Technology Development Parametric Cost and Schedule Modeling

ICEAA 2017 Professional Development & Training Workshop

Portland, Oregon

June 6-9, 2017

Chuck Alexander, PE, CPA

National Security Analysis Department

Technology, Programs and Cost

JOHNS HOPKINS
APPLIED PHYSICS LABORATORY

Overview

- **Background**
  - **Conceptual & Early Life Cycle Technology Development**
  - **Challenges of Technology Development Estimating**
  - **Industry Literature, Tools, and Data Search**
- **Model Development Methodology / Approach**
  - **Technology Development Project Dataset**
  - **Variable Selection and Data Modeling**
  - **Modeling Forms Investigated**
  - **Model Evaluation Criteria**
- **Cost Model Performance Results**
- **Cost Model Output and Applicability**
- **Schedule Modeling Results**
- **Cost and Schedule Model Variability Drivers**
- **Conclusions and Future Considerations**
- **Backup Slides**

Conceptual & Early Life Cycle Technology Development

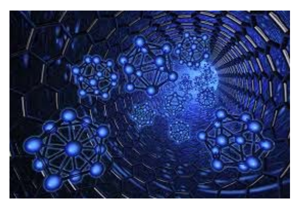
- **Numerous applications across DoD, Intel, Space and Civil sectors**
- **Breadth of focus areas and platforms including**
  - Sea – Ships / Submarines / Unmanned
  - Air – Aircraft / Airships / Unmanned
  - Space – Satellites / Spacecraft / Probes
  - Weapon Systems – Strategic / Tactical
  - Networks – Ground, Space and Marine Data / Communications / Sensors
  - Robotics / Automation / Nanotechnology
  - Information Technology / Electronics / Cyber
  - Military Strategy and Force Structure
  - Energy and Infrastructure
  - Warrior Armament
  - Healthcare

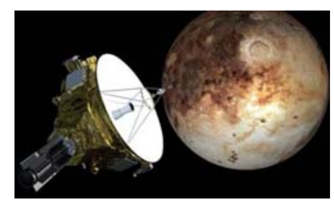






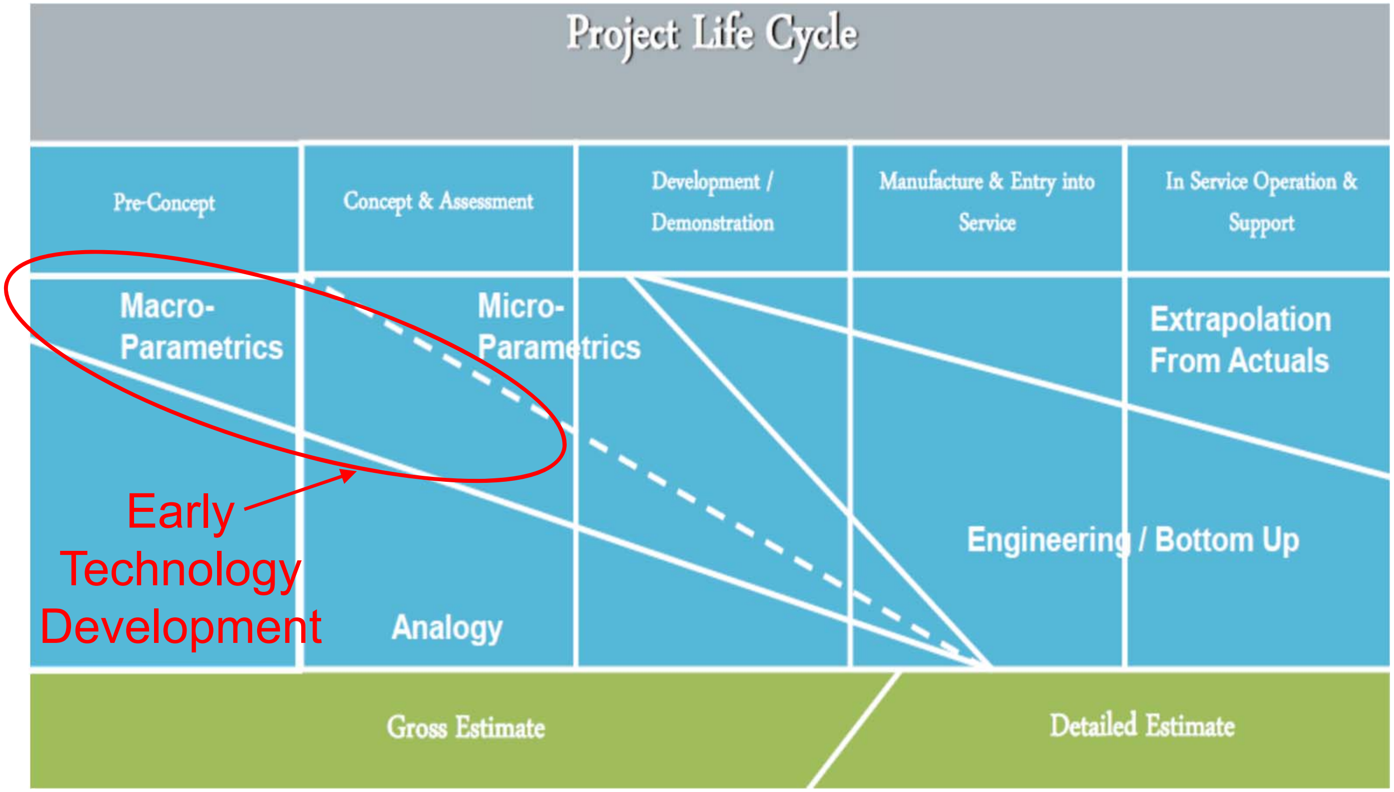






Challenges of Early Technology Development Estimating

- **Little or no analogous or comparable systems / applications**
- **High uncertainty and level of unknowns**
- **Lack of conceptual technical, engineering, design or performance parameters available to drive traditional micro-parametric models**



Source: "Macro-parametrics and the applications of multi-colinearity and Bayesian to enhance early cost modeling" - QinetiQ, Shermon & Barnaby, ICEAA 2015 Professional Development & Training Workshop

Industry Literature, Tool, and Data Search

- **Literature Search for Technology Development Cost and Schedule Estimating Methods / Models**
  - Various frameworks, analysis and modeling approaches have been proposed or developed
  - Research papers offer insightful analysis, methods and considerations for use of "macro level" parameters (e.g., Technology Readiness Level)
  - Deliver varying results but most are based upon limited data sets or focus on select technology areas / applications
- **Leading Technology Development Estimating Tools and Databases**
  - Available tools generally driven by detailed design, performance, and complexity "micro parameters", not available in early stages
  - General lack of available macro level parametric tools
  - Government sector repositories, databases and models primarily focused on Procurement or O&S Phases
  - Databases generally proprietary / protected or limited access

Technology Development Project Dataset

- Project dataset search conducted to develop broad-based technology development models
- NASA Technology Cost and Schedule Estimating (TCASE) tool identified and selected as resource for model development
  - Contained historical project cost, schedule and technical data with macro variables and project record quantities sought
  - Extensive core technology database containing over 2,900 project records with 164 available data fields across 14 broad-based technology areas (TA)

14+1 TECHNOLOGY AREAS
Table 2

TA #	Description
TA01	Launch Propulsion Systems
TA02	In-Space Propulsion Technologies
TA03	Space Power and Energy Storage
TA04	Robotics, Telerobotics, Autonomous Systems
TA05	Communication and Navigation
TA06	Human Health, Life Support, Habitation Systems
TA07	Human Exploration Destination Systems
TA08	Science Instruments, Observatories, Sensor Systems
TA09	Entry, Descent, and Landing Systems
TA10	Nanotechnology
TA11	Modeling, Simulation, Information Tech
TA12	Materials, Structures, Mechanical Systems, Manufacturing
TA13	Ground and Launch Systems Processing
TA14	Thermal Management Systems
(+) 1	Aeronautics

Model Development Methodology / Approach - Variable Selection and Data Modeling

- **Selection of Cost and Schedule dependent response variables**
  - **Total Development Cost ($) and Project Duration (months)**
  - **Continuous quantitative variables (i.e., cardinal numbers)**
- **Database fields with greatest potential as independent predictor variables for cost and schedule**
  - **System Hierarchy (SH) Level* (1 to 5);**
  - **Project Start / End Technology Readiness Level (TRL) (1 to 9);**
  - **Research and Development Degree of Difficulty (RD3) (Levels I to V);**
  - **Technology Area (TA1 to TA14);**
  - **System Characteristics;**
  - **Key Performance Parameters (KPPs);**
  - **Total FTEs (i.e., Full time Equivalents of project labor);**
  - **Capability Demonstrations**

** For modeling, the term **Hierarchy Rank** was used to represent the **SH Level***

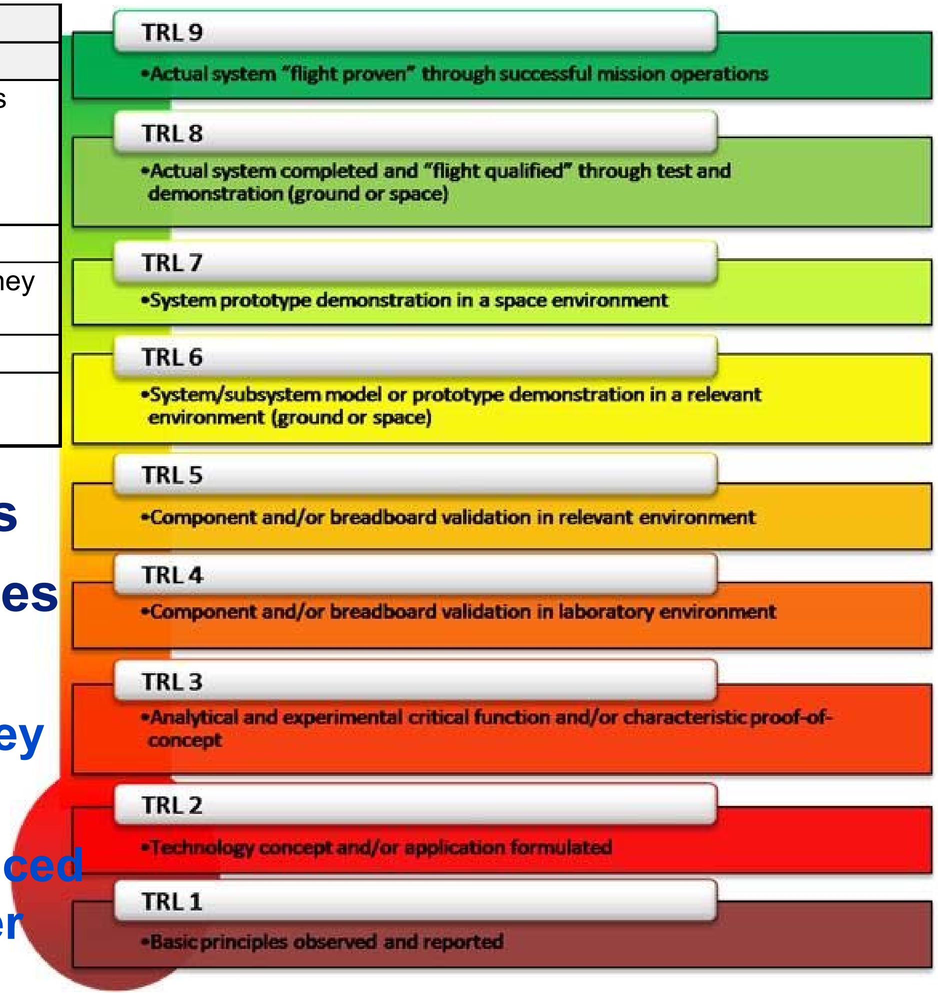
Model Development Methodology / Approach - Variable Selection and Data Modeling

After careful assessment two viable predictor candidates emerged

System Hierarchy Level*

System Hierarchy Table		
No.	**Tier**	**Definition**
5	System	An integrated set of constituent elements that are combined in an operational or support environment to accomplish a defined objective
4	Subsystem	A portion of a system
3	Assembly	A set of components (as a unit) before they are installed to make a final product
2	Component / Part	A portion of an assembly
1	Hardware / Material	An item or substance used to form a component

Technology Readiness Level*



- **Both ordinal categorical variables**
- **Other predictor variable candidates were eliminated based upon:**
  - **Insufficient project records with key predictor or response variables**
  - **Data relationship screening produced poor fit or overlap with other better suited variables**

**Source: NASA TCASE Training Guide and User Manual*

Data Modeling – Cost vs. TRL Transition Data

- **Investigation determined inadequate no. of observations in most TRL Start - End transition (TRL X-Y) categories for sufficient sample sizes**
  - **Only 5 of 36 TRL Start-End (TRL X-Y) categories contained "large" sample sizes (>30)**
  - **Cost curve fits developed for 14 of the 36 TRL X-Y categories (with >7 obs.)[1] produced erratic results**
- **TRL Improvement (TI) Level[2] was therefore examined and selected as viable alternative**
  - **Provided causal relationship & needed sample sizes**
  - **Resulted in consistent range across starting TRLs with adequate sample sizes for TI Levels 1 to 5**
  - **TI level 6 or greater appears to be extremely rare**

1. Small sample sizes < 8 observations demonstrated substantial volatility produced by limited inputs and considered too small to demonstrate statistical significance

2. TRL Start to TRL End difference, sometimes referred to as "TRL Transition Order" (e.g., TRL 3-5 is of Transition Order 2)

Start TRL	End TRL	TRL X-Y	No. Obs.
TRL Improvement Level 1			**177**
1	2	1-2	20
2	3	2-3	45
3	4	3-4	66
4	5	4-5	17
5	6	5-6	20
6	7	6-7	8
7	8	7-8	1
8	9	8-9	0
TRL Improvement Level 2			**133**
1	3	1-3	10
2	4	2-4	51
3	5	3-5	24
4	6	4-6	45
5	7	5-7	3
6	8	6-8	0
7	9	7-9	0
TRL Improvement Level 3			**63**
1	4	1-4	11
2	5	2-5	18
3	6	3-6	33
4	7	4-7	1
5	8	5-8	0
6	9	6-9	0
TRL Improvement Level 4			**22**
1	5	1-5	3
2	6	2-6	16
3	7	3-7	1
4	8	4-8	1
5	9	5-9	1
TRL Improvement Level 5			**10**
1	6	1-6	5
2	7	2-7	3
3	8	3-8	1
4	9	4-9	1
TRL Improvement Level 6			**0**
1	7	1-7	0
2	8	2-8	0
3	9	3-9	0
TRL Improvement Level 7			**0**
1	8	1-8	0
2	9	2-9	0
TRL Improvement Level 8			**0**
1	9	1-9	0
36		**Total**	**405**

Model Development – Modeling Forms Investigated

Hundreds of model variants under 4 primary forms

- Tailored curve fit function models
  - Over 20 functions* evaluated for ea. Cost and Schedule, TI and SH level
- Simple linear regression models
  - Single (univariate) and Composite (multivariate) predictor variables
  - Transformed predictor and/or response variables (up to 11 transformation types were evaluated for each variable combination)
- Multiple linear regression models
  - Multiple predictor - TI and SH Level
  - Transformed predictor and/or response variables (up to 11 types each)
- A range of nonlinear (NL) models
  - 21 forms for each predictor variable evaluated including polynomial, sigmoid & logistic curves, exponential & peak models, et.al.

** Beta, Chi-square, Erlang, Exponential, Gamma, Inverse Gaussian, LaPlace, Levy, Logistic, LogLogistic, Lognorm1/2, Pareto1/2, Pearson5/6, PERT, Raleigh, Triangular, Uniform, Weibull*

Model Evaluation / Selection Criteria

1. **Statistical Key Performance Measures (KPMs)**
   - **Error Variability and Dispersion Measures:**
     - Coefficient of Determination - R^2 and Adjusted R^2
     - Root Mean Square Error (RMSE)
     - Coefficient of Variation (CV)
   - **Statistical Significance Measures:**
     - F-ratio and t-stat (% of model terms with probability > |t|)
   - **Autocorrelation Measure:** Durbin-Watson test
   - **Data Reduction Measure:** Percent (%) of original data sample set unused
2. **Assessment of prediction model fit to actual sample data**
   - Various statistical measures and graphic data fit plots / charts
3. **Specific measures relevant to the particular model form**
   - Optimization methods for curve fits and measures applicable to linear and nonlinear models
   - VIF to measure multicollinearity for multiple regression models

Cost Model Performance Results - KPM

Cost Model KPM Results by model type

NOTIONAL PERFORMANCE RATING			
Good	Fair	Marginal	Poor

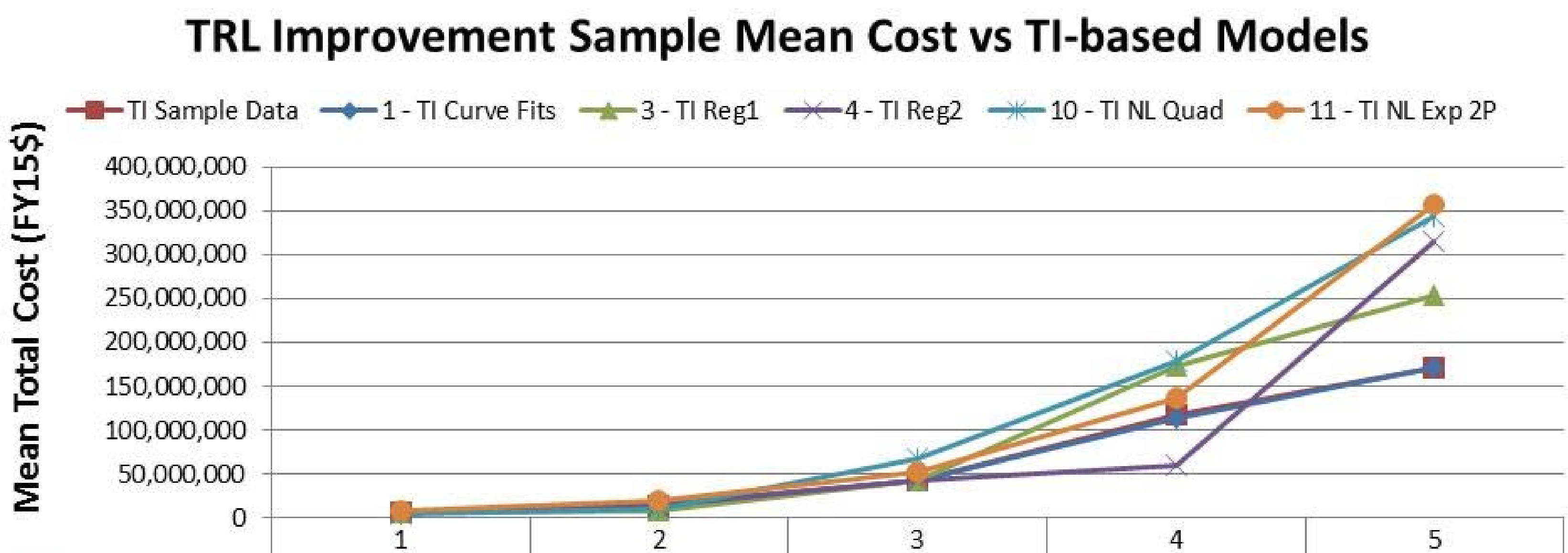
Mdl. No.	Model Form / Method	Predictor Variable Form	Key Performance Measures (KPM)								
			R-Sq	Adj R-Sq	RMSE (000's)	Coef. of Variation (CV)	F-ratio	Prob. > F	t-stat: % of terms w/ Prob. > |t|	Durbin-Watson Stat	Data Reduction (%)
1	Tailored Curve Fits	TI Level	N/A	N/A	40,929	0.736	N/A	N/A	N/A	N/A	2.5%
2	Tailored Curve Fits	SH Level	N/A	N/A	26,724	0.711	N/A	N/A	N/A	N/A	3.2%
5	Simple Linear Regression	SH Level	0.935	0.934	2,590	1.249	1893.2	<.0001*	75%	0.896	11.8%
6	Simple Linear Regression	SH Level	0.659	0.657	29,132	2.486	280.8	<.0001*	50%	1.275	3.5%
7	Composite Linear Regression	$[TI \times SH]^2$	0.772	0.771	38,324	1.526	719.5	<.0001*	100%	1.433	3.6%
8	Multiple Linear Regression	TI + SH	0.823	0.816	33,397	1.226	116.7	<.0001*	100%	1.757	5.0%
9	Multiple Linear Regression	$[TI + SH]^2$	0.788	0.780	2,621	0.617	90.4	<.0001*	50%	1.208	8.1%
10	Nonlinear - Quadratic	NL TI Level	0.610	0.609	32,685	1.606	N/A	N/A	N/A	N/A	15.3%
12	Nonlinear - Exponential 3P	NL SH Level	0.744	0.743	24,966	2.070	N/A	N/A	N/A	N/A	11.3%
13	Nonlinear - Gompertz 4P	NL SH Level	0.742	0.742	25,061	2.078	N/A	N/A	N/A	N/A	11.3%
	Note - TI level regression models 3 & 4 and TI NL model 11 were eliminated due to poor KPM results.										

- **Multiple linear regression models (8 & 9) performed well for KPMs alone**
- **Curve fit models (1 & 2) best replicated the underlying sample data central values and distribution shapes (see following slides)**

*Note: KPM categories that do not apply, cannot be generated, or are not available to a particular model form are shown as **N/A** for not applicable.*

Cost Model Output - TI Level Cost Models vs Actual Data

TI Model nos. 1, 3, 4, 10 and 11 vs TI Sample Project Data (means)



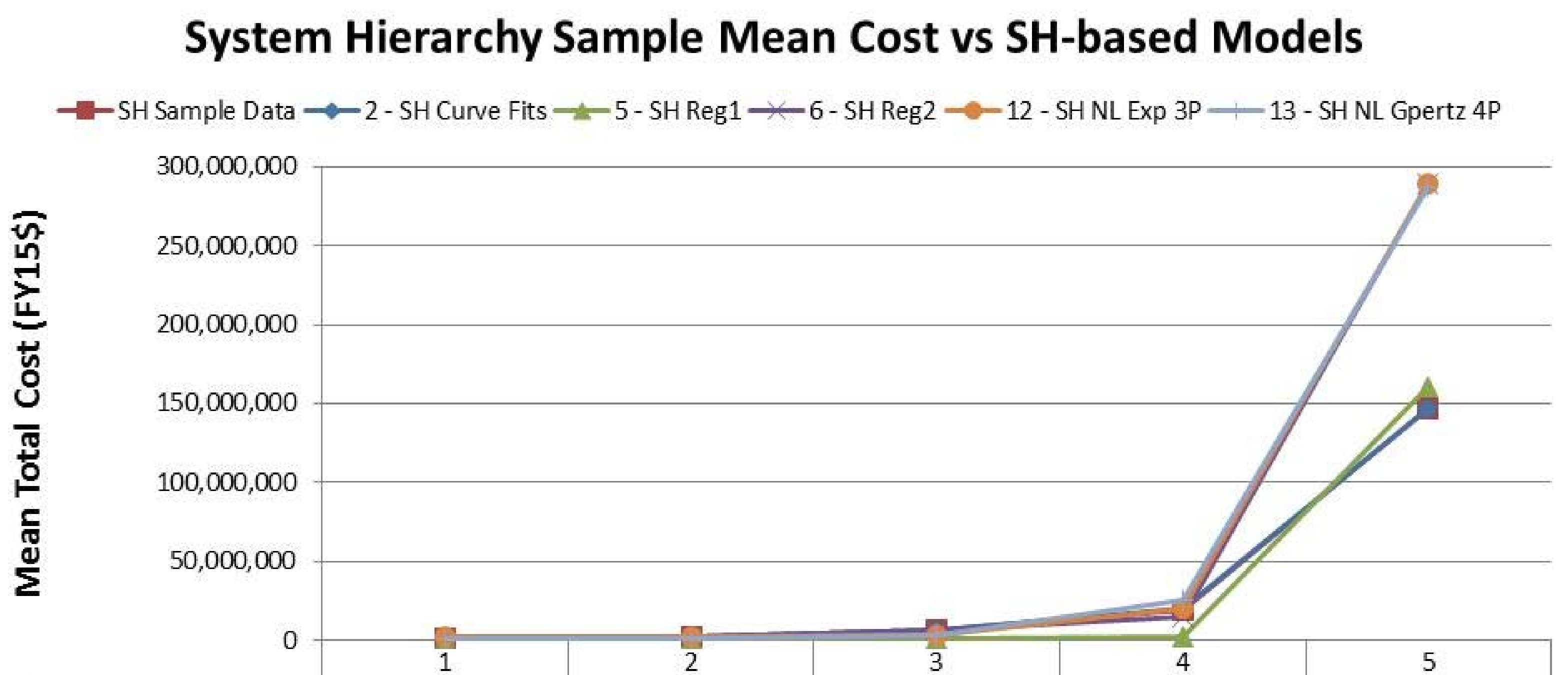
	1	2	3	4	5
TI Sample Data	5,385,091	13,715,880	41,948,963	117,418,045	171,881,182
1 - TI Curve Fits	6,098,593	14,886,701	41,949,255	113,853,889	171,879,982
3 - TI Reg1	5,385,091	8,989,340	41,948,963	173,522,584	254,272,144
4 - TI Reg2	5,385,091	13,715,880	41,948,963	59,337,393	314,353,878
10 - TI NL Quad	4,331,991	9,551,810	67,957,831	179,550,054	344,328,479
11 - TI NL Exp 2P	7,552,009	19,819,678	52,015,246	136,510,080	358,260,387

TRL Improvement Level

Note: Lines do not represent continuous functions but rather demonstrate the progression of model discrete ordinal values.

Cost Model Output - SH Level Cost Models vs Actual Data

SH Model nos. 2, 5, 6, 12 and 13 vs SH Sample Project Data (means)



	1	2	3	4	5
SH Sample Data	1,734,327	2,643,491	6,439,009	18,976,157	146,709,842
2 - SH Curve Fits	1,508,290	2,320,795	6,439,009	20,461,546	146,709,271
5 - SH Reg1	880,979	976,748	1,704,188	2,624,109	160,808,463
6 - SH Reg2	1,734,328	2,613,621	6,244,922	14,965,793	288,867,732
12 - SH NL Exp 3P	2,372,496	2,434,288	3,443,119	19,913,385	288,808,534
13 - SH NL Gpertz 4P	1,354,510	1,508,967	3,349,633	25,284,800	286,685,634

System Hierarchy Level

Cost Model Output - Summary Cost Curve Fit Model Statistics (Model nos. 1 & 2)

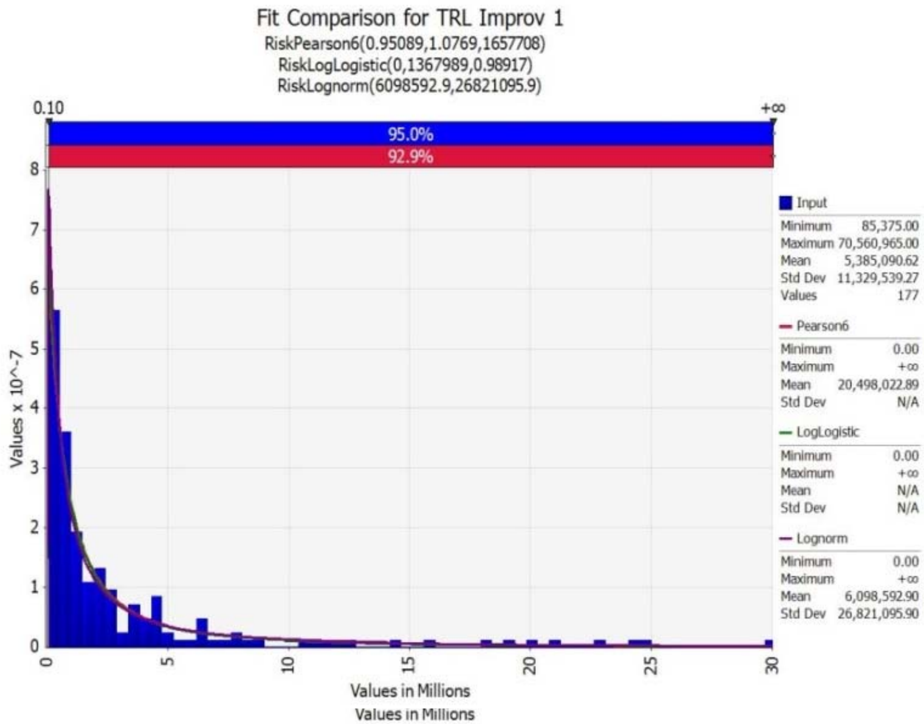
- **Lognormal, Gamma and LogLogistic functions produced best curve fit results across Models 1 (TI-level) and 2 (SH-level)**
- **Lognormal function selected for modeling uncertainty with regression and nonlinear models**

Predictor Level / Tier	Mean	Median	60th %ile	80st %ile	Curve Function Type
TRL Improvement Level					
TRL Improvement 1	6,098,593	1,352,186	2,098,994	5,827,153	Lognorm
TRL Improvement 2	14,886,701	2,937,018	4,636,000	13,379,843	Lognorm
TRL Improvement 3	41,949,255	17,585,237	28,194,724	68,557,068	Gamma
TRL Improvement 4	113,853,889	30,765,241	49,013,531	144,529,122	Lognorm
TRL Improvement 5	171,879,982	87,024,759	130,289,167	283,256,614	Gamma
Hierarchy Level					
Hardware / Software / Mat'l.	1,508,290	356,516	492,737	1,077,888	LogLogistic
Component / Part	2,320,795	427,230	600,295	1,366,661	LogLogistic
Assembly	6,439,009	855,392	1,308,794	3,661,668	LogLogistic
Subsystem	20,461,546	2,327,053	3,946,668	13,457,236	Lognorm
System	146,709,271	42,205,134	77,094,954	230,367,198	Gamma

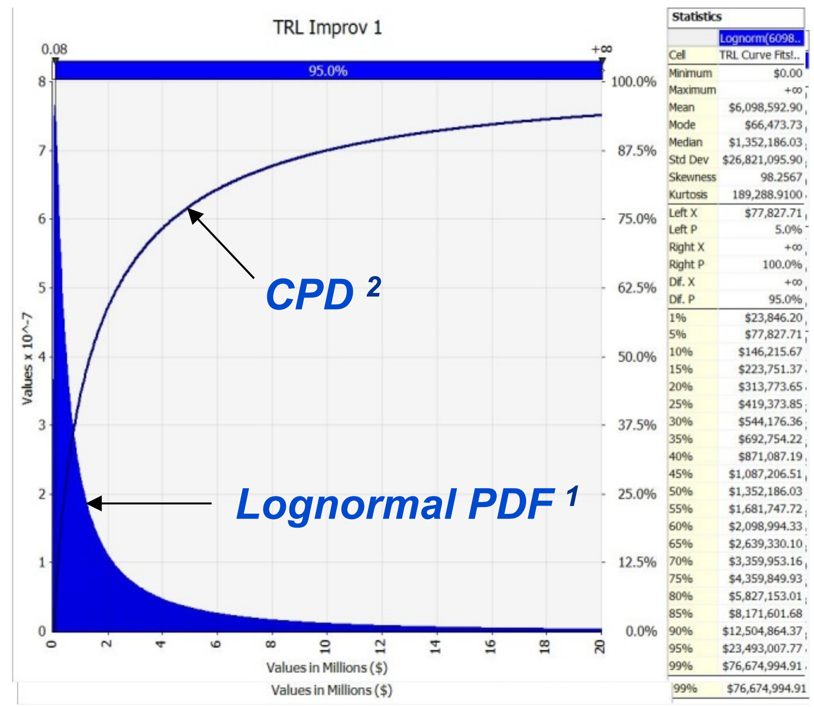
Curve fit model predicted costs in FY15$K

Cost Model Output - Sample Curve Fit Cost Model Uncertainty (Model no. 1)

Cost Model No. 1: TI Level 1 Sample Data with Higher Performing Curve Fits



Cost Model No. 1: TI Level 1 Selected Curve Fit Model



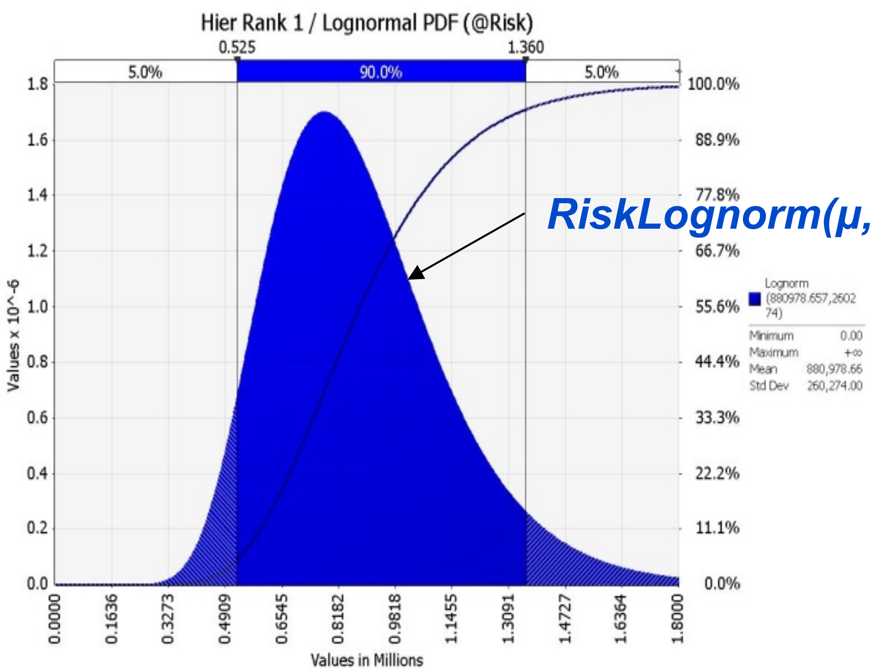
- Uncertainty PDFs were also developed for the *other* Curve Fit Model ordinal levels (see backup slides)
  - Cost Model no. 1 - TI Levels 2 to 5
  - Cost Model no. 2 - SH Levels 1 to 5

1. PDF - Probability Density Function
2. CPD - Cumulative Probability Distribution (a.k.a. ogive or "S-curve")

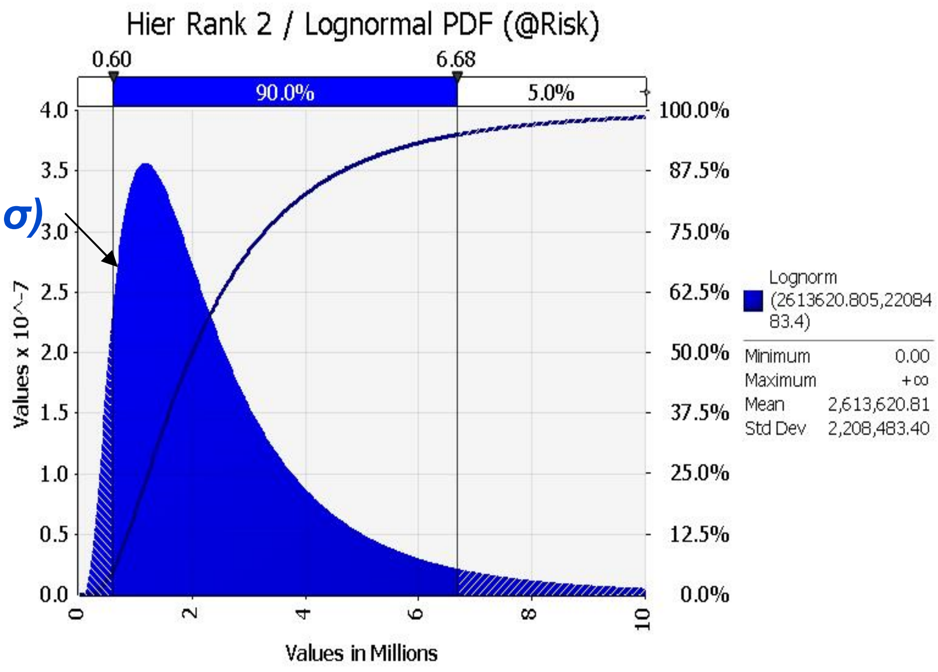
Cost Model Output - Sample Linear Regression Cost Model Uncertainty (Model nos. 5 & 6)

- **Sample Uncertainty PDF for Cost Model 5 at SH Level 1 and Cost Model 6 at SH Level 2 (FY$15)**
- **Uncertainty PDFs also developed for other SH Levels:**
  - **Cost Model 5 SH Levels 2, 3, 4, 5 and Cost Model 6 SH levels 1, 3, 4, 5**

Cost Model No. 5: SH level 1



Cost Model No. 6: SH level 2



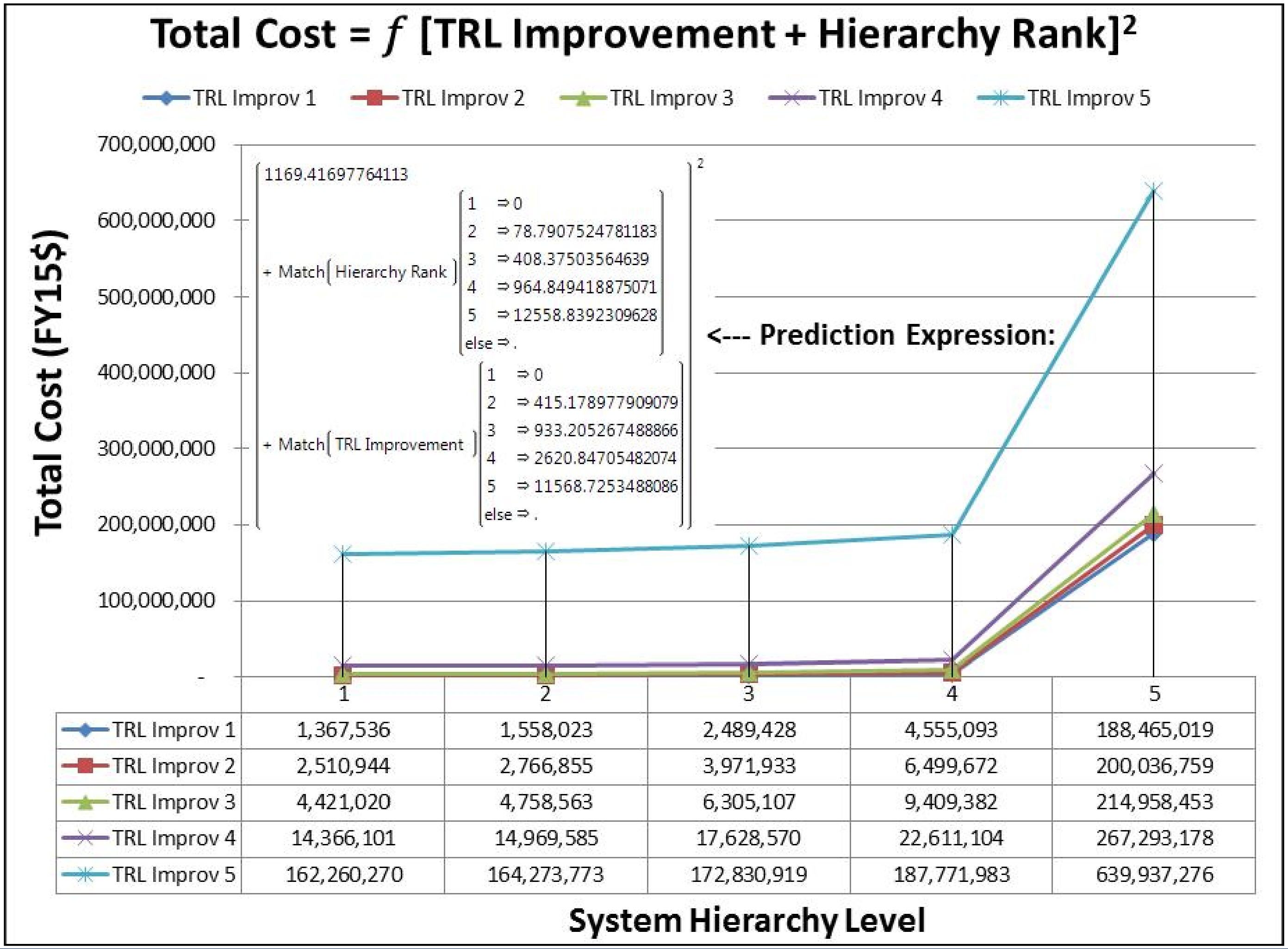
RiskLognorm(μ,σ)

Cost Model Output - Sample Multivariate Regression Cost Model (Model no. 9)*

- **Transformation – *ƒ* [squared ∑ of predictor variables]**
  - **Cost = c + (a • TI level + b • SH level)^2] (expression below)**

TI level more dominant than SH level based upon coefficients and total response at same ordinal levels up through level 4.

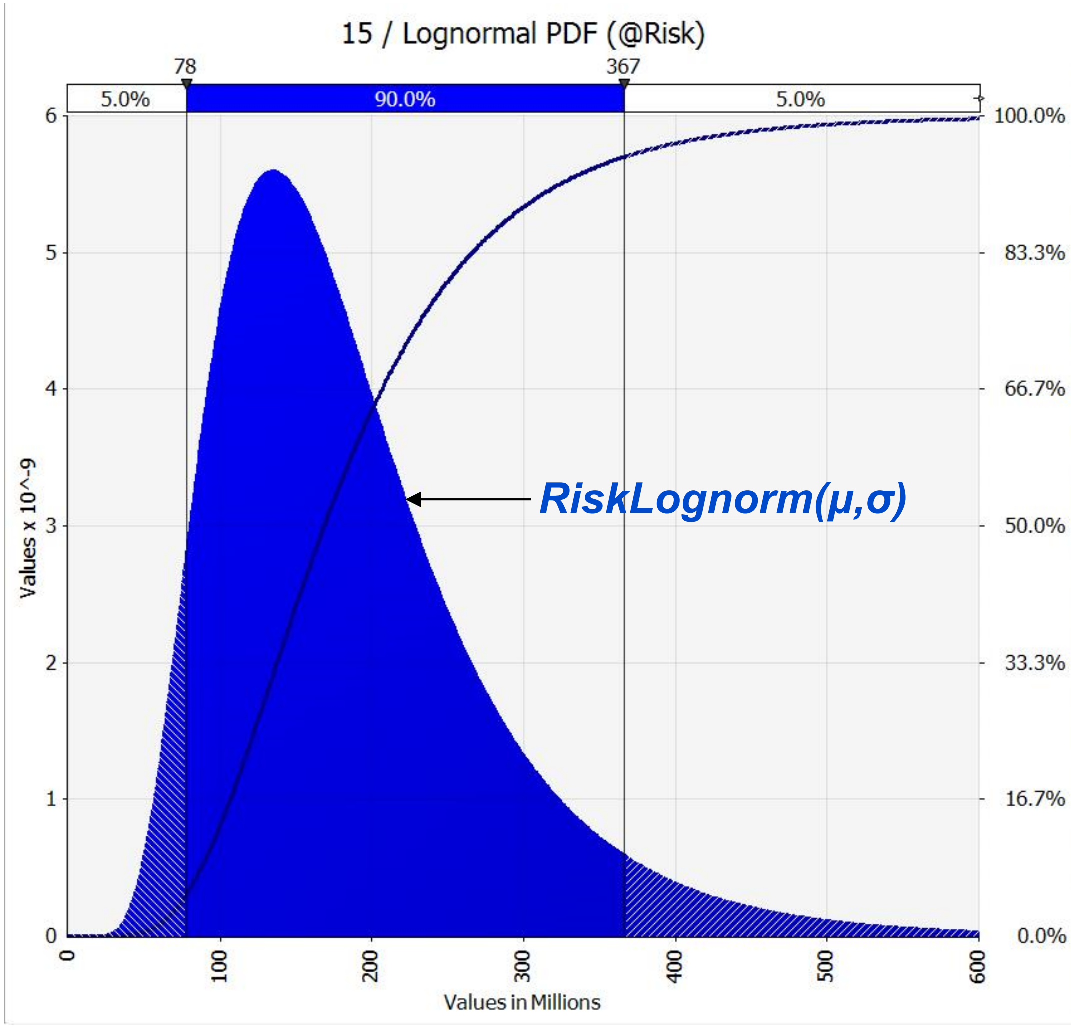
** Other Multivariate Regression Models 7 & 8 provided in backup slides*



	1	2	3	4	5
TRL Improv 1	1,367,536	1,558,023	2,489,428	4,555,093	188,465,019
TRL Improv 2	2,510,944	2,766,855	3,971,933	6,499,672	200,036,759
TRL Improv 3	4,421,020	4,758,563	6,305,107	9,409,382	214,958,453
TRL Improv 4	14,366,101	14,969,585	17,628,570	22,611,104	267,293,178
TRL Improv 5	162,260,270	164,273,773	172,830,919	187,771,983	639,937,276

Cost Model Output - Sample Multivariate Regression Cost Model Uncertainty (Model No. 9)

- **Composite Linear Regression (Model No. 9) – Sample PDF Uncertainty for TI Level 1 and SH Level 5 (24 other TI / SH Level PDFs also created)**

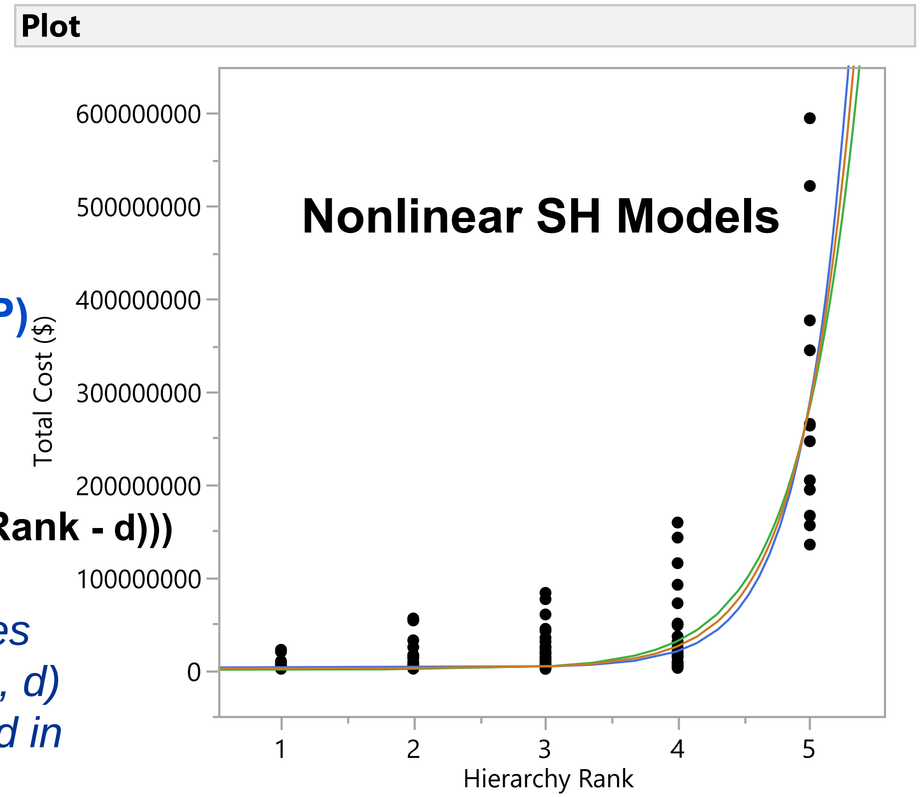


Statistics	Lognorm(1884..
Cell	TRL-Hier Sqrd R..
Minimum	0.00
Maximum	+∞
Mean	188,465,018.50
Mode	134,909,404.84
Median	168,591,073.65
Std Dev	94,168,670.48
Skewness	1.6237
Kurtosis	8.0268
Left X	77,552,163.35
Left P	5.0%
Right X	366,501,060.54
Right P	95.0%
Dif. X	288,948,897.19
Dif. P	90.0%
1%	56,217,130.79
5%	77,552,163.35
10%	92,062,110.35
15%	103,356,300.82
20%	113,312,770.39
25%	122,615,522.16
30%	131,618,795.63
35%	140,550,755.12
40%	149,586,150.71
45%	158,880,468.44
50%	168,591,073.65
55%	178,895,180.73
60%	190,010,572.37
65%	202,225,524.09
70%	215,949,021.41
75%	231,805,481.20
80%	250,836,247.47
85%	274,999,684.49
90%	308,736,677.98
95%	366,501,060.54
99%	505,592,329.53

Cost Model Output – Nonlinear Cost Models (nos. 10, 12 & 13)

- **Nonlinear TI Models [1, 2]**
  - **Cost Model no. 10 (TRL NL-Quadratic)**
    - **Cost = a + b x TI level + c x TI Level²**
- **Nonlinear SH Models [2]**
  - **Cost Model no. 12 (Hier NL-Exponential 3P)**
    - **Cost = a + b x exp(c x Hierarchy Rank)**
  - **Cost Model no. 13 (Hier NL-Gompertz 4P)**
    - **Cost = a + (b - a) x exp(-exp(-c x(Hierarchy Rank - d)))**

1. *Plot of TI models provided in backup slides*
2. *Values for expression parameters (a, b, c, d) and graph for all NL Cost models provided in backup slides*



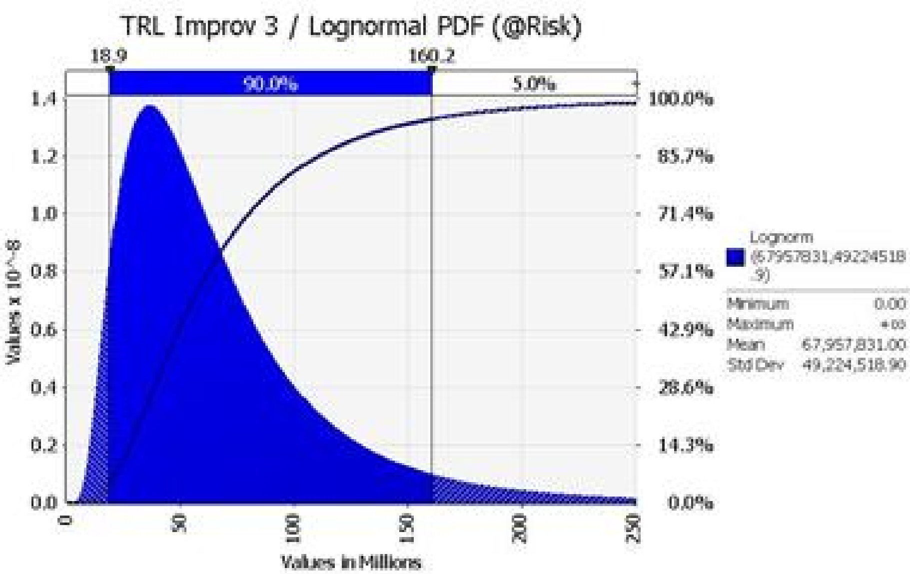
Nonlinear SH Cost Models

Model Comparison

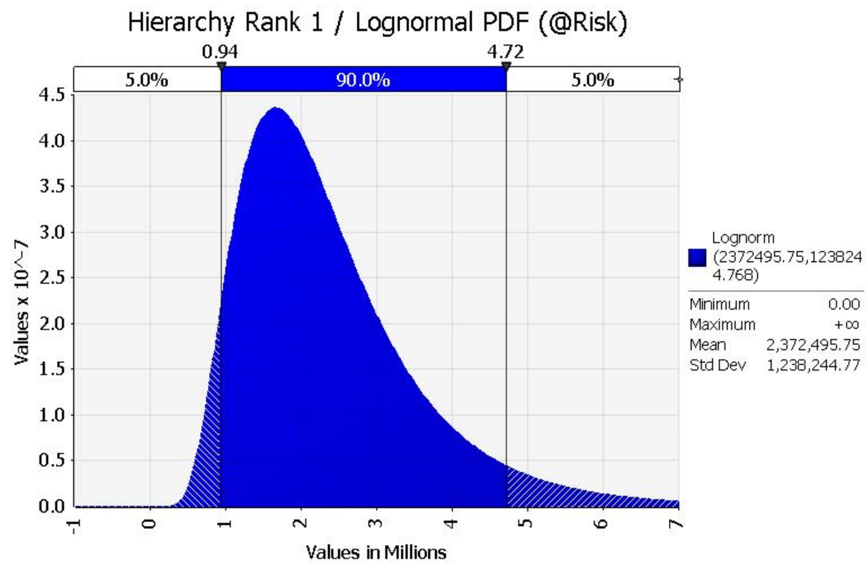
Model	AICc	AICc Weight	BIC	SSE	MSE	RMSE	R-Square
Exponential 3P	19748.669	0.9262116	19765.723	3.316e+17	6.233e+14	24966011	0.7436299
Gompertz 4P	19753.763	0.0725263	19775.061	3.335e+17	6.281e+14	25060990	0.7421611
Logistic 3P	19761.866	0.001262	19778.919	3.399e+17	6.389e+14	25275833	0.7372274

Cost Model Output – Nonlinear Cost Model Uncertainty (nos. 10 & 12)

NL Cost Model No. 10:
TI Level 3 Example PDF



NL Cost Model No. 12:
SH Level 1 Example PDF



- **All other Nonlinear Cost Model Ordinal Level Uncertainty PDFs (15 in total) also developed**
  - **Cost Model no. 10 (Quadratic) - TI levels 1, 2, 4, 5**
  - **Cost Model no. 12 (Exponential 3P) - SH levels 2 to 5**
  - **Cost Model no. 13 (Gompertz 4P) - SH levels 1 to 5**

Cost Model General Applicability

Mdl. No.	Model Form / Method	Predictor Variable Form	Model Performance and Technology Development Attributes				
			Best Project Sample Data Fit	Generally Higher KPM Performance	System Level Development (SH level 5)	Below System Level Development (SH Level 1-4)	Generally Higher Cost or Uncertainty Levels*
1	Tailored Curve Fits	TI Level	✓			✓	
2	Tailored Curve Fits	SH Level	✓		✓		
5	Simple Linear Regression	SH Level			✓		
6	Simple Linear Regression	SH Level			✓		✓
7	Composite Linear Regression	$[TI \times SH]^2$		✓	✓	✓	
8	Multiple Linear Regression	TI + SH		✓	✓	✓	✓
9	Multiple Linear Regression	$[TI + SH]^2$	✓	✓	✓	✓	✓
10	Nonlinear - Quadratic	NL TI Level				✓	✓
12	Nonlinear - Exponential 3P	NL SH Level			✓		✓
13	Nonlinear - Gompertz 4P	NL SH Level			✓		✓

** May be more applicable for higher risk or technology volatility developments*

Schedule Model Performance

- The same model forms were developed and assessed for schedule-based modeling
  - Ordinal Curve Fits, Linear Regression (Univariate and Multivariate including a range of transformations) and Non-linear
  - Dependent variable - Development Project Duration (months)
  - Independent Predictor variables - TI Level, SH Level and Project Spend Rate (investment $/mo.) added to augment analysis
- Results did not produce the same strength of relationship with the independent predictor variables as experienced with Cost
  - Exception was SH Level Curve Fit model (available KPM below)

Model No.	Fit Model Type	Single / Multiple Predictor Variable(s)	Predictor Type	Reference Model Name	Predictor Variable(s)	Key Performance Measures (KPM)				
						RMSE (months)	Coef. of Variation (CV)	No. of Available Obs.	No. of Applied Obs.	Data Reduction (%)
1	Tailored Curve Fits	Single	System Hierarchy Level	Hier Curve Fits	Hierarchy Rank	20	0.755	551	551	0.0%

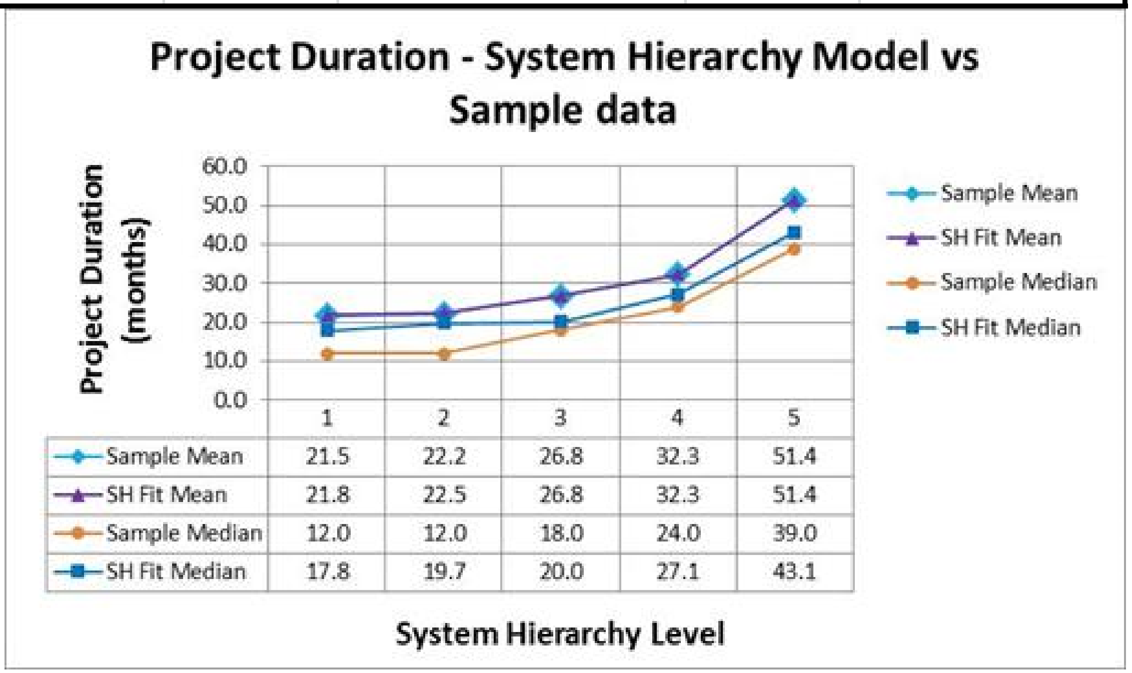
PERFORMANCE RATING			
Good	Fair	Marginal	Poor

Schedule Model Output - SH Level Schedule Model

Schedule Duration (months) vs. System Hierarchy Level Curve Fit

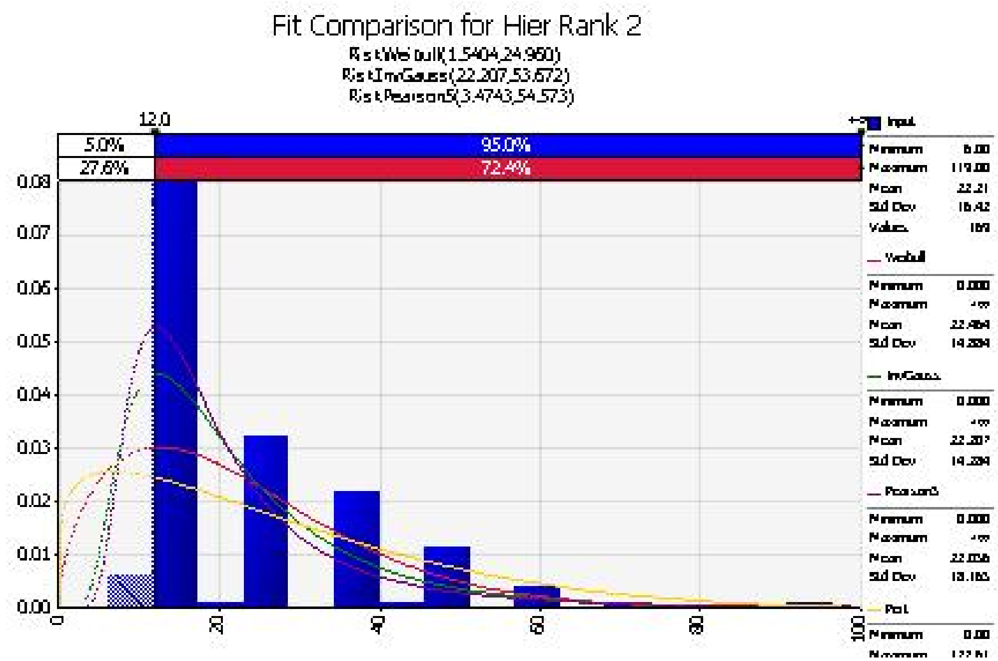
Predictor Level / Tier	Number of Observations	Mean	Median	60th %ile	80th %ile	Curve Function Type
System Hierarchy Level						
Hardware / Software / Mat'l.	98	21.8	17.8	20.4	28.5	Pearson5
Component / Part	169	22.5	19.7	23.6	34.0	Weibull
Assembly	173	26.8	20.0	24.4	38.6	InvGauss
Subsystem	86	32.3	27.1	32.7	48.3	Erlang
System	25	51.4	43.1	52.0	77.0	Erlang
	551	**0.0%**	**Data Reduction**			

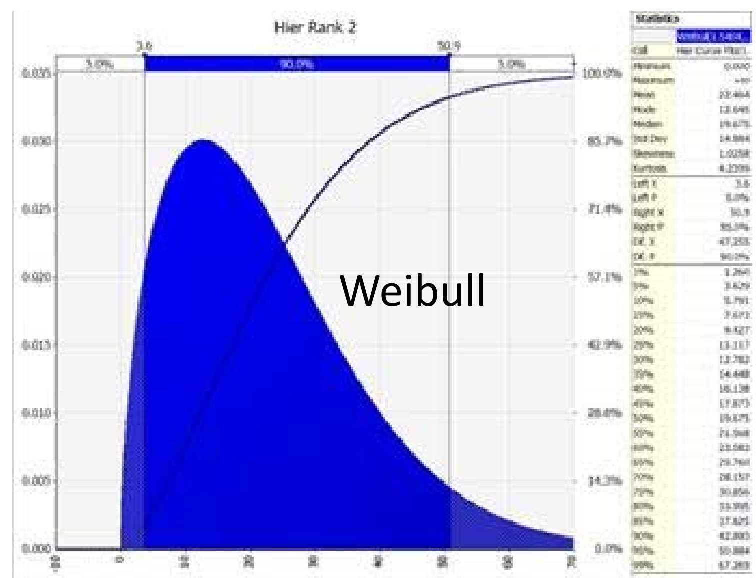
- **Consistent cost growth across key benchmark levels with best results from Pearson5, Weibull, Inverse Gaussian and Erlang distributions**
- **Summary chart demonstrates the closeness of fit to sample mean**



Schedule Model Output - Sample Curve Fit Model Uncertainty

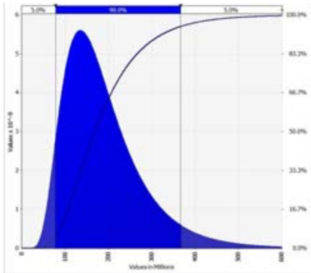
Example Schedule Curve Fits & Selected PDF – Project Duration (months) for System Hierarchy Level 2





Cost and Schedule Model Uncertainty Drivers

- **Range of Technologies in Project Data**
  - **Diverse TAs found in the database may contain varying considerations for R&D activities that can drive both cost and schedule**
- **TRL and SH Level Assessment Variability**
  - **TRL and SH level assessments are subjective qualitative valuations that can vary by source**



- **Cost Data Variability / Normalization**
  - **Scope and tracking of budgeting, cost accounting methods / categories, contractual CLINs, and indirect costs captured can vary across projects**
- **Source Data Characteristics**
  - **Data sample sizes are good but unexpected overabundance of smaller projects across higher predictor variable levels**
- **Model Forms**
  - **Output variability between or across model forms can be related to the nature of particular model relationship characteristics or constraints**

Conclusions and Future Work

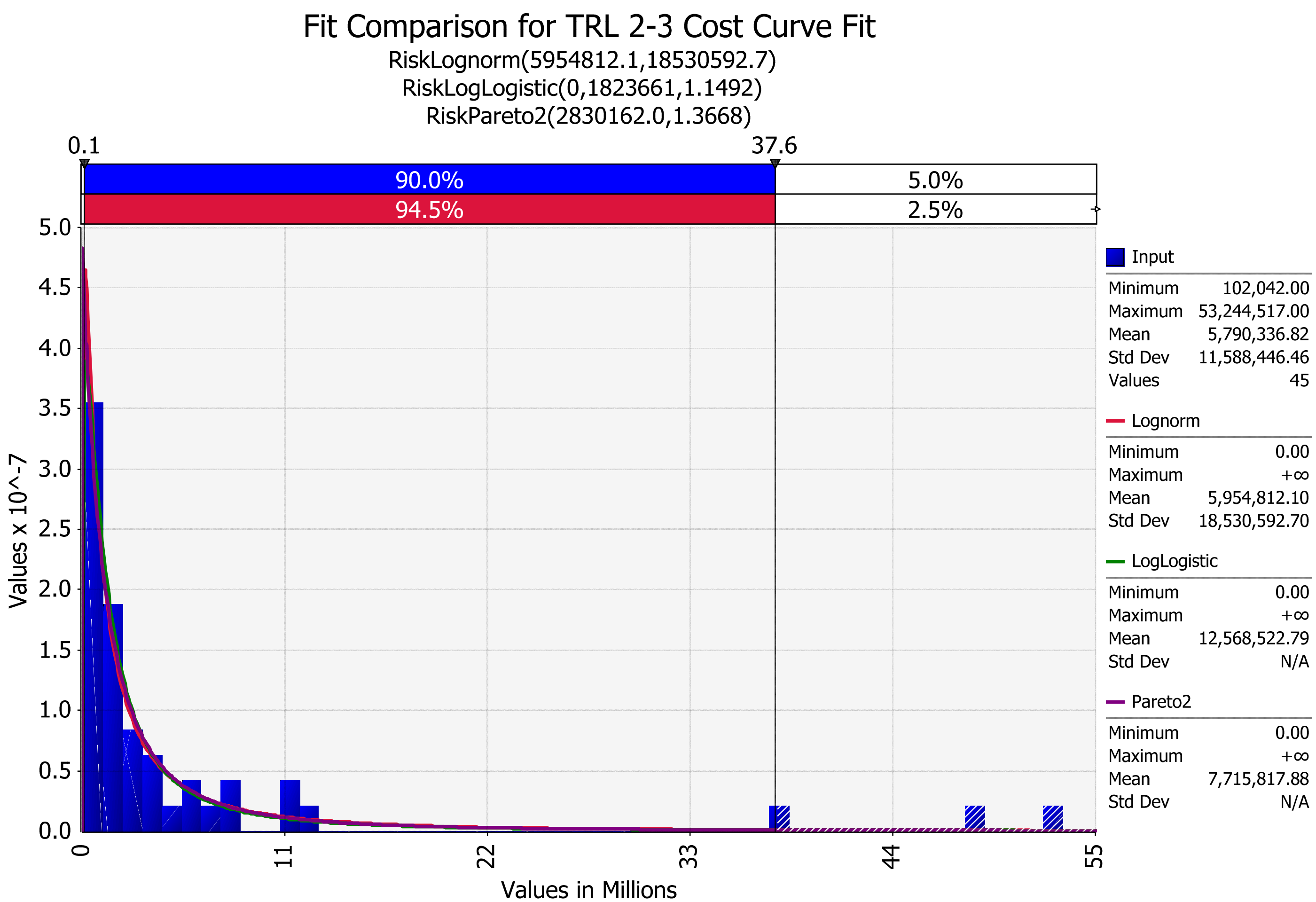
- TI and SH macro variable models produced good statistical KPM and goodness-of-fit characteristics but w/ significant variability
- Deliver forecasting value above very ROM estimates and SME opinion often applied in early technology development
- Other "macro-level" cost & schedule parameters to consider for early stage technology development estimating:
  - Research and Development Degree of Difficulty (RD3)
  - Capability Demonstrations
  - Advanced Degree of Difficulty (AD2)
  - System Readiness Level (SRL)
  - Integration Readiness Levels (IRL)
  - Implementation Readiness Level (ImpRL)
  - Manufacturing Readiness Level (MRL)
  - Macro-level technology performance or complexity factors

Questions?

email: chuck.alexander@jhuapl.edu

Additional Information

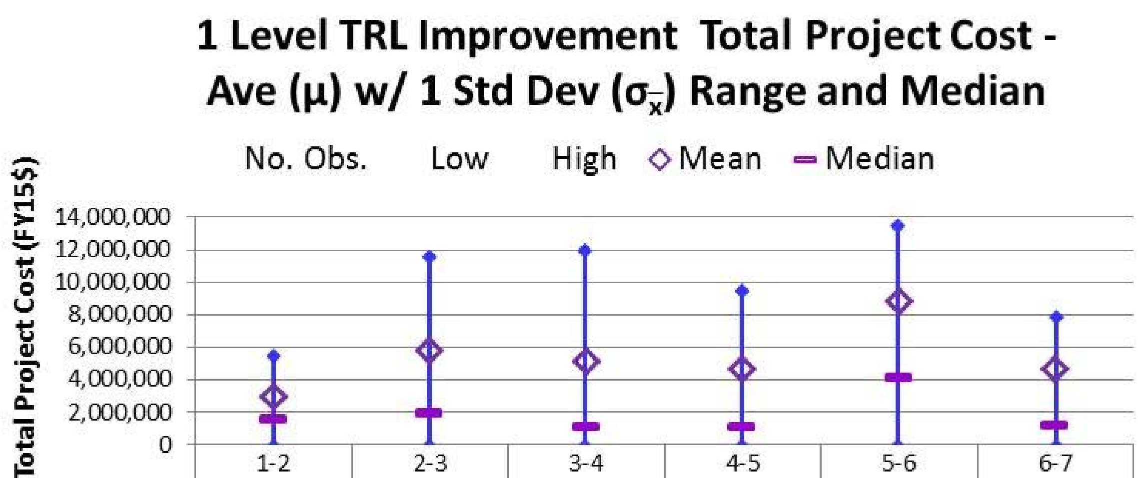
Sample TRL X-Y Transition Data for TRL 2-3



	Input
Minimum	102,042
Maximum	53,244,517
Mean	5,790,337
Mode	≈102,093.00
Median	1,846,495
Std Dev	11,588,446
Skewness	3.3134
Kurtosis	13.6324
Left X	102,148
Left P	5.00%
Right X	37,599,597
Right P	95.00%
Dif. X	37,497,449
Dif. P	90.00%
1%	102,042
5%	102,148
10%	138,048
15%	323,559
20%	482,812
25%	917,992
30%	977,118
35%	985,383
40%	1,300,893
45%	1,564,995
50%	1,846,495
55%	1,952,341
60%	2,400,217
65%	2,676,712
70%	4,132,726
75%	4,588,650
80%	6,425,581
85%	8,427,232
90%	11,368,804
95%	37,599,597
99%	53,244,517

Data Modeling – Cost vs. TI Level Relationship Screening

- **Data relationship screening for Cost vs. TI level showed stability of a relevant range across TRL Start-End (TRL X-Y) levels**
- **Representative example plot for for TI Level 1 is shown below (TI Levels 2 through 5 also assessed with similar results)**
- **This analysis plus other screening techniques supported the use of TI level as an independent predictor variable**



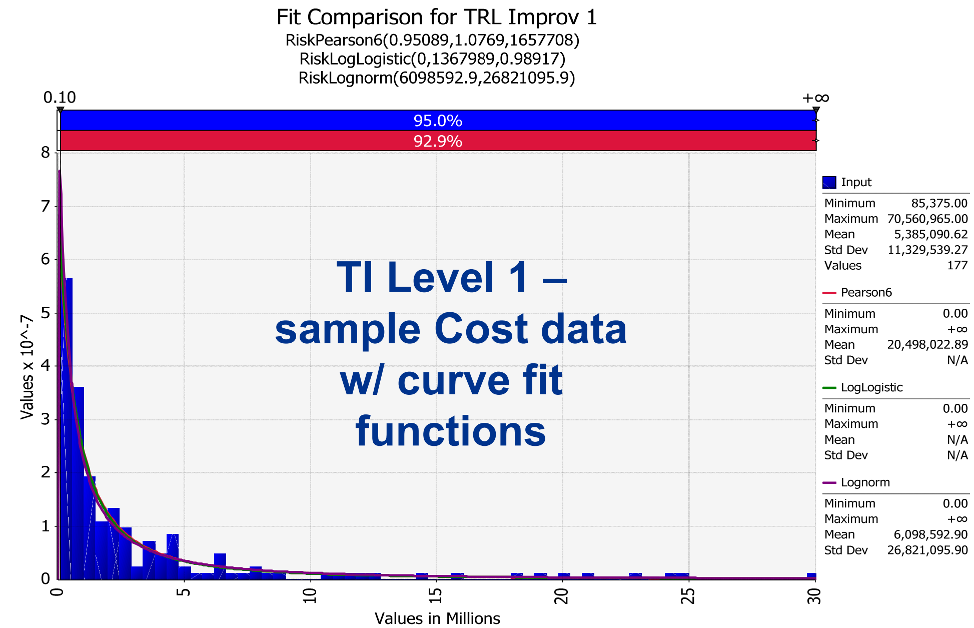
	1-2	2-3	3-4	4-5	5-6	6-7
No. Obs.	20	45	66	17	20	8
Low	$350,962	$-	$-	$-	$4,028,430	$1,370,923
High	5,583,833	11,584,560	12,024,121	9,498,699	13,561,723	7,904,513
Mean	$2,967,398	$5,790,337	$5,160,085	$4,680,995	$8,795,076	$4,637,718
Median	$1,477,615	$1,846,495	$988,260	$1,034,116	$4,087,391	$1,097,893

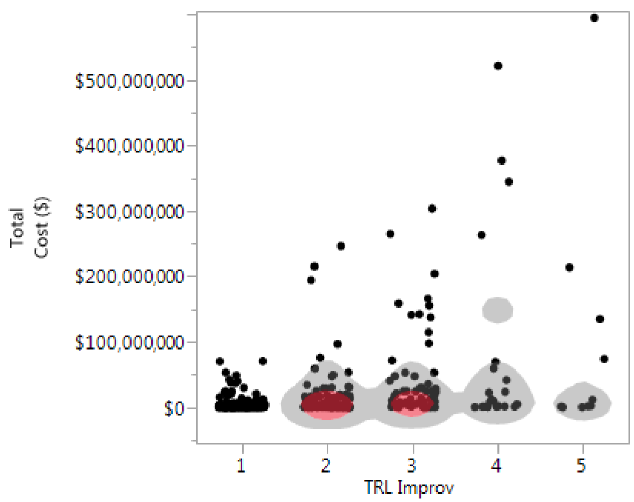
TRL Start-End (X-Y) Level

Data Modeling – Cost vs. TI Level Relationship

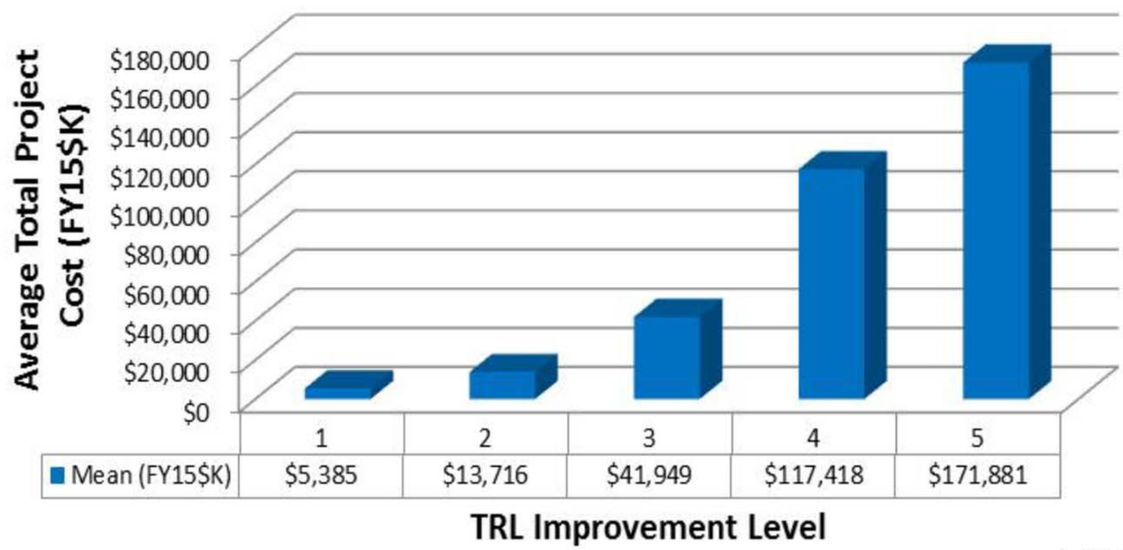
Project Cost (mean) vs. TI Level

- 405 project record dataset
- Direct relationship of Cost to TI level evident
- Geometric cost growth up through level 4, tapering off at level 5





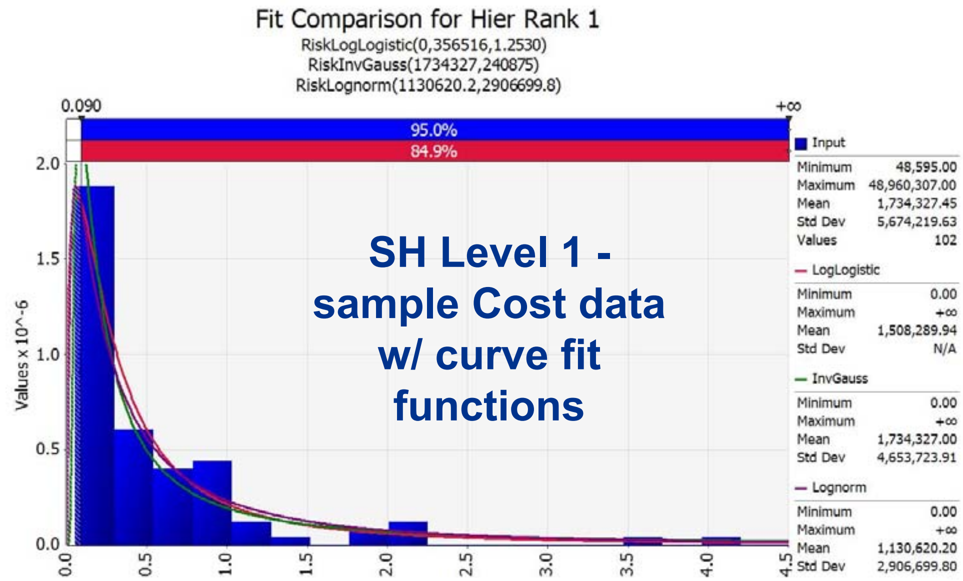
Scatterplot : Total Project Cost vs TI Level

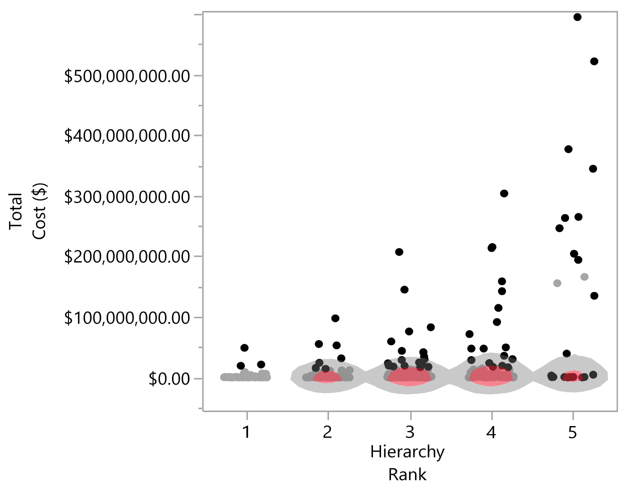


Data Modeling – Cost vs. SH Level Relationship

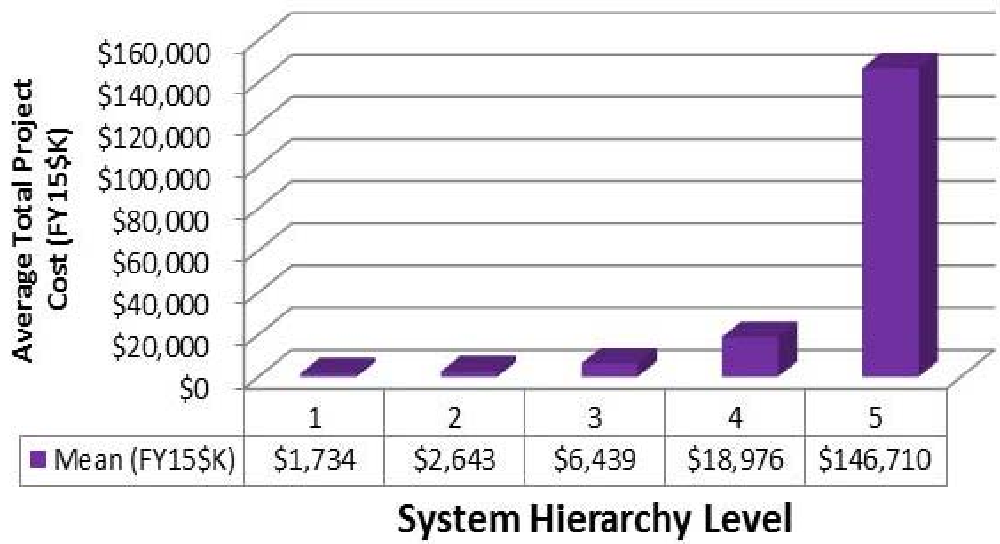
Project Cost (mean) vs. SH Level

- 603 project record dataset
- Gradual moderate growth up to Subsystem level (4)
- Dramatic increase at the System level (5) suggests possible exponential relationship





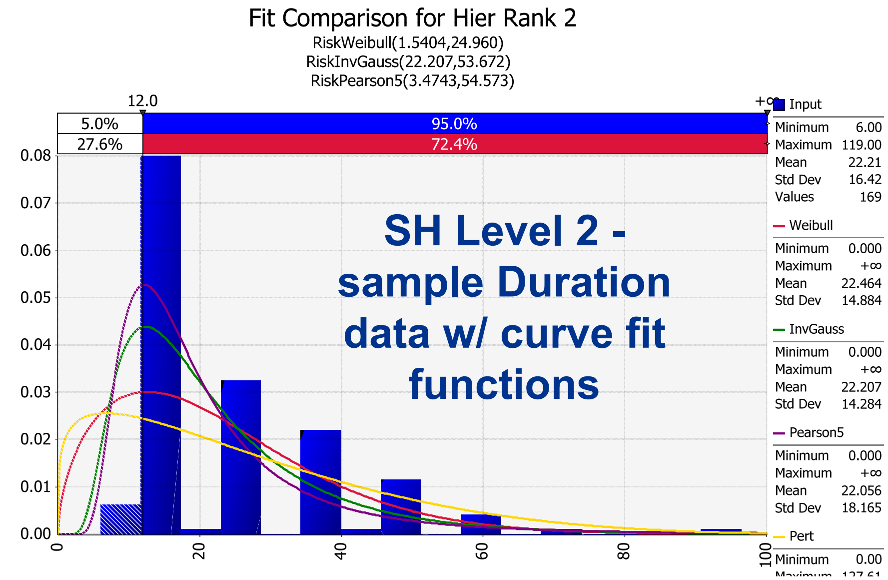
Scatterplot : Total Project Cost vs SH Level

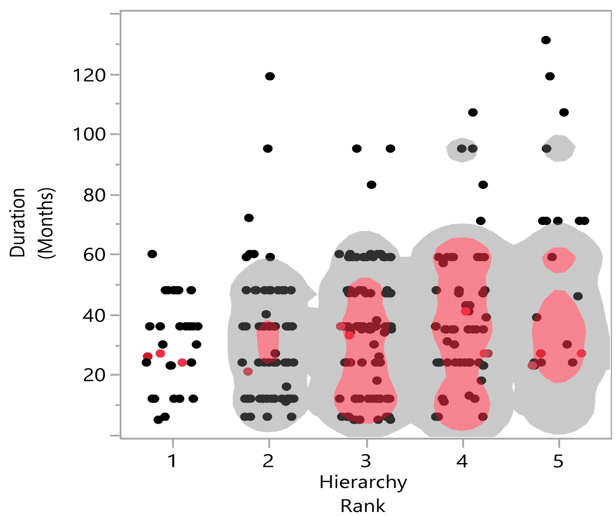


Data Modeling – Schedule vs. SH Level Relationship

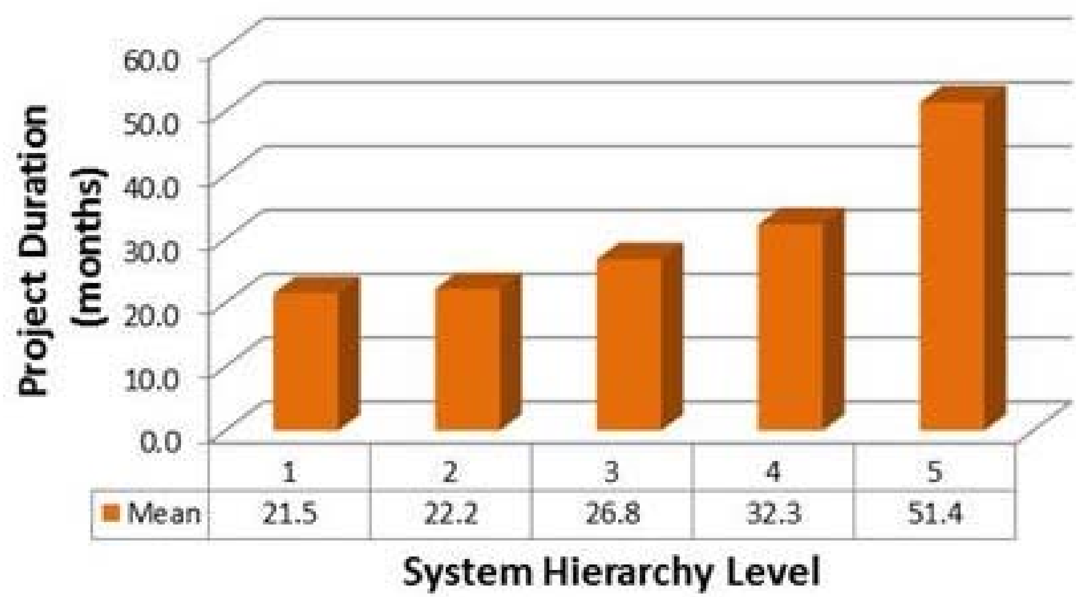
Project Duration (mean) vs. SH Level

- 551 project record dataset
- Direct relationship between SH level and Project Duration indicated





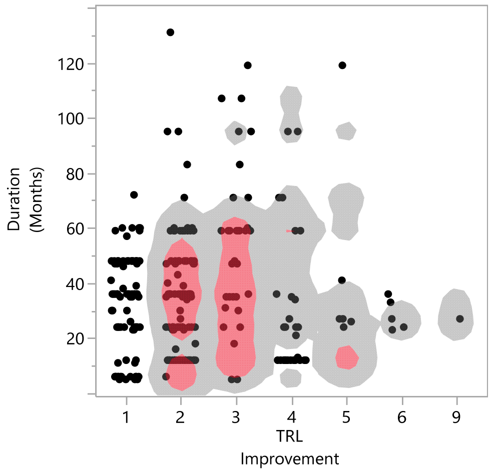
Scatterplot : Project Duration vs SH Level



Data Modeling – Schedule vs. TI Level Relationship

- **A poor affiliation between Project Duration and TI Level is indicated**
  - **Random data distribution, lack of obvious visual patterns, substantial nonparametric density areas**
  - **Moderate data correlation (r = 0.3238)**
  - **Columnar chart suggesting the mean project duration does not possess a continuous functional association with the TI level, peaking and then tailing off at level 3**

Scatterplot :Project Duration vs TI Level



Data Modeling – Cost and Schedule Model Datasets

- **Data set size limited by projects with valid[1] corresponding predictor and response variable values**
  - **TRL Start and End levels**
  - **System Hierarchy Level**
  - **Total Project Development Cost**
  - **Project Start and End Dates**
- **Resulting data sets available for modeling[2]**
  - **Total Project Cost vs. TI Level (405 / 395 available project records for cost / schedule models)**
  - **Total Project Cost vs. SH Level (603 / 551 available project records for cost / schedule models)**
  - **Total Project Cost vs. TI Level and SH Level (221 available project records for both cost and schedule models)**

1. Project records with zero, blank or erroneous values removed

2. Not all records for each data set had available project start or end dates so total number of records for schedule duration modeling was slightly less

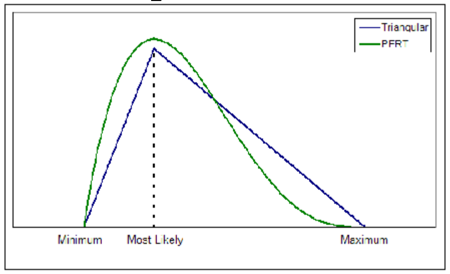
Key Performance Measure (KPM) Descriptions

- **Error Variability and Dispersion Measures:**
  - **Coefficient of Determination - R^2 and Adjusted R^2. Most commonly used measure of "goodness of fit". Relative measure of fit equal to the percent of the variation in the dependent variable (Y) explained by the independent variable (X) = SSR[1] / SST.**
  - **Root Mean Square Error (RMSE) – absolute measure of fit or accuracy based upon the differences between sample and population values predicted by a model.**
  - **Coefficient of Variation (CV) – RMSE for models, as applied here (vs. Standard Deviation used for individual variables), divided by mean of the Y-data, a unitless relative measure of estimating error (CV < 1 is considered low-variance and CV > 1 considered high)**
- **Statistical Significance Measures:**
  - **F-ratio - tests if the entire regression equation is valid (i.e., how well the statistical model is fitted to a sample data set).**
  - **t-stat - tests if the individual hypothesized predictor (X-variables) values are valid. t-stat represents the calculated difference represented in units of standard error. The % of expression terms with probability > |t| was applied as an overall measure.**
- **Autocorrelation Measure:**
  - **Durbin-Watson test - measures independence of regression residuals.**
- **Data Reduction Measure:**
  - **Percent (%) of original data sample set unused. The extent of selectivity in actual data set applied, measured as the % of available sample observations filtered out due to outliers, large residuals or non-core data, etc.**

1. SSR represents the sum of squares due to the regression and SST represents the sum of squares total.

Modeling Uncertainty

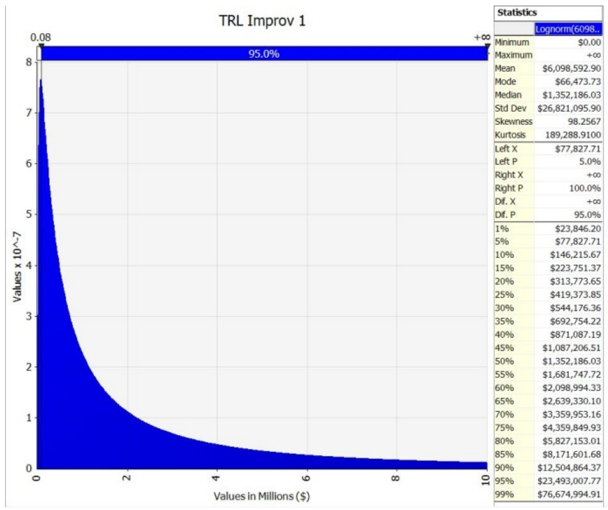
- **Uncertainty function evaluation**
  - **Lognormal, PERT, normal and triangular uncertainty PDFs evaluated**
    - **Inputs necessary to drive these functions (e.g., sample mean, min, max, mode, standard deviation, etc.) were available in most cases**
  - **Significant right-skewed PDFs found for actual Cost and Schedule TI and SH ordinal level sample data**
  - **Lognormal, Gamma and LogLogistic functions were generally highest performing across curve fits within the relevant data range**
- **Uncertainty function selection**
  - **PERT and Lognormal functions are generally considered superior to the triangular and normal distributions for modeling cost uncertainty**
    - Symmetrical normal function poor fit for right skewed data
    - Lognormal and PERT functions deliver natural, continuous distributions with less tendency to overemphasize direction of skew within normal planning range (50^{th} to 80^{th} %ile)

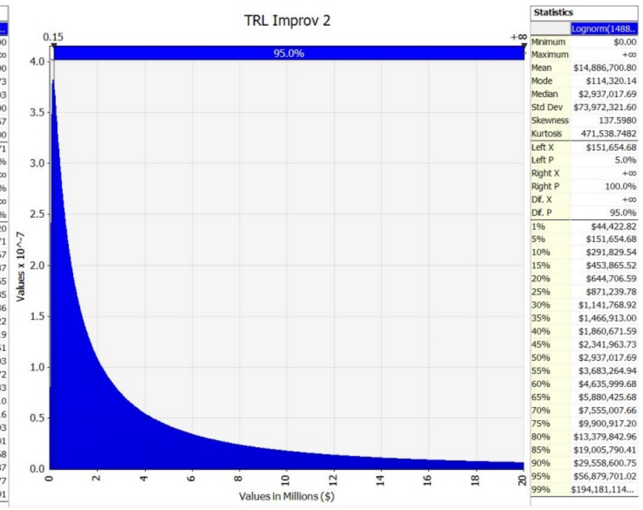


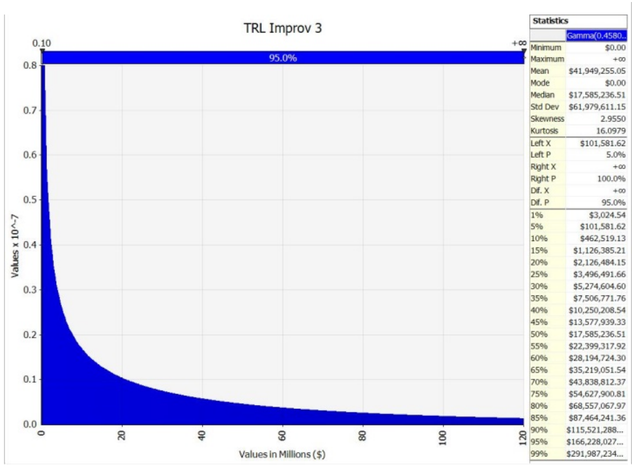
  - **Lognormal function was generally high performing across curve fits and also closely resembled other high performing Gamma and LogLogistic functions within the planning range**

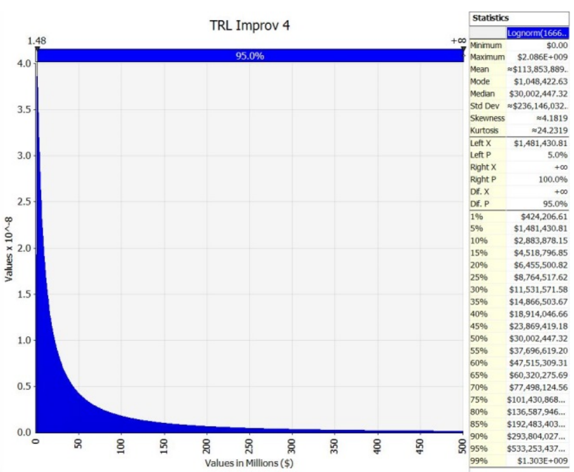
Cost Model No. 1 Selected Curve Fits – Total Project Cost (FY15$k) vs TRL Improvement Level

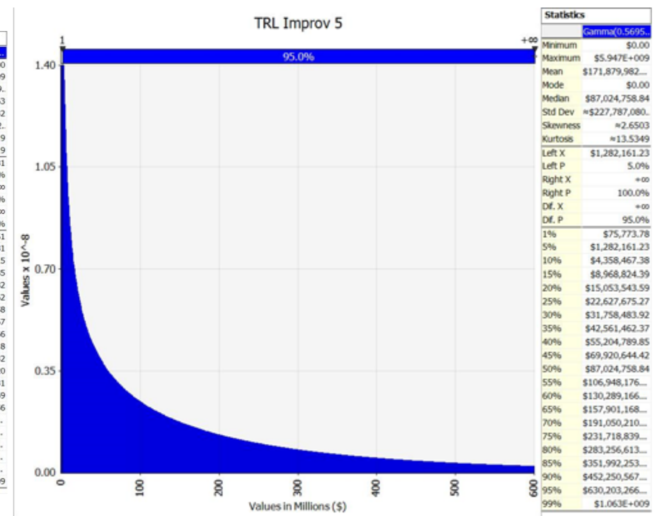
- **TI Level Curve Fit model PDFs with ventiles by ordinal level**
  - **TI levels 1, 2, 4 - LogLogistic**
  - **TI levels 3 & 5 - Gamma**





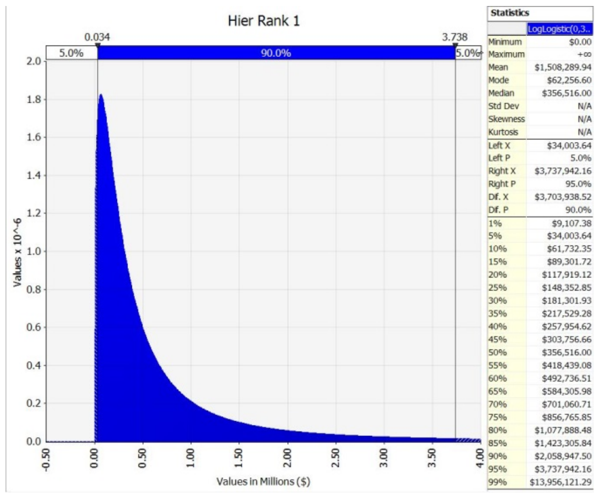


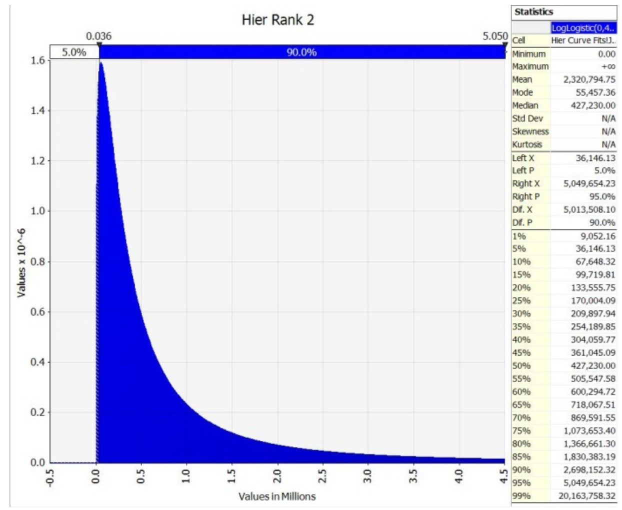


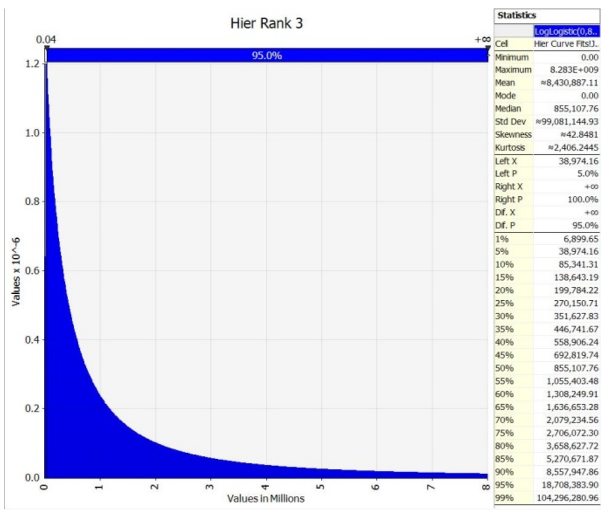


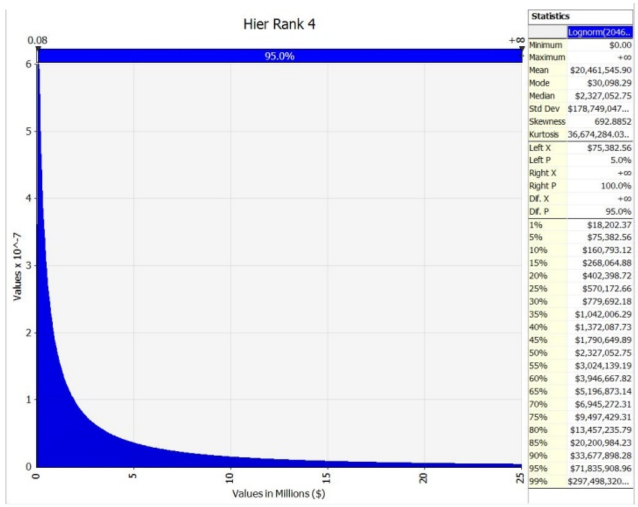
Cost Model No. 2 Selected Curve Fits – Total Project Cost (FY15$k) vs System Hierarchy Level

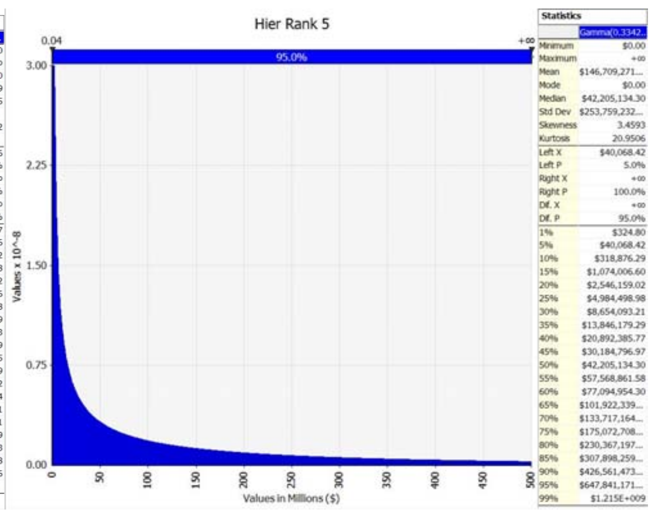
- **SH Level Curve Fit model PDFs with ventiles by ordinal level**
  - SH levels 1, 2, 3 - LogLogistic
  - SH level 4 - Lognormal
  - SH level 5 - Gamma





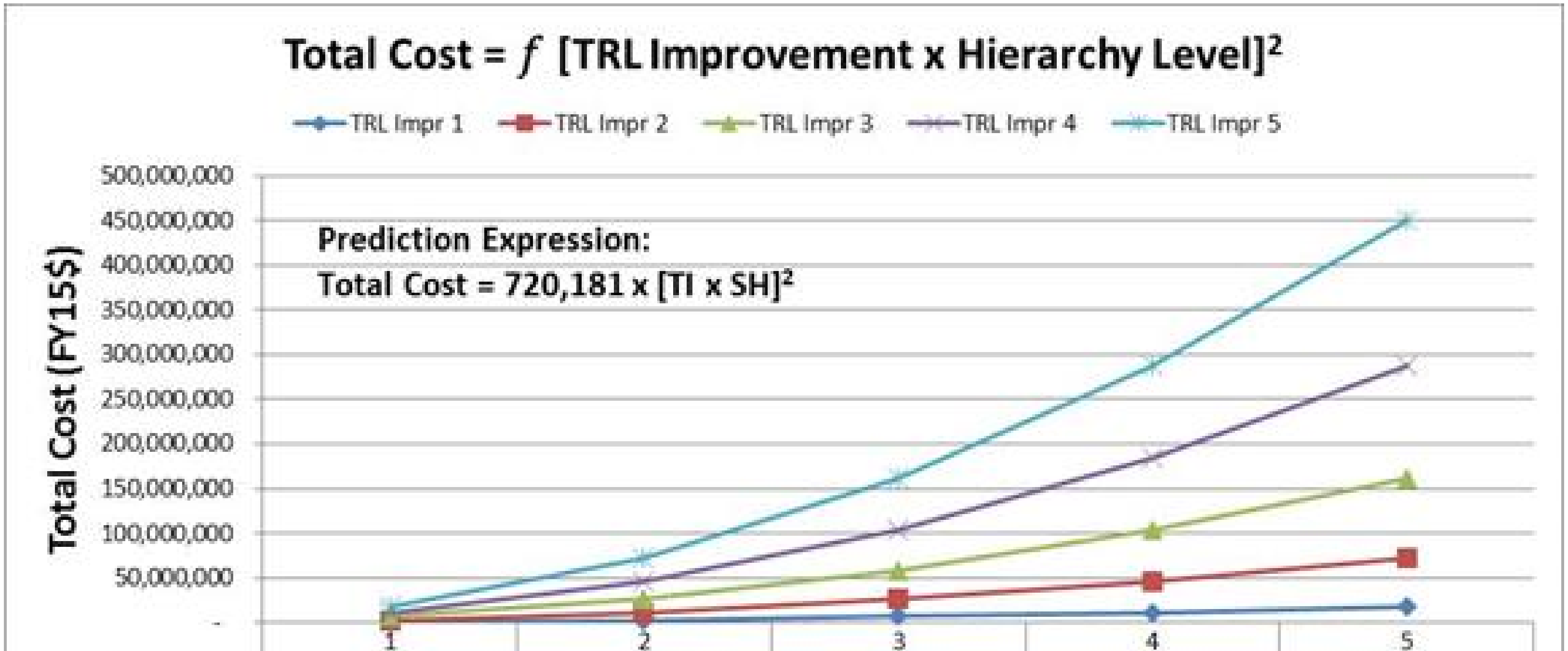






Multivariate Models – Multiple Regression Model no. 7

- **Product of predictors squared transformation (expression below)**
  - **Total Cost = f [TRL Improvement x Hierarchy Level]^2**

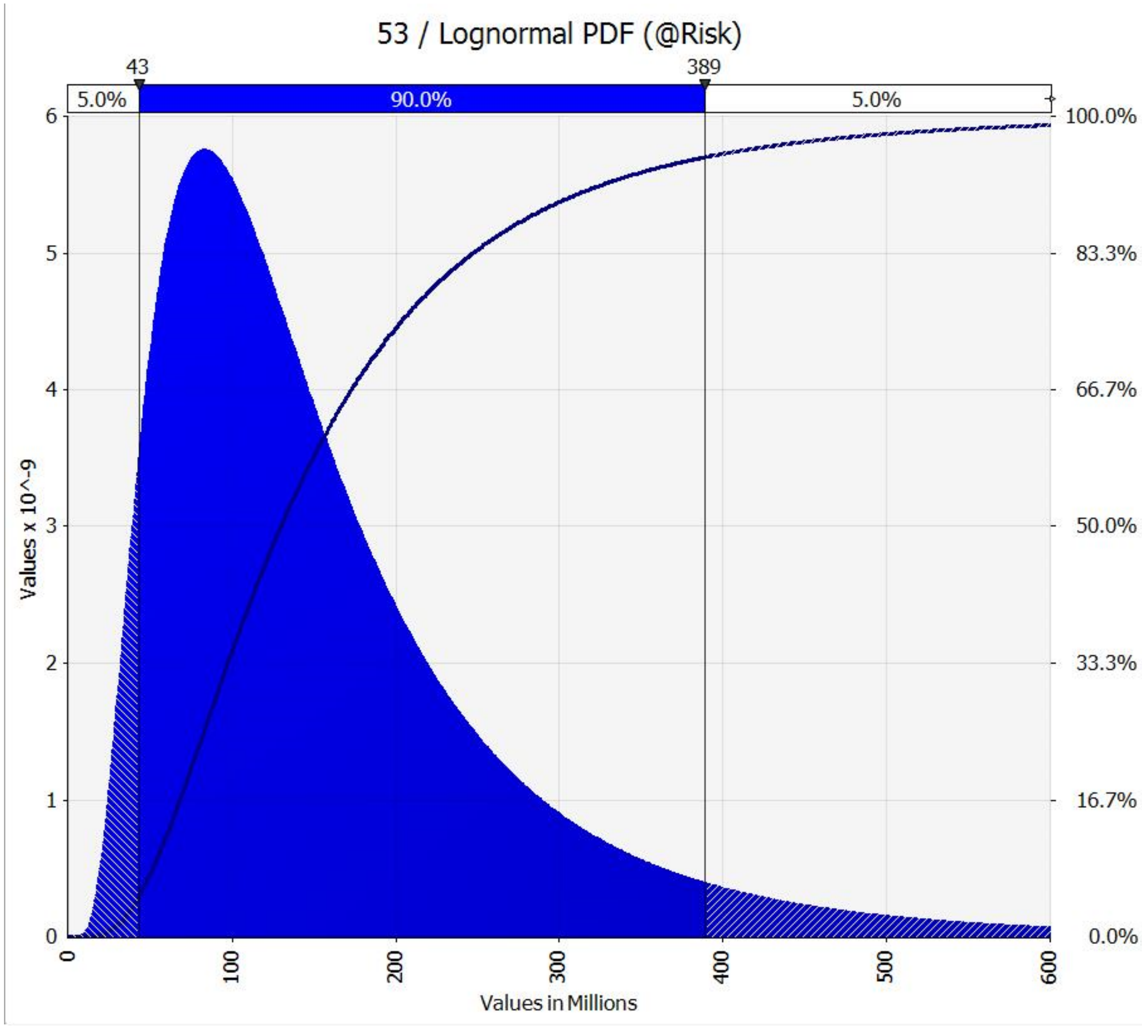


	1	2	3	4	5
TRL Impr 1	720,181	2,880,722	6,481,625	11,522,890	18,004,515
TRL Impr 2	2,880,722	11,522,890	25,926,501	46,091,558	72,018,060
TRL Impr 3	6,481,625	25,926,501	58,334,628	103,706,006	162,040,634
TRL Impr 4	11,522,890	46,091,558	103,706,006	184,366,233	288,072,238
TRL Impr 5	18,004,515	72,018,060	162,040,634	288,072,238	450,112,873

System Hierarchy Level

Multivariate Regression Cost Model Output - Sample Cost Model Uncertainty (Model No. 7)

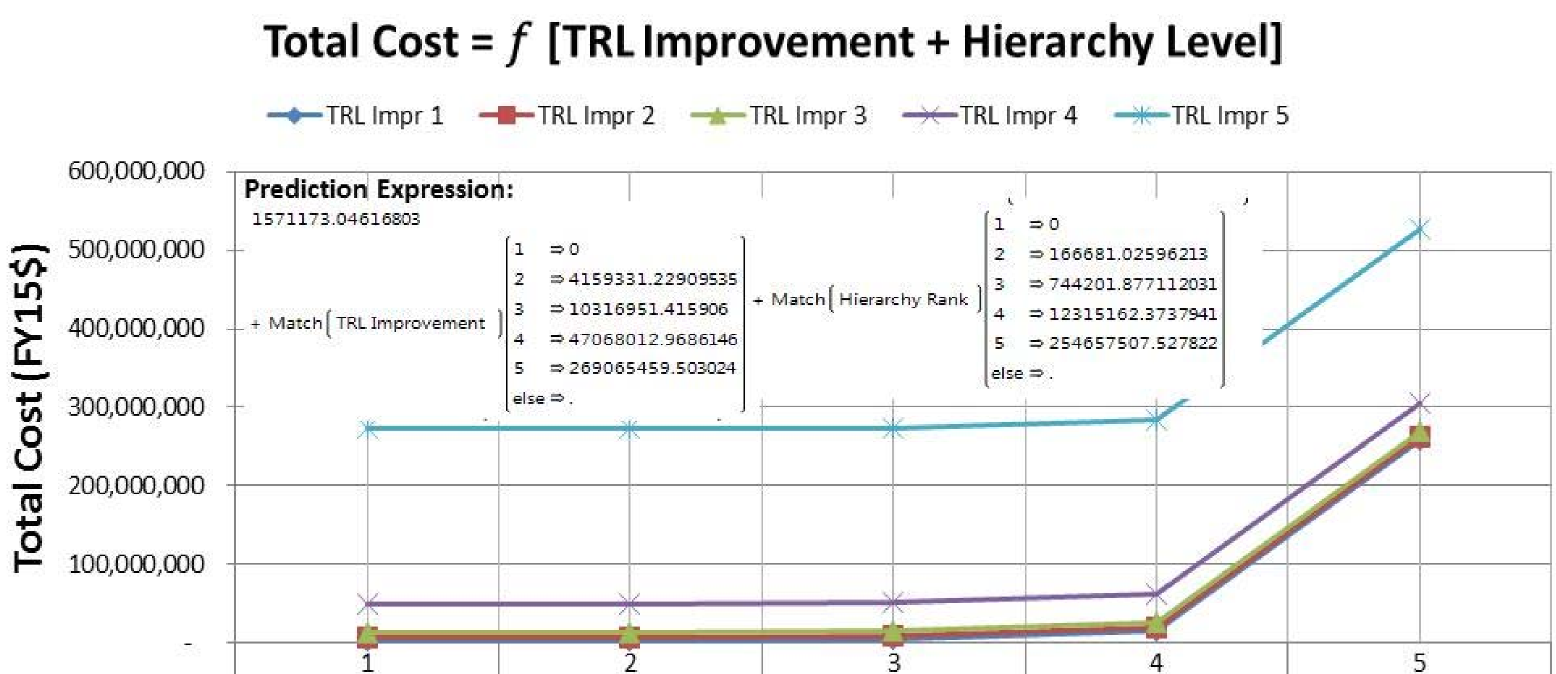
- **Composite Linear Regression (Model No. 7) – Sample PDF Uncertainty for TI Level 5 and SH Level 3 (24 other TI/SH Level PDFs also developed)**



Statistics	
	Lognorm(1620..
Cell	TRLxHier Sqrd7..
Minimum	0.00
Maximum	+∞
Mean	162,040,634.10
Mode	83,016,428.78
Median	129,659,389.27
Std Dev	121,460,472.00
Skewness	2.6699
Kurtosis	17.8886
Left X	43,232,030.58
Left P	5.0%
Right X	388,868,091.59
Right P	95.0%
Dif. X	345,636,061.01
Dif. P	90.0%
1%	27,426,861.96
5%	43,232,030.58
10%	55,101,217.76
15%	64,899,910.06
20%	73,916,012.25
25%	82,642,933.14
30%	91,354,671.59
35%	100,245,109.60
40%	109,479,985.45
45%	119,223,779.44
50%	129,659,389.27
55%	141,008,423.86
60%	153,558,270.56
65%	167,704,512.40
70%	184,025,150.91
75%	203,424,014.45
80%	227,441,344.77
85%	259,038,220.72
90%	305,103,188.45
95%	388,868,091.59
99%	612,959,559.57

Multivariate Models – Multiple Regression Model no. 8

- **Linear first order TI + SH predictor variable expression below**
  - **Constant intercept plus graduated TI and SH coefficients by level**



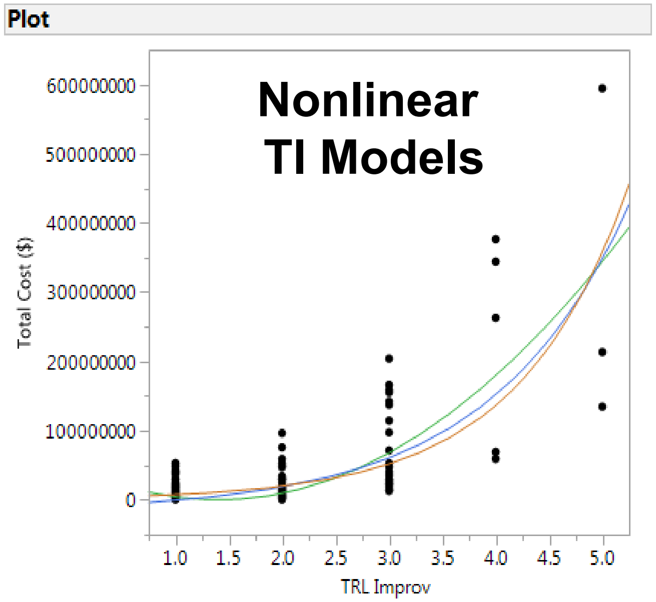
	1	2	3	4	5
TRL Impr 1	3,142,346	3,309,027	3,886,548	15,457,508	257,799,854
TRL Impr 2	7,301,677	7,468,358	8,045,879	19,616,840	261,959,185
TRL Impr 3	13,459,298	13,625,979	14,203,499	25,774,460	268,116,805
TRL Impr 4	50,210,359	50,377,040	50,954,561	62,525,521	304,867,867
TRL Impr 5	272,207,806	272,374,487	272,952,007	284,522,968	526,865,313

System Hierarchy Level

Cost Model Output – Nonlinear TI Cost Models

Nonlinear TI Level Cost Model

- **Nonlinear TI Models:**
  - **Cost Model no. 10 (TI NL-Quadratic)**
    - **Cost = a + b x TI level + c x TI Level^2**
  - **Cost Model no. 11 (TI NL-Exponential 2P)**
    - **Cost = a x (1 - b x exp(-c x TI level))**



Nonlinear TI Models

Model Comparison

Model		AICc	AICc Weight		BIC	SSE	MSE	RMSE	R-Square
Quadratic	green	12885.399	0.9999998		12900.644	3.643e+17	1.068e+15	32684768	0.6097699
Mechanistic Growth	blue	12916.095	2.16e-7		12931.34	3.983e+17	1.168e+15	34176060	0.5733479
Exponential 2P	orange	12929.287	2.95e-10		12940.738	4.163e+17	1.217e+15	34890388	0.5540222

Model Output – Nonlinear Cost Models

Model No.	Fit Model Type	Predictor Type	Single / Multiple Predictor Variable(s)	Predictor Variable / Parameter	Predictor Level / Parameter Name	Predictor Level / Parameter	Parameter Est. / No. Obs.	Prediction Estimate	Median	60th %ile	80st %ile	Mode
10	**Nonlinear - Quadratic**	**TRL Improvement Level**	**Single**	**TRL Improvement**	TRL Improvement 1	1		4,331,991	1,006,162	1,551,226	4,238,667	54,279
	Summary of Fit		*Ref Model Name:*	*TRL NL - Quadratic*	TRL Improvement 2	2		9,551,810	2,743,392	4,093,405	10,366,742	226,304
	AICc	12,885			TRL Improvement 3	3		67,957,831	55,036,648	64,879,635	95,067,227	36,097,482
	BIC	12,901			TRL Improvement 4	4		179,550,054	167,208,302	183,984,663	229,720,738	145,011,495
	SSE	3.643E+17			TRL Improvement 5	5		344,328,479	331,945,687	355,504,827	416,861,098	308,500,024
	MSE	1.068E+15		**Function**	Form	**Quadratic**						
	RMSE	32,684,768			Equation	Cost = a + b x TRL Improv + c x TRL Improv2				Lower 95%	Upper 95%	
	R-Square	0.6097699		**Parameters**	Intercept	a	52,298,374			36,460,718	68,136,030	N/A
	Coef. of Variation (CV) =	1.606			Slope	b	(74,559,484)			(90,696,944)	(58,422,023)	N/A
					Quadratic	c	26,593,101			23,012,184	30,174,017	N/A
						Total Applied	**343**	**15.3%**	**Data Reduction**			
11	**Nonlinear - Exponential 2P**	**TRL Improvement Level**	**Single**	**TRL Improvement**	TRL Improvement 1	1		7,552,009	7,358,007	7,795,826	8,915,566	6,984,827
	Summary of Fit		*Ref Model Name:*	*TRL NL - Exponential 2P*	TRL Improvement 2	2		19,819,678	19,071,071	20,460,087	24,088,178	17,657,615
	AICc	12,885			TRL Improvement 3	3		52,015,246	49,293,157	53,562,645	64,957,507	44,268,887
	BIC	12,901			TRL Improvement 4	4		136,510,080	127,045,553	139,852,153	174,791,139	110,039,589
	SSE	3.643E+17			TRL Improvement 5	5		358,260,387	326,495,146	364,157,622	469,225,970	271,164,357
	MSE	1.068E+15		**Function**	Form	**Exponential 2P**						
	RMSE	32,684,768			Equation	Cost = a x EXP(b x TRL Improv)				Lower 95%	Upper 95%	
	R-Square	0.6097699		**Parameters**	Scale	a	2877586.9			1,854,215	3,900,959	N/A
	Coef. of Variation (CV) =	1.714			Growth Rate	b	0.9648616			1	1	N/A
						Total Applied	**343**	**15.3%**	**Data Reduction**			
Model No.	**Fit Model Type**	**Predictor Type**	**Single / Multiple Predictor Variable(s)**	**Predictor Variable / Parameter**	**Predictor Level / Parameter Name**	**Predictor Level / Parameter**	**Parameter Est. / No. Obs.**	**Prediction Estimate**	**Median**	**60th %ile**	**80st %ile**	**Mode**
12	**Nonlinear - Exponential 3P**	**Hierarchy Level**	**Single**	**Hierarchy Rank**	Hardware / Software / Mat'l.	1		2,372,496	2,103,266	2,381,757	3,179,024	1,652,995
	Summary of Fit		*Ref Model Name:*	*Hier NL- Exponential 3P*	Component / Part	2		2,434,288	2,121,663	2,423,068	3,298,529	1,611,703
	AICc	19,749			Assembly	3		3,443,119	2,172,173	2,770,150	4,872,234	864,528
	BIC	19,766			Subsystem	4		19,913,385	5,784,901	8,615,831	21,727,389	488,200
	SSE	3.316E+17			System	5		288,808,534	59,985,261	93,997,585	266,732,460	2,587,698
	MSE	6.233E+14		**Function**	Form	**Exponential 3P**						
	RMSE	24,966,011			Equation	Cost = a+b x EXP(c x Hierarchy Rank)				Lower 95%	Upper 95%	
	R-Square	0.7436299		**Parameters**	Asymptote	a	2368463.9			-46553.5	4783481.2	N/A
	Coef. of Variation (CV) =	2.070			Scale	b	246.95741			-254.6162	748.53104	N/A
					Growth Rate	c	2.7927648			2.3862295	3.1993001	N/A
						Total Applied	**535**	**11.3%**	**Data Reduction**			
13	**Nonlinear - Gompertz 4P**	**Hierarchy Level**	**Single**	**Hierarchy Rank**	Hardware / Software / Mat'l.	1		1,354,510	945,930	1,172,449	1,930,128	461,332
	Summary of Fit		*Ref Model Name:*	*Hier NL-Gompertz 4P*	Component / Part	2		1,508,967	800,569	1,064,833	2,065,089	225,339
	AICc	19,754			Assembly	3		3,349,633	861,100	1,307,399	3,447,583	56,907
	BIC	19,775			Subsystem	4		25,284,800	5,501,148	8,562,946	23,924,522	260,400
	SSE	3.335E+17			System	5		286,685,634	71,940,045	109,630,329	291,585,161	4,530,021
	MSE	6.281E+14		**Function**	Form	**Gompertz 4P**						
	RMSE	25,060,990			Equation	Cost = a + (b -a) x Exp(-Exp(-c x(Hierarchy Rank - d)))				Lower 95%	Upper 95%	
	R-Square	0.7421611		**Parameters**	Lower Asymptote	a	8.24E+14			-6.76E+14	2.32E+15	N/A
	Coef. of Variation (CV) =	2.078			Upper Asymptote	b	1340361.5			-1157856	3838579	N/A
					Growth Rate	c	-2.477964			-2.780436	-2.175492	N/A
					Inflection Point	d	11.003063			11.003063	11.003063	N/A
						Total Applied	**535**	**11.3%**	**Data Reduction**			

APL

JOHNS HOPKINS

APPLIED PHYSICS LABORATORY