Proceedings of the 2015 Industrial and Systems Engineering Research Conference
S. Cetinkaya and J. K. Ryan, eds.

Highly Efficient Evaluation Design (HEED) for Comparing Algorithms Used to Detect Nuclear Materials

Abstract ID: 1750

Paul Kantor, Christie Nelson, Fred Roberts, William M. Pottenger
CCICADA, Rutgers University
Piscataway, New Jersey

Abstract

Radiological or nuclear materials are detected in vehicles by processing data from sensor systems. The key goals/objectives are to detect threats and control the number of "nuisance" alarms caused by non-threat materials. It is important to evaluate algorithms efficiently and fairly. Experiments can be done *in silico* or using real loads of cargo containing hidden radiation sources which is time-consuming and costly. In either mode, sensor data serve as inputs to the algorithms, whose outputs can be compared to ground truth. The goal is not to map the complete characteristics of an algorithm, but to compare algorithms. Therefore it is most efficient to concentrate on experimental configurations that reveal meaningful differences between the algorithms. The methods of Combinatorial Experimental Design are used to generate configurations that will find all situations in which one or two levels of key parameters reveal such a difference. In practice, an efficient set must be further reviewed by subject matter experts, who are tasked to specify configurations that would reveal meaningful differences among the algorithms. Experts may also assign importance values to the remaining configurations, based on other considerations related to the likelihood or consequences of the corresponding threat.

Keywords
Experimental design, nuclear detection, combinatorial experimental design

1. Introduction
We have developed and piloted an innovative extension of the concept of Combinatorial Experimental Design (CED) to the problem of scheduling efficient tests that will be adequate to support the Domestic Nuclear Detection Office (DNDO) in its evaluation of competing nuclear detection algorithms. These designs must cover all of the important levels or values of key variables, and must allow for the possibility of interaction effects, in which specific choices of two variables together cause an algorithm to exhibit inadequate performance, although neither of the specified values of the individual variables will reveal such an inadequacy. The research included expert elicitation of a vast array of significant variables, discussions with subject matter experts to reduce these to families; development of paper-and-pencil forms for elicitation towards experimental design, and pilot implementation of those methods in the context of algorithm selection.

The Domestic Nuclear Detection Office (DNDO) is a jointly staffed agency within the Department of Homeland Security and is the primary entity in the U.S. government for implementing domestic nuclear detection efforts. One way nuclear detection is used is in scanning vehicles, which can require coordination of sensor systems (physical hardware) and data processing algorithms. Primary scanning of cargo trucks can be done using Radiation Portal Monitors. Nuclear Detection algorithms must provide high screening performance, by detecting almost all threats, while holding down the number of false alarms. False alarms generate double costs – the cost of added inspection as well as costs imposed on national commerce by the delay of harmless shipments. It is important to detect Special Nuclear Materials (HEU; Plutonium; etc.) and materials that could be used in a Radiation Dispersal Device (RDD) or "dirty bomb."

Performance can be measured as a detection rate d. This rate depends on many factors, including the type of SNM, the mass (by type), shielding, masking, and background, and is a characteristic of the algorithm. A Nuisance Alarm

Rate *f* is the conditional probability that Naturally Occurring Radioactive Material (NORM) or legitimate medical isotopes will cause an alarm.

Two types of testing are injection testing and operational testing. Injection testing is done with computer experiments. These use synthesized signals created by combining the spectrum of a threat object with non-threat data taken from the stream of commerce and other scenarios. These depend on Monte Carlo simulations. However injection testing is much cheaper than operational testing. Operational testing can include field experiments, such as testing a limited number of scenarios. However, it is difficult to replicate some factors at a given site (e.g. high humidity in a desert)**.**

With injection testing, data from the sensors is presented as input to the algorithms, and outputs from the algorithms are then compared to the ground truth. The goal is to evaluate algorithms in a fair way. Secondary inspection (with handheld device) adds considerations of usability of the scanning device, which will not be considered here.

To design the right set of experiments for comparing algorithms we need to form discrete representations of several concepts. These are specified in terms of the measures of performance (MOP) and the factors that may affect it. (In other settings, MOPs can be called Measures of Effectiveness or simply “dependent variables.” The factors may be referred to as independent variables in other settings. Some factors are under experimental control. Others, while they may be important, must simply be measured as they occur naturally. We concentrate here on those that can be controlled in both simulation experiments and field experiments.) Examples of important factors are the type of SNM present in a cargo, and the size of the sample. Examples of MOPs are the nuisance rate and the detection rate.

We use the term configuration to refer to any specific set of values of the several factors. As an example, one factor may refer to the type of SNM, and take values which are the names of distinct isotopes. Another may refer to the mass of the sample, and may take values such as “small,” “medium,” or “large.” It is to be understood that these represent different amounts of mass, according to which SNM is considered. However, it is expected that subject matter experts (SMEs) will be able to define those levels or values in such a way that they have similar practical meanings both for security and for experimental design.

It is important to note that it may not be possible to have similar practical meaning both from the perspective of security, and of experimental design. In such a case two distinct factors or variables would be needed to describe the mass of material. For example, with regard to security an expert might be asked to specify “low” as the smallest quantity of a given SNM that could constitute an effective threat, taken alone. Similarly “high” might be elicited as the largest amount that an opponent might plausibly try to sneak past detection. If needed, “medium” could be elicited as some intermediate value, such as the arithmetic or geometric mean. On the other hand, if the goal of the elicitation is to discriminate among algorithms, there will likely be some levels which one would not expect any algorithm to resolve, and others which one would expect every algorithm to resolve. These two values define the range within which the elicited levels, low, (medium), and high should lie. In other words, in designing experiments we must also focus on configurations that will help us to distinguish among algorithms.

To complete the entire process of experimental design we need two more properties of the configuration. These are the “importance” of the configuration and the “cost of assessing the MOP” in the configuration. These will enter into an optimization calculation, if it is not possible to consider all of the configurations and repeat measurements in each one, enough times to distinguish among algorithms. To apply Combinatorial Experimental Design, we must map the MOPs into binary values. We examine this next.

2. Background – Combinatorial Experimental Design

This project used the process of Combinatorial Experimental Design (CED). The central idea behind combinatorial design testing is the application of experimental design to test generation as described in [1,2]. Each distinct element of a test input tuple is treated as a factor, and each distinct value of a factor is treated as a level. For example, a set of inputs that has 10 parameters with 2 possible values each would use a design appropriate for 10 factors with 2 levels for each factor. If the design covers two-way interactions, it ensures that every value (level) of every parameter (factor) is tested at least once with every other level of every other factor. The CED approach to pairwise coverage provides a very substantial reduction in the number of test cases when compared with traditional orthogonal designs.

The traditional approach of applying pairwise coverage is to use 2-way Orthogonal Array Designs. These are special cases of Resolution III designs. A design of resolution R is a design in which no p-factor effect is confounded with any other effect containing less than R-p factors. In a resolution III design main effects are not confounded with each other, but are confounded with pairwise interactions, and vice versa.

However, the traditional Orthogonal Arrays do not always exist, and in any case, they scale quadratically with the total number of levels of the factors. In contrast, the Combinatorial Designs introduced in [1,2] create pairwise designs that are much more parsimonious. For example, suppose there are ten factors of importance with, respectively, (3, 10, 3, 2, 3, 2, 2, 3, 3, 3) levels. Applying traditional factorial design techniques, to evaluate all the pairwise interactions among them requires a total of 495 different experimental configurations. The corresponding Combinatorial Design requires only 32 experimental configurations, a more than ten-fold reduction in experimental complexity. The reduction is possible because these designs do not determine the influence of each factor or pair of factors, but only indicate whether they have generated a significant effect. These designs can also be extended to higher *t*-way (e.g., 3-way) coverage of all factors/levels, should that be needed.

Crucial to the effectiveness of such CED is the fact that the outcome of the test must be a binary result. In the traditional application of CED, this result is either PASS (meaning that the software being tested functioned as it should) or FAIL (meaning that something did not work properly, and further investigation is required.

Test Result | Statistics

	FEATURE1	FEATURE2	FEATURE3	FEATURE4	FEATURE5	FEATURE6	FEATURE7	FEATURE8	FEATURE9	FEATURE10
1	A	1	50	false	2	false	false	2	2	2
2	B	1	165	true	3	true	true	3	3	3
3	E	1	5	false	1	true	false	1	1	1
4	A	2	165	true	1	false	true	2	3	1
5	B	2	5	true	2	false	false	3	1	2
6	E	2	50	true	3	false	true	1	2	3
7	A	3	5	false	3	true	true	2	1	3
8	B	3	50	false	1	true	false	3	2	1
9	E	3	165	false	2	true	true	1	3	2
10	A	4	5	true	3	false	false	3	2	1
11	B	4	50	false	1	true	true	1	3	2
12	E	4	165	false	2	false	false	2	1	3
13	A	5	5	true	3	false	false	1	3	2
14	B	5	50	false	1	true	true	2	1	3
15	E	5	165	true	2	false	false	3	2	1
16	A	6	5	true	3	false	false	1	3	2
17	B	6	50	false	1	true	true	2	1	3
18	E	6	165	true	2	false	true	3	2	1
19	A	7	5	true	3	false	false	1	3	2
20	B	7	50	false	1	true	true	2	1	3
21	E	7	165	false	2	false	true	3	2	1
22	A	8	5	true	3	false	false	1	3	2
23	B	8	50	false	1	true	true	2	1	3
24	E	8	165	true	2	true	false	3	2	1
25	A	9	5	true	3	false	false	1	3	2
26	B	9	50	false	1	true	true	2	1	3
27	E	9	165	false	2	false	false	3	2	1
28	A	10	5	true	3	false	false	1	3	2
29	B	10	50	false	1	true	true	2	1	3
30	E	10	165	true	2	false	true	3	2	1
31	B	3	5	true	3	false	true	3	3	2
32	B	2	5	false	3	true	true	3	3	3

Figure 1: CED example for 10 factors

The Combinatorial Experimental Design for our earlier example with 10 factors, with the levels specified in the text, is shown in Figure 1. The reader may verify that every pair of levels of any pair of factors appears in at least one of the configurations. Therefore, if any pairwise interaction should cause an algorithm to fail, that failure will be exposed in at least one of the configurations. The levels of the several factors are notional here, and do not correspond to specifics of the situation for detection algorithms.

Figure 1 shows the results obtained using the ACTS software to develop a very efficient design [3]. ACTS, which is freely available from NIST, is considerably less powerful than non-freeware such as the AETG software [4].

3. The Problem

For possible nuclear detection injection testing, data from the sensors is presented as input to the nuclear detection algorithms, and outputs from these algorithms are then compared to the ground truth. Two possible goals are acceptance of the technology and comparison of technology. For acceptance of the technology, the result of the test on each configuration is either PASS or FAIL, and an aggregate of those findings is used to accept or reject the technology. For comparison of technology, the result of a test on any pair of technologies is [No Meaningful Difference] or [One is Better], and an aggregate of these results is used to compare the technologies. Since the goal of the experiments is to evaluate or rank the algorithms and devices, the research presented here takes an end-to-end view of these experiments.

4. Elicitation with Subject Matter Experts

4.1 Pass Fail Levels

In order to apply CED for each measure of performance and each configuration we want to establish PASS and FAIL levels for each specific measure of performance (MOP). A PASS level means if the algorithm is performing this well, there is no urgent need to improve it. The level is adequate, given the proposed configuration. Another way of saying this is that if each of two algorithms is above this PASS level, then their performance on this MOP is not a reason to prefer one over the other. (Of course, there is always some bias in favor of the better performance. But it would be like asking for better "zero to sixty" on a car that is only used to drive over to the grocery store.) A FAIL level can be taken to mean if the algorithm does not exceed this level, then it cannot be considered acceptable for this configuration. Another way of saying it is if each of two algorithms is below this FAIL level, then their performance on this MOP is not a reason to prefer one over the other. (Of course, there is always some bias in favor of the less terrible performance. But that would be like asking whether one would prefer to have very bad sunburn or very bad poison ivy.)

4.2 Probe Values of the Factors

For each factor that has a meaningful influence on the MOP under consideration, the CED becomes most powerful if we can select a relatively small set of values at which to examine the algorithm. We call these "probe values." This will make the experimental design efficient, while retaining its effectiveness. The key idea is sketched in Figure 2.

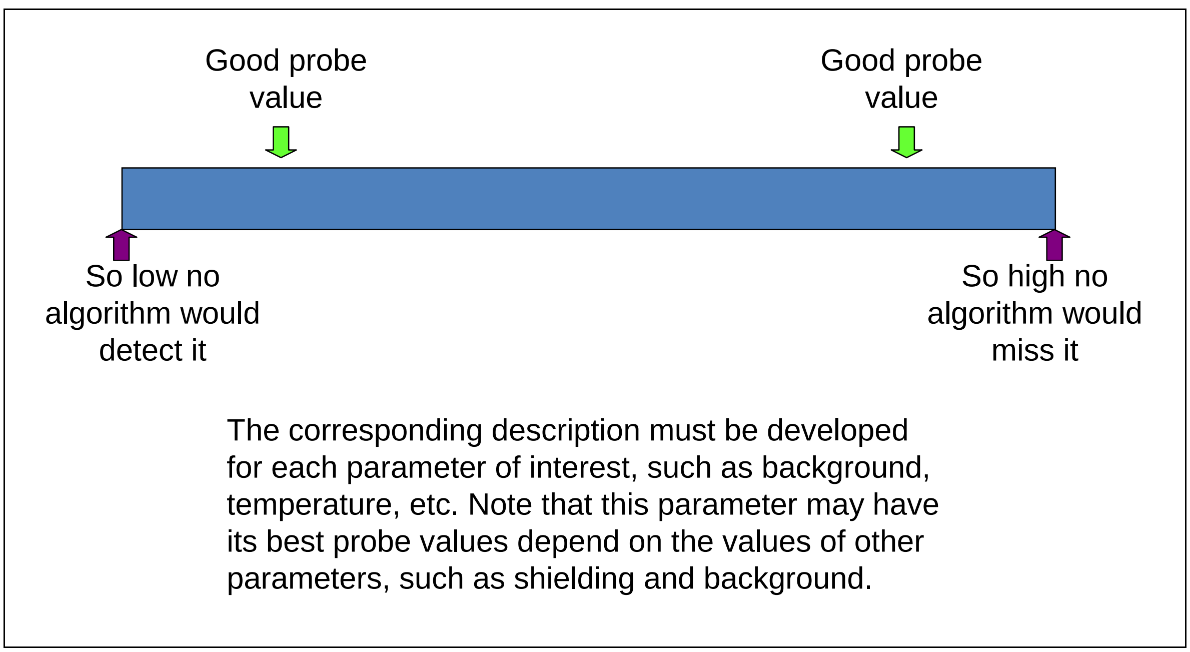

Figure 2: Good probe values for the strength of the source

Probe values should be near the ends of the "range of discrimination." Depending on the measure of effectiveness and the configuration, this might be defined (as shown in Figure 2) in terms of what is technically feasible. For other configurations it might be defined in terms of the reality of the threat, or of the magnitude of the consequences.

A good choice for a HIGH PROBE level of a factor value will be close to the edge of meaningful difference, perhaps 10% or 20% short of it. That is a judgment for the subject matter expert to make. Logically, there is no point in being at the extreme, since the results will be the same for all the algorithms, and a test would not be worth doing.

A good choice for the LOW PROBE level of a factor value will also be close to the edge of meaningful difference, perhaps 10% or 20% short of it. That is again a judgment for the subject matter expert to make. Again, there is no point in being at the extreme, since the results will be the same for all the algorithms, and a test would not be worth doing.

4.3 Elicitation Process

The search for the experimental design configurations involves three distinct analyses. First, any naturally continuous variables (such as the ROC, or the mass) must be made binary or categorical. Second, the methods of Combinatorial Experimental Design are used to cover all possible levels and pairs of levels of all the variables deemed to be important. The CED using ACTS [3] provides a table of configurations. Third, the table must then be reviewed by subject matter experts, tasked to assign a value and a cost to each of the configurations. When the cost data are available, the final stage of the analysis can be formulated as a set selection problem. The overall goal is to optimize the combined value of the selected subset of configurations, with the budget as a constraint. The technical formulation is a Binary Programming problem, which is NP complete.

Subject matter experts are needed for the elicitation process. First, with regard to any possible criterion for performance (e.g., detection rate; nuisance alarm rate; ease of use; tunability; robustness), we must work with SMEs to determine the PASS | FAIL level. Second, with regard to any of the variables that might affect performance (e.g., type of special nuclear materials; type of background; the weather; etc.), we must work with SMEs to determine which levels of the variable will make an effective probe of the system.

The elicitation of probe values and the PASS and FAIL levels is an iterative process. The logic is shown in Figure 3, and must be repeated for each factor and for each MOP.

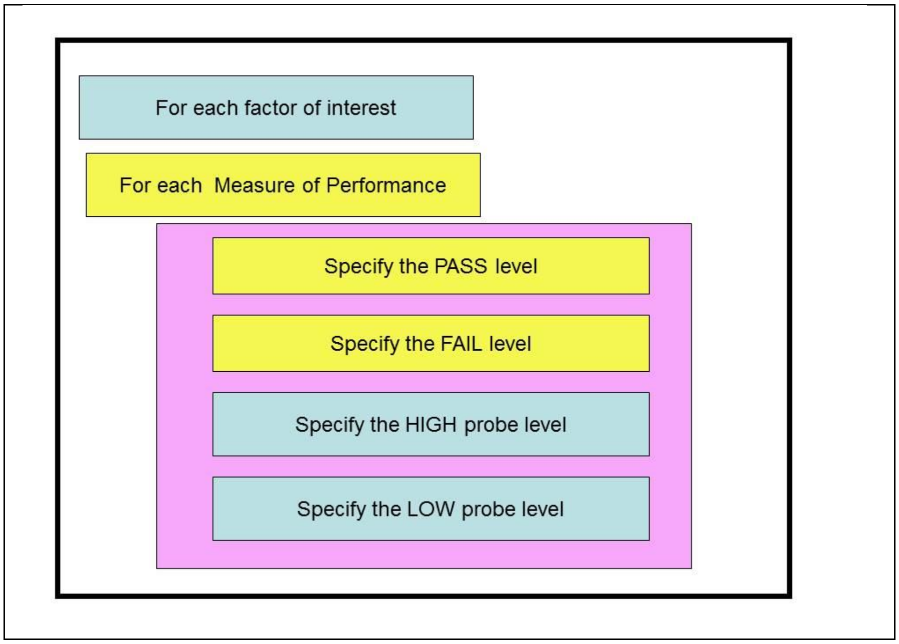

Figure 3: Elicitation plan for PASS|FAIL values of MOPs and HIGH|LOW values of continuous factors

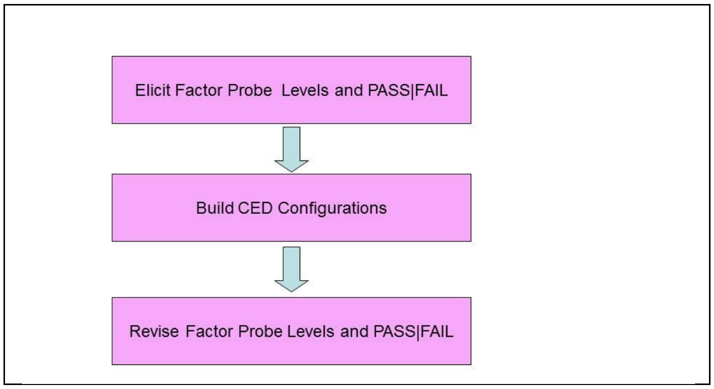

Figure 4: The iteration that adjusts factor levels and PASS|FAIL to reflect the specific configurations in which they are being assessed.

Factors:		
Special	[1, 2, 3, 4, 5]	
Normal	[N1, N2, N3]	
V1	[HI, LO]	
V2	[HI, LO]	
V3	[HI, LO]	

Test Case#	Special	Shield	V1	V2	V3	MOP-1		MOP-2	
						PASS	FAIL	PASS	FAIL
0	1	S1	LO________	LO________	LO________				
1	1	S2	HI________	HI________	HI________				
2	1	S3	LO________	HI________	LO________				
3	2	S1	HI________	LO________	HI________				
4	2	S2	LO________	HI________	LO________				
5	2	S3	HI________	LO________	HI________				
6	3	S1	LO________	HI________	HI________				
7	3	S2	HI________	LO________	LO________				
8	3	S3	LO________	LO________	HI________				
9	4	S1	LO________	HI________	HI________				
10	4	S2	HI________	LO________	LO________				
11	4	S3	LO________	LO________	HI________				
12	5	S1	LO________	HI________	HI________				
13	5	S2	HI________	LO________	LO________				
14	5	S3	LO________	HI________	LO________				

Figure 5: Illustration of an interactive layout to permit adjustments to a combinatorial design configuration

When the process described above has been completed, a suitable CED tool, such as AETG [4], or the ACTS [3] software can be used (during a short rest period for the SMEs) to generate an efficient set of configurations. We anticipate that there will be a few tens of such configurations. These are then presented to the SMEs, who are asked to refer to their previous estimates of what represent suitable values for HIGH|LOW and PASS|FAIL in these configurations. If they agree that no change is needed, the process is complete.

If they feel, given the possible interactions among the factors, that adjustments are needed, this can be done by any customary process of revision and agreement. Since the specific values are likely to be sensitive data, the elicitors may not be able to support this renegotiation. This process is shown in Figure 4.

To facilitate this process the output of the CED process can be modified to provide space to enter the specific numeric probe values of the factors and the corresponding PASS|FAIL levels for the MOPs. A notional example is shown in Figure 5.

Figure 5 shows an illustration of a combinatorial design with space for the SMEs to insert their own individual estimates and eventual consensus as to the specific numerical meaning of LO and HI or PASS and FAIL for the specified configuration. In this example we suppose that there are 5 varieties of SNM and 3 varieties of shield. Note that the logically minimal number of configurations (3x5=15) to cover all pairwise combinations of these two factors is also adequate to cover all pairwise combinations of the additional 3 variables called V1, V2, V3 with each other, and with the two leading variables.

5. Findings

5.1 The Elicitation Process with Subject Matter Experts

Pilot elicitations were done with three subject matter experts (SMEs) over a two day span. Together, the three SMEs well represented both the pilot concern (algorithms), and the future work of this project, a second phase, in which hand-held devices are to be assessed.

The elicitation process, which was done with each SME individually, had four steps:

(1) Elicit the names and descriptions of those factors that are important in that they may help to distinguish among algorithms all receiving a data stream from the same portal device. This produced a large list. The SME was then guided to score them, assigning 12 to the most important, with 5 meaning “sort of” important (e.g., it is on the list for completeness sake, but need not be included in the design). Most SMEs were not comfortable giving any score lower than 5.

(2) For each factor with a score greater than 5, SMEs were asked to generate representative values or “levels” for each factor. If that data (e.g., mass of a specific SNM) was sensitive information, the SME assigned a useful name such as “HI” or “MED”, and indicated how many possible levels they would use, but not the physical meaning of the level. They were permitted to choose as many or few values as they’d like, and which were deemed the most important. They were encouraged to narrow the focus so that the number of configurations did not explode.

(3) The factors and their levels were encoded into the ACTS software [3], producing a combinatorial experimental design. The SMEs, seeing the size of the design, would then work with us to narrow down the number of test configurations. Generally, they preferred to keep all of the factors, but narrow the number of levels of some factors. They were also able to identify constraints on the factors and levels, which further reduced the design size.

(4) The last step of the process, for which we did not have time during these pilot two-hour elicitation sessions, is to ask the SME to examine the table of configurations, and assess the importance (again with 12 being most important, 1 being least/not important).

5.2 Findings

The nature of the elicitation process is better understood by examining the summaries of three elicitations. Note that the factors were developed first, then the importance ratings, and finally the levels. In these elicitations, no SME gave a value lower than 6. Sample illustrative findings are shown in Tables 1-3. Factors and levels are lettered and numbered due to sensitive material. (Note, Factor A1 may or may not correspond to Factor B1, etc. They are simply numbered in value order by SME.)

Table 1. Illustrative Sample SME Elicitation

Value	**Factor**	**Levels**
13	A1	11
12	B1	4
10	C1	4
9	D1	2
8	E1	2
7	F1	2
6	G1	2
6	H1	2

Table 2. Illustrative Sample SME Elicitation

Value	**Factor**	**Levels**
12	A2	3
12	B2	12
11	C2	5
11	D2	13
10	E2	2
10	F2	2
9	G2	Initially 9, later narrowed to 3
9	H2	Initially 8, later narrowed to 2
6	I2	11

Table 3. Illustrative Sample SME Elicitation

Value	**Factor**	**Levels**
12	A3	15
12	B3	4
12	C3	4
8	D3	2

The SME corresponding to Table 1also discussed 10 additional possible factors that did not go into the CED. The SME corresponding to Table 2 discussed four additional possible factors, and the SME corresponding to Table 3 discussed eight additional possible factors that were not included in the CED.

On the basis of these three sessions we are confident that the elicitation process is challenging, but entirely feasible. One SME noted that the detailed scoring of the configurations would require access to additional data/information. Another proposed that a single hand-out which shows the flow of the elicitation process would help the SME to be oriented to this unfamiliar process. The most challenging part of the process seemed to be to determine what the factors actually are.

5.3 Recommendations

After performing a pilot set of elicitations, we have several recommendations. First, we recommend that the scoring of configurations should be done by asking the SME to "take home" the design and go through it, with possibly a second brief discussion to finalize the scoring. We also have noted a tendency to include every possible level of a factor, when perhaps one or two would suffice. Some SMEs find it challenging to "break apart" the problem and think about the factors separately rather than in physically natural combinations (e.g., SNM "A" shielded by shielding material "X" or masked by suitable masking material "Z"). As examples, one SME generated 44 configurations, another 132 configurations after having the list shortened from the original 170, and a third generated 36 configurations after having the list shortened from 84. Some of the SMEs were able to shorten their configurations by either adding constraints among the factors or reducing the number of levels for some of the factors. One alternative approach might invite the SME to propose specific combinations of factors. Those combinations could then be extended to a full design covering all meaningful pairwise interactions, using the CED software. It also might be valuable to conduct a group elicitation session to define the list of factors (without ranking them by importance), and then conduct individual elicitations with the SMEs on the same list of factors for ranking them by importance. In addition, now that the process has been piloted, it will be very valuable to build a software tool (in a spreadsheet) that would facilitate the process of naming the factors and defining the levels. Such a tool could invoke the ACTS software, facilitating a variant elicitation in which three factors are carried through to the generation level, permitting the SME to see where the process is headed.

6. Discussion and Conclusion

6.1 Security Considerations

During the elicitation process, the elicitors may not be allowed to know sensitive information, such as the specific numerical values of some of the factors, or the desired levels of performance. It is possible to have an iterative process in which several SMEs work together with the elicitation experts to discuss the problem and then the elicitors leave the room while the SMEs agree on a consensus value for the item under discussion.

6.2 Next Steps and Recommendations

We have developed and piloted an innovative extension of the concept of Combinatorial Experimental Design (CED) to the problem of scheduling efficient tests. These designs must cover all of the important levels or values of key variables, and must allow for the possibility of interaction effects, in which specific choices of two variables together cause an algorithm to exhibit inadequate performance, although neither of the specified values of the individual variables will reveal such an inadequacy. The research included expert elicitation of a vast array of significant variables to reduce these to families, development of paper-and-pencil forms for elicitation towards experimental design, and pilot implementation of those methods in the context of algorithm selection.

The concepts needed for the design (probe values, adequacy levels, and constraints on variables) were well understood by Subject Matter Experts, and three pilot sessions were informative and effective. The complete experimental design process will include a subsequent scoring of the experimental configurations by SMEs, which will provide a weighting or scaling that can be combined with an analysis of the importance of factors individually.

The method here, (AETG) [4], was developed for finding faults in software. Here we presented an extension to the assessment of technologies (algorithms or devices) for detecting and identifying radio-isotopes, but other applications are possible. This project has broader impacts as well. One example might be identifying which combination of algorithmic features leads to meaningful improvement in a recommender system. Another possibility is determining which combination of medications is most effective, etc. We propose that with suitable adaptation this approach, which eschews determination of specific effect sizes, provides a path to High Efficiency Experimental Design (HEED) for searching large spaces of design possibilities.

Acknowledgements

We would like to thank Dr. Jay Spingarn of DHS/DNDO who identified the problem, supported the research, and suggested the need for probe values, Dr. Detlof von Winterfeldt and colleagues at the CREATE Center, and Dr. Gariann Gelston and colleagues at PNNL. We thank Dr. Siddartha Dalal and Dr. Ashish Jain for introducing us to the notions of combinatorial experimental design. We also acknowledge support from DNDO Contract # HSHQDC-13-X-00069, DHS Contract #DHS-2009-ST-061-CCI002-06, and NSF Grant # 1247696 and 1142251.

References

1. Cohen, D., M., Dalal, S., Fredman M., and Patton, G.C. (1997). The AETG system: An approach to testing based on combinatorial design. IEEE Transactions on Software Engineering,23(7):437-444.
2. Dalal, S., and C. L. Mallows (1998). Factor-covering designs for testing software. Technometrics,40(3):234- 243.
3. ACTS Software. http://csrc.nist.gov/groups/SNS/acts/documents/comparison-report.html
4. Automated Efficient Test Generation System (AETG). User manual. http://aetgweb.argreenhouse.com/doc/
5. Dalal, S., and B. Han (2010) Detection of Radioative Material Entering Natrioianl Ports: A Bayesian Approach to Radition Portal Data. Ann. Applied Statistics, 1256-71
6. Dalal, S., A. Jain, N. Karunanithi, J. M. Leaton, and C. M. Lott (1998). Model-based testing of a highly programmable system. In F. Belli, editor, Proceedings of the Ninth International Symposium on Software Reliability Engineering, pages 174-178.
7. Dalal, S., and C. L. Mallows (1998). Factor-covering designs for testing software. Technometrics,40(3):234- 243.
8. Ely, J., Kouzes, R., Geelhood, B., Schweppe, J. and Warner, R. (2004). Discrimination of naturally occurring radioactive material in plastic scintillator material. IEEE Transactions on Nuclear Science 51 1672–1676.
9. Ely, J., Kouzes, R., Schweppe,. Siciliano, E., Strachan, D. and Weier, D. (2006). The use of energy windowing to discriminate SNM from NORM in radiation portal monitors. Nuclear Instrument and Methods in Physics Research Section A 560 373–387.
10. Hastie, T., Tibshirani, R., Friedman, J. (2008), The Elements of Statistical Learning, Springer Verlag, New York.
11. Lo Presti, C., D. Weier, KOUZES, R., SCHWEPPE, J.,(2006), Baseline suppression of vehicle portal monitor gamma count profiles: A characterization study, Nuclear Instruments and Methods in Physics Research A 562 (2006) 281–297
12. Lott, C., A. Jain and S. Dalal (2005). Modeling Requirements for Combinatorial Software Testing, ACM A-MOST Workshop, pp.
13. Knoll, G. (2010), Radiation Detection and Measurement, John Wiley and Sons, New York.
14. National Research Council. (2010) Evaluating Testing, Costs, and Benefits of Advanced Spectroscopic Portals for Screening Cargo at Ports of Entry: Interim Report (Abbreviated Version). Washington, DC: The National Academies Press.
15. Kantor, P., Dalal, S., Jain, A., Nelson, C., Roberts, F.. "Optimal Selection of Configurations to Test Radiation Detectors." INFORMS Annual Meeting. Nov 2014. San Francisco, CA.
16. Nelson, C., Kantor, P., Pottenger, W. M., Roberts, F. "Evaluating Algorithms for Processing Nuclear Detection Sensor Data." INFORMS Annual Meeting. Nov 2014. San Francisco, CA.