SANDIA REPORT

SAND2012-8989
Unlimited Release
Printed October 2012

An Initial Review of NE-KAMS Validation Data Quality Standards

V. Gregory Weirs

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Approved for public release; further dissemination unlimited.

Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

NOTICE: This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government, nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, make any warranty, express or implied, or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represent that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government, any agency thereof, or any of their contractors or subcontractors. The views and opinions expressed herein do not necessarily state or reflect those of the United States Government, any agency thereof, or any of their contractors.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: http://www.osti.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd
Springfield, VA 22161

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.fedworld.gov
Online ordering: http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online

SAND2012-8989
Unlimited Release
Printed October 2012

An Initial Review of NE-KAMS Validation Data Quality Standards

V. Gregory Weirs
Optimization and Uncertainty Quantification Department
Sandia National Laboratories
P.O. Box 5800, MS 1320
Albuquerque, NM 87185-1320
vgweirs@sandia.gov

Abstract

To meet the increasingly high expectations of fidelity and predictive capability of physics and engineering simulation tools, more stringent validation testing of such tools is required. One ubiquitous impediment to more rigorous testing is the lack of sufficiently detailed information on physical experiments. When experiments are not well characterized, the experimental data do not sufficiently constrain the numerical model and validation studies are inconclusive. The NE-KAMS validation data standards and requirements provide a framework to identify the information about an experiment that a computationalist will need in order to perform a definitive validation of a model of interest. In this work, the first assessments using the NE-KAMS standards and requirements are conducted. The objectives are to improve the standards and requirements and to identify the elements of an optimal assessment process.

Acknowledgment

We thank Steve Lomperski of Argonne National Laboratory for his tour of the MAX facility, for providing example data for this report, for his critical insight on local caffeine sources, and for his enthusiastic support of the assessment process.

Contents

1 Introduction . . . 7
1.1 The Current State of Validation Practice . . . 7
1.2 A Brief Introduction to NE-KAMS . . . 10
1.3 Objectives of Data Quality Assessment . . . 11
1.4 Objective of This Report . . . 12
2 An Assessment of ERCOFTAC Data . . . 13
2.1 Overview of the QNET-CFD Knowledge Base . . . 13
2.2 Assessment of Case UFR 3-30: 2D Periodic Hill Flow . . . 14
2.3 Remarks . . . 18
3 An Initial Assessment of the ANL MAX Data . . . 19
3.1 Experiment Overview and Objectives . . . 19
3.2 Initial Assessment . . . 20
3.3 Remarks . . . 25
4 Recommendations . . . 31
References . . . 33

Tables

1 MAX experimental data included in the assessment . . . 27

1 Introduction

As defined by the ASME and AIAA, 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" [1, 2]. In practice, validation most often involves a quantitative comparison between results of numerical simulations and measurements from a physical experiment. Despite the large amount of effort expended by experimentalists and modelers, such comparisons are often inconclusive because errors and uncertainties in the simulations and experiments are larger than the model accuracy requirements. That is, the model accuracy cannot be distinguished from other, known sources of uncertainty in the experiments and simulations, so conclusions about the sufficiency of the model cannot be drawn.

There are several contributors to this uncertainty that result from current practice of experimental science, as will be described below. While there are technical hurdles to be overcome, the biggest problem is that little of the information generated by and knowledge obtained from physical experiments is captured, documented, and disseminated to the broader community. The Nuclear Energy–Knowledge base for Advanced Modeling and Simulation (NE–KAMS) project seeks to address this shortcoming. NE-KAMS will provide a software system to catalog experimental and numerical benchmarks. A distinguishing feature will be that the information in the database will be assessed for its completeness from the point of view of model validation; the objective is to raise the standard of information reported about an experiment so that definitive conclusions about model validity may be reached.

The NE-KAMS document "Code Verification and Validation Standards and Requirements: Fluid Dynamics 1.0" [21], referred to as the "NE-KAMS Data Standards document" in this report, proposed a number of specific categories and completeness levels to guide the assessment of experimental validation (experimental) data. We describe initial assessments of two datasets following the Data Standards document. As the first such assessments, the emphasis is on the assessment process, strengths and weaknesses of the standards and the Data Standards document, and suggestions for improvements in the process and standards. One dataset is a good surrogate for both a legacy experiment, in that the the reported information is treated as final, and a discovery experiment, in that it is essentially documented as a journal article. In contrast, the second dataset is from a running experiment with a concurrent modeling and simulation effort.

1.1 The Current State of Validation Practice

As scientific computing has progressed, a number of different techniques for determining or estimating the error and uncertainty in simulation results have also been developed. Only a minimal summary of these techniques is given here, for the purpose of putting the current work in context. The definitive reference on the field is given by Oberkampf and Roy [16].

Numerical simulations are produced by software programs, or "codes," that compute approximate solutions to sets of partial differential equations or integro-differential equations. One reason

the solutions are approximate because the equations are discretized by some numerical method, which converts the differential equations into algebraic equations so they can be solved on a computer; the difference between the solutions to these two sets of equations is called numerical error. For the specific application, the boundary and initial conditions must be specified to the simulation code, and usually, submodels that describe material properties, constitutive relationships, and phenomena below the resolution of the simulation are also required. When the boundary and initial conditions are not known precisely, uncertainty is introduced into the simulation results because the code is solving a slightly, but perhaps critically, different problem than the one of interest. Finally, the submodels and the differential or integro-differential equations may not include all the important physical effects, or may not describe them properly; if solved without numerical error and with precise initial and boundary conditions, they still give an approximate solution to the physical system, and this error is called model error.

In scientific computing today, different techniques are used to estimate and quantify these different errors and uncertainties. Verification addresses numerical error, and uncertainty quantification (or sometimes uncertainty propagation) addresses input uncertainties in parameters that define, e.g., initial and boundary conditions and coefficients in constitutive models. In this report the focus is on validation, which is an assessment of the model error for a particular use. The model error is very difficult to estimate, because for problems of interest (those for which we rely on simulation codes to provide the solution) numerical simulations cannot be performed without numerical error or parameter uncertainties. The model error is estimated by comparing simulation and experimental results after accounting for numerical error, parameter uncertainties, and also errors and uncertainties in the experimental results. If any of these other errors or uncertainties are poorly quantified, the estimate of the model error is degraded.

Physical experiments are executed with a variety of objectives. One well-known purpose is the discovery of previously unseen physical phenomena, and the most famous experiments either uncovered a behavior that had not been observed before, or confirmed a previously unseen behavior that had been predicted by theory. Experiments to calibrate a model, such as determining the values of unknown model parameters, are much more common than discovery experiments. While quite useful in general, both of these classes of experiments are generally unsatisfactory for the purpose of model validation. The most important aspect of an experiment for validation is that the experiment is characterized completely and precisely. A complete characterization allows the inputs to numerical simulations to be specified without resorting to assumptions about the geometry of the experiment, the operation of the experiment, the initial and boundary conditions, and material properties. A precise characterization does not mean that there are no uncertainties in the experimental information, but that these uncertainties are are as accurate and as small as possible.

Most existing experimental data is poorly suited to validation. One reason, as mentioned above, is that most experiments have discovery, calibration, or other objectives. A second reason is that the most common way of encapsulating and distributing experimental information is in journal articles and technical reports; this mechanism works well for discovery experiments, and to lesser degree, calibration experiments. However, the amount of detailed information required for complete characterization of validation experiments overwhelms space limits in an article format, and the audience for such an article would generally be narrow, compared to articles describing dis-

covery experiments. Fortunately, computers and the internet provide better alternatives for storing and disseminating large amounts of data, and the next section will outline a proposal for software system to provide this specific capability.

The most successful validation activities are interactive collaborations between experimentalists and computationalists running numerical simulations. When simulations and validation experiments are concurrent and the participants frequently and actively exchange information, both the computationalists and experimentalists benefit from the synergy of the combined effort. It is essentially impossible to capture everything known or possible to learn from an experiment; in this sense, there is no truly complete characterization. However, interaction with computationalists identifies the most important missing information while the experimentalists are in best positioned to address the gaps with modified experiments or additional measurements. The impact of feedback between the experimentalists and computationalists is difficult to overstate, and equivalent information cannot be acquired from experiments that have been completed, or "legacy" experiments. For legacy experiments, information decreases as hardware is dismantled or upgraded, and the experimentalists move on to focus on different experiments. Gaps in legacy data cannot be addressed without reestablishing the experimental program.

The benefits of a "dynamic" validation process – one based on interactive collaboration between experimentalists and computationalists – points out another disadvantage of publications as the mechanism for distributing experimental data for validation. The time required to publish an article, from writing through submission and reviews to public availability, can be from months to years. Even if an experiment is maintained and running over the course of several publication cycles, the delay in feedback from computationalists that read the publications is a severe impediment to addressing gaps in the experimental data collected. While the internet is beginning to provide online journals with a more rapid review process and a shorter time from submission to publication, the focus is still on the traditional areas of research publishing and not on collaboration between experimentalists an computationalists or the dissemination of validation data.

While interactive collaborations are much more effective for model validation, in many situations legacy experimental data is the only option. A great deal of legacy data has been generated over decades, and repeating even a small fraction of the experiments in conjunction with simulation efforts represents an expensive investment. Some experiments cannot be repeated for environmental or public health reasons. Another reason that a less dynamic validation process must be considered is logistic: for small scale, single effects tests, collaboration is tractable to arrange, but larger scale experiments are more difficult to modify and adjust. They take longer to design and build and operate, involve more people, often pool funding across different organizations that may have different priorities, and are more likely to require build-to-order hardware or equipment, so the feedback cycle is necessarily slower. It is also worth mentioning that the numerical simulations can be slower to execute and analyze than the experiments, and can be responsible for a slower feedback cycle, but this is less common.

1.2 A Brief Introduction to NE-KAMS

A shortcoming of the current state of validation practice is that only a small fraction of the richness of data generated through physical experiments is captured, documented and disseminated. Experiments are expensive to conduct, and from a resource management viewpoint, there is a strong incentive to catalog the work done and the results obtained. As explained above, the current mechanism of journal articles and technical reports is reasonably successful for disseminating results of discovery experiments, but preserves only small fraction of the information generated and knowledge gained. At present, there is no widely adopted system for retaining detailed data from experiments that have been completed and dismantled or repurposed, and raw data from which article conclusions are drawn is inaccessible except in rare cases where an individual experimentalist has maintained it. The Nuclear Energy–Knowledge Base for Advanced Modeling and Simulation (NE–KAMS) project seeks to address this shortcoming, and in so doing provides a platform to enable and enhance Verification and Validation (V&V) of simulation codes for the extended nuclear energy enterprise.

The vision of NE-KAMS is described in several documents, including a high level requirements document [14] and a strategic plan [15]. The primary audience for the project is Nuclear Energy Modeling and Simulation, with an initial focus on computational fluid dynamics (CFD), and participation is expected from academic, national laboratory, regulatory, and industrial stakeholders. The NE-KAMS effort recognizes that existing data and knowledge bases, while useful and informative, are insufficient for rigorous model validation [7]. NE-KAMS seeks to improve on these efforts by developing and establishing standards and best practices for V&V and Uncertainty Quantification (UQ), establishing standards and procedures for assessing for verification data (test problems with known solutions) and validation data (physical experiments), and implementing a searchable, web accessible, access-controlled mechanism for maintaining and distributing this information.

One intended capability of NE-KAMS is to serve as a hub for collaboration between experimentalists and computationalists. Experimentalists would enter their data to NE-KAMS, either by providing links to their own systems or electronically transferring it to storage hosted by NE-KAMS, and computationalists would download it. Experimental data would also be assessed for its quality for use in model validation, and the assessment process would be somewhat similar to the review of journal publications. However, the data would be accessed from NE-KAMS by the reviewers, and when completed, their assessment reports would be uploaded to NE-KAMS for others to examine. The experimentalists might then respond to the assessment by providing additional information or modifying their experiment to address concerns raised; and the reviewers could then update their assessments.

At this point, NE-KAMS is at an early stage of development and an initial demonstration of the software and hardware infrastructure is being built. As such, this report on the assessment process does not assess the software aspect of NE-KAMS, just the process of a reviewer using the NE-KAMS Data Standards document to identify strengths and weaknesses in experimental data sets.

1.3 Objectives of Data Quality Assessment

Detailed information about an experimental is required for validation. The lack of data undermines comparisons with simulations – comparisons are unconstrained and the sufficiency of model can not be determined conclusively. If a model and the experiment are in conflict, detailed data is needed to determine the source of the conflict. The NE-KAMS Data Standards document provides a list of criteria to assess the completeness of information about an experiment. There are a number of categories and levels for each category, as listed below, but the key question the assessment seeks to answer is, how extensively can a computationalist (modeler or analyst) establish a model's sufficiency from the experimental data?

Some information is *necessary* for validation because without it, the analyst cannot specify the problem for the simulation code without making questionable assumptions, or assumptions about the values of inputs that are known to strongly affect the simulation results. Some *supplemental* information is not necessary to define the simulation, but is very useful for helping the modeler track down issues and inconsistencies between the the model and the experiment. Finally, some information is *peripheral* – while it might be useful in rare situations, it does not directly aid the analyst performing the validation of the model. Of course, these labels are points on a continuum, and are applied before validation has occurred and before any issues have been identified through validation. Nevertheless, expert judgement does provide some guidance. For each category the lower assessment levels corresponds to the necessary information, and higher levels to supplemental information.

A great deal of information about an experiment is relevant for validation, and the NE-KAMS Data Standards document [21] organizes this information into six attributes of the experiment: experimental facility, analog instrumentation and signal processing, boundary and initial conditions, fluid and material properties of the walls, test conditions, and measurement of experimental responses. These attributes were chosen to focus on information needed as input to the model, measurement of experimental responses from the model, and how measurements (both input and output) were processed by the experimentalist. For each attribute, there are four possible levels of completeness, with the lowest level indicating that minimal information is reported, and the highest level that extensive information is provided. The level indicates the completeness of the characterization, for the purpose of validation, for that particular attribute of the experiment. The Data Standards document contains more detail on the attributes and the type of information that corresponds to each level. The authors stress that, "A low completeness does not imply a poorly performed experiment, but rather, one for which limited information is reported" [21, p. 27].

One conceptual difficulty of the completeness levels is that the appropriate completeness level may depend on the fidelity of the physics modeling. For example, a simulation using a RANS model does not produce turbulent spectra as output, but simulations using LES models do. An experiment that provides measured turbulent spectra is helpful and improves the characterization for the LES model results, but the RANS model does not benefit from this information.

A similar difficulty, on which the Data Standards document is ambiguous, is the difficulty of rating the precision of the characterization of the experiment, because the sufficiency of the

experimental precision depends on many factors in the validation study. The principal factor is that validation is conducted in the context of an intended use of the model, and without knowing that intended use, no judgement on the sufficiency of the precision of experimental measurements can be made. This does not mean that the precision of measurements should not be estimated or reported; in fact it is critical for any validation study that they are, and the completeness levels reflect this. However, in this report the completeness levels will not consider the actual values of the precision.

1.4 Objective of This Report

The Data Standards document defines completeness levels and attributes for experiments, but they are as yet untried and untested. The main purpose of this report is to identify their strengths and weaknesses of the attributes and completeness levels and examine the assessment process. This is achieved by actually using the attributes and levels in assessments of two experimental data sets. The two datasets were chosen to represent different cases for the assessment process, one in which the data is viewed as complete ("legacy" data) and another in which the experiment is much less mature but continues to generate information. In addition, the first dataset provided the same types and amounts of information contained in journal articles.

The legacy dataset is case UFR 3-30, "2D Periodic Hill' Flow" [18] from the European Research Community on Flow, Turbulence, and Combustion (ERCOFTAC) QNET-CFD Knowledge Base [17]. Note that by "legacy" experiment, there is no suggestion that this data is obsolete or that interest in the data or specific flow situation has waned. The 2D Periodic Hill Flow has been studied for a couple decades, but the dataset examined is only three years old. The second experiment is the MAX experiment at Argonne National Laboratory, which provided prepublication example data. This experiment is just coming on line and has an associated modeling effort. The NE-KAMS software infrastructure for capturing the information from a running experiment is in an early stage of development; a key reason for working with the MAX experiment is because it provides example data to develop the NE-KAMS software system.

Since the NE-KAMS data standards (completeness levels and attributes) are new, no experiments have been documented with them in mind. As such, the assessments in this report may cast the example experiments in an artificially poor light. The information on the dataset from the QNET-CFD knowledge base, naturally, is targeted to meet the QNET-CFD reporting requirements, and these are quite different than the NE-KAMS data standards. The information on the MAX experiment is incomplete because the experiment is just coming on line; however, improving completeness levels as an experiment matures are anticipated by the NE-KAMS project. For these reasons, we emphasize the actual assessment levels assigned for these two datasets should be viewed somewhat skeptically in terms of representing all the information that could have been obtained with more effort. We also emphasize that the completeness levels are based on the detail of the information reported, and are not a judgement on the quality of the experiment or the abilities of the experimentalists; in fact, these datasets were chosen because they come from high quality experiments conducted by well regarded experimentalists.

2 An Assessment of ERCOFTAC Data

The European Research Community on Flow, Turbulence, and Combustion (ERCOFTAC) [4] hosts the QNET-CFD Knowledge Base [17], which is accessed through its wiki page. The QNET-CFD Knowledge Base contains test problems for computational fluid dynamics (CFD), and has evolved for approximately twenty years. In this section a basic description of QNET-CFD is given, and then a quality assessment of one dataset based on the NE-KAMS validation data completeness standards in the Data Standards document. Conclusions of the assessment on the actual test case should be taken with consideration of the context; as explained below, the objectives, and consequently the processes and criteria for quality, of the QNET-CFD Knowledge Base are different than those of NE-KAMS. However, the exercise of doing the assessment does provide valuable information on the NE-KAMS data standards and processes, which is the focus of this report.

2.1 Overview of the QNET-CFD Knowledge Base

The fundamental unit of the QNET-CFD Knowledge Base is a test case. The test cases are divided into two main groups, "Application Challenges" and "Underlying Flow Regimes." Application Challenges, or AC cases, are "realistic industrial test cases," while the UFR cases are broadly relevant across different applications. Within each of these groups, there are more specific subgroups. Also, there are two "Categories of Content" that distinguish the maturity and quality level of the test cases. A test case in the Gold domain "has been carefully checked and therefore satisfies high quality standards" and also includes a quality review as contributed by one or more reviewers. Test cases in the Silver domain are "less mature" and are "still under discussion and open for improvement." Some Silver test cases also get a Gold Star designation, that indicates that parts of the case qualify for the Gold domain. To access the Gold Domain test cases, a paid individual or institutional membership in ERCOFTAC is required.

Test cases are submitted by ERCOFTAC members. There are templates and guidance for putting information in the required formats. The information on each test case is categorized into the following sections. "Front Page" provides an abstract for the test case. "Description" includes the motivation and objectives for the test case and past work on the flow or similar configurations; references to published literature are provided. "Test Case Studies" specifies the problem in detail, including geometry and dimensions, initial and boundary conditions, and the metrics (system response quantities or SRQs) for comparison. It also includes a description of the physical experiment, its diagnostics, and the experimental SRQs. Finally, a description of the numerical simulations is given, including the code and the numerical method employed, any models required, the computational mesh and resolution studies, and the simulation SRQs. The "Evaluation" section discusses the comparison of the experimental and simulation results. Other details of interest that affect the comparison, or guidance to others that perform simulations of the test case, are in the "Best Practice Advice" section. Finally, there are a list of references and, for Gold cases, the reports from reviewers of the work. This structure is very much like a journal article, beginning with a literature search, moving through the details of what was done and making comparisons of experimental and numerical results, and then drawing conclusions. In many ways the "Quality

Review" provides referees' reports for public view when positive (the Gold cases.)

From this description it should be clear that there are some significant differences between QNET-CFD and what NE-KAMS proposes. NE-KAMS seeks to capture more detailed information about experiments, information well beyond that included in journal publications or technical reports. For validation purposes, NE-KAMS focuses on experimental data; numerical simulations may be included at some point, but for the purpose of illustrating good V&V practice rather than demonstrating the capabilities of particular models, methods, or codes. For QNET-CFD cases, however, simulations are central.

2.2 Assessment of Case UFR 3-30: 2D Periodic Hill Flow

The 2D Periodic Hill flow consists of a regular sequence of geometric perturbations on the lower wall of an otherwise canonical channel flow. The hills are smooth, with profiles defined by polynomials. While the geometry of the experiment is two-dimensional, the physics of the flow are not: the experiment is designed to study turbulent boundary layer separation from curved surfaces and subsequent reattachment at low and moderate Reynolds numbers. The flow configuration and this experiment in particular were designed for validation of turbulence models, but also to investigate the flow physics. A focus of earlier studies has been the relative performance of different subgrid (LES) models and different wall functions; separation and reattachment were found to be more sensitive to the latter. The work described in case UFR3-30 is heavily based on a journal article by the same authors [3], and the main objectives are to compare two different simulation codes (one for DNS and the other for LES) and new experimental data over a range of Reynolds numbers. The QNET-CFD case refers to this and other publications for greater detail. Unfortunately, the main reference on the physical experiments [19] is in German and difficult to obtain. In addition to the information on the wiki, a conference paper on the experiment [20] and the journal article [3] were considered for the assessment. No effort was made to obtain additional information from the experimentalists or the authors of the publications.

This Silver + Gold Star test case was chosen because it is one of the most recent cases in the Semi-Confined Flows category of Underlying Flow Regimes. At the same time, a history of similar experiments provides a depth of knowledge that informs and improves this experiment. The dimensions of the present experiment address problems that only became apparent after scrutiny of the earlier results. In particular, the distance between the hills was increased so that reattachment of the boundary layer occurs on a flat section of the lower boundary, rather than the windward side of the next downstream hill; the height of the channel and the Reynolds number of the flow were reduced to make simulations less expensive; and the spanwise extent of the experiment was extended so that periodic boundary conditions could be justified in that direction.

The rest of this section will proceed through the NE-KAMS attributes, and assign levels based on the information on the QNET-CFD wiki and some of the references therein. In the interest of time, not all of the citations were pursued, and this may result in lower assigned levels. The key point is the reasoning behind the assignment of the levels from the data and information used.

Experimental Facility: Level Zero

The functional operation of the experimental facility is minimally described overall, resulting in a completeness level of zero for the assessment. (The four levels are labelled zero, one, two, and three.) The path of the water through the facility is described, and a description of the design geometry is given; in particular, the location of the array of hills, and the long sections preceding and following the test section are described to justify the sufficiency of streamwise and spanwise periodicity assumptions of the test section. While the dimensions and locations of the flow straighteners in the part of the facility leading to the test section are provided, little information on the inlet and outlet reservoirs or on other flow conditioners is given. The only uncertainty information provided about the as-built geometry of the facility is a statement that the accuracy is within $0.01h$, where $h = 50$mm is the hill height.

Some information is reported that is more consistent with completeness level one. Each experiment is repeated six times, with the Particle Image Velocimetry (PIV) camera in a different location; this is done to avoid laser light reflections from the plexiglas test section walls. However, the camera locations and the viewing areas in the test section are only shown schematically. The variation of the water temperature is measured as part of each run, and this is used to derive the 1% uncertainty in the viscosity coefficient (which is temperature dependent) and subsequently affects the Reynolds number. The nominal flow rate is given, and the flow rate determines the Reynolds number of the flow; however, the range of actual Reynolds numbers across the six repeated experiments for each nominal Reynolds number cases is given in [3], information that can be used to infer the uncertainty in the flow rate or mean flow velocity.

The NE-KAMS Data Standards document lists calibration tests as useful information in the experimental facility section, but that assumes a scenario in which a wind tunnel is run with and without a model (e.g., of a vehicle or missile or ship) of interest. Then the calibration information consists of test data when the tunnel is run without the model, and the actual test data provides SRQs related to the model performance. For the 2D Periodic Hill Flow, there is no corresponding calibration analogy. Tests of streamwise and spanwise periodicity could be considered calibration data, but in this assessment they are classified as test conditions and discussed below.

Analog Instrumentation and Signal Processing: Level One

Several different diagnostics are used in the 2D Periodic Hill experiments. Pressure measurements are made with pressure taps in two sets of specified locations. To ensure the streamwise and spanwise periodicity of the mean flow, [3] seems to suggest that the experiment is run until the mean pressure variance is less than 1.0E-5 before the velocity data is recorded, although this is not entirely clear.

Mean and instantaneous velocities were measured using PIV. The most important specifications of the illuminating lasers are given, as are properties of the glass spheres used as the seed material. In addition, a procedure for combining data from six experimental runs to alleviate light reflections from the plexiglass walls is described, and some of the details of the PIV postprocessing algorithm

explain how "bad" velocity vectors are identified and what is used as surrogate data. The PIV results provide velocity vectors in a 2D plane, but at a limited number of points a laser doppler anemometry (LDA) was also used to measure the streamwise fluctuation velocity as a test of the PIV results [20].

The completeness level of the description of the PIV diagnostic is judged to be level one. While the information provided is helpful, the processing of image pairs to produce a field of velocity vectors is quite complex and an active field of research in and of itself, see, e.g. [12, 13, 5, 6]. No error estimates are provided for the main experimental responses (mean velocities and shear stresses.) For validation, more information in both of these areas is highly desirable. In isolation, the information on the pressure measurements and LDA diagnostics as described would be level zero, but these are supplementary diagnostics. The PIV system measures the experimental responses used for comparison to simulations.

Boundary and Initial Conditions: Level One

The design initial and boundary conditions are well described. The geometry of the hills is specified by several third-order polynomials [20] and the upper and side walls of the test section are flat plexiglas. Significant thought went into the design of the experiment to make the computationalist's job easier. In particular, the side walls are quite far apart and ten hills are regularly spaced along the bottom wall, so that in numerical simulations periodic boundary conditions may be applied in both the streamwise and spanwise directions. Furthermore, measurements were made and reported to provide evidence that these assumptions are satisfied. The mass flow rate is used to set the pressure gradient in the simulations in [3], and some measure of the uncertainty in this input is captured in the range of actual Reynolds numbers achieved for each nominal Reynolds number.

The data provided is sufficient for completeness level one. A number of improvements would be required for higher completeness levels. Regarding the geometry, measurements of the as-built facility would be required. The deviations from streamwise and spanwise periodicity would be useful, although these can probably be inferred from the data already provided, and it might be difficult for a computationalist to use that information without modeling the full span of the test section and the whole sequence of hills. For LES and DNS, it would also be useful to have a number of instantaneous velocity fields at the inflow and perhaps outflow planes to characterize turbulent spectra; this would allow comparisons of statistical information on the inflow between the experiments and simulations.

Fluid and Material Properties of the Walls: Level Zero

The main material used in the facility is polyvinylchloride (PVC), particularly the side walls and lower wall. For visual and diagnostic access, plexiglass is used for the cover (upper wall) and in the test section glass is used for the side walls. The hills are cast polyurethane, with a surface roughness of less than $1\mu m$. Epoxy resin was used to fill the joints between hills and PVC panels.

Few properties of the fluid are provided, but the water is a common material its properties should be easy to obtain by computationalists. The water is described as "filtered, decalcified, and chlorinated," and as the viscosity of water is known to be temperature dependent, the water temperature is continuously measured during each experiment.

A completeness level of zero is assigned for this attribute. Helpful additional information would include the location of the boundaries of the PVC, glass, and plexiglas and, as well as surface roughness measurements or information on each surface, not just the hills. While the seeding of the water with glass spheres (for the PIV and LDA diagnostics) may have a minimal effect on the properties of the water, the amount of seeding (mass or volume fraction) is not reported, nor is the effect, if any, on the glass spheres on the viscosity of the water.

Test Conditions: Level One

As described in the NE-KAMS Data Standards document, the Test Conditions attribute emphasizes detailed temporal variation of the conditions during an experiment. In this regard, the the only relevant information for the 2D Periodic Hill experiment is that the water temperature is continuously monitored during each experiment and the variation of less than 0.4K across experiments. This information underlies the reporting of the range of actual Reynolds numbers across experimental runs at a nominal Reynolds number [3]. Additional information that could have been included is the amount of time to establish steady state mean flow, and more detail on the temperatures and temperature variations for each run, rather than just the overall range across experiments.

Measurement of Experimental Responses: Level One

The experimental responses consist of mean streamwise and wall-normal velocities, and mean normal and shear stresses, i.e., $\langle U \rangle / u_b$, $\langle V \rangle / u_b$, $\langle u'u' \rangle / u_b^2$, $\langle v'v' \rangle / u_b^2$, and $\langle u'v' \rangle / u_b^2$, where u_b is the average streamwise velocity through the inflow plane (located over the peak of hill 7.) At each of four Reynolds numbers, these quantities are provided at approximately 60 locations in the wall-normal direction for 10 streamwise stations.

The NE-KAMS Data Standards document identifies spatial coverage and variety of experimental responses (SRQs) as important. In this experiment, essentially 2D data is provided (PIV data as measured in the streamwise-wall normal plane, reported at a number of streamwise and spanwise locations), and this is viewed favorably. The variety of responses is sufficient for completeness level one. Although the majority of the experimental responses are from PIV measurements, the few LDA results are both valuable as confirmatory evidence of the PIV data, and informative as a measure of the experimental uncertainty in at least one of the two techniques.

The information provided on the experimental responses could be improved in a number of ways. First, there is no uncertainty information provided. As approximately 10,000 image pairs were averaged for each of the six runs, there is some confidence that statistical noise is not an issue. However, the reported experimental responses are each a composite of the six runs and in

some sense refer to a single experiment, so it is not clear how repeatable the results are. As the mean velocity and stress data is primarily of interest for turbulence modeling, and in particular, LES and DNS, turbulence spectra or even the time dependent velocity fields would be valuable to computationalists and theorists. Note that this information is available from the data already collected, it is just not publicly accessible. Finally, the other diagnostic data (pressure tap and LDA velocity measurements) could also be made accessible.

2.3 Remarks

While there is some additional information on the QNET-CFD wiki (in particular the actual experimental responses are provided in tabular text files,) each case is essentially organized as a journal article. As such, this assessment points out the shortcomings of journal articles as documentation of validation experiments, and one can conclude that journal articles document experiments at completeness level one, at best. To the extent that legacy experiments are documented as journal articles, they suffer the same shortcomings, but technical reports and other forms of information retention could improve the completeness level.

Some attributes are more important than others. The most important attributes are "boundary and initial conditions," "fluid and material properties of the walls," and "measurement of experimental responses," because these are the inputs required by a simulation model and outputs that will be compared to the simulation results. Additionally, "analog instrumentation and signal processing," is almost always important for interpreting and understanding the experimental responses. It is difficult to imagine an experiment in which this information was not critical. On the other hand, for some experiments the "experimental facility" and "test conditions" attributes may be less important.

The amount of information to satisfy a particular completeness level also varies depending on the experiment. For example, in the 2D Periodic Hill flow, it is not clear how much information about the properties of water is really necessary or helpful beyond the temperature, purity, and viscosity. Similarly for the test conditions, if the flow rate and water temperature and their uncertainties are measured continuously and reported, what additional information would be useful?

3 An Initial Assessment of the ANL MAX Data

The MAX facility at Argonne National Laboratory (ANL) is just beginning to run experiments. All the information considered in this report has been obtained directly from the principal experimentalist, Steve Lomperski, following a visit to the facility [8]. The example data was provided primarily for the development of the NE-KAMS software system, and its use in this assessment was only a secondary consideration. The information provided is listed in Table 1. At this time the main public document on the experiment is a conference paper that is still in the publication process [9]; there are also a pair of internal reports and a draft of a second conference paper.

Since the information provided was intended for a different purpose than the assessment, for some of the attributes the experiment is not well represented. Obviously, if the experimentalist were asked for more information to address perceived gaps and responded with such information, the assigned completeness levels would be improved. However, no additional data was requested, and the assessment is based only on the information provided initially. For some SRQs, just a few data samples are provided, e.g., single images from a long time sequence, and the completeness level was assigned under the assumption that a complete time series was available. In general, more mature experiments will be better characterized and more information about them will be available. The NE-KAMS project anticipates that the completeness levels of running experiments will improve over time. The QNET-CFD Silver, Silver+Gold Star, and Gold domains explicitly provide a similar mechanism.

For these reasons, the assigned completeness levels are tentative. As mentioned earlier, the emphasis in this report is on the assessment process and the refinement of the attributes and completeness levels, rather than the levels assigned for the MAX experimental information. Finally, though the ability to exchange experimental and assessment information through the NE-KAMS software is not yet available, thought was given to this topic in the course of the assessment and is reflected in the remarks at the end of the section.

3.1 Experiment Overview and Objectives

The MAX experiment is intended to provide validation data for turbulent mixing and heat transfer in sodium cooled nuclear reactors. However, liquid sodium is opaque and difficult to diagnose, and air is used as the fluid instead. Water was also considered, but liquid experiments have greater infrastructure requirements and are more difficult to modify than gas experiments. The particular issue under investigation is "thermal striping," cyclic fluid/structure thermal interaction with the potential for thermal fatigue cracking and component failure in nuclear reactors.

The test section is essentially a 1m$\times$1.7m$\times$1m rectangular glass-sided box, with an aluminum base and lid. Two parallel jets enter the bottom of the test section from hexagonal channels, and either of the channels may be heated. The facility was designed to accommodate up to four inlet channels. The vertical jets interact and impinge on the lid of the test section, which is instrumented with a new diagnostic. A "thermal window" measures the surface temperature over a broad area

of the lid using a thermally sensitive paper viewed with an infrared (IR) camera. The flowfield between the jet entrance and the top of the test section is diagnosed with a PIV system. The channels are closer to one end of the rectangular base, and the air exits through a duct and filter on the lid at the other end of the rectangle.

The experimental facility is intended for validation from the outset. Also noteworthy is that numerical simulations were used in the design of the test section itself; that is, the feedback cycle between experimental and numerical results was initiated before the experimental facility was built.

3.2 Initial Assessment

Experimental Facility: Level Two

The operation of the MAX facility is well described. The facility design document ([11], artifact 46) describes the facility itself (formerly a containment for a research nuclear reactor), the principal diagnostics (PIV and thermal camera) in some detail, the auxiliary diagnostics (for flow rate, pressure and temperature of the inflow), the air handling system and heater, and finally, the data acquisition and control system. Particularly valuable is the identification of tradeoffs between various diagnostic performance objectives, with some specificity on how the expected flow conditions are mapped to technical requirements on hardware and operational parameters. The arguments behind the choices made provide at least initial uncertainty estimates, and these estimates can be checked after the experiment is running. For example, a subsequent facility startup report ([10], artifact 47) includes initial data on the flowrate, which oscillates somewhat more than expected because of feedback between the compressor and flow controllers. These two documents also include results from numerical simulations.

In light of the facility design and startup reports, much of the other information provided is considered supplemental. Several photographs of the test section while the experiment is running (artifacts 11-14) provide more of a working knowledge and feel for the scale of the experiment. Two versions of a movie (artifacts 15 and 16) show the locations of the PIV and IR cameras, the data planes for each, and sample data obtained. The virtual model of the air handling system helps the computationalist to understand all the processing steps of the fluid before it reaches the hex channels, although the image (artifact 3) would be more useful if the hardware components were labeled. (We recognize the equipment from the visit to the facility.) A few of the artifacts provided were not obviously useful, for example the circuit diagram for the laser safety shutoff (artifact 7) was not clearly relevant to a computationalist modeling the experiment; but there are also no negative consequences of including them.

Additional information might be useful for a computationalist. Some examples include measurements of the as-built geometry, and details on any flow straighteners, flow conditioners or diagnostics (e.g., hot wire anemometers) in the inlet channels. It would be useful to know how long it takes to establish a steady state mean flow in the test section, and how the temperature and, e.g., the humidity, in the test section change as heated, dried air from the inlet channels displace the air at (presumably) ambient conditions in the containment room. As turbulent heat transfer is

a central objective of the experiment, the thermal startup transient (in contrast to the more obvious fluid motion startup) is an important consideration.

Analog Instrumentation and Signal Processing: Level Two

The diagnostics were also well described, and most of the artifacts in Table 1 document the instrumentation and postprocessing. The facility design and startup documents provide information on the diagnostic hardware, software, and performance, as well as the rationale for choices made. Data sheets on the glass of the test section (artifact 8), the polyethylene glycol used as the seeding material (9), and the paper used for the thermal window (10) provide the technical performance characteristics relevant to the instrumentation.

Artifacts 17-27 contain calibration, testing, and certification reports from the manufacturers of many of the instruments; these are specific to the actual hardware delivered, as identified by model number but also serial number. In most cases error estimates or measurements are provided in these tests, and these may be of immediate interest to computationalists. Artfiacts 28, 29, and 31-40 are user's manuals for the instrumentation hardware and software. These describe how the instruments and software should be set up and calibrated, as well as data acquisition and analysis. LabVIEW software is used to control the experiment and the diagnostics; in the software, the operator establishes "virtual instruments" with particular settings that correspond to the physical instruments. Artifacts 42-45 are LabVIEW virtual instrument files which contain the specific settings for the MAX experiment.

The level of reporting required for completeness level three are quite high. If the postprocessing were completely described, one would be able to start with, e.g., pairs of images of the illuminated flowfield, and independently produce the instantaneous velocity field reported. The uncertainties and errors in the raw measurements of the diagnostics, in this example the pixel intensities of the image pairs, must also be estimated with significant justification that all error sources have been identified. This allows one to propagate uncertainties in the raw measurements through the processing steps to the reported system response quantities (i.e., the experimental responses.) The objective is to provide all of the information used to document the error estimates of the SRQs.

Dr. Lomperski also provided a draft of a paper on the tradeoff between the accuracy and the field of view (FOV) of the PIV diagnostic (artifact 63), and most of the data on which the paper is based. PIV is conceptually a simple technique, but requires a highly complex acquisition and processing system with a large number of factors that contribute to the accuracy of its results. While the tradeoff between the nominal spatial resolution and the FOV is straightforward to assess, the accuracy vs. FOV is not. The FOV study is essentially a research paper on one aspect of the PIV diagnostic technique, as opposed to a calibration test of the diagnostics per se, but it provides a good example of the information required for completeness level closer to three. Similar amounts of information are provided for both the PIV technique and the independent LDV diagnostic used as a reference. Raw data for the LDV is included, as are files containing the software and hardware settings and test conditions (artifact 64, a directory containing approximately 400 files). For the 200mm FOV case, artifact 65 contains a set of images establishing the length scale for the raw

images; artifact 66 contains raw data images; artifact 67 contains files from the PIV processing software that document the processing settings and steps; and artifact 68 contains intermediate and final results from the processing, including the instantaneous and final velocity fields. These four artifacts contain approximately 180 different files. Finally, artifact 62 is a spreadsheet containing all of the processed LDV and PIV data (for each of the four FOVs examined), consisting of mean velocities and their standard deviations as a function of distance from the inlet, for several experiments for each case. The spreadsheet also contains the plots of this data that went into the draft paper. This FOV study, including the other artifacts on the diagnostics described earlier, represents the highest level of completeness observed so far. The main way it could be improved would be to provide more information on the specific algorithms used for processing the image pairs to obtain the velocity vectors, in addition to the software processing information (artifact 67).

The FOV study demonstrates that the highest completeness level is achievable. For the experimental responses to achieve this level, the same types of information would be required for the actual experiments that produce the experimental responses. That is, the FOV study provides completeness level three information for the FOV study results, but a lower level of information for other experimental results at the facility, which are measured at different conditions.

Boundary and Initial Conditions: Level Zero

For the test section, the velocity boundary conditions on the glass walls are straight forward. However, the wall temperature, which provides the thermal boundary condition, is not, and this will have to be monitored as the experiment runs since heated air is provided by one of the inflow channels. There are also questions about the lid and the thermal window. The aluminum has been removed where the thermal window is located, so that it is visible to the IR camera. Consequently the computationalist will have two different thermal boundary conditions to apply to the different sections of the lid. While the aluminum has significant heat capacity compared to the air and should hold a quasi-steady temperature, the NOMEX paper that comprises the thermal window is very thin (50μm) and backed by air, so it is not clear how the computationalist should specify the temperature boundary condition of the window. Because it is so thin and relatively fragile, it is not clear how flat it remains during the experiment or across experiments. Finally, more detail is needed on how smooth or rough the window material is and the smoothness of the transition from the window to the aluminum lid at the boundary of the window, as these could influence the fluid and thermal boundary layer along the window and lid. The facility startup document [10] notes that an alternative window consisting of a 1μm film backed by a sapphire plate is being considered, and at least from a boundary condition perspective, the sapphire plate would be easier to to represent.

The inflow velocity conditions from the two jets are not well characterized, but this information would not be difficult to obtain from the existing diagnostics. In this regard, artifacts 50-53 provide examples of measured velocity data. With a sequence of measurements, profiles of the mean and fluctuating velocities can be obtained, or other quantities as needed by the computationalist such as the turbulent kinetic energy for RANS simulations. Although no data was provided, temperatures and pressures are also measured in the inlet channels. The computationalist would benefit from

measurements of the temperature fluctuations as a function of space and time in the inflow plane; however, it is not clear that such measurements are possible. At the outlet, the temperature and pressure are measured, although not provided here.

For this nominally steady state experiment, a minimal description of the initial conditions is the steady state mean velocity field. Mean fluctuation velocities would be even more helpful. Without this data, the computationalist must run a simulation from cold start through the transient development to the steady state, much as the experiment is conducted. However, the cost of the simulations is much much higher to go through this process than to begin with very nearly steady state conditions.

While the data provided is assessed at completeness level zero, the example data gives confidence that for the flowfield, level two could be achieved if time dependent velocity measurements were made near the inlet and at least mean velocity data were provided throughout the test section. The main concern is the characterization of the thermal environment, with little information on the heat transfer through the test section walls and the boundary conditions at the thermal window a particular concern. Note that the hex channels of the inlet extend into the test section by 136mm, and the walls of the channel are 3mm thick. These features can be difficult or expensive for a computationalist because the fine geometric scales must be resolved in numerical simulations.

Fluid and Material Properties of the Walls: Level One

Artifacts 8-10 provide material data on the glass of the test section, the seed material for the PIV system, and the NOMEX paper used in the thermal window. The source is the manufacturer, so independent confirmation of the most important properties for the experiment would be comforting. No data is provided on the aluminum used for the base and the lid of the test section. On the air, the most important properties are widely available, and the air handling equipment limits particulates, oil, and moisture content [10]. However, a characterization of the combined air/polyethylene glycol mixture with regards to viscosity and heat transfer parameters would be helpful to establish the veracity of the novel thermal window measurement technique.

Test Conditions: Level Zero

The test conditions category focuses on the conditions of the test section as a function of time. For steady, turbulent, laboratory flowfields, this aspect is often a minor concern, compared to experiments in unconfined environments or in facilities that are transient in nature (such as blowdown wind tunnels.) However, a key objective of the MAX experiment is the heat transfer to the thermal window, and from a thermal environment perspective, the tank no longer encloses the experiment. In the preliminary testing of the facility, the temperature difference between the two channels is 10K, but it is not clear if future experiments will be restricted to this difference. While one channel provided heated air, the temperature of the test section at the time the data is taken is critical – because there is some start up time to establish a steady flow, the air in the tank is probably at some intermediate temperature between that of the the hot and cold air entering from the channels.

Also, there is some small but possibly important heat transfer through the glass and aluminum walls of the test section, as well as through the thermal window itself. The temperature at the inlets and outlet are measured and are important to report. Test section wall temperatures would also be helpful, as would temperature measurements at several locations outside but nearby the tank.

The temperature difference and thermal environment are not expected to affect the velocity field very much. However, they introduce uncertainty into the conditions of the experiment. The computationalist will have a difficult time determining if differences in the results are due to the model itself or to a mismatch between the experimental and simulation conditions.

Measurement of Experimental Responses: Level Two

Since the MAX experiment is just coming on line, the amount of data for experimental responses is quite small; examples of the experimental responses were provided, but not full time sequences. For the PIV system, a pair of images (artifacts 48 and 49) and the corresponding snapshot of the velocity field (artifact 52) were provided; the mean velocity field was also provided as a velocity vector image and as a text file (artifacts 50 and 51.) Finally, a flooded contour image of the vertical velocity was given (artifact 53.) For a time sequence of images and corresponding instantaneous velocity vectors, the PIV data would be assessed at completeness level two if error estimates were provided for the velocities. The LDV system provides a second diagnostic to independently confirm the PIV measurements; data from the field of view study is taken as an example of the information that would be provided for the experimental responses.

For the 2D Periodic Hill flow, the experimental responses were profiles of mean velocities and stresses at a number of streamwise stations on the centerline of the channel. The experimentalist might consider providing similar profiles above the inlet channels for the MAX experiment; however, this would be just a convenience to the computationalist. From the instantaneous velocity fields, a number of different turbulence quantities can be computed including the mean velocities and stresses, so the velocity fields are more valuable in this regard; this contributes to the level two assignment.

The second class of experimental response is the temperature data from the IR camera and thermal window. Example data included images of the average temperature and the standard deviation of the temperature (artifacts 54 and 55), a text file with the average temperatures (artifact 56,) and finally, a movie of the instantaneous temperature measurements (artifact 57.) These artifacts were for the case in which the channel closest to the wall was heated; similar data was provided for the case where the channel closest to the center of the tank was heated (artifacts 58-61.) The temperature data is assessed at level one, for several reasons. First, there is no estimate of the error associated with the reported temperature fields. Second, the mean temperature is probably reliable, but the instantaneous temperature fields and the standard deviation of the temperature may have large uncertainties. As noted in the facility design and startup documents, the response time of the thermal window is 1Hz, two orders of magnitude less than the PIV system, principally because of the response time of the NOMEX paper. This response rate is probably not sufficient to capture the turbulent temperature fluctuations of the flow. However, with suitable characterization of the

NOMEX paper response, it might still provide useful validation data.

3.3 Remarks

The intent of the experiment is quite important and sets the context for the completeness levels. In particular, the intent of this experiment is to investigate turbulent mixing and heat transfer in liquid metals. However, the experiments are conducted with air for a variety of reasons. In this report, the assessment is from the view of investigating turbulent mixing and heat transfer in air, and it is the consumer of the assessment that must decide if air is a sufficient surrogate for liquid metal applications.

The thermal window is a new diagnostic, and there are many issues to be explored before its accuracy can be established. Therefore it will take some dedicated effort and time to reach a high completeness level for the temperature measurements on the lid of the test section. As for a new facility, the accuracy of and the refinement of operating procedures for new diagnostics improve as more experience is gained through testing and use.

The thermal environment raises many sources of uncertainty that may be small, but are difficult to quantify. Such sources of uncertainty are not usually associated with turbulence experiments, but a primary objective of the facility is to investigate phenomena that lead to thermal striping. The lack of clarity on the thermal environment and the introduction of new diagnostic technique are the sources of almost all the issues leading to low completeness levels (boundary and initial conditions, test conditions, and fluid and material properties of the walls.)

The amount of data generated by modern diagnostics can be overwhelming, and while NE-KAMS emphasizes that previous data capturing, documentation, and dissemination have been insufficient, it will not be possible or desirable to keep all the data produced by an running experiment. This leads to several remarks. First, data should be captured in an automated way if at all possible; in this case "data" refers to all the different types of artifacts associated with a particular experimental run, not just the primary diagnostic measurements. Information provided manually is less likely to be captured consistently. Second, to fully support running experiments with automated data capturing, NE-KAMS software will have to support some type of complementary automated uploading process. Third, NE-KAMS software will require a content management or version control system so that different datasets and information from the same experimental facility can be distinguished. Any changes to the set up of the experiment, i.e., cameras repositioned, different seeding material used, parameters in the postprocessing algorithms changed, etc., must also be tracked in this system. (This motivates the automated capture of information, first point.) Clear tracking and identification of which files pertain to each specific experiment are critical for consumers of the datasets (computationalists and reviewers.) Fourth, some criteria for prioritizing the information retained will have to be developed. For an experiment with a concurrent simulation effort, the associated computationalists will provide critical guidance, but different models and different codes may require different or additional information.

Artifacts from postprocessing software are welcome, but complete descriptions of the algo-

rithms that the software implements are also needed. Software artifacts are valuable because they document how the data was processed, even if the processing was not what the experimentalist intended. A different and perhaps more important use, though, is that with the software artifacts one could in principle analyze the data independently using the same software, and vary different software settings (parameters in the postprocessing algorithm) to examine their effects. Eventually, though, the software and the platform it runs on will be upgraded and the analysis will not be repeatable from just the software artifacts. This is why a complete description of the postprocessing algorithm is desirable. A potential barrier to such a description is that the software may be proprietary, and this raises the question of whether or not a high completeness level can be obtained if the postprocessing algorithm is not accessible.

Table 1: MAX experimental data included in the assessment

Artifact	Filename	Description
1	paper 12060.pdf	conference paper
2	cell 5 assy 8-12-08.pdf	virtual 3D model of the facility including control room
3	AIR_SYSTEM_6.jpg	virtual image of the gas handling system
4	Control room.pdf	floor plan of the control room
5	Controls Schematic.ppt	schematic of the controls of the experiment and diagnostics
6	Hex can cross section.emf	design document for inflow channel cross section
7	Laser Sentry Electrical3.ppt	schematic of the safety controls of the laser (PIV illuminator)
8	borofloat_33_e.pdf	manufacturer's technical data sheet for glass for test section walls
9	Carbowax PEG 400.pdf	material safety data sheet for PIV seeding material
10	Properties of NOMEX paper.pdf	manufacturer's technical data sheet for thermal window material
11	Front view.jpg	photo of test section
12	Flow with probe and camera.jpg	photo of test section
13	Flow with LDV.jpg	photo of test section
14	Equipment only.jpg	photo of test section
15	Fishtank movie 1080 QA A.avi	animation of running experiment, including PIV and IR data
16	Fishtank movie 1080 QA B.avi	animation of running experiment, including PIV and IR data
17	Agilent multimeter 1 of 2.pdf	multimeter factory calibration measurements
18	Agilent multimeter 2 of 2.pdf	multimeter factory calibration certification

Artifact	Filename	Description
19	Omega fine wire conformance.pdf	manufacturer certification of conformance for temperature measurement
20	Promass Coriolis.pdf	manufacturer flow meter calibration report
21	Rosemount reference transmitters.pdf	manufacturer calibration report for pressure transmitter
22	Setra pressure transmitters.pdf	manufacturer calibration report for pressure transmitter
23	Silver 660 9-22-10.pdf	manufacturer test and calibration report for IR camera
24	Silver 660 Infrared.pdf	manufacturer test and calibration report for IR camera
25	Teledyne HFC 307 2012.pdf	manufacturer test report for flow meter
26	Teledyne HFM 2009.pdf	manufacturer test report for flow meter
27	Teledyne HFMD301.pdf	manufacturer test report for flow meter
28	DL004U-O Cirrus User Manual.pdf	User's Manual for IR camera analysis software
29	DL005U-BN Cirrus Software Interfaces.pdf	Manual for IR hardware/computer interface
30	Sample flow and temperature data from LabVIEW.xlsx	spreadsheet of time dependent flow meter readings from LabVIEW
31	DynamicStudio Manual.pdf	data acquisition and analysis software manual
32	LV_Fundamentals.pdf	data acquisition and analysis software manual
33	LV_Getting_Started.pdf	data acquisition and analysis software manual
34	LV_Quick_Reference.pdf	data acquisition and analysis software manual
35	LV_Release_Notes.pdf	data acquisition and analysis software manual
36	LV_Upgrade_Notes.pdf	data acquisition and analysis software manual
37	Hardware.pdf	LDV hardware documentation
38	Installation.pdf	LDV hardware documentation
39	Software.pdf	LDV software documentation
40	TR-TM_Fiberoptic_Probes.pdf	LDV hardware documentation
41	200 mm day 1.dat	LabVIEW output data (same as 30 but in raw form

Artifact	Filename	Description
42	Flow control.vi	LabVIEW virtual instrument file
43	log data.vi	LabVIEW virtual instrument file
44	MAX globals.vi	LabVIEW virtual instrument file
45	thermocouples.vi	LabVIEW virtual instrument file
46	MAX Design Rev0.pdf	Experimental facility design report [11]
47	MAX startup tests.pdf	Experimental facility initial operation report [10]
48	Export.2rknkeee.000000a.jpg	first raw PIV image
49	Export.2rknkeee.000000b.jpg	second raw PIV image
50	Average.2rknl1rj.000000.emf	average velocity vector plot (xz pland PIV)
51	Average text.2rknl1rj.000000.txt	average velocity vector data file (xz plane PIV)
52	Vector snapshot.000000.emf	instantaneous velocity vector plot (xz plane PIV)
53	image003.jpg	average z-velocity magnitude image (xz plane PIV)
54	Hex1 24.6 Hex2 34.7 893 lpm 10 Hz AVG.emf	average temperature, image on lid (xy plane IR)
55	Hex1 24.6 Hex2 34.7 893 lpm 10 Hz RMS.emf	std. deviation of temperature, image on lid (xy plane IR)
56	Hex1 24.6 Hex2 34.7 893 lpm 10 Hz_export.asc	average temperature data on lid (xy plane IR)
57	Hex1 24.6 Hex2 34.7 893 lpm 10 Hz 120 s scale 26.5 to 29.0.avi	instantaneous temperature, movie on lid (xy plane IR)
58	Hex1 34.7 Hex2 24.8 893 lpm 10 Hz AVG.emf	average temperature, image on lid (xy plane IR)
59	Hex1 34.7 Hex2 24.6 893 lpm 10 Hz 120 s RMS.emf	std. deviation of temperature, image on lid (xy plane IR)
60	Hex1 34.7 Hex2 24.6 893 lpm 10 Hz_export.asc	average temperature data on lid (xy plane IR)
61	Hex1 34.7 Hex2 24.6 893 lpm 10 Hz 120 s scale 26.5 to 29.0.avi	instantaneous temperature, movie on lid (xy plane IR)
62	Accuracy vs FOV-1.xlsm	Spreadsheet of FOV study data and results
63	ATH12 San Diego draft 2.doc	Draft paper on FOV study
64	LDV raw data	Directory containing raw LDV data and LDV settings
65	200 mm FOV length scale	Directory of images with ruler to establish image length scale

Artifact	Filename	Description
66	200 mm FOV raw data at 700 Hz	Directory of raw PIV images
67	Dynamic Studio project files without images	Directory of PIV software analysis files
68	Sample processed images from within DS project file	Directory of velocity field data obtained from PIV software, including instantaneous and averaged velocity vectors and partially processed images with intermediate results

4 Recommendations

The two assessments performed, while not exhaustive, were useful for exploring the utility of the NE-KAMS attributes and completeness levels for validation data. Overall the attributes and completeness levels are well thought out and provide good summary metrics on the characterization of an experiment for the purpose of model validation. Nevertheless, there are situations in which the NE-KAMS Data Standards document [21] is ambiguous or gives the reviewer little guidance on how to assign a completeness level, and these are listed below. Some of these issues have been raised in the Remarks sections of the preceding chapters but are repeated here for a more cohesive summary.

- Adding an attribute or other category to the NE-KAMS data standards for the intent of the experiment should be considered. The intent of the experiment defines the context for the assessment. This would clarify, e.g., that the experiment is intended for validation of RANS models, rather than LES models, and consequently the reviewer could assign completeness levels for the signal processing and experimental responses appropriate to RANS models.

- The completeness levels 0–3 are best thought of as exponents of 10 of the amount of data provided, rather than points on a linear scale. That is, the difference in the amount of information provided to reach level two versus level one is about an order of magnitude, rather than double. Of course, the assigned levels are qualitative and no rigorous measurement of the number of bytes was performed to test this metric, but the amount of data involved in specifying a time series of instantaneous velocity vectors in a plane is significantly greater than that specifying the average velocity vectors, or the average stresses. For some attributes, or depending on the particular experiment, this trend does not hold, but for the experimental responses and signal processing it seems inescapable.

- Some attributes are less meaningful or applicable depending on the experiment. For the 2D Periodic Hill flow, we struggled to imagine what data would improve the test conditions information beyond an temperature vs. time data for each experimental run. Similarly, not much information would be needed to improve the completeness level for the fluid and material properties of the walls attribute. Finally, for the air in the MAX experiment and the water in the 2D Periodic Hill flow, there doesn't seem to be much to add beyond values and uncertainties for the purity level, relative water content of the air, temperature, and amount of seeding material for the diagnostics.

- It is difficult to identify *all* the information about an experiment that is critical. Most of it can be identified, but a particular item of information is only recognized as critical when it is needed but not available. A concurrent modeling and simulation effort is invaluable for identifying critical information, but is not foolproof.

- When there are multiple experimental responses, it might be helpful to allow the reviewer to assign a different completeness level to each. The particular motivating use case is two different diagnostic techniques measuring two independent (used loosely) responses, such as the PIV and IR diagnostics for the MAX experiment. The PIV diagnostic is well established,

while the thermal window is much less mature. A single, overall completeness level should still be assigned for the attribute (experimental responses, and/or analog instrumentation and signal processing,) because the diversity of responses contributes to a higher completeness level.

This work has also provided insight into the process of executing assessments. Both assessments were carried out by a single reviewer. For the 2D Periodic Hill assessment, there was no contact with the experimentalists or those involved with the corresponding simulations, and to a large degree, legacy experiments would be assessed in the same way. While the MAX experiment is on going and in principle allows for a different process, it was actually treated in the same way as the 2D Periodic Hill flow because of the early development status of the NE-KAMS software system. Nevertheless, significant thought was given to a more dynamic assessment process enabled by this software system, and some comments on desired NE-KAMS software system capabilities are given at the end of Section 3. Independently of these software capabilities, experience gained from conducting the assessments identifies some important points about the process and the reviewers.

- Better characterized experiments are more difficult to assess and require more time to assess. On the latter, more data is provided for better characterized datasets so there is more for the reviewer to consider. At the same time, identifying critical or helpful information that is not present is more difficult because the most important and obvious information has already been provided.
- In some previous NE-KAMS discussions, the journal article review process has been used as an analogy for the dataset assessment process. This analogy is useful because most experimentalists and computational scientists are familiar with the reviewing journal articles, but there are several differences. One difference is that reviewing a dataset can take more time than reviewing a journal submission. A second is that the intent is different; a journal article above all prioritizes new ideas, while the purpose of validation datasets is to provide a rigorous reality check (in a literal sense) to a physics or engineering model in numerical form. So, aspects of the journal article review process should be carried over to dataset assessment process, but not without understanding these differences.
- Multiple reviewers for each experiment are strongly recommended. Having multiple reviewers increases the chances that critical missing information is identified. As the consumers of the experimental data, there is a strong motivation for computationalists as reviewers, but an independent experimentalist's perspective provides valuable diversity and complementary technical expertise.
- For running experiments, the timeliness of assessments is important. The sooner experimentalists receive feedback, the sooner they can address gaps in the information captured. Providing partial assessments as they become available, rather than releasing an assessment report only when all attributes have been assessed, should be considered. One source of the delay in the journal publication process is that reviewers are essentially uncompensated volunteers, so reviews are rarely a high priority activity. Some form of accountability for timeliness on the part of reviewers is highly desirable, at least for running experiments.

References

[1] Guide for the verification and validation of computational fluid dynamics simulations. Technical Report AIAA-G-077–1998, American Institute of Aeronautics and Astronautics, Reston, VA, 1998.

[2] Guide for verification and validation of computational solid mechanics. Technical Standard V&V 10–2006, American Society of Mechanical Engineers, New York, NY, 2006.

[3] M. Breuer, N. Peller, Ch. Rapp, and M. Manhart. Flow over periodic hills – numerical and experimental study in a wide range of Reynolds numbers. *Computers & Fluids*, 38:433–457, 2009.

[4] European research community on flow, turbulence, and combustion (ERCOFTAC). `http://www.ercoftac.org/`.

[5] C. J. Kähler, S. Scharnowski, and C. Cierpka. On the resolution limit of digital particle image velocimetry. *Exp. Fluids*, 52:1629–1639, 2012.

[6] C. J. Kähler, S. Scharnowski, and C. Cierpka. On the uncertainty of digital PIV and PTV near walls. *Exp. Fluids*, 52:1641–1656, 2012.

[7] H. B. Lee. A survey of existing V&V, UQ, and M&S data and knowledge bases in support of the nuclear energy – knowledge base for advanced modeling and simulation (NE-KAMS). Technical report INL/EXT-11-24275, Idaho National Laboratory, Idaho Falls, Idaho, 83415, September 2011.

[8] S. Lomperski, 2012. Private communication.

[9] S. Lomperski, E. Merzari, A. Obabko, W. D. Pointer, and P. Fischer. The MAX facility for CFD code validation. In *Proceedings of ICAPP '12*, 2012.

[10] S. Lomperski, E. Merzari, D. Pointer, P. Fischer, and A. Obabko. MAX startup tests. Technical report ANL-NE 10/28, Argonne National Laboratory, September 2010.

[11] S. Lomperski, D. Pointer, S. Wiedmeyer, P. Fischer, and A. Obabko. Multi-dimensional turbulent gas mixing experiment MAX design report. Technical report ANL-NE 09/27, Argonne National Laboratory, September 2009.

[12] R. Mei. Velocity fidelity of flow tracer particles. *Experiments in Fluids*, 22:11–13, 1996.

[13] A. Melling. Tracer particles and seeding for particle image velocimetry. *Meas. Sci. Technol.*, 8:1406–1416, 1997.

[14] K. C. Mousseau, R. W. Johnson, and H. Lee. High level requirements for the nuclear energy – knowledge base for advanced modeling and simulation (NE-KAMS). Technical report INL/EXT-11-23412, Idaho National Laboratory, Idaho Falls, Idaho, 83415, September 2011.

[15] K. C. Mousseau, R. W. Johnson, and H. Lee. Strategic plan for nuclear energy – knowledge base for advanced modeling and simulation (NE-KAMS). Technical report INL/EXT-11-22049, Idaho National Laboratory, Idaho Falls, Idaho, 83415, October 2011.

[16] W. L. Oberkampf and C. J. Roy. *Verification and Validation in Scientific Computing*. Cambridge University Press, 2010.

[17] QNET-CFD. `http://www.ercoftac.org/products_and_services/wiki/`.

[18] C. Rapp, M. Breuer, M. Manhart, and N. Peller. UFR 3-30: 2D periodic hill flow. `http://www.ercoftac.org/products_and_services/wiki/`. ERCOFTAC-QNET-CFD.

[19] Ch. Rapp. Experimentelle studie der turbulenten strömung über periodische hügel. Mitteilungen 75, Fachgebiet Hydromechanik, TUM, 2009.

[20] Ch. Rapp, F. Pfleger, and M. Manhart. New experimental results for a LES benchmark case. In V. Armenio, B. Geurts, and J. Fröhlich, editors, *Direct and Large-Eddy Simulation VII*, volume 13 of *ERCOFTAC Series*. Springer Netherlands, 2010.

[21] V. G. Weirs, K. C. Mousseau, R. W. Johnson, H. Lee, W. L. Oberkampf, B. Smith, J. Fort, and W. J. Rider. Nuclear energy – knowledge base for advanced modeling and simulation (NE-KAMS) code verification and validation data standards and requirements: Fluid dynamics version 1.0. Technical report INL/EXT-11-23419, Idaho National Laboratory, Idaho Falls, Idaho, 83415, September 2011.

DISTRIBUTION:

1 Hyung B. Lee (electronic copy)
Bettis Laboratory
P. O. Box 79
West Mifflin, PA 15213

1 Weiju Ren (electronic copy)
Oak Ridge National Laboratory
One Bethel Valley Road
P. O. Box 2008 Building 4500-S, MS-6155
Oak Ridge TN 37831-6155

1 MS 0779 Kim Mousseau, 6222 (electronic copy)

1 MS 0899 Technical Library, 9536 (electronic copy)

Sandia National Laboratories