A Comprehensive Validation Methodology for Sparse Experimental Data

Ryan B. Norman
University of Tennessee, Knoxville, Tennessee

Steve R. Blattnig
Langley Research Center, Hampton, Virginia

February 2010

NASA STI Program . . . in Profile

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA scientific and technical information (STI) program plays a key part in helping NASA maintain this important role.

The NASA STI program operates under the auspices of the Agency Chief Information Officer. It collects, organizes, provides for archiving, and disseminates NASA's STI. The NASA STI program provides access to the NASA Aeronautics and Space Database and its public interface, the NASA Technical Report Server, thus providing one of the largest collections of aeronautical and space science STI in the world. Results are published in both non-NASA channels and by NASA in the NASA STI Report Series, which includes the following report types:

- TECHNICAL PUBLICATION. Reports of completed research or a major significant phase of research that present the results of NASA programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counterpart of peer-reviewed formal professional papers, but having less stringent limitations on manuscript length and extent of graphic presentations.

- TECHNICAL MEMORANDUM. Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.

- CONTRACTOR REPORT. Scientific and technical findings by NASA-sponsored contractors and grantees.

- CONFERENCE PUBLICATION. Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or co-sponsored by NASA.

- SPECIAL PUBLICATION. Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.

- TECHNICAL TRANSLATION. English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services also include creating custom thesauri, building customized databases, and organizing and publishing research results.

For more information about the NASA STI program, see the following:

- Access the NASA STI program home page at *http://www.sti.nasa.gov*

- E-mail your question via the Internet to help@sti.nasa.gov

- Fax your question to the NASA STI Help Desk at 443-757-5803

- Phone the NASA STI Help Desk at 443-757-5802

- Write to:
  NASA STI Help Desk
  NASA Center for AeroSpace Information
  7115 Standard Drive
  Hanover, MD 21076-1320

NASA/TP–2010-216200

A Comprehensive Validation Methodology for Sparse Experimental Data

Ryan B. Norman
University of Tennessee, Knoxville, Tennessee

Steve R. Blattnig
Langley Research Center, Hampton, Virginia

February 2010

Acknowledgments

Contents

List of Figures

List of Tables

Abstract

A comprehensive program of verification and validation has been undertaken to assess the applicability of models to space radiation shielding applications and to track progress as models are developed over time. The models are placed under configuration control, and automated validation tests are used so that comparisons can readily be made as models are improved. Though direct comparisons between theoretical results and experimental data are desired for validation purposes, such comparisons are not always possible due to lack of data. In this work, two uncertainty metrics are introduced that are suitable for validating theoretical models against sparse experimental databases. The nuclear physics models, NUCFRG2 and QMSFRG, are compared to an experimental database consisting of over 3600 experimental cross sections to demonstrate the applicability of the metrics. A cumulative uncertainty metric is applied to the question of overall model accuracy, while a metric based on the median uncertainty is used to analyze the models from the perspective of model development by analyzing subsets of the model parameter space.

1   Introduction

With NASA's vision for space exploration emphasizing human exploration beyond low Earth orbit (LEO), radiation exposure concerns have become increasingly important. As exploration moves beyond LEO to radiation environments where much less data is available, models will be more heavily relied upon to make decisions regarding vehicle shield design and overall mission planning. This reliance on model results makes it critical to effectively validate these models in order to determine their accuracy and reliability. In order to ensure confidence in the validation assessment, reliable and repeatable processes must be implemented. A validation assessment does not lead to credible results if the validation process is unreliable or if the data or analyses are not sufficient to make the appropriate conclusions. See Oberkampf and Barone [1], Oberkampf et al. [2], and Blattnig et al. [3] for further discussion on assessing model results.

The degree of confidence in a model is not only an issue of accuracy, but also of the rigor and completeness of the assessment itself. An essential aspect of a comprehensive validation effort is the development of configuration-controlled verification and validation (V&V) test cases. Configuration control (also called configuration management) is a process in which consistency is established for a product (i.e. a model or a software suite), and any changes made to the product are tracked. The effects of the changes on the product are documented, and therefore, problems caused by changes to the product can be backtracked. The models reviewed in this paper have been placed under configuration control with the criteria that the V&V test cases will be run when significant changes are made to those models. This approach allows accuracy to be tracked across the relevant range of applications and avoid situations where model changes intended for a specific calculation or application actually decrease the overall accuracy. In addition, it will enable more complete accuracy assessments to measure progress against goals as the models and codes are updated and as new data become available for validation. It also helps ensure model results are repeatable.

To properly assess the accuracy of a model, quantitative validation metrics [1] need to be developed that are targeted at the applications of interest. A validation metric is a mathematical operator that quantitatively measures the difference between a model and experiment for an output value or system response quantity (SRQ) of interest. In general, a validation metric can take any SRQ as an input. Different metrics or different SRQs may be required for different types of applications and experimental databases. Additionally, ease of use and simplicity need to be taken into account when developing such metrics. Not only will researchers need to be able to use and understand such metrics when validating models, but also, the results of the validation activity will need to be communicated to managers and decision makers. In that regard, this paper uses two nuclear fragmentation models as a case study of the validation strategy currently being implemented in several of NASA's space radiation projects.

An important component of a comprehensive validation program is the experimental database used to validate the models. The database assembled for validation should be as comprehensive as possible so that all phenomena of interest to the problem are represented. This process serves to help model developers understand what is essential to the problem and how well the models represent the phenomena. To validate the nuclear models used by NASA in space radiation applications, a thorough literature survey was performed, and all available experimental datasets containing nuclear fragmentation total cross sections were assembled. Nuclear fragmentation total cross sections are the SRQs of the nuclear models used in space radiation applications. At the conclusion of the literature search, an experimental database of over 3600 cross sections from 30 different publications with data taken at 8 different experimental facilities from around the world was assembled for validation.

2   Nuclear Models used for Space Radiation

Understanding nuclear interactions of heavy ions is of fundamental importance to the safe and reliable exploration of space. Outside the safety of Earth's atmosphere and geomagnetic field, there is a nearly isotropic background of high energy, fully ionized radiation called galactic cosmic rays (GCR). The GCR spectrum consists mainly of protons and helium ions but includes all naturally occurring isotopes. A significant drop in flux of GCRs occurs for ions heavier than nickel. Prolonged exposure to GCR radiation can have dangerous health effects [4].

The effects of radiation on microelectronics is also of significance. As solid state devices have decreased in size, the device current of the microelectronic is lowered. As a consequence, less energetic particles, which are more plentiful, can generate errors in which one or more logic bits have their state changed. The errors caused by cosmic ray interactions can have adverse effects on computer systems integral to space operations.

To understand and quantify the risk to astronauts from space radiation exposure, a computer software suite is used which takes as input a given radiation environment, then uses an appropriate radiation transport algorithm to transport the radiation through material. This transport process requires cross sections to determine how the radiation is modified by the intervening materials. A measure of risk is then determined from the flux of radiation remaining after the transport process [5, 6].

Nuclear fragmentation, the process through which energetic heavy ions interact with target

nuclei and break apart, is a significant contributor to the nuclear interactions of heavy ions. We define heavy ions as fully ionized atomic nuclei with charge greater than helium. In addition, the possibility of multiple heavy ion fragments being created makes this an interesting and complex physical process. From a radiation shielding point of view, this must be accounted for when determining radiation exposure to astronauts and sensitive electronics.

To account for the modification of the radiation environment through interactions with intervening spacecraft shielding and human tissue, a nuclear fragmentation cross section database is used by radiation transport codes. A cross section is defined in terms of a beam of incident particles impacting a target and the number of particles scattered from the beam. The number of particles scattered per unit time, N, is given as

$$N = \mathcal{L}\sigma, \tag{1}$$

where $\mathcal{L}$ is the beam luminosity or flux (number of particles per unit time passing through a unit area transverse to the beam), and σ is the cross section. Eq. (1) gives a working definition of a cross section.

The dimensional unit of cross section is area, and is understood to represent the effective area removed from the incident particles through interaction and is a measure of the likelihood of an interaction occurring. The usual unit of a cross section in nuclear physics is a millibarn (mb), which is equal to 10^{-27} cm^2. A cross section is essentially a probability of a certain type of reaction taking place. Note that the particles considered herein are fundamentally quantum mechanical in nature, so any reaction that does not violate conservation laws has a nonzero probability of occurring. Also, even with perfectly known initial conditions, the results of a scattering experiment are still probabilistic. Specifically, consider a flux of particles of type P and energy E, $\mathcal{L}_P(E)$, hitting a target. The projectile has some probability of breaking up and producing a fragment of type F. The number of fragments of type F produced per unit time can be found from Eq. (1) if we know the cross section for production of fragment F from particle P interacting with a given target. Note that because the problem is quantum mechanical and fundamentally probabilistic, this only holds in the limit where the flux goes to infinity. The fundamentally probabilistic nature of the problem is due to the fact that predictions in quantum mechanics only pertain to probability distributions which describe the dynamics of ensembles of particles. The probabilistic nature is not, however, because of variations in initial conditions but due to a fundamentally probabilistic nature of the interaction.

An accurate understanding of the nuclear fragmentation process is important to accurately determine the biological effect of cosmic radiation on astronauts. The uncertainty of models used to generate the nuclear fragmentation database used in radiation transport codes must be quantified in order to understand the total uncertainty of the radiation transport codes. Once the total uncertainty of the radiation transport codes is quantified, then the contribution of radiation transport code uncertainty to mission risk can be determined.

Two models of nuclear fragmentation commonly used in space radiation transport codes, NUCFRG2 [7] and QMSFRG [8], have been chosen to demonstrate the validation process. However, due to lack of access to the source code for QMSFRG, a table of values generated from QMSFRG was used. Interpolation was used to produce cross sections not found in the database. All conclusions drawn from this paper apply only to the database and not the original model

since the error due to interpolation is not known exactly. However, the use of the database version of QMSFRG is sufficient to demonstrate the validation process and the metrics used. Further discussion of each model is presented below.

2.1 The NUCFRG2 Model

NUCFRG2 [7] models the projectile fragmentation in heavy ion collisions using the classical, geometric abrasion-ablation formulation [9]. In the abrasion-ablation formalism, a heavy ion collision is modeled in two stages. During the abrasion stage, a portion of the projectile is sheared away during the interaction with the target. The residual nucleus (called the prefragment) is then left in an excited state and decays during the ablation stage through particle emission. The details of the interaction between the projectile and target rely only on the projectile energy and relative physical overlap between the target and projectile nucleus.

During the ablation stage of projectile fragmentation, the prefragment is left in a highly unstable state. NUCFRG2 assumes the charge ratio of nucleons left in the prefragment is equal to that of the charge ratio of the projectile prior to interaction. This approximation is only appropriate at high energies where the projectile does not have time to react to the target. A crude visualization of the prefragment after the abrasion stage is to think of a spheroid with a channel gouged out of it. In addition to abrasion and ablation, NUCFRG2 includes electromagnetic dissociation for the production of nucleons. Further details concerning the NUCFRG2 model can be found in Wilson et al. [7].

2.2 The QMSFRG Model

The formalism used in the quantum multiple scattering fragmentation (QMSFRG) model for nuclear fragmentation is based on the multiple scattering series as formulated by Glauber [10] for two heavy ions. The multiple scattering series is solved using the impulse and eikonal approximations [8] to obtain a closed-form solution to the abrasion stage cross section spectrum. QMSFRG assumes the fragmentation reaction occurs through an abrasion stage producing a prefragment, which is outside the region of overlap, and a fireball fragment produced from the projectile-target overlap region. Following the abrasion stage, an ablation stage occurs where the prefragment becomes the final projectile fragment through nuclear de-excitation. The ablation stage is described by a stochastic process using a Master equation for de-excitation through particle emission [8]. In addition to abrasion and ablation processes, QMSFRG also includes cluster knockout, electromagnetic dissociation for the production of nucleons, and a coalescence model of light ion ($A \leq 4$) formation.

Access to QMSFRG through a machine executable file or source code was not available during our analysis. QMSFRG produced a cross section table that used linear interpolation over eight energy points (25, 75, 150, 300, 600, 1200, 2400, 7200 MeV/nucleon) and linear interpolation or extrapolation over six target isotopes (^{1}H, ^{12}C, ^{16}O, ^{27}Al, ^{40}Ca, ^{56}Fe) for values not found in the database. All conclusions drawn from this paper apply only to the interpolated database version. The error introduced by using the interpolated database version of QMSFRG has been estimated as 5%. This is a rough estimate and a discussion of the uncertainty due to the use of the database is found in Appendix A.

4

3   Model Validation

There are multiple aims when performing a validation assessment. One possible aim is to determine the overall accuracy of the model for the application of interest. Another possible aim is to produce information on how to improve a model, determine which assumptions are correct and what effects are important. To demonstrate the two metrics developed in this work, each metric will be applied to a separate component of the nuclear fragmentation validation problem. The validation problem is divided into two subproblems. The first is a global assessment using cumulative uncertainty distributions as metrics in which the accuracy of the model is measured against all available experimental data. The global assessment attempts to quantify how well a model can be expected to perform for the application of interest. The second subproblem is a more detailed set of comparisons to data divided into different energy regions, projectile types, target types, and fragments produced. By performing the analysis as a function different variables of the validation space (e.g. energy, projectile type, etc.) the analysis can focus on details that are useful to researchers interested in improving models. Due to the differing goals of model development and overall accuracy assessment, the metrics applied to them should have these goals in mind. For model development, the metric should be flexible enough to handle dividing the phase space into regions which are functions of the independent variables that span the phase space. In addition, the metric must give detailed enough information to allow model developers and analysts to determine domains of applicability for the model and possible model improvements needed to better represent the experimental data. For overall accuracy, the validation process should quantify the model accuracy over the entire experimental database and should have the ability to separate the uncertainty due to the model from the uncertainty due to the experiment. Also, the validation process needs to have the ability to be summarized in a relatively simple form so that decisions can be made in a relatively objective manner and the results can be communicated effectively to those who will make decision based on them.

One of the keys to this overall strategy of validation is to have both the models and the validation benchmarks, which include the data and the software needed to make the comparisons, under configuration management. Then, as models are modified or new data are published, the validation software suite can identify the global effect of the changes. Performing validation in this manner helps to prevent models developing biases towards data subsets or improvements that decrease model uncertainty in a local region but cause global model uncertainty to increase. Validation comparisons can then be used to choose between models for particular applications, help model developers be certain that "improvements" in models are actually global improvements, track accuracy against programmatic goals, and give researchers information needed to improve models. In addition, the ability to separate the uncertainty due to the model from the uncertainty due to the experimental measurement will allow an assessment to be made as to whether making additional, more accurate measurements would be useful if measurement uncertainty is the dominating factor.

The experimental database assembled for validation should ideally cover the complete phase space of interest to validation with appropriate resolution for each independent variable. Therefore, in the case of nuclear fragmentation models used in space radiation shielding analysis, the experimental database should closely represent the galactic cosmic ray background which has an

energy spectrum that encompasses approximately 5 orders of magnitude (10 MeV - 10^5 MeV), a projectile distribution of 128 isotopes of elements from hydrogen to nickel, and contain data from targets commonly used in spacecraft shielding [4]. In addition, information on all possible fragments produced from a given projectile-target combination are important. This is a vast phase space to cover for validation purposes.

In order to assess the accuracy of the nuclear fragmentation models, a database of experimental nuclear fragmentation total inclusive cross sections was assembled. The first task in the validation activity was to perform a comprehensive literature search. All experimental data in the open literature were assembled and a database was created which consists of over 3600 cross sections with 25 distinct projectile isotopes from ^{10}B to ^{58}Ni, a projectile kinetic energy range of 90 MeV/nucleon to 14500 MeV/nucleon, and both elemental and compound targets ranging from hydrogen to uranium [7, 11–39]. The experimental data were taken from 30 different publications which described experiments done at 8 different facilities around the world, including the Joint Institute for Nuclear Research in Dubna, Russia, the Heavy Ion Medical Accelerator in Chiba, Japan, and the Super Proton Synchrotron at CERN near Geneva, Switzerland.

Measurement uncertainty reported for most experimental data sets is a combination of systematic and statistical uncertainty with some experiments only reporting statistical uncertainty. Overall, the systematic uncertainty reported in the experimental papers tended to be of the same order as the reported statistical uncertainty.

Table 1 gives an overview of the experimental database assembled for validation of the nuclear models. There is a distinct lack of data in the low energy ($\leq$ 300 MeV/nucleon) and high energy ($>$ 5000 MeV/nucleon) regions for most projectile-target combinations. In addition, there is a lack of a consistent, robust energy range for many projectiles. There is also a lack of data for large values of the charge removed from the projectile. For a good number of experiments, only cross sections for fragments with charge equal to half the projectile or larger were reported.

One of the motivations for the validation process is to quantify the uncertainty of the nuclear models. Understanding the uncertainty in all components of any system is important to understanding all the uncertainty involved in any decision based on that system. In any validation effort, however, the experimental data is only taken at representative points in the phase space, and the representative points may not even span the complete phase space. Many times, the validation process only attempts to test the consistency of a model with the experimental data. On the other hand, the validation process outlined in this paper is attempting to determine the model's accuracy, which is much different than testing the consistency of a model with experimental data. For the application of space radiation, the models will be relied on for making decisions for regions of the phase space that are not covered by experimental data. Therefore, quantifying the uncertainty in a way that does not simply measure the consistency of the model with the experiments is important. This is the reason the validation process must have the ability to separate the uncertainty due to the experiment from the model uncertainty.

One issue that is not addressed in this paper is the completeness of the experimental data set in covering the validation phase space. A large portion of the experimental data assembled for validation did not have space radiation applications in mind when the experiments were performed. Many of the experimental measurements were concerned with discovering new fundamental physics or attempting to shed new light on fundamental physics problems. In ad-

dition, due to the goals of these experiments and large relative cost of the experiments, there are very few experimental results that were reproduced by other experiments. However, there were a series of experiments carried out by a group at Lawrence Berkeley National Laboratory which were commissioned by NASA for use in validating models used in the space radiation problem [11–15]. The phase space for validation is immense, and therefore, measurements were typically only taken at representative sample points. While the experiments commissioned by NASA were chosen to cover the validation phase space, there may exist more measurements in certain regions of phase space than in others, or there may be a complete lack of data in regions which can bias the representation. In addition, certain portions of the phase space will be more important to the applications under consideration. The completeness of this and other nuclear fragmentation datasets is currently being investigated and will be published in another work. One way this problem could potentially be addressed is to perform sensitivity analyses for particular applications and develop weights that could be used in the validation metrics. These issues, however, are beyond the scope of this paper and will be addressed in future work.

3.1 Cumulative Uncertainty Metrics

When a large experimental database is available for use in model validation, it may be useful to consider a method of validation that can evaluate the entire validation database with a single metric. This allows the overall model accuracy to be characterized for all available data. The overall accuracy of a model is an important quantity when, for instance, trying to distinguish between models for use in an application or when the model results are used as input to another system and the uncertainty of the entire system is needed. A simple method to measure the accuracy is presented here which separates the contribution due to experimental uncertainty from the model uncertainty. This metric is used over the entire database and is used to produce a distribution of model uncertainties. A cumulative uncertainty metric based on an absolute uncertainty distribution has been developed and is presented below.

For most experiments in the database, experimental uncertainty was reported as a combination of statistical and systematic uncertainty. In these cases, the contribution of statistical uncertainty compared to systematic uncertainty was not explained. Therefore, the experimental uncertainty was treated as completely epistemic uncertainty or uncertainty due to lack of knowledge. As a result, the experimental data were treated as an interval such that the "true value" is assumed to lie somewhere within the interval without any additional information as to where within the interval the "true value" may lie.

3.1.1 Cumulative Absolute Uncertainty

Assume model accuracy assessment is taking place at a set of discrete points in a phase space $\{x_i\}$. Given a set of experimental values, $\{E(x_i)\}$, where $E(x_i)$ represents the experimental value at a specific phase space point x_i, with experimental uncertainty $\{\epsilon(x_i)\}$, and a set of model values $\{M(x_i)\}$, we define difference functions at each point in the set

$$D^+(x_i) \equiv M(x_i) - [E(x_i) + \epsilon(x_i)],\qquad(2)$$

7

Table 1: Overview of the experimental database [7, 11–39] assembled for validation.

Projectile	Energy Range (MeV/nucleon)	Distinct Energies	Targets
^{58}Ni	338	1	H
^{56}Fe	330-1880	14	H,He,Li,Be,C,Al, S,Cu,Ag,Sn,Ta,Pb,U
^{52}Cr	343	1	H
^{48}Ti	1000	1	H,C,Al,Cu,Sn,Pb
^{40}Ca	357-763	4	H
^{40}Ar	90-1650	13	H,Be,C,Al,KCl, Cu,Ag,Sn,Pb
^{36}Ar	213-1050	9	H,Be,C,Al,Cu,Ag,Sn,Pb
^{35}Cl	650-1000	2	H,C,Al,Cu,Sn,Pb
^{32}S	365-3650	8	H,C,Al,Cu,Ag,Pb
^{28}Si	290-14500	15	H,C,Al,Cu,Ag,Sn,Pb
^{27}Al	582	1	H,He,C
^{24}Mg	309-3650	8	H,C,Al,Cu,Ag,Pb
^{23}Na	461	1	H,C
^{22}Ne	377-894	3	H
^{20}Ne	468-1057	5	H,C,Al,Cu,Ag,Sn,Pb
^{21}O	557	1	C
^{20}O	585	1	C
^{19}O	635	1	C
^{18}O	573	1	C
^{17}O	629	1	C
^{16}O	290-13500	9	H,He,Be,C,Al,Cu,Ag,Pb
^{14}N	441-516	4	H,He
^{12}C	112-3660	27	H,He,Be,C,CH$_2$,H$_2$O, Al,Cu,Ag,Sn,Pb
^{11}B	326-561	2	H,He,C

$$D^-(x_i) \equiv M(x_i) - [E(x_i) - \epsilon(x_i)]. \qquad (3)$$

$D^+(x_i)$ is the difference between the model and maximum value of the experiment, while $D^-(x_i)$ is the difference between the model and the minimum value of the experiment. Both are illustrated in Fig. 1. In our present analysis, $E(x_i)$ would be an experimental cross section for a given projectile-target combination at a defined incident energy producing a given fragment (e.g. 100 MeV/nucleon ^{12}C+^{27}Al→^{11}C+X, where X is anything else) and $M(x_i)$ is the model cross section for that process given by NUCFRG2 or QMSFRG.
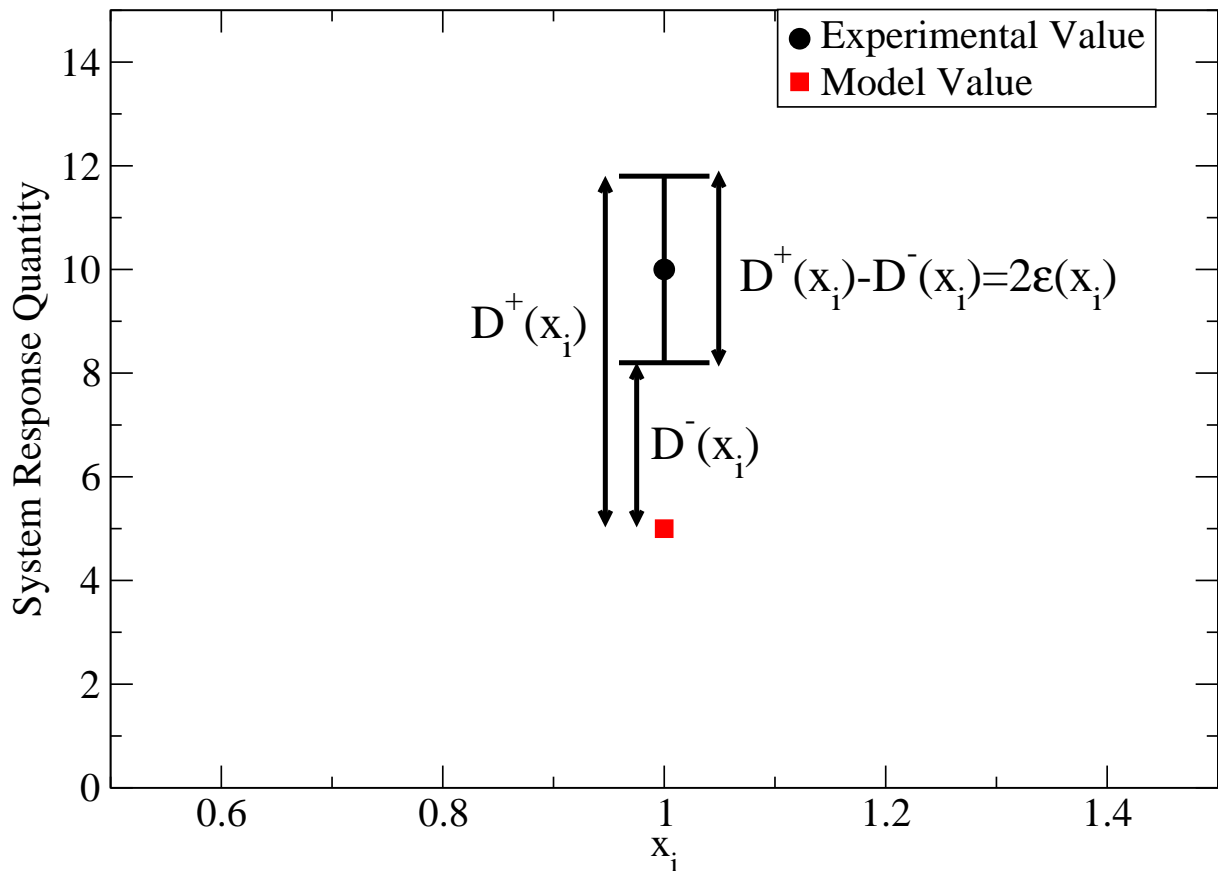
Figure 1: An illustration of the functions $D^+(x_i)$ and $D^-(x_i)$ for arbitrary model and experimental values.

We define the absolute model uncertainty U at the measurement point x_i as

$$U(x_i) \equiv \text{MAX}\left(|D^+(x_i)|, |D^-(x_i)|\right), \qquad (4)$$

where MAX denotes taking the maximum value of the two possibilities. U is an absolute uncertainty because it is positive definite and carries the same units as the model and experiment. It is important to understand the choice of the maximum value as opposed to the minimum value in Eq. (4). If the minimum value was used in Eq. (4), the metric would have the undesirable quality that when the uncertainty in the experiment increased, the metric would show that model was more accurate. This is because using the minimum value would correspond to testing for consistency of the model with the data. Choosing the maximum value, however, answers the question of how accurate the model is proven to be using the data. Therefore, the maximum value is used when we define the uncertainty in Eq. (4).

It should be noted that when the model result falls within the range of the experimental uncertainty, all the uncertainty is due to that of the experimental data. Also, note that in the limit

when the experimental uncertainty, ϵx_i, goes to zero, $U(x_i)$ reduces to the traditional definition of absolute uncertainty, $U(x_i) = |M(x_i) - E(x_i)|$. Since the uncertainty in the experimental measurement is available for all data in this work, a metric was created to quantify the experimental uncertainty. Since the experimental data were treated as an interval, the uncertainty in the experimental measurement is defined as

$$U^e(x_i) \equiv |D^+(x_i) - D^-(x_i)| = 2\epsilon(x_i). \tag{5}$$

Note that this method does not give information about the contribution to model uncertainty due to areas where there is a lack of experimental data.

Separating the uncertainty due to the experiment from the model uncertainty is important because it allows the validation analysis to account for the inherent uncertainty in the experiment without unjustly penalizing the model [1]. Moreover, separating the uncertainty in the model from the uncertainty due to the experiment helps determine if more accurate experiments would be useful to model validation.

The absolute model uncertainty, as defined above, has the same dimensional units as the model and experiment. In the case of nuclear fragmentation, an absolute uncertainty is useful when assessing the impact of the model on biological quantities such as radiation dose, since the models are used as inputs to radiation transport codes (i.e. HZETRN [5] and HETC-HEDS [40]) which are concerned with the biological impact of radiation. Because the risk involved with exposure to radiation is typically dependent on the absolute value of the exposure, discussion is restricted to absolute uncertainty. As a complement to the cumulative absolute uncertainty metric and for the sake of completeness, a cumulative relative uncertainty metric is developed in Appendix B.

A cumulative uncertainty distribution has been developed based on the idea of the cumulative distribution function. In general, the cumulative distribution function, $D(x)$, gives the probability that a random variable has a value less than or equal to x. If some random value y is chosen, the fraction of the database which has uncertainty less than or equal to y is defined as

$$\text{Fraction of Data} \equiv \frac{1}{n} \sum_{U=0}^{U_{\max}} P(U \leq y), \tag{6}$$

where n is the number of experimental data points, U is the absolute model uncertainty defined in Eq. (4) with the dependence of the specific point in phase space x_i suppressed, $U_{\max}$ is the largest value of uncertainty, and

$$P(U \leq y) \equiv \begin{cases} 1 & \text{if } U \leq y \\ 0 & \text{if } U > y \end{cases}. \tag{7}$$

The cumulative absolute uncertainty distribution is then given by the cumulative absolute uncertainty, y, versus the fraction of data. The results of the cumulative absolute uncertainty distributions for NUCFRG2 and interpolated QMSFRG are shown in Figs. 2 and 3, along with the cumulative distribution for the uncertainty due to the experiment. The cumulative
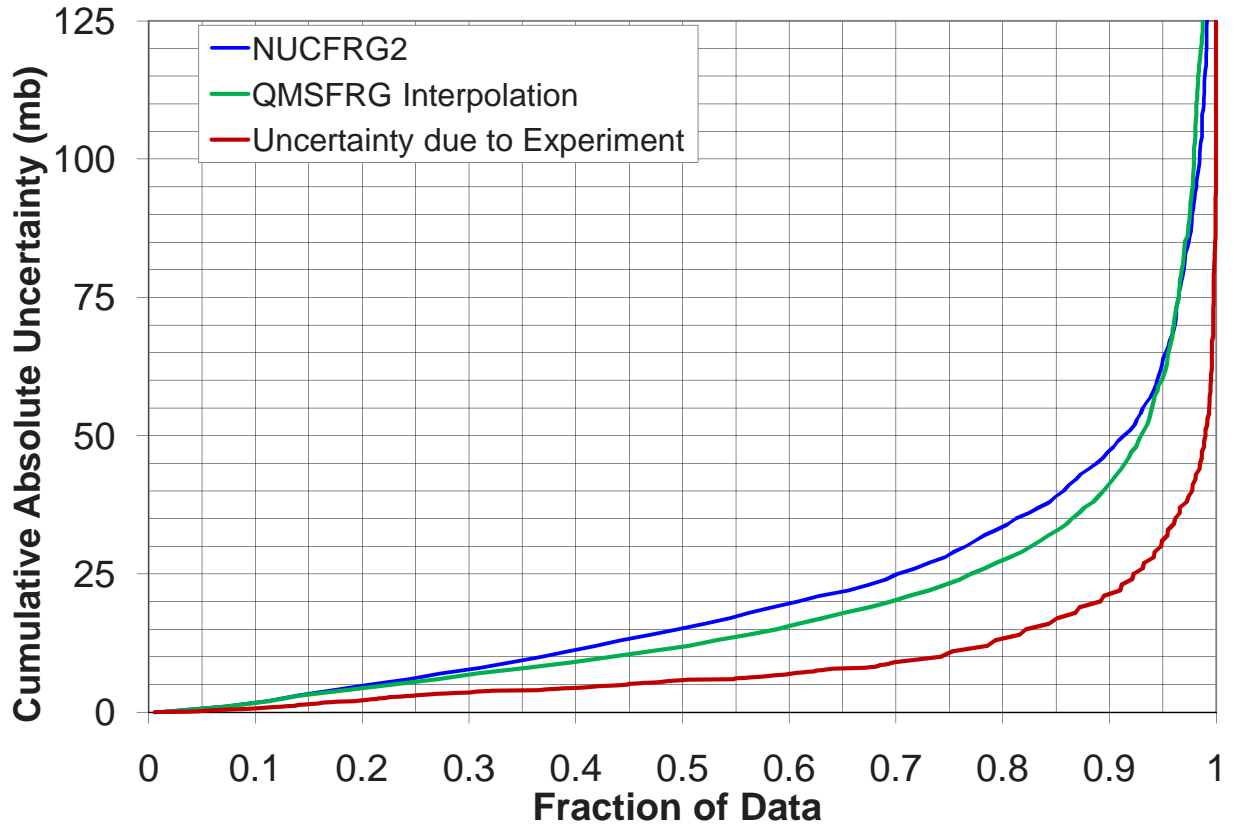
10

Figure 2: Cumulative absolute uncertainty distributions for NUCFRG2 and interpolated QMS-FRG compared to the experimental database, along with the distribution for the cumulative uncertainty due to experiment.

distribution for the uncertainty due to the experiment is found by substituting U^e for U in Eqs. (6) and (7). From Fig. 2, one might be tempted to conclude that QMSFRG is a more accurate model of the experimental database than NUCFRG2. For instance, at .50 fraction of data, there is an absolute uncertainty of 15 mb for NUCFRG2 and 12 mb for QMSFRG, while the uncertainty due to experiment at .50 fraction of data was 6 mb. For all values of fraction of data, however, the uncertainty due to the experiment is larger than the difference between the models. Therefore, within the accuracy of the experiments, we cannot conclude that either model is more accurate. In addition, if the larger values of the fraction of data are investigated, as shown in Fig. 3, QMSFRG has more outliers compared to NUCFRG2.

The more traditional metric of mean absolute error (MAE), which is defined as

$$MAE \equiv \frac{1}{n} \sum_{i=1}^{n} |M(x_i) - E(x_i)|, \tag{8}$$

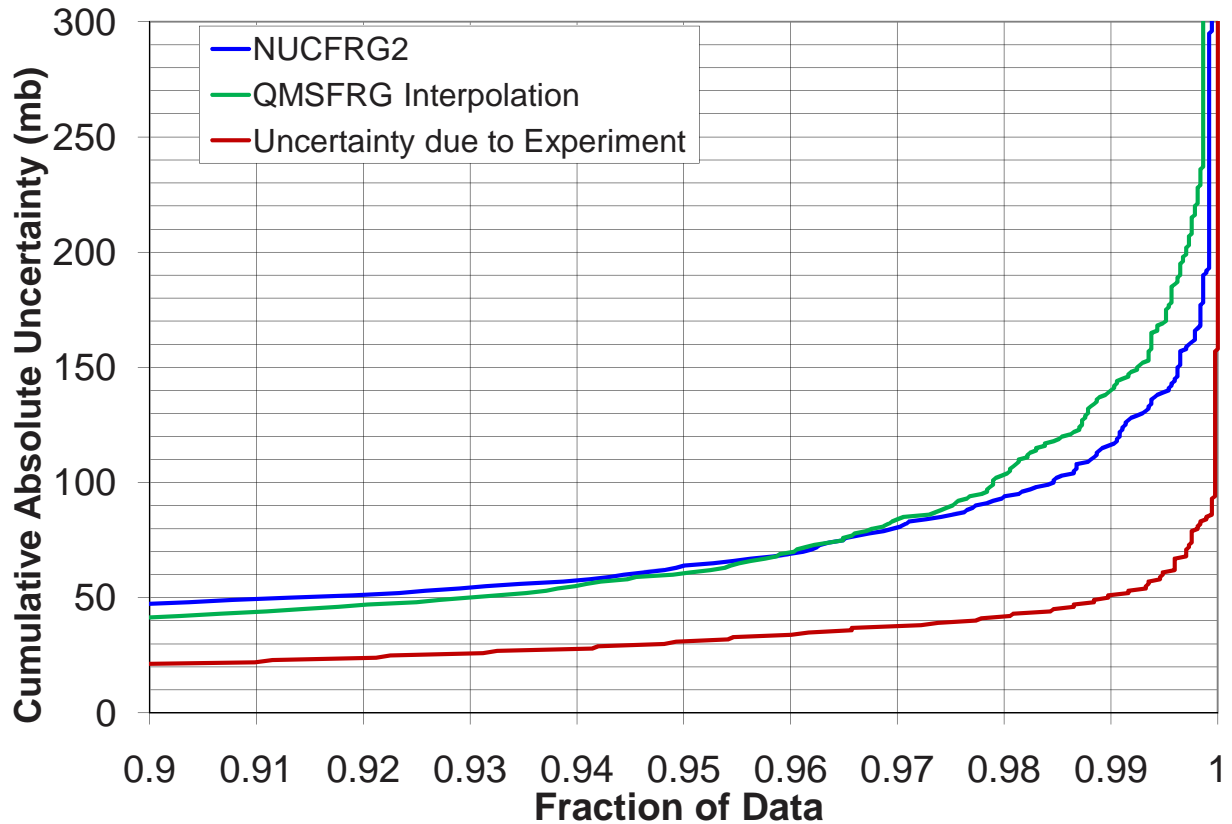is presented here as a contrast to the cumulative absolute uncertainty distribution. Traditional

Figure 3: Same as Fig. 2 with focus on the .9 to 1.0 fraction of data.

metrics, such as MAE, characterize uncertainty as a scalar value. The cumulative absolute uncertainty distribution, however, allows for much more flexibility in characterizing a model versus experiment. Not only does one have flexibility to choose the value of the fraction of data if one wishes to have a scalar value for uncertainty, but the cumulative absolute uncertainty distribution gives information for how the uncertainty varies over the whole experimental database. This functional behavior can be a powerful tool when evaluating uncertainty in a model especially when leveraged for use in an uncertainty propagation analysis. For instance, the cumulative absolute uncertainty functions for NUCFRG2 and interpolated QMSFRG could be used in an uncertainty propagation analysis for the radiation transport code HZETRN to determine which model contributed more to the overall uncertainty. The development of the cumulative uncertainty metric was motivated, in part, for use in uncertainty propagation. The functional form of the metric makes it easy to sample to propagate uncertainty.

As a comparison to the cumulative absolute uncertainty at .50 fraction of data, the results of the MAE for both NUCFRG2 and interpolated QMSFRG are presented. The differences between the models correlate well with the MAE which gives a value of 16 mb for NUCFRG2 and 15 mb for interpolated QMSFRG with a mean experimental uncertainty of 5 mb.

Analysis of the absolute uncertainty for both models found the difference between the models was smaller than the uncertainty due to the experiment. Therefore, neither model can be considered more accurate for the available data. However, an uncertainty propagation analysis using the cumulative absolute uncertainty functions might provide enough information to determine if one model is superior to the other.

3.2 Validation metric based on median statistics

In order to analyze the models more closely and identify regions of model disagreement with the experimental data, the validation phase space was split into subsets and each subset was analyzed with a newly developed metric based on robust statistical methods. Identifying regions of disagreement and agreement with experimental data allows model developers to define regions of model applicability and should give clues on how to extend and modify their model to better represent the experimental data. In the case of small or sparsely covered experimental datasets, the size of the database does not allow the use of the cumulative uncertainty distributions developed earlier. In addition, an outlier in the validation metric can quickly overwhelm the analysis when small or sparse experimental databases are used. The issue of outliers is at the center of the consideration of how to validate and choose a validation metric appropriately.

Using an established measure of uncertainty based on mean values (e.g. root mean square error) has the unwanted effect of magnifying the contribution of outliers. Traditionally, there are a few ways to deal with outliers: leave them as they are, correct, or delete them. If the outliers are left in the analysis, they skew the mean value which may nullify any meaningful conclusions. If they are altered (e.g. using a weighting function) or deleted (e.g. rejection rule) from the analysis, either method requires a clear justification that is very easily criticized [41].

A proposed solution to the problem of outliers in a validation effort utilizing a small or sparse experimental database based upon the use of the median value in place of the mean is presented. Given an ordered set $\{Y_i\}$ with n elements, the median of $\{Y_i\}$ is defined as

$$\tilde{Y} \equiv \begin{cases} Y_{(n+1)/2} & \text{if n is odd} \\ \frac{1}{2}[Y_{n/2} + Y_{(n+1)/2}] & \text{if n is even} \end{cases}. \tag{9}$$

The median is a more robust measure of the central tendency than the mean [41] and is used to develop a validation metric appropriate for small or sparse experimental datasets. Instead of developing a cumulative uncertainty distribution, as was done previously in Section 3.1, the functions are utilized to represent model uncertainty as a number. Although flexibility is lost when the uncertainty is distilled from a distribution to a number, the applicability of this validation metric to sparse experimental datasets for which a meaningful cumulative uncertainty distribution cannot be created is an important quality.

The new validation metric, the median uncertainty (MU), is defined as

$$MU \equiv \tilde{U} - \tilde{U}^e, \tag{10}$$

where $\tilde{U}$ is the median value of the set of total uncertainty values ,$\{U(x_i)\}$, and $\tilde{U}^e$ is the median of the set of the uncertainty due to the experiment values, $\{U^e(x_i)\}$. $U(x_i)$ and $U^e(x_i)$ are defined in Eqs. 4 and 5, respectively.

The median uncertainty defined in Eq. (10) is a measure of the consistency of the model with the experimental data. The aim of this metric is model development; therefore, consistency with the experimental data is what the metric should quantify. The median uncertainty, combined with the uncertainty due to the experiment, allows for an analysis of whether model validation would benefit more from resources invested in model development or investing in more precise measurements for the data currently available. Note that the question of whether additional data are needed where there are currently no data is not addressed by this metric.

3.3   Application of the Median Metric

The complete experimental database of nuclear fragmentation cross sections described previously has been divided into different subsections and analyzed. An analysis is presented using the median validation metric defined in Eq. (10) and using the same models and experimental database discussed earlier. The metric is applied as a function of the following parameters: projectile beam element, charge removed from the projectile ($\Delta Z \equiv Z_{\mathrm{projectile}} - Z_{\mathrm{fragment}}$), fragment charge, projectile energy, and target. When the validation metric is examined as a function of these parameters, differences that may not have been apparent when looking at the overall accuracy across the entire experimental database may become clear. By segregating the database in this way, the deficiencies in the models and experimental data can be identified. Additionally, areas that are important from the perspective of space radiation hazards can be identified.

When the models were analyzed as a function of projectile energy, both models were found to have the smallest median uncertainty at low projectile energies ($E_{\mathrm{proj}} \leq 300$ MeV/nucleon). The results of this analysis are shown in Table 2. Both models had an increasing median uncertainty at higher projectile energies. Interpolated QMSFRG seemed to have a steadily increasing median uncertainty with energy, while NUCFRG2 saw an almost factor of 2 increase from the $E \leq 300$ MeV/nucleon region to the 300 MeV/nucleon $< E_{\mathrm{proj}} \leq 1000$ MeV/nucleon region. It should also be noted that the median uncertainty due to the experiment for the $E \leq 300$ MeV/nucleon energy region was about twice as large as either of the model median uncertainties.

Fig. 4 shows the results of the median uncertainty analysis for NUCFRG2 and interpolated QMSFRG compared to the experimental data as a function of projectile element. The median uncertainty of the models defined in Eq. (10) is shown along with the median of the uncer-

Table 2: Median uncertainty for NUCFRG2 and interpolated QMSFRG along with the uncertainty due to the experiment for the experimental database [7, 11–39] shown for 3 projectile energy domains.

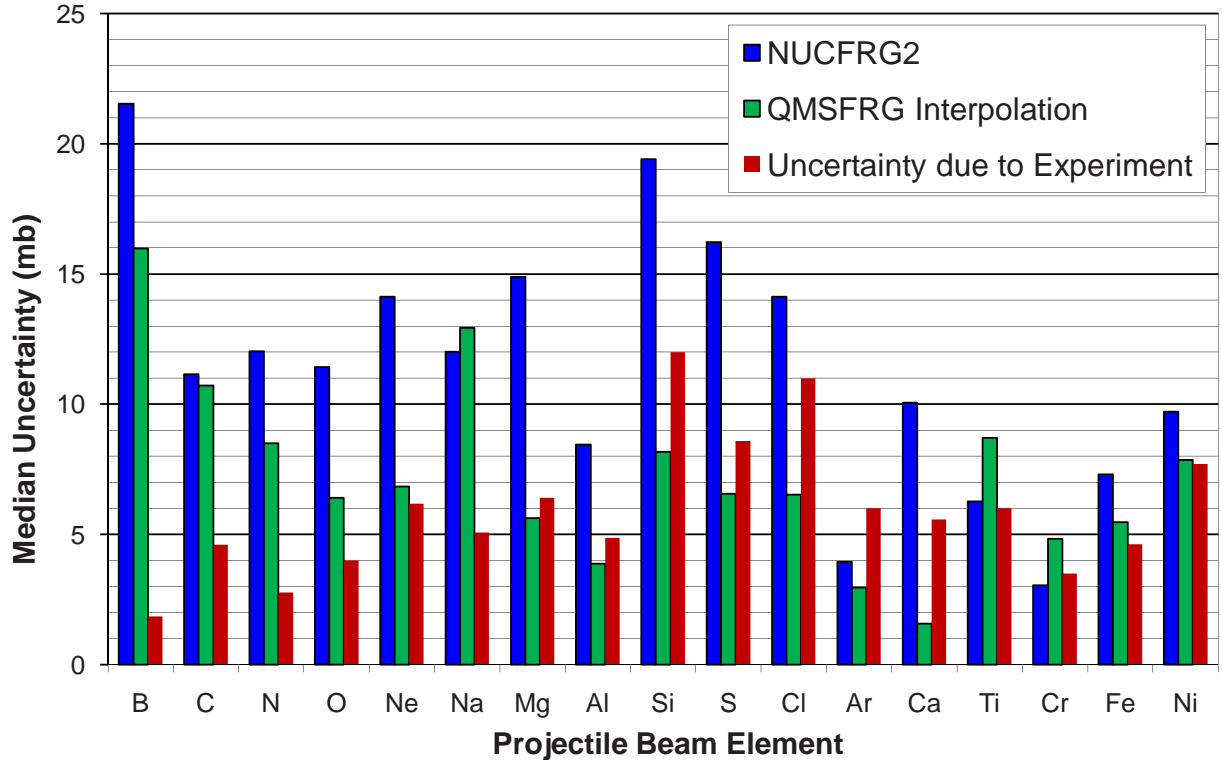| Energy Region (MeV/nucleon)	Median Uncertainty (mb)		
	NUCFRG2	QMSFRG Interpolation	Experiment
$E_{\mathrm{proj}} \leq 300$	5.0	3.9	8.8
$300 < E_{\mathrm{proj}} \leq 1000$	8.9	5.4	6.0
$E_{\mathrm{proj}} > 1000$	9.5	6.4	6.2

Figure 4: Median uncertainty for NUCFRG2 and interpolated QMSFRG along with the median uncertainty due to the experiment for the experimental database [7, 11–39] as a function of projectile type.

tainty due to the experiment. For the carbon-nitrogen-oxygen (CNO) group, which are a large flux component to the GCR environment, NUCFRG2 and interpolated QMSFRG were found to have comparable median uncertainties for carbon projectiles, while interpolated QMSFRG represents the data better for nitrogen and oxygen projectiles. In addition, it was observed that NUCFRG2 had an approximately constant median uncertainty for the CNO group, while interpolated QMSFRG saw a steadily decreasing median uncertainty with increasing mass for the CNO group. For silicon projectiles, the median uncertainty for NUCFRG2 is more than twice the median uncertainty for interpolated QMSFRG. For iron, which is also an important component of the GCR environment, there is a small (relative to most other projectile types) and comparable median uncertainty for both interpolated QMSFRG and NUCFRG2. In general, it was found that both models represent the experimental data better for heavier projectiles, $Z{\geq}18$, than for lighter projectiles. For NUCFRG2, the decreased median uncertainty for heavy projectiles, $Z{\geq}18$, compared to lighter projectiles comes from two sources. First, for larger mass nuclei, the energy levels of the nucleons which make up the nuclei are much more closely spaced in energy compared to smaller mass nuclei. Nuclear structure effects due to the energy levels of the nucleons will therefore have less of a role in the interactions of heavier nuclei. The second
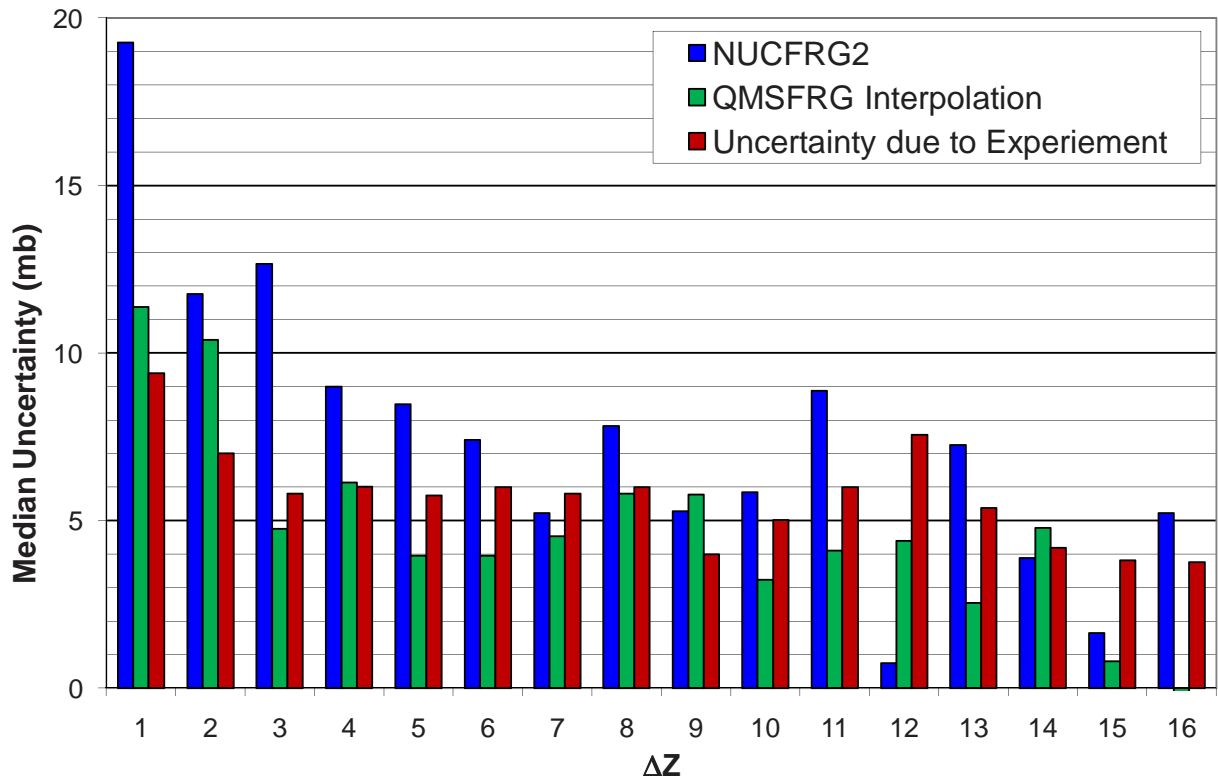
Figure 5: Same as Fig. 4, except as a function of ΔZ, the charge removed from the projectile.

source of decreased median uncertainty is related to the component of the model describing ablation which relies on statistical physics. This approach needs a large ensemble of particles to be accurate. This condition is not realized for lighter projectiles.

In Fig. 5, we show the results of our analysis based on the charge removed from the projectile, ΔZ. In general, we found for small values, $\Delta Z \leq 6$, interpolated QMSFRG was the better fit to the data. We also found that both models have the largest median uncertainty at $\Delta Z = 1$ with the median uncertainty of NUCFRG2 decreasing on average as the amount of charge removed from the projectile increased. In contrast, interpolated QMSFRG had a significant drop in median uncertainty for values of $\Delta Z \geq 3$ compared to $\Delta Z \leq 2$. The median uncertainty of interpolated QMSFRG is fairly constant between $3 \leq \Delta Z \leq 14$. Interpolated QMSFRG had a very small median uncertainty for $\Delta Z \geq 15$. For $\Delta Z = 16$, all the median uncertainty for interpolated QMSFRG is due to the experimental uncertainty.

An analysis based on the charge of the fragment is shown in Fig. 6. For a fragment charge of $Z = 25$, there is an anomalously large median uncertainty for both interpolated QMSFRG and NUCFRG2. This uncertainty is dominated by the data from iron projectiles and correlates with the previously found large median uncertainty for $\Delta Z = 1$. A very large median uncertainty was also found for $Z = 9$ and $Z = 6$ fragments for NUCFRG2. These uncertainties are likely due to important shell structure effects. The cross section for $Z = 9$ fragments has been found
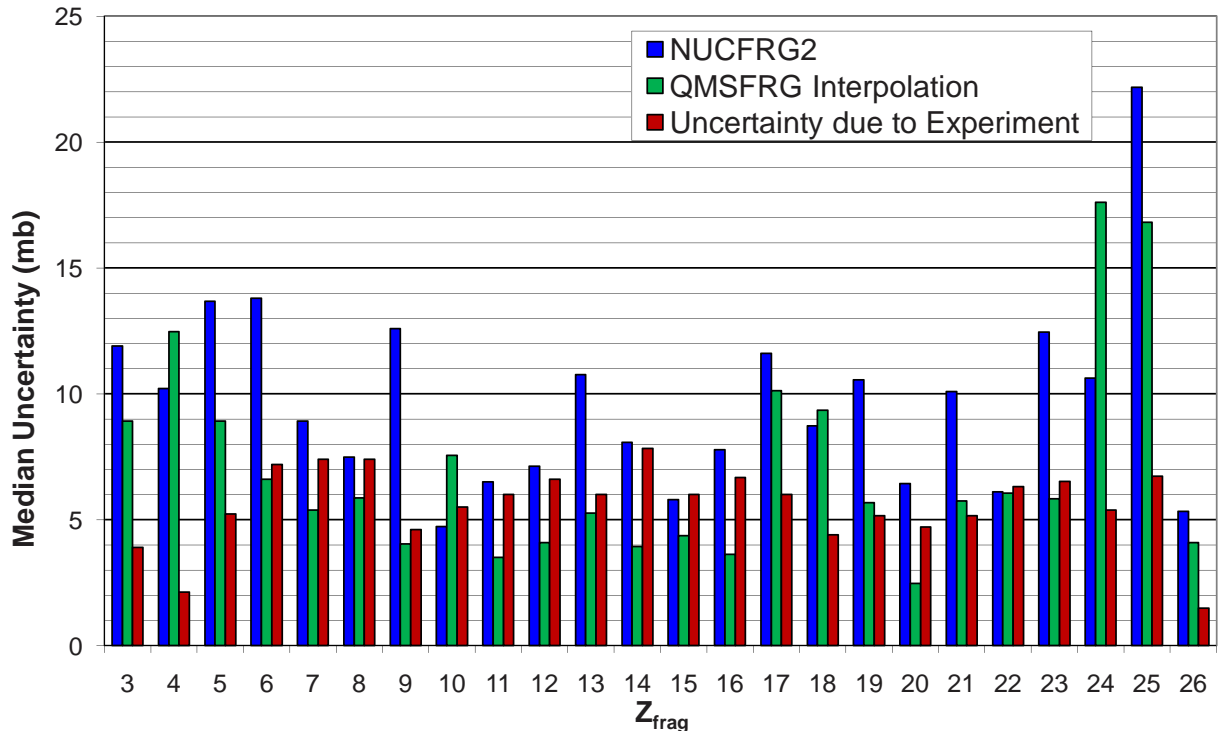
16

Figure 6: Same as Fig. 4, except as a function of the fragment charge, Z_{frag}.

to be a minimum value for most systems [12, 14, 15]. In the theory of shell structure of the nucleus, $Z = 8$ is what is known as a magic number and indicates increased stability of $Z = 8$ fragments produced in heavy ion fragmentation. Therefore, if a total charge of 9 is going to be produced from a heavy ion interaction, experiments indicate that it may be more likely to produce a combination of fragments (i.e. an oxygen and a proton fragment) rather than one fluorine fragment. For carbon fragments ($Z = 6$), shell structure should also play a role in that carbon has a closed $1p_{3/2}$ shell for protons. This closed shell causes carbon production to be slightly more likely than odd charged, nearest neighbor fragments. This is known as the "odd-even effect" and is seen in many projectile-target combinations. In addition, the lack of the odd-even effect in NUCFRG2 contributes to the increased uncertainty of the odd charged fragments with $Z \geq 17$. This complicated shell structure effect is not included in the classical model found in NUCFRG2, but is accounted for in the QMSFRG model.

An analysis based on the type of target is shown in Fig. 7. Both models follow the general trend of increasing median uncertainty with increasing target mass. The increasing median uncertainty with increasing target mass is likely due to a combination of effects. The first likely contribution is due to the larger cross sections for heavier targets. Secondly, it is known [7] that the simplified nuclear matter distributions utilized in NUCFRG2 contribute to the increased median uncertainty when compared to heavy targets. The QMSFRG model, however, accounts for density dependent corrections to the cross sections and utilizes accurate single body wave
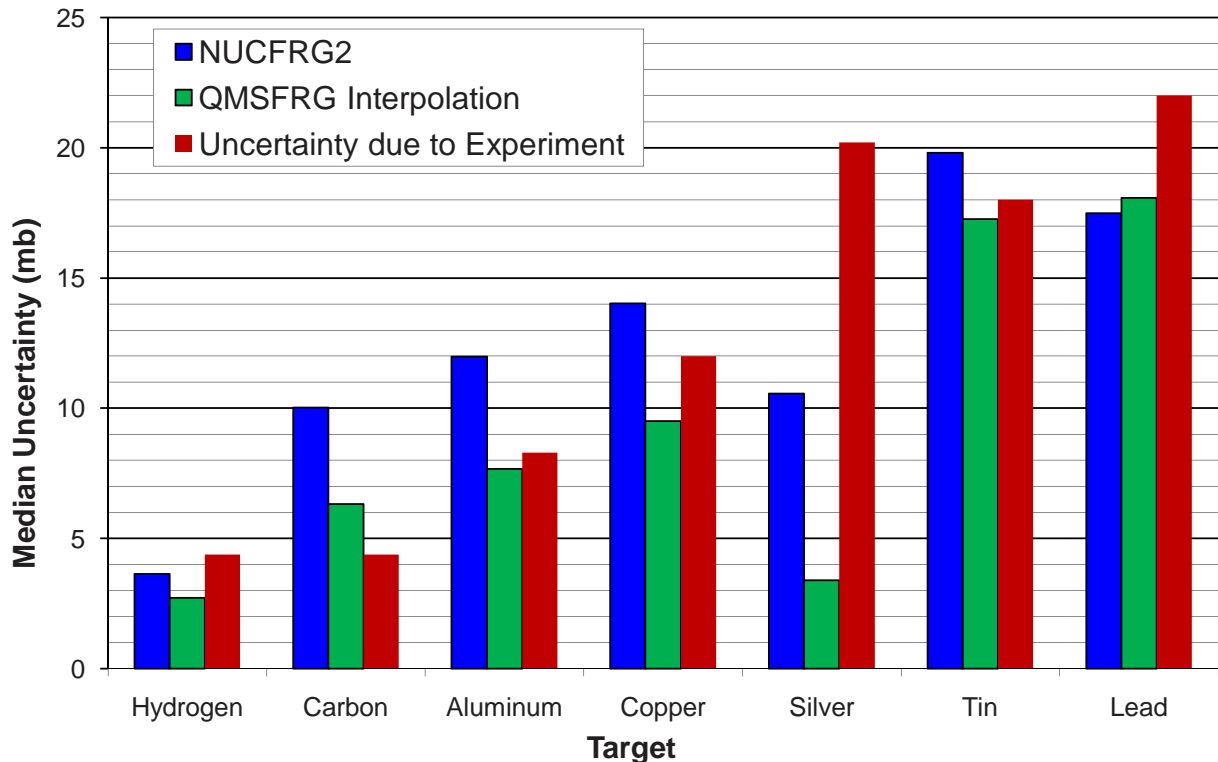
Figure 7: Same as Fig. 4, except as a function of target.

functions to build up the total nuclear wave function. Therefore, it is reasonable to assume that extrapolation of the database version of QMSFRG for targets of mass larger than Cu likely introduces an increasing error.

Analysis of the subregions of the validation phase space based on the median uncertainty metric allows for some recommendations for future model development. A large median uncertainty for the $\Delta Z=1$ cross sections implies investigating improvements to both models for this physical process. In addition, an increasing median uncertainty with larger energy and target mass signals the need for improved energy dependence and an improved description of nuclear matter in the models. The lack of the odd-even effect in NUCFRG2 did not overly affect the model's accuracy compared to interpolated QMSFRG, but including the odd-even effect would likely decrease the median uncertainty of NUCFRG2.

4 Summary and Conclusions

The need for a well defined process of verification and validation is essential to NASA's vision of a safe and reliable exploration of space. A comprehensive experimental database containing all data of interest to the validation program is also essential to a successful verification and validation effort. For this paper, a comprehensive literature search yielded an experimental

database with more then 3600 experimental nuclear fragmentation cross sections. The experimental database contained 25 distinct projectile isotopes from ^{10}B to ^{58}Ni, a projectile kinetic energy range from 90 MeV/nucleon to 14500 MeV/nucleon and both elemental and compound targets ranging from hydrogen to uranium. It is important to note that this should still be considered a sparse database, since the GCR background spans more than 5 orders of magnitude in energy and contains 128 isotopes of elements from hydrogen to nickel.

Two validation metrics were developed in this paper and applied to different components of the validation of nuclear physics models used by NASA. The first metric was used to quantify the overall accuracy of the nuclear physics models for the entire experimental database assembled for validation. To determine the overall accuracy, a cumulative absolute uncertainty distribution was introduced. The functional nature of the cumulative absolute uncertainty distribution allows for a quantitative assessment of the accuracy of a model compared to experiment, for flexibility in determining the desired confidence level of the results, and can be leveraged for uncertainty propagation in the radiation transport codes which utilize these nuclear models. When the overall accuracy of the models was analyzed using the cumulative absolute uncertainty distributions, an absolute model uncertainty of 15 mb and 12 mb was found at .50 fraction of data for NUCFRG2 and interpolated QMSFRG, respectively. The uncertainty due to experiment at .50 fraction of data was found to be 6 mb. For all values of the fraction of data, the uncertainty due to experiment was larger than the difference in uncertainty between the models. This fact led to the conclusion that neither model can be considered more accurate given the current experimental data available. Although the models were found to be quantitatively similar in their accuracy, fundamental qualitative differences in the models are known to exist.

A second type of metric was developed in this work based on median statistics which was appropriate for use with sparse or small experimental databases and it was applied to the task of model improvement and analysis. This metric was applied to the analysis of subsets of the experimental database based on different variables in the validation phase space and was developed specifically to be insensitive to outliers which may skew the analysis. Regions of model improvement and regions of model agreement for both models were identified using the median uncertainty based metric. The importance of quantum effects and shell structure to the nuclear fragmentation process was found through comparison of the quantum multiple scattering model QMSFRG to the classical, geometric approach of NUCFRG2. The importance of shell structure was most clearly shown for the median uncertainty analysis as a function of fragment charge where the median uncertainty of NUCFRG2 for fluorine and carbon fragments, along with odd charged fragments with $Z \geq 17$, was disproportionately large compared to other fragments. In general, our analysis found a steadily increasing median uncertainty with increasing target mass and increasing projectile energy for both models. In addition, a small median uncertainty was found for iron projectiles compared to other projectiles for both models and a general trend for the models to fit the data better for higher charge and mass projectiles. Lastly, a relatively large median uncertainty for $\Delta Z = 1$ fragments was found, where ΔZ is the difference between the projectile and fragment charge.

Given our analysis of the subsets of the validation phase space based on the median uncertainty metric, some recommendations for future model development can be made. A large median uncertainty for the $\Delta Z = 1$ cross sections was found and investigating improvements to

both models for this physical process is recommended. Improved energy dependence and an improved description of nuclear matter used in the models should be investigated to rectify the increasing median uncertainty with larger energy and target mass, respectively. The lack of the odd-even effect in NUCFRG2 did not overly affect the model's accuracy compared to interpolated QMSFRG, but including the odd-even effect would likely decrease the median uncertainty of NUCFRG2.

These results and conclusions are tempered by the completeness of the experimental data set. The application of these models for use in space radiation analysis creates a high-dimensional phase space of validation parameters. The GCR background spans more than 5 orders of magnitude in energy and contains 128 isotopes of elements from hydrogen to nickel. An analysis must contain data for all targets commonly used in spacecraft shielding and all possible fragments produced from interactions. The amount of time, effort, and money required to blanket this phase space with experiments is prohibitive, and therefore, the experimental database assembled for validation, though it consists of all known, published experimental data sets, does contain gaps. The question of whether the data are complete enough to capture all the important physics is currently being investigated. Further analysis the authors hope to perform is a sensitivity analysis to identify subregions of the phase space important for particular applications. The authors would also like to investigate the feasibility and effect of weighting schemes for the important subregions of phase space. In addition, the authros would like to carry out an uncertainty propagation study using the cumulative absolute uncertainty distributions for the radiation transport codes which use these nuclear models.

References

[1] W. L. Oberkampf and M. F. Barone, Measures of agreement between computation and experiment: Validation metrics, J. Comput. Phys. **217**, 5 (2006).

[2] W. L. Oberkampf, M. Pilch and T. G. Trucano, Predictive capability maturity model for computational modeling and simulation, Sandia Report SAND2007-5948 (2007).

[3] S. R. Blattnig, L. L. Green, J. M. Luckring, J. H. Morrison, R. K. Tripathi and T. A. Zang, Towards a credibility assessment of models and simulations, in *49th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, Shaumburg, Illinois, (2008).

[4] National Council on Radiation Protection and Measurements, Information needed to make radiation protection recommendations for space missions beyond low-Earth orbit, NCRP Report no. 153, (2006).

[5] J. W. Wilson, R. K. Tripathi, F. F. Badavi and F. A. Cucinotta, Standardized radiation shield design methods: 2005 HZETRN, in *Proceedings of the 36th Conference on Environmental Systems (ICES)*, Norfolk, Virginia, (2006).

[6] F. A. Cucinotta, M.-H. Y. Kim and L. Ren, Evaluating shielding effectiveness for reducing space radiation cancer risks, Radiat. Meas. **41**, 1103 (2006).

[7] J. W. Wilson, R. K. Tripathi, F. A. Cucinotta, J. L. Shinn, F. F. Badavi, S. Y. Chun, J. W. Norbury, C. J. Zeitlin, L. Heilbronn and J. Miller, NUCFRG2 An evaluation of the semiempirical nuclear fragmentation database, NASA Technical Paper 3533 (1995).

[8] F. A. Cucinotta, M.-H. Kim, S. I. Schneider and D. M. Hassler, Description of light ion production cross sections and fluxes on the Mars surface using the QMSFRG model, Radiat. Environ. Biophys. **46**, 101 (2007).

[9] J. D. Bowman, W. J. Swiatecki and C. F. Tsang, Abrasion and ablation of heavy ions, Lawrence Berkeley Laboratory Report LBL-2908 (1973).

[10] R. J. Glauber, *Lectures in theoretical physics* (Interscience, New York, 1959), p. 315.

[11] C. Zeitlin, A. Fukumura, L. Heilbronn, Y. Iwata, J. Miller and T. Murakami, Fragmentation cross sections of 600 MeV/nucleon ^{20}Ne on elemental targets, Phys. Rev. C **64**, 024902 (2001).

[12] C. Zeitlin, L. Heilbronn, J. Miller, S. E. Rademacher, T. Borak, T. R. Carter, K. A. Frankel, W. Schimmerling and C. E. Stronach, Heavy fragment production cross sections from 1.05 GeV/nucleon ^{56}Fe in C, Al, Cu, Pb, and CH_2 targets, Phys. Rev. C **56**, 388 (1997).

[13] C. Zeitlin and S. Guetersloh and L. Heilbronn and J. Miller and A. Fukumura and Y. Iwata and T. Murakami, Fragmentation cross sections of 290 MeV/nucleon and 400- MeV/nucleon ^{12}C beams on elemental targets, Phys. Rev. C **76**, 014911 (2007).

[14] C. Zeitlin, A. Fukumura, S. B. Guetersloh, L. H. Heilbronn, Y. Iwata, J. Miller and T. Murakami, Fragmentation cross sections of ^{28}Si at beam energies from 290 A MeV to 1200 A MeV, Nucl. Phys. A **784**, 341 (2007).

[15] C. Zeitlin, S. Guetersloh, L. Heilbronn, J. Miller, A. Fukumura, Y. Iwata, T. Murakami, L. Sihver and D. Mancusi, Fragmentation cross sections of medium-energy ^{35}Cl, ^{40}Ar, and ^{48}Ti beams on elemental targets, Phys. Rev. C **77**, 034605 (2008).

[16] W. R. Webber, J. C. Kish and D. A. Schrier, Individual charge changing fragmentation cross sections of relativistic nuclei in hydrogen, helium, and carbon targets, Phys. Rev. C **41**, 533 (1990).

[17] W. R. Webber, J. C. Kish and D. A. Schrier, Individual isotopic fragmentation cross sections of relativistic nuclei in hydrogen, helium, and carbon targets, Phys. Rev. C **41**, 547 (1990).

[18] J. R. Cummings, W. R. Binns, T. L. Garrard, M. H. Israel, J. Klarmann, E. C. Stone and C. J. Waddington, Determination of the cross sections for the production of fragments from relativistic nucleus-nucleus interactions. I. Measurements, Phys. Rev. C **42**, 2508 (1990).

[19] G. D. Westfall, L. W. Wilson, P. J. Lindstrom, H. J. Crawford, D. E. Greiner and H. H. Heckman, Fragmentation of relativistic ^{56}Fe, Phys. Rev. C **19**, 1309 (1979).

[20] D. L. Olson, B. L. Berman, D. E. Greiner, H. H. Heckman, P. J. Lindstrom and H. J. Crawford, Factorization of fragment-production cross sections in relativistic heavy-ion collisions, Phys. Rev. C **28**, 1602 (1983).

[21] C. E. Tull, Relativistic heavy ion fragmentation at HISS, Lawrence Berkeley Laboratory Report LBL-29718 (1990).

[22] A. N. Golovchenko, J. Skvarč, N. Yasuda, M. Giacomelli, S. P. Tretyakova, R. Ilić, R. Bimbot, M. Toulemonde and T. Murakami, Total charge-changing and partial cross-section measurements in the reactions of $\sim 110 - 250$ MeV/nucleon ^{12}C in carbon, paraffin, and water, Phys. Rev. C **66**, 014609 (2002).

[23] S. Momota et al., Production of projectile-like fragments at intermediate energies, Nucl. Phys. A **701**, 150 (2002).

[24] C. Brechtmann and W. Heinrich, Measurements of elemental fragmentation cross section for relativistic heavy ions using CR39 plastic nuclear track detectors, Nucl. Instr. and Meth. B **29**, 675 (1988).

[25] C. Brechtmann and W. Heinrich, Fragmentation Cross Sections of ^{32}S at 0.7, 1.2 and 200 GeV/Nucleon, Z. Phys. A **331**, 463 (1988).

[26] C. Brechtmann, W. Heinrich and E. V. Benton, Fragmentation cross sections of ^{28}Si at 14.5 GeV/nucleon, Phys. Rev. C **39**, 2222 (1989).

[27] C. Brechtmann, H. Drechsel, J. Beer and W. Heinrich, Cross sections for the production of fragments with Z≥8 by fragmentation of 9≤Z≤26 nuclei, Nucl. Tracks **12**, 361 (1986).

[28] F. Flesch, G. Iancu, W. Heinrich and H. Yasuda, Projectile fragmentation of silicon ions at 490 A MeV, Radiat. Meas. **34**, 237 (2001).

[29] A. Leistenschneider et al., Fragmentation of unstable neutron-rich oxygen beams, Phys. Rev. C **65**, 064607 (2002).

[30] G. Iancu, F. Flesch and W. Heinrich, Nuclear fragmentation cross-sections of 400 A MeV ^{36}Ar and ^{40}Ar in collisions with light and heavy target nuclei, Radiat. Meas. **39**, 525 (2005).

[31] Y. P. Viyogi et al., Fragmentations of ^{40}Ar at 213 MeV/Nucleon, Phys. Rev. Lett. **42**, 33 (1979).

[32] A. Korejwo, T. Dzikowski, M. Giller, J. Wdowczyk, V. V. Perelygin and A. V. Zarubin, The measurement of isotopic cross sections of ^{12}C beam fragmentation on liquid hydrogen at 3.66 GeV/nucleon, J. Phys. G: Nucl. Part. Phys. **26**, 1171 (2000).

[33] A. Korejwo, M. Giller, T. Dzikowski, V. V. Perelygin and A. V. Zarubin, Isotopic cross-sections of ^{12}C fragmentation on hydrogen measured at 1.87 and 2.69 GeV/nucleon, J. Phys. G: Nucl. Part. Phys. **28**, 1199 (2002).

[34] M. Caamano, D. Cortina-Gil, K. Summerer, J. Benlliure, E. Casarejos, H. Geissel, G. Munzenberg and J. Pereira, Production cross-sections and momentum distributions of fragments from neutron-deficient ^{36}Ar at 1.05 A GeV, Nucl. Phys. A **733**, 187 (2004).

[35] C. N. Knott et al., Interactions of relativistic neon to nickel projectiles in hydrogen, elemental production cross sections, Phys. Rev. C **53**, 347 (1996).

[36] C. N. Knott et al., Interactions of relativistic ^{36}Ar and ^{40}Ar nuclei in hydrogen: Isotopic production cross sections, Phys. Rev. C **56**, 398 (1997).

[37] P. Napolitani, K.-H. Schmidt, A. S. Botvina, F. Rejmund, L. Tassan-Got and C. Villagrasa, High-resolution velocity measurements on fully identified light nuclides produced in ^{56}Fe+hydrogen and ^{56}Fe+titanium systems, Phys. Rev. C **70**, 054607 (2004).

[38] S. E. Hirzebruch, W. Heinrich, K. D. Tolstov, A. D. Kovalenko and E. V. Benton, Fragmentation cross sections of ^{16}O between 0.9 and 200 GeV/nucleon, Phys. Rev. C **46**, 1487 (1992).

[39] D. Sampsonidis, E. Papanastassiou, M. Zamani, M. Debeauvais, J. C. Adloff, B. A. Kulakov, M. I. Krivopustov and V. S. Butsev, Fragmentation cross sections of ^{16}O, ^{24}Mg, and ^{32}S projectiles at 3.65 GeV/nucleon, Phys. Rev. C **51**, 3304 (1995).

[40] L. W. Townsend, T. M. Miller and T. A. Gabriel, HETC radiation transport code development for cosmic ray shielding applications in space, Rad. Prot. Dos. **116**, 135 (2005).

[41] J. W. Müller, Possible advantages of a robust evaluation of comparisons, J. Res. Natl. Inst. Stand. Technol. **105**, 551 (2000).

Appendix A   QMSFRG Interpolation

Table 3: QMSFRG cross sections reported in Ref. [8] published 2007.

2100 MeV/u ^{16}O+^{1}H $\to$	QMSFRG (mb)	Interpolation (mb)	Percent Error
n	284.1	292.4	2.96
p	310.2	312	.58
^{2}H	40.3	41.3	2.48
^{3}H	17.3	16.7	3.47
^{3}He	17.3	16.7	3.47
^{4}He	168.9	172.0	1.84
^{16}O+^{12}C $\to$			
n	2697	2717	.74
p	2902	2906	.14
^{2}H	459.6	503.1	9.46
^{3}H	140.4	130.7	6.91
^{3}He	103.7	95.4	8.00
^{4}He	457.7	438	4.30
^{16}O + Cu $\to$			
n	6459	6493	.53
p	6941	6944	.04
^{2}H	817.4	872.1	6.69
^{3}H	250.3	225.2	10.03
^{3}He	194.7	174.5	10.37
^{4}He	695.6	639.5	8.06
		RMS Error	5.68%

The version of QMSFRG used in the analysis of this paper is not the standard version. Access to standard QMSFRG through a machine executable file or source code was not available. Without access to source code, a QMSFRG produced cross section table was used that applies linear interpolation over eight energy points (25, 75, 150, 300, 600, 1200, 2400, 7200 MeV/nucleon) and linear interpolation/extrapolation over six target isotopes (^{1}H, ^{12}C, ^{16}O, ^{27}Al, ^{40}Ca, ^{56}Fe).

Table 3 shows the published version of some QMSFRG fragmentation cross sections (labeled QMSFRG in the table) from Ref. [8] and the results from the QMSFRG interpolation database (labeled Interpolation) along with the percent error between between the two. The QMSFRG cross sections results were published in 2007, and this is the most recent paper found with published values of QMSFRG cross sections explicitly provided.

Table 4 shows the published version of some QMSFRG fragmentation cross sections (labeled QMSFRG in the table) from Ref. [11]. For the reactions ^{20}Ne +^{12}C $\to Z$ and at ^{20}Ne + Al $\to Z$ at 600 A MeV, exact results from the QMSFRG code used to generate the cross section database

Table 4: QMSFRG cross sections reported in Ref. [11] published 2001.

600 A MeV	QMSFRG (mb)	Interpolation (mb)	Percent Error
^{20}Ne+^{12}C $\rightarrow Z$			
9	93	96	3.23
8	125	114	8.8
7	96	96	0
6	125	127	1.6
5	58	39	32.76
^{20}Ne + Al $\rightarrow Z$			
9	117	125	6.84
8	158	143	9.49
7	121	119	1.65
6	153	155	1.31
5	66	46	30.3
^{20}Ne + Cu $\rightarrow Z$			
9	134	148	10.45
8	206	161	21.84
7	131	135	3.05
6	185	178	3.78
5	54	53	1.85
		RMS Error:	13.75%

are shown. Definite differences between the the database and the published results exist, most notably for boron and oxygen fragments. These differences are most likely due to changes to the QMSFRG code over time.

Additionally, the results from 2001 [11] (Table 4) are summed over isotopes for a given value of the fragment charge. This summation, when compared to an isotopic analysis such as that presented in Ref. [8], tends to average out the error, and specific isotopic inconsistencies that might have shown up in a more rigorous RMS error analysis are lost.

In summary, to definitively determine the error caused by the interpolation/extrapolation code compared to the full QMSFRG version would require published cross sections, over a range of energies, incident beams, and targets with isotopic resolution of many different fragments. Without access to source code, the authors are unable to decouple the error introduced from interpolation from the error due to the QMSFRG model.

Appendix B   Cumulative Relative Uncertainty

A cumulative relative uncertainty metric is presented here. Although not of importance to the present analysis, a relative uncertainty can be of interest in a validation program. For instance, in the case of high linear energy transfer (LET) radiation having a small flux compared to other radiation present in the analysis. An absolute value of uncertainty based on the flux, however, might under-represent the importance of this radiation from a biological point of view. In the case of high LET radiation or any other SRQ which might be better characterized by a relative value, a validation metric based on relative differences might be more appropriate. If a relative quantity is of interest, Eqs. (2) and (3) can be redefined as

$$D^+ \equiv \frac{M(x_i) - [E(x_i) + \epsilon(x_i)]}{E(x_i) + \epsilon(x_i)}, \tag{B.1}$$

$$D^- \equiv \frac{M(x_i) - [E(x_i) - \epsilon(x_i)]}{E(x_i) - \epsilon(x_i)}, \tag{B.2}$$

with the relative model uncertainty defined as

$$U(x_i) \equiv \text{MAX}\left(|D^+(x_i)|, |D^-(x_i)|\right), \tag{B.3}$$

which is consistent with Eq. (4). The relative uncertainty due to the experiment is given by

$$U^e(x_i) \equiv |D^+(x_i) - D^-(x_i)|. \tag{B.4}$$

It should be noted that the difference functions D^+ and D^-, defined in Eqs. (B.1) and (B.2) respectively, are not relative to the same value. The difference functions were defined in this way to be consistent with the interval representation of the data discussed in Section 3.1. This inconsistency in the difference functions, however, leads Eq. (B.4) to have a dependence on the model value. Expanding Eq. (B.4) gives,

$$U^e = \frac{2\epsilon M}{E^2 - \epsilon^2}, \tag{B.5}$$

where the dependence on the validation phase space point x_i has been omitted. The linear dependence of U^e on the model value M can lead to some confusion in how to interpret Eq. (B.5) since each model will have a distinct value of U^e for the same experimental data. One interpretation of U^e for the relative case is the contribution of the experimental measurement uncertainty relative to E weighted by the ratio of M/E. This interpretation is the first order approximation of Eq. (B.5) when the ratio ϵ/E is small, which should be the case for most "good" experiments. Eq. (B.5) can then be interpreted as a measure of the importance of the uncertainty due to the experiment relative to the both the experiment and the model. In addition, it is important to remember that the uncertainty due to the experiment should always be used in conjunction with the total uncertainty from Eq. (B.3) since the uncertainty due to the experiment will be largely used to determine if more precise experiments are needed to help determine model accuracy.
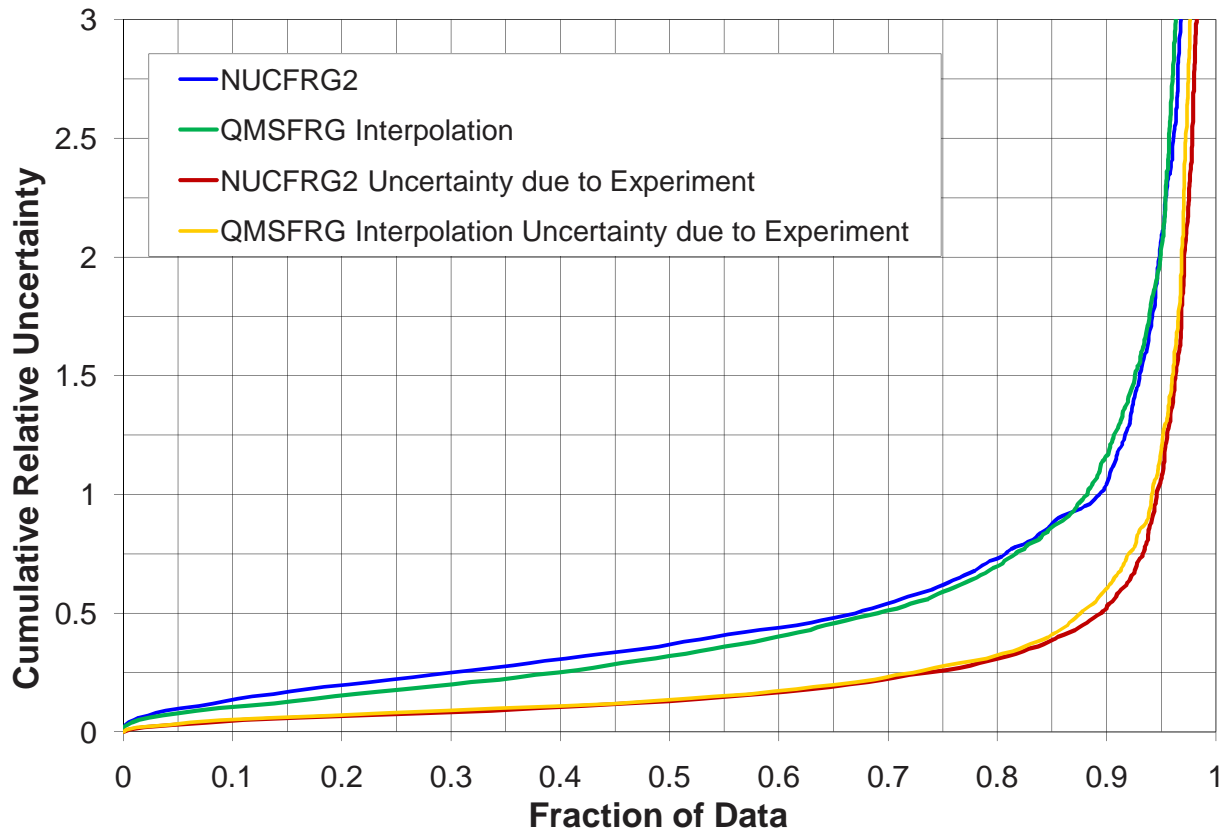
Figure 8: Cumulative relative uncertainty distributions for NUCFRG2 and interpolated QMSFRG compared to the experimental database, along with the distributions for cumulative uncertainty due to the experiment for both models.

To develop the cumulative relative uncertainty distribution for uncertainty, the same procedure outlined in Section 3.1.1 is followed. Fig. 8 shows the results of the relative uncertainty distributions for NUCFRG2 and QMSFRG compared to the experimental database. At a .50 fraction of data, NUCFRG2 has a 23% relative uncertainty due to the model while QMSFRG was found to have a relative uncertainty of 17%.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
01-02-2010	Technical Publication	

4. TITLE AND SUBTITLE

A Comprehensive Validation Methodology for Sparse Experimental Data

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Norman, Ryan B.; Blattnig, Steve R.

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

651549.02.07.01

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

NASA Langley Research Center
Hampton, VA 23681-2199

8. PERFORMING ORGANIZATION REPORT NUMBER

L-19828

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

National Aeronautics and Space Administration
Washington, DC 20546-0001

10. SPONSOR/MONITOR'S ACRONYM(S)

NASA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

NASA/TP-2010-216200

12. DISTRIBUTION/AVAILABILITY STATEMENT

Unclassified - Unlimited
Subject Category 93
Availability: NASA CASI (443) 757-5802

13. SUPPLEMENTARY NOTES

14. ABSTRACT

A comprehensive program of verification and validation has been undertaken to assess the applicability of models to space radiation shielding applications and to track progress as models are developed over time. The models are placed under configuration control, and automated validation tests are used so that comparisons can readily be made as models are improved. Though direct comparisons between theoretical results and experimental data are desired for validation purposes, such comparisons are not always possible due to lack of data. In this work, two uncertainty metrics are introduced that are suitable for validating theoretical models against sparse experimental databases. The nuclear physics models, NUCFRG2 and QMSFRG, are compared to an experimental database consisting of over 3600 experimental cross sections to demonstrate the applicability of the metrics. A cumulative uncertainty metric is applied to the question of overall model accuracy, while a metric based on the median uncertainty is used to analyze the models from the perspective of model development by analyzing subsets of the model parameter space.

15. SUBJECT TERMS

Space radiation; Model validation; Nuclear models

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			STI Help Desk (email: help@sti.nasa.gov)
U	U	U	UU	34	19b. TELEPHONE NUMBER *(Include area code)* (443) 757-5802

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18