US Army Corps of Engineers®

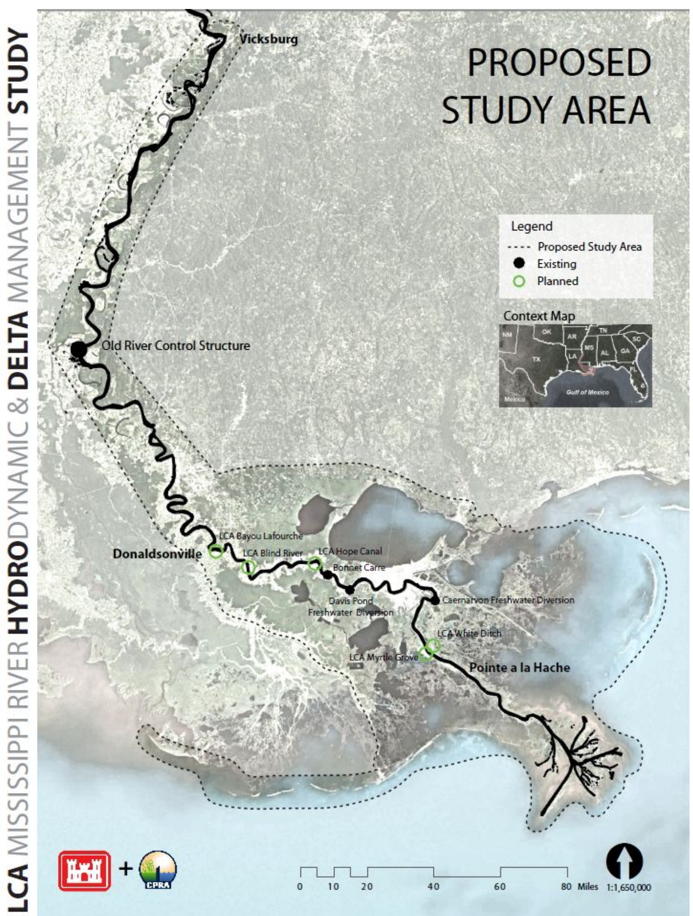
COASTAL PROTECTION AND RESTORATION AUTHORITY
CPRA

Louisiana Coastal Area Program Mississippi River Hydrodynamics and Delta Management Study

Models Performance Assessment Metrics and Uncertainty Analysis

Ehab A. Meselhe, PhD, PE and Mallory D. Rodrigue, MS, EI, PE

December 2013



Acknowledgements

The report was reviewed by the numerical modeling teams of the Mississippi River Hydrodynamic and Delta Management Study (MRHDM). The report was also reviewed by the Independent External Peer Review panel of the MRHDM study. Finally, this report was reviewed by Denise Reed (Chief Scientist) and edited by Kathleen Hastings of The Water Institute of the Gulf.

Models Performance Assessment Metrics and Uncertainty Analysis

Ehab A. Meselhe, PhD, PE

Contractor: The water Institute of the Gulf
301 N. Main St., Suite 2000
Baton Rouge, LA 70825

Mallory D. Rodrigue, MS, EI

Subcontractor: Fenstermaker
1100 Poydras Street, Suite 1550
New Orleans, LA 70163

Final report

Prepared for Louisiana Coastal Protection and Restoration Authority

Under LCA Mississippi River Hydrodynamic and Delta Management Project

Abstract

This report describes the development of the metrics for assessing model predictive performance component of the Mississippi River Hydrodynamic and Delta Management Study (MRHDM). Hydrodynamic models will be used for both components of the MRHDM study to provide an evaluation of the current conditions and future conditions under various environmental scenarios that capture plausible variations in conditions such as sea-level rise and subsidence, and assess the impact of implementing restoration strategies. There are uncertainties inherently associated with these numerical predictive models. As such, the intent of this report is to describe an uncertainty analysis approach to evaluate the confidence level in the models' predicting ability. The outcome and benefit of the proposed uncertainty analysis is to inform decision makers on how model uncertainties affect the assessment and feasibility of proposed restoration and protection strategies.

Contents

Illustrations

Figures

Tables

Preface

This study was conducted for the U.S. Army Engineer District, New Orleans and the State of Louisiana as part of the Louisiana Coastal Area Mississippi River Hydrodynamic and Delta Management Study. The project managers for the U.S. Army Engineer District, New Orleans were Bill Hicks and Daimia Jackson and the Plan Formulator was Cherie Price. The study managers for the State of Louisiana were Carol Parsons Richards, Austin Feldbaum, Elizabeth Jarrell, and Wes LeBlanc. Ehab Meselhe from The Water Institute of the Gulf and Barb Kleiss from the Corps of Engineers, Mississippi Valley Division, were the Technical Leads.

Ehab Meselhe and Mallory Rodrigue conducted this study from August 2012 to December 2013. The principal investigators for this study were Ehab Meselhe of The Water Institute of the Gulf and Mallory Rodrigue of Fenstermaker.

1 Background of the Study

The Louisiana Coastal Area, Louisiana Ecosystem Restoration Study was recommended to Congress by a Chief of Engineers report dated January 31, 2005, that called for a coordinated, feasible solution to the identified critical water resource problems and opportunities in coastal Louisiana. The MRHDM Study focuses on the use and applications of tools that can evaluate both the existing conditions of the Mississippi River and any potential local and system-wide impacts of proposed changes to the system, such as large scale river diversions. The Mississippi River hydrodynamic component of the feasibility study focuses on impacts to the Mississippi River. This component will: (1) evaluate the Mississippi River system from old River Control Structure to the Gulf of Mexico, (2) develop a comprehensive numerical modeling system to assess potential restoration alternatives, and (3) determine the availability of fresh water, sediment, and nutrients for restoration usage without compromising flood control and navigation missions. The Mississippi River Delta Management component focuses on impacts to the receiving areas.

Hydrodynamic models will be used for both components of the MRHDM study to provide an evaluation of the current conditions, future conditions under various environmental scenarios that capture plausible variations in conditions such as sea-level rise and subsidence, and assess the impact of implementing restoration strategies. There are uncertainties inherently associated with these numerical predictive models. As such, the intent of this report is to:

- Provide a clear definition and distinction between terminologies such as model calibration, validation, sensitivity analysis, performance metrics, and uncertainty analysis;
- Provide a framework of these various components including model setup, calibration/validation, sensitivity analysis, and uncertainty analysis, of the overall modeling effort;
- Design an approach to quantify the models' uncertainties;
- Discuss guidelines to evaluate the ability of these predictive models to support and inform decision making regarding proposed coastal restoration and protection strategies.

2 Purpose

This report describes the development of the Metrics for Assessing Model Predictive Performance component of the Mississippi River Hydrodynamic and Delta Management Study (MRHDM). This assessment ensures transparency in the models' performance evaluation. It also provides objective quantifiable measures for evaluating the models' performance to minimize subjectivity in the performance assessment. Further, this report outlines an uncertainty analysis approach to evaluate the confidence level in the models' predicting ability and how these modeling tools can support and inform decisions regarding restoration strategies.

This is a critical component of the overall MRHDM Study as it provides uncertainty bounds on the models' results, allowing disclosure of risks—and associated assumptions—and uncertainties related to model performance, thus facilitating the decision-making process. An external review panel will provide an independent opinion on the models' performance and utility based on the metrics and uncertainty analysis developed through this task. Further, the assessment of the models' performance will provide valuable feedback on areas of potential improvements needed to achieve the study's objectives. Additionally, this analysis will provide a baseline with which to evaluate performance improvements made through the field measurements collected through the broader MRHDM effort.

The performance metrics developed herein will help establish credibility of the models while also helping set realistic expectations regarding model outputs, when considering the complexity of the Lower Mississippi River and deltaic system, and the limitations of numerical modeling tools in general. The model performance assessment procedure presented here should supplement (and not replace) other guiding standard documents such as the American Association of Civil Engineers (ASCE) Manual 110 (2008).

It should be emphasized that data used to evaluate the model performance also include uncertainty. Uncertainty in field observations directly affects the model input and output information. As such, the performance metrics presented here should not be viewed as pass/fail criteria. Rather, it should be used to gain insights into the models' performance and the level of confidence the models' predictions should be viewed

3 Approach

The process of setting up and using numerical models involves several steps and phases (Habib, 2012, Kleindrfer et al., 1998, Lall et al., 2002, and Oreskes et al., 1994). Currently, there is no universal consensus on the terminology of each step or phase of such a process. Therefore, this report proposes a convention to be used among the various modeling and analysis teams within the MRHDM study. Figure 1 provides an overall outline of a possible version of a numerical modeling process. It also shows the sequence and transfer of information from one step to another.

Figure 1. Sequence and transfer of information among the various components of a numerical modeling effort.

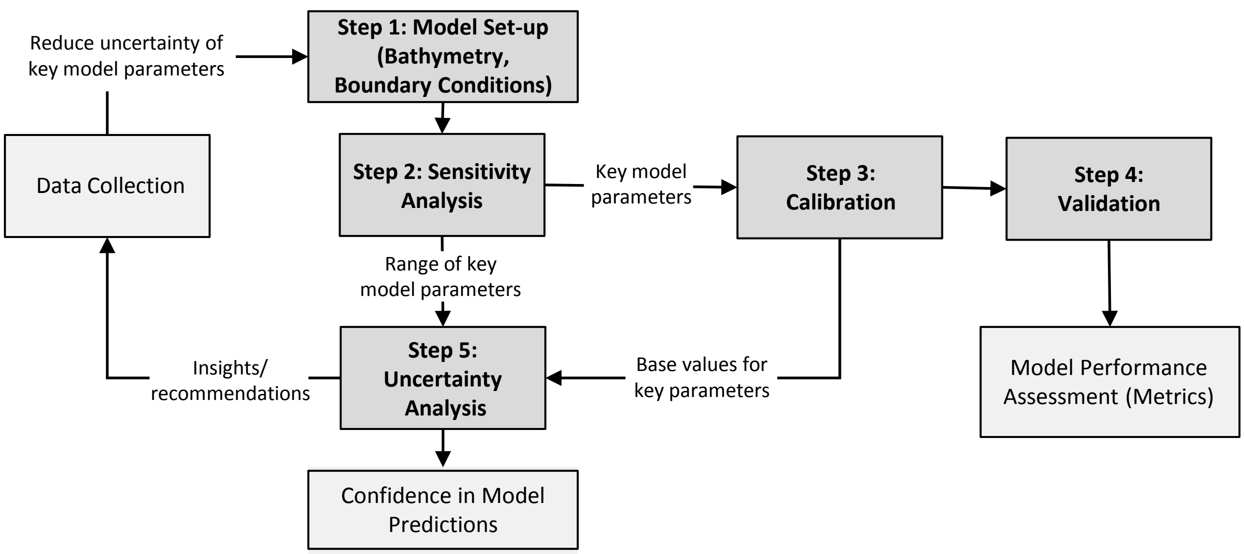


Figure 1 assumes that the models used have already been verified to comply with the standard governing flow equations (e.g. mass conservation). This step is typically performed for newly developed models. For this study, only mature and widely accepted models are being utilized. As such it is assumed that these models conform to the standards. Further, a comprehensive description of each of the steps is beyond the scope of this report. Thus, only a brief overview of each component is provided here. Typically, the first step of a modeling effort—after gathering the necessary data for key model parameters—is to set up the model. That entails defining the model's spatial extent (often referred to as the model domain), selecting the type and location of the model boundaries as defined by the area of interest, and designing a computational grid. The spatial resolution (or spacing between computational points or sections) is selected to capture the variations in the bathymetry and topography, and to provide sufficient information at critical locations among other factors. The model boundary conditions are typically defined based on field observations. Once the model setup is complete, a sensitivity analysis should be conducted (Step 2). The modeling teams would identify a list of parameters perceived to be important but whose precise value is not known. Numerical simulations would be conducted as part of the sensitivity analysis while varying the values of each of the parameters identified by the modeling teams. The range of variation for each parameter is established from published literature and is to be within accepted physical limits. Through these numerical simulations, key model parameters that have a strong influence or impact on the model output should be identified. Although it is not the focus of this report, it is worth noting that sensitivity analysis is also used to examine the adequacy of spatial resolution of the computational grid designed during the model setup phase. The goal is to ensure that the model results are no dependent on the spatial resolution. This sometimes is referred to as "grid independence analysis."

The key model parameters identified through the sensitivity analysis should be the focus of Step 3, the calibration effort. These key model parameters are fine-tuned until the model output compares well to field/laboratory observations (Hammons & Shelden, 2012). During calibration, parameters are adjusted to meet a desired response. The model calibration process is discussed later in this report. Through the calibration process, a base value is established for each key parameter identified in the sensitivity analysis phase. Once this base value is established, no further changes to the key model parameters are permitted. In Step 4—and using these base values—additional model simulations are to be performed under different conditions, or time periods, than those used in the model calibration phase. This is called model validation. Validation differs from calibration in that parameters are no longer adjusted to optimize the fit, and that the observed data are independent of the data used during calibration. This process provides circumstantial evidence of acceptable model performance (ASCE Manual 110). Both graphical and statistical metrics can be used to assess the model performance and how well it replicates the natural system being modeled. Full details about the performance metrics recommended for MRHDM are described later in this report. Based on the models' performance established through the metrics, they can be declared to be calibrated and validated.

The list of key model parameters and the base value and range assigned to each of these are necessary to perform the uncertainty analysis for Step 5. The uncertainty analysis presented here should be performed using the calibrated and validated models. A full section devoted to defining and discussing uncertainty analysis is provided later in this report. In brief, the goals of the uncertainty analysis are to:

- Establish the confidence bounds in the numerical models predictions and assess the model's ability to discern the impacts of restoration projects;
- Identify areas or elements of weakness and high uncertainties in the models that mask the model predictions and limit their use;
- Provide insights and recommendations to reduce such uncertainties. These recommendations can be incorporated in the design of future data collection programs, laboratory experiments, and other research activities. Improving the knowledge of key model parameters is one the most effective mechanisms to reduce model uncertainties.

Before introducing the details of establishing the performance metrics and the uncertainty analysis, the general philosophy and terminology of modeling assessment strategies are discussed below.

4 Philosophy

There has been an ongoing debate between scientists and modelers regarding the nature and legitimacy of numerical model validation. This debate comprises (1) a philosophical debate, which calls into question whether or not model validation is even possible; (2) a terminology debate, which attempts to redefine the intent of validation; and (3) a predictability debate, which questions whether or not a model can be used for forecasting purposes. The intent of this overview is not to expand on or even summarize these debates. Rather, it will highlight key schools of thought and discussions and then present a brief summary. This section also includes additional definitions for sensitivity analysis, calibration, and uncertainty analysis, which are sometimes misused or misunderstood.

Konikow and Bredehoeft (1992) suggest that the concept of model validation must be viewed in a philosophical context, and argue that there are two primary schools of philosophical thought for model validation. The first states that theories or hypotheses can be proved right or wrong on the basis of experimentation. This philosophy suggests that it is possible to validate or invalidate a model. The second school of thought suggests that theories or hypotheses in science can only be proved wrong through experimentation (Oreskes, 1998; Konikow & Bredehoeft, 1992; Kleindorfer et al., 1998). In terms of model validation, this philosophy suggests that validation is never possible because there may always be another set of circumstances that disprove the model's conceptual theories (Oreskes, 1998; Konikow & Bredehoeft, 1992). Additionally, it has been argued that to say a model is validated is to claim that the model contains certain truths. As such, model validation is considered impossible (Rykiel, 1996). However, Rykiel (1996) points out that scientific truth is relative to the knowledge of the modeler at the time the model was created.

4.1 Terminology

Most scientists and modelers that believe validation is not possible tend to refer their audience to the common-use dictionary definitions of the word valid, i.e., being true or logically correct. Therefore, stating that a model is validated means to them that the model is authenticated or legitimized (Oreskes, 1998; Konikow & Bredehoeft, 1992). Consequently, their argument is that models cannot be valid because there are always uncertainties or flaws in the conceptual or mechanical design of the models (Oreskes, 1998). Oreskes suggests that excessive positive language surrounds model validation and proposes that scientists and modelers should use neutral terms such as evaluate to allow for both positive and negative results (Oreskes, 1998; Oreskes et al., 1994). Other attempts to avoid using the term valid when referring to models are to use synonyms such as verify, confirm, or corroborate (Rykiel, 1996; Oreskes et al., 1994). The result of applying these terms is a reduction in the public's confidence in the model, or not to claim more accuracy than is possible to achieve. However, as Rykiel (1996) points out, all these terms mean the same thing. Opposed to using repetitive synonyms, Rykiel (1996) suggests changing the meanings or semantics of the words truth and valid to better fit the modeling world. For example, Rykiel (1996) redefines the Merriam-Webster definition of truth, i.e., the quality or property of being in accord with fact or reality, and translates it into a modeling definition "consistent with available data." Other examples are provided, but the overall concept of changing the meanings is the same.

An important concept for the public, policy makers, modelers, and scientists to understand is that "one cannot determine the meaning of a technical term simply by inquiring about its common meaning or less, its etymology" (Roache, 2009). Similar to Rykiel's concept of amending the common-use definitions to fit the modeling world, Roache (2009) suggests that verify and validate are technical terms and therefore should be defined in a technical context. This logic seems reasonable and as such will be adopted here for the MRHDM study. The technical definitions of verification and validation as specified by ASCE (Roache, 2009) will be adopted here. Verification is defined as "solving the equations right" (Roach, 2009). This means modelers have a duty to test that equations are coded correctly and that correct results are achieved. Verification is achieved by performing hand calculations, debugging the code, performing dye tests, or modeling simplified systems with known analytical solutions. Validation is defined as "solving the right equations" (Roache, 2009). This means that modelers have a duty to ensure that the equations are applicable to the problem at hand. Validation is achieved by showing, either graphically or statistically, that the model results can match the previously observed data. As a result, model verification should precede model validation.

4.2 Reliability of Predictions

Models are developed and used to help policy makers and decision makers better understand the environment and provide recommendations based on science and experience. Therefore, the discussion among scientists and modelers becomes focused on whether or not a numerical model can be used for predictive purposes. When a model is calibrated and then validated to replicate past natural conditions, does that mean it can predict future conditions? Some argue that model forecasting is not possible. Their argument is that the uncertainties embedded in the numerical models prevent them from having any predictive capabilities (Oreskes, 1998; Konikow & Bredehoeft, 1992). Oreskes et al. (1994) argue that a calibrated model, whose results match or fit a previously observed dataset, cannot predict future conditions due to the dynamic property of the natural environment. Haag and Kaupenjohann (2001) agree that predictive capacity is not guaranteed; however, they argue that simulation models can at least "outline a space of possibilities or of potentiality," that can be used to better understand the system being modeled. Konikow and Bredehoeft (1992) also suggest that calibration does not guarantee that all future conditions, or stresses, can be accurately replicated. Further, Konikow and Bredehoeft (1992) indicate that "predictions should be cast in a probabilistic framework with confidence limits bounding the predicted response." Rykiel (1996), on the other hand, argues that model prediction is possible and can be proved through validation testing. Oreskes (1998) emphasizes that "prediction is not as important as it is often thought to be." In summary, recommendations by decision makers based off model predictions can be better supported if modelers determine the uncertainty limitations of the model outputs.

4.3 Summary

It is unreasonable to assume that perfect validation is achievable, meaning that the natural environment can rarely be perfectly replicated through numerical models. It is acknowledged here that there will always be sources of error or uncertainty in any model. However, perfect validation is not essential for the purposes of numerical modeling. Certain levels of imperfection in the validation process are acceptable. It is reasonable to agree with Rykiel (1996) and adopt the pragmatic declaration that "a model only needs to be good enough to accomplish the goals of the task to which it is applied." In order to build confidence in the model for the decision makers and practitioners, modelers should provide: (1) the model's purpose or intended use, (2) the acceptability criteria, or limits, of the model validation, and (3) the operational context or assumptions within the model. Therefore, the use of model performance metrics to establish those limits of acceptability for the validation process is recommended. It is also recommended to determine uncertainty bounds for critical model parameters to support the decision-making process.

5 MRHDM Goals and Objectives

It is essential to establish the overall study goals and objectives. As stated above, defining the goals and objectives is critical to identifying the physical processes and model output variables on which to focus. Simple agreement between model output and field measurements measured by standard statistics, although useful, is not necessarily meaningful. It is critical to link the model performance and the uncertainty bounds of the model predictions to the measures that directly define the project goals and objectives. The goal and objectives of the MRHDM study are briefly described below:

Goal: Reconnect Mississippi River water, sediment, and nutrient resources to the surrounding basins, sufficient to provide a sustainable coastal ecosystem that allows for the coexistence of navigation and flood risk reduction.

Objectives:

- Re-establish natural deltaic processes to restore the maximum number of acres of wetlands and sustain habitats in the long term.
- Maintain dynamic diversity of the coastal wetland ecosystem delta-wide over time

6 Models' Performance Metrics

Model calibration can be a never-ending process. There is always another combination of parameters that will produce a better fit to the observed data. However, is a perfect fit necessary for the model to be informative and helpful to meet the project objectives? In most cases, perfection is not necessary, nor is it possible. Uncertainties in the model input (e.g., bathymetry, boundary conditions, etc.), physical parameters (e.g. roughness), numerical parameters (e.g. diffusion coefficients) and model simplifying assumptions—even in combination—do not allow for perfect calibration or replication of the natural system. Therefore, it is important to establish a set of metrics to help identify acceptable model performance in such a way as to support decision making regarding project implementation.

In 2012, FTN Associates established a set of model performance metrics for the LCA Medium Diversion at the Myrtle Grove study (Hammons & Shelden, 2012). These metrics were intended to establish acceptable model performance using three goodness-of-fit statistics. Data scarcity, data uncertainties, and inconsistencies must be considered when the performance metrics are used or applied (Hammons & Shelden, 2012). As a result, the parameters with the most observed data were given the most stringent criteria and vice versa. For example, there are typically more robust records of stage data than discharge/velocity, salinity, or sediment data. Expanding upon the metrics recently established by Hammons and Shelden (2012), metrics for modeling the Mississippi River discharge/velocity and sediment are suggested below.

Model performance metrics were developed for MRHDM for application to one-dimensional, two-dimensional, and three-dimensional models. The following list specifies the model outputs that were used to create these model performance metrics for the MRHDM study. It should be emphasized again that the performance metrics should not be used to assess whether a model 'passes' or 'fails'. Rather, they should be used to gain insights into the model performance, identify potential weaknesses, and guide efforts to gather additional information to improve the performance.

- One-Dimensional Models

    - River water depth (hourly or daily depending on availability of data);
    - Water discharge (riverine or tidal) or velocity within the river channel and leaving the river through all major passes (daily averages);
    - Cross sectional load of suspended sediment (daily or annual depending on availability of data). Alternatively, cross sectional average concentration could be used;
    - Cross sectional total (suspended plus bed) Load for sand, silt, and clay fractions (daily or annual);
- Two-Dimensional Models
    - River water depth;
    - Water discharge (riverine or tidal) or velocity within the river channel and leaving the river through all major passes (daily averages);
    - Cross sectional load of suspended sediment (daily or annual depending on availability of data). Alternatively cross sectional average concentration could be used;
    - Cross sectional total (suspended plus bed) Load for sand, silt, and clay fractions (daily or annual);
    - Depth averaged velocity within the river channel and leaving the river through major passes;
    - Depth average salinity concentration;

- Three-Dimensional Models
    - Vertical and transverse profiles:
        - Velocity
        - Sediment, both coarse and fine material concentrations
    - Bed change quantities (erosion and deposition)
    - Cross sectional load of suspended sediment (daily or annual depending on availability of data);
    - Cross sectional total (suspended plus bed) Load for sand, silt, and clay fractions (daily or annual);
    - Vertical salinity concentration profile.

6.1 Developing the Model Performance Metrics

The objective of this section is to establish metrics of acceptable model performance. This section is based on, and expands upon, Hammons and Shelden (2012). These metrics were created for three goodness-of-fit statistics: (1) the root mean square error (RMSE) percentage, (2) the Pearson product-moment correlation coefficient, and (3) bias. The following section presents a definition for these statistics and their corresponding model performance metrics that will be adopted for the MRHDM study. It is important to note that these performance metrics can also be used for the validation process.

6.2 The Root Mean Square Error

The RMSE is a measure of the variation of predicted or modeled data to observed data (Legates and McCabe, 1999). RMSE is estimated as the square root of the average of the squared residuals, where the residuals are the differences between the predicted and observed data. The RMSE percentage is calculated as follows:

$$\boldsymbol{RMSE}\% = \sqrt{\frac{\sum_{i=1}^{n}(\boldsymbol{P_i}-\boldsymbol{O_i})^2}{\boldsymbol{n}}} * \frac{\boldsymbol{n}}{\sum_{i=1}^{n}\boldsymbol{O_i}} * \mathbf{100}\% \qquad \text{A1}$$

where:

P = predicted value
O = observed value
n = number of observations

A small RMSE percentage corresponds to a better fit between the predicted and observed data. Table 1 shows the model performance metrics for RMSE percentage of one-, two-, and three-dimensional models. The tables are organized into two targets: high and low. The high (desired) target represents a very good match between the model predictions and the field observations, whereas the low (acceptable) target represents a moderate match. If a model does not meet the acceptable target, it does not mean that the model is not useful nor does it imply that insights cannot be gained from such a model. Rather, it should be acknowledged and taken into account during analysis and interpretation. These ranges are not intended to be rigid metrics to assess performance; rather, they should be viewed as guidelines. It should be noted that water depth has been used in the RMSE calculations instead of stage (water level). Since stage is references to an arbitrary datum, it is not possible to express the results in percent. As such, water depth is being proposed as an alternative to evaluate the models' performance.

Table 1. Root Mean Square Error Metrics for One-, Two-, and Three-dimensional Models.

One-dimensional Models		
Model Output	Target – Desired	Target – Acceptable
River Water Depth	< 15% for all stations	not applicable, will need to use direct units
Total Suspended Concentration	< 33% for all stations	< 50% for 50% of stations
Total (susp. plus bed) Load for Coarse Fraction	< 33% for all stations	< 50% for 50% of stations
Total (susp. plus bed) Load for Fine Fraction	< 33% for all stations	< 50% for 50% of stations
Water discharge	< 20% for all stations	< 20% for 50% of stations
Two-dimensional Models		
Model Output	Target - Desired	Target - Acceptable
River Water Depth	< 15% for all stations	< 15% for 80% of stations
Depth Average Salinity*	< 20% for all stations	< 40% for 50% of stations
Total Suspended Concentration	< 33% for all stations	< 50% for 50% of stations
Total Load for Coarse Fraction	< 33% for all stations	< 50% for 50% of stations
Total Load for Fine Fraction	< 33% for all stations	< 50% for 50% of stations
Velocity (depth average)	< 20% for all stations	< 20% for 50% of stations
Water Discharge	< 20% for all stations	< 20% for 50% of stations
Three-dimensional Models		
Model Output	Target - Desired	Target - Acceptable
Velocity (vertical profile)	< 20% for all stations	< 30% for 50% of stations
Velocity (transverse profile)	< 20% for all stations	< 30% for 50% of stations
Coarse Sediment (vertical profile)	< 33% for all stations	< 50% for 50% of stations
Fine Sediment (vertical profile)	< 33% for all stations	< 50% for 50% of stations
Total Load for Coarse Fraction	< 33% for all stations	< 50% for 50% of stations
Total Load for Fine Fraction	< 33% for all stations	< 50% for 50% of stations
Vertical Salinity concentration profile	< 20% for all stations	< 40% for 50% of stations

* If phase errors are detected, and since it may amplify RMSE errors and present unrealistic model performance metrics, it is recommended to compare maximum and minimum or daily mean salinities instead of higher frequencies such as hourly.

6.3 Pearson Product-Moment Correlation Coefficient

The Pearson product-moment correlation coefficient, r, is a measure of the phasing between the predicted and observed data (Legates and McCabe, 1999). It does not take into account the amplitude of the residuals, but rather how well the peaks and troughs of the curves line up. The Pearson product-moment correlation coefficient is calculated as follows:

$$\boldsymbol{r} = \frac{\sum_{i=1}^{n}(\boldsymbol{P_i}-\bar{\boldsymbol{P}})(\boldsymbol{O_i}-\bar{\boldsymbol{O}})}{\sqrt{\sum_{i=1}^{n}(\boldsymbol{P_i}-\bar{\boldsymbol{P}})^2}\sqrt{\sum_{i=1}^{n}(\boldsymbol{O_i}-\bar{\boldsymbol{O}})^2}} \qquad \text{A2}$$

where:

P = predicted value
$\bar{P}$ = mean of predicted values
O = observed value
$\bar{O}$ = mean of observed values
n = number of observations

The value of r ranges is from -1.0 to +1.0, where a value of +1.0 is preferred. The following tables show the model performance metrics for the Pearson product-moment correlation coefficient of one-, two-, and three-dimensional models, respectively. Table 2 is organized in a similar fashion to the tables for the RMSE. Some of these metrics can be assessed only if the data are available. These metrics are not applicable, or at least not reliable, if data are scarce or the sample size is small (e.g. short record, or few data points). Such decision would be made by the modeling teams.

Table 2. Correlation Coefficient Metrics for One-, Two-, and Three-dimensional Models.

One-dimensional Models		
Model Output	Target - Desired	Target - Acceptable
River Water Depth	> 0.9 for all stations	> 0.9 for 80% of stations
Total Suspended Concentration	> 0.5 for all stations	> 0.5 for 50% of stations
Total (susp. plus bed) Load for Coarse Fraction	> 0.5 for all stations	> 0.5 for 50% of stations
Total (susp. plus bed) Load for Fine Fraction	> 0.5 for all stations	> 0.5 for 50% of stations
Water discharge	> 0.8 for all stations	> 0.7 for 50% of stations

Two-dimensional Models		
Model Output	Target - Desired	Target - Acceptable
River Water Depth	> 0.9 for all stations	> 0.9 for 80% of stations
Total Suspended concentration	> 0.5 for all stations	> 0.5 for 50% of stations
Load for Coarse Fraction	> 0.5 for all stations	> 0.5 for 50% of stations
Load for Fine Fraction	> 0.5 for all stations	> 0.5 for 50% of stations
Depth averaged salinity	> 0.7 for all stations	> 0.5 for 50% of stations
Velocity (depth average)	> 0.8 for all stations	> 0.7 for 50% of stations
Water Discharge	> 0.8 for all stations	> 0.7 for 50% of stations

Three-dimensional Models		
Model Output	Target - Desired	Target - Acceptable
Velocity (vertical profile)	> 0.75 for all stations	> 0.75 for 50% of stations
Velocity (transverse profile)	> 0.75 for all stations	> 0.75 for 50% of stations
Coarse Sediment (vertical profile)	> 0.5 for all stations	> 0.5 for 50% of stations
Fine Sediment (vertical profile)	> 0.5 for all stations	> 0.5 for 50% of stations
Load for Coarse Fraction	> 0.5 for all stations	> 0.5 for 50% of stations
Load for Fine Fraction	> 0.5 for all stations	> 0.5 for 50% of stations
Vertical Salinity concentration profile	> 0.7 for all stations	> 0.5 for 50% of stations

6.4 Bias

Assessing the model bias is important in order to ensure that the model is not consistently over- or underestimating critical quantities. Or at least, if the model has inherent bias that cannot be corrected, it is important for such bias to be documented and taken into account during the analysis. Bias can be calculated as follows:

$$\boldsymbol{Bias} = \frac{\bar{P} - \bar{O}}{\bar{O}} \qquad \text{A3}$$

where:

$\bar{P}$ = mean of the predicted values
$\bar{O}$ = mean of observed values

Table 3 shows the model performance metrics for the bias for one-, two-, and three-dimensional models, respectively. It should be noted that bias could be positive (overestimation) or negative (underestimation). As such, the values listed in the table apply to the magnitude of the bias regardless of the sign. The table is organized into two ranges with an ideal range (goal), which is desirable, and a range limit, which is the limit of accepta-ble performance.

Table 3. Bias Metrics for One-, Two-, and Three-dimensional Models.

One-dimensional Models		
Model Output	Target - Desired	Target - Acceptable
River Water Depth	< 10 for all stations	< 10 for 80% of stations
Total Suspended Concentration	< 20 for all stations	< 20 for 50% of stations
Total (susp. plus bed) Load for Coarse Fraction	< 20 for all stations	< 20 for 50% of stations
Total (susp. plus bed) Load for Fine Fraction	< 20 for all stations	< 20 for 50% of stations
Water discharge	< 15 for all stations	< 15 for 50% of stations
Two-dimensional Models		
Model Output	Target - Desired	Target - Acceptable
River Water Depth	< 10 for all stations	< 10 for 80% of stations
Total Suspended concen.	< 20 for all stations	< 20 for 50% of stations
Load for Coarse Fraction	< 20 for all stations	< 20 for 50% of stations
Load for Fine Fraction	< 20 for all stations	< 20 for 50% of stations
Salinity (depth average)	< 20 for all stations	< 20 for 50% of stations
Velocity (depth average)	< 15 for all stations	< 15 for 50% of stations
Water Discharge	< 15 for all stations	< 15 for 50% of stations
Three-dimensional Models		
Model Output	Target - Desired	Target - Acceptable
Velocity (vertical profile)	< 15 for all stations	< 15 for 50% of stations
Velocity (transverse profile)	< 15 for all stations	< 15 for 50% of stations
Coarse Sediment (vertical profile)	< 20 for all stations	< 20 for 50% of stations
Fine Sediment (vertical profile)	< 20 for all stations	< 20 for 50% of stations
Load for Coarse Fraction	< 20 for all stations	< 20 for 50% of stations
Load for Fine Fraction	< 20 for all stations	< 20 for 50% of stations
Vertical Salinity concentration profile	< 20 for all stations	< 20 for 50% of stations

7 Uncertainty Analysis

In the previous section of this report, guidelines for calibrating and validating the numerical models used in the MRHDM study were introduced. However, a well-calibrated and validated model may or may not be adequate to provide insights and support decisions regarding coastal restoration and protection strategies. Decision makers need to know the reliability of the predictions determined by the numerical models. As such, it is important to establish quantitative measures to define the word "reliable." For example, it is helpful to formulate the model predictions in the following manner: "The numerical model predicts, with 80% confidence, that a sediment diversion at River Mile "x", would build "y" acres of land after 50 years." It is also critical to provide assurance that these predictions incorporate environmental factors such as sea-level rise, subsidence, sediment loading, and runoff volumes. It is strongly emphasized that the models' prediction assumes a specific trend for these environmental factors. These trends are projections based on the best available science, and will be explored through scenario analysis and simulations. However, if such trends do not materialize in the future, the models' predictions would not be accurate and it would not be the responsibility of these hydrodynamic, salinity, and sediment models. Such statements cannot be made without a carefully designed uncertainty analysis. This section introduces an approach that would quantify the uncertainty bounds of the models' predictions.

In this section, we closely follow the uncertainty approach of Lall et al. (2002) that was also adopted in Louisiana's 2012 Coastal Master Plan*. Similar to the master plan, and as indicated previously, the MRHDM study relies on numerical predictive models to simulate the current conditions of the Lower Mississippi River, the delta, and the receiving basins. The models will also be used to analyze the future conditions under plausible variations in environmental conditions such as sea-level rise and subsidence. Further, the models will be used to evaluate the local and systemwide cumulative impacts of proposed restoration strategies over short- and long-term temporal scales. However, it is expected that such predictive tools include uncertainties. These uncertainties are defined as the disparity between model predictions and reality that result from error propagation through the model (Lall et al., 2002). Uncertainties result from: (1) in-

* http://www.coastalmasterplan.louisiana.gov/

formational uncertainties in specifying boundary and initial conditions, (2) structural uncertainties including errors in model structure and variability of observed values over different spatial/temporal scales than model, and (3) numerical errors in the model algorithm. There are numerous sources that contribute to these uncertainty components, including:

- Outdated, insufficient, inaccurate, or unrepresentative input data (bathymetry, topography, freshwater inflow volumes, sediment load, constituents load, etc.);
- Poor or incomplete knowledge of the pertinent physical processes represented in the predictive models;
- Approximations and numerical assumptions in the numerical schemes;
- Imperfect characterization of numerical and physical parameters in the formulations utilized in the models.

The data collection program of the MRHDM study is specifically designed to address the first and second bullets listed above. Although it is not expected to obtain a perfect set of field observations, the data collection program is the logical approach to minimize these uncertainty sources, especially if such a data collection program is maintained beyond the duration of this study. The study team adopted a multiple-models philosophy to help overcome and understand the limitations and constraints of the approximations and numerical assumptions (see third bullet). These intermodels applications will provide invaluable insights into the impact of these approximations and assumptions on capturing the spatial and temporal patterns of the pertinent physical processes studied herein. Hence, the focus of the uncertainty analysis will be on the fourth bullet from the list above, specifically the uncertainties in the models parameters.

As such, the proposed uncertainty analysis approach for the MRHDM study is founded on:

- Determining a clear set of performance measures (acres of land built, reduction of flood elevations, etc.) as indicators for the response of the natural system (Lower Mississippi River and its receiving basins) to restoration projects;
- Establishing the key model parameters (as reflected in the sensitivity analysis) with the strongest influence on the response of the Lower Mississippi River and its receiving basins to restoration projects;
- Identifying and analyzing the impact of uncertainties in specifying these parameters on the selected performance measures.

The term “parameter” refers to any model term that controls the relationship between model inputs (drivers) and outputs (response). Parameters could be a simple numerical value, table, or a mathematical relationship.

7.1 General Outline of the Uncertainty Analysis Approach

The approach outlined below intends to determine uncertainties of the key model parameters influencing the performance measures. The performance measures are identified as the prime factors for selecting the restoration strategies; that is, they are proxies to the project objectives. Quantifying the uncertainty boundaries associated with model predictions would allow decision makers to evaluate whether there is potential risk involved in a given restoration strategy. The approach adopted from Lall et al. (2002) is described below:

- **Step 1**: Determine the performance measures most appropriate as proxies for the project objectives. Examples of a performance measures include: (1) surface area of land created as a result of a sediment diversion, (2) shoaling/erosion in the river channel, (3) changes in currents (as a proxy of impact on navigation), and (4) change in flooding to nearby communities, etc. These are only examples; actual measures should be formulated and agreed on by the project team and managers in the early stages.
- **Step 2**: Identify the model outputs that impact the performance measures identified in Step 1. For example: velocities, stage, water discharge, sediment concentration, and erosion/shoaling quantities.
- **Step 3:** Perform a sensitivity analysis to determine which model parameters or functional relationships influence the model outputs identified in Step 2. The sensitivity analysis should result in a set of independent (uncorrelated) parameters. The modeling team should focus on the parameters that have the most impact on the model output identified in Step 2 to minimize the number of numerical simulations needed to complete the uncertainty analysis. It might be challenging to identify a set of parameters that are fully—in the formal statistical sense— uncorrelated. Examples of parameters that might result from this sensitivity analysis include roughness parameters, coefficients in the sediment transport formulations, coefficients in the morphological formulations, and numerical diffusion coefficients.
- **Step 4:** For each parameter selected in Step 3, the modeling teams should identify a range that the parameter is likely to fall. This range reflects the degree of uncertainty associated with each parameter. The

range should respect accepted values of these parameters in the literature.

- **Step 5:** Design an array of numerical simulations based on combinations of values assigned to the parameters identified in Steps 3 and 4 to reflect the specified parameter uncertainties. The combinations of parameter values should be selected to produce the widest possible range in the model outputs and performance measures identified in Step 1.
- **Step 6:** Based on the results of Step 5, the modeling teams (or an assigned analyst) would construct empirical probability distributions for each of the selected performance measures. These approximate probability distributions are used to:
    - Assess the impact of model uncertainties on the prediction of performance measures;
    - Evaluate whether—and how such—uncertainties can affect the selection of a specific restoration strategy or determine whether a restoration strategy is or is not feasible.

7.2 Detailed Methodology and Procedure

7.2.1 Selection of Performance Measures

The selection of the performance measures should be done through full participation of the technical teams, the study planners, and the managers. It is also understood that environmental factors such as various rates of sea-level rise and subsidence, will be investigated through dedicated numerical simulations. A tentative set of performance measures is included below. This list will be revised and refined:

- Surface area built in the receiving basins through sediment diversions;
- Change in flooding (flood depth) of nearby communities;
- Sediment erosion/deposition volume in the river channel;
- Change in velocities in the river channel (proxy to navigation interests);

These and other performance measures would be considered as objectives or targets. The performance measures would then dictate the critical and relevant model parameters. These parameters directly control the model output used to calculate the performance measures. However, the impact of these model parameters on the model output and subsequently on the performance measures is uncertain. As such, this analysis is designed to quantify the uncertainty bounds of the performance measures as propagated through the model parameters.

7.2.2 Critical Model Outputs

The modeling teams would be required to identify critical model outputs for each of the performance measures previously identified. The modeling teams should provide insight on the possible impact of these outputs on the performance measures (e.g., higher diverted sediment load increases surface area of land built). This is necessary to design the attributes of the uncertainty simulations, which will be discussed in subsequent sections. The preliminary list of model outputs that are believed to influence the performance measures (along with a list of model parameters affecting these model outputs) is provided in Table 4.

Table 4. Examples of key model parameters affecting the model output and subsequently influencing the performance measures. Note: sediment coefficients in the table are placeholder variables listed here for demonstration purposes only. They still need to be identified by the modeling teams.

Output	Parameter with Uncertainty
Stage	Bed Roughness
Salinity	Bed Roughness
	Diffusion Coefficient
Sediment	Settling velocity
	Sediment formulations coefficients
	Sediment substrate parameters
	Morphological parameters
Velocity	Bed Roughness
	Turbulence model parameters

As shown in the Table 4, there are parameters that influence stage, salinity, sediment, and velocity outputs. Specifying a value for each of these parameters is subject to uncertainty. In the model calibration and uncertainty analysis report, the modeling teams should provide synopses of each of these parameters, how a value has been assigned, and more importantly, how a range has been established for each of these parameters. Table 5 provides an example of outlining the range of each of these parameters, and it indicates the value that was identified during the calibration procedure to produce the best comparison with the available field observations (referred to here as the “Base” value). Some or all of these parameters are spatially variable. As such, these ranges should be established for each parameter value for each spatial zone or region. The five samples defining the range of each parameter will be utilized in designing the parameter combinations used in the uncertainty simulations.

Table 5. Example of the uncertainty range of the key model parameters that influence the models’ output.

	Parameter Setting				
Key Model Parameters	**Low**	**Medium Low**	**Base**	**Medium High**	**High**
Bed roughness					
Diffusion coefficients					
Sediment settling velocity					
Sediment formulations coefficients					
Sediment bulk density					
Turbulence model parameters					

7.3 Methods

The previous sections outlined the process of identifying: (1) a set of performance measures that are critical and can serve as proxies for the MRHDM objectives, (2) a set of key model parameters that affect the selected performance measures, and (3) a range for each of the key model parameters listed in item (1) above. These ranges are intended to capture the uncertainty in these parameters. Ideally, one should establish a full probability distribution of these key parameters. However, that might be impractical due to the complexity of the physical processes involved in this massive modeling effort. Key parameters are not precisely known, and as such, can potentially propagate uncertainty in the model output/predictions. Further, due to the large spatial and temporal scales included in this study, it may not be possible to perform sufficient numerical experiments to establish such a probability distribution. Hence, establishing a range and sampling within that range for each key parameter might be the practical and acceptable approach for the MRHDM study.

7.3.1 Design of Uncertainty Analysis Realizations

Next, a simulation experiment composed of multiple model simulations (referred to here as "realizations" to distinguish them as specific numerical simulations used in the uncertainty analysis). These realizations are based on sampling from the key model parameter values identified in Table 5. It is important to note here that values assigned to the key model parameters were selected such that the various combinations produce the widest possible range in the model output predictions. For example, parameters should be aligned to produce the most erosion in one uncertainty analysis realization, and in another case they should be aligned to produce the most accretion. This would produce the widest possible range of model output and provide insights regarding the possible uncertainty range.

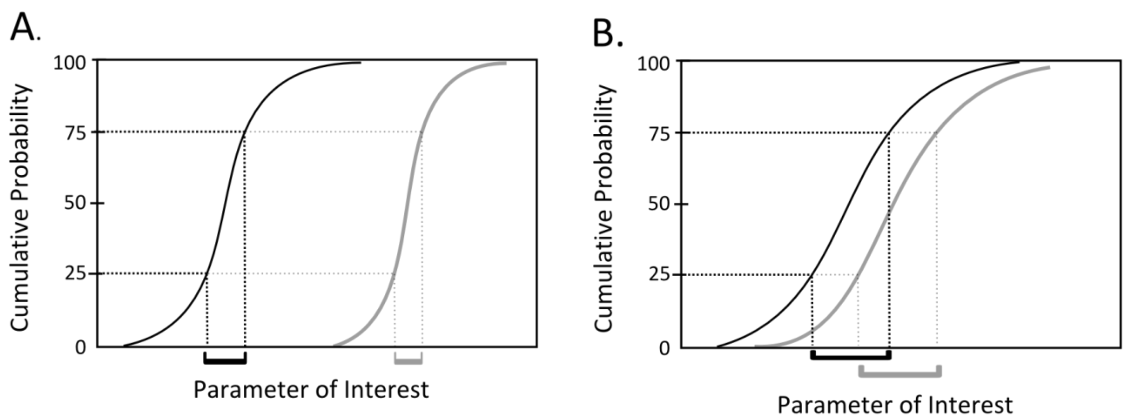
The performance measures identified here, for instance, sediment load through a diversion (or acres of land built as a result of diverting such sediment load) would be used to understand and quantify the uncertainty—or associated risk—of the models predictive ability. For example, the uncertainty realizations described above would produce values for such performance measures as the sediment load through a diversion. A reasonable quantity to focus on could be annual sediment load through the diversion. Analyzing the results of the five realizations described above over a span of 50 years would provide insights regarding the impact of model uncertainties on land building prediction as an example of the performance measures.

7.3.2 Construction of Cumulative Distribution Functions (CDF)

Continuing with the example of acres of land built, a cumulative distribution function is calculated from the results of these uncertainty realizations. Graphical analysis provides insight on the probability that the amount of acres built is less than a specific value. The probability of being greater (i.e., exceedance) than such a value can be easily calculated. However, the CDF analysis provides information far more important than the simple extraction of the probability of exceedance. For example, the width of each CDF curve provides a measure of the impact of model-induced uncertainties on the ability to predict a specific value of acres built. The CDF curve can also be used to provide the acreage of land built with a confidence range of 5%-95% or 25%-75%. It can also be used to determine the median value (50% probability in the CDF curve).

One of the central questions of decision makers pertains to the ability of the model(s) to show a discernible impact of a certain restoration strategy or whether that impact would be masked or shadowed by uncertainties. The proposed uncertainty analysis herein is specifically designed to address this and similar questions. An ideal case can been seen in Figure 2a, where the 25%-75% confidence bound for a hypothetical parameter of interest under one restoration option is completely separate from the confidence bound under the other hypothetical option. This is an ideal case where the impact of a restoration action was clearly not masked by model uncertainty. In Figure 2b, the confidence bounds of the two options overlap in this hypothetical scenario. In this case, the impact of the restoration action is masked by the model uncertainty.

Figure 2: Cumulative probability functions (CDF) for a hypothetical parameter of interest. The figure demonstrates scenarios in which model uncertainty does not mask impact of restoration action (left panel) and does mask (right panel).

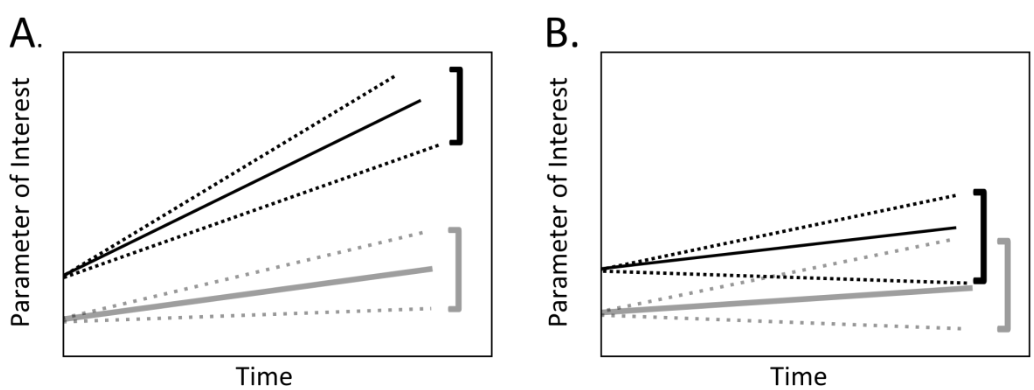


An alternative—and perhaps a simpler—indicator of the ability of numerical models to assess the impact of a restoration strategy is to determine the median of the distribution of the base case and cases with projects. It would be of value to identify if the median value of the "with project" falls outside the CDF curve of the base condition which would imply that the model prediction of the impact of a project is not masked by uncertainties. At times, the model uncertainty would mask the ability of the model to provide a discernible prediction. If the CDF curves of the 'with' and 'without project' exhibit excessive overlap, that would imply that the impact of a specific project or strategy is minor or the model is unable to provide a discernible prediction compared to the base case.

7.3.3 Temporal Variations of Uncertainty

The variability of the uncertainty bounds around the numerical models' predictions is of great interest to decision makers. To fully assess the feasibility of a restoration project or strategy, it is important to quantify its benefits not only at the present time but also in the future. As such, it is of value to generate CDF curves at incremental years in the future. For example, during a 50-year simulation, one can construct CDF curves at years 5, 10, 15… 45, and 50. Examining such curves would provide insights regarding whether the uncertainty bounds are stable or whether they grow as time passes. An important indicator that should be tracked is the temporal variation in the median prediction (50th percentile) and whether the "with project" scenario falls within or outside the uncertainty bounds of the "without project" case. Figure 3 shows how the uncertainty bounds might grow with time. Despite this growth, Figure 3a is an example of how the model prediction for one restoration action (black lines) was distinct from the other (grey lines). Solid line indicates the median; whereas the dashed lines reflect the 25 and 75 percentiles. It is critical to ensure that the model uncertainty are not large enough to partially or completely mask and limit the ability of the model to provide discernible impacts on the natural system from a restoration or protection strategy (Figure 3b).

Figure 3: Temporal propagation of uncertainties for a hypothetical parameter of interest under two restoration actions (black and grey lines). The display bounds represent the 25th and 75th percentiles (dashed lines) and 50th percentile (solid line). Figure 3a demonstrates distinct impacts between the two restoration scenarios while Figure 3b demonstrates partial overlap.



7.4 Closing Remarks

As discussed earlier in this report, the model prediction uncertainty in external model drivers (e.g., sea-level rise, subsidence, inflow sediment load, and water volume), is addressed through model simulations specifically dedicated to examine the response of the model output to such variability. In this report, the focus is on the uncertainty due to imperfect knowledge of key model parameters. The proposed uncertainty analysis would examine the impact of model uncertainties on the prediction of measures such as sediment load diverted from the Mississippi River. It would provide a reliable approach to assess the feasibility of various restoration projects/strategies, and it may provide the ability to rank or prioritize them. Although ideally the uncertainty analysis could/should be applied to all alternatives, due to budget and time constraints, it is recommended here to perform the uncertainty analysis on the FWOP and the Final Array of projects.

The outcome and benefit of the proposed uncertainty analysis is to inform decision makers on how model uncertainties affect the assessment and feasibility of proposed restoration and protection strategies. Ultimately, the proposed uncertainty analysis would provide insights as to whether the predicted changes in the natural system, considering various restoration strategies, are discernibly different from those under FWOP conditions. Such insights and information is quite valuable to decision makers to assess the viability of various restoration and protection strategies.

8 REFERENCES

American Society of Civil Engineers (ASCE). 2008. Sedimentation engineering: Processes, measurements, modeling and practice , ASCE Manuals and Reports on Engineering Practice No 110. Marcelo. H. Garcia, ed.

Haag, D. & Kaupenjohann, M. (2001). Parameters, prediction, post-normal science and the precautionary principle – A roadmap for modelling for decision-making. *Ecological Modelling, 144*, 45-60.

Habib, E. (2012). Model uncertainty analysis technical report. Appendix D26 in Louisiana's 2012 Coastal Master Plan. Coastal Protection and Restoration Authority: Baton Rouge, LA.

Hammons, M. & Shelden, J. (2012). Draft model acceptability metrics for numerical modeling of LCA medium diversion at Myrtle Grove with dedicated dredging, Plaquemines Parish, Louisiana. *Draft Technical Memorandum (Rev.).* FTN No. 3111-110.

Kleindorfer, G.B., O'Neill, L., & Ganeshan, R. (1998). Validation in simulation: Various positions in the philosophy of science. *Management Science, 44*, 1087-1099.

Konikow, L.F. & Bredehoeft, J.D. (1992). Ground-water models cannot be validated. *Advances in Water Resources, 15*, 75-83.

Lall, U., Phillips, D.L., Reckhow, K.H., & Loucks, D.P. (2002). Quantifying and communicating model uncertainty for decision making in the everglades. *Report of the Comprehensive Everglades Restoration Plan's Model Uncertainty Workshop, US Army Corps of Engineers and South Florida Water Management District*, West Palm Beach, FL.

Legates, D.R., & McCabe, G.J., Jr. (1999). Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. *Water Resources Research, 35(1)*, 233 -241.

Oreskes, N. (1998). Evaluation (not validation) of quantitative models. *Environmental Health Perspectives, 106*, 1453-1460.

Oreskes, N., Shrader-Frechette, K., Belitz, K. (1994). Verification, validation, and confirmation of numerical models in the Earth sciences. *Science, 263*, 641-646.

Roache, P.J., 2009. "Terminology and Definitions", Chapter II in *Verification and Validation of 3D Free-Surface Flow Models*. Environmental and

Water Resources Institute (EWRI) of the American Society of Civil Engineers (ASCE).

Rykiel, E.J., 1996. Testing ecological models: the meaning of validation. *Ecological Modelling, 90,* 229-244.