19th International Conference on Modelling and Applied Simulation
17th International Multidisciplinary Modeling & Simulation Multiconference

ISSN 2724-0037 ISBN 978-88-85741-48-5 © 2020 The Authors.
DOI: 10.46354/i3m.2020.mas.005

Effective Use of Verification and Validation Resources

Steven Darcy[1,*], Daniel Maxwell [2] and Andrew Loerch[1]

[1] George Mason University, 4400 University Drive MSN 4A6, Fairfax, VA, 22030, USA
[2]KaDSci LLC, 031 University Dr., Suite 100, Fairfax, VA, 22030, USA

[*]Corresponding author. Email address: sdarcy1@gmu.edu

Abstract

A model has force if it can positively and constructively influence decision makers. To a large extent, the force a model has on a decision maker is determined by the confidence they have in the model. Ideally that confidence is a consequence of the model performing well in a high-quality verification and validation process. This paper presents a generalization of the Line of Sight Evidential Reasoning and Analysis methodology. This methodology is used to construct a Bayesian network which codifies our belief about how specific verification and validation activities change a model's force on a decision maker. Using the Bayesian network we rank verification and validation activities based on how much impact they have on the model's force. We then show how the ranking of verification and validation activities can change based on the intended use of the model. In other words, the Generalized Line of Sight Evidential Reasoning and Analysis methodology provides a quantitative basis for allocating V&V resources and this allocation may be different for different decision makers or model uses.

Keywords: Validation of OR Computations, Decision analysis, Finance, OR in banking, Quality Control, Simulation

1. Introduction

Data can be thought of as low-level individual observations and measurements. Related pieces of data can be organized to produce information. Information can be analyzed, understood, and explained to produce knowledge (Waltz 2003). There is a corollary in evidential reasoning. Evidence based reasoning has been defined as gathering evidence and using it to draw conclusions about matters of interest (Schum). Let's say we want to solve a crime. From and evidential perspective we would start by gathering some evidence and then drawing some intermediate conclusions. From an information perspective we can think of this as gathering data and producing information. Based on the intermediate conclusions we may be compelled to repeat this process and continue in an iterative fashion until we solve the crime. From an information perspective this is gathering additional data and developing new information until we have the knowledge we seek. We can think of evidence is data; intermediate conclusions are information; and, solving the crime is knowledge; and I will use these terms interchangeably throughout this paper.

In the above example we start with evidence and reach a conclusion about a matter of interest. The data leads us to knowledge. But can knowledge lead us to data? For example, I want to know whether or not a computer model is suitable for use in my study. Under these circumstances we would start by examining existing evidence such as the model documentation and documents describing how the model has been validated. After this examination we might conclude that given our study the model has not been sufficiently tested. In cases such as this we would want to perform additional tests; and, the tests we perform should help us make our determination. Stated more generally, given a matter of interest and the conclusion we wish to draw it is natural to ask the question what evidence should I gather?

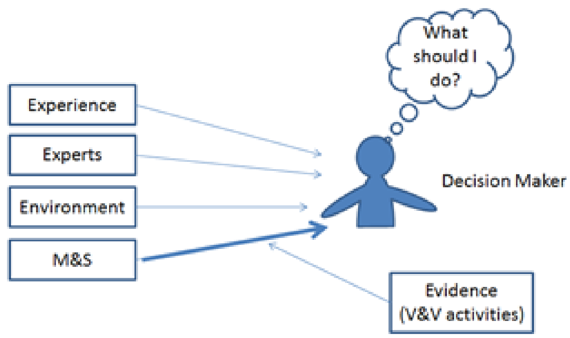

Figure 1. Using Models to Inform a Decision

A model has force if it can positively and constructively influence decision makers (DMs). To a large extent, the force a model has on a DM is determined by the confidence they have in the model. Ideally that confidence is a consequence of the model performing well in a high quality verification and validation (V&V) process. However, all V&V efforts are necessarily constrained by time and resources. This gives rise to the question 'How can I best allocate limited V&V resources to expose the model's strengths or weaknesses. We want to help the DM answer the question 'how much force should the model have?' This paper presents a method for answering this question. We do this using a generalization of the Line of Sight Evidential Reasoning (LSERA) (William Bunting 2012) methodology. Line of Sight (LoS) is defined as the ability to make a probabilistic inference about how performing (or not performing) a test will in influence our belief about the validity of the model. This is accomplished by building an evidence based Bayesian network (BN) which captures our belief about the model's validity. Information entropy (or Shannon entropy) (C. E. Shannon 1948) is used to analyze the BN. We present an example of how to apply the generalized LSERA methodology using a decision support model called EXOGENiUS (Anamaria Berea, Daniel Maxwell 2017). The EXOGENiUS system is used to inform investment decisions in early stage startups.

The rest of this paper is organized as follows:

- In the **Background** section we introduce verification and validation, Bayesian networks, information entropy, and the generalized LSERA methodology;
- In the **Assessing EXOGENiUS** section we apply the generalized LSERA methodology to the EXOGENiUS model to produce a ranked list of V&V activities.

2. Background

2.1. Verification and Validation

Model verification is often defined as "ensuring that the computer program of the computerized model and its implementation are correct". Model validation is usually defined to mean "substantiation that a computerized model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model" (Sargent 2010).

Upon completion of model V&V many organizations (the US Department of Defense (DoD) in particular) formally "Accredit" a model. When this happens, the model has been approved to use for a "specific purpose." (VV&A) However, no definition of the term "specific purpose" is usually provided and its interpretation can be very broad. In the DoD, specific purposes can include things such as "campaign analysis", "force structure analysis", and simulation based testing of systems." Once V&V'd, many different analytic studies may be performed with the model; and, each of these studies could inform a unique set of decisions being made by various DMs. What the DM really wants to know is this: "Given the decision I am trying to make, how much force should the model have in informing my (our) decision?"

A critical difference between a model having "force" and being "V&V'd" is this: a model can be V&V'd upon its completion. However, we cannot establish how much force a model should have until the specific decision being informed is known. This implies that given any decision/decision maker combination, a model may have more or less force. This further implies that for each decision we need to make a separate assessment of the model's force. This leaves us in a bind as we are seldom given sufficient time and resources to completely re-validate the model. The generalized LSERA methodology can be quickly applied and used to identify V&V tasks which give us the most insight into model force.

While there are dozens of possible types of V&V activities; in this paper we consider only the following eight V&V techniques (Sargent 2010):

- **Predictive Validation**: In this type of validation, a model would be used to predict (forecast) the performance of one or more companies. These predictions would then be compared to actual company performance. We consider predictive validation to be the strongest (most diagnostic) of the V&V techniques presented; however, this type of validation is a continuous process in which new data is used to both update and validate the model.
- **Historical Data Validation**: Historical data validation differs from predictive validation in that it relies on data which has already been collected. If

adequate historical data exist (e.g., data collected on companies) then a model's predictions (based on historical data) could be compared to the actual company performance. Historical data validation is nearly as strong as predictive validation without the substantial waiting period. However, it depends on the existence of accurate historical data, which may not exist.

- **Comparison to Other Models**: a model's results can be compared to the results produced by competing models. Variations in the results can then be explored. This technique is only as robust as the set of competing models. If no quality competing models exist then this may not be a good choice.
- **Parameter Variability – or Sensitivity Analysis**: Here we systematically change the input values of a model and examine the effect on the model predictions. A design of experiment (a matrix of carefully chosen inputs) is often used with this technique. This method can be used to determine which model inputs have the greatest effect on the model outputs.
- **Traces**: The effect of different model inputs are traced through the system node by node. This is particularly easy to do in BNs.
- **Degenerate Tests**: Here we would attempt to devise combinations of inputs which would cause a model's behavior to degenerate. For example, in a queueing model the average number in the queue of a single server should continue to increase over time when the arrival rate is larger than the service rate.
- **Extreme Condition Tests**: The model's outputs should be plausible when given unlikely or extreme inputs. This technique is useful for understanding the range of valid inputs for the model.
- **Face Validity**: Here we ask subject matter experts whether or not the model and or model results are reasonable. This is a weak (not very diagnostic) technique which very much depends on the quality of the experts. However, it is used quite frequently because it has the unique advantage of getting a larger community of experts familiar with (and hopefully supporting) the model.

2.2. Bayesian Networks

'A Bayesian network is a representational device that is meant to organize one's knowledge about a particular situation into a coherent whole.' (Darwiche 2014) In a Bayesian Network (BN) we use a graph to capture two important aspects of our belief. Nodes represent random variables and capture the information we consider relevant to our decision. Arcs represent conditional dependencies between nodes and capture our belief about the relationships which exist between the information. A BN can be drawn as a graph which must be directed and acyclic. A fully specified Bayesian Network contains quantitative probability judgments that reflect the joint probability distribution over all of the variables in the model and support the execution of Bayes Rule. There are two large advantages of using a Bayesian Network for this application of LSERA: 1) it allows us to clearly visualize the relationship between our hypotheses about model validity and the applicability of the V&V techniques under consideration, and 2) BNs are computationally very efficient.

The standard form of Bayes rule is:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} \quad (1)$$

P(A) is the prior probability of event A. P(B) is the prior probability of event B. P(A|B) is the likelihood of event A given that event B has happened.

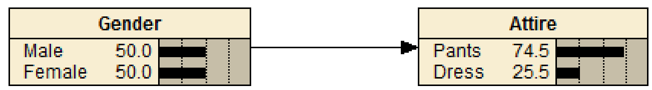

Figure 2. Simple Bayesian Network

The probability, P(B|A), is the posterior probability of event B given that event A has already occurred. Here is a simple example with two variables, gender and attire. Suppose we know that 99% of the male population wears pants and the remaining 1% wears a dress. We also know that 50% of the female population wears pants and the remaining 50% wears a dress. If our population is evenly divided between men and women, we can calculate that 74.5% (.99*.5 + .5*.5) of the population wears pants using the chain rule of probability. These probabilities are shown in the simple Bayesian network (Norsys) shown in Figure 2.

Now we observe that a person is wearing pants and we want to know if they are male or female. Using equation (1) we can calculate this posterior probability using Bayes rule:

$$P(Male|Pants) = \frac{P(Pants|Male)P(Male)}{P(Pants)} = \frac{0.99 \times 0.5}{0.745}$$
$$= 0.664$$

This result can also be calculated using our simple BN by entering evidence into the Attire node. When we do this the posterior probabilities are calculated in the Gender node. This is shown in **Errore. L'origine riferimento non è stata trovata.** by the updated probabilities.

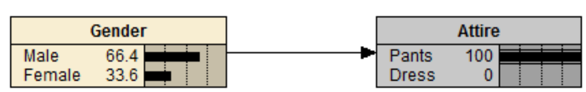

Figure 3. Simple Bayesian Network with Evidence

3. Background Information Entropy (IE) and Mutual Information (MI)

This paper addresses the topic of ranking V&V activities. We will do this by determining which V&V

activities provide us with the most information about model force. Said another way, we will rank V&V activities by how diagnostic they are. Information entropy H(X) is a measure of how much information will be produced if the value of a discrete random variable X becomes known (C. E. Shannon 1948). If X has the possible values {x_1, x_2, ⋯x_n } with a probability mass function P(X) then H(X) is computed as follows:

$$H(X) = -\sum_{i=1}^{n} P(x_i) log_2 P(x_i) \qquad (2)$$

IE reaches its maximum value for *X* when all possible values x_i are equally likely. When this happens, we can say that no information as to the value of *X* is available. Conversely, IE will reach its minimum value of zero when we know the value of *X*. At this point we can say that we have perfect information as to the value of *X*. **Errore. L'origine riferimento non è stata trovata.** shows IE calculations for a random variable for the case when we have no information and the case when we have near perfect information.

Random variable X with no information available		
Values	P(X=xi)	log2P(xi)P(xi)
x1	0.33	-0.53
x2	0.33	-0.53
x3	0.33	-0.53
	H(X) =	1.58

Random variable X with near perfect information		
Values	P(X=xi)	log2P(xi)P(xi)
x1	0.998	-0.003
x2	0.001	-0.010
x3	0.001	-0.010
	H(X) =	0.02

Figure 4. Information Entropy Calculations

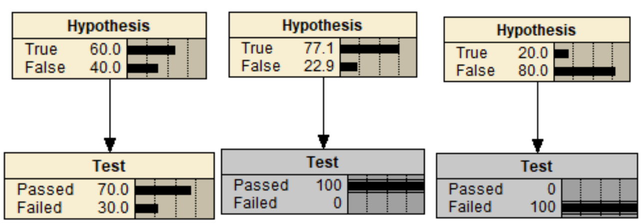

Figure 5. A Simple Bayesian Network

Consider the simple BNs shown in **Errore. L'origine riferimento non è stata trovata.**. Knowing whether or not the test has passed or failed can change our belief about whether the Hypothesis is true or false. The expected change in the hypothesis' IE as a result of knowing the value of the test is called mutual information (MI). MI is a measure of the IE reducing potential the test. We will use MI to rank possible V&V activities based on their ability to reduce the hypothesis' IE. (In this case the "Force" a model should have.)

state	P(x)	Entropy
TRUE	0.6	0.44
FALSE	0.4	0.53
	H(Hyp)	0.97

state	P(x)	Entropy
TRUE	0.771	0.29
FALSE	0.229	0.49
	H(Hyp|Test=Pass)	0.78

state	P(x)	Entropy
TRUE	0.2	0.46
FALSE	0.8	0.26
	H(Hyp|Test=Fail)	0.72

Figure 6. Entropy Calculations

Using equation **Errore. L'origine riferimento non è stata trovata.** the entropy calculations for the simple BN's Hypothesis node are shown in **Errore. L'origine riferimento non è stata trovata.**. When the test is passed the reduction in entropy is *0.97 − 0.78 = 0.19*. When the test fails the reduction in entropy is *0.97 − 0.71 = 0.25*. Recall our prior belief is the test will pass 70% of the time. Using this we calculate *MI = 0.7 * 0.19 + 0.3 * 0.25 = 0.21*. The formal definition is shown in equation **Errore. L'origine riferimento non è stata trovata.**).

$$H(X|Y) = \sum_{y} P_Y(y) \left[-\sum_{x} [P_{X/Y}(x/y) log_2 P_{x/y}(x|y)] \right] \qquad (3)$$

3.1. Line of Sight Evidential Reasoning Analysis

In 2012, William Bunting defended his thesis 'Line of Sight Evidential Reasoning Analysis (LSERA)'. When the Federal Government acquires IT systems, it expects these systems to provide a reasonable return on investment. Bunting provided a method to calculate the likelihood that an IT system will in fact provide the expected return (William Bunting 2012). Bunting writes:

Federal agencies are modernizing at an increasing rate and, before authorizing funds for each specific modernization investment, the Office of Management and Budget (OMB) requires agencies to show alignment to agency mission outcomes. OMB defines the Federal Enterprise Architecture Performance

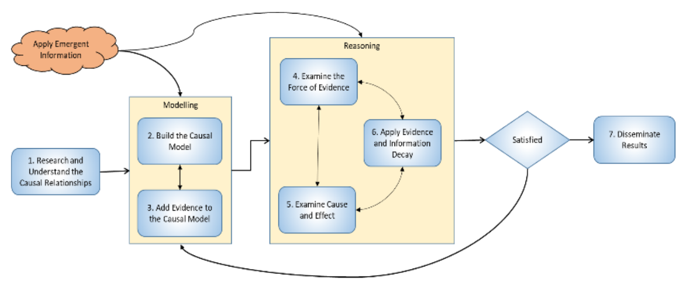

Figure 7. Updated LSERA Process

Reference Model (PRM) "Line of Sight" as a means to show this alignment. This research formulated the Line of Sight Evidential Reasoning Analysis (LSERA) method to provide a principled method for PRM Line of Sight analysis that improves reasoning under uncertainty, the examination of evidentiary force, and the inquiry into results of contemplated alternatives.

4. Generalized LSERA Process

Bunting's LSERA process assesses the likelihood an IT acquisition is aligned with the goals of a federal agency. This assessment is based on existing performance measures which have been established by the federal agency in question. However, there is no guarantee the agency's existing performance measures are sufficient to support the LSERA modeling process. Said another way, the agency may not be collecting the data needed to effectively assess whether or not the IT system is aligned with its goals. The generalized LSERA process presented here allows us to make this assessment.

4.1. Step 1: Research the Driving Argument

Our goal is to inform the DM about the credibility of the model they intend to use to inform their decision. To do this we need a compelling argument supported by evidence. In this step we examine the model and available documentation to develop an initial argument. The argument can be reviewed by SMEs and the DM and refined iteratively as required.

4.2. Step 2: Build the Belief Model

Construct a BN which codifies our belief about the argument. Review the BN with the DM and SMEs. Verify the BN reasonably represents our belief about the argument relationships.

4.3. Step 3: Add Evidence to the Belief Model

Here we add nodes to the BN which represent the evidence (V&V activities) which could be (or have been) done.

4.4. Steps 4, 5, and 6: Reasoning

In steps 4, 5, and 6 we examine the inferences and refine the model. We can rank evidence nodes using mutual information calculations. We can review these ranking with SMEs. This is an iterative process and we can loop back to earlier steps. For example, examination of the inference model may reveal details which cause us to update our understanding of the causal relationships (Step 1).

4.5. Step 7: Disseminate Results

Here we produce a prioritized list of data for collection. We call this the best next test (BNT).

4.6. Apply Emergent Information

The situation about which we are reasoning will change over time. For M&S V&V we expect this change to be slow. A new version of the M&S may be released every 3 months. These types of changes require us to refine our LoS model following the above methodology.

5. Assessing EXOGENiUS

5.1. EXOGENiUS Overview

EXOGENiUS is a Bayesian model used to inform investment decisions in early (angel) stage startups. The model is an influence diagram that combines historical data on startups and elicited expert judgment on the factors and relationships among those factors that impact the likelihood a startup company will succeed and an investor will get a return on their investment. (Anamaria Berea, Daniel Maxwell 2017) The underlying model is informed by a series of twenty questions whose answers are mapped to states in the probabilistic nodes of the Influence Diagram. Entrepreneurs answer these questions as part of a self-assessment and investors (or evaluators) as part of their investment decision in the early stage company. Applications of EXOGENiUS have included support to individual angel investors, start-up funding competitions, and educational outreach programs that for entrepreneurs. The size of the investment provided at angel stage is normally in the \$50K - \$250K range. Moreover, the base rate of startup failure is approximately 90% within five years, so successful investors efficiently assess many companies and invest in only a few.

Here is a step by step description of how the generalized LSERA process was applied.

5.2. Step 1: Research the Driving Argument

EXOGENiUS can be used for two different purposes. The first purpose is to help an angel stage investor decide whether or not to invest in a startup company. The second purpose is to help a startup company understand how to both make itself more attractive to investors and improve its likelihood of success. These will become the two central hypotheses of the generalized LSERA model.

EXOGENiUS produces three primary measures that are represented as value nodes in the influence diagram: the Value Proposition Measure, the Exit Potential Measure, and the Business Execution Measure. Supporting the three primary measures are secondary measures in five key business areas: Technology/Offering, Market, The Team, Financial/Capital, and Company Infrastructure. An examination of the structure of the EXOGENiUS model shown in **Errore. L'origine riferimento non è stata trovata.** is useful for identifying the key relationships.

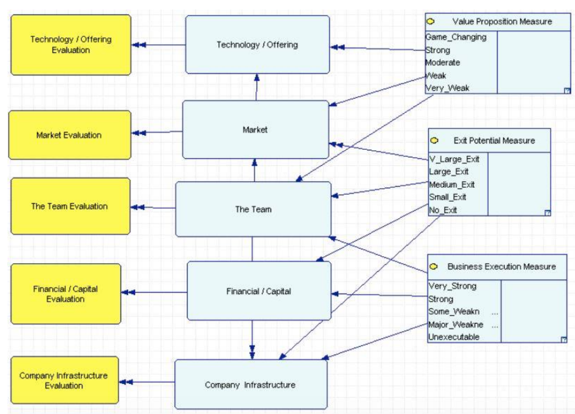

Figure 8. EXOGENiUS Overview

5.3. Step 2: Build the Belief Model

Structuring the Model

Figure 1 shows the structure of the LSERA belief model. This model was constructed using information obtained from model documentation, model subject matter experts (SMEs), and the EXOGENiUS model itself. The model was then reviewed with SMEs and revised.

Building the Conditional Probability Tables

Consider a Bayesian network where all the nodes are discrete with the same set of m states $\mathbf{S} = \{s_1, s_2, \ldots, s_m\}$. Let c be a child node with n parents $\mathbf{P}=\{p_1, p_2, \ldots, p_n\}$. Let the force each parent exerts on the child be $\mathbf{F} = \{f_1, f_2, \ldots, f_n\}$. Let the leak parameter λ represent a measure of our uncertainty. The conditional probability table (CPT) of c is based on the Cartesian product of the parent's states. This means c will have m^n rows in its CPT and $m^n(m-1)$ probabilities are required to completely specify c's CPT. Eliciting these probabilities is challenging in a non-trivial BN.

The following heuristic simplifies the elicitation, is used to calculate c's CPT, and requires only the elicitation of **F** and λ. When one of the child node's state becomes known it has the effect of moving the parent state in the same direction without short circuiting the remaining children. We denote the parent's states in the Cartesian product $\mathbf{PS} = \{ps_1, ps_2, \ldots, ps_n\}$. We denote the states in the child CPT $\mathbf{CS} = \{cs_1, cs_2, \ldots, cs_m\}$. We must calculate the probability of **CS** for each row in c's CPT. To do this we first calculate how much force each of the parents is exerting on each child state $\mathbf{F}\rightarrow\mathbf{S} = \{F\rightarrow cs_1, F \rightarrow cs_2, \ldots, F \rightarrow cs_m\}$.

$$F \rightarrow cs_m = \sum_{1}^{n} f_n \; where \; ps_n = cs_m \tag{4}$$

For the row in question we can now calculate the probability for each element of **S**:

$$P(cs_i|\boldsymbol{F} \rightarrow \boldsymbol{S}) = \frac{\lambda}{m} + \frac{(F \rightarrow cs_i)}{\sum \boldsymbol{F} \rightarrow \boldsymbol{S}}(1-\lambda) \tag{5}$$

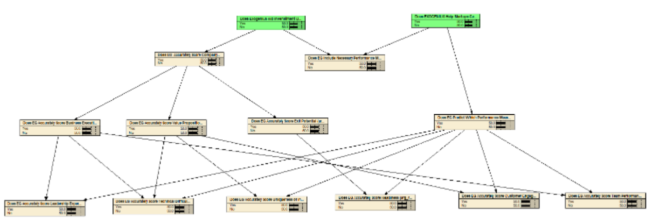

Figure 1. LSERA Belief Model

Table 1 shows sample calculations for a child node with 2 parents. All parent nodes have 2 states s1 and s2. The child node has 3 states s1, s2, and s3. In this example F = {2, 3} and λ=10%. The Cartesian product of the parent's states is displayed in the first 2 columns. Columns 3-5 are calculated using equation **Errore. L'origine riferimento non è stata trovata.**, and columns 6-8 are calculated using equation **Errore. L'origine riferimento non è stata trovata.**. A complete set of CPTs for the causal model is presented in Appendix A.

Table 1 shows sample calculations for a child node with 2 parents. All parent nodes have 2 states s1 and s2. The child node has 3 states s1, s2, and s3. In this example F = {2, 3} and λ=10%. The Cartesian product of the parent's states is displayed in the first 2 columns. Columns 3-5 are calculated using equation **Errore. L'origine riferimento non è stata trovata.**, and columns 6-8 are calculated using equation **Errore. L'origine riferimento non è stata trovata.**. A complete set of CPTs for the causal model is presented in Appendix A.

5.4. Step 3: Add Evidence to the Belief Model

Recall that we are choosing between the 8 different types of V&V activities shown in table 2. Each of these V&V activities could be performed on any of the causal nodes. **Errore. L'origine riferimento non è stata trovata.** shows the completed LSERA BN. Each of the orange evidence nodes in the outside ring represents a type of V&V activity which could be performed on a specific part of the EXOGENiUS model. As you can see, there are a lot of V&V activities from which to choose.

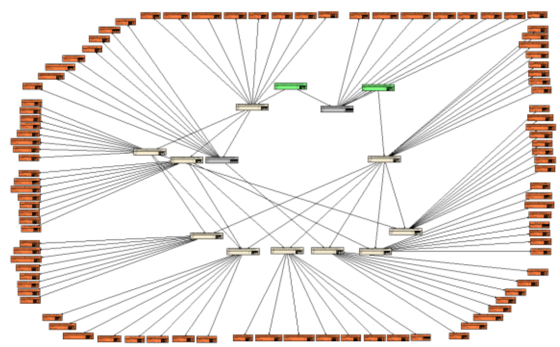

Figure 10. LSERA Belief Model with Evidence

The CPTs for each of the evidence nodes are computed in the same way as the causal nodes. However, when there is only 1 parent the force F has no impact on the child CPT. This makes the CPT a function of the leak parameter λ. As λ get smaller the diagnostic value of the V&V activity increases. Table 2 and Table 3 show CPT calculations for two different values of λ.

Table 1. Example CPT Calculations

Parent States		Parents Force on Child's State			Child CPT		
ps_1	ps_2	F -> s_1	F -> s_2	F -> s_3	s_1	s_2	s_3
s1	s1	5	0	0	95.0%	5.0%	5.0%
s1	s2	2	3	0	41.0%	59.0%	5.0%
s1	s3	2	0	3	41.0%	5.0%	59.0%
s2	s1	3	2	0	59.0%	41.0%	5.0%
s2	s2	0	5	0	5.0%	95.0%	5.0%
s2	s3	0	2	3	5.0%	41.0%	59.0%
s3	s1	3	0	2	59.0%	5.0%	41.0%
s3	s2	0	3	2	5.0%	59.0%	41.0%
s3	s3	0	0	5	5.0%	5.0%	95.0%

Table 2. CPT for λ = 10%

Parent States	Parents Force on Child's State			Child CPT		
ps_1	F -> s_1	F -> s_2	F -> s_3	s_1	s_2	s_3
s_1	1	0	0	93.3%	3.3%	3.3%
s_2	0	1	0	3.3%	93.3%	3.3%
s_3	0	0	1	3.3%	3.3%	93.3%

Table 3. CPT for λ = 50%

Parent States	Parents Force on Child's State			Child CPT		
ps_1	F -> s_1	F -> s_2	F -> s_3	s_1	s_2	s_3
s_1	1	0	0	66.7%	16.7%	16.7%
s_2	0	1	0	16.7%	66.7%	16.7%
s_3	0	0	1	16.7%	16.7%	66.7%

5.5. Steps 4, 5, and 6: Reasoning

Recall that the EXOGENiUS model has two different uses. The first use is to help a venture capitalist decide whether or not to invest in a startup company. The second use is to help a startup company understand how to make itself more attractive to venture capitalists. The intended use of the model will impact how V&V activities are ranked. **Table 5** shows how the V&V activities are ranked for each of the two model uses.

In this paper we are interested in situations where V&V time and resources are limited, so the overhead associated with applying the generalized LSERA methodology must be manageable. The generalized LSERA model presented here is high level; and, it was created with about 8 hours of developer time and 2 hours of SME review time. Time and resources permitting, this high level model could be expanded to include additional details. In addition, BNs are highly transparent and interactive. This enables real time what-if analysis. We can easily answer questions such as 'How will my V&V rankings change if I perform test A?

Table 5. Ranking of V&V Activities for EXOGENiUS

	Does Exogenius Aid Investment Decisions?		Does Exogenius Help Startups Get Funding?	
Rank	Description	Mutual Information	Description	Mutual Information
1	Predictive Validation on Does Eg Accurately Score Company Performance?	0.0729	Predictive Validation on Does Eg Predict Which Performance Measure Will Improve Funding?	0.0729
2	Predictive Validation on Does Eg Accurately Score Exit Potential?	0.0590	Historical Data Valication on Does Eg Predict Which Performance Measure Will Improve Funding?	0.0506
3	Predictive Validation on Does Eg Accurately Score Business Execution?	0.0590	Comparison to Other Models on Does Eg Predict Which Performance Measure Will Improve Funding?	0.0324
4	Predictive Validation on Does Eg Accurately Score Value Proposition?	0.0590	Parameter Variation on Does Eg Predict Which Performance Measure Will Improve Funding?	0.0182
5	Historical Data Valication on Does Eg Accurately Score Company Performance?	0.0506	Predictive Validation on Does Eg Include Necessary Performance Measures?	0.0182

shows the values of λ which were used. A complete set of CPTs for the evidence nodes is presented in appendix B.

Table 4. λ Values for V&V Activities

V&V Technique	λ
Predictive Validation	0.3
Historical Data Validation	0.35
Comparison to Other Models	0.4
Parameter Validation	0.5
Traces	0.55
Degenerate Tests	0.65
Extreme Condition Tests	0.7
Face Validity	0.8

5.6. Steps 4, 5, and 6: Reasoning

Recall that the EXOGENiUS model has two different uses. The first use is to help a venture capitalist decide whether or not to invest in a startup company. The second use is to help a startup company understand how to make itself more attractive to venture capitalists. The intended use of the model will impact how V&V activities are ranked. **Table 5** shows how the V&V activities are ranked for each of the two model uses.

In this paper we are interested in situations where V&V time and resources are limited, so the overhead associated with applying the generalized LSERA methodology must be manageable. The generalized LSERA model presented here is high level; and, it was created with about 8 hours of developer time and 2 hours of SME review time. Time and resources permitting, this high level model could be expanded to include additional details. In addition, BNs are highly transparent and interactive. This enables real time what-if analysis. We can easily answer questions such as 'How will my V&V rankings change if I perform test A?

Table 5. Ranking of V&V Activities for EXOGENiUS

	Does Exogenius Aid Investment Decisions?		Does Exogenius Help Startups Get Funding?	
Rank	Description	Mutual Information	Description	Mutual Information
1	Predictive Validation on Does Eg Accurately Score Company Performance?	0.0729	Predictive Validation on Does Eg Predict Which Performance Measure Will Improve Funding?	0.0729
2	Predictive Validation on Does Eg Accurately Score Exit Potential?	0.0590	Historical Data Valication on Does Eg Predict Which Performance Measure Will Improve Funding?	0.0506
3	Predictive Validation on Does Eg Accurately Score Business Execution?	0.0590	Comparison to Other Models on Does Eg Predict Which Performance Measure Will Improve Funding?	0.0324
4	Predictive Validation on Does Eg Accurately Score Value Proposition?	0.0590	Parameter Variation on Does Eg Predict Which Performance Measure Will Improve Funding?	0.0182
5	Historical Data Valication on Does Eg Accurately Score Company Performance?	0.0506	Predictive Validation on Does Eg Include Necessary Performance Measures?	0.0182

6. Discussion and Conclusions

In this paper we explained why a model's force cannot be established until both the decision and the decision maker are known. We then presented a generalization of the Line of Sight Evidential Reasoning and Analysis methodology. We then used the generalized methodology to construct a Bayesian network which codified our belief about how specific verification and validation activities change a model's force on a decision maker. Using the Bayesian network, we then showed how the rank of verification and validation activities differs based on the intended use of the model.

Our methodology provides a quantitative basis for allocating V&V resources. For simple cases, the methodology present here can be implemented in a few hours. The generalized LSERA BN can be iteratively refined as time allows. Going forward, as the model is used to inform new decisions the BN can be easily modified or largely reused.

We do not attempt to answer the question 'How much V&V is enough?' However, based on the beliefs codified in the BN, we do make a probabilistic assessment about how much force the model has. Future research may be able use this fact to develop on overall metric for model credibility. Then for any given use of the model, this metric could be used to decide when enough V&V has been performed.

In this paper, the generalized LSERA methodology was applied to a model (EXOGENiUS) which had already been build. However, we believe the methodology could be applied early in the model development life cycle before model construction begins. By consulting DMs who are potential users of the model, a generalized LSERA BN could be developed to reveal short falls in model requirements or design. We could establish and maintain LoS between the model development process and the intended DM's needs over the entire model development life cycle.

Appendix A. CPTs for the Belief Model

The following CPTs were calculated using method described in the *Building the Conditional Probability Tables* section. The Leak parameter λ = 0.1, and all parents were assumed to have equal force. As a result, all nodes with the same number of parents will have the same CPT values.

Node: Does EG Accurately Score Company Performance?

Parents:

P_1: Does Exogenius Aid Investment Decisions?

CPT:

Parents			CPT	
P_1	F -> s_1	F -> s_2	Yes	No
Yes	1	0	95%	5%
No	0	1	5%	95%

Node: Does EG Accurately Score Business Execution?

Parents:

P_1: Does EG Accurately Score Company Performance?

CPT:

Parents			CPT	
P_1	F -> s_1	F -> s_2	Yes	No
Yes	1	0	95%	5%
No	0	1	5%	95%

Node: Does EG Accurately Score Value Proposition?

Parents:

P_1: Does EG Accurately Score Company Performance?

CPT:

Parents			CPT	
P_1	F -> s_1	F -> s_2	Yes	No
Yes	1	0	95%	5%
No	0	1	5%	95%

Node: Does EG Accurately Score Exit Potential?

Parents:

P_1: Does EG Accurately Score Company Performance?

CPT:

Parents			CPT	
P_1	**F -> s_1**	**F -> s_2**	**Yes**	**No**
Yes	1	0	95%	5%
No	0	1	5%	95%

Node: Does EG Predict Which Performance Measure Will Improve Funding?

Parents:

P_1: Does EXOGENiUS Help Startups Get Funding?

CPT:

Parents			CPT	
P_1	**F -> s_1**	**F -> s_2**	**Yes**	**No**
Yes	1	0	95%	5%
No	0	1	5%	95%

Node: Does EG Include Necessary Performance Measures?

Parents:

P_1: Does Exogenius Aid Investment Decisions?
P_2: Does EXOGENiUS Help Startups Get Funding?

CPT:

Parents				CPT	
P_1	**P_2**	**F -> s_1**	**F -> s_2**	**Yes**	**No**
Yes	Yes	2	0	95%	5%
Yes	No	1	1	50%	50%
No	Yes	1	1	50%	50%
No	No	0	2	5%	95%

Node: Does EG Accurately Score Leadership Experience?

Parents:

P_1: Does EG Accurately Score Business Execution?
P_2: Does EG Predict Which Performance Measure Will Improve Funding?

CPT:

Parents				CPT	
P_1	**P_2**	**F -> s_1**	**F -> s_2**	**Yes**	**No**
Yes	Yes	2	0	95%	5%
Yes	No	1	1	50%	50%
No	Yes	1	1	50%	50%
No	No	0	2	5%	95%

Node: Does EG Accurately Score Uniqueness of Innovation?

Parents:

P_1: Does EG Accurately Score Value Proposition?
P_2: Does EG Predict Which Performance Measure Will Improve Funding?

CPT:

Parents				CPT	
P_1	**P_2**	**F -> s_1**	**F -> s_2**	**Yes**	**No**
Yes	Yes	2	0	95%	5%
Yes	No	1	1	50%	50%
No	Yes	1	1	50%	50%
No	No	0	2	5%	95%

Node: Does EG Accurately Score Readiness?

Parents:

P_1: Does EG Accurately Score Exit Potential?
P_2: Does EG Predict Which Performance Measure Will Improve Funding?

CPT:

Parents				CPT	
P_1	**P_2**	**F -> s_1**	**F -> s_2**	**Yes**	**No**
Yes	Yes	2	0	95%	5%
Yes	No	1	1	50%	50%
No	Yes	1	1	50%	50%
No	No	0	2	5%	95%

Node: Does EG Accurately Score Customer Engagement?

Parents:

P_1: Does EG Accurately Score Value Proposition?
P_2: Does EG Predict Which Performance Measure Will Improve Funding?

CPT:

Parents				CPT	
P_1	**P_2**	**F -> s_1**	**F -> s_2**	**Yes**	**No**
Yes	Yes	2	0	95%	5%
Yes	No	1	1	50%	50%
No	Yes	1	1	50%	50%
No	No	0	2	5%	95%

Node: Does EG Accurately Score Team Performance?

Parents:

P_1: Does EG Accurately Score Business Execution?
P_2: Does EG Predict Which Performance Measure Will Improve Funding?

CPT:

Parents				CPT	
P_1	**P_2**	**F -> s_1**	**F -> s_2**	**Yes**	**No**
Yes	Yes	2	0	95%	5%
Yes	No	1	1	50%	50%
No	Yes	1	1	50%	50%
No	No	0	2	5%	95%

Node: Does EG Accurately Score Technical Difficulty?

Parents:

P_1: Does EG Accurately Score Business Execution?
P_2: Does EG Accurately Score Value Proposition?
P_3: Does EG Predict Which Performance Measure Will Improve Funding?

CPT:

Parents					CPT	
P_1	P_2	P_3	F -> s_1	F -> s_2	Yes	No
Yes	Yes	Yes	3	0	95%	5%
Yes	Yes	No	2	1	65%	35%
Yes	No	Yes	2	1	65%	35%
Yes	No	No	1	2	35%	65%
No	Yes	Yes	2	1	65%	35%
No	Yes	No	1	2	35%	65%
No	No	Yes	1	2	35%	65%
No	No	No	0	3	5%	95%

Appendix B. CPTs for the Evidence Nodes

These nodes represent types of V&V which could be performed against the EXOGENiUS model. The CPTs represent the belief that some tests are stronger (able to reveal more information about EXOGENiUS) than others. These CPTs are present in order of their strength.

Node: Predictive Validation

Parent	CPT	
P_1	Valid	Not Valid
Yes	80%	20%
No	20%	80%

Node: Historical Data Validation

Parent	CPT	
P_1	Valid	Not Valid
Yes	75%	25%
No	25%	75%

Node: Comparison to Other Models

Parent	CPT	
P_1	Valid	Not Valid
Yes	70%	30%
No	30%	70%

Node: Parameter Variation

Parent	CPT	
P_1	Valid	Not Valid
Yes	65%	35%
No	35%	65%

Node: Traces

Parent	CPT	
P_1	Valid	Not Valid
Yes	62%	38%
No	38%	62%

Node: Degenerate Tests

Parent	CPT	
P_1	Valid	Not Valid
Yes	60%	40%
No	40%	60%

Node: Extreme Condition Tests

Parent	CPT	
P_1	Valid	Not Valid
Yes	58%	42%
No	42%	58%

Node: Face Validity

Parent	CPT	
P_1	Valid	Not Valid
Yes	55%	45%
No	45%	55%

References

VV&A, 12/9/2009: DoD Instruction 5000.61. Available online at http://www.dtic.mil/whs/directives/corres/pdf/500061p.pdf.

Anamaria Berea; Daniel Maxwell (2017): A Bayesian Model for Investment Decisions In Early Ventures.

C. E. Shannon (1948): A Mathematical Theory of Communication. In Bell Labs Technical Journal 27 (3), pp. 379–423.

Darwiche, Adnan (2014): Modeling and reasoning with Bayesian networks. New York, N.Y.: Cambridge University Press.

Norsys: Netica. Available online at https://www.norsys.com/netica.html.

Sargent, Robert G. (2010): Verification and validation of simulation models. In : WSC 2010, Winter Simulation Conference. Baltimore, MD, USA, pp. 166–183.

Schum, D.: The Evidential Foundations of Probabilistic Reasoning.

Waltz, Edward (2003): Knowledge management in the intelligence enterprise. London: Artech House.

William Bunting (2012): Enterprise Line of Sight Analysis Using Evidential Reasoning. George Mason University, Fairfax, VA.