M&C 2017 - International Conference on Mathematics & Computational Methods Applied to Nuclear Science & Engineering, Jeju, Korea, April 16-20, 2017, on USB (2017)

Maximum Likelihood Estimators and Extreme Value Theory for Uncertainty Quantification in Criticality Analysis

Max Shepherd, Geoff Dobson, Ben Lindley, Chris Baker, Simon Richards,* Paul Smith

ANSWERS Software Service, Amec Foster Wheeler, Kings Point House, Queen Mother Square, Poundbury, Dorchester, Dorset, DT1 3BW, United Kingdom.
simon.richards@amecfw.com

Abstract - *Uncertainty quantification is an important topic in criticality analysis. In this paper two approaches are investigated, one based on the use of Maximum Likelihood estimators for the estimation of bias in calculations and one based on Extreme Value Theory which is investigated in the context of assessing tolerance limits on k_{eff}. The Maximum Likelihood estimator is a very general approach, and here single and multiple parameter models are investigated for both continuous and discrete parameters. The discrete parameter case can be related to groups of experiments which share common apparatus, which could lead to correlations in the experimental bias. The statistical models and their results are compared with linear regression models where appropriate. The Extreme Value Theory approach is applied to a simple model, and the method is compared against the commonly-used Wilks approach.*

I. INTRODUCTION

In criticality analysis, it is important to account for various sources of uncertainty to ensure subcriticality. Sources of uncertainties include:

- code bias;
- manufacturing tolerances;
- nuclear data; and
- statistical uncertainties resulting from use of the Monte Carlo method.

In recent years, there has been a trend towards quantification of these uncertainties, with the goal of reducing conservatisms and better quantifying safety margins. In this paper, methods for quantifying these uncertainties are developed, with application to the UK industry standard MONK® Monte Carlo code [1], developed by Amec Foster Wheeler.

Bias refers to a systematic difference between calculations from the criticality code and experimental results. Experiments from a validation database can be used to estimate any bias in the code for physically-similar systems to the chosen application. Classification systems are available to assist with experiment selection, though these must be combined with engineering judgement. In this paper, Maximum Likelihood Estimation (MLE) based approaches are developed for bias estimation, and applied to the MONK validation database. MLE is a statistical technique which chooses statistically-optimal models to explain observed data. In this context, these techniques do not require considerable computational expense and can be used to validate engineering judgement.

For manufacturing tolerances and nuclear data uncertainties, a sampling method can be used, where covariances on engineering and nuclear data parameters are defined, a sampling scheme is specified, and a number of cases are sampled. A recent example of such an approach is where the nuclear data libraries themselves are generated using a sampling process [2]. Along with statistical uncertainties, this results in a distribution of k-effective (k_{eff}) values. These can be used to derive a tolerance interval on the probability that a percentage of samples will exceed a given value of k_{eff}. In this paper, Extreme Value Theory (EVT) is applied to this problem, giving potential benefits over other approaches. EVT is the branch of statistics associated with understanding rare events – exactly what we are interested in here.

II. MAXIMUM LIKELIHOOD ESTIMATORS

MONK has a validation database consisting of a wide variety of comparisons between experiment and calculated k_{eff} values with different nuclear data libraries. Systems span a range of fuel compositions, fuel types, geometries, fuel to moderator ratios and other physical parameters. Useful properties are also tabulated such as energy of neutrons causing capture and fission. The objective of the current work is to use the discrepancies between experiment and calculated values to estimate the bias in calculations for criticality problems. The natural framework for this is to use MLE. This comprises:

- assuming an underlying (parameterized) family of distributions that characterize observed data; and
- solving for the value of the parameters in that family of distributions that make the observed data most likely.

Given some observations $\boldsymbol{x_i}$ from a probability distribution with parameters $\boldsymbol{\theta}$, the likelihood $\mathcal{L}$ for a given i is:

$$\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{x_i}) \equiv \mathbb{P}(\boldsymbol{x_i}|\boldsymbol{\theta})\,. \tag{1}$$

For a set of n independent, identically-distributed observations $\boldsymbol{x_1}, \ldots, \boldsymbol{x_n}$, the likelihood function is:

$$\mathbb{P}(\boldsymbol{x_1}, \ldots, \boldsymbol{x_n}|\boldsymbol{\theta}) = \prod_{i=1}^{n} \mathbb{P}(\boldsymbol{x_i}|\boldsymbol{\theta}) = \mathcal{L}(\boldsymbol{\theta}|\boldsymbol{x_1}, \ldots \boldsymbol{x_n})\,. \tag{2}$$

Because of the form of distributions commonly assumed, it is often more convenient to work with the log likelihood

function, which can be justified since log is monotonic, so we solve:

$$\underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmax}} \log \mathcal{L}(\boldsymbol{\theta}|\boldsymbol{x_1},\ldots\boldsymbol{x_n}), \tag{3}$$

where $\boldsymbol{\Theta}$ is the set of all possible values of $\boldsymbol{\theta}$. That is, we choose the parameter values that make the observed data most likely. Here argmax, applied to some function, gives the points of the domain of the function at which the function is maximized – in other words, the arguments of the maxima. Note that the parameter values may not be uniquely defined by this process. We assume that the underlying data have a normal distribution, since:

- Normal distributions are uniquely characterized by their mean and variance (and generally this is the only information we have);
- we can do online updating with a Normal distribution as it is self-conjugate; and
- the equations to solve are smooth and well behaved.

We assume that the bias b_i (defined as calculated value − experimental value) for each of i different experiments is linearly dependent on m known parameters z_{ij}, where $j = 1,\ldots,m$. In this case, we may express the bias as:

$$b_i = \gamma + \sum_j P_j z_{ij} + \epsilon_i, \tag{4}$$

where

- $\gamma \sim \mathcal{N}(\mu_0, \sigma_0^2)$ is a normally distributed constant offset to the bias;
- $P_j \sim \mathcal{N}(\mu_j, \sigma_j^2)$ are normally distributed coefficients of the assumed linear dependence; and
- $\epsilon_i \sim \mathcal{N}(0, \rho_i^2)$ is the distribution of the error of experiment i with quoted experimental uncertainty ρ_i.

From here it is possible to derive the distribution of b_i and hence the log likelihood function that we wish to maximize:

$$\sum_i \Big(-\frac{1}{2}\log(2\pi s_i^2) - \frac{(b_i - m_i)^2}{2s_i^2} \Big), \tag{5}$$

where

- $m_i = \mu_0 + \sum_j \mu_j z_{i,j}$; and
- $s_i^2 = \sigma_0^2 + \sum_j \sigma_j^2 z_{i,j}^2 + \rho_i^2$

Although it is possible to differentiate the log likelihood function, the equations that result do not appear to be amenable to analytic solution so we solve this maximization problem numerically.

It is worth here comparing the classical linear regression model with the MLE model described above, in order to highlight the similarities and the differences.

The basic linear regression model comes from considering the following model.

$$b_i = \gamma + \sum_j P_j z_{ij} + \epsilon, \tag{6}$$

where in this case

- γ and P_j are constants; and
- ϵ is the distribution of the error. It is not necessary to assume this to be normally distributed, but if that assumption is made, then the MLE estimators for the model are the same as the usual least squares estimators.

So it can be observed that this linear regression model is equivalent to the MLE model, but with all the σ terms set to zero, and with the ϵ term made independent of the experiment i.

An interpretation of the σ terms in the MLE model is the degree to which the underlying linear model is consistent with the observations, taking into account the experimental uncertainty. If the model is consistent with the observations (taking into account the experimental uncertainty) we expect small values of the σ terms to be fitted. As inconsistencies are introduced, the MLE fit will need increased values of the σ terms to compensate.

The σ terms can be combined to give a scalar function of the parameters (z_j), which we refer to as the model uncertainty, given by:

$$\sqrt{\sigma_0^2 + \sum_j \sigma_j^2 z_j^2} \tag{7}$$

A small value for the model uncertainty does not imply a small uncertainty on the bias (since there is still the experimental uncertainty to take account of), but it would indicate an underlying consistency between the model and the observations. Larger values may indicate a limitation to the underlying model's ability to represent the observations. The model uncertainty will be included in the results presented later.

Two MLE approaches are considered in this paper:

- Parameter-based MLE, where bias is a function of experimental set-up (e.g. energy of neutrons causing capture). This is essentially the model described above.
- Experiment-based MLE. Typically, experiments can be grouped, where multiple experiments are performed using the same apparatus. It could therefore be assumed that experiments in each group have an associated constant bias, which could occur as a result of correlations between the experiments for example. Some small adjustments need to be made to the model for Parameter-based MLE to achieve this. Details are given when we describe the method in more detail below.

Simple examples are now presented for each approach based on the MONK validation database.

1. Parameter-Based MLE

A. Single Parameter-Based MLE

In this section, a single-parameter MLE is first performed to derive the bias for selected cases in the MONK validation database, based on the log of the mean energy of neutrons causing capture. This is not in itself necessarily a good means of estimating bias, but is used to show how MLE behaves in

different regimes, and serves as a starting point for developing a more appropriate model.

Firstly, a small set of 9 cases, where the energy of neutrons causing capture is very similar is considered (Fig. 1, top). Fits to the bias based on a simple mean, a linear regression analysis and MLE are calculated. Unlike least squares regression, MLE factors in the experimental error. In this case, any potential dependency on the variable, log energy of neutrons causing capture, is masked by the experimental uncertainty. The results are very consistent, with the variation easily explained by experimental uncertainty. In this case the σ terms from the MLE model are effectively zero, which can be interpreted as indicating that there is no need for the model to allow additional uncertainty on the bias in order to explain the experimental results. The lines given by the mean, the linear regression model and MLE are all consistent with the experimental results, with the MLE and linear regression lines being effectively coincident (which is why the MLE line is not easily visible).

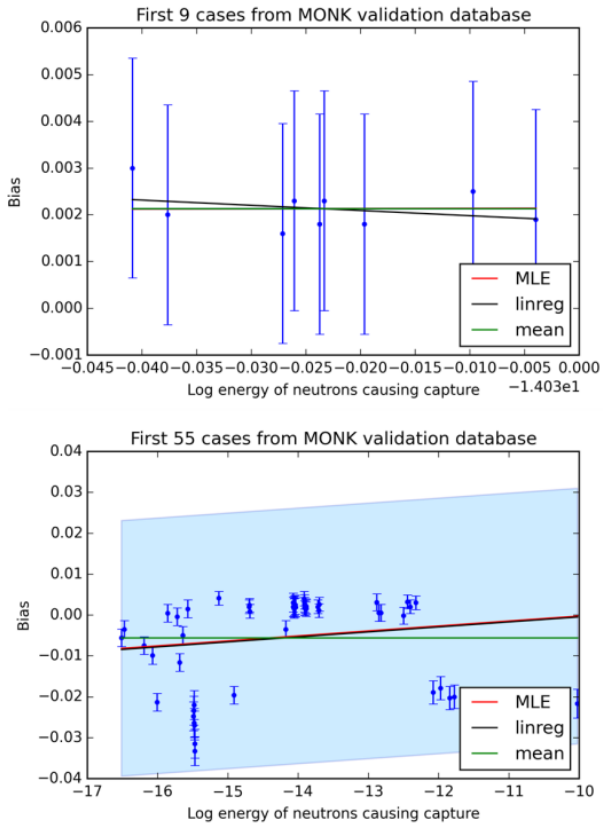

Fig. 1. MLE for tests with 9 (top) and 55 (bottom) cases.

Next, a larger set of 55 cases is considered (Fig. 1 bottom). Here, there is a stronger variation of bias with energy of neutrons causing capture, and the MLE is in good agreement with the linear fit as the spread on the data points is much larger than the individual error.

The figure also includes the model uncertainty at the three standard deviation level (shown in blue highlight in Fig. 1). As noted above, the model uncertainty is derived from the MLE estimators, but without the quoted experimental uncertainty included. Specifically, these highlighted error ranges are calculated as

$$\pm 3\sqrt{\sigma_0^2 + \sigma_1^2 z^2} \tag{8}$$

These error ranges, which were not visible in the MLE for the 9 cases, arise from the need for the MLE model to allow for the variation in the observed results which cannot be explained by the experimental uncertainty alone. The "most likely" explanation, given the constraints of the model is for there to be uncertainty in the coefficients themselves. Here this modelling uncertainty is dominated by the uncertainty associated with the constant term in the model, and this can be observed visually as the spread of the modelling uncertainty appears constant throughout the range.

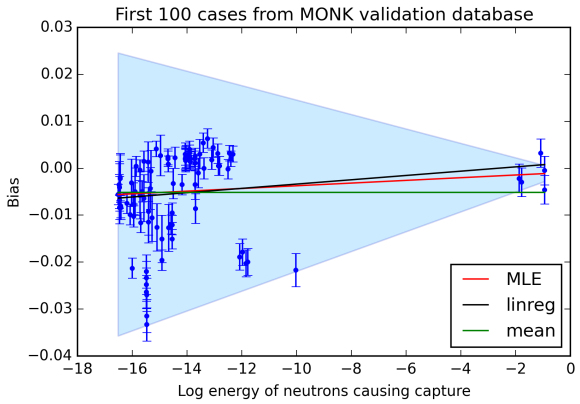

Fig. 2. MLE for tests with 100 cases.

If we extend the number of cases from the MONK validation database to 100, then a different picture emerges, as shown in Fig. 2. Now there appears to be two distinct regions of data as a function of the log energy of neutrons causing capture. The behaviour of the MLE differs from both the mean and linear regression model. The MLE model effectively tries to give the best of both worlds. In the left hand side of the plot, the linear model makes sense but in the right hand side of the plot, the mean makes more sense (based on relative spread of data points and experimental error).

The model as presented here is a little unsatisfactory as it artificially forces a minimum model uncertainty at an arbitrary energy of 1 unit. Invariance under a change in energy units could be recovered by adding an additional term into the model

$$b_i = \gamma + \sum_j P_j(z_{ij} - z_{0j}) + \epsilon_i\,, \tag{9}$$

where

- z_{0j} is a fitted offset to the z coordinate

If this is done, then a slightly modified graph results, but what is still clear is that the model chosen (that is, a model assuming an underlying linear relationship with a single variable) is insufficient to describe the data. We therefore consider a more general model.

B. Multiple Parameter-Based MLE

The extended model now includes three parameters:

- H:Fissile material ratio;
- energy of neutrons causing capture; and
- energy of neutrons causing fission.

Fig. 3 shows the multiple parameter MLE model, applied to the first 27 cases from the MONK validation set in the top figure, and the first 30 cases in the bottom figure.

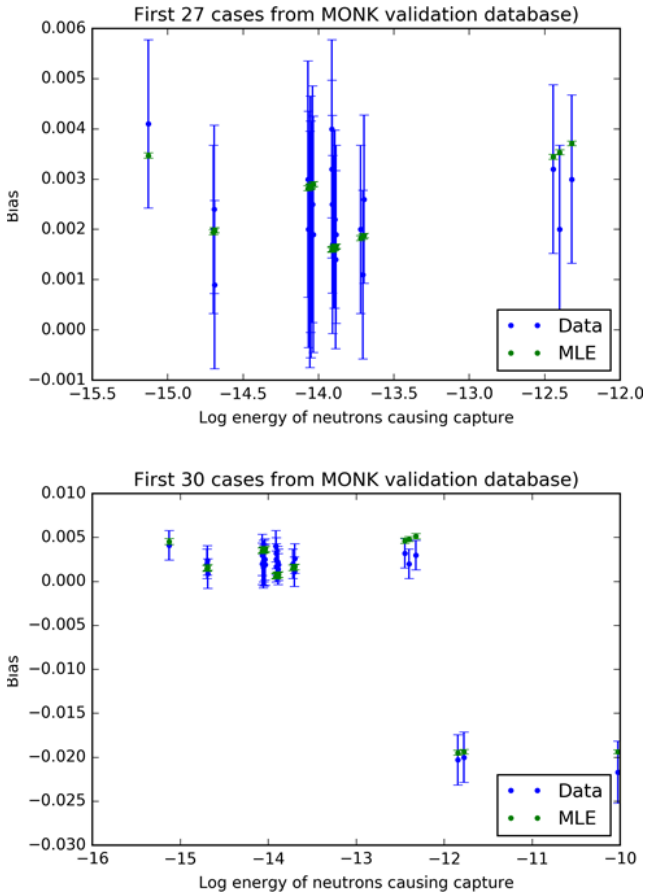

Fig. 3. Multi-parameter MLE compared to data with uranium only cases (top) and additional plutonium cases (bottom).

As the bias is now being cast as a function of multiple parameters, it is no longer possible to simply plot best line fits, and so the presentation of the results differs a little compared to the single parameter cases above. The multiple parameter graphs show the observed expermental result together with the associated uncertainty in blue, and the model predicted bias, together with the model uncertainty (at the one standard deviation level) in green. Once again we emphasize that the model uncertainty does not indicate the overall uncertainty on the bias, rather the additional uncertainty required to give consistentency between the model and the observations.

Going from 27 to 30 cases introduces a plutonium based system, so a small minority of plutonium-based experiments are added to a dataset containing cases which are otherwise uranium based. (Note these plutonium cases were not included in the one parameter model comparisons).

For the first 27 cases the predicted bias and model uncertainty is small. The introduction of the plutonium based systems challenges the MLE model as implemented, since the model does not include any dependency directly based on material composition. As expected, the errors on the predicted biases increase, but the predicted biases for the uranium systems remain sensible. This gives evidence that the approach is robust against occasionally misidentifying cases that the chosen model cannot explain.

Finally in this section, we increase the size of the database again, now to include 60 cases (Fig. 4). The number of uranium cases still outweighs the number of plutonium cases. The predicted error on the bias increases markedly but for many of the uranium systems, the prediction is still very good.

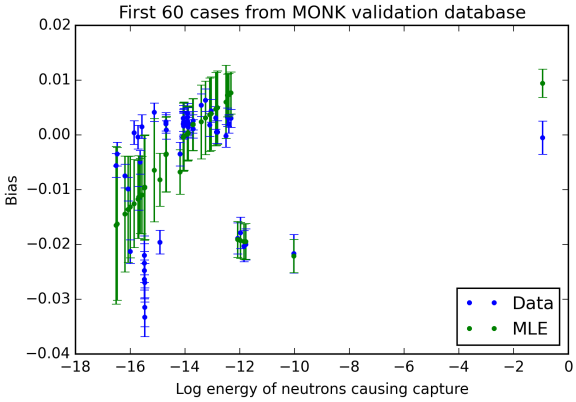

Fig. 4. Multi-parameter MLE for the first 60 MONK validation cases.

We now present a modified MLE approach.

2. Experiment-based MLE

In the Experiment-based MLE approach we model the bias in the form.

$$b_{g,i} = b_0 + b_g + \epsilon_{g,i}\,, \tag{10}$$

where

- g is the experimental group;
- i refers to a specific experiment in an experimental group;
- $b_0 \sim \mathcal{N}(\mu_0, \sigma_0^2)$ is a normally distributed overall bias that is independent of the experimental group;
- $b_g \sim \mathcal{N}(\mu_g, \sigma_g^2)$ are normally distributed biases that are specific to the experimental group to which the experiment belongs; and
- $\epsilon_{g,i} \sim \mathcal{N}(0, \rho_{g,i}^2)$ is the distribution of the error of experiment i in experiment group g.

The parameter estimation for the Experiment-based MLE is similar to that of the Parameter-based MLE. In the Experiment-based MLE the function to be maximized is now:

$$\sum_{g}\sum_{i=1}^{ne_g}\Big(-\frac{1}{2}\log(2\pi s_{g,i}^2)-\frac{(b_{g,i}-m_g)^2}{2s_{g,i}^2}\Big), \qquad (11)$$

where

- ne_g is the number of experiments in experiment group g;
- $m_g = \mu_0 + \mu_g$; and
- $s_{g,i}^2 = \sigma_0^2 + \sigma_i^2 + \rho_{g,i}^2$.

The principal changes to the formulation compared to the Parameter-based approach is that rather than all coefficients potentially featuring for each experiment, we now just have the overall bias and the bias attributed to the experimental group to which an individual experiment belongs. The approach to optimizing the parameters is the same as before. However the approach we take to demonstrating the effectiveness of the method differs. The discussion below tries to capture the idea that as the experimental base is extended we would hope to see a clearer division between the bias attributable to each individual experimental group and that which belongs as an overall bias. We proceed as follows by considering subsets of the experimental groups. We consider n experimental groups, each of which may contain a number of individual experiments. For this set of n experimental groups there are $2^n - 1$ possible subsets (not including the empty set).

The MLE problem is solved for progressively-larger subsets of the overall set of experimental groups, and the bias is derived based on the number of subsets included.

To illustrate this we use a setup of experiments based on 9 experimental groups which here are labelled as cases 1–9 (see Table I). Each group has multiple individual experiments associated with it, ranging from four to twelve. These cases were taken from the MONK validation database, with some artificial anomalies added for testing purposes as described below.

Case	No of expts	Description and additional comments
1	9	2.35 wt% ^{235}U enriched UO_2
2	10	4.738 wt% ^{235}U enriched UO_2
3	8	4.31 wt% ^{235}U enriched UO_2
4	8	2.46 wt% ^{235}U enriched UO_2
5	9	4.46 wt% ^{235}U enriched damp U_3O_8
6	12	4.74 wt% ^{235}U enriched UO_2
7	9	Repeat of case 5 with added bias
8	4	Benchmark LEU-COMP-007
9	4	Benchmark LEU-COMP-039

TABLE I. Cases from the MONK Validation database.

Fig. 5 shows the behaviour of the estimated experimental group bias (that is the μ_g terms) for each case as the size of the subgroup increases. To be precise, the MLE procedure is applied to every subgroup of a particular size that also contains the particular group, and the average of the estimated bias over all these subgroups has been plotted. The subgroup will include all individual experiments corresponding to the experimental group included in the subset.

Additionally an overall bias is plotted. This is the component of the bias that is not attributed (by the model) to a particular group (that is, it is based on the μ_0 term). Again it is actually the average over all subgroups of a particular size.

The figure also shows the standard deviation plotted against subgroup size. This is analogous to the bias plot, but uses the predicted σ_0 and σ_i terms. As expected, the biases appear to be converging on fixed values.

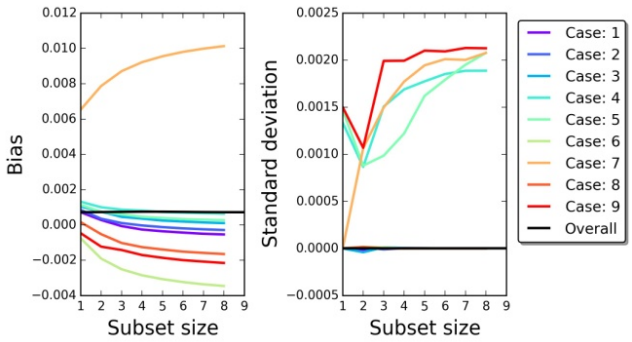

Fig. 5. Biases calculated for subsets of a given fixed size.

For the data in Fig. 5, Cases 5 and 7 are the same, except that Case 7 has an artificially-added bias of 0.01. As apparent from Fig. 5, this bias can be recovered by MLE. These cases also have the same standard deviation.

III. EXTREME VALUE THEORY

When assessing the likelihood of k_{eff} being less than unity, accounting for the uncertainties described above, it is common to run a large number of simulations in order to estimate the probability that the system is critical. Possible approaches to defining this probability are discussed below.

1. Tolerance Limit Based Approach

Without knowing anything about the distribution of k_{eff} it is possible to follow the approach of Wilks [3] for calculating a tolerance range (L_1, L_2) for x from a sample:

For a given method of calculating tolerance limits, how large should our sample be in order that the proportion P of the universe included between L_1 and L_2 has an average value a, and will be such that the probability is at least p that x will lie between two given numbers, say b and c?

The method proceeds by letting L_1 be the greatest of the r smallest values in the sample, and L_2 the smallest of the r largest values, and analysing the distribution of the proportion of the universe represented by the tolerance range.

The approach is quite general and results in distributions from the beta family. That is, distributions of the form

$$f(p) \propto p^{(a-1)}(1-p)^{(b-1)}\,, \qquad (12)$$

where f is the probability density function.

Here a and b are the parameters of the distribution, and completely specify the distribution.

Although not required generally for beta distributions, for this particular application a and b will always be integers, and arise quite naturally from a partition of the observed results.

In [3] the result is derived by considering a multinomial distribution and generates a beta distribution with parameters $a - 1 = n - 2r$ and $b - 1 = 2r - 1$.

Here we only consider the proportion of the universe given by all values up to the upper bound of the tolerance range (i.e. ignoring the lower bound). In this case the beta distribution parameters are $a - 1 = n - r$ and $b - 1 = r - 1$.

In the special case where $r = 1$, the distribution just reverts to the following familiar distribution

$$f(p) = n * p^{(n-1)}, \tag{13}$$

with Cumulative Distribution Function

$$F(p) = p^n. \tag{14}$$

(This form of expression can arise from the probability of success in n successive independent Bernoulli trials, where the probability of success in an individual trial is p, but note here it is being used as a distribution for p, not n.)

This distribution therefore gives the probability that the range up to the maximum observed value covers the proportion p of the universe.

The method makes no assumptions regarding the distribution f, other than mild constraints regarding continuity. The result itself is independent of the distribution f.

Very often this expression is used to determine the sample size required to give a certain probability that a certain proportion of the universe is sampled. For example a probability of 95% that at least 95% of the universe is sampled may be achieved with a sample size of 59, and using the maximum of the sample. This comes from evaluating

$$\int_{0.95}^{1} n * p^{(n-1)} dp = 1 - 0.95^n$$

and finding the value of n so that the integral exceeds 0.95.

If we were to repeat the experiment of sampling 59 values from a distribution a large number of times, then we would expect at least 95% of the observed maxima to be above the 95$^{\text{th}}$ percentile of the distribution, regardless of what the underlying distribution was.

We now turn our attention to a slightly different question:

Based on our observed values of k_{eff}, what is the probability that a given proportion of the universe is less than unity (or some other specified limit)?

Note that the variation in k_{eff} is the variation in calculations resulting from uncertainty in input parameters.

With this question in mind, the Wilks formula where we take the maximum k_{eff}, is only directly applicable if the maximum observed k_{eff} is actually unity.

However if the maximum value is less than unity, then we can still infer that the probability is less than $1 - p^n$ that a given proportion of the universe (represented by p) is less than unity.

If the maximum value is greater than unity then we do not get a bound on the probability based on the maximum value. However it is still possible to take the location of the maximum observed k_{eff} that is less than unity, and use the appropriate beta distribution for this location using the expressions involving r and n above.

Evaluations are shown in Table II for various r and n, where the proportion of the universe, p, has been set to 0.999.

	$r = 1$	$r = 2$	$r = 3$
n=100	0.0952079	0.00463807	0.000150376
n=1000	0.632305	0.264241	0.0802093
n=10000	0.999955	0.999503	0.99724

TABLE II. Probability of 99.9% of the universe being below the top r^{th} value from a sample size of n.

Having described the Wilks approach, we now turn our attention to an alternative approach to finding tolerance limits.

2. Distribution of Maxima

Given that a known distribution of k_{eff} is not guaranteed physically, an alternative is to work with the data in such a way as to guarantee a known distribution mathematically. It is a well known result in Extreme Value Theory (EVT) (see for example [4]) that, if $X_1, X_2, \ldots$ are independent identically-distributed random variables, and $M_n = \max(X_i)$, then if a sequence of pairs of real numbers (a_n, b_n) exists with each a_n greater than zero such that Eq.(15) is true:

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq x\right) = F(x) \tag{15}$$

then $F(x)$ is either a Gumbel, Frechet or Weibull distribution; these are often grouped into the generalized extreme value distribution which has the form:

$$F(x; \mu, \sigma, \xi) = \exp\left(-\left[1 + \xi\frac{x - \mu}{\sigma}\right]^{-\frac{1}{\xi}}\right) \tag{16}$$

where μ, σ and ξ are parameters describing the distribution.

That is, if the distribution of maxima is suitably normalized by a set of (a_n, b_n), and tends towards a limiting distribution, then we know what form that distribution takes.

We can observe that it is necessary to include some form of normalization via the sequence of pairs (a_n, b_n), since without these then M_n will approach a degenerate distribution allowing only one value for the random variable. Loosely this is a theorem applying to maxima that is the counterpart of the central limit theorem for averages, and even in the case of the central limit theorem there is a need for some form of normalization since the variance of the distribution of averages approaches zero as the size of the set of values forming the average increases.

The existence of the sequence of pairs can either be justified or just assumed, but in practice this means that if we can generate a sequence of maxima from the k_{eff} values then we can guarantee that that they will obey the generalized extreme value distribution and we can use MLE, or some other fitting method, in order to solve for the parameters of the distribution.

Two common methods [4] for generating the sequence of maxima are:

- Method of block maxima (MOBM) – the sequence of maxima generated by considering the calculated k_{eff} values block-wise and returning the maximum of each block; and

- Peak Over Threshold (POT) – the sequence of maxima generated by taking all values over a given, user defined threshold. Note that, in this case, the numbers in fact follow a Pareto distribution rather than a generalized extreme value distribution; however, this distinction is not important in the context of this paper.

Note that both of these processes contain a degree of arbitrariness – the first on the choice of block size and the second on the value of the threshold. So, the process of calculating the probability of the system being critical based on the k_{eff} data goes as follows:

1. Calculate a sequence of maxima using the MOBM.

2. Fit this to a generalized extreme value distribution.

3. Calculate a return time for $k_{\text{eff}} = 1$; that is, calculate the expected number of blocks before we should see a value of $k_{\text{eff}} > 1$.

4. This defines an exponential process with a known rate constant. We can then use this exponential process to calculate the probability of getting $k_{\text{eff}} > 1$ within a fixed number of samples (say $n = 1000$).

A simple mathematical example is used here to illustrate the approach, where:

$$k_{\text{eff}} = \frac{1.2Z}{Z + \alpha X + \alpha Y}\,, \tag{17}$$

with α being a user-defined parameter and X, Y, Z being random variables, in this case taken from beta distributions. Although this model is intended for illustrative purposes, we can clearly identify Z as analogous to a fission term, the factor 1.2 as the average number of fission children, and the X and Y terms as representing absorption. It is also clear that as α increases (analogous to increasing the amount of absorption), the probability of the event that $k_{\text{eff}} > 1$ will occur decreases.

In the following, a sample of size, n_{samp}, was generated for this problem, and used to calculate the probability that there is a 0.1% chance of getting $k_{\text{eff}} > 1$, based on calculations using EVT with MOBM. The value of α is then varied, with the same set of sampled values for X, Y, and Z used for each value of α.

For clarity, we go through the EVT approach here in sufficient detail as to allow the reader to repeat the method. Here we used the the *Mathematica* [5] package to provide some of the numerics – alternative numerics packages could have been used.

The steps are as follows:

1. take n_{samp} random values for each of X, Y, Z. Here we used distributions $X \sim \beta(2,4)$, $Y \sim \beta(9,2)$ and $Z \sim \beta(4,4)$;

2. calculate the corresponding n_{samp} values of k_{eff} by substituting into
$$k_{\text{eff}} = \frac{1.2Z}{Z + \alpha X + \alpha Y}$$
for a specific value of α;

3. partition the k_{eff} values into bins of size 20, and take the maximum of each (so, for example $n_{samp} = 1000$ would give 50 values, $n_{samp} = 10000$ would give 500 values);

4. fit the generalized extreme value distribution to the set of maximum values. For this work we have used the `EstimatedDistribution` function from the *Mathematica* package, using the default method for parameter estimation, which is in fact to maximize the log-likelihood function;

5. set $\lambda = -\log(F(1))$ where F is the Cumulative Distribution Function for the fitted generalized extreme value distribution; and

6. finally evaluate $1 - G(50)$ where G is the Cumulative Distribution Function for the exponential distribution with parameter λ (or equivalently, evaluate $\exp(50\lambda)$), to give the probability of obtaining a consecutive sequence of 1000 ($= 50 \times 20$) k_{eff} values with $k_{\text{eff}} < 1$.

From the simulation, the probability that there is a 0.1% chance that $k_{\text{eff}} > 1$, using n_{samp}=1000, is shown in Fig. 6.

As α increases, the EVT shows an increasing probability that $k_{\text{eff}} > 1$.

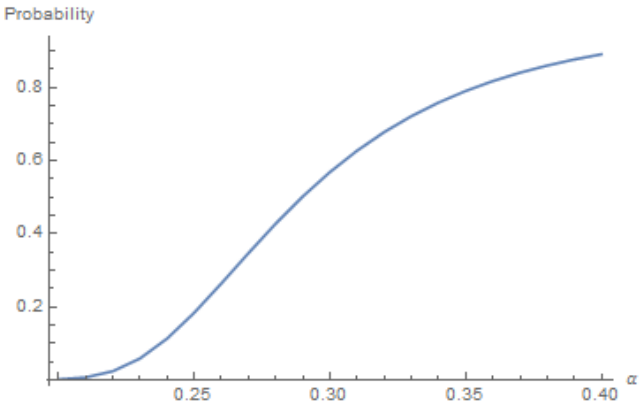

Fig. 6. Probability of there being a 0.1% chance of getting $k_{\text{eff}} > 1$ as calculated by EVT – each run using the same set of random numbers.

Using 1000 sampled values (for n_{samp}) gives a manageable number of evaluations to test the implementation, but it is apparent by, for example, rerunning cases with different random number seeds, that 50 evaluations of the maxima is insufficient to fit the extreme value distribution in this instance. This can be seen in Fig. 7, where different random number seeds are chosen to start the process.

This leads us to investigate how large a sample size is needed in order to give a stable representation of the extreme value distribution. It was observed that once the number of

maxima reached about 1500 (rather than the 50 used above), the calculation stabilized, as can be seen in Fig. 8 and Fig. 9 which use 500 and 1500 maxima respectively. 1500 maxima correspond to a value of n_{samp} of 30000.

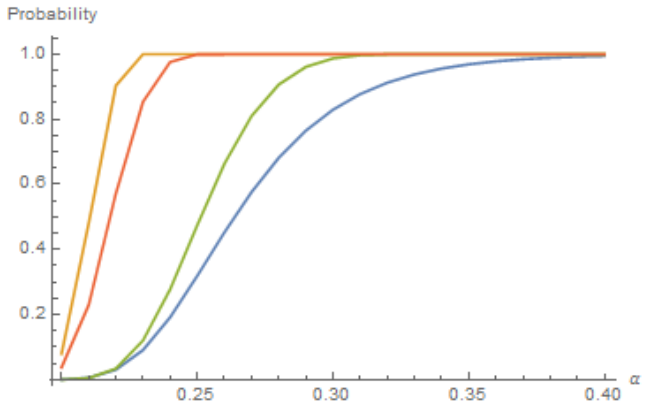

Fig. 7. Probability of there being a 0.1% chance of getting k_{eff}> 1 as calculated by EVT – each case using different random number seeds.

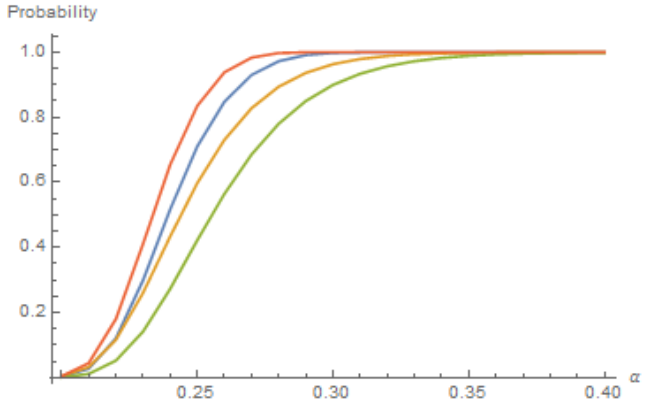

Fig. 8. Probability of there being a 0.1% chance of getting k_{eff}> 1 as calculated by EVT – each case using different random number seeds, with 500 maxima.

3. Discussion

The EVT approach appears to offer a meaningful way to assign a probability to the event that there is a (specified) low probability of observing $k_{\text{eff}} > 1$. However there are still questions and caveats as detailed below:

- What method should be used to generate the list of maxima: MOBM, POT or some other method?
- For whichever method is chosen, there are arbitrary parameters (the number of bins, the height of the threshold etc.) –what is the correct/optimal choice of these parameters? How does the choice of these parameters affect the calculated probability?
- The example has used an off-the-shelf fitting routine. Are there alternative approaches that might be more suitable?

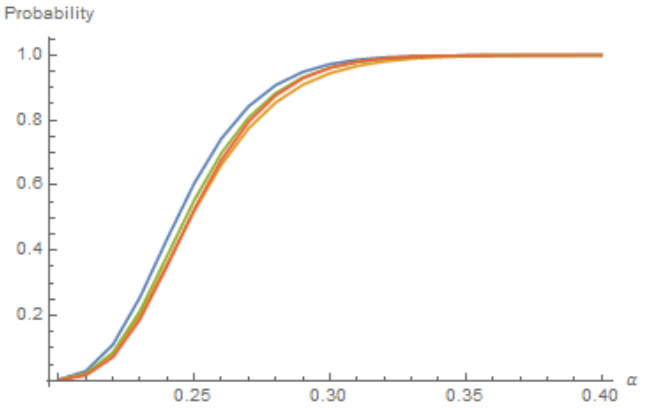

Fig. 9. Probability of there being a 0.1% chance of getting k_{eff}> 1 as calculated by EVT – each case using different random number seeds, with 1500 maxima.

The above list relates to potential improvements in the EVT implementation. We also note another potential issue where care is needed:

- Is it a valid assumption to consider the k_{eff} values as independent, identically-distributed variables or a time series? In particular, if there is a structure to the variation of the input parameters to the calculations then this will lead to a non-random structure in the calculated k_{eff} values.

The probabilities generated by the EVT method as presented are different from those given by the Wilks method. In the EVT method they are the probability of obtaining (in this case) 1000 successive values of $k_{\text{eff}} < 1$, or more generally $(1-q)^{-1}$ successive values, where q is the fraction of the universe required to be less than unity. In the case of Wilks, the probability supplied is the probability that the range contains a fraction q of the universe. This means the two measures are not directly comparable. Further work is needed to investigate the relative merits of the approaches, particularly in the light of the potential improvements that could be made in the EVT approach from the above.

To round off this discussion, we apply the Wilks approach to the simple mathematical example given by Eq. (17). Fig. 10 shows the probabilities calculated by the Wilks method that in 99.9% of the universe k_{eff} is less than unity (that is the probability that 0.1% of the universe has k_{eff} greater than unity), and plots these probabilities against α as in the previous figures. It is calculated once for a sample of 1000 k_{eff} values (as calculated using the test model), and once using 30000 k_{eff} values. The number of occurrences of $k_{\text{eff}} > 1$ is counted, and this is used to select the appropriate beta distribution.

The plots can be considered as lower bounds since the probabilities are evaluated for the highest observed value of k_{eff} that is less than unity. This can be illustrated by the right hand side of the figure in the 1000 evaluations case. Although the probability stays well below 1, this is because there are no observations of k_{eff} exceeding unity at the right hand side of the figure.

As the number of observations increases (or evaluations in the case of our simple model), there is a much more defined

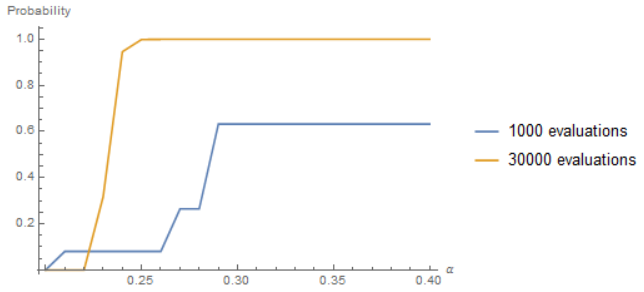

Fig. 10. Probability of there being a 0.1% chance of getting $k_{\text{eff}} > 1$ as calculated by the Wilks approach.

range of α corresponding to the transition where 99.9% of the universe is below unity.

IV. CONCLUSIONS

This paper outlines the mathematical basis for two methods to address uncertainty quantification for Monte Carlo radiation transport, with particular application to criticality analysis.

The Maximum Likelihood Estimator (MLE) has been applied to the problem of bias estimation based on a validation database. Using this approach, which does not require considerable computational expense, bias can be estimated based on pre-existing data and experiment selection for bias estimation based on engineering judgement can be validated. MLE is shown to be robust against occasionally misidentifying cases that the chosen model cannot explain.

Two approaches to the model parameterization have been explored: one where the bias is modelled as varying continuously with some physical parameter, the other where the bias is considered to vary in a discrete way, depending for example on aspects of the experimental facility. This provides a framework for a more formal, rigorous calculation of bias estimation for criticality analysis, in particular for the MONK Monte Carlo code. further work includes analysing the methods in more detail, and testing against a wider range of data in the MONK validation database.

An approach to calculating the probability that k_{eff} is greater than 1, or some subcritical limit, has been identified based on EVT, with clear relevance to the field of criticality analysis. This method has the potential to provide much higher levels of certainty on reported results than other methods and has been demonstrated for an example mathematical problem.

REFERENCES

1. S. RICHARDS, M. SHEPHERD, A. BIRD, D. LONG, C. MURPHY, and T. FRY, "Recent developments to MONK for criticality safety and burnup credit applications," in "International Conference on Mathematics & Computational Methods Applied to Nuclear Science & Engineering," Jeju, Korea (April 2017).
2. D. ROCHMAN, A. KONING, S. VAN DER MARCK, A. HOGENBIRK, and D. VAN VEEN, "Nuclear Data Uncertainty Propagation: Total Monte Carlo vs. Covariances," *Journal of the Korean Physical Society*, **59**, 1236–1241 (2011).
3. S. WILKS, "Determination of sample sizes for setting tolerance limits." *Ann. Maths. Stats*, **12**, *1*, 91–96 (1941).
4. M. GILLI and K. EVIS, "An Application of Extreme Value Theory for Measuring Financial Risk," *Computational Economics*, **27**, *1*, 1–23 (2006).
5. *Mathematica 11.0*, Wolfram Research, Inc. (2016).