Proceedings of the IMAC-XXVII
February 9-12, 2009 Orlando, Florida USA

Developing Adequacy Criterion for Model Validation Based on Requirements

Randy L. Mayes
Experimental Mechanics, NDE and Model Validation Department
Sandia National Laboratories*
P.O.Box 5800 - MS0557
Albuquerque, NM, 87185
rlmayes@sandia.gov

NOMENCLATURE

RMS	Root Mean Square value of a vector of numbers
MAC	Modal Assurance Criterion

ABSTRACT

Model validation assesses the usefulness of a model for its intended purpose. It is often difficult to develop a quantitative measure to determine if the model is acceptable for the intended purpose. In this paper, it is assumed that the model will have systematic differences from the validation test data. The ultimate requirement is that the model be good enough to provide guidance for an engineering decision that must be made. It is then useful. A rationale must be developed that relates one or more response features to a requirement that is tied to the decision that will be made with the model predictions. At that point, one or more adequacy criteria need to be set in advance of the validation prediction of the response feature. One common mistake is to select a feature that is common in the industry, but cannot be easily related to the requirement for the model. Another common mistake is setting the adequacy criteria too stringently so that the model fails the validation even though it is useful for guiding the decision. Minimizing the number of response features is desirable for making the validation decision more straightforward. The rationale for selection of features and adequacy criteria should be agreed upon by the decision maker, analyst, test engineer and other stakeholders.

INTRODUCTION

In the field of structural dynamics, "model validation" is a common phrase. However, the meaning of the phrase is sometimes not so clear. In the 80's, the structural dynamics community often used the words "model validation", "model updating" and "model correlation" interchangeably. Often, these terms meant that the parameters of a finite element model had been tuned to produce a "better match" of the eigenvectors and eigenvalues of the model with modal test mode shapes and frequencies. In some circles, perhaps this is still the meaning associated with model validation, but model validation technology is advancing. One popular definition is that "Model validation is the process of determining the degree to which a computer model is an accurate representation of the real world from the perspective of the intended model application [1]". In general, the "accurate representation of the real world" comes from some type of test data. Here we take the approach that the model must make a blind prediction of the test response, and then be compared with the response to see if it matches within some predetermined bounds, to decide if it is valid. There is a high probability of the model predictions being systematically different from the test data, however, that does not necessarily destroy the value of the model.

*Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the U.S. Department of Energy under Contract DE-AC04-94AL85000.

George Box[2] is credited with the statement that "All models are wrong, but some are useful". He went on to say "... that all models are wrong; the practical question is how wrong do they have to be to not be useful." This paper attempts to provide guidance to answer the practical question.

Let us focus on implementing three major elements of the validation definition presented in the previous paragraph [1]. Perhaps the most important element of the definition focuses on the **purpose** of the validation based on the “perspective of the intended model application." The customer for a model validation effort is a decision maker who wants to know, “For the engineering decision I must make, is the information from the model valid or not?” This leads to the second element of the definition "to determine the degree to which a computer model is an accurate representation". Another way to restate George Box’s practical question is “How far can the model deviate from reality and still provide information that is valid for my decision?” The goal is to determine a response feature and adequacy criterion that provides the decision maker with confidence that the model either can or cannot be used for the engineering decision to be made. The need is to provide a "Go/No Go" determination on "the degree to which a computer model is an accurate representation". This Go/No Go determination should be based in some sense on the *requirements* for the model, and that is the focus of this paper. The third major element of the definition is the "accurate representation of the real world". Validation utilizes test data as the "accurate representation of the real world". The test data have uncertainty that should be quantified as well so that the decision maker has an idea how different the test data may be from reality. The model can be thought of in a perfect sense as our best understanding of the response for the intended application. The *requirement* here is to determine if that understanding is good enough to be useful in making the engineering decision.

DEFINITIONS

In order to be clear about developing a process that relates the model to its requirements in the validation process, some definitions will be set here to enable one to meet the validation goal.

- *Response measures* are the quantities the model is to predict [3]. (e.g. acceleration, stress, temperature)
- *Response features* are the results of some mathematical operation on the response measures. (e.g. Maximum acceleration, RMS value of multiple accelerations, Fourier transform of a time history, modal frequency)
- *Metrics* are the standard of measure of the response features for comparing the model with test data.
- *Adequacy criteria* provide the required maximum acceptable difference between the validation experiment and the computational model response features.

If these quantities can be established and agreed upon, then the validation process can produce fruitful results. Note that we allow for a decision that the model is not valid in this work, so that the decision maker knows that the model should not be utilized to make the engineering decision.

GETTING ADVANCE AGREEMENT WITH THE VALIDATION TEAM

A validation strategy must have agreement among all the players who are providing input for the model validation process or ultimately using the result. Before the validation process is attempted, getting agreement with the validation team is an important step, and one that is often neglected. An undesirable scenario at the end of the process is to have the decision maker or the experimentalist saying to the analyst "How does this information give us confidence that we can use this model to make a decision?" The analysts, experimentalists, decision maker, managers and other stakeholders must agree that there is a rationale that is acceptable to them to establish validation of the model.

INTENDED APPLICATION OF THE MODEL - Establish the Requirements

Before deciding on response features, the validation team must be clear on the intended application for the model. A model may be perfectly acceptable for one application and not good enough for another application. The application should be as specific as possible because that makes it easier to determine a few concise requirements for the model. The decision maker should be able to state the requirement in a way that the rest of the team understands. Most validation problems are sufficiently difficult that the decision maker cannot state the requirement at the beginning of this process. The decision maker may not be a specialist in structural dynamics.

The decision maker must discuss with the validation team what is desired from the model, and then the team can inform the decision maker about what can actually be learned from a model and possible validation experiments. Once the purpose or intended application of the model is clear and the validation team can state the requirement (which may not be quantitative), the team must work through the following selections to develop a rationale that ensures the decision maker will get his Go/No Go decision on the validation of the model.

RESPONSE MEASURES

The analyst and experimentalist may predict and measure many quantities, but they must agree on what must ultimately be compared for validation. Deciding on the response measures is the first step in relating the validation to requirements. Obviously, the response measures predicted by the model must correspond one for one to those of the test. At Sandia National Laboratories a common statement that is made about validation is that "You must model what you test, and test what you model". For example, if the response measures include acceleration time histories, there are cases where the model must include the accelerometer mounting block mass and associated accelerometer dimensional offsets to get a good estimate of the test acceleration.

The response measures must be chosen with the validation requirement in mind. Some thought may also be required for the response feature, before the response measures are finalized.

RESPONSE FEATURES

The response features may be the response measures themselves, but often response features are a result of some mathematical operation performed on the response measures. For example, the response feature might be the root mean square (RMS) of an acceleration time history. The response feature should be a quantity that can be strongly related to the requirements by quantitative values. It is often easier to relate a response feature to the requirements than all the response measures. For example, in one validation, the maximum acceleration at the attachment location for a particular component was chosen as the response feature. This reduces perhaps thousands of acceleration time history values down to one number that is much easier to relate to a requirement. Here are some guidelines for choosing meaningful response features.

1. Response features must be well defined.
2. The number of response features should be kept to a minimum.
3. Multiple response features should be prioritized as to their importance.
4. A clear rationale should relate the response features to the requirements.

The first guideline is the most important. Some well defined response feature(s) must be chosen. Oftentimes, no response feature is chosen at all and validation is attempted by comparing hundreds of responses, assuming that these comparisons will automatically ensure validation. Sometimes such ambiguous results fail to provide meaningful information from the modeling effort, even though significant resources may have been expended. Even if one chooses a terrible response feature for the validation with poor rationale, usually at the end it will be clear that this was a terrible choice, and at least something has been learned for the next validation process. The second guideline is important because having fewer response features makes the validation decision easier. Also, uncertainty quantification on the response feature can be performed if many responses are condensed down to a much smaller number of features. An example of this might be a smoothed frequency response function, in which hundreds of values of response at different frequencies are reduced down to a dozen values, each with an associated standard deviation. The third guideline will help the validation team to decide what is important and what is not. The last guideline is perhaps the most difficult, since it is used to quantify the requirement. This rationale should be understood and agreed upon by every member of the validation team. A common mistake is to choose a response feature that is common in the industry, rather than one that relates to the requirements. For example, modal assurance criterion (MAC) is commonly calculated in structural dynamics, but can the team show the rationale that relates the MAC to the requirement?

ADEQUACY CRITERIA

Once the response features are chosen that relate to the requirements, the final step is to decide how close the model prediction should be to the test result to be good enough to use for decision making. This step is not easy. One mistake that is often made by analysts and experimentalists is to overconstrain the adequacy criteria. We want our models to be perfect, and we are sometimes too analytical about differences. Perhaps a better way to approach it is to repeat George Box's quote in the introduction, “How far can the model deviate from reality and still provide information that is valid for my decision?” How bad can the model be and still be useful? When the "how bad can it be" question is answered, the adequacy criteria have been found. Here are some guidelines for good adequacy criteria.

1. Adequacy criteria must be well defined.
2. Adequacy criteria should not be artificially restrictive.
3. Adequacy criteria should relate the response features to requirements.

The first guideline is the most important. Sometimes response features have been chosen, but there are no clear adequacy criteria. This results in ambiguous conclusions, and nothing is learned. The way this author learned the second guideline, "not to choose artificially restrictive adequacy criteria" was by choosing artificially restrictive adequacy criteria for one validation. After the team declared that the model had failed the validation, it became clear that we had chosen overly restrictive criteria, so we learned from our mistake. Sometimes validation criteria are set based on statistical tests of significance for differences between the model and experiment response features. Rarely, do these differences in any way relate to the requirements for the model. They can be overly restrictive. Statistical uncertainty quantification does provide value. For example, one may find that the uncertainty in the test and/or the uncertainty in the model is so large that the required model accuracy is dwarfed by these uncertainties. Although this may not be the desired result, at least it shows that the model should not be considered as a definitive guide for making a decision. Use both uncertainty quantification and the requirements together to decide what the adequacy criteria should be. Finally, the third guideline should quantitatively implement a rationale that relates the validation to the requirements.

STEPS TO MODEL VALIDATION

Urbina, Paez, et al [3] have provided guidelines for the model validation process that this author has found useful. Figure 1 gives their diagram of the process. They have given steps to the validation process which are repeated here.

- Preliminary Steps
    - Specify model use/purpose (what decision is to be made)
    - Specify response measures (what the model predicts)
    - Specify validation features and metrics and comparison domain
    - Specify calibration experiments
    - Specify validation experiments
    - Specify adequacy criteria
- Perform calibration experiments/Calibrate model parameters
- Validation
    - Perform experiment
    - Make predictions
    - Calculate metrics/compare with adequacy criterion
- Subsequent Action
    - Not valid – Reformulate model/Additional calibration
    - Valid – Make predictions

The process outlined here is based on the idea that the validation data are not used for calibration. However, it also takes into account the practice that the validation data may be used for calibration or as the basis for reformulating the model if the initial validation fails. To complete such a validation, there would theoretically need to be another (and different) validation test data set for final validation. If this author were to add one item to the list above, it would be that the step to calibrate model parameters would also be utilized to debug the model. Sometimes there are unintentional errors in the model with its thousands of detailed inputs, and it is definitely valuable to exercise the model with some calibration experiments to try to assure that errors are exposed and corrected before the final validation prediction.

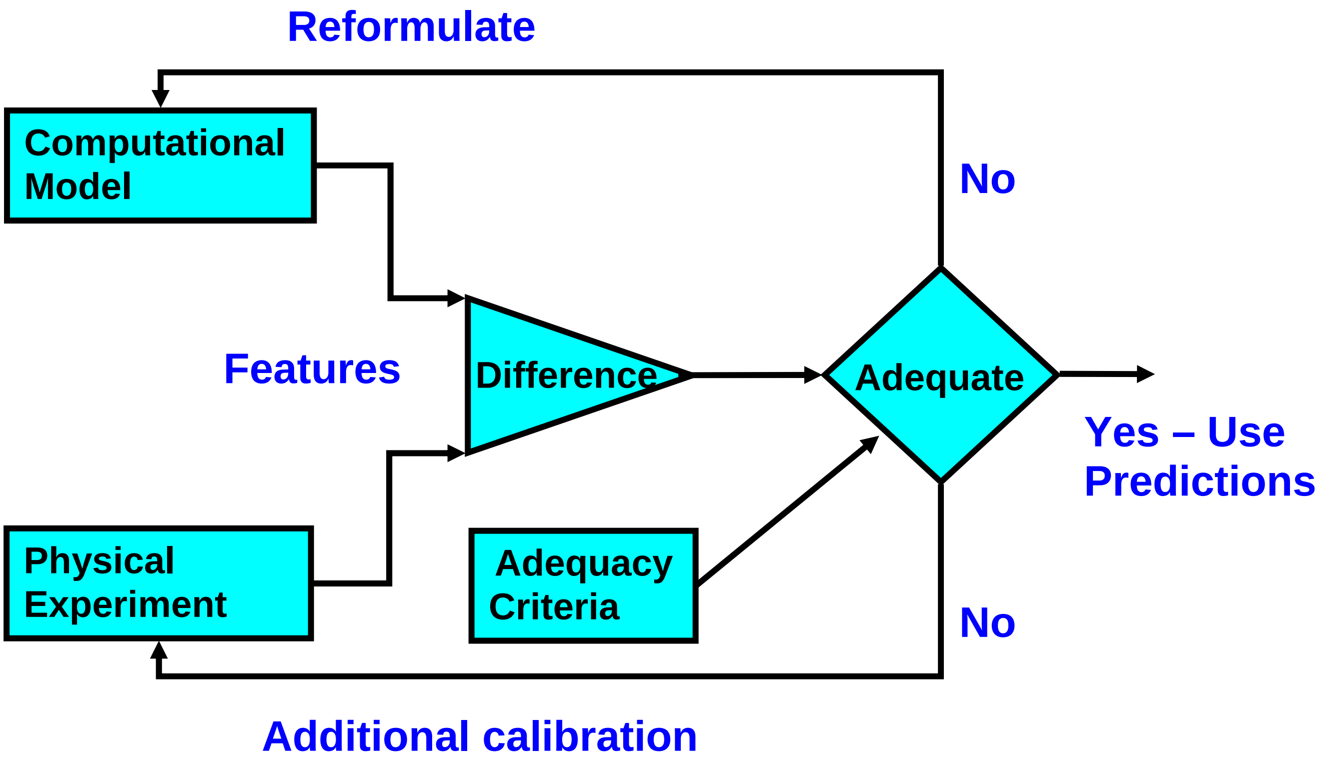


Figure 1 - Validation Process

AN EXAMPLE

An example may help to firm up the principles presented here. A FE model of a complete delivery system was to be used to provide specifications for shock testing of components within the delivery system for a specific shock environment that the system could experience. The component test specifications were to be in terms of shock spectra for a specific low frequency haversine shock input. The experimentalists, analyst, system lead and management agreed to the following rationale after many meetings providing dialog on the validation process. Since the duration of the test haversine was already determined by limitations of the component shock test facility, the level of the component test shock spectrum could be set simply by the peak acceleration at the component mounting location low pass filtered to the frequency content of the capability of the shock test facility. The stakeholders determined that some significant overprediction by the model was acceptable, but significant underprediction was unacceptable. Calibration of the model to subsystem modal tests was utilized to adjust certain joint stiffnesses and debug the model. A validation ground test was designed which was not the same as the real environment, but exercised the system to the same levels and with similar physics. Acceleration time history measurements were collected from the test and predicted by the model. The peak acceleration from the filtered time history (the response feature) at each component attachment location was compared between the model and test in a blind prediction. The adequacy criterion was that the model must predict the filtered peak acceleration within a band. The high end of the band was twice the experimental acceleration plus the uncertainty on the test measurement. The low end of the band was the experimental acceleration minus the uncertainty on the experiment. There were several component attachment locations, and the model passed the adequacy criteria at all locations except two.

THE ALLEGORY OF VALIDATION BASED ON REQUIREMENTS

Once upon a time there was an important engineering decision to be made. Decision Maker did not know what decision to make. As decision time drew near, Decision Maker sought more information to help with the decision. He knew that someone had a model that was supposed to predict something, so he looked up the phone number of Analyst, called Analyst, and set up a meeting to see if the model could help with the decision. Analyst told

Decision Maker how much work had been expended to generate the model and what it could predict. Analyst showed very colorful vugraphs that represented virtual reality. Decision Maker asked if Analyst could prove that the model was any good. Analyst said he thought it was very good for making vugraphs. Decision Maker was uncomfortable so he looked up the phone number of Experimentalist and called him to come to the meeting. Decision Maker asked Experimentalist if the model was any good. Experimentalist did not know, but he said he could make a measurement on the prototype hardware. Meanwhile, Stakeholder was very interested in the decision, heard about the meeting and came to it. Stakeholder could see that the meeting was going nowhere and offered a suggestion. Stakeholder had just been to a short course on model validation and told them about the concept of validating the model based on requirements for the model. They were all very interested, so they had more meetings to define the purpose of the model, translate the purpose into requirements, and relate the requirements to response features with adequacy criteria. There was much arguing during the meetings, but finally they agreed upon and built a prefabricated Bridge to the Engineering Decision shown in Figure 2.

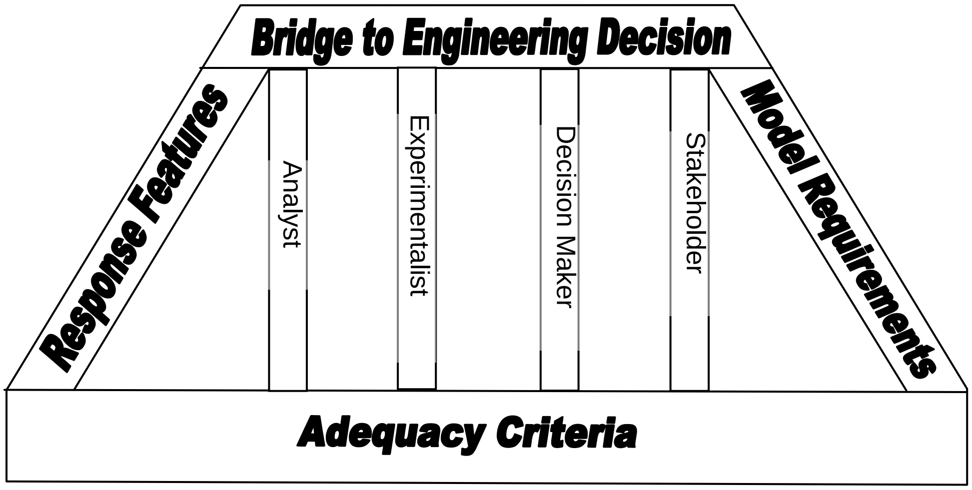


Figure 2 - Model Validation Bridge to Engineering Decision

The Bridge was designed to span the Gap between the model prediction and reality for a validation experiment they had worked hard to develop. They had struggled with the question of how long to make the bridge, but had decided together upon adequacy criteria that determined its length. Analyst made a prediction with the model on which such hard work had been expended. Finally, they sent out Experimentalist to measure the response in the validation experiment and everyone held their breath to see if the prefabricated bridge would span the Gap shown in Figure 3.

Alas, the bridge did not span the Gap, and Analyst, Experimentalist and Stakeholder were crushed.

However, Decision Maker was an optimist and looked on the bright side. At least Decision Maker knew the model should NOT be used for this decision. And during Validation Team interactions, Decision Maker had come to respect the opinions and capabilities of Analyst, Experimentalist and Stakeholder. Soon another project arose in which another important engineering decision had to be made. This time Decision Maker called the Validation Team together early in the project, and they forged another prefabricated Bridge. This time they easily defined the purpose and requirements of the model. They were able to spend more time focusing on how they could relate the requirements to response measures and features as well as the adequacy criteria. From the last exercise, Decision Maker and Stakeholder had learned that the adequacy criterion had really been overconstrained a bit, so they all agreed that the new bridge could be a little longer. Analyst had learned that one subsystem of the model had some significant uncertainties, so Analyst pled with Decision Maker to budget for a quick modal test of the subsystem so that a very uncertain parameter could be calibrated. Analyst just knew that this would improve the model. Decision Maker was convinced. After the calibration effort, Analyst made a prediction, and again they sent out Experimentalist to measure the response in the new validation experiment. Experimentalist had learned from the last experience that the measurements had a bias error. Experimentalist had developed a method to estimate this error, so when the experimental measurement was made, Experimentalist subtracted out the bias error. This time when they measured the Gap, the Bridge could span it!

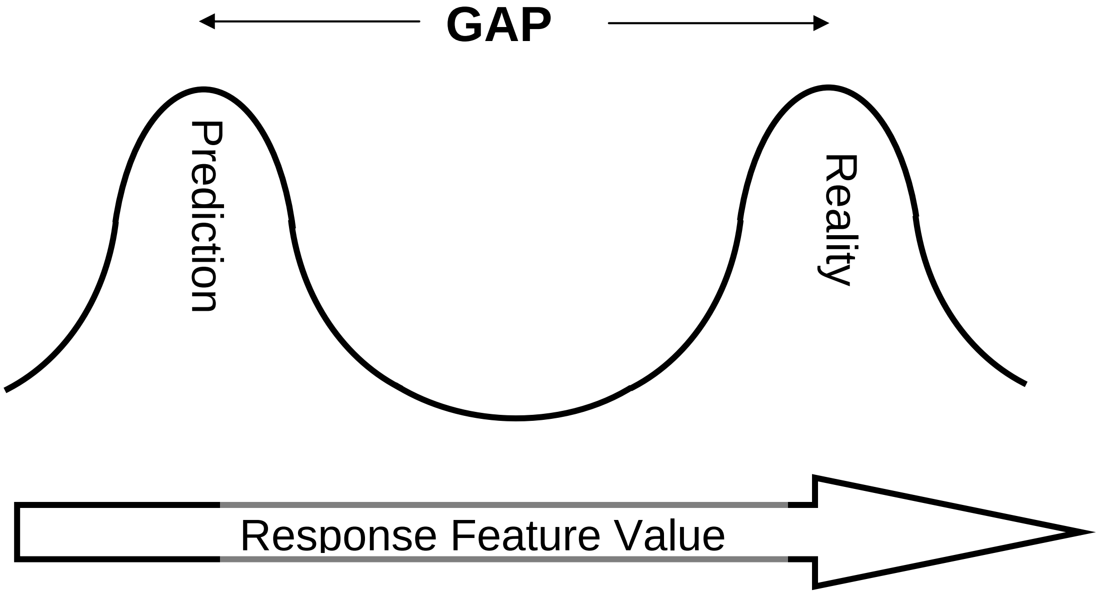
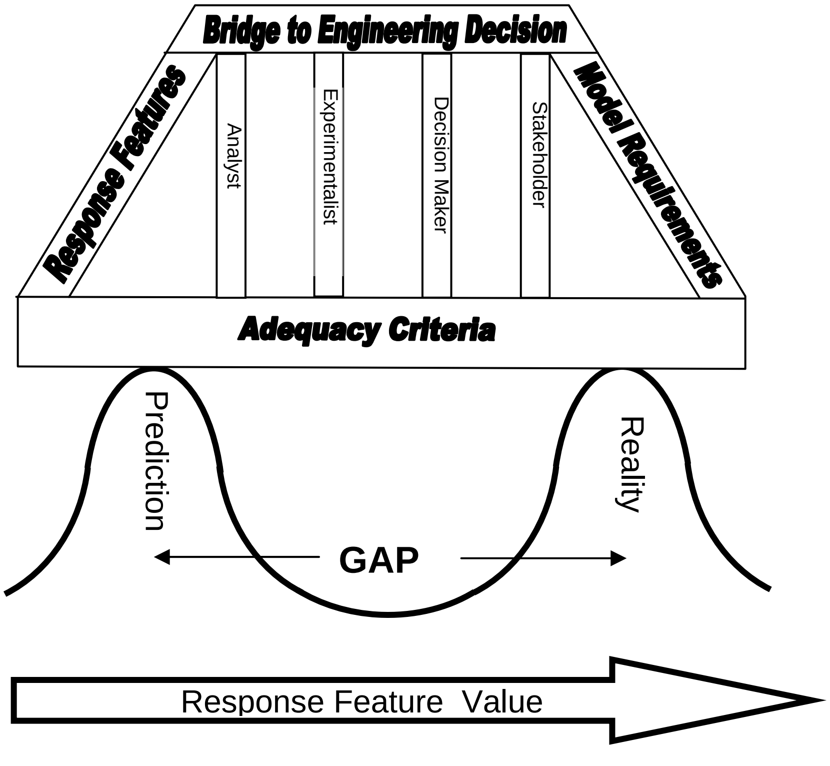


Figure 3 - The Gap Between Prediction and Reality

Since Analyst was a cowboy from Wyoming, he invited everyone over to his house that night, killed the fatted calf, and served barbecue. The team rejoiced together and even wrote up an abstract on validation for the next conference. And Decision Maker was very happy, since the model could help make the decision.

THE END



CONCLUSIONS

Modeling simulates our understanding of a process. A definition of model validation is utilized to develop a rational process of determining whether the understanding represented in the model is good enough that the model can be utilized for making an important engineering decision. Sometimes validation efforts are undertaken which do not have rationale for deciding whether the model is useful. The results can be ambiguous. The validation team should develop the rationale that relates the model validation to requirements, which should be derived from the intended purpose of the model. This team should include the analyst, experimentalist, decision maker and stakeholders in the validation process. Uncertainty quantification on the test results as well as the model should be utilized along with requirements to determine if the model is useful. Steps for model validation as well as guidelines for selecting response measures, response features and adequacy criteria are provided.

REFERENCES

[1] AIAA (American Institute of Aeronautics and Astronautics), *Guide for the Verification and Validation of Computational Fluid Dynamics Simulations*, AIAA-G-077-1998, Reston, VA, American Institute of Aeronautics and Astronautics.

[2] Box, George E. P., Draper, Norman R., *Empirical Model-Building and Response Surfaces*, 1987 p. 424, 74, Wiley. ISBN 0471810339

[3] Urbina, A., Paez, T.L., Rutherford, B., O'Gorman, C., Hinnerichs, T., Hunter, P., "Validation of Mathematical Models: An Overview of the Process", Proceedings of the 2005 SEM Conference and Exposition on Experimental and Applied Mechanics, Paper 210, June 2005.