hydroanalytic computing. More work needs to be done now to validate the uncertainties and uses of such approaches.

Addressing the science and technology disrupters will require investments in model development and research while also attracting a new type of work force and subject matter experts. Leading organizations are not waiting to be disrupted but are actively incorporating these revolutionary technologies into and within their existing methodologies.

TRANSPARENCY AND COLLABORATION

Transparency and coproduction are critical to keep pace with others, assure products are viable and meet the needs of the decision makers. For instance, many of the leading organizations reviewed embraced community modeling and open source access. Deltares' mantra for their software is "Dare to Share."[1] Open science and open data are a requirement in US government as evidenced in an order from the Obama administration (Holdren 2013) and enacted by Congress (Public Law No. 115-436 2019). Sharing findings with other innovators at conferences, workshops and in peer-reviewed journals provides opportunities for further collaboration and credibility. Collaboration can stretch limited resources while improving tools and models.

A Department of Homeland Security (DHS) Science and Technology Directorate (S&T) sponsored workshop on Rethinking Flood Analytics, held at the University of North Carolina, Chapel Hill in 2017, concluded that transformational changes to our current state of modeling practice are contingent upon embracing transdisciplinary approaches and coproducing with end-users. The workshop also concluded that state-of-the-art analytics and leveraging technology would be required to meet the challenges of the future (Knight et al. 2017). To embrace coproduction and transparency, several leading organizations have built the process of client acceptance into their development strategy. As can be seen by many of the more recent research funding opportunities, transdisciplinary approaches are more often being awarded grants. This is critical as the complex problems described above will, by necessity, require a diverse set of disciplines to generate solutions.

To maintain credibility as world-class leaders in hydroanalytic modeling and solutions, organizations must be prepared to not only address the technical challenges associated with a changing planet, but also adapt their strategies to compete in a global marketplace. They must be committed to providing credible science and engineering solutions that withstand sociopolitical scrutiny while informing resilient solutions to grand challenges. To stay competitive in this rapidly evolving market with other national and international hydroanalytic laboratories, agencies and consultants, successful organizations will need to modernize their modeling and data portfolio by incorporating transformational science and technology, embracing collaborative and transdisciplinary approaches and measuring their progress through a robust evaluation framework.

Attendees of the ADCIRC Users Group Meeting, an annual gathering of practitioners and end users, discuss the model in between presentations at a 2019 gathering. ADCIRC is used by many agencies as a tool to predict the impacts of hurricane storm surge and flooding on coastal communities. *Photo by Chris A. Johns / Coastal Resilience Center of Excellence.*

Chapter 3

LANDSCAPE REVIEW OF HYDROANALYTIC PRACTICES

This chapter captures a snapshot of an ever-evolving landscape of current practices in evaluating hydroanalytic information. While highlighting and extracting the best practices, the review also uses examples to explain weaknesses or gaps. Further, many of the methodologies and practices reviewed relate to the technical merit of the information, but it was clear that users, uses and institutional oversight were equally important in an evaluation framework. Acceptance or buy-in of modeling tools and data by the client or end-user requires their engagement early and often and a transparent framework for sharing. The institutional guidance — policies/governance/business processes/human resources — must establish clear and authoritative directives, be built on an organizational framework that streamlines functions and by organizations, and those that lead them, that embrace a culture of change and innovation.

Therefore, best approaches to manage and standardize processes, evaluate quality and communicate uncertainties for evolving and future hydroanalytical information were synthesized into three foundational areas: user engagement, technical requirements and institutional guidance (See Figure 1). The best ideas and lessons learned on evaluation practices reviewed in these foundational areas were considered by the author as important to the long-term viability and sustainability of hydroanalytic tools. Institutional practices are shared through examples that highlight policies and standards, programmatic evaluations, organizational governance and business processes.

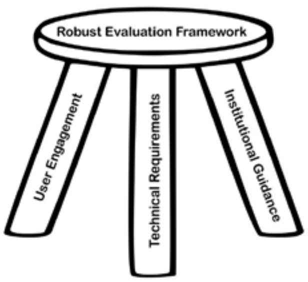


Figure 1. Evaluation Framework Stool

USER ENGAGEMENT

Starting with the user leg of the stool, the following are explored — the users and their functional needs, the fitness-of-use of the models to the problem, the acceptance of the tool(s) and the capacity and capability of the users or their organizations. The latter includes a discussion on how to assess or certify the qualifications of professionals and organizations.

Know the Users

Successful organizations that develop and apply hydroanalytic information understand their users. Users can range from the practitioners that apply models to officials who depend on outputs to make strategic decisions. The broad range of users of hydroanalytic information may include agency regulators, academia, urban planners, real estate developers, legal teams, homeowners, elected officials, other modelers, contractors and other service providers. The requirements established by the end-user and the skills and resources available to those users drive development of the tools needed and dictate their quality, complexity and performance. This may seem obvious, but a hydroanalytic tool may have been originally developed with one purpose or user in mind and applied in other areas that may or may not fit the original intent of the software. Further, the model may be applied by persons who do not understand its use and/or are not skilled in using it. These could lead to misapplication and impact the credibility of the model software developer. That is why it is critically important to learn more about the product's current and prospective users/customers. Who is or will be using the product?

As an example of widely used models, Corps hydrology, hydraulic and coastal models were developed to accomplish the Corps' mission, but the uses have expanded well beyond Corps projects and Corps field users. Who are the Corps model software users? National Weather Service (NWS) uses Corps models in many of their river forecasts for inundation mapping. Federal Emergency Management Agency (FEMA) and the Federal Energy Regulatory Commission (FERC) include certain Corps models as acceptable for use by their contractors and field offices in evaluation of activities relative their agency missions, such as floodplain mapping or dam safety, respectively (FERC 2014). More profound, but not inventoried by the Corps, is the use of these models by academia, contractors, state transportation agencies and other governmental organizations around the world.

FEMA, one external user of Corps products, sets the guidelines and standards for cooperating technical partners and contractors in conducting flood insurance studies and mapping floodplains under the National Flood Insurance Program (NFIP). FEMA maintains a list of approved FEMA Hydraulic Numerical Models, hydrologic models and coastal numerical models based on the Guidance for Flood Risk Analysis and Mapping: Accepting Numerical Models for Use in the NFIP (FEMA 2018). These models must meet the minimum NFIP requirements of the code of regulations (44 CFR 65.6 (a)(6)) for flood hazard mapping activities. FEMA does not technically review or test models but depends on approved federal and non-federal agencies to certify the models. Many of the models on the approved lists are Corps models. Because of the broad national application of FEMA products to over 21,000 communities, this means that a large portion of the nation's flood mapping is influenced by the application of Corps models.

The large portfolio of Corps model users is a positive indication of the importance and general acceptance of Corps models. Yet the technical validation of these models is largely based on an internal process that aligns a certain modeling software to the planning or engineering requirements of Corps studies. Further, the investment strategy for Corps modernization and model management is largely driven by internal requirements and does not consider the benefits of those investments beyond Corps use (USACE June 11, 2011). That is not to say that these models are not appropriate for FEMA use, but FEMA should take care that these approved models continue to meet their mission and quality needs.

The general acceptance and use of agency-developed models and products by others makes it even more important to assure that these models are rigorously evaluated and managed. The developing agency and the user's credibility are both tested when these products are applied beyond the purpose for which they were originally developed without evaluating their fitness-of-use (see below). Therefore, it is critical to understand the current and potential users and uses as modeling and data tools transfer from development to mainstream application. User studies and customer surveys are one way to help organizations learn more about the product users and their evolving requirements, provide a pathway for continuous product improvement and help expand the credibility of the market brand.

Fitness-of-use

How will the model or data be used? In the case of hydroanalytic information, the uses are many and include forensic studies, real-time forecasting, project design, strategic planning, community planning, floodplain management, response and preparedness, investment prioritization, commodity routing, insurance rate setting, risk assessment, risk communication, infrastructure operations, property valuation and the list goes on. One subset of uses of hydroanalytic models and products, flood analytics and mapping products is shown in the following list:

- Real-time forecasting
- Risk assessment
- Emergency planning and preparedness
- Emergency response and recovery
- Strategic planning
- Infrastructure operations
- Project design
- Investment prioritization
- Urban planning
- Risk communication
- Insurance rate setting
- Floodplain management
- Forensic studies
- Resilience planning

Clear articulation of the model and data use is the first step in building the right product. Additionally, the selection or development of a modeling tool should assure the accuracies between the data and the modeling outcomes are aligned.

A more refined question might be, what decisions will be made with this tool? The answer to that question will dictate the need for, say, speed versus accuracy or complexity versus simplicity. It will

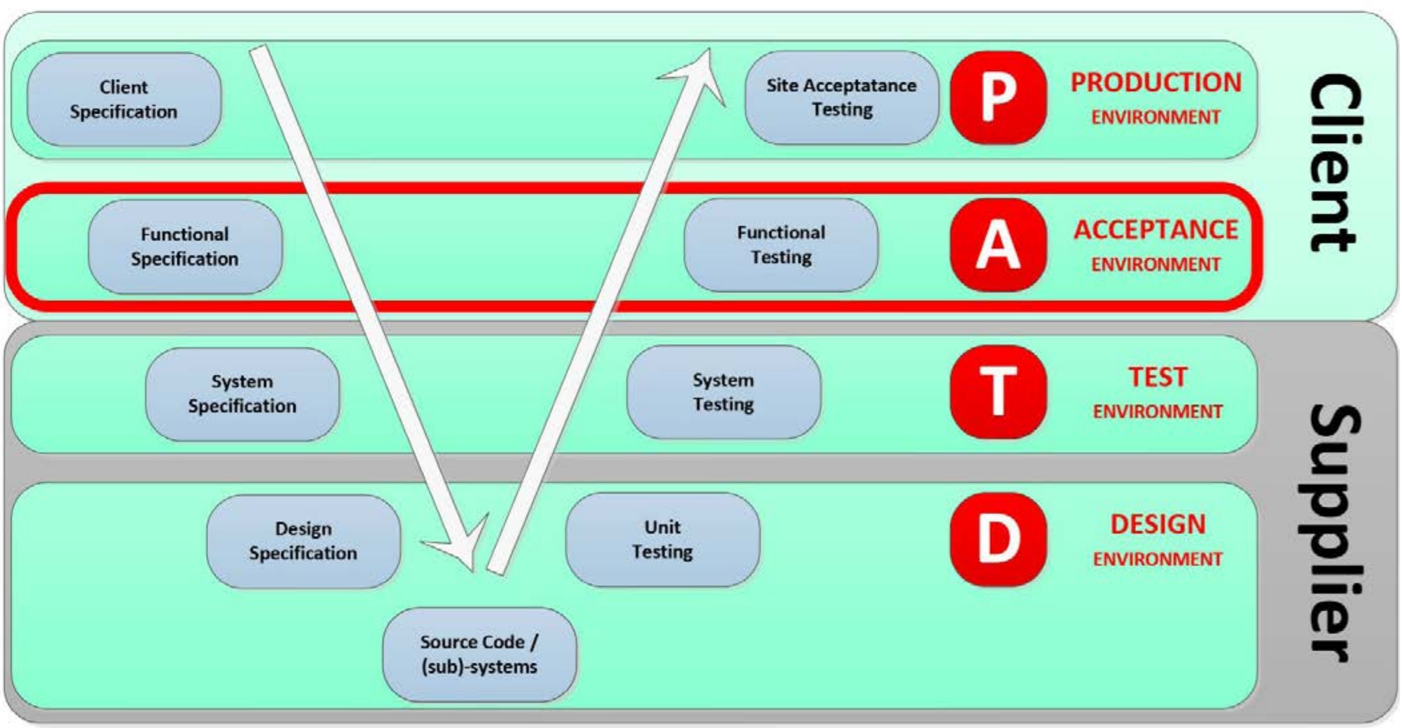


Figure 2. Concept model of client acceptance shared with permission by Deltares (deJong 2018).

also drive the accepted tolerance level for uncertainty. Scientists and engineers often operate in the slow thinking mode, characterized by Daniel Kahneman in his 2011 book Thinking, Fast and Slow, as logical and analytic (Kahneman 2011). By their very nature, they strive for accuracy and precision, but sometimes what the user needs is a "good enough" answer. For example, some military decisions and emergency operations may be driven by time-critical factors where data provided through models or forecasts must meet the timing needs of the user. If the best answer arrives too late, it is of little value.

Other requirements that dictate accuracy or the level of analysis, tie to the criticality or safety requirements of the outcome and less to the timing. Using hydrology and hydraulic (H&H) models to determine the design height and loading for a dam or sea wall requires much more certainty and accuracy to assure safety of property and people, than using H&H models to conduct a reconnaissance level planning study. In addition to the agreement of accuracy between model and data, selection of the best tool should consider the appropriate accuracy needed for the decision. If the wrong answer arrives on time, there can be adverse consequences.

The actual physical processes or the outcome of interest being modeled is a priority in model selection. Many traditional hydroanalytic models, such as hydrodynamic codes, are grounded on physics-based approaches whereas other methodologies may rely more on data-driven solutions or empirical assessments. Yet, even the most sophisticated approaches can accumulate uncertainties in the modeling chain and/or overlook a significant error source when applied. As discussed above, understanding the aleatory or epistemic nature of the problem to be solved, can influence the solution methods. Remember, we don't know what we don't know. As the Validation, Verification and Uncertainty Quantification (VVUQ) process demonstrates, in the technical requirements section of the report, a model may be better at calculating one quantity of interest, say for instance wave height, while not meeting the level of certainty of another parameter such as water velocity. The use of ensemble modeling (to be discussed later in the report) or probabilistic analysis over standard deterministic approaches gives users a range of potential answers that can help bound possible outcomes and frame uncertainty.

Along with the type of decisions being made is the usability of the products. Matching the right tool to the right problem also means matching it to the right system. NIST led an international consortium to develop a Common Industry Format (CIF), a reporting standard, for evaluating the usability of commercial software products. According to the website, there are four steps in usability:

- Environment which system will be used
- Purpose of system
- How the system fulfills requirements
- Evaluation of system

There are pros and cons to using widely accepted models across many applications. For example, there has been considerable federal investment in the FEMA mapping products since the establishment of the NFIP along with widespread availability of these products over a considerable portion of the US. The maps and the flood studies supporting them provide a baseline for communities that participate in the NFIP to use these plans for floodplain management and land-use planning. The positive aspects of this national effort are the availability of large datasets, widely accepted processes and results, centralized funding and programmatic governance. The downside is that the product development is decentralized (by regions and locality), the quality varies according to best available information at time of product development, the data can be hard to access, and the results may be contested. Because of the widespread availability, the NFIP products are used for analysis and decisions beyond the intended application, such as property value assessments, litigation of losses or detailed infrastructure planning.

Whether used for forecasting, forensics, planning, design or operations, the fitness-of-use of a model software to a specific application and the availability and quality of data should be evaluated prior to selecting the best tool for a problem. Importantly, that selection must support and be supported by the ultimate end-user. The UK Environment Agency is moving toward a scoring system that evaluates the quality of various models for their fitness-of-use for certain application areas. This approach will be further discussed in technical requirements.

Acceptance and Tech Transfer

Modeling organizations must identify what services and products are needed by their customer. The organization must be able to articulate the user's problem or requirement (see fitness-of-use), understand the capability and capacity of the user (see below) and have user buy-in. To achieve the latter, successful organizations maintain close relationship with their customers. Iowa Institute of Hydraulic Research (IIHR), a state-supported laboratory, does not have a formal user buy-in process, but works very closely with customers throughout the project to assure that IIHR is providing the support, services or products needed[2]. This builds a level of trust and credibility for the organization. Others use more formal processes. Deltares uses Development, Testing, Acceptance and Production (DTAP) as key phases of its approach. In particular, the acceptance cuts across all the areas and is part of a strategy they developed and call the V-model (Figure 2). This acceptance strategy or a similar one would be an excellent model to use when building complex models for operational use.

Other large-scale studies such as the Coastal Louisiana Master Plan and the U.S. Army Corps of Engineers North Atlantic Coast Comprehensive Study (NACCS) relied on extensive stakeholder engagement to identify the most pressing issues and inform their model frameworks. In the 2017 Coastal Master Plan, the outreach and engagement strategy were designed to develop public confidence and support for the technical framework as a decision-making tool for the Louisiana citizens (Speyrer and Gaharan 2017). For the NACCS study, visioning and partnership meetings were held throughout the region to create a shared vision and framework for addressing coastal flood risk management (CDM Smith 2014).

Transition of research to applications has a long history of being a difficult process and is often portrayed comically as the valley of death — the place where all good ideas go to die. Yet, modeling activities do advance and are improved, so there is also a history of success. Organizations such as NASA and DoD have employed transition readiness levels (TRLs) to assure that technologies are thoroughly tested in a stepwise fashion and are operationally ready prior to use (DoD 2010).

Research organizations regularly employ external review of their studies, portfolios and/or their laboratories. As an example, quadrennial laboratory reviews are conducted in NOAA's Office of Oceanic and Atmospheric Research by external experts to "evaluate quality, relevance and value of research and development to both internal and external interests, and help to strategically position the laboratory for science planning in the future."(NOAA-OAR 2007)

Sometimes the acceleration of demonstration or research tool(s) is necessary to meet a critical societal need or technical crisis. This was true post-Hurricane Katrina in the acceptance and use of more advanced models to conduct the forensic studies led by the Interagency Performance Evaluation Team (USACE 2009). The study prompted the use of robust, complex tools that included a 2D circulation model, ADCIRC, a nearshore wave model, STWAVE and a new characterization of storm surge frequency, the Joint Probability Method. The acceptance of these approaches for this urgent study led to a broader acceptance and use by the CoP. The new modeling approaches helped to inform the design of the Hurricane Storm Damage Risk Reduction System in New Orleans and were later identified as preferred methods for coastal mapping products at FEMA. External subject matter experts were a critical part of the acceptance process to assure the modeling techniques provided best possible fitness-of-use.

Whether simple or complex, the most effective product delivery or tech transfer occurs in environments that are conducive to coproduction. Colocation of the client and developer helps. For instance, NOAA's National Severe Storm Lab is co-located in Norman, Okla. with the University of Oklahoma's National Weather Center and a National Weather Service Forecast Center. Similar arrangements occur at the National Water Center in Tuscaloosa with the University of Alabama, and at Deltares where they are in close proximity to the Technical University of the Delft. These co-located facilities are conducive to the interactions and trust needed to build better products. In absence of such arrangements, organizations must make it a priority to engage the client early and often, as shown in the Deltares V diagram (previous page) or enter into agreements such as TRLs.

User acceptability is an important requirement. Operational organizations that accept the "readiness" of the product must be prepared to defend its viability as they are financially, and often legally, responsible for its use. There are viable transition pathways for new products that can be driven by necessity but grounded in expert evaluation and review.

Building Capacity and Capability

Understanding the user also means understanding their capacity and capability to use the model outputs or to run the model. Capacity building improves the credibility of the giving and receiving organizations. While all software and models are not transferrable (some software may require the developer or a highly skilled modeler to provide a service), building capacity and capability is still an important goal to fulfill. It has also been demonstrated that even complex modeling approaches can ultimately be transitioned as software and hardware improve and users obtain more skills.

Corps district offices have long depended upon in-house capabilities to perform basic hydrologic and hydraulic (H&H) modeling using Corps tools to support their local and regional projects. The Corps Hydrologic Engineering Center (HEC) has been an important organization in delivering the H&H products and providing training and support for both Corps district and external users of their products such as HEC-RAS or HEC-DSS. Since the establishment of the U.S Waterways Experiment Station in Vicksburg, Miss., in 1929 following the 1927 flood on the Mississippi River (Cotton 1979), the Coastal and Hydraulics Laboratory has been the organization within the Corps to address complex hydroanalytic problems using its extensive physical and numerical capabilities. The CHL continues to support development and use of sophisticated state-of-the-art software and tools to support the Corps mission. Many of these tools ultimately transfer to practitioners within and external to the Corps. The transition to field offices depends on the capability of the modelers in the district offices and the computing capacity accessible to run higher fidelity and complex models. CHL, like HEC, provides model support and training for their models to build capacity and acceptance. The CHL can provide the hydroanalytic services to support complex projects.

FEMA also has an interest in building capacity and capability. The Cooperating Technical Partners Program was designed to provide federal grants to support building that capacity and capability at the local level. According to FEMA's website, the agency has more than 240 partnership agreements in place. Partners have an opportunity to participate, if eligible and awarded grants, in an assortment of activities related to the flood mapping program. Where technical capabilities may be lacking, FEMA supports eligible partners that can provide outreach and education. Yet most of the modeling and data analysis done for the NFIP is still provided by contractors. Providing the models and data used to develop the Flood Insurance Rate Maps and conduct the Flood Insurance Studies to local governments and/or their consultants for continued improvement and use would be ideal, but local communities may not be able to utilize them. Growing the CTP program would help build local capacity and capability. Local authorities could then capitalize on the rich understanding that comes with operating, advancing and applying a product that they would continue to use for planning and operations.

Some specific approaches used to build capabilities and capacity in applying and using hydroanalytic models and data are user groups, software boot camps, model "fairs," competitions and more traditional training and outreach. Deltares hosts annual software users' group meetings where users from around the globe share their experiences through peer-to-peer networking. The users improve their own skills and share lessons learned with others. According to Deltares, another benefit is the unsolicited positive testimonials provided by the product users. Others may host software fairs or training academies, like the Danish Hydraulic Institute, DHI. The ADCIRC CoP comes together for user groups meetings, boot camps and career fairs to expand capabilities among users. Often hosted at universities and managed by a third-party consultant, these activities are popular ways to get ADCIRC users up to speed fast. Traditional training and outreach can also be important tools for existing and would-be users of specialty software.

Credentialing Organizations and Professionals

Many technical and professional organizations outside of the hydroanalytic CoP provide user certifications and/or credentialing. Consider the Project Management Institute, which provides training courses for project managers that lead to certifications such as the Program Management Professional (PMP). Other organizations such as the American Academy of Water Resource Engineers inducts experienced and licensed water resource engineers who qualify and continue to take training, the opportunity to be certified as a Diplomate Water Resource Engineer. The Association of State Floodplain Managers offers a certification program for Certified Floodplain Managers (CFM). These certifications, to name a few, add value to the individual's own credentials and can help propel their careers. Yet, this study could not find evidence of any certification or certifying organization for the broad category of hydroanalytic modeling professionals. This may be due in part to difficulty in developing a process that would accommodate such a highly variable array of hydroanalytic tools and the highly variable level of skills needed to apply them.

There were a few examples of hydroanalytic related categorical certifications and software-specific user certifications. The International Society of Catastrophe Managers (ISCM) provides two credentials, Certified Specialist in Catastrophe Risk (CSCR) and Certified Catastrophe Risk Management Professional (CCRMP) that are targeted to insurance catastrophe risk professionals. The curriculum for these certifications covers topics from insurance fundamentals and ethics to the basics of catastrophe modeling and is not tied to a specific vendor or software. Like other certifications, fees and testing are applied to various phases of the credentialing process. Deltares, as a software-specific example, offers an organizational certification program for Delft3D, their three-dimensional hydrodynamic and sediment transport/morphology code. They charge a fee for the initial certification and for renewals. These examples demonstrate that there is an interest by individuals, consultants and organizations seeking grants and/or contracting funds to go through the requirements and pay for a credential that distinguishes themselves from their competition.

Taking software credentialing a step further, organizations and agencies that rely heavily on external model providers may consider incorporating a requirement in their requests for proposals that successful applicants meet certain proficiency and training requirements. For modeling activities in the Netherlands, the use of approved software by certified organizations is encouraged. While Quality Management Plans can provide some reassurances for procuring hydroanalytic products, credentialing/certificating organizations and/or professionals would be a more explicit way to assure quality.

During this study, professionals who both developed and used models expressed their concerns that the quality of modeling results, regardless of the quality of the model, were explicitly tied to the capability and proficiency of the modeler. As noted, there is a void and few examples of credentialing individuals and/or organizations that apply specifically to hydroanalytic tools. As data and software continue their move to open-source and shared platforms, there is a growing need that the credibility of those who apply the models and generate data be validated.

TECHNICAL REQUIREMENTS

Many good practices for technically evaluating and managing hydroanalytic information and other derivative products are presented. The practices ranged from qualitative criteria, such as documentation, to robust quantitative processes such as VVUQ. While evaluation of hydroanalytic information varied across the CoP, there are clearly some best practices that could provide a basis for more consistent standards and applications. Software and data management practices, qualitative and quantitative criteria and hybrid approaches to understanding product quality are discussed with examples.

Software and Data Management

Fundamental practices that were broadly accepted among software and data managers include sharing codes and data, accepting and implementing data and model standards, and managing and curating data and models with attention to metadata and version control.

Open Source/Open Data

The advantages of sharing code were explored and described by an early adopter, Eric Raymond. In an essay and later a book, *The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary*, Raymond describes the advantages of having many people working on a code versus the traditional top-down approach of restricting access to just a few code developers (Raymond 1999). Code sharing provides an opportunity to have an expanded community of experts add value to the model or algorithms within the model. Having open access to source code and open data helps transfer, share and advance the modeling technology and the data.

The practice of using open source or community modeling was observed in several successful organizations and academic institutes. The community nature of the products allows the products to improve and evolve and increases transparency. For some organizations, the positive public image of sharing code offset any minimal economic gains to sell license or protect intellectual property. While some federal agencies still embrace proprietary code or only provide executable software, the NWS supports sharing code on many of the products such as WRF-hydro and the National Water Model. Academic institutes such as the University of North Carolina and the University of Iowa rely on open sharing of hydroanalytic code to give faculty, students and research collaborators an opportunity to make advancements in their areas of expertise and contribute to the body of knowledge[3].

Lack of transparency and/or proprietary codes can mask faulty algorithms and limit rigorous peer review. As examples, with the introduction of AI and ML in analyzing disaster risk, several companies have come under scrutiny recently for misrepresenting exposure and impacts. A 2018 article in the Insurance Journal cautions that many of the models used by insurers for flooding result in widely variable estimates of average annual losses (AAL). An evaluation of three different commercial models gave AAL estimates of a specific ocean front property that ranged from $30 to $20,000 (Howard 2018).

Like software, open data provides a fundamental basis for developing hydroanalytic information. In 2013, President Obama, through the Office of Science and Technology Policy, issued an executive order making open and machine-readable the new default for government data (Holdren 2013). The open government data act was codified in law on January 14, 2019, as the *Foundations for Evidence-Based Policymaking Act* (Public Law 2019). In compliance with the executive order, this triggered various agencies to make their data available to the public. As a recent result of this act, FEMA set up a website to access their NFIP flood insurance claim data in machine-readable formats. The release of these data has already triggered new studies and produced findings that can help reduce the gap in flood insurance coverage, as noted by the First Street Foundation, *The Power of Data Transparency* (Costello 2019). In turn, First Street Foundation continues sharing the data and results through a software interface or application programming interface (API).

The National Weather Service and Global Positioning Systems have long practiced an open data policy that has led to great advances in navigation and forecasting systems by other public and private-sector users. NOAA requires that its data and products can be used by other governmental agencies, the private sector, the public and global community while also being compliant with certain technical specification and privacy provision. They state (NOAA 2007):

> *Information will comply with recognized standards, formats, and metadata descriptions to ensure data from different observing platforms, databases, and models can be integrated and used by all interested parties.*

As new analytic techniques become more mainstream, it is increasingly important that organizations open their data and software for review. Not only is open-source desirable, it is being mandated as federal practice by DoD, DHS and others where security is not compromised. Sharing codes and algorithms among advanced researchers opens the door to transparency and transdisciplinary approaches to solving the next generation problems. This does not preclude the need, and even makes it more imperative, to have rigorous quality assurance/quality control, licensing and version control and regular verification and validation processes in place.

Computer Code and Code Development

In discussion with modelers, code languages seemed largely tied to application and the preferences and skills of the code developer. There are many software preferences in the hydroanalytic community. As examples, FORTRAN, JAVA, C++ were used for more complex coding such as hydrodynamics and early warning systems. Python and Matlab were primarily used as statistical and data analysis tools. R is a popular open source programming tool for new development or basic use. The use of multiple languages by an organization in developing new software is not an issue, but some organizations chose to migrate the languages to a common code when software was transitioned to an operational platform so it could be managed and updated by skilled software engineers for better product control and efficiency.

To manage code development and workflow, organizations use various methods. Deltares uses a sophisticated flow management process centered around the user that includes software packages that monitor version control, manage issues, conduct code review and automate build and test (Baart 2018).

HR Wallingford developed pyxis to "provide our expert numerical modeling teams with a more efficient way to manage and audit models." Pyxis, a community model, is a graphical user interface with built-in quality assurance that runs on MS Windows and is used to manage workflow (HR Wallingford 2019).

Software may need to be tailored to meet the needs of a specific project or client. When special modeling products are needed for operational use, the software team at Deltares employs a software development process used in Agile Project management practice called SCRUM (Stackify 2017). This process helps fast track software design by breaking down work items and iterating on progress. The product owner or client is a key player in the development. Waterfall methodology is also a technique used to develop and design software systems. It follows a more traditional linear method of requirements, design, implementation, verification and maintenance.

Data Standards

Numerous data standards exist for hydroanalytic and geospatial data. For example, the Corps uses (Hierarchical Data Format) HDF5®, a binary file format originally created by the National Center for Supercomputing Applications (NCSA), University of Illinois, to store large amounts of numerical data. Another commonly used format for meteorology and climatology data is NetCDF, developed by the University Corporation for Atmospheric Research (UCAR). Standards are also set by national organizations such as US Geologic Survey (USGS) and the National Institute of Standards and Technology (NIST). Internationally, the Committee on Data of the International Council for Science (CODATA) advances open science through better availability of data. CODATA recognizes that integration of interdisciplinary data will require common data formats. Standard data formats help promote data sharing and collaboration and streamline management and use.

Data quality is as important as access and standard formats. An excellent example of quality assurance/ quality control of real-time streamflow observations is found in an article in the NHWC Transmission (Bushnell 2018). Color codes are used to describe whether data have passed critical real-time quality control tests and are deemed adequate to use as preliminary data. Data providers and users can see if data are flagged that have not been QC-tested or if the information on quality is not available or highly suspect. More will be discussed on quality later.

Management and Curation

Lord, et. al. (2004) described the importance of keeping data, which requires a resource commitment. This involves three activities: curation, archiving and preservation. Curation is the active management of data from creation through continuous updating. Archiving assures data is properly selected and stored and is accessible. Preservation is a subset of archiving to assure certain data can still be accessed even as technology changes. A common problem with hydroanalytic and/or environmental research data, even if published, has been that it is not electronically curated for easy access by future users. Also, since hydrodynamic codes can consume and generate huge volumes of data it is resource-intensive to manage. However, investing in data curation could save money in the long run by eliminating redundant data collection or costly model runs and providing accessibility to rich datasets for new analytic approaches. Determining best strategies for keeping, archiving or sun-setting data and models will be important in modernizing hydroanalytic information.

The American Meteorological Society recently issued a statement calling for the community of meteorology to adopt best practices in data management (BAM 2019). They make a case for a standard set of principles due to the growing complexity and increasing volume of observations and model data. In other data communities, data management and stewardship has led to the development of the FAIR guiding principles for scientific data: findable, accessible, interoperable and reusable (Wilkinson 2016). And as an example of good practice, NOAA requires all the organizational elements to submit a data management plan (NOAA 2015). This is also an essential requirement for many federal and academic grants.

Qualitative Criteria

When selecting or accepting models and software, the literature review uncovered a heavy reliance on more qualitative rather than quantitative assessments. Common qualitative requirements included software and data accessibility, graphical user-interfaces and visualization, documentation and training and technical support. Peer review was considered a desirable requirement for many review processes as were code reliability and quality and the skill set of the modelers. A history of successful applications and general acceptance by the respective CoP added credibility to the product when being considered or accepted for use. Fitness-of-use (how well-suited the model was for its application) was important, but acceptable levels of accuracy were not often specified to assure the model could reproduce the processes of interest. Physics- or process-based methodologies were generally deemed better approaches more often than data-driven results because of the acceptance of the science behind the modeling approach.

Qualitative criteria included, but were not limited to the following assessments:

- Is the software/data accessible (open-source, freeware, ease-of-use)?
- How current is the software?
- Is there a user's manual?
- Does it have a Graphical User Interface or visualization to simply access and use of information?
- Does the software owner provide technical support?
- Is training available?
- Is there documentation that the code and algorithms have been validated or verified?
- Has the product been peer reviewed?
- Are there demonstrated use cases that align with application (fitness-of-use)?
- Does the provider have a good reputation and demonstrated success in applying the software?

These questions can form a simple binary checklist or made more rigorous by rating the quality of each requirement. Sub-criteria or more specific categorizations can add more granularity to a qualitative assessment. As an example of categorizations, Deltares in the development of their Delta Model set requirements for functionality, usability, consistency, acceptance, and accuracy (Ruijgh 2015). Additionally, expert elicitation can be used to strengthen the evaluations and provide credibility to rating the products.

The U.S. Army Corps of Engineers (USACE), Hydrologic, Hydraulic and Coastal Community of Practice (HH&C CoP) maintains a software inventory of approximately 70 tools, data and models (TMDs) that are considered approved or allowed for USACE use. The inventory is evaluated based on a binary review of some of the above criteria — whether a software has a Graphical User Interface, technical support, a user's manual, training and documentation of validation and verification. The location and management of the TMDs is largely decentralized — most of them at the Coastal and Hydraulics Laboratory with others at the Hydrologic Engineering Center and the Environmental Laboratory. For continuity in evaluation, the HH&C inventory list is guided by the Enterprise Standard (ES) 08101, Software Validation for the Hydrology, Hydraulics and Coastal Community of Practice (USACE June 11, 2011) and establishes the process to validate software for use in planning and engineering studies. It defines model, software, software certification and validation as follows:

> ***Model.*** *An application or implementation of a piece of software created for a specific purpose.The construction and implementation of a model usually requires an Agency Technical Review (ATR) and at times an Independent External Peer Review (IEPR) to ensure the native software was implemented appropriately. The Planning ECs (i.e., 1105-2-407, 1105-2-412) define planning models as any models and analytical tools that planners use to define water resources management problems and opportunities. Their definition suggests software rather than models (see Software definition).*
>
> ***Software.*** *The native code (programs) that informs a computer what to do and the associated user interface that allows the required and optional parameters to be introduced, e.g. HEC-RAS. Software is certified or validated; models receive an ATR to ensure the software is implemented correctly.*
>
> ***Software Certification.*** *A corporate determination that a piece of software is a technically and theoretically sound and functional tool that can be applied during the planning, engineering and design process by knowledgeable and trained staff for purposes consistent with the software's purposes and limitations. Certification implies that a prescribed and detailed set of rules/procedures were followed and documented to ensure that the software is technically and theoretically sound.*
>
> ***Software Validation.*** *A corporate determination that a piece of software is a technically and theoretically sound and functional tool that can be applied during the planning, engineering and design process by knowledgeable and trained staff for purposes consistent with the software's purposes and limitations. Validation implies that a prescribed set of rules/procedures were followed and documented to ensure that the software is technically and theoretically sound. This process is less rigorous than software certification and relies on the experience of a team of experts to make the validation determination.*

In large agencies like the USACE, the standardization of process can be complicated by the need to address requirements for various functional areas. In the USACE, the various functional areas – planning, engineering and construction and operations and maintenance, have led to separate guidance documents. The Planning CoP for instance has an approval process for assuring the quality of a planning model that has a different purpose and uses a different review process (USACE 2012) from the HH&C SOP. While it is important to distinguish technical requirements based on use, it is often the same models and software that are being used. According to a conversation with members of several CoPs within the USACE, a more unified process or umbrella approach is under consideration. The processes and steps for review encourage a standardized approach within each CoP or functional area, but ultimately, the final approvals are based largely on qualitative criteria and internal agency review. This is not to say that technical soundness is not important. In fact, the main criterion for approval or certification of planning models is technical soundness, here described as “the ability of the model to represent or simulate the processes and/or function it is intended to represent” (USACE March 31, 2011).

FEMA’s list of accepted models, as stated earlier, are provided on their website, but a disclaimer states they have not been individually evaluated by FEMA for their technical soundness. In discussions with FEMA model reviewers, models used by their contractors and accepted for use by FEMA are generally reviewed based on their suitability to develop flood insurance rate maps, robust documentation, demonstrated validity of datasets, validation to known datasets, use by others and successful applications. They also consider and prefer skilled practitioners and peer-reviewed models. They expect the narrative provided to them for review to explain why the chosen model is best for the application and how the model processes best support the problem. FEMA also requires product providers to meet certain guidelines and standards (FEMA 2014). However, at the end-of-the day, FEMA’s acceptance of a computer program used for hydrologic or hydraulic analyses is ultimately driven by the Code of Federal Regulations which states under 44 CFR 65.6 (a)(6), the computer program must meet all of the following criteria:

(i) It must have been reviewed and accepted by a governmental agency responsible for the implementation of programs for flood control and/or the regulation of flood plain lands. For computer programs adopted by non-Federal agencies, certification by a responsible agency official must be provided which states that the program has been reviewed, tested, and accepted by that agency for purposes of design of flood control structures or flood plain land use regulation.

(ii) It must be well-documented including source codes and user’s manuals.

(iii) It must be available to FEMA and all present and future parties impacted by flood insurance mapping developed or amended through the use of the program. For programs not generally available from a Federal agency, the source code and user’s manuals must be sent to FEMA free of charge, with fully-documented permission from the owner that FEMA may release the code and user’s manuals to such impacted parties.

The organizations reviewed had variations in the consistency of their evaluation processes whether based on qualitative or quantitative criteria. Some organizations have vetting processes that use consistent criteria for every model evaluation, some have evaluation practices that vary according to functional areas and others implicitly build criteria into their peer review process. Many of these vetting processes depend on the review of the technical information provided by the software/model developer but often do not specify quantitative requirements.

Quantitative Criteria

There are many good practices that use quantitative criteria and techniques, particularly when testing the validity of the software or the accuracy of the results. Benchmark testing, test beds, uncertainty quantification, statistical analysis and product scoring represent some of the more accepted practices.

Benchmarking

Organizations known for their hydroanalytic models consider quantitative evaluation strategies important to the credibility of the modeling results. Benchmarking models — that is, evaluating the performance of the software or its outputs against some standard — is commonly practiced among the hydroanalytic CoP. However, the practice varies in consistency, complexity and scope. Historically a modeler might verify a model’s computational algorithms against known analytic solutions or datasets, calibrate model parameters or coefficients to a specific known event and then validate the model under different physical scenario(s) or event(s). For purposes of this report, verification is a test performed to determine if the model fits its mathematical or analytical description. Validation implies the model accurately represents a real-world application.

The USACE Hydrologic Engineering Center (HEC) completed a verification and validation (VV) study of their most widely used models, HEC-RAS 1D and 2D unsteady flow models (Brunner 2018). In the first verification, the process compared hand calculated water surface profiles to results from the HEC-RAS model for a simplified case. The verification process continued comparing HEC-RAS outputs to more sophisticated analytic solutions. The model was then validated to real world datasets. The comparative differences in these computed water surface elevations to observations can vary depending upon the complexity of the natural conditions. Due to the aleatory nature of the problem, the models do not exactly replicate nature, and therefore assumptions must be made which often occur in the initial parameterization of variables (calibration), such as friction. Then the process requires two steps: calibration and validation.

At the Coastal and Hydraulics Laboratory, a coastal model, the Coastal Modeling System (CMS), was evaluated using verification, calibration and validation to compare three critical aspects: wave, flow and sediment. The summary report is accompanied by three companion reports that speak to the technical rigor of the verification and validation (VV). The VV of the CMS was based on the processes detailed in the literature and often used in practice. The VV study was conducted to determine if the CMS has the right capabilities, and is mathematically correct, and whether the code is robust and the model solutions consistent. Like the riverine VV, the CMS used analytical and empirical solutions to verify the model, laboratory experiments for the first validation and data from field studies for the second validation. By testing against many different laboratory and field datasets, the models could be tested for their ability to reproduce accurate results and identify the limits of where they were applicable (Demirbilek 2011).

A similar quantification process has been used by the UK Environmental Agency in 2004 and 2013 to benchmark a suite of 1D and 2D hydraulic models. In the 2004 report (Crowder 2004), the benchmarking study was conducted to assure the three approved riverine software packages, ISIS HR Wallingford software, MIKE 11 DHI Software, and HEC-RAS USACE software, appropriately met their agency’s need for flood modeling. They conducted tests for numerical accuracy, software capability and reproducibility. The first two tests largely focused on verification testing of analytical solutions and basic hydraulic features. The reproducibility component of the testing aligned most with validation to real world or observed experimental results for other more complex features. In the benchmarking tests conducted in 2013, thirteen two-dimensional hydraulic codes were tested on their ability to model more complex realities such as flooding a disconnected water body or dam break (Neelz 2013). The primary value of these studies is the independent third-party review of multiple software products. The independence allows model developers and those acquiring or using models to see how the product performs on specific tests and compare with similar models.

Benchmarking is an evidence-based process for helping compare capabilities among different models. It is an important tool for understanding existing models and assessing new ones. In general, however, there are no universal set of bench tests for various categories of hydroanalytic models. Further, test cases often do not represent the real complexity of nature or are not applicable to large scale basins or near-field applications.

Sensitivity Analysis

Benchmarking can require complicated testing, particularly with multi-dimensional modeling tools. As noted, these tests may not quantify uncertainty. Ideally, large numbers of model evaluations would help identify and quantify risk. This can be costly and time-consuming, so another method often used by modelers is to conduct sensitivity analyses. That is, by varying boundary conditions, grid size, time steps or other input parameters, the modeler can test their impact on the most important output variables. This helps identify sources of uncertainty in modeling data and frames the boundaries for the outcomes. Recent studies on coastal flood risks conducted at HR Wallingford propagated the most relevant sources of uncertainty through the modeling chain to determine the impact on the outputs. Using sensitivity analysis, they found the biggest source of uncertainty on predictive flood elevations was the wave overtopping component of the modeling chain.[4]

Test Beds

Observational and experimental data are critical components for developing, verifying and validating algorithms and models. The use of test beds that provide those observations and allow for experimentation can provide a platform where algorithms and theories can be replicated and go through rigorous testing. The National Weather Service has long valued this approach as evidenced in their Hydrometeorology Testbed (HMT) dating back to 2009. An expanding set of sensors and meteorology stations that monitor physical processes has allowed scientists to test algorithms and models that improve forecasts of extreme precipitation. Likewise, the Coastal Model Test Bed (CMTB) is executed with data from CHL's Field Research Facility (FRF) in Duck, N.C., and automates evaluation of coastal numerical models. Test bed users have grown organically, bringing their models and algorithms for evaluation with real-time and historic data at the FRF. The software being developed is open source to encourage collaboration and flexibility without additional costs. The collaborative and community-based approach currently only applies to the limited range of data associated with its location on the east coast but could be replicated with other institutions as partners and/or at other locations (Bak 2018). The newly established US Coastal Research Program and its collaborators including the U.S. Naval Research Laboratory, USGS, University of Southern California, and Scripps Institution of Oceanography provide support to this effort.

Test beds are a tool that can be fundamental to the technical evaluation of models and algorithms. Expanding to a broader set of hydroanalytic environments, testbeds could allow users to evaluate their models, improve accuracy and quantify uncertainty. The consortia of collaborators that use the test bed(s) creates a transparent platform that could ultimately lead to certification of models of various kinds (hydrologic, tidal, riverine, etc.). Therefore, the hydroanalytic CoP should continue to explore how to expand and institutionalize the test bed certification process and seek a sustainable funding source.

Verification and Validation and Uncertainty Quantification

Taking benchmarking to a more complex level, a relatively new area of practice has emerged around Validation, Verification and Uncertainty Quantification. The American Society of Mechanical Engineers has a special journal and a standards committee for VVUQ. They define VVUQ as follows (see also definitions below):

> ***Verification*** *is performed to determine if the computational model fits the mathematical description.*
>
> ***Validation*** *is implemented to determine if the model accurately represents the real-world application.*
>
> ***Uncertainty quantification*** *is conducted to determine how variations in the numerical and physical parameters affect simulation outcomes.*

While the ASME standards are not focused on hydroanalytic modeling per se, there are areas related to VV of machine learning and computational fluid dynamics. The VVUQ practices as outlined by ASME also are computationally intensive and designed to decrease the number of costly physical tests (ASME no date).

The National Academies of Science (National Research Council 2012), examined the best practices and scientific gaps for VVUQ considering the large-scale computational simulations of physical processes that are the basis of important and complex decision making. The quantities of interest (QOI) — quantities representing a physical parameter of interest — must be identified and the VVUQ designed to characterize the differences between computed QOIs and true QOIs. In hydrodynamic models this usually means water surface elevation, velocity, and discharge or wave height, period and direction. These variables could also include ice flow, debris flow, sediment transport, morphology change and others. Quantification may demonstrate that a model is excellent for one QOI but poor at others. The paper presents best practices in quantifying uncertainty and best concepts for reducing uncertainty for modeling practices that introduce error such as input conditions, reduced-order models, parameter approximations and ensembles.

In a report by the Naval Post Graduate School (Blais 2008), a tool was under development to build templates that standardize verification, validation and accreditation for DoD models. Schematizations and standards using XML for metadata were proposed. The automation of templates was designed to support a military standard for models and simulation, MIL-STD- 3022 (DoD 2008). The use of DoD standards is not unprecedented in USACE Civil Works R&D; the SHOALS program[5] was guided by MIL-STD-2167A, Defense System Software Development (DoD 1988).

Considering the above examples, VVUQ could provide a robust computational platform for modernization of evaluation techniques, particularly for complex hydroanalytic models.

More on Uncertainty Quantification and Reduction

Benchmarking is designed to test how well software replicates analytic and real-world processes. Sensitivity analysis is another methodology, described above, often used by modelers to understand what variables, inputs and components drive uncertainty. Both can include statistical tests that do not necessarily quantify the uncertainty but can indirectly lead to improving the certainty. This section speaks to the metrics and methods that rate performance or help quantify uncertainty.

Statistical tests such a root mean square error (RMSE) and bias are often used to compare observations with predictions. Using these statistics is a starting point for quantifying uncertainty, allowing a simple way to characterize and compare a suite of models or inform suitability for a certain use. But the tests alone may overlook the uncertainty that is propagated throughout the model chain and can fail to identify sources of errors in the data. A single value test may lead to false confidence in performance.

Testing to characterize performance relative to a baseline is called skill scoring. The Brier's Skill Score, Figure 3 (accessed from the Deutscher Wetterdienst, German Meteorological Service) is a statistical index that compares the model prediction to the actual forecast or relative difference between types of forecasts. A perfect score is 1. The meteorological CoP has for some time used skill scoring in quantifying performance of their weather forecasts. In the U.K., the use of skill scores has expanded to evaluate the performance of coastal morphology numerical models and to strengthen user confidence in these types of hydroanalytic models (Sutherland et al. 2004).

The NWS is also expanding the use of skill scores. They have historically evaluated their meteorological models and predictions using forecast verification metrics. The measurement of these scores is not only a part of the VV process, but also helps track forecasting improvements and is a transparent way to communicate progress.

NWS is now starting to apply this approach to river forecasting. A few of the metrics and definitions are as follows:

Bias. *The difference between the mean of the forecasts and the mean of the observations*

Continuous Ranked Probability Score (CRPS). *Measures the integrated squared difference between the cumulative distribution function of the forecasts and the corresponding cumulative distribution function of the observations. A perfect score is 0.*

Skill Score. *Measure of the relative improvement of the forecast over some benchmark forecast. A perfect score is 1.*

Data science and computational performance and accessibility have enabled more complex statistical approaches to quantifying and reducing uncertainty. For instance, ensemble modeling has long been used in weather and climate forecasting, it is expanding to other hydroanalytic areas such as hydrologic, river forecasting and geomorphologic modeling. An ensemble approach, simply stated, can be the statistical merging of different independent models or assimilation of different datasets. The concept is that by using more than one approach or model, the ensemble will arrive at a better solution. As data sources are assimilated and the model is trained, the uncertainty is reduced.

Probabilistic risk assessments, Monte Carlo analysis and Bayesian approaches are also increasingly being used to understand uncertainty in addressing hydroanalytic problems. Moving from a deterministic to probabilistic approach has helped to communicate and identify uncertainty. For instance, the use of Probabilistic Risk Assessment (PRA) is currently moving mainstream in flood hazard assessments by the Nuclear Regulatory Commission, FEMA, the U.K. Environment Agency and academia. Numerous scholarly papers and studies have been conducted to demonstrate the usefulness in helping identify the dominate sources of uncertainty based on variable model assumptions and data sources.

Overall Quality and Performance

Often a combination of qualitative and quantitative criteria, standards and expert elicitation are applied to evaluate the quality and use of individual models or to apply to a portfolio of modeling products. Identifying critical user and technical requirements (whether qualitative or quantitative) and establishing a process for rating or scoring the criteria results in a hybrid approach to evaluating hydroanalytic information. The following describes key methodologies for taking a hybrid approach to assessing models and model products in the flood and coastal modeling CoP that were reviewed during this study.

Model Review Assessment

Benchmarking tests described earlier by the U.K. tested 13 hydraulic modeling software packages for the ability of models to simulate standard analytic, laboratory or observed datasets. Participation in the U.K. testing was voluntary; therefore, it is likely model developers felt confident that their packages could stand up to the others. The report documented pass or fail on some requirements, showed graphical comparisons of outputs (such as discharge or water level), and recorded runtimes. It did not pick "winners and losers" or seek to develop a composite score or index. By presenting the results, the end-user (in this case U.K. agencies) could make their own decisions whether a package was suitable for, say, large basin-scale inundation or dam failure impacts (Neelz 2013).

In the case of identifying the best models for a specific application, FEMA requested Compass Production and Technical Services Joint Venture to review 26 hydrologic and hydraulic models (some of which were included in the U.K. benchmark study) for their prestorm forecasting capability to improve emergency preparedness and response. Ten evaluation factors were used, weighted according to importance and graded on a five-point scale: simulation runtime, hydrologic input parameters, hydraulic input parameters, output manipulation requirements, ability to use existing 1D and 2D datasets, flexibility of input parameters, replicability, ease of model

$$BS = \frac{1}{N} \sum_{i=1}^{N} (p_i - a_i)^2, (0 \leq BS \leq 1)$$

$$BSS = \frac{BS_c - BS_v}{BS_c}, (BSS \leq 1)$$

Figure 3. Equations for Brier Score (BS) and Brier Skill Score (BSS)
(See website for variable definitions)

modifications, engineering community's familiarity with the model and reliability of the model. The scores for each were rated from 1 (least desired option) to 5 (best option). While scored by knowledgeable experts and informed by literature (such as the U.K. report), the results were subjective in nature, particularly since the standards were not quantitative. The results did allow for identifying likely best options for further validation and testing. A more objective detailed statistical analyses of the top three models then compared the difference in computed and observed water surface elevations (COMPASS 2019).

U.K. Flood Risk Assessment Portfolio Analysis Using a Confidence Index

In support of the U.K.'s National Flood Risk Assessment (NaFRA), a methodology was prepared to quantify uncertainty of both local and national flood assessments. NaFRA is conducted by the Environment Agency using local data and expertise to assess the likelihood of both river and coastal flooding across England and Wales.[6] This assessment in turn helps to inform Long Term Investment Scenarios (LTIS) which then help optimize and prioritize national investments in flood risk reduction over time (United Kingdoom, Environment Agency 2014). The U.K. is in the process of updating their national assessment (NaFRA2) to develop a scalable and less fragmented risk analysis that bridges local and national flood assessments and further reduces their uncertainty.

NaFRA (Sayers 2011) basically assesses and assigns scores for data quality and model performance for various flood typologies and generates an index for each. It takes the methodology several steps further by first quantifying uncertainty and then developing a more qualitative score for confidence in communicating with the public. The methodology is conducted in nine steps briefly described here and found in "Measuring Confidence in NaFRA Outputs" report (Sayer 2011):

Step 1 assigns flood area typologies based on the physical and risk characteristics of flood areas. These can include the location of the primary source of flooding, the general basin characteristics, stream complexity, presence or absence of flood defenses, and complexity of pathways such as in urban areas. **Steps 2 through 5** rate individual scores to various parameters for data quality and model performance respectively and then generate a Data Quality Index (DQI) and a Model Performance Index (MPI). **Step 6** combines the DQI and MPI into a Confidence Index (CI) having a scale of 1 to 25 (where a high score is associated with low quality). **Step 7** quantifies uncertainty in the estimated probability of flooding (based on CI). In **Step 8** a qualitative five-star rating is developed from the CI. This qualitatively rates flood risk assessments using one star as very unlikely to be locally reliable to five stars as highly likely to be locally reliable. Finally, in **Step 9**, the five-star rating is translated to likely uses — national investment planning, community or regional general planning, community or regional detailed planning and individual property choices.

By applying this methodology across the U.K. to evaluate existing mapping and assessment products, this approach could help 1) assess the confidence and quality in local to national products used in the NaFRA2; 2) identify areas of needed investment in mapping/modeling updates; and 3) communicate flood risk and investment needs to the public.

U.K. Coastal Modeling Quality Evaluation and Standards

The coastal modeling and forecasting group of the Environment Agency, U.K., has developed Technical Guidance (TG) for flood modeling standards and an evaluation process that provides a quality score for open coast and estuary flood models (U.K. Environment Agency 2017). Considering computational hydraulic models as assets that should be maintained and improved, the TG establishes a framework that helps to score coastal models based on target national standards. A target standard qualitatively assigns a letter — A – Design, B – Appraisal, C – Strategic or U – Unsatisfactory — for minimal suitability for a specific use. The process evaluates sub-elements in three component areas: source data, flow pathway data and model build. Each coastal/estuarine model is evaluated, and a score is developed that ties back to a target standard. This process was developed to ensure consistent assessment and that standards apply across flood models to generate better flood maps and predictions. It also helps to identify gaps and areas of needed improvement.

FEMA's Coordinated Needs Management Strategy and Quality Standards

FEMA manages a large national portfolio of flood risk products covering approximately 22,000 communities participating in the National Flood Insurance Program and is required to re-evaluate all flood hazard studies every 5 years. A reporting goal used in the National Flood Insurance Program is to maintain an inventory of maps where 80% of the total miles of FEMA flood hazard studies are assessed as having New, Validated, or Updated Engineering (NVUE) miles. (Miles are determined along stream centerlines or coastal shorelines.) NVUE-compliant means that either a new study is underway, following FEMA's guidelines and standards, or that the existing study passes all critical elements and some of the secondary criteria reviewed in the Coordinated Needs Management Strategy (CNMS) (FEMA 2019).

In FEMA's Guidelines and Standards Master Index for flood risk analysis and mapping there are more than 600 standards defining how studies and information are to be obtained, analyzed and curated to meet requirements under the federal code of regulations and according to FEMA's policies. These are largely instructional and only a few set quantitative criteria for items such as minimum vertical and horizontal elevation resolution or geospatial positional accuracy.

Contractors and mapping partners are required to implement these standards and to conduct and resolve any issues in an eight-step quality review. The following steps are taken, and in some cases modified for simplicity from the guidelines (FEMA 2014):

QR1: The draft FIRM database shall be uploaded to the MIP for auto-validation and must pass before QR2 is conducted.

QR2: The preliminary FIRM database shall be uploaded to the MIP for auto-validation and must pass before QR3 is conducted.

QR3: The preliminary FIS Report, FIRM, and SOMA shall be reviewed using standardized checklists after the work has been self-certified as meeting FEMA standards. The FIS Report, SOMA, FIRM and FIRM database shall not be issued as preliminary until written certification is provided indicating that all issues cited at this review were properly addressed and resolved.

QR4: This review validates the Proposed FHD Notice, Appeal Period Docket, and 90-day Start Letter(s). If a 90-day appeal period is required, the proposed flood hazard determination notice information must be entered into the FHD Notices on the Web tool. An approved docket must be received from FEMA prior to the issuance of the 90-day Start Letter(s).

QR5: The FIRM database shall be auto-validated in the MIP and a visual review shall be conducted using standardized checklists to compare the FIRM database to the printed FIRM and all cited issues must be resolved before the LFD will be distributed.

QR6: This review validates the LFD prior to the distribution of the final products. As part of the "Prepare LFD Docket" MIP task, the LFD Summary Sheet/Docket, FEDD Files, and LFD Questionnaire must be prepared and submitted, concurrent with QR5 and QR7. All cited issues must be resolved before the LFD will be distributed.

QR7: The final FIS Report, FIRM and associated paperwork shall be reviewed using standardized checklists before delivery to the MSC and all cited issues must be resolved before the LFD will be distributed.

QR8: A review of the FIS Report, FIRM, MSC paperwork, and delivery manifest shall be conducted by the FEMA Map Service Center using standardized checklists and all cited issues must be resolved before delivery of the final products to the end users.

The Coordinated Needs Management Strategy (CNMS) is the system for collecting and managing attributes, tracking and reporting the official NVUE miles and many of the reporting requirements in the guidelines above. According to FEMA's website:

> *The CNMS validation assessment process evaluates a flood hazard study against 17 factors used to represent possible Physiological, Climatological, and Engineering methodology (PCE) changes that may have occurred since the date the FIRM took effect and its original study date. These may include changes in land use, new/removed bridges and/or culverts, and recent floods. Each study is assigned a validation status — Valid, Unverified, Unknown or Assessed.*

A Valid assignment means the miles meet the NVUE requirements. Unverified is assigned to a study that does not pass the critical and secondary elements of the validation checklists. If information is not available for existing studies resulting in an incomplete evaluation, the status is Unknown. Currently unmapped areas that have been considered for a new study are assigned a status of Assessed. Regardless of status, the flood hazard information on FEMA's Flood Insurance Rate Maps is still regarded as the regulatory standard for floodplain management (FEMA 2019).

The CNMS manages many data attributes relevant to the flood studies. Technical data include information such as the name of the hydraulic or hydrologic model, its accessibility for reuse, validation status and whether it meets certain critical and secondary requirements, such as when updates were made to the maps, changes in gage record or discharge, changes in the stream or basin such as channel improvements, structures and increases in impervious areas. The data provide information to determine and prioritize needed improvements to the mapping. This extensive database considers technical information

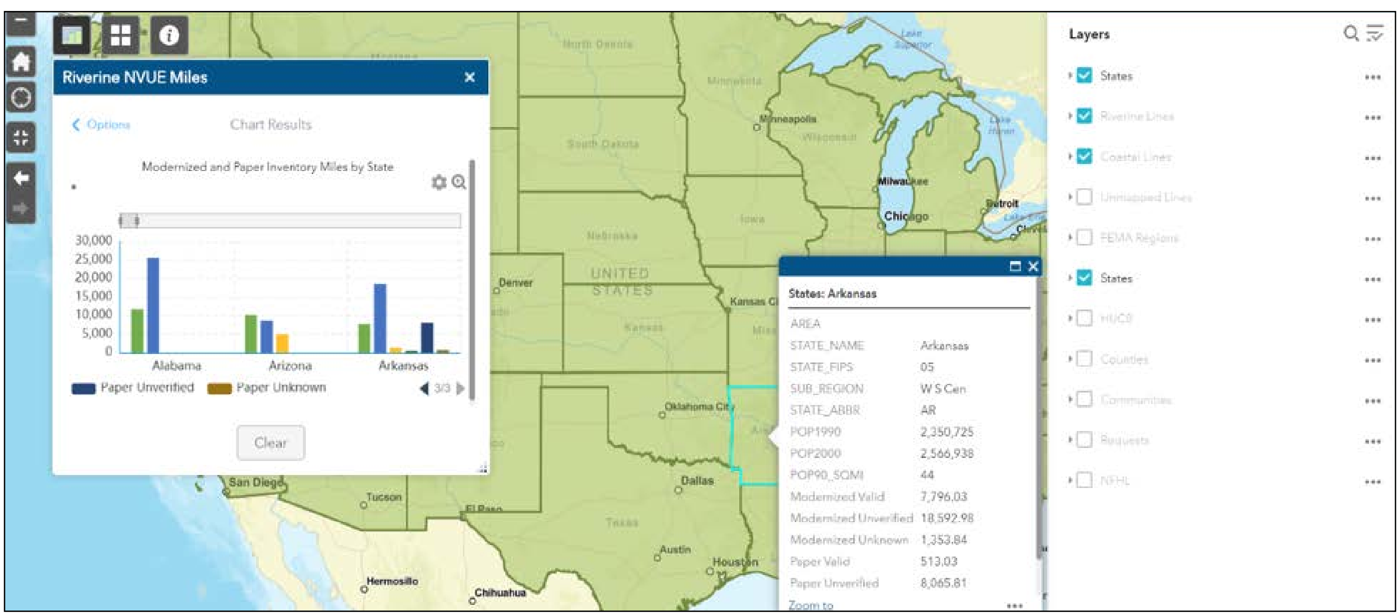


Figure 4. CNMS Viewer (accessed from FEMA's CNMS website.)

but largely in a qualitative manner (ex. Was a detailed study conducted for the area? Was there an addition/removal of a reservoir with more than 50 acre-ft storage per square mile?) and can identify areas of unknown information and trigger next steps or analysis. A viewer is provided (see Figure 4) to review information in the database by several delineations (state, regions, hydrologic basin, etc.).

While the database provides an excellent opportunity to help manage FEMA's mapping portfolio, an audit by the Office of the Inspector General in 2017 stated that the financial management and tracking of flood map projects must improve to achieve its target of 80% NVUE. The report suggested FEMA improve the quality reviews provided by mapping partners and keep life cycle cost estimates for mapping updated to better inform program decisions (DHS-OIG 2017).

As can be seen, there are many excellent practices to help assess and evaluate the technical quality of hydroanalytic information. The practices and methodologies reviewed included rigorous data management and standards, validating, verifying software and models, quantifying uncertainty and managing large portfolios of information. There are numerous gaps particularly in standardization of practice and consistent implementation. The inconsistencies in practice exist not only across the hydroanalytic CoP but within the agencies themselves. Some of these issues could be reduced by implementing good institutional practices.

INSTITUTIONAL GUIDANCE

Policies and Standards

The institutional policies that an organization implements can improve their credibility and transparency while making them more effective and efficient. Best practices in some of these areas became evident while researching modeling strategies.

Good data standards and practices often are the result of good policies. As described above, the U.S. government enacted the Foundations for Evidence-Based Policymaking Act (Public Law 2019), making data open and in machine readable format. Also discussed above, NOAA has established detailed policy requirements for data standards and accessibility of environmental data. In a recent consolidation, alignment and update to its policies, the NWS strategically organized their directives (NWS 2018) under one system for easier access and use.

As federal agencies and organizations improve and implement new data and modeling policies, they should strive to conform to the highest level of practical standards following accepted practices at international and national levels. To be useful and achieve the outcomes intended, the policies must be resourced and enforced.

Evaluation

Progressive, first-class organizations also embrace evaluation. They employ self-evaluation practices such as annual reporting, external boards and special reviews to assure they are keeping pace and meeting the highest level of expectations.

Organizations like Deltares prepare and publish outward facing documents such as their annual report and R&D Highlights annual reports (Deltares 2016) that share their strategic interests and accomplishments. The R&D reports highlight the latest advances in modeling and data and how they are transforming critical mission areas and supporting clients. While these reports are good marketing tools to share with current and potential clients, the retrospective review provides the opportunity to self-evaluate their project portfolio and consider the balance of their investments between strategic research, applied, development, knowledge transfer and consultancy. Further they follow the Protocol for the *Monitoring and Evaluation of Applied Research Organizations in the Netherlands* to review their institution by an independent evaluation committee every four years. The evaluation of research at Deltares and other technical institutions is based on three main criteria: quality, impact and vitality (Deuten et al. 2015).

Similarly, NOAA's Oceanic and Atmospheric Research established a review policy (NOAA-OAR 2007) to evaluate its laboratories on a staggered four-year schedule. Similar to the above protocol, the OAR calls for independent scientific reviews to evaluate quality, relevance and value of R&D to both internal and external interests. Both reviews are designed to inform strategic and budgetary planning priorities.

Special advisory committees to federal agencies provide a higher-level opportunity to impact organizational change. There are numerous examples relevant to modeling portfolios and research. The University Corporation for Atmospheric Research (UCAR) subcommittee, the UCACN Model Advisory Committee (UCAR 2015) provided recommendations to inform NOAA's modeling strategy across NOAA line offices. Key recommendations from UCAR were: reduce complexity of modeling systems by phasing out redundant or obsolete models; use an evidence-driven approach to decision-making and model system development; develop a unified collaborative modeling strategy across NOAA; create a Chief Scientist position to coordinate modeling; leverage external capabilities (academia and private sector); enhance HPC capabilities; develop comprehensive and detailed vision document for predictive capabilities; and execute strategic plans informed by stakeholder requirements.

FEMA's Technical Mapping Advisory Council provides guidance to FEMA on their floodplain mapping products including recommendations for enhanced quality assurance and control (QA/QC). As part of the QA/QC process they want FEMA to have a "proactive process that allows new vendors/ data suppliers, model developers an avenue for evaluation of their products for potential use by FEMA for the purposes of Flood mapping, both regulatory and non-regulatory products" (FEMA 2016). The 2015 report of the TMAC recommended that quantifying accuracy and uncertainty of future conditions was critical for their flood risk products. USACE Federal Advisory Committees, the Coastal Engineering Research Board and the Environmental Advisory Board, provide recommendations that help guide and inform the future of ERDC modeling activities. ERDC also uses programmatic reviews of specific research areas by engaging USACE field office users and external experts. Recommendations from these advisory committees can have broader application to other model and data organizations.

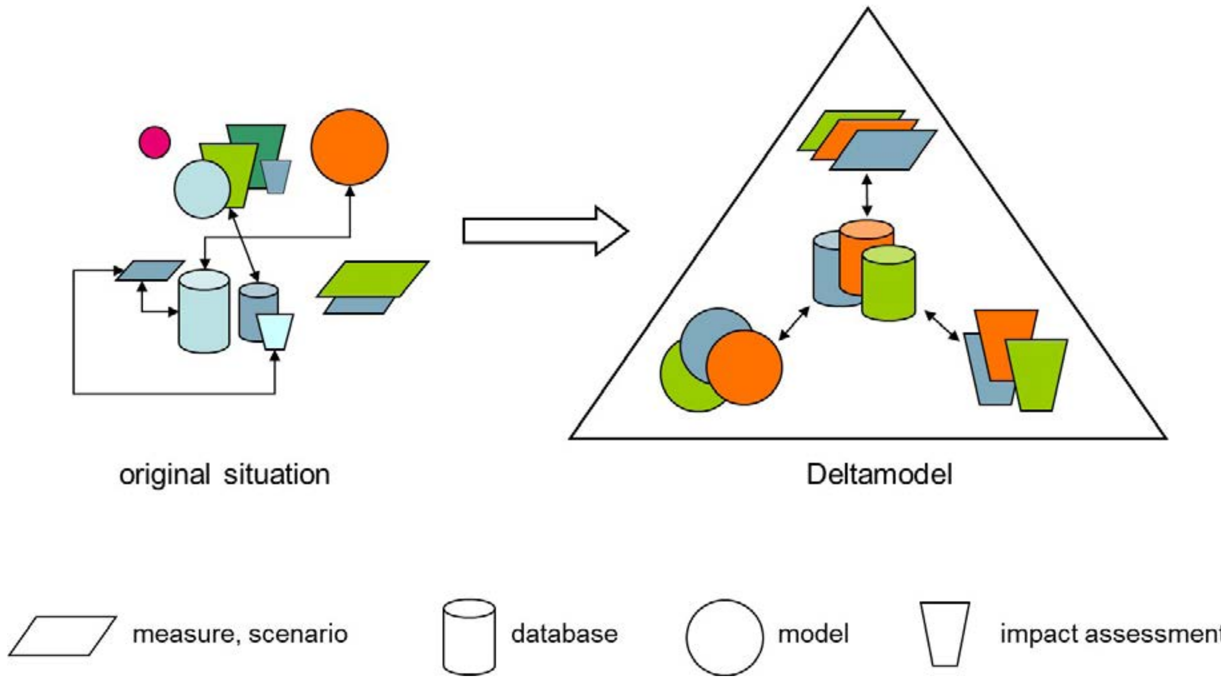


Figure 5. Schematic demonstrating the conceptual framework for the Deltamodel (Ruijgh 2015).

Governance

Culture can be a leading factor on whether an organization is world-class. Culture starts at the top with leadership, but is also driven by clear definitions of authorities, roles and responsibilities. The organizational structure can support the authority structures and ease the path to success. This study did not examine which organizational structures led to better performance. However, several large governmental providers of hydroanalytic information, including the USACE and Deltares, use a matrixed organizational approach whereby technical or engineering specialties support functional mission areas or end-users.

Changing directions in institutional approaches can be driven by governmental directives and national or international programs. As an example, to modernize software in support of the Delta Programme, Deltares developed the Delta Model to have a master model schema for managing water across the country, automating workflow of a range of models and evaluating scenarios for long-term planning (Ruijgh 2015). The model is organized and united under one operational framework, unlike the disparate models that had been used, to assure consistency in its management of water. Figure 5 demonstrates the conceptual framework for consolidation of enterprise software. The Delta model is an example of a framework for building interoperability and better management of the best state-of-the-practice tools.

In discussions with several organizations, it was noted there is often a trade-off between centralized processes and requirements and decentralized execution. However, it was clear that in large organizations where management of the modeling portfolio is splintered across numerous elements, it was difficult to capitalize on collective strength due to stovepipes, competing or conflicting interests and/or simply a lack of awareness of what others were doing. A decentralized approach also led to less-than-optimal investments in sustaining products. Top-down approaches provide the institutional structure and policies to support the whole organization but can also limit creativity that could bubble up from innovation at the ground level. Centralized approval also led to one-size-fits-all modeling paradigms or reduced the portfolio of approved models. Those approaches may be perfectly appropriate for organizations that have limited geographic scope or more focused application needs, but in large geographically dispersed agencies where hydroanalytic information is highly impacted by terrain or meteorology, the portfolio of hydroanalytic approaches must be larger. Finding the balance between top-down control and bottom-up execution is difficult to achieve. A review of more organizations could provide lessons-learned and best approaches.

Business Processes

This study was not designed to analyze business processes, but several ideas for improvement were reviewed.

Portfolio Asset Management

The hydroanalytic modeling and data portfolio of an organization should be managed like any set of real-property assets. The critical elements of the portfolio are the computing platforms and hardware, data storage (inputs and outputs), software and code and importantly, the in-house experts needed to manage the portfolio. Good asset management strategies include maintaining an active inventory of modeling/data assets; conducting periodic assessments of condition and usability of software and hardware; estimating and budgeting for life-cycle costs to include software and hardware operational requirements, upkeep and maintenance; and setting the conditions for replacement, archiving or disposal.

Just like real property assets become aged and/or obsolete, so can the hardware platforms and code that support hydroanalytic information. There are a variety of poor management practices or externalities that can jeopardize the enterprise software and systems that are critical to operational success. For instance, slow bureaucratic acquisition processes that can lead to hardware and software obsolescence before it can be fully operationalized; dependence upon proprietary software that may no longer be supported by a vendor; attempts to "patch" code that cannot keep pace with advancing high performance computing platforms or capabilities; cyber-security requirements that can constrain operating budgets and timelines; and the inability of existing systems to be interoperable with others. These are just a few of the reasons to embrace a life-cycle asset management process.

A more unusual asset element in hydroanalytic tools, data and models is the expert-in-the-loop. There is often a model history that includes a research code or software product that was developed by an expert in the field. As a unique algorithm or a parallel software is generated to test new theories or computational paradigms, advancements are made to the body of knowledge in hydroanalytics. While many of the best algorithms and codes are merged into existing or evolve into new software, considerations need to be made to archive or dispose of software when there is no corporate memory, existing capability or funding stream to support it. Understanding when to discontinue use or dispose of software, data and tools is as important as identifying when to continue investing in the highest priority assets.

For agencies that develop, manage and use hydroanalytic information, an asset management strategy is essential. As a good practice, the USACE HH&C CoP works closely with the ERDC and HEC to review the current set of tools, models and data (TMD) in the hydrologic, hydraulic and coastal portfolio that best meet the requirements they have established for approval to apply on Corps projects. The approved list also informs how model maintenance funds will be allotted to ERDC and HEC to provide continual upkeep and support. The U.K. Environment Agency also manages their portfolio of modeling assets (as described above) to better inform their investments (UK-EA 2017). For agencies that require suppliers and partners to deliver products based on hydroanalytic information, such as FEMA, it is critical that the list of approved models is kept current and is supported by the organizations that develop and manage that asset.

Buying What You Need

Developing software is resource intensive, but there are merits in having internal control and capabilities to improve and use the product. But when appropriate, acquisition or incorporation of models and data developed externally or that include Commercial-off-the-Shelf (COTS) products can be a wise business decision. This can often be counter to the do-it-yourself or independent culture of an organization. Breaking that culture, however, can lead to transformational changes. As an example, the NWS has a strong history of internal development of its own tools but made a corporate decision a decade ago to adopt the Forecast Early Warning System developed and used by the Netherlands for their river forecasting system (Roe et al. date unknown). The NWS brought the key developer to the U.S. and embedded him within the organization to build the system[7] that is now operational across the U.S. and that has formed the basis for further enhancements. Ultimately when determining whether to buy or build, a long-range perspective is needed to assess potential impacts to the organization's operation.

Revenue

Organizations such as ERDC, Deltares, DHI and HR Wallingford have capabilities and facilities that external organizations might be willing to purchase. Large physical facilities have the obvious draw for conducting reimbursable studies and specialized facilities can be used to test equipment or validate new algorithms and models. To encourage use of these facilities, U.S. federal laboratories have unique technology transfer capabilities such as Cooperative Research and Development Agreements (CRADAs) and Testing Service Agreements (TSAs) that can improve their ability to leverage their resources with others.

Credentialing and certifications are also activities that can provide revenue. As discussed above, Deltares offers its software for free, but encourages organizations, through a fee, to be model-certified. Credentialing modelers and organizations could provide credibility to the use of models outside the organizations in which they were developed while also providing market exposure and revenue.

Some companies choose to charge subscription fees for their software products, such as Flo2d. While that approach may not be appropriate for federal agencies, they could consider a subscription service internally to support model maintenance or receive reimbursement for specialized training. Both practices are used by the USACE. Based on criteria set at USACE Headquarters, models that are managed and updated at the ERDC and HEC receive nominal funds for upkeep.

Adapting to Technological Change

There are often technical challenges to modernizing modeling and data strategies. Any strategies will need to consider IT challenges related to cybersecurity, compliance with various privacy laws, maintaining platform resilience and incorporating hardware advances. Adhering to these IT challenges can critically impact cost-effective management strategies and must be included in the development of a life-cycle business strategy and budget. High-performance computing advances and a new generation of satellites and sensors will require that strategies keep pace with the hardware and code that can incorporate an increasing volume of data. New technologies could make existing computational methods obsolete, giving importance to adaptable and flexible organizations.

Branding

Finally, branding is clearly a practice among model competitors. Distinguishable symbols or icons for each of their software models and a clean entry to their web "store" helps make it easier for customers to navigate and assess the value of their products. Like HR Wallingford, The Danish Hydraulic Institute or Deltares, easy access to software and services on a website can draw users. Other ways to market the brand were discussed in building user capability such as software fairs and training, and there are other opportunities such as exhibitions at conferences, presentations and use of social media.

Chapter 4

EMERGING HYDROANALYTICS

World-class problems and transformative technologies are causing revolutionary changes in data and model analytics. Integrating modeling systems, incorporating new analytic methods and including non-traditional data will be necessary to address such areas as the consequences of disasters, the energy-water nexus, adaptation to climate change and managing aging and inadequate infrastructure. Quantifying non-traditional benefits and costs, inclusion of cascading effects, and compound events will add to the complexity. Hydroanalytic modeling approaches are expanding to include not just economic consequences but social behaviors using techniques such as gaming theory, agent-based modeling and multi-agent systems. Organizations around the world are turning to big data, advanced modeling approaches and hybrid or coupled analytics to model complex problems. The following discussion along with a few examples reflect how hydroanalytic modeling is adapting to these challenges.

INCORPORATING AND MERGING TECHNIQUES

Modeling techniques using big data are already entering the field of hydroanalytic analysis. In areas such as the energy-water nexus, and flood and climate modeling, new or multiple techniques are being applied. In a 2018 article on modeling the energy-water nexus (Zaidi et al. 2018), the authors explore the challenges and opportunities of using the vast amount of data that is available in both the water and energy sectors. Using the process-based approaches which often represent the physical systems and the data-driven approaches, they look at the pros and cons of both. While physics-based approaches can be more reliable predictive tools when there is a complete understanding of the system, they are often computationally intensive and subject to miscalibration or rely on estimation of parameters. Data-driven approaches can be used to quantify uncertainty, and can be quicker to develop and used to integrate different process-based approaches. They are, however, data-intensive and subject to errors and misleading results from missing data, disparate sources of spatial and temporal data, heterogeneity and lack of data standards (Zaidi et al. 2018).

Likewise, machine learning methods are increasingly contributing to flood analysis and prediction. The methods are improving the performance of existing models through coupled and hybrid approaches and helping to generate more cost-effective solutions. Mosavi et.al in a 2018 paper, reviewed many approaches in the literature and compared the lead times in forecasting floods. While limited to lead time predictions, they conclude that hybrid approaches using two or more ML methods, data decomposition techniques and ensemble approaches improve predictions (Mosavi et al. 2018). Like physics or process-based approaches, ML/AI techniques must be validated and tested. Authors caution that the ML algorithms are only as good as their training.

To assess consequences, organizations are using numerous methods for coupling natural hazards with socioeconomic impacts. HR Wallingford, for instance, has developed software the Life Safety Model (LSM) that represents not only the flooding hazard but the fate of receptors such as people, vehicles and buildings in the path of the flood. Using an agent-based method, the LSM helps communicate and prepare for emergency scenarios. Similarly, FEMA with its partners use a non-proprietary software called HAZUS to model the consequences of natural hazards. This GIS-based program couples the probability of the hazard with its physical, economic, and social losses. Coupling economic models with hydroanalytic information is not uncommon either. The USACE uses the analytical modeling software, Beach-Fx, that employs event-based simulations combining meteorology, coastal engineering and economics to estimate storm damages and identify cost effective shore protection alternatives.

Advances in hydroanalytics in disaster management may mean simplifying existing methods, or conversely, expanding the level of details. End-user requirements, as discussed earlier, may trade speed for accuracy. For instance, a flood depth regression methodology was developed to provide a cost effective and rapid prediction of flood extents (Longenecker et al. 2019). This method is designed to make it easier for non-modelers to use and for emergency managers who need information fast. Other methods, such as two-dimensional surge models, added complexity to help reduce uncertainties.

Moving from deterministic approaches to probabilistic risk assessments allows for a way to look at the probably of plausible outcomes over a range parameter estimates. The use of Probabilistic Risk Assessment (PRA) is currently being applied by the Nuclear Regulatory Commission, FEMA, the U.K. Environment Agency and others. Even with PRA, there is still much to learn about uncertainty. Variations in input parameters such, as Manning's n for the hydraulic models, sources and level of accuracy of available data, such as elevation data, or the algorithm selected to assess impacts, such as depth-damage curves, can influence the overall results. The importance of accurate elevation data was described above in the paper by Kulp and Strauss (2019) that demonstrated a three-fold underestimation of the impacts of sea level rise. Another paper demonstrates the high variability in results when comparing model assumptions for probabilistic flood risk modeling (Winter et al. 2017). In this paper, five aspects of model assumptions were compared to a reference simulation. It identified a large variability in accuracy, particularly when selecting a damage function. Numerous scholarly papers and studies have been conducted to demonstrate the usefulness in helping identify the predominant sources of uncertainty based on variable model assumptions and data sources.

Insurance and catastrophic modeling organizations often use data analytics coupled with physics-based modeling to estimate their risk exposure. Improving modeling capabilities in this area will be critical for providing better policies for those at risk. Flood risk modeling is a primary area of hydroanalytic modeling for some insurers and reinsurers. Catastrophic models, which often employ hybrid approaches, are advancing through many companies like AIR, KatRisk,

Satellite image of Hurricane Matthew, October 4, 2016, shortly after it made landfall on Haiti as a Category 4 storm

RMS, CoreLogic and others and could plausibly leap-frog traditional approaches to physics-based flood risk models. An alternative approach would be to use catastrophe modeling to provide insight and help drive improvements in more traditional hydroanalytic modeling. FEMA is currently exploring catastrophic modeling capabilities for the National Flood Insurance Program.

An article by the founder of Jupiter Intelligence, Rich Sorkin (Sorkin 2018), offers that traditional stationarity-based risk models are not useful to mitigate risks to specific assets nor aid in design. Employing dynamic models that use big data analytics and flexible computational architecture, Jupiter Intelligence's proprietary software combines two modules, FloodScore and ClimateScore, to predict real-time flooding for operations and for long-range planning scenarios at the asset level (Jupiter Intelligence 2018).

CONTINENTAL AND LARGE REGIONAL SCALE MODELING

The above demonstrate techniques and emerging concepts that are influencing hydroanalytic methodologies. The following are examples of how hydroanalytic modeling is expanding to cover large geographic domains. In the fields of climate and atmospheric science this is not a new concept, but it is an evolving area for flood, coastal and ecosystem modeling.

NOAA operates a high-resolution National Water Model (NWM). The model, like other products NWS has developed and operates, is transforming water information services and streamflow forecasting. The model mathematically represents different physical processes to better integrate water predictive capabilities across the entire country. It uses meteorological forcing data to make best use of thousands of gage and radar precipitation observations. The simulations are updated hourly. Surface routing is conducted on a 250m grid and incorporates 2.7 million stream reaches. This real-time analysis of stream flow and other surface and near-surface hydrologic information can provide short range forecasts from hourly to 18 hours. Medium-range forecasts, updated four times daily, can extend out to 10 days and long-range produces a 30-day, 16-member ensemble forecast. The NWM is providing information that complements the current NWS river forecasts at more than 4,000 locations in the continental U.S. and provides information at locations where traditional forecasts are not available (NOAA-NWS no date).

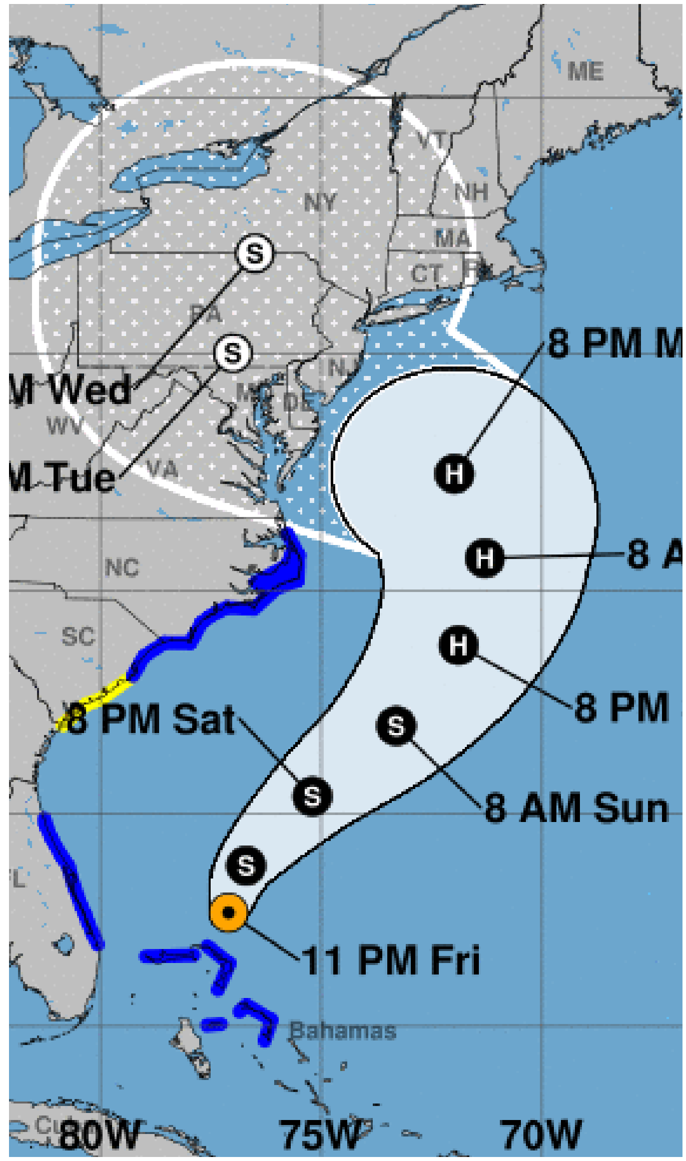
In June 2020, the First Street Foundation released a continental-scale analysis of flood risk in the U.S. The model provides a national risk assessment that includes fluvial, pluvial and coastal flooding. Additionally it calculates a flood score risk for individual properties, FloodFactor.com. The analysis is driven by a comprehensive flood exposure model for the entire contiguous U.S. developed by Fathom-US. A peer reviewed article describes the development of a 30m resolution flood hazard model and its validation to existing FEMA flood maps (Wing et al. 2017). It uses a 2D hydrodynamic model based on only publicly available data. It compares outputs to higher quality local maps and USGS data. The goal was to be able to simulate within reasonable accuracy and at lower cost, flood risk in areas of the country not previously mapped or studied. According to their website, Fathom hopes that their model can provide advanced flood risk and exposure information at competitive prices to help support uses such as insurance ratings. This model is not officially used by NWS or FEMA but demonstrates an opportunity to rethink flood risk analytics.

Through a mandated requirement after Hurricane Katrina, the state of Louisiana began the hard work of developing a comprehensive coastal master plan for its vulnerable and disappearing coastal areas. With each iteration of the plan (the most current is 2017) the large-scale regional modeling to support multibillion-dollar alternatives has evolved into a sophisticated and integrated suite that includes hydrodynamics, geomorphology, ecosystems, economics and planning (Coastal Protection and Restoration Authority 2017). While not fully integrated, the effort represents one of the largest scale coastal and riverine watershed modeling efforts including multi-disciplinary science, economic consequences and the inclusion of sociocultural systems. Its planning tool helps to merge across modeling disciplines and is being used by decision makers to evaluate engineering and policy alternatives related to ecosystem restoration and flood risk reduction.

ENSEMBLE MODELS

Ensemble modeling is no longer just being used in weather and climate forecasting; it is expanding to other hydroanalytic methods such as hydrologic, river forecasting and geomorphologic modeling. As stated before, an ensemble approach can statistically merge different independent models or assimilate different datasets. By using more than one approach or model, ensembles can improve our understanding of a problem and help us make better decisions. Hurricane forecasting models have long used ensemble modeling approaches, generating the often shown "spaghetti model" plots. The dynamic models solve governing equations of the atmosphere on a global scale. The NWS National Hurricane Center typically builds a consensus model that closely resembles the average track, of the top six global models, to use as its forecast. Even when the consensus model is out-performed by some of the better ones- ECMWF, GFS and GFDL- it still has a good track record (Masters 2019). Ensembles and consensus approaches are gaining traction in the world of hydrologic forecasting and hydraulic modeling.

As directed by the National Research Council with the goal of better-informed water decisions, the NWS began development of a hydrologic ensemble forecast service (HEFS) for their River Forecasting Centers (Demargne 2013). The system incorporates climatological and weather forecasts from downscaled models into the system while adjusting for bias and calculating uncertainty. To create the ensemble flow forecast requires close cooperation with the National Center for Environmental Predictions at NWS to hindcast significant amounts of historical data in the (Global Ensemble Forecast Service) GEFS model to verify HEFS. To minimize disruptions in operations, NWS has used a careful and deliberate process to move towards operationalizing the HEFS. Using ensemble data from the meteorological models and ingesting them into the hydrologic models helps



Track forecast cone for Hurricane Sandy beginning October 26, 2012

to quantify and propagate uncertainty. This model will transform operations providing river forecasts from one hour to one year. The model was developed to help the state of New York manage its reservoirs by producing longer range forecasts and quantifying uncertainty and is now transitioning to the NWS River Forecast Centers. The significance of HEFS lies in its ability to quantify uncertainty using an ensemble approach while also reducing the subjective nature of the forecasts in the RFS (Wells 2018).

The usefulness of ensemble modeling has also been demonstrated in hydraulic modeling. Zarzar et al. (2018) explain how it can be done using two H&H models, iRIC and HEC-RAS. These models that are typically used in a deterministic mode are calibrated to historical events, and then using 11 ensemble precipitation forecasts, generate 11 streamflow forecasts. Results are compared and agreement identified using an app to view the flood threat.Ensemble approaches can help quantify and reduce uncertainties and communicate those uncertainties to decision makers.

Chapter 5

BUILDING BLOCKS TO IMPLEMENTATION

Many existing concepts and practices for evaluation have been reviewed for this study. The synthesis of this information is described in this chapter as building blocks toward implementation. Each building block represents an essential consideration in the development and implementation of a robust hydroanalytic evaluation strategy. The building blocks as described in this final chapter include:

- Synthesis of Best Practices (User, Technical and Institutional)
- The Model Evaluation Cycle (Roles and Responsibilities)
- Product and Portfolio Quality Assessment
- Certification of Modelers and Models

SYNTHESIS OF BEST PRACTICES

As discussed throughout the report a robust evaluation framework must include meaningful user engagement, sound technical requirements and responsible institutional guidance (see Figure 1). A summary of the best practices is provided here.

User Engagement Practices

Best practices in user engagement as shared by model developers and providers include identifying and maintaining an inventory of product users, conducting customer surveys, evaluating fitness-of-use of proposed methodologies prior to conducting studies, maintaining a close relationship between the provider and the client to assure acceptance, providing training and support to the end-user to build capacity and capability, continually learning and reviewing state-of-the-art modeling methodologies to stay current with emerging technologies and, ideally, encouraging certification or credentialing of organizations and people who provide modeling services. Similarly, users and end-clients should be prepared to question methodologies by evaluating fitness-of-use, setting quality requirements and expectations in their requests for services and planning for and being available to participate in the product development to build confidence in accepting the product. These best practices can help both providers and users identify the strengths and weaknesses of the final products, identify opportunities for growth and improvement, and help to prepare for challenges in maintaining credibility and market share.

Technical Practices

The second leg of the evaluation stool is to develop and implement evidence-based technical requirements. Shared data, standards for data collection, management and curation are essential elements to assure quality in data management and analytics. Software should be subject to vigorous verification processes to assure the computational code is free of errors and adequately solves the algorithms. Open source and community shared codes add value to this process by allowing access to more experts who can review and improve the software.

Models should be subject to both qualitative review and quantitative validation. Good practices in qualitative review include documentation of technical soundness (ideally through rigorous validation and verification processes), a user's manual, an accessible user interface, the availability of technical support and training, peer review and the demonstration of applicability through use cases. Good quantitative practices include benchmarking (comparing capabilities among models and to a standard), VVUQ (a more rigorous approach for complex models), testbeds, skill scoring and ensemble approaches. Uncertainties should be quantified whenever possible throughout the modeling chain – input data, at each modeling step, for each quantity of interest, output data/statistics and decision support tools.

Further, the confidence and ability to relatively compare hydroanalytic approaches requires an integrated or hybrid approach to evaluations. Combing qualitative and quantitative methods and setting targets for acceptability based on usage can provide a consistent approach to scoring or rating products. Particularly when applied over a portfolio of hydroanalytic assets, product ratings can help to evaluate fitness-of-use and quality, provide a more robust approach to managing the life cycle and investment strategies of these assets and more readily communicate their quality. Concepts for product and portfolio assessments will be discussed below.

Institutional Practices

Successful organizations that have a credible and sustainable role in hydroanalytics have built a viable institutional infrastructure. Components are good leadership, clearly articulated and enforced policies and procedures, well-defined organizational roles and responsibilities, a culture of evaluation, strong business acumen and most importantly, qualified and enthusiastic people. Starting at the top, leadership is paramount to guiding quality products, customer service, resource management and embracing change. An organizational structure that supports functional missions and ties authority to clearly articulated roles and responsibilities leads to a more efficient and effective delivery of quality services. Institutions should strive to implement a uniform set of policies and standards for data and model management that conform to the highest standards practicable and avoid stove-piped variations on criteria and processes.

As was demonstrated through numerous examples, evaluation practices are a cornerstone for most hydroanalytic organizations. These include internal and external organizational, administrative, and/or product reviews conducted on a recurring basis (such as laboratory or program area reviews), under a special advisory committee (such as the CERB or TMAC) and/or adhering to certain criteria (such as quality, impact and vitality).

USER SPECS	SOFTWARE DEVELOPMENT	MODEL APPLICATION	DECISION SUPPORT	CURATION AND MANAGEMENT
• Acceptance (U) • Speed (T) • Accuracy (U) • User Capacity (U) • Adheres to Laws (I) • Fitness-of-use (T,U) • Ease-of-use (T)	• SCRUM (U,T) • Test Beds (U) • Version Control (I) • Open Source (I) • Peer Review (T)	• VVUQ (T) • Benchmark (T) • Data Standards (T) • Model Performance (T,U) • Skill Score (T) • Accuracy (T,U)	• Visualization (U) • Understandable (U) • Timeliness (U) • Credibility (T) • Operational Acceptance (I) • Confidence in Results (T,U)	• Documentation (I,T) • Standards (T) • Accessibility (U,T)

Table 1. Examples of good practices in each step of the model cycle. The letters support the evaluation framework legs of the stool: U-User, T-Technical, I-Institutional

To financially manage and sustain a hydroanalytic portfolio, organizations should have good business processes in place. Life-cycle asset management practices should be applied to hydroanalytic assets — software/data, hardware and people — particularly when these assets drive the fundamental decisions of an organization such as investments in Corps water infrastructure projects or execution of financially and nationally impactful programs such as FEMA's National Flood Insurance Program. Adapting to technological changes requires an agile organization that is monitoring the landscape for opportunities (new hardware platforms) and identifying challenges (cybersecurity requirements). Other business processes that can improve the sustainability of an organization and its role in hydroanalytic analysis are the cost-effectiveness of developing in-house versus purchasing software and services, the ability to use the product to leverage or produce revenue and branding its products to assure visibility in the marketplace.

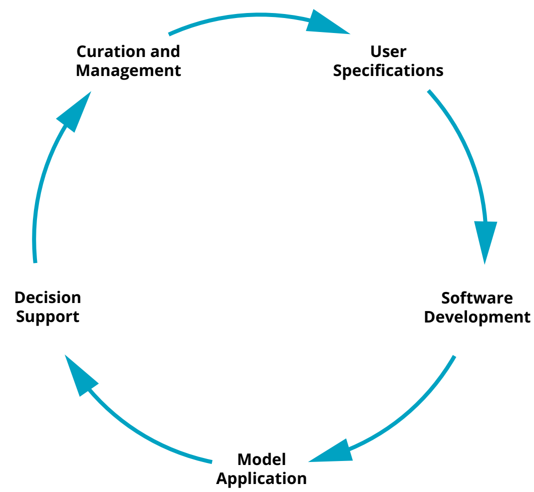


Figure 6. The Model Evaluation Cycle

THE MODEL EVALUATION CYCLE

Though rather simplistic, the model evaluation cycle shown in Figure 6 identifies a series of steps from conception to development to use and curation. Beginning with an initial set of user specifications, a software either needs to be selected, developed or modified; then it must be applied to the specific domain; the results or the model are transferred to the user or decision-maker and ultimately stored or saved for future use or modifications. A second loop around the cycle implies a refresh and may include modifications to the software or a version update, new data inputs to the model or a new domain application, a different set of decisions or refinement of previous decisions and again, documentation of the changes and curation. Importantly, within each step and for each loop, quality should be evaluated, and uncertainty captured.

Examples of best practices for technical requirements (T), user engagement (U) and institutional guidelines (I) in each step of the cycle are provided in the Table 1. These do not reflect all the evaluation processes that could be used but rather some sample activities. It is important that regardless of the step, the processes be identified to assure quality is managed throughout the cycle.

While Table 1 shows examples of what is evaluated in each step of the cycle, Table 2 highlights who is accountable. Assigning someone the responsibility and giving them the authority and resources to execute the evaluation will help deliver a quality product. The responsibility may fall on different people and organizations for each step, or one or two people and/or organizations may execute the entire cycle. But for simplicity in defining roles and responsibilities for evaluation, the responsibilities are delegated to the software developer, the modeler who applies the software and the end-user or decision-maker. Some organizations may have a separate person or division for software and data management and curation; however, each responsible party has a role in that process.

The model evaluation cycle is useful in identifying the requirements and accountability for a single product or project as it moves through steps from inception to completion. A different quality assessment process is required for the evaluation of a portfolio of hydroanalytic products.

PRODUCT AND PORTFOLIO QUALITY ASSESSMENT

Large organizations and agencies often are responsible for a portfolio of products that develop over time and are used for multiple purposes. As part of a portfolio asset management strategy, a process is needed to conduct a condition assessment of the inventory of products. Reviewing and managing quantitative and qualitative data

PRODUCT DEVELOPER	MODELER / APPLICATION USER	END-USER
• Applies necessary Quantitative measures (VVUQ, Benchmarking, test beds, skill scores, computational performance, etc) • Applies Qualitative measures (documentation, user's manual, accessibility, tech support, peer review, use cases) • Engages end-user in development of new software applications • Provides proof of passing required Certification Process • Software curation/version control	• Demonstrates experience using model-use cases • Has proper accessibility and ability to use required Computer Platform • Has Modeling Credentials • Has Education/training appropriate to problem area • Assures technical requirements are satisfied • Manages Data quality and model execution • Captures data and model uncertainties • Curates study applications and data inputs/outputs	• Sets the Acquisition Specs • Accepts end-products and results • Assures Quality is meeting expectations • Manages Quality of Product(s) within Portfolio • Assures Fitness-of-Use (reviews certifications, use-cases and quality) • Sets Performance Standards • Applies results with understanding of limitations and uncertainties • Curates and manages enterprise product data

Table 2. Sample Quality Responsibility Chart

for many parameters, such as described in the NaFRA methodology or in FEMA's CNMS, provides a robust way to track current conditions and identify areas of improvement. As demonstrated by the methodology developed in the U.K. of assigning confidence scores, a scorecard could help assess the confidence and quality in local to national products, identify areas of needed investment in hydroanalytic information, inform programmatic and capital investment strategies and communicate requirements to the public.

While technical criteria can be used to assess the quality of hydroanalytic information, they may not be able to provide information that can influence sociocultural, environmental or economic decisions. For instance, a portfolio of coastal surge models and data can be technically assessed for the speed at which it provides timely forecasts or whether it can provide a probabilistic hazard assessment to design seawalls. Those assessments alone can drive usage and inform investment decisions. However, understanding the characteristics of the ecosystems, built systems and communities that are impacted could be equally or more important and could cause a shift in investments to target a specific geographic or sociocultural regions or alternatively to improve algorithms that reflect a specific causal mechanism. The following chart provides a conceptual perspective of a portfolio of flood risk tools. It expands upon the technical assessment concepts of the NaFRA and CNMS to address community vulnerability and future conditions. This expanded assessment could inform, not only the characterization of flood risk, but other objectives such as investments in community capacity building or development standards.

In the conceptual portfolio evaluation framework in Table 3, three assessment areas are assumed to be critical to informing decisions within an organization that provides flood risk information — Technical Quality Assessment (TQ), Community Vulnerability Assessment (CV) and Future Conditions Assessment (FC). Technical quality clearly drives the validity of the hydroanalytic products. In this example five sample categories of assessment are shown for discussion purposes: Qualitative, Quantitative, DQI, MPI and Usage. All categories may not be needed. Some of the quantitative and qualitative criteria could roll into a data quality or model performance index, similar to the NaFRA methodology. The specific criteria in any of these categories could be assigned a sub-score with each category contributing to a total or weighted composite score. For instance, under Qualitative Criteria sub-categories might have binary criteria such as: Is the code open source? Is there a user's manual? Is the software supported? The quantitative criteria could describe the level of benchmarking or VVUQ and rate its quality based on expert review, or as in the NaFRA methodology, an index could be developed for several contributing variables. The last category in the TQ assessment should clarify best uses for the product.

Increasingly, it has been demonstrated that flooding impacts vulnerable communities more than thriving communities where resources, capacity and capability are limited to prepare, respond, recover and mitigate their flood risk. Therefore, it may be more important to identify where investments in hydroanalytic data can help bring focus and support to more vulnerable communities. The CV includes a number of potential data sources of variables that could be considered. Each portfolio owner would want to develop these to best serve their mission. FEMA has recently developed a National Risk Index. In addition to showing geographic risk, this tool could be used to prioritize or target opportunities to deliver flood risk products to vulnerable communities. If a national risk or social vulnerability index is not available, there are other data sources and proxies to assess community needs and capabilities such as the availability of local colleges or a tax supported engineering department, the history of repetitive flooding or insurance claims. Like the technical assessment, a composite score or index could help prioritize flood product investments, mitigation grants and the need for technical support. It would be a way to provide needed resources and information to communities most at need.

The last assessment area considered in this example is about future conditions (FC). The hydroanalytic information for flood products has probabilistically been driven by historical precipitation and flooding. For certain types of analysis, such as actuarially-based flood insurance premiums, this is an appropriate of these hazard probabilities. However, for floodplain management, long-term capital investments or development, and design of flood risk-reduction projects, it is essential to understand and characterize future flood risk. Hydrologic conditions can be impacted by changes to the runoff characteristics in the basin and by the non-stationarity of precipitation due to climate change. Further, development of the built environment in high-risk areas and other climate change impacts such as sea level rise and extreme events can escalate the consequences. The expected or projected future flood scenario assessment should have analysis and indicators that help inform the need to address these issues. The USACE Engineering Technical Letter for nonstationarities (USACE 2019) and the user guide could help to triage areas in the country that have a higher probability of impact due to non-stationarity in annual discharges based on climate change. The National Urban Change Indicator, a tool developed in collaboration by ESRI and Maxar Technologies, can provide information on persistent human-related and urban changes over the U.S. It could help identify where risk might be increasing due to development. A higher level of investment in assessing future flood risk by using tools such as this will not only improve the quality of the products, but will broaden the use of an organization's flood products to help communities develop building and zoning standards, adaptively design risk reduction alternatives and prepare for flood risk.

The example provided here of potential elements to assess a flood risk tool portfolio is just one approach to consider in developing a portfolio evaluation framework. Assessment areas, categories and scoring in the portfolio framework can all be tailored to the decision requirements of the hydroanalytic portfolio manager. The framework should provide a consistent way of combining multiple criteria (both quantitative and qualitative) over a complex list of parameters to better understand the needs and quality of the inventory.

The evaluation framework should also provide a transparent way to roll up the assessment in a composite product that helps to communicate value. A product-scoring concept or confidence index like that presented here or the one used in the U.K. can help local decision-makers and the public understand the quality and best uses of their current flood products. Table 4 provides a modified version of the NaFRA methodology using a five-star rating to represent qualitatively the best uses of products available. Such a rating process, while easy to understand, could also unintentionally convey winners (5-star) and losers (1-star). Therefore, the derivations of the scoring should be objectively based on the above assessment and transparently explained to those that might use it.

CERTIFICATION OF MODELERS AND MODELS

As was noted earlier, there is a void and few examples of credentialing individuals and/or organizations that apply to hydroanalytic tools. Professionals in the modeling business expressed concern that the use of models by unqualified or marginally qualified individuals and organizations could lead to misleading outputs and poor decision-making. Poorly executed model applications can also impact the personal or organizational reputation of the data or software tool owner, hindering the willingness to openly share data and codes. For the organizations and agencies that rely heavily on external model providers, bad applications can lead to poor public credibility, wasting of resources to make corrections and costly liability issues. To minimize misuse, software developers and owners could offer their own certifications, convene users for training or support and facilitate user-group engagement. Agencies that must acquire model and data products should be specific about the qualifications and proficiency of the modelers in their contracting specifications and requests for proposals. A more holistic approach (not just model-by-model) would be to offer professional credentialing for certain categories of hydroanalytic tools. For instance, a non-profit, professional institute or laboratory could administer the training and develop and evaluate the experience requirements for a certain class of models such as hydrologic, coastal circulation or hydrodynamic models.

While it seemed logical that software should meet some standards to be considered certified for hydroanalytic applications, the reality of doing that is complicated by the various requirements of end use and the lack of standards. As examples, FEMA maintains a list of accepted models to use in developing flood insurance rate maps. The hydroanalytic tools needed for this product may not appropriately translate to some other use, say near-field hydrodynamics to assess fish survivability. The acceptance of models for FEMA is driven by the code of regulations governing the National Flood Insurance Program which allows the acceptance of models used by other federal agencies, such as the USACE. Conversely, the USACE self-certifies its own models for the specific functional use categories within the USACE such as for planning, engineering and operations – not flood insurance rate maps. For many uses, acceptance of a modeling product is ultimately an agreement between the modeling organization and the end-user. None of these approaches fully addresses the best practices captured in this study for consistently and objectively applying both quantitative and qualitative criteria.

As with the credentialing of professionals who apply hydroanalytic tools, there is no third-party organization that currently provides hydroanalytic certification services. But there are also no universal standards or set VVUQ processes agreed upon by the hydroanalytic CoP that could be used to provide that stamp of approval. In the absence of set standards or certifying organizations, Table 5 offers a tiered approach to assuring the selected modeling or data analytic tools meet the needs of an organization and its functional use. The nomenclature of enterprise, preferred and allowed follows the validation categories in USACE guidance (USACE June 2011) where *enterprise* is software mandated or required by the agency, *preferred* is software used by the specific CoP (here HH&C) and *allowed* is niche software good enough for a specific use. Tier 1 would apply to complex software that is operationalized across an organization such as the DeltaModel in the Rijkswaterstaat or the HEFS in NWS. It is imperative that these type models pass stringent quality assurance checks. In Tier 2, experts could be convened to assess new updates to commonly used or previously certified enterprise or preferred models. The third tier represents simpler tools, minor adjustments to existing tools or approval of single use or niche tools.

For instance, the HEC-RAS model is an "enterprise" level tool for the USACE and FEMA. Minor updates by USACE might only require a moderate internal review, Tier 3. However, if a new complex H&H model were to be adopted as an enterprise tool by USACE or FEMA, it should undergo a more thorough process of evaluation, warranting a Tier 1 review.

TECHNICAL QUALITY ASSESSMENT	COMMUNITY VULNERABILITY ASSESSMENT	EXPECTED FUTURE SCENARIO ASSESSMENT
• Qualitative Criteria Index (check list) • Validity/Uncertainty Index (based on Quantitative Benchmarking or VVUQ) • Data Quality Index • Model Performance Index • Usage recommendations	• FEMA National Risk Index (Includes a social vulnerability index) • Proxies for vulnerable communities: population demographics, date of last study, recent disasters, disaster relief funds provided • Properties with repetitive flooding and multiple insurance claims • Insurance penetration/affordability • Community capacity (population size, engineering department, area college or university) • Community users (local college or consultants)	Climate Change • Check for stationarity (USACE ETL 1100-2-3) • SLR and precipitation updated to latest USGRCP climate assessment • Considers heat and drought with flood Economic Development • National Urban Change Indicator • Urban/Rural • Identification of cities rebounding • Areas suitable for sustainable, resilient neighborhoods • Co-benefits/Quality of Life
• **Composite Score**	• **Composite Score**	• **Composite Score**

Table 3. Conceptual Portfolio Evaluation Framework for Flood Risk Management

SCORE	TECHNICAL DESCRIPTOR (TIED TO TECHNICAL QUALITY ASSESSMENT)	EXAMPLE BEST USE
★★★★★	Highest level of granularity and certainty in data and model performance. Includes future projections	Design of infrastructure, urban development planning at site-specific location
★★★★	Very good quality data/models with locally reliable information and future projections	Floodplain management and recovery planning. Detailed risk damage assessments
★★★	Good quality and locally reliable models and data	Insurance rate setting, hazard assessment or community planning
★★	Limited local data. Most analysis based on regional, national or dated information	Risk communication, quick or first tier hazard assessment
★	Regional or continental scale estimates, not reliable for local assessment	Strategy level assessment of general risk and investment needs

Table 4. Example of a Quality-to-Usage Scorecard for flood risk products. (Adapted from Sayer 2011).

LEVEL OF REVIEW	MODEL COMPLEXITY	USE (ENTERPRISE, PREFERRED, ALLOWED, SINGLE USE OR NICHE)	PROCESS FOR EVALUATION
Tier 1 (Extensive)	Highly complex software, software containing new analytical approaches and not previously certified	Mandated or enterprise level software	Third party certification based on qualitative and quantitative criteria
Tier 2 (Moderate)	New algorithms or major update to currently acceptable complex models	Models generally preferred or recommended for use by agency or previously certified	Expert elicitation with review of existing certifications and review of modifications and peer reviewed publications
Tier 3 (Minor)	Standard or previously certified models with minor revisions or expired certifications	Enterprise, Preferred or Allowed	Internal expert review based on documentation

Table 5. Proposed tiered approach to hydroanalytic quality reviews

Chapter 6

NEXT STEPS

This study reviewed and described many strategies for evaluating hydroanalytic information and the organizations that develop and manage them. While there are many methodologies and standards being developed and used in various areas of hydroanalytics, standardization and implementation across the CoP is needed. Codifying current practices and incorporation of new practices in the management of hydroanalytic software, data and models could lead to transparent development of hydroanalytic products, improved lifecycle maintenance, curation and updates of these products, increased credibility among end-users and decision-makers in their application and more consistent contracting specifications for best acquisition strategies. A robust evaluation framework should include the following key steps.

First, recognize that leadership is critical for setting the right example and building the right environment for change from the top down and bottom up. For the U.S., this could mean more extensive interagency collaboration among the USACE, NOAA, DHS S&T, FEMA, USGS and others. Internationally, leaders in hydroanalytic development and use should join in the development of broader standards and practices. At the grass roots, the innovative methodologies that are being developed and tested every day should be acknowledged as they contribute to the body of knowledge. By leveraging resources and fostering coordination and partnering with academia, international organizations and governments, industry and not-for-profit organizations, next generation analytics will provide a transformative and credible way for decision makers to address grand challenges. It will take champions at all levels in these organizations to bring about change.

With so many proven and emerging methodologies for hydroanalytics, a second key step toward a robust evaluation framework is to harmonize the evaluation standards and practices across the hydroanalytic CoP. Because the hydroanalytic community covers a broad range of disciplines, lessons learned from one area will need to be applied and merged with others.

Third, to track progress, organizations should set goals or targets for success. Organizations should establish detailed and clearly defined metrics for managing performance in each leg of the evaluation framework and/or throughout the modeling cycle. These metrics can be driven and tied to the technical quality scoring concepts described above as well as to the user needs and institutional performance. While the primary objective of tracking performance is for self-improvement, it can also be a window for how others view your organization. As stated before, best-of-class is not based on how we view ourselves but how others view us. Therefore, establishing and tracking performance metrics can provide a transparent and objective way for others to see our accomplishments. It can be a dashboard for progress that might incorporate various general measures such as:

Aerial view of container ship underway

- Performing regular external programmatic and policy evaluations and tracking implementation of recommendations
- Establishing requirements and maintaining lists of credentialed modelers
- Performing regular user/client satisfaction surveys
- Tracking revenue from contractual projects and research
- Maintaining a model inventory and performing regular certification or quality checks
- Tracking advances in development and/or completion of model improvements

Fourth, the evaluation process must be institutionalized and implemented. One place to start could be leveraging current federal interagency mechanisms to establish a governance platform for broad implementation of hydroanalytic standards. On a smaller scale, individual organizations can require hydroanalytic evaluation as they build new and/or update or modernize their current portfolio of hydroanalytic tools.

Finally, transparency and a willingness to share intellectual property requires a change in culture that can be difficult within small organizations but that gets even harder across large agencies and multi-organizational groups. Historical practices of isolated development and unnecessary competition may take decades to overcome, but changes are forthcoming as a new wave of transdisciplinary innovators joins to collectively improve the state-of-the-practice. Sometimes transformation happens in the face of a dire or dramatic event. Waiting for disaster to harmonize our efforts is not the best option. The time seems right for the hydroanalytic community to come together and embrace the changes needed to develop and implement an evaluation framework.

REFERENCES AND FOOTNOTES

American Society of Civil Engineers. (2017). ASCE Infrastructure Report Card. (web accessed April 13, 2020)

American Society of Mechanical Engineers. Verification, Validation and Uncertainty Quantification (VVUQ). asme.org/codes-standards/publications-information/verification-validation-uncertainty (website accessed April 15, 2020)

Bak, Spicer. (October 2018). *Coastal Model Test Bed. US Army Engineer Research and Development Center, Coastal and Hydraulics Laboratory, Coastal Inlets Research Program. [White Paper]. Unpublished. Duck, NC.*

Baart, Arthur. (November 30, 2018). *Deltares Software.* [PowerPoint presentation to Sandra Knight while visiting Deltares].

Blais, Curtis. (December 2008). *Naval Post Graduate School, Toward a Verification, Validation, and Accreditation (VV&A) Ontology.* NPS-MV-08-003

Brunner, G.W., Sanchez, A., Molls, T, Parr, D. (April 2018). *HEC-RAS Verification and Validation Tests. US Army Corps of Engineers, Hydrologic Engineering Center.* [Research Document, RD-52]. Davis CA.

Bulletin of the American Meteorological Society. (December 2019). Best Practices for Data Management, A best Practice Statement of the American Meteorological Society, Adopted by the AMS Council September 27, 2019. Vol. 100, No.12, pp 2673–2674.

Bushnell, Mark. (July 2018). Quality Assurance/Quality Control of Real-Time Oceanographic Data (QARTOD) developing Manual for Real-time Quality Control of Stream Observations- A Guide to QA/QC for Stream Flow Observations in Rivers and Streams. *The NHWC Transmission*

CDM Smith. (June 2014). North Atlantic Coast Comprehensive Study Visioning Meetings Summary, Final Report, report submitted to US Army Corps of Engineers, Baltimore.

Coastal Protection and Restoration Authority. (June 2, 2017). Louisiana's Comprehensive Master Plan for a Sustainable Coast. Published on behalf of John Bel Edwards, Governor, State of Louisiana by OTS-State Printing.

COMPASS PTS JV (January 2019). Rapid Response Flood Modeling Final Report, prepared for US Department of Homeland Security, FEMA Mitigation Branch. (website accessed April 15, 2020).

Costello, C., Ensor, C., Chadwick, S. (July 23, 2019). The Power of Data Transparency, article from the First Street Foundation.firststreet.org/flood-lab/research/data-transparency (website accessed April 14. 2020)

Cotton, Gordon. (June 1979). *A History of the Waterways Experiment Station: 1929-1979*. Retrieved from apps.dtic.mil/dtic/tr/fulltext/u2/a635951.pdf

Crowder, R. Pepper, A., Whitlow, C., Sleigh, A., Wright, N. and Tomlin, C. (June 2004). *Benchmarking of hydraulic river modelling software packages,* [R&D Technical Report: W5-105/TR0] prepared by Bullen Consultants for Defra/Environmental Agency, Flood and Coastal Defence R&D Programme. UK Department for Environment, Food and Rural Affairs (Defra)/Environmental Agency. Retrieved from gov.uk/government/organisations/department-for-environment-food-rural-affairs

Danish Hydraulic Institute, DHI dhigroup.com

de Jong, Martijn. (December 10, 2018). *Deltares, V-model RWS Ministry of Infra Netherlands.* [Slide provided to Sandra Knight].

Deltares (2016). *R&D Highlights 2016.* deltares.nl/app/uploads/2015/02/RD-Highlights-2016_lowres.pdf (website accessed February, 19, 2019). The Netherlands.

Deltares. oss.deltares.nl/ Opening page to Deltares software. (website accessed February 19, 2019)

Demargne, J. et al. (2013). The Science of NOAA's Operational Hydrologic Ensemble Forecast Service. *Bulletin of the American Meteorological Society.* 95. 10.1175/BAMS-D-12-00081.1.

Demirbilek, Z. and Rosati, J.D. (December 2011). *Verification and Validation of the Coastal Modeling System.* [Technical Report ERDC/CHL TR-11-10, US Army Engineer Research and Development Center, Coastal and Hydraulics Laboratory]. Vicksburg, MS.

Department of Defense (DoD) (2010). Defense Acquisition Guidebook, Technology Readiness Levels in the Department of Defense.

Department of Defense (DoD) (2009). "Clarifying Guidance Regarding Open Source Software (OSS)," Department of Defense (DoD), Chief Information Officer (CIO). dodcio.defense.gov/Portals/0/Documents/FOSS/2009OSS.pdf

Department of Defense (DoD). (2008). *MIL-STD-3022. DoD Standard Practice, Documentation of Verification, Validation, and Accreditation (VV&A) for Models and Simulations*. (website accessed February 4, 2019) acqnotes.com/Attachments/MIL-STD-3022%20Documentation%20of%20VV&A%20for%20Modeling%20&%20Simulation%2028%20Jan%2008.pdf

Department of Defense (DoD) (1988). *MIL-STD-2167A Defense System Software Development.*

Department of Homeland Security, Office of Inspector General. (September 27, 2017). FEMA Needs to Improve Management of Its Flood Mapping Programs. OIG-17-110

Deuten, J., Koens, L., Faasse, P. and van der Meulen, B. (December 16, 2015). *Protocol for the monitoring and evaluation of the Applied Research Organizations in the Netherlands.* Rathenau Instituut. The Netherlands

Federal Emergency Management Agency (FEMA) (November 2019). Coordinated Needs Management Strategy (CNMS) Technical Reference: CNMS Database User's Guide. (website accessed April 15, 2020)

Federal Emergency Management Agency (FEMA). (February, 2018). *FEMA Policy Standard for Flood Risk Analysis and Mapping,* [FEMA Policy #FP 204-078-1 (rev7)].

Federal Emergency Management Agency (FEMA). (February 2018). Guidance for Flood Risk Analysis and Mapping: Accepting Numerical Models for Use in the NFIP. Retrieved from: fema.gov/media-library-data/1537888714503-3f556061b84f26b86d286513240e8dd4/Guidance_Accepting_Numerical_Models_for_Use_in_the_NFIP_(Feb_2018).pdf (website accessed April 13, 2020)

Federal Emergency Management Agency (FEMA). (June 2016). ***Technical Mapping Advisory Council, National Flood Mapping Program Review, June 2016*** (website accessed December 27, 2018).

Federal Emergency Management Agency (FEMA). (December 2015). *Technical Mapping Advisory Council, Future Conditions Risk Assessment and Modeling, December 2015*

Federal Emergency Management Agency (FEMA). (November 2014). Retrieved from ***Guidelines and Standards: Master Index*** (website accessed April 15, 2020)

Federal Energy Regulatory Commission (FERC). (Draft 2014). Engineering Guidelines, Risk-Informed Decisions Making, Chapter R21, Dam Breach Analysis. ferc.gov/industries/hydropower/safety/guidelines/ridm/eng-guide/chapter-R21.pdf (website accessed January 25, 2019)

Federal Energy Regulatory Commission (FERC). (Draft 2014). *Engineering Guidelines, Risk-Informed Decision Making, Chapter R19, Probabilistic Flood Hazard Analysis.* ferc.gov/industries/hydropower/safety/guidelines/ridm/eng-guide/chapter-R19.pdf (website accessed January 25, 2019)

FLO-2D flo-2d.com

Fox, C and Ukumen, G. (2011). Distinguishing Two Dimensions of Uncertainty, Chapter 1 in Brun, Keren, Kirkeboen and Montgomery book on Perspectives on Thinking, Judging, and Decision Making.

Holdren, John P. (February 22, 2013). *Appendix C. Office of Science and Technology Policy 2013 memorandum: Increasing Access to the Results of Federally Funded Scientific Research.* [Executive office of the President, Office of Science and Technology Policy].

Howard, L.S. (July 24, 2018). U.S. Flood Models Still in Infancy, So Underwriters Must be 'Cautious consumers.' *Insurance Journal.* amp-insurancejournal-com.cdn.ampproject.org/c/s/amp.insurancejournal.com/news/national/2018/07/24/495798.htm

HR Wallingford, pyxis. hrwallingford.com/software/pyxis (website accessed February 19, 2019)

Iowa Institute for Hydroscience and Engineering, IIHR iihr.uiowa.edu

Jupiter Intelligence Flood Score jupiterintel.com/wp-content/uploads/2018/06/Jupiter-FloodScore-Overview_5-31-18.pdf?61dbcc&61dbcc

Kahneman, Daniel. (2011). *Thinking, Fast and Slow.* 1st Ed. USA. Farrar, Straus and Giroux. New York.

Knight, S., Richardson, T., Lenhardt, C., Graeden, E. Luettich, R. (2017). *Rethinking Flood Analytics: Proceedings from the 2017 Flood Analytics Colloquium.* Co-sponsored by DHS Science and Technology and the Office of University Programs, Hosted by RENCI and CRC, UNC Publications. coastalresiliencecenter.unc.edu/files/2018/09/CRC-FloodAnalytics-Colloquium-Report.pdf

Kulp, S.A., Strauss, B.H. (2019). New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding. *Nature Communication* 10, 4844 (2019). doi.org/10.1038/s41467-019-12808-z

Longenecker, H., Graeden, E., Kuskiewicz, D, et. al. (July 2019). A rapid flood risk assessment method for response operations and nonsubject-matter-expert community planning. Journal of Flood Risk Management, CIWEM, Wiley.

Lord, P., Macdonald, A., Lyon, L, Giaretta, D. (2004) From Data Deluge to Data Curation. *The Digital Archiving Consultancy Limited and the Digital Curation Center. Journal of Documentation 67 (2):*214-2237. Doi:10.1.1.111.7425.

Masters, Jeff. (2019). *Weather Underground. Hurricane and Tropical Cyclones, Hurricane Forecast Computer Models*. wunderground.com/hurricane/models.asp (website accessed February 5,2019)

Mosavi, A., Ozturk, P., Chau, K. (October 27, 2018). Flood Prediction Using Machine Leaning Models: Literature Review. *Water* 2018, 10, 1536:doi:10.3390/w10111536.

National Academy of Sciences, National Academy of Engineering, and Institute of Medicine. (2007). *Rising Above the Gathering Storm: Energizing and Employing America for a Brighter Economic Future*. Washington, DC: The National Academies Press. doi.org/10.17226/11463.

National Institute of Standards and Testing (January 2017). National Voluntary Laboratory Accreditation Program nist.gov/nvlap/accreditation-vs-certification (website accessed April 13, 2020)

National Institute of Standards and Testing (NIST). NIST Software Usability. nist.gov/industry-impacts/software-usability (website accessed April 14, 2020)

National Oceanic and Atmospheric Administration (NOAA).No date. National Water Model. water.noaa.gov/documents/wrn-national-water-model.pdf (website accessed April 15, 2020)

National Oceanic and Atmospheric Administration (NOAA). (January 1,2015) Data Management Planning Procedural Directive, v.2.0.1 nosc.noaa.gov/EDMC/PD.DMP.php; nosc.noaa.gov/tpio/main/aboutdma.html

National Oceanic and Atmospheric Administration (NOAA). (July 2007). *NAO 216-112: Policy on Partnerships in the Provision of Environmental Information,* corporateservices.noaa.gov/ames/administrative_orders/chapter_216/216-112.html (website accessed April 14, 2020)

National Oceanic and Atmospheric Administration (NOAA), Office of Oceanic and Atmospheric Research (OAR). (July 23, 2007). OAR Directive, *Laboratory Science Reviews. OAR No. 001-07.* NOAA, Silver Spring, MD

National Research Council. (2012). Assessing the Reliability of complex Models: Mathematical and Statistical Foundations of Verification, Validation, and Uncertainty Quantification. *The National Academies Press,* Washington, DC

National Weather Service (NWS). (April 20, 2018). *NWS Policy Directive 1-10, Administration and Management, Managing the Provision of Environmental Information.*

Neelz, S., Pender, G. (2013). Benchmarking of 2D Hydraulic Modelling Packages (2013 refresh). Report- SC120002/R, published by UK Environment Agency.

Public Law No. 115-436. (January 14, 2019). *Foundations for Evidence-Based Policy Making Act.* 115th Congress.

Raymond, Eric. (1999). The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary. *O'Reilly Media. ISBN 1-56592-724-9*

Roe, J., Dietz, C., Restrepo. P., Halquist, J., Harman, R., Horwood, R., Olsen, B, Opitz, H., Shedd, R., Welles, E. *NOAA's Community Hydrologic Prediction System.* (website accessed December 28, 2018). training.weather.gov/nwstc/CHPS/roe.pdf.

Ruijgh, E.F., Kroon,T. Snippen,E. (May 2015) *The Dutch Delta model to support national policy analysis.* [Conference paper for the International Water Resources Association, World Water Congress XV]. Edinburgh, Scotland.

Sayers, P., Hunter, N., Panzeri, M., Neve, P., et.al. (2011). Measuring Confidence in NaFRA Outputs: Stage 1 – Method. Report prepared for UK Environment Agency

Sorkin, Rich, (2018). *A new Era in Modeling Catastrophic Risk. Jupiter Intelligence.* brinknews.com/a-new-era-in-modeling-catastrophic-risk

Spreyrer, N. and Gaharan, A. (April 2017). 2017 Coastal Master Plan, Appendix G: Outreach and Engagement, Coastal Protection and restoration Authority.

Stackify. (May 11, 2017). *What is Scrum? How Does Rugby Help Software Developers Create Quality products.* stackify.com/what-is-scrum (website accessed February 19, 2019)

Sutherland, J., Peet, A.H. and Soulsby, R.L. (2004). "Evaluating the performance of morphodynamic modelling". Coastal Engineering, 51 (8-9): 917-939. doi.org/10.1016/j.coastaleng.2004.07.015

United Kingdom, Environment Agency (December 13, 2017). Technical Guidance (Draft) Open Coast Hydraulic Modeling Assessments.

United Kingdom, Environment Agency (2014). Flood and coastal erosion risk management: Long-term investment scenarios (LTIS) 2014.

University Corporation for Atmospheric Research (UCAR), (December 2015). Report of the UCACN Model Advisory Committee.

US Army Corps of Engineers (USACE). (July 26, 2019). Guidance for detection of nonstationarities in annual maximum discharges. [Engineer Technical Letter 1100-2-3 issued April 28, 2017 with errata sheet July 26, 2019]

US Army Corps of Engineers (USACE). (February 2012). Assuring Quality of Planning Models- Model Certification/Approval Process, Standard Operating Procedures. *National Planning Centers of Expertise.*

US Army Corps of Engineers (USACE). (June 11, 2011). *Software Validation for the Hydrology, Hydraulics and Coastal Community of Practice.* [Enterprise Standard (ES) -08101].

US Army Corps of Engineers (USACE). (March 31, 2011). *Assuring Quality of Planning Models.* [Engineer Circular 1105-2-412. Expires 31 March 2013].

US Army Corps of Engineers (USACE). (June 2009). Performance Evaluation of the New Orleans and Southeast Louisiana Hurricane Protection System: Final Report of the Interagency Performance Evaluation Task Force, Volume I- Executive Summary and Overview.

Wells, Ernie. (December 11, 2018.) *National Weather Service Hydrologic Ensemble Forecast Service.* [PowerPoint presentation provided to Sandra Knight].

Wilkinson, MD., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data 3:160018* doi: 10.1038/sdata.2016.18

Wing, O., Bates, P., Sampson, C.,Smith, A., Johnson, K., Erickson, T. (September 2017). Validation of a 30 m resolution flood hazard model of the conterminous United States, Water Resources Research, 53, AGU Publications, 7968–7986, doi:10.1002/2017WR020917.

Winter, B., Schneeberger, K., Huttenlau, M., Stotter, J. (November 2017). Sources of uncertainty in a probabilistic flood risk model. Nat Hazars (2018) 91:431-446, doi.org/10.1007/s1 1069-017-3135-5

Zaidi, S., Chandola, V., Allen, M., Sanyal, J., Stewart, R., Bhaduri, B. & Manamay, R. (2018). Machine learning for energy-water nexus: challenges and opportunities, *Big Earth Data, 2:3,* 228-267, DOI: 10.1080/20964471.2018.1526057 **tandfonline.com/doi/full/10.1080/20964471.2018.1526057**

Zarzar, Christopher M., Hossein Hosseiny, Ridwan Siddique, Michael Gomez, Virginia Smith, Alfonso Mejia, and Jamie Dyer, (2018). A Hydraulic Multimodal Ensemble Framework for Visualizing Flood Inundation Uncertainty, *Journal of the American Water Resources Association* (JAWRA) 54 (4); 807-819**.** doi.org/10.1111

Footnotes

1. Based on Conversations with Deltares on November 30, 2018.
2. Conversation with Dr. Larry Weber, IIHR, December 19, 2018.
3. Based on conversations with various experts from these organizations.
4. Materials provide by Ben Gouldby, HR Wallingford, August 13, 2019.
5. Per conversation with Jeff Lillycrop, December 28, 2018.
6. Per Conversations with UK Environment Agency, August 16, 2019.
7. Based on conversations with both Deltares and NWS.

waterwonks llc

THE UNIVERSITY *of* NORTH CAROLINA *at* CHAPEL HILL

Homeland Security

Science and Technology

COASTAL RESILIENCE CENTER

A U.S. Department of Homeland Security Center of Excellence