

It's All Rocket Science: On the Equivalence of Development Timelines for Aerospace and Nuclear Technologies

Maxime Crépin
Arts et Métiers ParisTech
maxime.crepin@ensam.eu

Bernard El-Khoury
Massachusetts Institute of Technology
bkhoury@mit.edu

C. Robert Kenley
Massachusetts Institute of Technology
kenley@mit.edu

Copyright © 2012 by Maxime Crépin and Massachusetts Institute of Technology. Published and used by INCOSE with permission.

Abstract. Early in the lifecycle of a system development, systems engineers must execute trade studies to allocate resources between different research and development efforts that are developing technologies to be deployed into the system, and they must prepare risk management plans for the selected technologies. We have been developing a statistical model for schedule and cost uncertainty based on a small number of inputs that are quite objective and are already integrated with technology readiness assessment. An algorithm that transforms Technical Maturity (TM) scores from Department of Energy projects into a Technology Readiness Level (TRL) score was created, allowing us to add data from a US Department of Energy to an existing set of data from NASA. We statistically tested whether the two samples (i.e. the DoE and NASA datasets) were randomly drawn from the same population and concluded that the transition times for developing aerospace and nuclear technologies are very similar.

Introduction

The research is part of the TRL-Based Modeling for Cost and Schedule Uncertainty project, which is being executed by the Massachusetts Institute of Technology under sponsorship the US Naval Postgraduate School. The project's goal is to provide a simple, yet powerful, forecasting method that is applicable to all acquisitions of advanced technology in DoD and other US federal agencies. It aims to develop a statistical model for schedule and cost uncertainty based on a small number of inputs that are quite objective and already integrated with technology readiness assessment, making them more likely to be used by program managers. The model would help allocate resources between different technologies, and better manage the cost and schedule risk of the developing technologies.

We selected Technology Readiness Level (TRL) as the principal model input parameter, because it is commonly used for technology readiness assessment. TRL is a discrete scale used by major US agencies, such as NASA, DoD, and DoE, and many of the world's major technology-intensive companies to assess the maturity of technologies (materials, components, devices, etc.) that are being developed for incorporation into a system. A low TRL (1-2) indicates a technology that is still at a basic research level, while a high TRL (8-9) indicates a technology has been incorporated into a system that is in use. Table 1 indicates the formal definitions of the TRL levels used by DoE (US Department of Energy, 2009) and NASA (Mankins, 1995):

Table 1: Definition of Technology Readiness Levels for DoE and NASA

TRL	DoE TRL Definition	NASA TRL Definition
1	Basic principles observed and reported in white papers, industry literature, lab reports, etc. Scientific research without well-defined application.	Basic principles observed and reported
2	Technology concept and application formulated. Issues related to performance identified. Issues related to technology concept have been identified. Paper studies indicate potentially viable system operation.	Technology concept and/or application formulated
3	Proof-of concept: Analytical and experimental critical function and/or characteristic proven in laboratory. Technology or component tested at laboratory- scale to identify/screen potential viability in anticipated service.	Analytical and experimental critical function and/or characteristic proof-of-concept
4	Technology or Component is tested at bench-scale to demonstrate technical feasibility and functionality. For analytical modeling, use generally recognized benchmarked computational methods and traceable material properties.	Component and/or breadboard validation in laboratory environment
5	Component demonstrated at experimental scale in relevant environment. Components have been defined, acceptable technologies identified and technology issues quantified for the relevant environment. Demonstration methods include analyses, verification, tests, and inspection.	Component and/or breadboard validation in relevant environment
6	Components have been integrated into a subsystem and demonstrated at a pilot-scale in a relevant environment.	System/subsystem model or prototype demonstration in a relevant environment (ground or space)
7	Subsystem integrated into a system for integrated engineering-scale demonstration in a relevant environment.	System prototype demonstration in a space environment
8	Integrated prototype of the system is demonstrated in its operational environment with the appropriate number and duration of tests and at the required levels of test rigor and quality assurance. Analyses, if used support extension of demonstration to all design conditions. Analysis methods verified. Technology issues resolved pending qualification (for nuclear application, if required). Demonstrated readiness for hot startup	Actual system completed and “flight qualified” through test and demonstration (ground or space)
9	The project is in final configuration tested and demonstrated in operational environment.	Actual system “flight proven” through successful mission operations

Although TRL is commonly used, it is not common for agencies and contractors to archive and make available data on the timeline to transition between TRLs. The aim of this paper is to develop an algorithm that transforms Technical Maturity (TM) scores that were collected on a Department of Energy project into Technology Readiness Level scores as shown in Figure 1. For a given technology, the algorithm takes some of the TM parameters, such as Process Maturity (PM) or Hardware Equipment Maturity (EQ), as inputs to generate a TRL score compatible with the DoE’s nuclear TRL scale. Being able to transform the Technical Maturity scores into Technology Readiness Level scores allows us to add data from the Department of Energy to the database used for the broader project. This would help increase the statistical significance of the model by providing a larger TRL data pool.

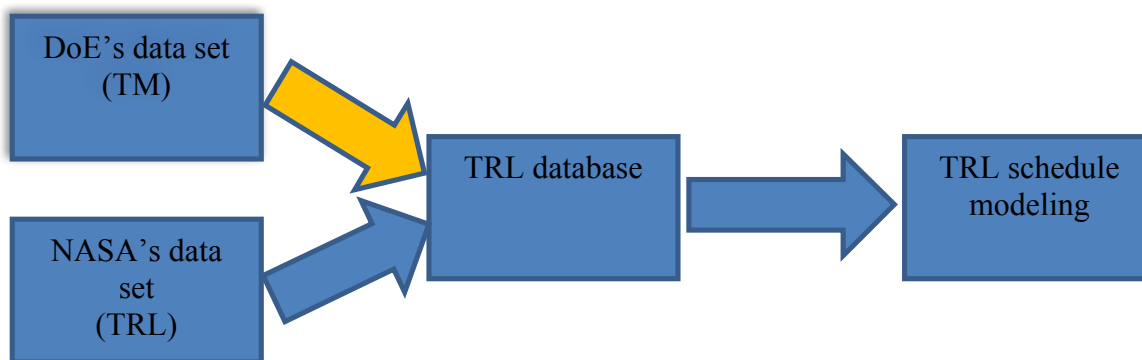


Figure 1: Merging the DoE TM Data into the TRL Database

Presentation of the Data

The NASA data source is a case study done by the Systems Analysis Branch at NASA's Langley research center looking at typical times aeronautical technologies take to mature. The data was collected through interviews with NASA personnel. Table 2 shows the full NASA data set, which captures the transition times from one TRL to the next. The indicated transition times are in years.

Table 2: TRL Data from NASA

Transition	Carbon-6 Thermal Barrier	Direct To	Fiber Preform Seal	Low Emissions combustors	Nondestructive Evaluation	Tailless Fighter	Thrust Vectoring Nozzle	Electro Explosive DeIcing	Engine Monitoring Systems	Flow Visualization	Fly-by-Light	GA Wing	Graphite Fiber Stator Vane Bushings (Tribology)	Particulate Imaging Velocimetry	Propfan development	Runway Grooves	Surface Movement Advisor	Supercritical Wing	Tiltrotor Technology
12	0.4	0.2	1	1	0.5	3	0.3	0.5	0.5	5	2.5	0.5	1.9	2	2.5	0.7	0.8	1.5	3
23	0.4	0.1	1.5	1	1	1	0.3	0.5	0.5	1	5	0.5	1.9	4	1	0.2	0.3	1	1
34	0.4	0.1	1.5	1	1	1	0.4	1	0.5	3.5	7.5	0.5	1.9	2.5	1.5	0.2	0.3	1	1
45	0.5	1.1	1.5	2	1	1	2	1	0.5	1	4	3	1.9	3	2.5	0.2	0.35	1	1
56	0.2	0.1	6	4	1	2	2	1	1	0.5	1.5	0.5	1.9	0.5	1	0.2	0.35	1	22
67								6	0	0.5	1.5	1.5	1.9	0.8	2.5	1	0	1	8
78								0.5	5	0.5	1.5	3	1.9	0.3	6	1	1.2	12	0
89								5.5	0	0.5	1.5	4	1.9	0.3	1	1	0.1	1	11

In general terms, the technical maturity of a technology can be viewed as a function of many different maturities, such as maturity of the requirements, processes, equipment, facilities, schedule, personnel, and safety. Kenley and Creque (1999) showed that, out of a set of seven parameters, four were particularly relevant in determining an overall technical maturity score for technologies that were being developed by DoE laboratories to stabilize nuclear materials: process maturity (PM), hardware equipment maturity (EQ), facility readiness (FAC) and operational safety readiness (SAFT). The TM score was generated as a weighted average of the maturity scores that were assessed for each of the seven parameters. Table 3 summarizes the definitions of the four most relevant parameters identified by Kenley and Creque. A parameter score of 0 means that a technology is in use and meets all requirements; while a score of 10 means that it is still in a pre-conceptual phase. The progression from concept to operational readiness is a typical complete cycle. However, steps can be skipped if evidence allows developers to do so.

Kenley and Creque used DoE data from the 94-1 R&D plans (US Department of Energy, 1995, 1996, 1997, and 1998), which were prepared by a Research Committee chartered by the Nuclear Materials Stabilization Task Group in order to fill gaps in the technology base needed to accomplish the stabilization of excess nuclear materials that were in the midst of

being produced for weapons and were no longer needed for this purpose when the Cold War concluded. After 1998, Kenley extracted updates to the data from other plans such as the 2000-1 plan (US Department of Energy, 2002) and formal letters from DoE to the Defense Nuclear Facilities Safety Board. The DoE data set is shown in Table 4.

Table 3: Scales for Determining Technical Maturity

Level	Hardware equipment maturity (EQ)	Facility readiness (FAC)	Safety readiness (SAFT)	Process Maturity (PM)
10	Equipment requirements not yet defined.	None available. New facility required.	New facility or facility restart required.	No currently identified solutions meet requirements
9	New design. Conceptual design completed.		NEPA process complete.	Design concept /technology application formulated
8	Experimental system. Cold demonstrated.			Cold feasibility demonstrated
7		Facility available. Major modifications required (new glove boxes, seismic mods).	Contractor Operational Readiness Review (ORR) complete.	
6	Experimental system. Hot demonstrated.			Hot feasibility demonstrated
5	Commercially equipment available. Requires modification.	Facility available. Moderate modifications required (modify glove boxes and equipment).	Contractor Readiness Assessment (RA) complete.	End-to-end design (flowsheet) complete
4	Integrated end-to-end equipment designs completed.			Cold prototype demonstrated at-end-use site
3	Cold prototype demonstrated.	Facility available. Minor modifications required (existing glove boxes and minor equipment mods).	DOE ORR complete (awaiting Secretary of Energy signature).	
2			DOE RA complete (awaiting Site Manager signature).	Hot prototype demonstrated at end-use site
1	Hot prototype demonstrated.			
0	Equipment in use processing the given material.	Facility operating. No modifications required.	Facility Authority to Proceed issued or within authorization basis.	Process integrated into operations

Table 4: DoE Technical Maturity Data

R&D Plan Technology	TM Evaluation Date	PM	EQ	FAC	SAFT	Years to Go	Date Operational
Bagless Transfer System - FB	Sep-91	4	3	0	0	2.00	Sep-93
Calciner with Full Batch TGA	Apr-94	6	0	0	5	3.76	Jan-98
Cementation	Sep-92	2	0	3	9	0.08	Oct-92
Charcoal Treatment	Sep-92	1	1	0	0	1.00	Sep-93
Digital Radiography	Sep-91	2	0	2	5	1.54	Mar-93
Electrolytic Decontamination	Sep-91	4	6	7	0	2.25	Dec-93
Electrolytic Decontamination	Sep-92	4	3	7	9	1.25	Dec-93
HB Phase II Solution Conversion	Aug-94	4	3	5	0	2.09	Sep-96
Nitric Acid Soluble Bags	Sep-91	5	1	3	3	2.08	Oct-93
Pipe Component	Sep-92	6	6	2	3	0.33	Jan-94
Polycube Pyrolysis	Jul-94	4	0	0	0	3.67	Mar-98
Precipitation - Magnesium Hydroxide	Sep-91	2	0	0	8	0.17	Nov-92
Precipitation - MgOH HAN	Jan-95	0	0	0	0	1.71	Sep-96
Pretreatment of RFETS SS&C - SRS	Sep-91	5	7	5	5	2.58	Apr-94
Pretreatment of RFETS SS&C - RFETS	Sep-92	2	1	5	5	1.58	Apr-94
Pu238 Storage Container - SRS/LANL	Sep-91	2	1	0	0	1.08	Oct-92
Pu239 Standard Container - RFETS	Sep-92	5	0	3	9	4.12	Oct-96
PuSPS - Packaging - RFETS	Sep-92	4	3	0	0	4.79	Jun-97
PuSPS - Packaging - LLNL	Apr-94	5	0	3	5	2.83	Feb-97
Pyrochemical Salt Oxidation	Sep-91	2	1	0	0	2.33	Jan-94
Scrub Alloy Processing	Sep-91	6	1	3	9	5.56	Mar-97
SS&C Stabilization	Sep-91	2	1	7	5	2.08	Oct-93
Thermal Stabilization - HAN	Apr-94	2	1	3	9	3.19	Jun-97
Thermal Stabilization - RFETS	Sep-91	2	4	3	9	1.37	Jan-93
Trapping of Uranium Hexafluoride	Sep-91	6	1	3	9	1.17	Nov-92
Vertical Calciner	Sep-91	8	8	5	7	4.00	Sep-95
Vertical Calciner	Sep-93	0	0	0	10	0.00	Sep-95

Developing the Correspondence between Technical Maturity and Technology Readiness Level for Nuclear Technologies

Qualitative Evaluation. Table 5 shows the correspondence between the DoE TRL definitions and the PM and EQ scale definitions for technical maturity. The rows that are not highlighted indicate the cases where the definitions were analogous enough to establish a correspondence between the TRL and TM scales without quantitative analysis. For the highlighted rows, there was no clear qualitative correspondence between the definitions (hence the use of the quantitative statistical methods to evaluate different correspondence possibilities).

The TRLs of the DoE technologies in Table 4 never exceed 7, because the technologies were operationally deployed as one-of-a-kind prototypes. Because of stringent safety requirements for nuclear operations, the prototypes were certified via a safety review prior to operations; however, the levels of performance-testing rigor and quality assurance that are typical of systems that are manufactured in large quantities were not necessary to allow the prototype to be used in operations and the TRL levels of 8 and 9 were not achieved.

Table 5: Correspondence between TRL, PM and EQ

TM	EQ	PM	TRL	DoE TRL Definition
10	Equipment requirements not yet defined	No currently identified solutions meet requirements	1	Basic principles observed and reported in white papers, industry literature, lab reports, etc. Scientific research without well-defined application.
9	New design. Conceptual design completed.	Design concept / technology application formulated	2	Technology concept and application formulated. Issues related to performance identified. Issues related to technology concept have been identified. Paper studies indicate potentially viable system operation.
8	Experimental system. Cold demonstrated.	Cold feasibility demonstrated	3	Proof-of concept: Analytical and experimental critical function and/or characteristic proven in laboratory. Technology or component tested at laboratory- scale to identify/screen potential viability in anticipated service.
6	Experimental system. Hot demonstrated.	Hot feasibility demonstrated	4	Technology or Component is tested at bench-scale to demonstrate technical feasibility and functionality. For analytical modeling, use generally recognized benchmarked computational methods and traceable material properties.
5	Commercially equipment available. Requires modification.	End-to-end design (flowsheet) complete	5	Component demonstrated at experimental scale in relevant environment. Components have been defined, acceptable technologies identified and technology issues quantified for the relevant environment. Demonstration methods include analyses, verification, tests, and inspection.
4	Integrated end-to-end equipment designs completed.	Cold Prototype demonstrated at end-use site	6	Components have been integrated into a subsystem and demonstrated at a pilot-scale in a relevant environment.
3	Cold prototype demonstrated			
2		Hot prototype demonstrated at end-use site		
1	Hot prototype demonstrated.			
0	Equipment in use processing the given material.	Process integrated into operations	7	Subsystem integrated into a system for integrated engineering-scale demonstration in a relevant environment.

Methods to Establish the Correspondence between TRL and TM. Multiple methods were evaluated to establish the correspondence between TRL and technical maturity for TRL 4, 5, and 6. Some of the methods use only PM and EQ as inputs, others use all of the four of the relevant parameters (PM, EQ, FAC, and SAFT). The definitions for FAC and SAFT are only remotely related to the definitions for TRL, whereas the definitions of PM and EQ were closely related to the definitions for TRL, and thus, were used alone in some methods. To generate a TRL score compatible with the DoE's nuclear TRL scale from these inputs, we had two possible approaches: (1) generating first an overall TM score that would then be transformed into a TRL score and (2) transforming each of the parameters' scores into TRL scores and then generating an overall TRL score from them. We dropped the first approach of generating first an overall TM score that would then be transformed into a TRL score quite early in the research, as it clearly had lower results.

In both cases, we used a correspondence table, such as Table 6. In Table 6, some TRL entries are yellow, and others are orange. The orange cells correspond to fixed values, because the definition of the corresponding TRL clearly matches a TM score definition (TRLs 1, 2, and 3 respectively closely matched the definitions of TM scores 10, 9, and 8 for parameters PM and EQ, while TRL 7 matched a TM of 0 for the reasons explained above, see Table 3). As for the yellow cells, the definitions were more flexible (the correspondence was less direct); therefore, those cells were used as input variables, allowing the mapping of the

TM values to the TRL values in the correspondence table to be changed in order to improve the results of the methods.

Table 6: Example Correspondence Table for Mapping TM to TRL

TM	TRL
10	1
9	2
8	3
7	4
6	5
5	5
4	6
3	6
2	6
1	6
0	7

We first used multiple correspondence tables that individually map each TM parameter’s scores to TRL, as shown in Table 7. This multi-dimensional variation increased computational complexity while it did not significantly improve the statistical scores compared to using one correspondence table, such as Table 6, that maps all TM parameters to the same TRL. The TM scales definitions are somewhat coherent in that a PM of 4 indicates approximately the same level of relative advancement on a 0 to 10 scale as an EQ of 4, and suggests that PM and EQ should share a common correspondence table with the TRL scale. Using a common correspondence table allowed us to simplify the computations by dramatically decreasing the number of combinations.

Table 7: Example Multiple Correspondence Table for Mapping TM to TRL

PM	TRL		EQ	TRL
10	1		10	1
9	2		9	2
8	3		8	3
7	4		7	4
6	4		6	4
5	4		5	5
4	5		4	5
3	5		3	5
2	5		2	6
1	6		1	6
0	7		0	7

Once a correspondence is established for mapping TM to TRL for the DoE data, the times to transition from a given TRL (call it TRL_{begin}) to another given TRL (call it TRL_{end}) for the DoE data set is compared statistically to the times to transition from TRL_{begin} to TRL_{end} for the NASA data set. Every method has two forms for comparing transition times: a form where the transition times from one TRL score to another are not processed, the raw-data method, and a form where every data point undergoes a log transformation. The

log-transformation reduces the larger values and separates the smaller ones, making the distribution of the data more closely match a normal probability distribution.

Methods Tested. Thirty-three methods were tested, and they can be grouped as shown in Table 8.

Table 8: Methods Tested

Approach Mapping	Minimum and Maximum	Weighted Arithmetic Mean	Weighted Geometric Mean	Weighted Harmonic Mean	System Failure Approach
Ceiling		3;8;13;16	20;23	26;29	31
Floor		4;9;14;17	19;22	25;28	33
Round		5;10;15;18	21;24	27;30	32
Other	1;2;6;7;11;12				

The minimum methods were tried because we thought that taking the minimum value among the parameters' score as the TRL score would make sense. These methods are conservative, they consider that to reach a TRL score, you need to have all the parameters at least equal to this score, even if some other parameters scores are higher. On the other hand, the maximum methods, by considering that the TRL score is equal to the best parameter score, are the less conservative ones.

To try to balance the degree of conservatism, weighted averages were then introduced. But as the TRL scores are integers, the weighted averages needed to be transformed into integers. To do that, three methods were used: the ceiling of weighted average, which corresponds to the less conservative approach, the floor of weighted average which corresponds to the conservative approach and the rounded weighted average, which is in between.

The weighted arithmetic mean gives more influence to the biggest values of the datasets, while the weighted geometric mean and even more the weighted harmonic mean reduce the influence of these values for the benefit of the smallest values.

The system failure approach, which is part of reliability engineering, is especially well adapted for our research. Let p_1, \dots, p_n be the failure rates of different components of a system. If all the components are required to be operational so that the global system can run, then the global failure rate is $1 - \prod_{i=1}^n (1 - p_i)$. This approach has been applied to our case by replacing p_1, \dots, p_n by PM/7 and EQ/7 or PM/7, EQ/7, SAFT/7 and FAC/7 (the idea being, that the higher a parameter TM score is, the less mature the "component" is, and the higher the probability of failure is). The global failure rate is the TRL score divided by seven, in this case.

The binary methods ensue from the observation that the SAFT and FAC scores are less relevant in technology maturity terms, and have a reduced spectrum of values. These methods all use PM, EQ, FAC and SAFT. The difference is that binary scores are generated from SAFT and FAC scores, which can only be equal to two different values. To find the right way to generate the binary scores, the medians of the FAC and SAFT TM scores were determined. Then the scores below median were given the value 0, while the other scores were given the value of the TRL score corresponding to the median.

Evaluation and Comparison of the Methods

The comparison between the different methods was done using DecisionTools Suite's StatTools and @RISK, which are add-ons to Microsoft Excel. These add-ons allowed us to perform statistical tests on the data from DoE projects processed by the methods used to

generate TRLs in comparison with the data from NASA projects. We looked at transitions from TRL 5 to TRL 7 (transition 5-7) and from TRL 6 to TRL 7 (transition 6-7). The choice not to look at TRL score 4 for the statistical tests was made because the definitions of PM and EQ do not match well with the DoE's definition of TRL score 4. Indeed, in the nuclear projects, TRL 4 does not really exist, and could almost be included in TRL 5.

Statistical Tests. As mentioned above the comparison between different methods is based on statistical tests. We are statistically testing whether the two samples (i.e. the DoE and NASA datasets) were randomly drawn from a same population. A classical way of doing this is by testing for the equality of means using the T-test. The same can be done with fewer assumptions by using nonparametric tests such as the Mann-Whitney test or the Fisher exact tests of equality of medians.

For each test, the p-value was used as the sole comparison criterion: the higher the p-value, the more likely the two samples were drawn from a same population. This is different from the classical use of p-values in hypothesis testing where they are only used to *reject* the hypothesis that the sample are from the same population (the p-value is used to check if the hypothesis can be rejected by comparing it to the pre-chosen significance level).

Hence if dataset A has a higher p-value than dataset B when compared to the NASA dataset, then NASA and dataset A are more likely to have been drawn from the same population than NASA and dataset B. Let us note however, that the p-value 0 to 1 scale is skewed, meaning that we get more information from smaller p-values by that indicate the two samples are surely not from the same population, while a high p-value only means that we cannot reject the hypothesis that they come from the same population. We mostly used an elimination process that rejects low p-values, or surely bad results, by simultaneously using the three statistical tests, and other auxiliary criteria (counts of the transitions to TRL 7 and TRL 8, counts of transitions 5-7 and 6-7, sensitivity analysis, and simultaneous non-rejection of 5-7 and 6-7 transitions).

The screenshot in Figure 2 shows the duration for the TRL 5-7 and 6-7 transitions for both the DoE data (in red) , processed by the tested method, and NASA data (in purple). It also shows their means. Tests' p-values for TRL transitions 5-7 and 6-7 can be seen in the tables under the corresponding points in the graph.

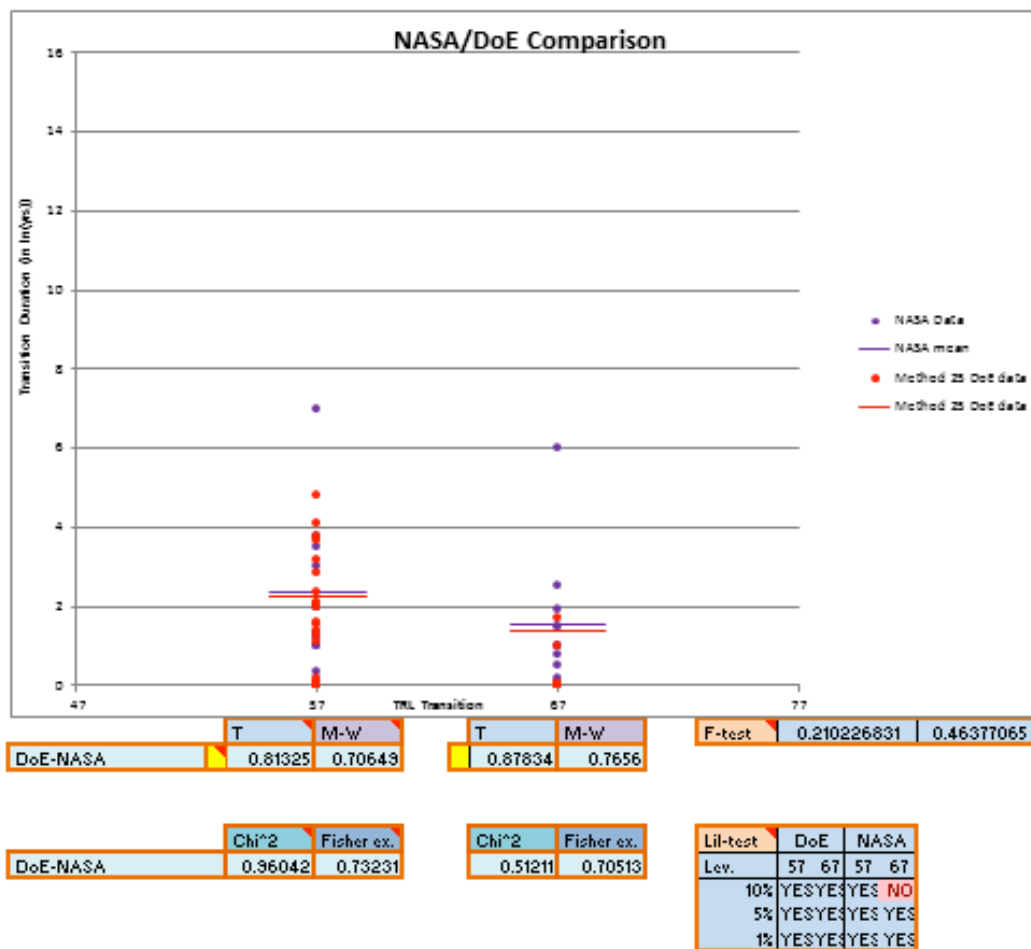


Figure 2: Screenshot Comparing Transition Times for the Two Populations

Student’s T-test. The Student’s T-test assesses whether the means of two groups are statistically different from each other, that is to say it provides a two-population location test of the null hypothesis that the means of two normally distributed populations are equal. It needs to be handled with caution: for every case you need to make sure that

1. There is no rejection of normality using the Lilliefors test for the DoE or NASA data sets (in Figure 2, the case where normality is rejected is indicated by a "NO" in the Lil-test table)
2. There is equality of variances using the Fisher’s test, (in Figure 2, the F-test table has the p-value for evaluating the equality of variances)

Let us be more specific about this second point: there are two types of T-tests: the homoscedastic T-test and the heteroscedastic one. A sequence of random variables is homoscedastic if all random variables in the sequence have the same finite variance.

It is better to use the homoscedastic T-test because it is more powerful. To know which test should be used, a Fisher’s test for the null hypothesis that two normal populations have the same variance is run: if the F-test p-values are small enough (less than 0.05), this means the variances are significantly different, and it is then advised to use the heteroscedastic T-test.

When the heteroscedastic T-test is used instead of the homoscedastic one, we hope that (1) if the T-test p-value was very large, that it will stay large and not get in the red zone, and (2) if the T-test p-value very small, that it will not get very larger and stay in the red zone.

The results of Student’s T-test are seen under the “T” cells in Figure 2 (on the left for

transition 5-7 and on the right for transition 6-7). When condition 1 is not respected, or condition 2 gives mixed results, then move on to the adjacent Mann-Whitney test.

Mann-Whitney. The Mann-Whitney test is the nonparametric analog of the Student's T-test, and is based on ranks. It assesses whether two independent samples of observations have equally large values. It is more robust to outliers than Student's T-test, making it more widely applicable than the T-test. Unlike the parametric t-test, this non-parametric test makes no assumptions about the distribution of the data (e.g., normality). It will give results similar to the T-test, and should be considered when the pre-conditions for the T-test are not met.

Selection

Our criteria to select the methods allowed us to progressively eliminate the methods. Figure 3 summarizes the results of the process to down-select methods.

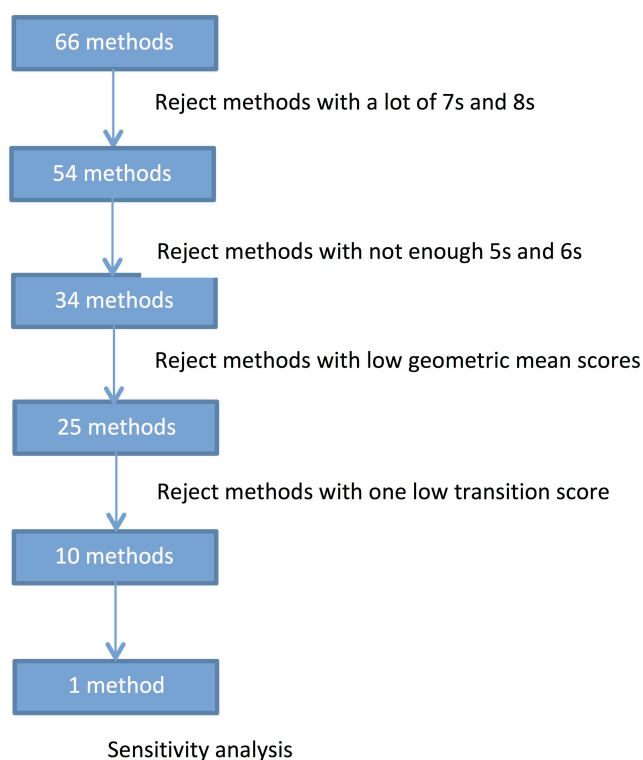


Figure 3: Down-selecting To Find the Best Method

Throughout this method elimination process, the principle of parsimony was used: when two methods had similar results, we always went for the method using the least inputs.

First of all, the methods that gave the most TRL scores of 7 and 8 were rejected. Indeed, these TRL scores cannot be reached for the DoE technologies considered, because the maximum TRL score of 7 was defined as the boundary condition of all TM parameters being 0 when the technology became operational, and the data that we are comparing is the transition times prior to becoming operational. We chose to eliminate the methods giving more than three TRL scores of 7 or 8.

Second, the methods which gave the least TRL scores of 5 and 6 were rejected, because having few such scores decreases the significance of the tests undergone by the methods, which rely on transitions 5-7 and 6-7. We chose to eliminate the methods giving less than 16 TRL scores of 5 or 6.

Then we calculated the geometric mean of the 5-7 and 6-7 transitions test results for each of the three tests, and rejected the methods that gave the lowest geometric means. This

allowed us to take both transitions into account during the elimination process. Once we had kept only the methods with the highest geometric mean scores, we eliminated the methods that had low scores for one of the transitions.

These different rejections led to five log methods and five raw-data methods. An interesting observation is that they were actually, with one exception, the same methods in their log and raw-data forms. This comes from the fact that two out of the three tests that were used were non-parametric tests based on rank, and the non-parametric tests were used instead of the T-test whenever normality was violated.

Finally, to choose the best method out of these ten methods, a sensitivity analysis was run with @RISK. By introducing variation to the control parameters' weightings (normal distribution with 0.1 standard deviation), but also to the TM-TRL correspondence scale (a special algorithm was developed to randomly move the position of the 4-5 and the 5-6 transitions in the scale), we generated distributions of the test results summarized in Table 9 and Table 10.

Table 9: Sensitivity Analysis Results for Raw-data (Non-Log) Method

	Non-Log														
	M1 Non-Log			M4 Non-Log			M18 Non-Log			M19 Non-Log			M25 Non-Log		
	5%	95%	Norm.mode	5%	95%	Norm.mode	5%	95%	Norm.mode	5%	95%	Norm.mode	5%	95%	Norm.mode
DoE-NASA / T 57	0,2037966	0,8680483	1,023260659	0,2832202	0,9813812	#VALUE!	0,3283842	0,9771305	1,303117919	0,1560792	0,9813812	0,823580499	0,1459566	0,9813812	0,973451859
DoE-NASA / M-W 57	0,2535563	0,9104338	0,557044969	0,2084497	1	0,954144976	0,2120284	0,9781417	0,838560602	0,1733676	1	0,442651784	0,1469011	1	0,828139827
DoE-NASA / Fisher ex. 57	0,2307692	1	0,421900643	0,2392939	0,9	0,622974688	0,1272727	0,7368421	0,967730105	0,2392939	1	0,426628587	0,2392939	1	0,962668358
DoE-NASA / T 67	n/a	n/a	#VALUE!	n/a	n/a	#VALUE!	n/a	n/a	#VALUE!	n/a	n/a	#VALUE!	n/a	n/a	#VALUE!
DoE-NASA / M-W 67	0,0930461	0,7656017	1,138347143	0,0723302	1	0,347546871	0,0924903	0,8798024	1,27014433	0,0638655	1	0,537124187	0,084994	1	0,836717677
DoE-NASA / Fisher ex. 67	0,0921246	0,8333333	0,502820504	0,068937	0,8333333	0,283515573	0,1490398	1	0,429379206	0,0657393	1	0,669392391	0,0921246	0,8333333	0,951322097
DoE-NASA / T GM	n/a	n/a	#VALUE!	n/a	n/a	#VALUE!	n/a	n/a	#VALUE!	n/a	n/a	#VALUE!	n/a	n/a	#VALUE!
DoE-NASA / M-W GM	0,2356205	0,7715675	0,800335289	0,2033015	0,7347786	0,928464485	0,2272331	0,8066453	1,383330946	0,1957271	0,7314496	0,800670677	0,2004432	0,748979	1,34075057
DoE-NASA / Fisher ex. GM	0,2417927	0,9128709	0,671327934	0,2062015	0,7939191	0,508180998	0,1826317	0,7694837	0,791107031	0,1921201	0,9486833	0,595473771	0,213552	0,9128709	1,02755618

Table 10: Sensitivity Analysis Results for Log Method

	Log														
	M1 Log			M4 Log			M18 Log			M25 Log			M32 Log		
	5%	95%	Norm.mode	5%	95%	Norm.mode	5%	95%	Norm.mode	5%	95%	Norm.mode	5%	95%	Norm.mode
DoE-NASA / T 57	0,332498	0,9969882	0,80071283	0,1245221	0,9305906	1,154480839	0,1174343	0,8800382	0,708860578	0,0188269	0,9305906	#VALUE!	0,273028	0,9957346	1,287646519
DoE-NASA / M-W 57	0,2535563	0,9104338	0,557044969	0,2110157	1	0,957248125	0,2299602	0,9753959	0,861821388	0,1469011	1	0,907802955	0,2540932	1	1,012530978
DoE-NASA / Fisher ex. 57	0,2307692	1	0,421900643	0,2213622	0,9	0,606513779	0,1272727	0,7368421	0,967730105	0,2086957	1	0,850456603	0,2086957	0,7033818	0,832049206
DoE-NASA / T 67	0,696315	0,696315	#VALUE!	0,3509972	0,7206926	1,701017474	0,0444038	0,8734698	0,978597612	0,3509972	0,7206926	1,883482954	0,3509972	0,8108544	1,36750786
DoE-NASA / M-W 67	0,0930461	0,7656017	0,747626714	0,0723302	1	0,347546871	0,0924903	0,8512821	1,317884563	0,0846903	1	0,836440072	0,0723302	1	0,347546871
DoE-NASA / Fisher ex. 67	0,0921246	0,8333333	0,843739429	0,068937	0,8333333	0,283515573	0,1490398	1	0,429379206	0,0728016	0,8333333	0,927151593	0,068937	0,8333333	0,283515573
DoE-NASA / T GM	n/a	n/a	#VALUE!	0,3267211	0,6408733	2,43509352	0,1372803	0,828391	0,958250213	0,3267211	0,6478828	#VALUE!	0,3384723	0,76499	1,79357149
DoE-NASA / M-W GM	0,2356205	0,7715675	0,800335289	0,2033015	0,7347786	0,928464485	0,2200375	0,8015188	1,378408604	0,2004432	0,748979	1,403757051	0,2029139	0,7214199	0,951691228
DoE-NASA / Fisher ex. GM	0,2417927	0,9128709	0,671327934	0,2062015	0,7939191	0,508180998	0,1816207	0,7694837	0,789746494	0,2091613	0,9128709	0,978900182	0,2107628	0,7656053	0,538291346

The 5% and 95% values are the boundaries of the range where 90% of the values belong. The higher the 5% value is, the better (or less sensitive) the method. The “normalized mode” value is equal to the value obtained without the sensitivity analysis divided by the difference between the 95% and 5% values. The higher it is, the better. The 5% values or the normalized modes values that were significantly higher (better) with a method are highlighted in orange, and those that are significantly lower (worse) are highlighted in green in Table 9 and Table 10.

Three methods stood out as the best: M18 Raw-data, which is the method using rounded weighted averages with PM, EQ, FAC and binary SAFT, and M25 in its Log and Raw-data forms. M25 uses the floor of weighted harmonic mean with PM and EQ.

The results from the three methods were significantly different and M25 uses only two inputs; therefore, we chose M25 as the best method. Between log and raw-data form for M25, the differences were not significant, so we chose the raw-data form, which is easier to use. The resulting correspondence tables for M25 to map PM and EQ to TRL are shown in Figure 4. The TRLs obtained according to the observed combinations of PM and EQ are displayed in Table 11.

PM	TRL	EQ	TRL
10	1	10	1
9	2	9	2
8	3	8	3
7	4	7	4
6	4	6	4
5	4	5	4
4	5	4	5
3	5	3	5
2	5	2	5
1	6	1	6
0	7	0	7

Figure 4: Correspondence Table for M25

Table 11: Result of Applying M25 Correspondence Tables to Observed Data Points

		EQ											
		10	9	8	7	6	5	4	3	2	1	0	
PM	10												
	9												
	8			3									
	7												
	6					4					4	5	
	5				4						4	5	
	4					4			5			5	
	3												
	2							5			5	5	
	1										6		
	0												6

Conclusion

With the appropriate mapping, we observe that DoE’s and NASA’s transition times are very similar. It leads us to think that statistically we find the same timelines for the acquisition of advanced technologies across the different agencies. The statistical tests’ p-values for the chosen method are especially good as shown in Figure 5.

Transition 57				Transition 67			
		T	M-W			T	M-W
DoE-NASA	0	0.813246	0.706485	0	0.878343	0.765602	
		Chi^2	Fisher ex.			Chi^2	Fisher ex.
DoE-NASA		0.96042	0.732308		0.512111	0.705128	

Figure 5: P-values for Chosen Method

Table 12 shows the transition times in years for DoE’s dataset with the TRLs calculated using the chosen methods. Note that transition times covering more than one cell indicates that no observed data was reported for a transition from TRL to TRL+1, but that the transition time from TRL to TRL+2 or to TRL+3 was reported.

Table 12: Observed Transition Times Based on TRLs Derived Using Chosen Method

Transition	Bagless Transfer System - FB	Calcliner with Full Batch TGA	Cementation	Charcoal Treatment	Digital Radiography	Electrolytic Decontamination	HB Phase II Solution Conversion	Nitric Acid Soluble Bags	Pipe Component	Polycube Pyrolysis	Precipitation - Magnesium Hydroxide	Precipitation - MgOH HAN	Pretreatment of RFETS SS&C - SRS	Pretreatment of RFETS SS&C - RFETS	Pu238 Storage Container - SRS LANL	Pu239 Standard Container - RFETS	PuSPS - Packaging - RFETS	PuSPS - Packaging - LLNL	Pyrochemical Salt Oxidation	Scrub Alloy Processing	SS&C Stabilization	Thermal Stabilization - HAN	Thermal Stabilization - RFETS	Trapping of Uranium Hexafluoride	Vertical Calcliner
12																									
23																									
34																									
45						1																			
56	2	3.76	0.08		1.54	1.25	2.09	1.08	1.33	3.67	1.17		2.58	1.58	1.08	4.12	4.79	2.83	2.33	3.56	1.08	3.19	1.37	1.17	2
67				1								1.71													0.92

The research developed statistical models for schedule uncertainty estimation that use Technology Readiness Levels as the only user-provided inputs with the goal of providing a simple, yet powerful, method that is applicable to all acquisitions of advanced technology in DoD and other federal agencies.

Data-driven models that can be applied easily and effectively can play a major role in identifying key technical areas that are associated with cost and schedule uncertainties. Current approaches to modeling uncertainty in cost and schedule require multiple inputs that can be quite subjective and are not integrated with technology readiness assessments that usually are performed separately as part of the technical risk management effort. A modeling method that uses fewer inputs that are less subjective and integrated with technology readiness assessment will be more likely to be used by program managers. This will provide more a credible analysis that enhances evaluation of alternatives, identifies risk areas, and ultimately provides benefits that derive from the program managers’ ability to intervene earlier in the life cycle either to select technologies that are aligned to cost and schedule constraints, or to produce more realistic plans for expenditures and schedules that aligned to the maturity of technologies that promise increased performance.

We evaluated data previously collected on DOE and NASA projects, and compared NASA and DOE data on the times to transition from TRL 5 (component demonstration in a relevant environment) and TRL 6 (subsystem demonstration in a relevant environment) to TRL 7 (system prototype demonstration in a relevant environment). Statistically speaking, research and development timelines for nuclear materials processing technologies based on chemical engineering R&D approaches and techniques are the same as for flight technologies that use aerospace engineering R&D approaches and techniques. This supports our thesis that the approach and even the data and models can be used across the spectrum of industries and agencies that are engaged in R&D of advanced technologies.

Acknowledgement

This publication was developed under work supported by the Naval Postgraduate School (NPS) Acquisition Research Program Assistance Agreement No. NN00244-10-I-0070 awarded by the U.S. Fleet Industrial Supply Center, San Diego (FISC-SD). It has not been formally reviewed by NPS. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. FISC-SD and NPS. The FISC-SD and NPS do not endorse any products or commercial services mentioned in this publication.

References

Kenley, C. R., and T. R. Creque. 1999. “Predicting Technology Operational Availability Using Technical Maturity Assessment.” *Systems Engineering* 2: 198-211.

- Mankins, J. C. 1995. "Technology Readiness Levels." White paper, NASA Office of Space Access and Technology. <http://www.hq.nasa.gov/office/codeq/trl/trl.pdf>.
- US Department of Energy. 1995. *Research and Development Plan*. Washington, DC (US): Nuclear Materials Stabilization Task Group Research Committee.
- . 1996. *Research and Development Plan*. Washington, DC (US): Nuclear Materials Stabilization Task Group Research Committee.
- . 1997. *Research and Development Plan*. Washington, DC (US): Nuclear Materials Stabilization Task Group Research Committee.
- . 1998. *Research and Development Plan*. Washington, DC (US): Nuclear Materials Stabilization Task Group Research Committee.
- . 2002. *An Implementation Plan for Stabilization and Storage of Nuclear Material*. Washington, DC (US): Office of Environmental Management.
- . 2009. *Technology Readiness Assessment Guide*. Washington, DC (US): Department of Energy.

Biography



Maxime Crépin is a graduate student pursuing a master's degree in Mechanical and Industrial Engineering at Arts et Métiers ParisTech, where he also received a B.S. in Engineering. He will be simultaneously be pursuing a master's degree in a graduate program in the United States that is yet to be determined. His research interests are in the area of policy and systems engineering as they apply to technology cost and schedule forecasting.



Bernard El-Khoury is a graduate student at the Massachusetts Institute of Technology where he is pursuing a master's degree in the Technology and Policy Program. He is simultaneously pursuing a master's degree in Industrial Engineering at Ecole Centrale Paris, where he also received a B.S. in Engineering. His research interests are technology cost and schedule forecasting, and power systems modeling.



C. Robert Kenley holds a Ph.D. and M.S. in Engineering-Economic Systems from Stanford, a M.S. in Statistics from Purdue, and a S.B. in Management from MIT. He has over 30 years experience in systems engineering in the research and development of advanced socio-technical systems in the aerospace, nuclear, and healthcare domains. He has provided clients with insight and understanding of systems problems as an independent consultant, and he currently is a research associate at MIT. He is a published author of several papers and journal articles in the fields of systems engineering, decision analysis, Bayesian probability networks, and applied meteorology.