A Probabilistic Approach to System Maturity Assessment

Weiping Tan, Jose Ramirez-Marquez, and Brian Sauser*

Stevens Institute of Technology, School of Systems and Enterprises, Systems Development and Maturity Laboratory, Castle Point on Hudson, Hoboken, NJ 07030

ABSTRACT

Prescriptive metrics have been widely accepted and used in engineering management to assess the progress and success of engineering efforts. However, these types of metrics have two major challenges: human subjectivity and confidence in data estimates. In this paper we use a system-focused prescriptive metric entitled System Readiness Level (SRL) [a function of Technology and Integration Readiness Levels (TRL/IRL)], to propose a probabilistic approach to estimating the development maturity of a system of interest. In order to reduce the subjective impact, we propose a probabilistic method by assigning probability distributions to the estimated TRLs and IRLs to reflect the reality of evaluators’ subjective estimates. Based on these probability distributions of TRLs and IRLs, a Monte-Carlo simulation methodology is used to assess the maturity status of the whole system and its components. An illustrative example is examined to show the proposed methodology and to investigate its implication to engineering management. The paper concludes with a discussion of the added value of this new methodology, its limitation, and future work. © 2010 Wiley Periodicals, Inc. Syst Eng 14: 279–293, 2011

Key words: prescriptive metric; subjectivity; Technology Readiness Level; Integration Readiness Level; System Readiness Level

1. INTRODUCTION

For any engineering project, the primary goal is to provide its customer(s) with a product (or system) at the desired performance level within a predetermined budget and schedule. In order to evaluate the development and success of these efforts, engineering managers will use metrics to monitor a project’s progress against a baseline. This routinely results in what Íniesta [1994, p. 1] specifies as “what is not measured is not controlled; what is not tracked is not done.” Therefore, metrics are commonly needed to measure project progress and demonstrate the magnitude of achieved performance [Griffin, 1997] while allowing for a successful evaluation that becomes a necessary precursor of improvement [Tervonen and Isakka, 1996]. As an example, Nambisan [2002] showed that the lack of process rigor and metrics resulted in numerous software project failures, and thus recommended that companies, especially those who want to evolve from start-ups to stars, hold positive attitude towards process and metrics.

While the types of metrics used in engineering projects can be as numerous and diverse as the projects themselves, generally, metrics can be categorized into two types: descriptive or prescriptive [Fan and Yih, 1994; Harjumaa, Tervonen, and Salmela, 2008; Tervonen and Isakka, 1996]. Descriptive metrics are objective metrics describing natural phenomena and facts, and they are derived from analyzing the charac-
teristics and behavior of natural processes and objects. For example, the physical, electrical, strength, and optical metrics of a material can be derived from observations. Prescriptive metrics conversely are derived from descriptive metrics along with relevant information (i.e., knowledge, regulatory codes, standards), which specify the status that an objective can be considered accomplished. A prescriptive metrics’ approach can be understood as the combination of descriptive metrics and relevant, valid state information, which associates semantics with its measures, and thus adds a reference dimension into the metric. For simplicity, descriptive metric is whatever its value is, and prescriptive metric is whatever it is plus an interpretation of what state it represents. For example, human body temperature, blood pressure, and pulse are prescriptive. This added reference dimension has already engendered the requirement for the derivation of prescriptive metrics, which are used to effectively and economically manage the quality assurance activities that are considered to be important for success [Fan and Yih, 1994].

While the intention of assessing an engineering project with metrics is to use valid input data, for prescriptive metrics, estimates are predominately formulated by subject matter experts that can be greatly influenced by traits in human nature (e.g., seeking out information that supports our existing point of view, to give disproportionate weight to the first information received, or to make choices in a way that justifies past choices) [Hitt, Miller, and Colella, 2005]. By comparing the selection of reference points of R&D professionals between those from public institutes and from private institutes, Lee and Shin [2000] found that egocentric biases and personal goals play a large role in human beings’ evaluation process. Since such cognitive bias is involved in assessment, subjectivity is more or less inherent in our estimation, and it is very hard to avoid its influence [Yan et al., 2006]. By saying subjectivity in this paper, it means the variation of viewpoints towards the same thing (e.g., the maturity of a technology) from different people. Hu, Huang, and Zhang [2007] state that human subjectivity can influence and undermine the effectiveness of evaluation. Conversely, Redmill [2002b] specifies that subjectivity in analysis and decision-making should be recognized and allowed, since exposure could be beneficial when subjectivity is arbitrary and thus reduced when better understanding is spent to improve accuracy. Therefore, it is necessary to identify and mitigate the risk resulting from subjectivity when prescriptive metrics are used for project evaluation. In general, if more information is given in advance, there will be less subjectivity, and vice versa.

Besides the consideration of subjectivity, the reliability of the estimation has also drawn much attention in the application of prescriptive metrics. For estimations that use firmly established and well-understood metrics, point estimates may be enough to communicate the status. However, this is not always the situation; when the object being estimated is complex or the metric being used is new and arguable, point estimates are less reliable. In such a situation, estimators may be requested to provide information, which can quantify the confidence level or reliability of their estimation. Oliveira et al. [Braga, Oliveira, and Meira, 2007; Oliveira and Meira, 2006] pointed out in their work on software effort estimation that it is important to provide a confidence interval to quantify the reliability for the estimation. As the requirement for reliable estimation is very common in evaluation with prescriptive metrics, confidence intervals are desirable information that should be provided when possible.

Therefore, in this paper we will demonstrate a more rigorous approach to the evaluation of a set of prescriptive metrics used in engineering projects for assessing the maturity of system’s development. Commonly referred to as readiness levels (i.e., Technology Readiness Level, Integration Readiness Level, and System Readiness Level), these prescriptive metrics, each derived for its particular purpose, are widely acknowledged, but attention needs to be paid to the problems that have evolved during their evolution and application (i.e., they are error-prone, human-intensive, and subjective) [Sauser and Ramirez-Marquez, 2009; Yacoub and Ammar, 2002]. Thus, this paper proposes a probabilistic methodology and applies it to this set of readiness levels. First, we will briefly describe the history of various readiness level metrics and methods to employ them. By examining the shortcomings of the current methods for estimation, we emphasize the need to develop a modified methodology to lower the impact of human subjectivity on estimation and to formulate a more confident estimation of these readiness level metrics. We then describe our approach, demonstrate its application and value with a case example, and discuss the potential implications of this approach to engineering management. We conclude with considerations and limitations of our approach and provide suggestions for future work in this area.

2. DEFINITION AND APPLICATION OF READINESS LEVELS

2.1. Readiness Levels

Using metrics to measure system performance is deemed by some to be critical for supporting decision making and people motivation [Chiesa et al., 2008]. Morris and Pinto [2004] echoed that a suite of performance metrics that provides direct line of sight feedback on current project performance and anticipated future success is one of the three key elements to ensure projects are consistently done right. The other two elements consist of an effective means of learning from experience on projects and portfolio and program management processes for fully resourcing projects. Thus, for techniques in prescriptive metrics, Smith and Winterfeldt [2004] state that they can allow people to make better decisions by using normative models, but with knowledge of the limitations and descriptive realities of human judgment. In systems engineering management, we often rely on prescriptive or soft metrics to make informed decisions, but these metrics are measured through subjective judgment, are relatively easy to derive, and require a complementary rationale that explains their assessment [Dowling and Pardoe, 2005].

While using metrics contributes to management success, using unjustified metrics can cause miscommunication, and even good metrics can cause unjustified decisions if used in an improper way or in an unfit situation. Blackburn and Valerdi [2009] have performed a thorough review of the evolution of metrics knowledge, identified several common
metric selection mistakes, and provided guidance in the formation and execution of metrics for systems engineering decision making.

Given its pragmatic and successful application, the prescriptive and soft metric of Technology Readiness Levels (TRL) have been used within agencies of the United States (U.S.) government to assess the maturity of evolving technologies prior to incorporating them into a system or sub-system. While TRL is widely adopted in the U.S., it has also been used by several other government or international organizations, such as Canada, the United Kingdom, Japan, and the North Atlantic Treaty Organization (NATO). Each of these organizations has tailored the TRL metric to satisfy their own requirements for implementation. Table I is a comparison of how four of these organizations have interpreted the TRL metric. It should be noted that, in this paper, maturity means how ready a technology/integration/system is for its application in an intended environment.

While the use of TRL has increased, there have been many attempts to identify alternative readiness (maturity) levels that will complement TRL. For example, the Department of Defense (DoD) has developed a Manufacturing Readiness Level which focuses on manufacturing risks and maturity [Cundiff, 2003]. Because integration has been shown to be a complex effort in its own right [Jain et al., 2008], two independent efforts, within the National Aeronautics and Space Administration (NASA) and Stevens Institute of Technology, have developed a set of integration maturity metrics (i.e., Integration Readiness Level) [Gove, Sauser, and Ramirez-Marquez, 2007; Rasky, 2003; Sauser et al., 2009a]. While many of the readiness level metrics have focused on hard-

Table I. Sample Technology Readiness Level Definitions

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>N/A</td>
<td>N/A</td>
<td>Basic research with future military capability in mind</td>
<td>Basic research with future military capability in mind</td>
</tr>
<tr>
<td>1</td>
<td>Basic principles observed and reported</td>
<td>Initial concept verified against first principles and evaluation criteria defined.</td>
<td>Initial concept verified against first principles and evaluation criteria defined.</td>
<td>Initial concept verified against first principles and evaluation criteria defined.</td>
</tr>
<tr>
<td>2</td>
<td>Technology concept and/or application formulated</td>
<td>Technology concept and/or application formulated</td>
<td>Technical options evaluated and parametric ranges are defined for design</td>
<td>Technology concept and/or application formulated</td>
</tr>
<tr>
<td>3</td>
<td>Analytical and experimental critical function and/or characteristic proof-of-concept</td>
<td>Analytical and experimental critical function and/or characteristic design</td>
<td>Success criteria and technical specifications are defined as a range</td>
<td>Analytical and experimental critical function and/or characteristic design</td>
</tr>
<tr>
<td>4</td>
<td>Component and/or breadboard validation in laboratory</td>
<td>Component and/or breadboard validation in laboratory environment</td>
<td>Fuel design parameters and features defined</td>
<td>Component and/or &quot;breadboard&quot; validation in laboratory/field (e.g., ocean) environment</td>
</tr>
<tr>
<td>5</td>
<td>Component and/or breadboard validation in relevant environment</td>
<td>Component and/or breadboard validation in relevant environment</td>
<td>Process parameters defined</td>
<td>Component and/or &quot;breadboard&quot; validation in a relevant (operating) environment</td>
</tr>
<tr>
<td>6</td>
<td>System/subsystem model or prototype demonstration in a relevant environment (ground or space)</td>
<td>System/subsystem model or prototype demonstration in a relevant environment</td>
<td>Fuel safety basis established</td>
<td>System/subsystem model or prototype demonstration in a realistic (operating) environment or context</td>
</tr>
<tr>
<td>7</td>
<td>System prototype demonstration in a space environment</td>
<td>System prototype demonstration in an operational environment</td>
<td>All qualification steps completed and fuel is licensed</td>
<td>System prototype demonstration in an operational environment or context (e.g., exercise)</td>
</tr>
<tr>
<td>8</td>
<td>Actual system completed and &quot;flight qualified&quot; through test and demonstration (ground or space)</td>
<td>Actual system completed and &quot;flight qualified&quot; through test and demonstration</td>
<td>Reactor full-core conversion to new licensed fuel completed</td>
<td>Actual system completed and qualified through test and demonstration</td>
</tr>
<tr>
<td>9</td>
<td>Actual system &quot;flight proven&quot; through successful mission operations</td>
<td>Actual system &quot;flight proven&quot; through successful mission operations</td>
<td>Routine operations with licensed fuel established</td>
<td>Actual system operationally proven through successful mission operations</td>
</tr>
</tbody>
</table>

ware, NASA and DoD have also developed a similar set of metrics for application to software development and separately proposed a Software Readiness Level. Beyond these metrics, there have been a number of other developments leading to a proliferation of readiness levels, and this trend has all indications of continuing; drawing more and more effort and attention. Table II summarizes some of the various readiness level metrics that have been developed.

While most of the readiness levels function in isolation, a group from the Systems Development & Maturity Laboratory at Stevens Institute of Technology have developed a System Readiness Level (SRL) which is intended to assess current and future readiness (or developmental maturity) of a system by incorporating both the current TRL and an IRL (also proposed by the same group; see Table III) [Magnaye, Sauser, and Ramirez-Marquez, 2009; Sauser et al., 2008a, 2006].

Noting that technologies and the integrations among them are the two basic elements to the construction of a system, the SRL is defined as the function of TRLs of the technologies and IRLs of the integrations. It is one of the first instances that a single readiness level metric has been defined based on the combination of other existing readiness levels, thus also providing a new way to combine readiness level metrics. Other efforts have shown correlations between readiness levels, but not quantitatively combining individual readiness levels to determine a single readiness level metric. One aspect of its importance is that it gives credibility to the quantitative combination of readiness levels and opens the potential to further expand SRL by incorporating other readiness level metrics, such as Manufacturing Readiness Level and Software Readiness Level.
### Table II. Sample Readiness Level Metrics

<table>
<thead>
<tr>
<th>Readiness Metric</th>
<th>Organization</th>
<th>No. of levels</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manufacturing Readiness Level</td>
<td>DoD</td>
<td>10</td>
<td>[DoD, 2005]</td>
</tr>
<tr>
<td>Integration Readiness Level</td>
<td>NASA</td>
<td>5</td>
<td>[Rasky, Kolodziej, Farkas and Dutro, 2003]</td>
</tr>
<tr>
<td>Integration Readiness Level</td>
<td>Stevens Institute of Technology</td>
<td>9</td>
<td>[Sauser, Gove, Forbes and Ramirez-Marquez, 2009]</td>
</tr>
<tr>
<td>Countermeasure Readiness Level</td>
<td>NASA</td>
<td>9</td>
<td>[Institute, 2006]</td>
</tr>
<tr>
<td>Business Readiness Level</td>
<td>Northrop Grumman</td>
<td>10</td>
<td>[Leonard, 2008]</td>
</tr>
<tr>
<td>Innovation Readiness Level</td>
<td>University of Cambridge</td>
<td>6</td>
<td>[Lan]</td>
</tr>
<tr>
<td>System Readiness Level</td>
<td>Stevens Institute of Technology</td>
<td>5</td>
<td>[Sauser, Ramirez-Marquez, Magnaye and Tan, 2008]</td>
</tr>
</tbody>
</table>

### Table III. Integration Readiness Levels [Gove, 2007; Gove, Sauser, and Ramirez-Marquez, 2007]

<table>
<thead>
<tr>
<th>IRL</th>
<th>Definition</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>Integration is Mission Proven through successful mission operations.</td>
<td>IRL 9 represents the integrated technologies being used in the system environment successfully. In order for a technology to move to TRL 9, it must first be integrated into the system, and then proven in the relevant environment, so attempting to move to IRL 9 also implies maturing the component technology to TRL 9.</td>
</tr>
<tr>
<td>8</td>
<td>Actual integration completed and Mission Qualified through test and demonstration, in the system environment.</td>
<td>IRL 8 represents not only the integration meeting requirements, but also a system-level demonstration in the relevant environment. This will reveal any unknown bugs/defect that could not be discovered until the interaction of the two integrating technologies was observed in the system environment.</td>
</tr>
<tr>
<td>7</td>
<td>The integration of technologies has been Verified and Validated with sufficient detail to be actionable.</td>
<td>IRL 7 represents a significant step beyond IRL 6; the integration has to work from a technical perspective, but also from a requirements perspective. IRL 7 represents the integration meeting requirements such as performance, throughput, and reliability.</td>
</tr>
<tr>
<td>6</td>
<td>The integrating technologies can Accept, Translate, and Structure Information for its intended application.</td>
<td>IRL 6 is the highest technical level to be achieved, it includes the ability to not only control integration, but specify what information to exchange, labels units of measure to specify what the information is, and the ability to translate from a foreign data structure to a local one.</td>
</tr>
<tr>
<td>5</td>
<td>There is sufficient Control between technologies necessary to establish, manage, and terminate the integration.</td>
<td>IRL 5 simply denotes the ability of one or more of the integrating technologies to control the integration itself, this includes establishing, maintaining, and terminating.</td>
</tr>
<tr>
<td>4</td>
<td>There is sufficient detail in the Quality and Assurance of the integration between technologies.</td>
<td>Many technology integration failures never progress past IRL 3, due to the assumption that if two technologies can exchange information successfully, then they are fully integrated. IRL 4 goes beyond simple data exchange and requires that the data sent is the data received and there exists a mechanism for checking it.</td>
</tr>
<tr>
<td>3</td>
<td>There is Compatibility (i.e. common language) between technologies to orderly and efficiently integrate and interact.</td>
<td>IRL 3 represents the minimum required level to provide successful integration. This means that the two technologies are able to not only influence each other, but also communicate interpretable data. IRL 3 represents the first tangible step in the maturity process.</td>
</tr>
<tr>
<td>2</td>
<td>There is some level of specificity to characterize the interaction (i.e. ability to influence) between technologies through their interface.</td>
<td>Once a medium has been defined, a &quot;signaling&quot; method must be selected such that two integrating technologies are able to influence each other over that medium. Since IRL 2 represents the ability of two technologies to influence each other over a given medium, this represents integration proof-of-concept.</td>
</tr>
<tr>
<td>1</td>
<td>An Interface between technologies has been identified with sufficient detail to allow characterization of the relationship.</td>
<td>This is the lowest level of integration readiness and describes the selection of a medium for integration.</td>
</tr>
</tbody>
</table>
Mathematically, the procedure for calculating the SRL is as follow (assuming \( n \) technologies within the system):

a. Normalize the \([1, 9]\) scale Raw TRLs and IRLs into \((0, 1)\) scale TRLs and IRLs; denote them by matrices:

\[
[TRL]_{n \times 1} = \begin{bmatrix} 
TRI_1 \\
TRI_2 \\
\cdots \\
TRI_n 
\end{bmatrix}_{1 \times 1} \rightarrow \begin{bmatrix} 
TRI_1 \\
TRI_2 \\
\cdots \\
TRI_n 
\end{bmatrix}_{n \times 1}, \quad (1)
\]

\[
[IRL]_{n \times n} = \begin{bmatrix} 
IRL_{11} & IRL_{12} & \cdots & IRL_{1n} \\
IRL_{21} & IRL_{22} & \cdots & IRL_{2n} \\
\cdots & \cdots & \cdots & \cdots \\
IRL_{n1} & IRL_{n2} & \cdots & IRL_{nn} 
\end{bmatrix}_{n \times n} \rightarrow \begin{bmatrix} 
IRL_{11} & IRL_{12} & \cdots & IRL_{1n} \\
IRL_{21} & IRL_{22} & \cdots & IRL_{2n} \\
\cdots & \cdots & \cdots & \cdots \\
IRL_{n1} & IRL_{n2} & \cdots & IRL_{nn} 
\end{bmatrix}_{n \times n}, \quad (2)
\]

where \( IRL_{ij} = IRL_{ji} \). When there is no integration between two technologies, a Raw IRL value of 0 is assigned; for integration of a technology to itself, a Raw IRL value of 9 is used, that is Raw IRL = 9. See Table III for the criteria of assigning IRL of each level.

b. Component SRL matrix is the product of TRL and IRL matrices:

\[
[SRL]_{n \times 1} = [Norm]_{n \times n} \times [IRL]_{n \times n} \times [TRL]_{n \times 1}
\]

\[
[SRL_1] = \begin{bmatrix} 
1/m_1 \\
0 \\
\cdots \\
0 
\end{bmatrix} \times \begin{bmatrix} 
0 \\
1/m_2 \\
\cdots \\
1/m_n 
\end{bmatrix} = \begin{bmatrix} 
0 \\
1/m_2 \\
\cdots \\
1/m_n 
\end{bmatrix}, 
\]

\[
[SRL_2] = \begin{bmatrix} 
0 \\
0 \\
\cdots \\
1/m_n 
\end{bmatrix} \times \begin{bmatrix} 
1/m_1 \\
0 \\
\cdots \\
0 
\end{bmatrix} = \begin{bmatrix} 
0 \\
1/m_2 \\
\cdots \\
1/m_n 
\end{bmatrix}, 
\]

\[
[SRL_n] = \begin{bmatrix} 
0 \\
0 \\
\cdots \\
1/m_n 
\end{bmatrix} \times \begin{bmatrix} 
1/m_1 \\
0 \\
\cdots \\
0 
\end{bmatrix} = \begin{bmatrix} 
0 \\
1/m_2 \\
\cdots \\
1/m_n 
\end{bmatrix}
\]

\[
[IRL_{11}] \times [TRL_1] + IRL_{12} \times [TRL_2] + \cdots + IRL_{1n} \times [TRL_n] \\
[IRL_{21}] \times [TRL_1] + IRL_{22} \times [TRL_2] + \cdots + IRL_{2n} \times [TRL_n] \\
\cdots \\
[IRL_{n1}] \times [TRL_1] + IRL_{n2} \times [TRL_2] + \cdots + IRL_{nn} \times [TRL_n]
\]

\[
= \frac{(IRL_{11} \cdot TRL_1 + IRL_{12} \cdot TRL_2 + \cdots + IRL_{1n} \cdot TRL_n)/m_1}{(IRL_{21} \cdot TRL_1 + IRL_{22} \cdot TRL_2 + \cdots + IRL_{2n} \cdot TRL_n)/m_2} \\
\cdots \\
= \frac{(IRL_{n1} \cdot TRL_1 + IRL_{n2} \cdot TRL_2 + \cdots + IRL_{nn} \cdot TRL_n)/m_n}{(IRL_{n1} \cdot TRL_1 + IRL_{n2} \cdot TRL_2 + \cdots + IRL_{nn} \cdot TRL_n)/m_n}, \quad (3)
\]

where \( m_i \) is the number of integrations of technology \( i \) with itself and all other technologies, and \([Norm] = diag[1/m_1, 1/m_2, \ldots, 1/m_n] \) to normalize the SRL from \((0, m)\) scale to \((0, 1)\) scale for consistency; thus \([Norm] = diag[1/m_1, 1/m_2, \ldots, 1/m_n] \).

c. Composite SRL is the average of all component SRLs:

\[
Composite\ SRL = \frac{SRL_1 + SRL_2 + \cdots + SRL_n}{n} = \frac{1}{n} \sum_{i=1}^{n} SRL_i \quad (4)
\]

See Sauser et al. [2008b] for a more detailed description of how to calculate and apply the SRL.

### 2.2. Current Approaches to Implement Readiness Levels

Just as the ways that agencies or organizations have adopted the TRL metric or created new readiness levels have been diverse, so have the ways that they employ these metrics. Yet, it is necessary to notice that few of them detail the process for the final estimation of readiness levels. For example, the most descriptive is the DoD with its use of TRL. After the DoD began its adoption of the TRL metric, much effort was invested in applying the metric to technologies in ongoing programs and projects. To support this, a series of Technology Readiness Assessment (TRA) Deskbooks [DoD, 2003, 2005b, 2009] have provided the guidance for performing technology maturity assessments prior to incorporating these technologies into systems in defense programs. For example, Figure 1 is a depiction of the process described in the TRA Deskbook [DoD, 2009] showing the major proposed steps to assess the maturity of critical technology elements (CTE1) of an ongoing program in the DoD.

Bilbro [2007, p. 5] states that a technology assessment should be structured within the framework of the product-oriented Work Breakdown Structure (WBS) because the framework “breaks the ‘problem’ down into systems, subsystems and components that can be more accurately assessed, and it provides the results of the assessment in a format that can readily be used in the generation of program costs and schedules.” While this guidance is useful for assessment of technology maturity, it does not provide details for use in real systems. According to a report prepared by Gragettinger et al. [2002] at Carnegie Mellon Software Engineering Institute, approaches for readiness level implementation among agencies are quite broad, which range from a formal software tool to more informal face-to-face discussions between stakeholders.

In a summary, the common approaches to performing readiness level assessments are:

1. Individual estimation: A subject matter expert assesses the maturity of a technology.
2. Group discussion estimation: holding a meeting or conference to discuss the technology maturity.
3. Individual-group estimation: Subject matter experts first perform independent estimations and then discuss these independent estimates in a collective manner to arrive at a consensus of a single estimate.

For each of these methods, guidance such as readiness level descriptions, checklists, and/or assessment tools (e.g., TRL calculator developed by William Nolte at AFRL [Nolte, Ken-
nedy, and Dziegiel, 2003]) has been used individually or in combination.

### 2.3. Limitations of Current Approaches

Despite any rigor to the formulation of an estimate, they still inherit the constraints of the three forms of estimation: educated guess, analogy, and standards [Kerzner, 2009]. Estimators who are given expert opinion or best engineering judgment in the situation when there is not enough knowledge of what is being estimated or not enough time to generate a good estimate perform an educated guess. Analogy means that estimators can do estimation based on comparison and experience with previous work. Though there are some attractive benefits of this method (e.g., relatively easy to execute, can be done on most things, requires limited amounts of data, easily modified), the very weakness of this method is its heavy dependency on the skills of the estimator. Finally, estimation according to standards can be used for relevant estimation if there are standards within the domain developed by some organizations. The key issue with using this method is the requirement of assuring the internal processes in a particular organization against standards, among which there is human subjectivity involved.

Equally, even with the number of processes and methods employed to assess the readiness (or maturity) level, they all attempt to arrive at a unique number. Essentially, all of them rely on the same underlying assumption, that maturity is deterministic. By trying to arrive at consensus, the deterministic result absorbs a lot of subjectivity, which is inherent in human estimation [Bilbro, 2007; DoD, 2003, 2005, 2005b; Graettinger et al., 2002]. Likewise, there are several factors contributing to the subjectivity that impacts the reliability of the estimation. First, ambiguous words used in the readiness level definitions, such as “relevant environment” and “breadboard,” can lead to various interpretations in different applications. Second, the data and additional qualitative information which are usually provided by the engineering managers and/or program office to form the basis of an assessment may not be thorough or in the matching format for performing an assessment.

While using a unique deterministic number for estimation can be relatively easy and potentially adequate to communicate the status of a technology, integration, or system, all stages of the estimation process, including the calculators used, involve subjectivity. There is always uncertainty, the need for judgment, considerable scope for human bias, and inaccuracy [Redmill, 2002a]. Therefore, Galway [2007] suggests a way of using multiple independent experts and Wallsten et al. [1997] suggest combining information from multiple sources prior to taking action for many real-world decisions. However, neither of them suggests a method to combine the inputs from multiple experts or resources. Scholz and Hansmann [2007] point out that a method in which expert judgments are aggregated or integrated is of importance to research while performing risk and decision analysis, and they suggest that risk judgment and uncertainty, in practice, rely on the whole probability distribution of outcomes. Sauser and Ramirez-Marquez [2009] use the assumption of deterministic estimation for TRLs and IRLs in their work, and state that it is more rational to assume that the evaluation of TRLs and IRLs follows a probabilistic form.

While the SRL discussed earlier explores a new way to define readiness/maturity as a function of component TRLs and IRLs, the issue of estimation subjectivity and reliability still remains. Besides assuming the deterministic value for TRL/IRL, the methods proposed in these papers, Magnaye, Sauser, and Ramirez-Marquez [2009], Sauser et al. [2008a, 2008b, 2006], and Sauser and Ramirez-Marquez [2009], just provided a unique value for the SRL of the system in consideration and were unable to elaborate the reliability of the estimation result. One of the main objectives to estimate the maturity status of a system is to provide decision-makers with sound system maturity data. Since more reliable data contributes to better decisions, it is better to place a confidence level on the information about the estimated values [Redmill, 2002a].

To address these limitations, this paper proposes a probabilistic approach to system readiness assessment, which will be described and illustrated in following sections.

### 3. METHODOLOGY FOR SRL CONFIDENCE

#### INTERVAL ESTIMATION

Although the real status of an object (technology, integration, etc.) to be estimated can never be known, we believe the estimate from subject matter experts will mostly approximate the real value. However, it emerges as an issue when different experts have different estimates on the same technology or integration. Therefore, to overcome the current shortcomings of SRL computation, this paper develops a new SRL estimation approach based on the assumption that the evaluation of the TRL/IRL follows a probabilistic form. Based on this assumption the new approach incorporates information provided by system evaluators by using the relative frequency of the TRL/IRL values generated as a probability distribution to combine every evaluator’s judgment of the technology/integration maturity. That is, the dispersion in the TRL and IRL estimates can be represented in the values that are calculated. Information that shows the degree of precision that accompanies a SRL estimate can be useful to a decision-maker when determining the risk-return tradeoffs. Based on these assumptions, a Monte-Carlo simulation approach is applied to yield an estimation of the probability distribution of the SRL. Note
that while the methods that the evaluators use for TRL/IRL estimation can be various, the evaluators must be properly schooled in the use of TRL/IRL and be encouraged to follow the procedures mentioned in the previous section.

In essence, the approach provides a new SRL confidence calculation by replacing the original deterministic values for TRL and IRL assessment with data obtained from different system evaluators and using such data to generate a discrete probability distribution by calculating the relative frequencies for each TRL and IRL of the system of interest. It is important to mention that these probability distributions represent the views of the different system evaluators regarding the maturity of the different technologies and integrations in the system. Thus, the value of SRL is no longer deterministic. Furthermore, the values of SRL will vary depending on the specific evaluator.

However, engineering managers cannot be expected to make assessment about system maturity based on potential conflicting information, such as multiple SRL values. Thus, one could provide a range of likely values (i.e., an interval) regarding the SRL that a manager could be more comfortable working with. To do this, a Monte-Carlo simulation approach has been developed to calculate both component SRLs and a system’s composite SRL based on the probability distributions obtained for each of the TRLs and IRLs. Based on this simulation approach, a SRL data set is obtained describing the probability distribution of the system’s SRL. Furthermore, percentile intervals are calculated for both component SRLs and the system’s composite SRL in order to provide the management team with statistical information. Contrary to the deterministic SRL, these percentile intervals can provide engineering managers statistical knowledge about the current maturity status of the system and more flexibility for further decision-making. Besides the overall information, the [5%, 95%] percentile interval bound is graphically presented here to show the relationship between the composite SRL and component SRLs, which is beneficial for leveraging resources for the system development.

As illustrated in Figure 2, the application of the proposed approach encompasses three steps: assignment of frequency distributions, Monte-Carlo simulation, and the calculation of SRL intervals. Each step is described in the following sections.

### 3.1. TRL and IRL Frequency Distributions Assignment

As specified before, the most common methods to perform estimations are human-intensive and subjective. The resulting TRL and IRL values are the products of the tradeoffs among several contractors or stakeholders. Since different opinions always exist among individuals, it is inevitable that variations will exist in the estimates produced by the evaluators. These variations can provide insight into the real maturity levels of the system’s technologies and integrations, and thus can benefit engineering managers by accounting for uncertainty during the technology acquisition process based on their tolerance for risk. Currently, SRL computation [Magneye, Sauser, and Ramirez-Marquez, 2009; Sauser et al., 2008a, 2006; Sauser and Ramirez-Marquez, 2009] assumes discrete numbers (average, median, mode, etc.) as the actual estimate for TRL/IRL. Such an assumption can miss inherent information on variation and thus potentially provide an unlikely estimate. The new proposed method addresses such issue by considering different TRL/IRL levels that are estimated from the evaluator’s data. This first step calculates the relative frequencies for each TRL/IRL and describes the probability distributions of the system’s technology and integration elements.

Let $s$ denote the number of evaluators to examine a system with $n$ components. Each technology and potential integration between any two technologies may be estimated with several readiness level numbers which fall in the range of $[0, 9]$. Let $f_{ij}$ indicate the number of evaluators assigning a readiness level of $k$ to technology $i$, and $f_{ijk}$ the number of evaluators assigning a readiness level of $k$ to the integration between technology $i$ and $j$. To obtain the probability distribution, calculate the relative frequency for each TRL/IRL as

$$p_{i,k} = f_{i,k}/s,$$

$$p_{ij,k} = f_{ij,k}/s,$$

where $i, j = 1, \ldots, n; k = 0, 1, \ldots, 9$.

To clarify, consider a simple system with three technologies and two integrations as represented in Figure 3. It is assumed that the maturity of the system’s technologies and integrations have been analyzed by 20 different evaluators with results as described in Table IV.

Because this is the full information about the technologies and integrations of the system, it can therefore be considered as a good representative of the component status, and it is suitable to assume this to be the input data for further SRL estimation. The frequency distribution of the TRL/IRL is shown in Table V.

![Figure 3. System with three technologies (1, 2, and 3) and two integrations.](Image)
3.2. Monte-Carlo Simulation Approach for SRL Calculation

Based on the components frequency distributions described previously, a Monte-Carlo (MC) simulation approach is then implemented to generate a set of SRL values. MC simulation relies on repeated random sampling to analyze problems with uncertainty. Briefly, there are three steps to implement MC simulation:

1. Define a domain of possible inputs.
2. Generate inputs randomly from the domain, and perform a deterministic computation on them.
3. Aggregate the results of the individual computations into the final result.

The first step to implement MC simulation for SRL estimation is to determine the domain of possible inputs of TRLs and IRLs. Based on the TRL/IRL probability distributions obtained in Step 1, MC simulation is implemented to generate readiness levels for TRL/IRL. Then, for each MC simulation run, component SRLs and the system’s composite SRL are calculated via the SRL formula. This process is repeated until a predetermined number of iterations, L, is reached. Moreover, in each run the corresponding component SRLs and system composite SRL are stored in data sets. The simulation runs are completed, the data sets are then used for statistical analysis.

In order to make the procedure simple and clear, it is first necessary to introduce the following transformation:

\[ x_{ik} = \begin{cases} 1, & \text{if } TRL_i = k \\ 0, & \text{otherwise} \end{cases}, \quad x_{jk} = \begin{cases} 1, & \text{if } IRL_{ij} = k \\ 0, & \text{otherwise} \end{cases} \]

where \( i, j = 1, 2, \ldots, n; k = 0, 1, 2, \ldots, 9 \).

Based on these binary variables, each of the possible normalized TRLs and IRLs within the system can be obtained as:

\[ SRL_i = \frac{1}{m_i} \left[ \frac{1}{9} \sum_{k=0}^{9} k x_{ik} \right] \]

\[ IRL_{ij} = \frac{1}{9} \sum_{k=0}^{9} k x_{jk} \]

where \( m_i \) is the total number of data sets for component \( i \), \( D_k \) is the probability of \( TRL_i \) assuming level \( k \).

Meanwhile, let us define:

\[ p_{ik} \] the probability of \( TRL_i \) assuming level \( k \).

\[ p_{ijk} \] the probability of \( IRL_{ij} \) assuming level \( k \).

\( D_1, D_2, \ldots, D_n \) and \( D \) are the storage data sets for component \( SRL_1, SRL_2, \ldots, SRL_n \) and \( SRL \) from the MC simulation.

After 10,000 iterations, the simulation results for all component SRLs and system composite SRL are stored in data sets \( D_1, D_2, D_3 \) and \( D \).

3.3. Simulation Results Analysis

Once the \( L \) simulation runs are completed, the data sets \( D_1, D_2, \ldots, D_n \) and \( D \) can be analyzed to provide representative figures of merit for point estimates and uncertainty of the corresponding SRL, such as the mean and the standard deviation, respectively. For further insight about the behavior of the data sets, the method herein calculates percentiles for the data sets. The \( p \)th (here we calculate the 0th, 5th, 10th, \ldots, 95th, 100th) percentile is a value such that at least \( p \) percent of the simulation data are less than or equal to this value and at least \((100 - p)\%\) of the simulation data are greater than or equal to this value. Steps to calculate the \( p \)th percentile are as follows:
1. Arrange the data in ascending order (smallest value to largest value).

2. Compute an index \( c = L \times (p/100) \), where \( p \) is the percentile of interest and \( L \) is the number of data attained from simulation.

3. If \( c \) is not an integer, round up and the next integer greater than \( c \) denotes the position of the \( p \)th percentile; if \( c \) is an integer, the \( p \)th percentile is the average of the values in positions \( c \) and \( c + 1 \).

To graphically describe the behavior of the composite SRL based on the input of TRL/IRL frequency distributions, the frequency chart and cumulative frequency chart can be developed according to data set \( D \). Similarly, one can obtain \( \alpha \)-level confidence intervals around the SRL estimate. For instance, an \( \alpha = 10\% \) confidence interval around the composite SRL would indicate with 90\% confidence a likely value for system maturity.

For the system in Figure 3, Table VII shows the SRLs’ statistics, and Table VIII shows the complete information of SRL percentiles. The mean of composite SRL from 10,000 iterations is 0.55, and the \( \alpha = 10\% \) confidence interval is \([0.50, 0.60]\).

4. ILLUSTRATIVE EXAMPLE

The following example is executed to obtain a maturity assessment of an actual system using the proposed approach. This case example was selected because it has been studied in multiple sources using the readiness level metrics presented in this paper [Gove, 2007; Sauser et al., 2008a; Sauser and Ramirez-Marquez, 2009]. While the system is real, the data used in the evaluation are notional.

4.1. Case Overview

After many years of exceptional service, NASA had been considering a technical solution to repair the Hubble Space Telescope (HST) as it had far surpassed its expected lifetime. Its gyroscopes were approaching the end of their lifecycle, its batteries were nearing their lower levels of acceptable per-

---

**Table VI. Example 1 Parameters**

<table>
<thead>
<tr>
<th>( n )</th>
<th>( L )</th>
<th>( m_1 )</th>
<th>( m_2 )</th>
<th>( m_3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>10,000</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

Note: \( m_i \) is the number of integrations of technology \( i \) including its integration to itself.

**Table VII. Example 1 SRL Statistics**

<table>
<thead>
<tr>
<th></th>
<th>SRL</th>
<th>SRL1</th>
<th>SRL2</th>
<th>SRL3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.548</td>
<td>0.514</td>
<td>0.483</td>
<td>0.648</td>
</tr>
<tr>
<td>St dev</td>
<td>0.030</td>
<td>0.030</td>
<td>0.035</td>
<td>0.046</td>
</tr>
<tr>
<td>Minimum</td>
<td>0.453</td>
<td>0.420</td>
<td>0.366</td>
<td>0.519</td>
</tr>
<tr>
<td>Maximum</td>
<td>0.662</td>
<td>0.630</td>
<td>0.621</td>
<td>0.790</td>
</tr>
<tr>
<td>5%</td>
<td>0.501</td>
<td>0.463</td>
<td>0.424</td>
<td>0.574</td>
</tr>
<tr>
<td>95%</td>
<td>0.604</td>
<td>0.574</td>
<td>0.551</td>
<td>0.736</td>
</tr>
</tbody>
</table>

**Table VIII. Example 1 SRL Percentiles**

<table>
<thead>
<tr>
<th>Percentiles</th>
<th>SRL</th>
<th>SRL1</th>
<th>SRL2</th>
<th>SRL3</th>
<th>Percentiles</th>
<th>SRL</th>
<th>SRL1</th>
<th>SRL2</th>
<th>SRL3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0%</td>
<td>0.453</td>
<td>0.420</td>
<td>0.366</td>
<td>0.519</td>
<td>55%</td>
<td>0.552</td>
<td>0.519</td>
<td>0.490</td>
<td>0.648</td>
</tr>
<tr>
<td>5%</td>
<td>0.501</td>
<td>0.463</td>
<td>0.424</td>
<td>0.574</td>
<td>60%</td>
<td>0.552</td>
<td>0.519</td>
<td>0.490</td>
<td>0.648</td>
</tr>
<tr>
<td>10%</td>
<td>0.507</td>
<td>0.475</td>
<td>0.436</td>
<td>0.593</td>
<td>65%</td>
<td>0.557</td>
<td>0.519</td>
<td>0.490</td>
<td>0.660</td>
</tr>
<tr>
<td>15%</td>
<td>0.521</td>
<td>0.481</td>
<td>0.453</td>
<td>0.605</td>
<td>70%</td>
<td>0.558</td>
<td>0.519</td>
<td>0.494</td>
<td>0.667</td>
</tr>
<tr>
<td>20%</td>
<td>0.524</td>
<td>0.481</td>
<td>0.453</td>
<td>0.611</td>
<td>75%</td>
<td>0.573</td>
<td>0.519</td>
<td>0.510</td>
<td>0.685</td>
</tr>
<tr>
<td>25%</td>
<td>0.529</td>
<td>0.506</td>
<td>0.481</td>
<td>0.611</td>
<td>80%</td>
<td>0.578</td>
<td>0.531</td>
<td>0.519</td>
<td>0.691</td>
</tr>
<tr>
<td>30%</td>
<td>0.530</td>
<td>0.506</td>
<td>0.481</td>
<td>0.611</td>
<td>85%</td>
<td>0.598</td>
<td>0.531</td>
<td>0.519</td>
<td>0.704</td>
</tr>
<tr>
<td>35%</td>
<td>0.531</td>
<td>0.519</td>
<td>0.461</td>
<td>0.636</td>
<td>90%</td>
<td>0.583</td>
<td>0.556</td>
<td>0.523</td>
<td>0.704</td>
</tr>
<tr>
<td>40%</td>
<td>0.543</td>
<td>0.519</td>
<td>0.481</td>
<td>0.648</td>
<td>95%</td>
<td>0.604</td>
<td>0.574</td>
<td>0.551</td>
<td>0.735</td>
</tr>
<tr>
<td>45%</td>
<td>0.551</td>
<td>0.519</td>
<td>0.486</td>
<td>0.668</td>
<td>100%</td>
<td>0.662</td>
<td>0.630</td>
<td>0.621</td>
<td>0.790</td>
</tr>
<tr>
<td>50%</td>
<td>0.552</td>
<td>0.519</td>
<td>0.486</td>
<td>0.648</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

Figure 4. System composite SRL probability density function. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Figure 5. System composite SRL cumulative distribution function. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]
formance, and the fine-guidance sensors had already begun failing. If HST was not serviced soon and the batteries ran out or navigational and attitude control was lost, certain instruments onboard could have been permanently damaged either due to low temperatures or direct exposure to sunlight. In order to repair HST, NASA had to perform a Service Mission 4 (SM-4) to keep HST operating into the future. The problem was that HST had not been serviced since before the Columbia disaster and the Columbia Accident Investigation Board (CAIB) requirements for shuttle operations would impact a HST SM-4 since time and fuel considerations might not allow the crew to safely dock to the International Space Station (ISS) if damage was found on the heat shield. To combat this problem, a robotic servicing mission (RSM) had been suggested, thereby reducing cost and the risk to human life [King, 2004; Lanzerotti, 2005; Mattice, 2005]. Figure 6 represents a subsystem concept diagram of the RSM.

4.2. New Methodology for Maturity Assessment

In order to gain information about the maturity status of this system, a board of relative experts (here we assume the population is 100) is set to evaluate its current TRLs and IRLs. Table IX shows the evaluation details with the notional data. Using the probability distribution assignment method proposed in the last section, Table X presents the TRL/IRL probability distributions of this system.

Setting $L = 50,000$, Table XI lists the parameters for simulation.

For each iteration of the Monte-Carlo simulation, $SRL_1$, $SRL_2, \ldots, SRL_6$ and composite SRL are stored in the data sets of $D_1, D_2, \ldots, D_6$ and $D$, respectively. After the simulation, 50,000 simulation data are obtained and stored in each data set. Figures 7 and 8 show us the probability density distribution and cumulative distribution function of the composite SRL.

Based on the simulation data, further analysis provides us with the statistics and the percentile details of all SRLs (see Tables XII and XIII).

From the simulation, the mean for the composite SRL is 0.51 with a standard deviation of 0.02. Since it is calculated from significant number of simulation iterations, the mean is sufficient to represent the maturity status of the system. The standard deviation gives us a brief notion about the dispersion of the SRL over its whole possible range. For this particular case with the inputs of TRL/IRL probability distribution, the possible minimum and maximum composite SRL are 0.44 and 0.59. The 90% confident interval for composite SRL is [0.48, 0.54], which indicates 90% of the time the composite SRL will fall in this interval.

4.3. Implications to Engineering Management

To understand developmental maturity, a framework for describing the maturity of a system comparative to its lifecycle is needed. Sauser et al. [2006a] performed a mapping of the SRL index to four system’s lifecycles. They later refined their results with the analysis of additional systems from DoD, Lockheed-Martin, NASA, and Northrop Grumman for which random maturity assessments were created to assess the sensitivity of the SRL index to a system’s lifecycle. This analysis created a recalibration of the mapping initially introduced in Sauser et al. [2006] and refined in Sauser et al. [2008]. Table XIV represents our calibration of the SRL index to the current DoD Acquisition Management System as described in DoD Instruction 5000.02 [DoD, 2005a]. We will use this mapping to serve as a tool for explaining the implications our work to systems engineering.

From the simulation result, the point estimate (mean) for RSM composite SRL is 0.51, and the 90% confidence percentile interval is [0.48, 0.54], which indicates the current RSM system is at some point between the technology development and engineering and manufacturing development phases according to Table XIV. As stated, the main objective during this phase is to reduce technology risks and determine appropriate set of technologies to integrate into a full system.
This is typically a program initiation point for systems development.

While the composite SRL is important, it is also critical to understand the interdependencies of the component SRLs to the composite. Therefore, Figure 9 plots the 90% confidence interval of the system composite SRL and all component SRLs from the simulation. Briefly, the 90% confidence intervals of components SRL
\(_1\), SRL
\(_2\), and SRL
\(_3\) are higher than the composite SRL, indicating these technologies are ahead in the progress of maturity, while those of components SRL
\(_4\), SRL
\(_5\), and SRL
\(_6\) are somewhat lower than the composite SRL, indicating that the maturity of these technologies and their integrations are behind the whole system development. While the previous work in SRL may have concluded this, our approach would indicate that components SRL
\(_2\), SRL
\(_4\), and SRL
\(_6\) would be considered to be in the same phase of development as the composite SRL since there is an overlap based on the 90% confidence interval. Recall the assumption of

Table X. RSM TRL/IRL Frequency Distribution

<table>
<thead>
<tr>
<th>Technology</th>
<th>TRL</th>
<th>Probability (p)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>8</td>
<td>0.15</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>9</td>
<td>0.05</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>6</td>
<td>0.10</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>7</td>
<td>0.15</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>7</td>
<td>0.25</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>7</td>
<td>0.20</td>
</tr>
</tbody>
</table>

Table XI. System Parameters

<table>
<thead>
<tr>
<th>n</th>
<th>L</th>
<th>m_1</th>
<th>m_2</th>
<th>m_3</th>
<th>m_4</th>
<th>m_5</th>
<th>m_6</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>50,000</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>

Table XII. RSM SRL Statistics

<table>
<thead>
<tr>
<th>SRL</th>
<th>SRL1</th>
<th>SRL2</th>
<th>SRL3</th>
<th>SRL4</th>
<th>SRL5</th>
<th>SRL6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.507</td>
<td>0.639</td>
<td>0.577</td>
<td>0.597</td>
<td>0.435</td>
<td>0.374</td>
</tr>
<tr>
<td>St dev</td>
<td>0.018</td>
<td>0.032</td>
<td>0.026</td>
<td>0.027</td>
<td>0.029</td>
<td>0.025</td>
</tr>
<tr>
<td>Minimum</td>
<td>0.435</td>
<td>0.514</td>
<td>0.448</td>
<td>0.475</td>
<td>0.325</td>
<td>0.278</td>
</tr>
<tr>
<td>Maximum</td>
<td>0.587</td>
<td>0.786</td>
<td>0.682</td>
<td>0.719</td>
<td>0.556</td>
<td>0.488</td>
</tr>
<tr>
<td>5%</td>
<td>0.478</td>
<td>0.584</td>
<td>0.534</td>
<td>0.556</td>
<td>0.391</td>
<td>0.335</td>
</tr>
<tr>
<td>95%</td>
<td>0.537</td>
<td>0.691</td>
<td>0.620</td>
<td>0.645</td>
<td>0.481</td>
<td>0.417</td>
</tr>
</tbody>
</table>

Table XIII. RSM SRL Percentiles

<table>
<thead>
<tr>
<th>Percentile</th>
<th>SRL</th>
<th>SRL1</th>
<th>SRL2</th>
<th>SRL3</th>
<th>SRL4</th>
<th>SRL5</th>
<th>SRL6</th>
</tr>
</thead>
<tbody>
<tr>
<td>0%</td>
<td>0.435</td>
<td>0.514</td>
<td>0.448</td>
<td>0.475</td>
<td>0.325</td>
<td>0.278</td>
<td>0.340</td>
</tr>
<tr>
<td>5%</td>
<td>0.478</td>
<td>0.584</td>
<td>0.534</td>
<td>0.556</td>
<td>0.391</td>
<td>0.335</td>
<td>0.352</td>
</tr>
<tr>
<td>10%</td>
<td>0.484</td>
<td>0.597</td>
<td>0.546</td>
<td>0.559</td>
<td>0.399</td>
<td>0.343</td>
<td>0.370</td>
</tr>
<tr>
<td>15%</td>
<td>0.498</td>
<td>0.605</td>
<td>0.549</td>
<td>0.568</td>
<td>0.407</td>
<td>0.349</td>
<td>0.389</td>
</tr>
<tr>
<td>20%</td>
<td>0.502</td>
<td>0.609</td>
<td>0.552</td>
<td>0.577</td>
<td>0.412</td>
<td>0.352</td>
<td>0.395</td>
</tr>
<tr>
<td>25%</td>
<td>0.505</td>
<td>0.617</td>
<td>0.562</td>
<td>0.577</td>
<td>0.416</td>
<td>0.358</td>
<td>0.395</td>
</tr>
<tr>
<td>30%</td>
<td>0.508</td>
<td>0.630</td>
<td>0.568</td>
<td>0.583</td>
<td>0.428</td>
<td>0.361</td>
<td>0.407</td>
</tr>
<tr>
<td>35%</td>
<td>0.508</td>
<td>0.634</td>
<td>0.568</td>
<td>0.583</td>
<td>0.428</td>
<td>0.364</td>
<td>0.407</td>
</tr>
<tr>
<td>40%</td>
<td>0.508</td>
<td>0.634</td>
<td>0.568</td>
<td>0.583</td>
<td>0.432</td>
<td>0.370</td>
<td>0.407</td>
</tr>
<tr>
<td>45%</td>
<td>0.508</td>
<td>0.634</td>
<td>0.571</td>
<td>0.596</td>
<td>0.436</td>
<td>0.370</td>
<td>0.407</td>
</tr>
<tr>
<td>50%</td>
<td>0.508</td>
<td>0.634</td>
<td>0.574</td>
<td>0.602</td>
<td>0.436</td>
<td>0.370</td>
<td>0.426</td>
</tr>
<tr>
<td>55%</td>
<td>0.508</td>
<td>0.638</td>
<td>0.577</td>
<td>0.602</td>
<td>0.436</td>
<td>0.380</td>
<td>0.426</td>
</tr>
<tr>
<td>60%</td>
<td>0.512</td>
<td>0.642</td>
<td>0.586</td>
<td>0.602</td>
<td>0.436</td>
<td>0.383</td>
<td>0.444</td>
</tr>
<tr>
<td>65%</td>
<td>0.514</td>
<td>0.654</td>
<td>0.590</td>
<td>0.602</td>
<td>0.444</td>
<td>0.386</td>
<td>0.444</td>
</tr>
<tr>
<td>70%</td>
<td>0.517</td>
<td>0.668</td>
<td>0.593</td>
<td>0.608</td>
<td>0.449</td>
<td>0.389</td>
<td>0.444</td>
</tr>
<tr>
<td>75%</td>
<td>0.519</td>
<td>0.667</td>
<td>0.593</td>
<td>0.617</td>
<td>0.457</td>
<td>0.389</td>
<td>0.444</td>
</tr>
<tr>
<td>80%</td>
<td>0.522</td>
<td>0.667</td>
<td>0.596</td>
<td>0.620</td>
<td>0.461</td>
<td>0.395</td>
<td>0.444</td>
</tr>
<tr>
<td>85%</td>
<td>0.526</td>
<td>0.667</td>
<td>0.602</td>
<td>0.627</td>
<td>0.461</td>
<td>0.398</td>
<td>0.444</td>
</tr>
<tr>
<td>90%</td>
<td>0.530</td>
<td>0.679</td>
<td>0.611</td>
<td>0.630</td>
<td>0.469</td>
<td>0.407</td>
<td>0.463</td>
</tr>
<tr>
<td>95%</td>
<td>0.537</td>
<td>0.691</td>
<td>0.620</td>
<td>0.645</td>
<td>0.481</td>
<td>0.417</td>
<td>0.481</td>
</tr>
<tr>
<td>100%</td>
<td>0.587</td>
<td>0.786</td>
<td>0.682</td>
<td>0.719</td>
<td>0.556</td>
<td>0.488</td>
<td>0.519</td>
</tr>
</tbody>
</table>
TRLs and IRLs; these statuses do match with the assumption, that is, those technologies that assume relatively lower TRLs and IRLs result in lower component SRL in the simulation, and vice versa. Sauser and Ramirez-Marquez [2009] emphasized the importance of resource allocation for engineering management and proposed a model with an evolutionary algorithm aiming to identify an optimal solution for resource allocation. The information from Figure 9 can also benefit engineering manager insights for resource allocation from another perspective. Besides providing a more convincing maturity estimation about the HST RSM system, information from Figure 9 can provide us the details about both the system and all the components. For example, if the next step is to leverage the system development, then prioritized resource may first flow to the development of component SRL4 because it is the least mature. Another approach would be to evaluate the length of these bars which indicate the 90% dispersion range of the associated SRLs, in which a wider range is a representation of a higher variation within such a SRL. In the HST RSM system, the length of component SRL1 and SRL5 are relatively larger, indicating there is more uncertainty in the maturity assessment of these two technologies and their integrations. It may be possible that resources need to be allocated to reduce this uncertainty, or even investigate the validity of the assessment.

Furthermore, the methodology proposed in this paper provides a systems maturity assessment along with estimation of the variation, which is very important for engineering managers to assess a systems’ associated risk. For example, engineering managers can make more informed decisions, based on their personal or the organization’s tolerance for risk, if the SRL values that they are evaluating can reflect the dispersion in the TRL and IRL values produced by all stakeholders. That is, a risk-adverse decision-maker may tend to use the lower-bound estimates of SRL while a risk-tolerant manager may favor the upper-bound values. As an example, some scenarios are summarized in Table XV.

When the complexity or the uncertainty or both of a project rise, the difficulty in assessing their maturity increases, and so can the degree of disagreement on the assessment. Thus, although the variation in the evaluation of legacy technologies used in a system may not be significant (i.e., TRLs and IRLs assumption for the illustrative example), it is probably significant in systems of greater complexity within which there are emerging technologies with high uncertainty. Since the main objective of the development of SRL metrics is to add value to maturity assessment of systems, system maturity estimation with confidence level will provide much more insights in such situations.

**Table XIV. SRL to Acquisition Management System Lifecycle**

<table>
<thead>
<tr>
<th>SRL</th>
<th>Acquisition Phase</th>
<th>Definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.90 to 1.00</td>
<td>Operations &amp; Support</td>
<td>Execute a support program that meets operational support performance requirements and sustains the system in the most cost-effective manner over its total lifecycle.</td>
</tr>
<tr>
<td>0.80 to 0.90</td>
<td>Production &amp; Deployment</td>
<td>Achieve operational capability that satisfies mission needs.</td>
</tr>
<tr>
<td>0.50 to 0.80</td>
<td>Engineering and Manufacturing Development</td>
<td>Develop system capability or (increments thereof); reduce integration and manufacturing risk; ensure operational supportability; reduce logistics footprint; implement human systems integration; design for production; ensure affordability and protection of critical program information; and demonstrate system integration, interoperability, safety and utility.</td>
</tr>
<tr>
<td>0.20 to 0.50</td>
<td>Technology Development</td>
<td>Reduce technology risks and determine appropriate set of technologies to integrate into a full system.</td>
</tr>
<tr>
<td>0.10 to 0.20</td>
<td>Materiel Solution Refinement</td>
<td>Assess potential materiel solutions and satisfy the phase-specific entrance criteria for the next program milestone.</td>
</tr>
</tbody>
</table>
Table XV. Potential Engineering Management Tolerances Based on Risk Conditions

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>CONDITION</th>
<th>SRL VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LOWER</td>
<td>UPPER</td>
</tr>
<tr>
<td>RISK ATTITUDE</td>
<td>Averse</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Tolerant</td>
<td></td>
</tr>
<tr>
<td>CONSEQUENCES</td>
<td>High</td>
<td>X</td>
</tr>
<tr>
<td>OF FAILURE</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>PAYMENT</td>
<td>Owner</td>
<td>X</td>
</tr>
<tr>
<td>DECISION</td>
<td>Contractor</td>
<td></td>
</tr>
<tr>
<td>STRATEGY</td>
<td>Conservative</td>
<td>X</td>
</tr>
</tbody>
</table>

Note: an ‘X’ represents preferred choice

5. CONCLUSION

The use of prescriptive metrics in engineering management has been a proven and successful practice for many organizations. As a benchmark for making decisions, metrics can perform an integral part of management activities for indicating performance or effectiveness of risk, quality, and maturity; identifying critical parameters; establishing milestones to assess progress; providing direction for risk management/mitigation; and sustaining entry and exit criteria for major milestones. With the widespread use of readiness level metrics, it becomes imperative that one finds more rigor in how these metrics are used to make a strategic planning decision in systems development. This paper reviewed various readiness level metrics and some of the approaches to applying them for estimations. The need to decrease the impact of human subjectivity on estimation and to provide a more reliable estimation emerged by examining these approaches. Focusing on the SRL metric, which incorporates TRL and IRL, we proposed a probabilistic methodology for TRL/IRL estimation and a Monte-Carlo simulation approach for an SRL confidence estimation. The probabilistic methodology combines the thorough estimation from all estimators to mitigate the influence of subjectivity. The Monte-Carlo simulation approach presented here achieves an estimation with confidence intervals for all component SRLs and the system’s composite SRL. While the point estimate (mean) of SRL provides engineering managers a basic assessment of the readiness level of the system of interest, the interval estimate of SRL provides them the degree to which they can be confident about the system readiness status. Moreover, the methodology has the potential to leverage the development of the whole system by comparing the component SRLs with the composite SRL.

While we believe the approach and methodology presented in this paper is a step in the right direction to decrease the impact of subjectivity on estimation and to provide a more reliable estimation with confidence intervals, it is not without its limitations. Thus, we conclude with some of these limitations and opportunities for future research:

First, the underlying assumption in using a probabilistic approach to combine all evaluators’ estimation of TRLs and IRLs is that all the evaluators involved have effectively performed their estimation; however, it is not always the fact in practice since some evaluators may not perform their estimation accurately due to limited time or information. As a result, the methodology needs further work to take this into consideration.

Second, the other assumption is that every estimator only provides a unique number for his/her estimation on one technology or one integration point. However, with incomplete or ambiguous data, even one estimator may be unsure about the maturity of what is being estimated, and he/she may guess several numbers with likelihood that the estimation from one estimator is already a probability distribution. In such a situation, the methodology needs to be adjusted to take into account combining this type of estimation and the type we assumed in this paper.

Third, further study needs to be done on how to determine the number of iterations to run the Monte-Carlo simulation. It would be beneficial to shorten the simulation iterations because running more iterations requires more time and resources. However, we need a large number of iterations in order to get a steady SRL estimate and minimize the SRL distortion that can result from inadequate iterations. Therefore, tradeoffs must be made under this dilemma. The authors suggest monitoring the dynamic frequency distribution during the simulation process; the process can be terminated when the relative frequency distribution is stable or the time to run the simulation has reached its allowable limit.

Finally, the illustrative example in the paper is to show how to use the methodology, but the data are notional. For the methodology to be useful in reality, we need to further investigate it with real data from real systems. Though this is somewhat difficult to carry out, the authors do plan to devote future effort to this issue.

REFERENCES


Systems Engineering DOI 10.1002/sys
R. Magnaye, B. Sauser, and J. Ramirez-Marquez, Using a system maturity scale to monitor and evaluate the development of systems, 6th Annu Acquisition Symp, Monterey, CA, 2009, pp. 95–112.
Weiping Tan received his B.E. with first class honor in 2006 in Automation from Beijing Institute of Technology, and received his M.E. in Engineering Management from Stevens Institute of Technology in 2009. He is currently pursuing his Ph.D. in Engineering Management at Stevens Institute of Technology in the School of Systems and Enterprises. His research interest is system maturity assessment. He was elected to Epsilon Mu Eta—the Engineering Management Honor Society, is the Vice-President for the INCOSE Stevens Student Chapter, and received the Brian Mar Best Student Paper Award at the 2009 INCOSE International Symposium.

Jose Emmanuel Ramirez-Marquez is an Assistant Professor of the School of Systems and Enterprises at Stevens Institute of Technology. A former Fulbright Scholar, he holds degrees from Rutgers University in Industrial Engineering (Ph.D. and M.S.) and Statistics (M.S.) and from Universidad Nacional Autonoma de Mexico in Actuarial Science. His research efforts are currently focused on the reliability analysis and optimization of complex systems, the development of mathematical models for sensor network operational effectiveness, and the development of evolutionary optimization algorithms. In these areas, Dr. Ramirez-Marquez has conducted funded research for both private industry and government. Also, he has published more than 50 refereed manuscripts related to these areas in technical journals, book chapters, conference proceedings, and industry reports. Dr. Ramirez-Marquez has presented his research findings both nationally and internationally in conferences such as INFORMS, IERC, ARSym, and ESREL. He is an Associate Editor for the International Journal of Performability Engineering, currently serves as director of the QCRE division board of the IIE, and is a member of the Technical Committee on System Reliability for ESRA.

Brian J. Sauser is an Assistant Professor in the School of Systems and Enterprises at Stevens Institute of Technology. He holds a B.S. from Texas A&M University in Agricultural Development with an emphasis in Horticulture Technology, an M.S. from Rutgers, The State University of New Jersey in Bioresource Engineering, and a Ph.D. from Stevens Institute of Technology in Project Management. Before joining Stevens, he spent more than 12 years working in government, industry, and academia both as a researcher/engineer and director of programs. His research interests are in theories, tools, and methods for bridging the gap between systems engineering and project management for managing complex systems. This includes the advancement of systems theory in the pursuit of a biology of systems, system and enterprise maturity assessment for system and enterprise management, and systems engineering capability assessment. His work on system maturity assessment has been adopted as a decision support tool by organizations within NASA, U.S. Army, U.S. Navy, Northrop Grumman, and Lockheed Martin (http://www.SystemReadinessLevel.com).