Development and Performance Testing of a Nuclear Material Testbed Digital Twin

T. Stockman[1*], A. Lintereur[1**], R. Lakis[1***]

[1]Los Alamos National Laboratory, Los Alamos, NM, 87545, USA

[*]tstockman@lanl.gov, [**]alintereur@lanl.gov, [***]rlakis@lanl.gov

Abstract

The Dynamic Material Control program at Los Alamos National Laboratory was established to develop advanced nuclear material monitoring techniques for the LANL Plutonium Facility. As part of this effort, a nuclear material testbed was built to be a representative platform of a nuclear processing facility. The testbed consists of four mock gloveboxes, each instrumented with four ^{3}He neutron detectors. In addition to being a platform for testing measurement capabilities, the testbed is also being used to establish advanced data analysis methods, including a digital twin. The digital twin is trained on a set of measurements taken with a source of known strength in prescribed positions within the testbed gloveboxes. With these measurements, models are constructed which enable real-time continuous unattended testbed monitoring. A single-source template matching scheme is implemented to search the database of prescribed measurements for patterns similar to in situ measured count rates, and an iterative deconvolution algorithm enables multiple sources to be simultaneously identified and localized as they move throughout the testbed.

In this work, we detail the benchmarking of the algorithms in the digital twin. We examine the pros and cons of experiment-trained data driven digital twin modeling applied to a non-idealized environment. We compare the performance of our data driven models to other methods, such as reduced order analytical modeling and high-fidelity Monte Carlo physics modeling. The performance analysis includes quantitative metrics like computational speed and accuracy, as well as qualitative metrics like ease of implementation and resilience to change.

Introduction

A strong, complex, and dynamic background can present a significant challenge for accurate assays of nuclear material in an operating nuclear facility; however, the need to move material to a shielded measurement location introduces extra cost, contamination risk, and dose to workers. The Dynamic Material Control (DYMAC) collaboration at Los Alamos National Laboratory (LANL) is working to address this challenge via several different methods. One particular effort, which is currently housed in the DYMAC Nuclear Material Testbed (Testbed), is exploring the application of a large array of detectors spread across a room. While this system has low absolute efficiency, it can be used by sophisticated software to determine the real time background in the local area. With the real time background known, it can be dynamically subtracted so that high fidelity measurements can be taken even while strong sources are present and dynamically moving.

Source localization is referred to as an "inverse" modeling problem, and performing this type of solving in real time based on data driven models from a physical system is often referred to as a "digital twin". The problem is fairly straightforward for a single source, but deconvoluting the background that multiple sources create in a complex environment is difficult. In solving this inverse problem, an iterative scheme of guessing source locations and then simulating them to successively improve guesses is implemented. As such, it is vital to have a fast and accurate forward model which can predict detector count rates based on given source strengths and locations. Typically, a tool like MCNP [Werner 2017] would be an excellent candidate for this task, but the detailed radiation transport MCNP performs comes at the cost of relatively high computational expense. In this work, we explore data driven forward modeling methods which are less accurate than high fidelity models like MCNP, but are able to provide solutions much quicker, which enables them to be used in a real time iterative scheme for background deconvolution.

These forward models generally fall into two categories: purely data driven models (DDM) and reduced order models (ROM), though at points there is overlap. For the purpose of this work, DDM will refer to a model which does not simulate any physics and purely relies on an experimental database to infer new simulation results. By contrast an ROM will account for some physics with low fidelity and use an experimental database to calibrate various constants in those physics models. In this work, two DDM models and one ROM model are developed and compared in terms of speed, accuracy, and application.

Methodology

1. Testbed

The Testbed is a 720 cm x 720 cm room where radioactive sources can be introduced and manipulated and various kinds of hardware can be installed and reconfigured. For the work presented here, the Testbed is configured with four aluminum framed mock gloveboxes labeled A, B, C, and D as illustrated in Figure 1. Gloveboxes A and B are “half size” with a footprint of 76 cm x 244 cm, and Gloveboxes C and D are “full size” with a footprint of 156 cm x 244 cm. Also in the room is a housing for data acquisition electronics, a wooden workbench with a PC, and a large screen where real time information about the Testbed is displayed.

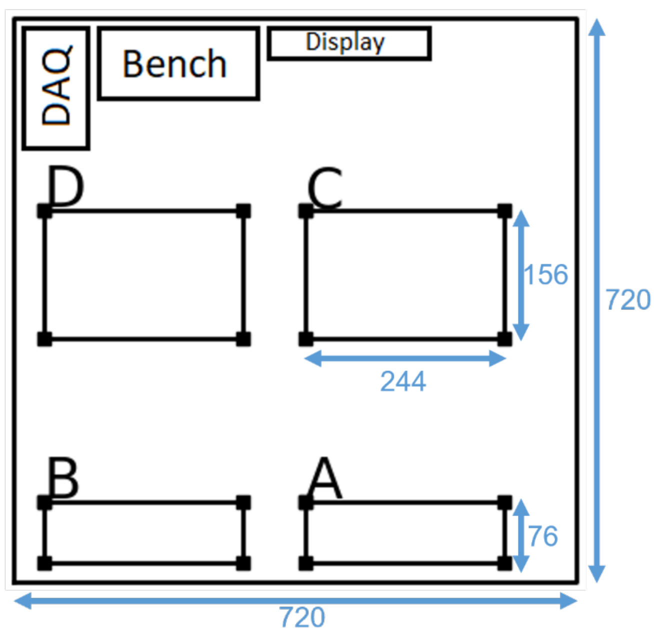

Figure 1. Testbed floorplan illustrating four gloveboxes labeled A, B, C, and D. Squares on the corners of the gloveboxes represent the locations of He-3 neutron detectors.

Each glovebox has a He-3 neutron detector mounted to each corner, for 16 total detectors in the room. Each detector consists of a 2.54 cm diameter, 50.8 cm long tube filled with He-3 and CF_4, pressurized to 4 atm. The tube is centered in a square cross-section block of HDPE, wrapped in a 0.04 cm thick sheet of cadmium. Aluminum housing provides structural support to mount the detectors to the gloveboxes.

The signal from each of the 16 detectors is fed to an Advanced List Mode Module (ALMM) which collects the data in list mode, meaning that every neutron detected in the testbed is recorded with an associated timestamp [Newell 2017]. The primary advantage of list mode data collection is that it allows for post processing analysis of neutron coincidences, but in this work we will not be using that feature. Instead, a program called Apricot takes the list mode data and updates a CSV file which provides an average count rate for each detector in 1 second bins.

2. Sources

Cf-252 sealed sources were used for all of the measurements; data was collected over the span of several months, which significantly changed the source strength due to the short 2.6 year half-life of Cf-252. For this reason, the neutron emission rate of each source will be described with its associated measurement.

3. Detector Benchmarking

A set of measurements was performed with each detector in a fixed location relative to a source to confirm that every detector had similar efficiency. The average count rates from these measurements varied less than 1% from maximum to minimum.

4. Template Measurements

A set of 144 measurements, referred to as Template Measurements, were performed in the testbed which span the floor space of each glovebox in a roughly uniform grid, illustrated in Figure 2. These measurements were performed with a single Cf-252 source with a neutron emission rate of 7.31×10^5 n/s which was iteratively moved to each location on the grid and measured for 15 minutes.

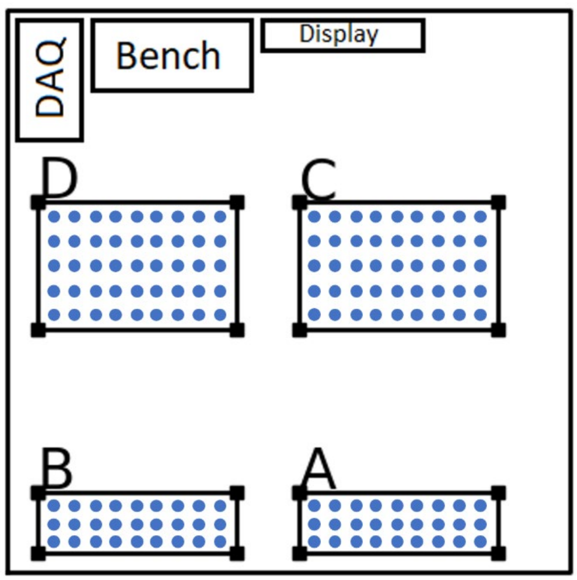

Figure 2. Blue dots indicate the location of each of the Template Measurements.

These measurements serve several purposes. First, they provide some characterization of the physical environment. Features like wall thickness and composition play a significant role in the reflection and moderation of neutrons in the Testbed, so some detectors will appear to have higher efficiency than others. When the detectors are mounted in their respective locations we note a difference in response of up to 20% when the same source is placed the same distance from each detector.

The second purpose of these measurements is to serve as a database for DDM and ROM which will be described in more detail in other sections.

5. Performance Characterization Measurements

A set of measurements, referred to as the Performance Characterization Measurements, were performed in the testbed at pseudo-random locations. The purpose of these measurements was to explore a variety of source configurations within the Testbed to characterize the performance of various models. To serve this purpose, it was decided that five measurements with four sources in the Testbed simultaneously, twelve measurements with two sources in the Testbed simultaneously, and sixteen measurements with a single source in the Testbed would be performed. To mitigate bias from the people selecting sources and locations, random number generation in Microsoft Excel was used to select locations within the gloveboxes and the sources to be used in each measurement.

The 252-Cf sources used in this set of experiments will be referred to as W, X, Y, and Z and had neutron emission rates of 5.96×10^5, 7.87×10^4, 7.09×10^4, and 5.17×10^4 n/s respectively. These sources were chosen to span a dynamic range of at least 10 (W and Z) and to have at two sources which were similar though not identical in strength (X and Y).

6. Template DDM

The first DDM constructed from these measurements is essentially a lookup table. To refine the resolution of this model, a 5 cm x 5 cm grid of artificial measurements was generated by interpolating between the Template Measurements. To construct the model, the average count rates from each of the interpolated Template Measurements was normalized to the respective measured source strength, creating a table of absolute detector efficiencies with associated point source locations within the Testbed.

With this table, a user can specify a source strength (n/s) and position within the gloveboxes in the Testbed, and the model can identify the nearest location in the table and multiply the associated row of absolute efficiencies for each of the 16 detectors by the user-specified source strength to get a predicted count rate in each detector from that source and position.

Further, superposition can be used to model multiple sources simultaneously. Provided the detectors are not being significantly affected by dead time, the counts in any given detector from multiple sources can simply be modeled as the sum of the counts expected from each source individually. So if the user specifies multiple sources, the model will predict count rates for each source individually and then sum them to model the simultaneous response.

7. Linterp DDM

The next DDM created, referred to as the Linterp DDM, is similar to the Template DDM in that it exclusively uses the Template Measurements as a database to predict count rates in the detectors, but instead of shifting the user-specified location to an exact location which exists in its database, Linterp DDM interpolates the predicted detector efficiencies linearly from the nearest measurement locations in the Template Measurements. Just like the Template DDM, this model uses superposition to model count rates from multiple sources as the sum of the count rates from each source individually.

The motivation for this model is that it would trade some speed for accuracy. Performing an interpolation for every source was expected to be more computationally burdensome than simply looking up a location in a table, but it would also allow the model to simulate locations at a finer resolution than the prescribed mesh of measurements used for the Template DDM.

8. Curve Fit ROM

Where the Template DDM and Linterp DDM are purely data driven models, the Curve Fit ROM assumes some simple physics to model detector count rates beyond the constraints of the Template Measurements. The physics model assumed is similar to the inverse square law, which would indicate that the count rate in each detector should decrease as the square of the distance between the detector and the source. However, the tall, thin geometry of our detectors deviates from this model significantly at the distances encountered within the Testbed.

Thus, to create this model, a curve fit was implemented which takes the form:

$$C = \frac{A}{d^B}$$

Where C is the count rate in a detector (n/s), d is the distance from the source to the detector, and A and B are fitting parameters which were solved to be 17,000 and 0.96, respectively. The data for this curve fit was assembled from the Template Measurements, creating a data point from each of the 16 count rates and detector distances in each of the 144 measurements for a total of 2304 data points. The curve fit and data are illustrated in Figure 3. Note that the counting statistics error on the data points shown in Figure 3 is smaller than the size of the dots which represent the measurements. Data points which deviate significantly from the trend are a result of environmental factors and illustrate the impact those factors can have on count rate.

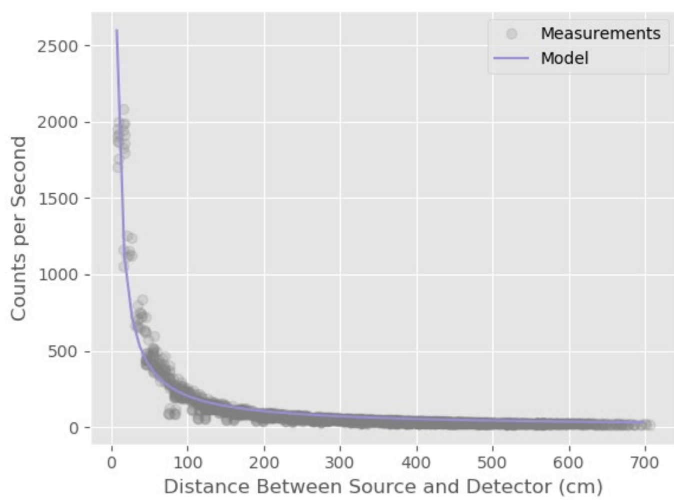

Figure 3. Curve Fit ROM illustrated on top of the data it was fit on.

It was noted previously that the environment has a significant effect on count rates. For example, detectors near walls tend to have higher count rates than detectors near the center of the room. This model ignores those environmental factors by aggregating every measurement together. This means it will be less accurate in predicting the count rates for any specific detector, but it gains the flexibility to predict the count rates from a detector in any arbitrary location. This is a significant advantage over the previously mentioned DDMs which can only model count rates for the detectors in their respective locations at the time the Template Measurements were taken.

9. PyMAC Digital Twin

Where the previously mentioned models have been "forward" models, predicting the detectors' count rates in response to sources placed in prescribed locations, the PyMAC Digital Twin (hence referred to simply as PyMAC) is an "inverse" model which takes measured detector count rates as an input and predicts the corresponding configuration of source positions and strengths. Previous work developing this model can be found here [Stockman 2021], and another inverse solver of this nature here [Ren 2021].

This inverse solve is performed by discretizing the Testbed into zones, each of which has associated detectors and a single point source modeled inside its boundaries. If the predicted neutron emission rate of the source in a given zone is below a prescribed threshold or negative, it is rounded to 0 strength, and assumed to be an artifact of the solving methodology. The location and dimensions of these zones can be arbitrarily selected, but for this implementation the borders of each glovebox are chosen to be the boundaries of each zone.

PyMAC first examines the count rates of each detector in each zone. These count rates are normalized by the maximum count rate among them and then compared to similarly normalized count rates within the Template Measurements. A source is then assumed to exist in the location corresponding to the Template Measurement with count rates closest to what was measured. The strength of the source is determined by averaging the ratio of the measured count rates to the Template Measurement count rates and associated Template source strength. We refer to this algorithm as Template Matching.

If we assumed only a single source existed in the testbed at any time, this methodology alone might be sufficient. However, multiple sources create background from one zone to another such that the source emission rate predicted from Template Matching will be too high because the count rates in a given zone are higher than would be expected from the single source. Thus, the predicted position of the source will be biased toward the direction of the background. For this reason, a scheme is implemented to iteratively model the background from the source in each zone and make increasingly accurate predictions by subtracting the modeled background each time a new prediction is made. The iterative algorithm works as follows:

1. In each zone, Template Matching uses the local zone count rates to predict a source location and strength in that zone assuming no other sources are present in the Testbed.
2. A new prediction is made in each zone, but this time the sources predicted in the previous step are assumed to exist. When a particular zone is being considered, the background from each external source is modeled with the previously described Template DDM. The background from each outside source is subtracted from the detectors associated with the zone of interest, and a new Template Matching prediction is made with the background-subtracted counts.
3. The previous step is iterated until a solution converges to within a prescribed tolerance.

Options for modifying the convergence rate have been considered but not yet thoroughly explored. For example, instead of fully subtracting the background predicted from each outside source, a damping factor could be applied which only subtracts a prescribed percentage of the predicted background. This could help mitigate the "bouncing" behavior induced by a poor initial source prediction. Future work may consider this approach, but the Template Matching and Template DDM algorithms used in this scheme are sufficiently fast that performance increases would be imperceptible on the current time scale.

Results and Discussion

This section will describe the performance of the models in both speed and accuracy. To assess the speed of the three forward models (Template, Linterp, and Curve Fit) they were run through 100,000 simulations on a typical laptop and timed. To assess performance, all three models were used to simulate the source strengths and positions in the Performance Characterization measurement set. Model error will be described according to the equation:

$$Error = \frac{modeled - measured}{measured}$$

where "modeled" is the simulated count rate of each individual detector for a given scenario, and "measured" is the measured count rate of the same detector in the corresponding measurement configuration. With 16 detectors per measurement and 33 total measurements in the Performance Characterization set, the errors for each model are an average of 528 total comparisons.

1. Forward Models

The average absolute error of the Template, Linterp, and Curve Fit models was 18.4%, 11.7%, and 21.7% respectively. The distribution of these errors is illustrated in Figure 4, but note there are some outliers of as high as 300% error which are not plotted in the histograms. The most accurate of these three methods was Linterp which was the intended performance; recall that the development goal for Linterp was to trade some speed for more accuracy than the Template DDM. Template's accuracy was second best, though it is worth noting this has a direct function with the mesh spacing in the template database it reads from. In this case, the measurements used were on the order of 20 cm apart with an interpolated mesh resolution of 5 cm. A finer mesh of measurements and a finer mesh of interpolated points could increase accuracy at the cost of some speed and experiment time. The least accurate of all three models was the Curve Fit method, which is due to the fact that it is a very rough approximation of detector efficiency averaged over all 16 detectors. By contrast, the other two methods are able to model every detector individually.

Figure 4. Error for the three forward model methodologies compared to the Performance Characterization measurements.

The average runtimes for all three methods were 4.39 µs, 258 µs, and 21.8 µs respectively as illustrated in Figure 5. Template was the fastest as it is performing a very simple lookup operation. Curve fit was the next fastest, because it is computing a simple analytical equation which is only slightly more computationally burdensome. Linterp was the slowest by a wide margin, though at only 258 µs it is certainly still fast enough for use in a real time algorithm.

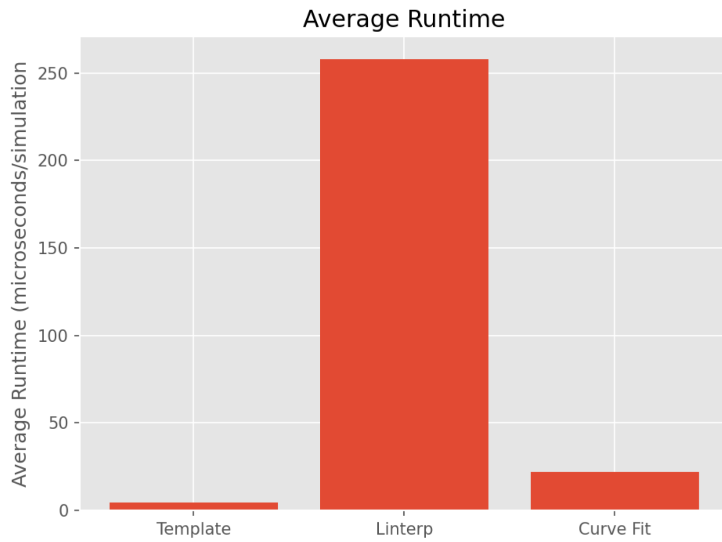

Figure 5. Average runtime for the three forward model methodologies.

An important note for the Curve Fit model is that it was not designed specifically to predict count rates in these detectors for given sources. Instead, it was designed as a visualization model optimized almost entirely for speed and versatility so that it could be run thousands of times per second to create contour plots for visual display in real time. An example of this visualization is shown in Figure 6. The other two models are unable to predict counts for any location except the exact locations of the 16 detectors as they were placed in their training set, so Curve Fit is ideally suited to give a fast, though slightly less accurate, approximation of the real time radiation background in the room.

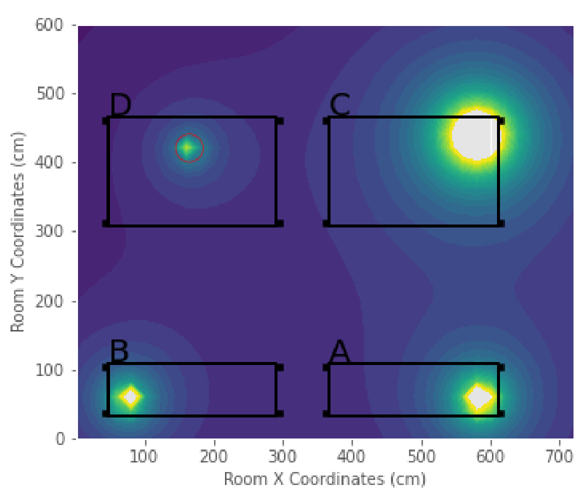

Figure 6. Visualization of the radiation background produced by four sources in the Testbed.

2. Digital Twin

The PyMAC Digital Twin was similarly benchmarked for speed in several trials and for accuracy compared to the Performance Characterization measurements. After 1,000 localization runs, PyMAC's average runtime was 0.091 s. Figure 7 illustrates PyMAC's accuracy in terms of error on source strength (n/s) and source position with colors indicating the approximate strength of the sources and symbols representing how many simultaneous sources were being localized. The average error over these trials was 5.55% on source strength and 8.0 cm on source position.

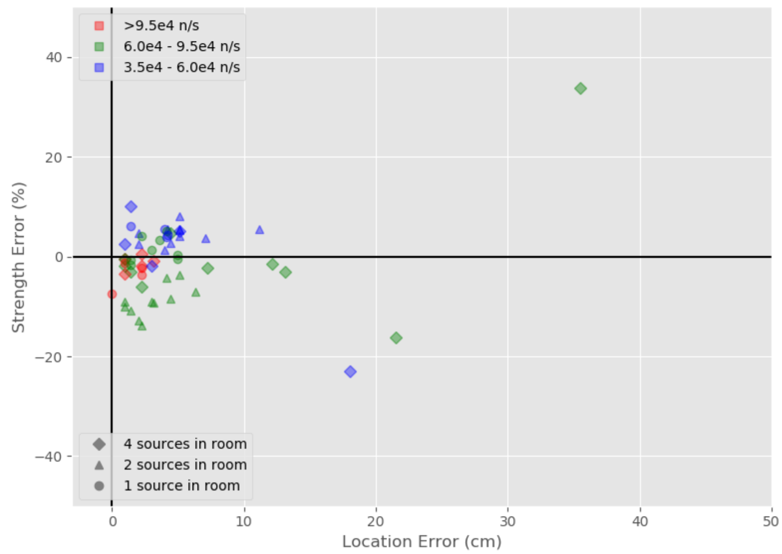

Figure 7. PyMAC's error by source strength and number of simultaneous sources.

Future work will more optimize the performance of these algorithms and test them in a wider variety of circumstances. Each model shows some outliers which significantly sway the averaged errors reported in this work and which may hold clues to regimes of poor algorithm performance. Overall, the PyMAC Digital Twin is performing well deconvoluting the signal from multiple sources in real time while determining the background each one produces in the room.

Acknowledgements

This work was funded by the U.S. National Nuclear Security Agency, NA-191, under the Dynamic Material Control (DYMAC) collaboration.

References

Werner, C. J. (editor), "MCNP Users Manual - Code Version 6.2", Los Alamos National Laboratory, report LA-UR-17-29981 (2017).

Newell, M., Rothrock, R., and Henzlova, D., “Demonstration of the Advanced List Mode Module”, *INMM Annual Meeting Conference Proceeding,* Indian Wells CA, July 2017. LA-UR-17-25577 (2017).

Stockman, T., et. Al., “Facility scale in-situ source localization and assay via a sparse 3He neutron detector array: enhancing nuclear material control and accounting in nuclear fuel cycle facilities”, *INMM Annual Meeting Conference Proceeding*, Virtual, August 2021. LA-UR-21-28490 (2021).

Ren, C., et. Al., “Source term estimation via combined sparse convex optimization and maximum likelihood estimation for nuclear material accounting”, *INMM Annual Meeting Conference Proceeding*, Virtual, August 2021. LA-UR-21-21634 (2021).