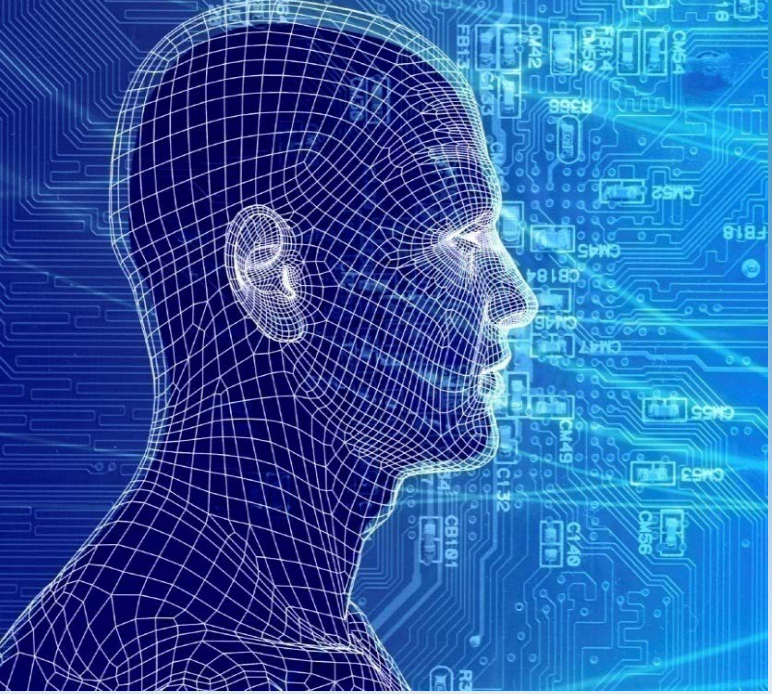


Joining Effort and Duration in a Probabilistic Method for Predicting Software Cost and Schedule

Mike Ross
President & CEO
r2ESTIMATING, LLC
325 Woodyard Road
Lakeside, Montana 59922
(o) 480.710.4678 (f) 480.488.8420
mike.ross@r2estimating.com
http://www.r2estimating.com

Outline

- Problem and Background
- Model Summary
- Regressing CDERs
- Probability in CDER Mathematics
- Joint and Conditional Probability Defined
- Examples
- References

Challenge

- **When cost uncertainty analyses are presented to decision-makers, questions often asked are**
  - ***"What is the chance the system can be delivered within cost and schedule?"***
  - ***"How likely might the point estimate cost be exceeded for a given schedule?"***
  - ***"How are cost reserve recommendations affected by schedule risk?"***
- **During the past thirty years, techniques from univariate probability theory have been widely applied to provide insight into P(Cost ≤ x1) and P(Schedule ≤ x2).**
- **Although it has long been recognized that a system's cost and schedule are correlated, little has been applied from multivariate probability theory to study joint cost-schedule distributions.**

(Garvey, 2000)

r2™

Background

- **USAF (Marvin Sambur / Peter Teets memo of 2004)**
- **NASA Cost Estimating Handbook**
- **Past ISPA/SCEA Conferences**
  - Panel of Experts Discussion
  - Presentations
- **SSCAG Risk Working Group (Tim Anderson, Eric Druker, et. al.)**
  - ***Parametric***
    - ***"Disjoint Cost and Schedule distributions are conflated into a bivariate distribution through the injection of correlation between the two"***
  - Buildup
  - Estimate to Complete Projection
- **Joint Confidence Level (JCL)**
  - NASA
  - USAF AFCAA

r2 TM

Notation Convention

Use of Fonts

Numeric Constant -- Times New Roman : 54.0905

Simple Variable -- Times New Roman Italic : x

Function -- Times New Roman Bold Italic : $\boldsymbol{f}\left(x\right)$

Vector, Matrix, List, Array -- Times New Roman Bold : $\mathbf{X}$

Random Variable -- Arial Bold Italic : $\boldsymbol{\mathsf{X}}$

Use of Symbols

$\equiv$	The left operand is defined as (assumed equivalent to) the right operand
$\cong$	The left operand is approximated by the right operand
□	The left operand estimates the right operand
$\in$	The left operand is an element (member) of the right operand
$\propto$	The left operand is proportional to the right operand
$\wedge$	Logical AND operator
$\neg$	Logical NOT operator
†	Indicates specific calibration to Aerospace 2004 : Military Ground C / C++ data set

Variable Dictionary

$$\text{Letting X} \in \left\{ \begin{array}{l} \text{E (effort), T (duration), S (software size), D (difficulty), I (intensity),} \\ \omega \text{ (inefficiency), } \Psi \text{ (work), } \Omega \text{ (staffing), P (productivity)} \end{array} \right\}$$

$\boldsymbol{X} \equiv$ Random variable of X; range and distribution of possible outcomes

$\mathbf{X} \equiv$ List of outcomes of X; $\boldsymbol{X} \cong \mathbf{X}$

$\mathbf{X}_{[i]} \equiv$ The list $\mathbf{X}$ indexed by i

$X_i \equiv$ The i^{th} element (member) of list $\mathbf{X}$; $X_i \in \mathbf{X}$

$\bar{X} \equiv$ The expected (mean) value of $\boldsymbol{X}$ or $\mathbf{X}$

$X \equiv$ Some specific value of X

$\alpha_E, \alpha_T, \alpha_S, \alpha_{ET} \equiv$ Effort, duration, size, and effort-duration tradeoff nonlinearities (exponents)

$a_1, a_2, a_3 \equiv$ First, second, and third regression exponents

$\mathcal{E}af \equiv$ Effort adjustment factor; data set life cycle to all-all life cycle

$\mathcal{T}af \equiv$ Duration adjustment factor; data set life cycle to all-all life cycle

r2™

r2 Software Estimating Framework (r2SEF) CDERs – General System of Equations

Work Relation

Resources Applied ↔ Work Performed

$$\textbf{Effort}^{EffortExponent} \times \textbf{Time}^{TimeExponent} = \textbf{Difficulty} \times \textbf{Size}^{Size\ Exponent}$$

$$E^{\alpha_E} T^{\alpha_T} = D S^{\alpha_S}$$

Intensity Relation

Cost ↔ Schedule

$$\textbf{Effort} = \textbf{Intensity} \times \textbf{Time}^{Intensity\ Exponent}$$

$$E = I T^{\alpha_{ET}}$$

(Ross, 2008a), (Ross, 2008b)

r2SEF CDERs – Calibration Process (3 regressions)

Inefficiency Regression $(y = bx^a)$ **Yields a_1** **Yields ω based on a_1**

$$\left[\boldsymbol{E} \propto f(\boldsymbol{S})\right]_{\langle \text{dataset name} \rangle} \rightarrow \left[\mathbf{E} = \omega\bar{\omega}\mathbf{S}^{a_1}\right]_{\langle \text{dataset name} \rangle} \rightarrow \left[\;= \frac{\mathbf{E}}{\mathbf{S}^{a_1}}\right]_{\langle \text{dataset name} \rangle}$$

Intensity Regression $(y = bx^a)$ **Yields a_2** **Yields I based on a_2**

$$\left[\boldsymbol{E} \propto f(\boldsymbol{T})\right]_{\langle \text{dataset name} \rangle} \rightarrow \left[\mathbf{E} = \bar{I}\mathbf{T}^{a_2}\right]_{\langle \text{dataset name} \rangle} \rightarrow \left[\mathbf{I} = \frac{\mathbf{E}}{\mathbf{T}^{a_2}}\right]_{\langle \text{dataset name} \rangle}$$

Difficulty Regression $(y = bx^a)$ **Yields a_3** **Yields D based on a_3**

$$\left[\boldsymbol{\omega} \propto f(\boldsymbol{I})\right]_{\langle \text{dataset name} \rangle} \rightarrow \left[\boldsymbol{\omega} = \bar{D}\mathbf{I}^{a_3}\right]_{\langle \text{dataset name} \rangle} \rightarrow \left[\mathbf{D} = \frac{\boldsymbol{\omega}}{\mathbf{I}^{a_3}}\right]_{\langle \text{dataset name} \rangle}$$

By Substitution

$$\frac{\mathbf{E}}{\mathbf{S}^{a_1}} = \mathbf{D}\left(\frac{\mathbf{E}}{\mathbf{T}^{a_2}}\right)^{a_3} \rightarrow \mathbf{E}^{1-a_3}\mathbf{T}^{a_2 a_3} = \mathbf{DS}^{a_1}$$

Letting

$$\alpha_E \equiv 1 - a_3, \quad \alpha_T \equiv a_2 a_3, \quad \alpha_S \equiv a_1, \quad \alpha_{ET} \equiv a_2 \quad \rightarrow$$

$$\left[\mathbf{E}^{\alpha_E}\mathbf{T}^{\alpha_T} = \mathbf{DS}^{\alpha_S}\right]_{\langle \text{dataset name} \rangle}$$

Work Equation

$$\left[\mathbf{E} = \mathbf{IT}^{\alpha_{ET}}\right]_{\langle \text{dataset name} \rangle}$$

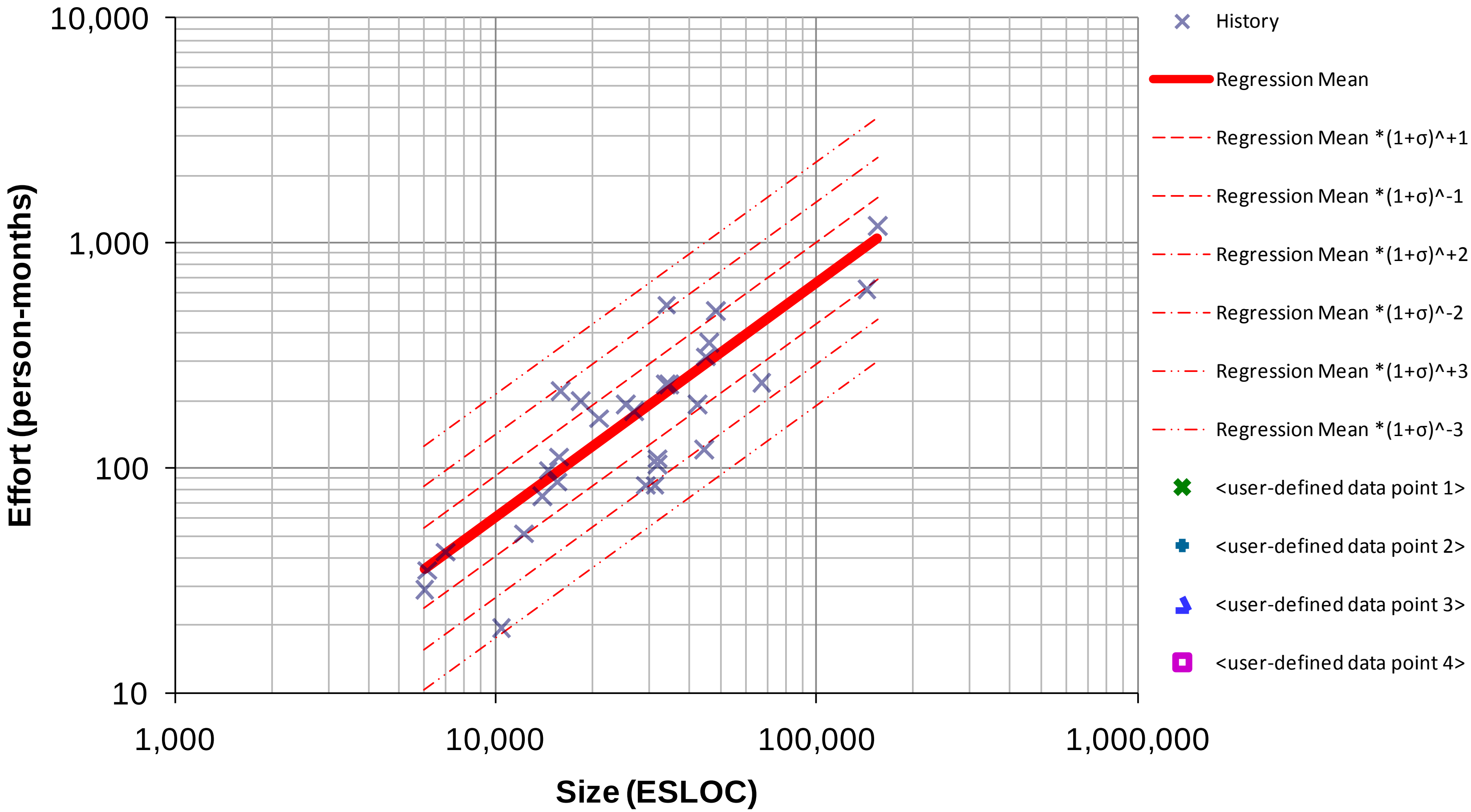
Intensity Equation

Inefficiency Regression

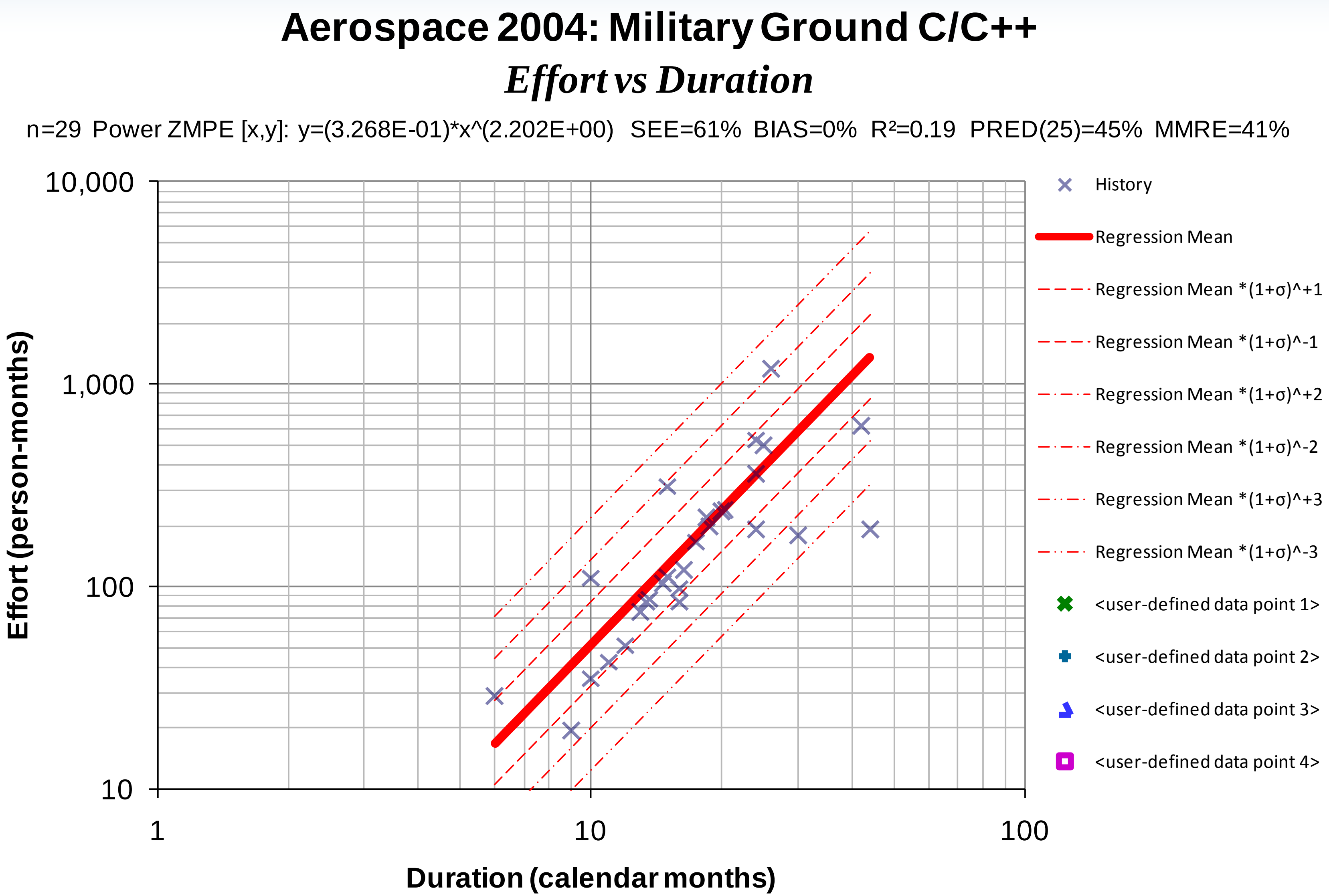
Aerospace 2004: Military Ground C/C++

Effort vs Size

n=29 Power ZMPE [x,y]: y=(4.333E-03)*x^(1.037E+00) SEE=51% BIAS=0% R^2=0.75 PRED(25)=48% MMRE=36%



Intensity Regression

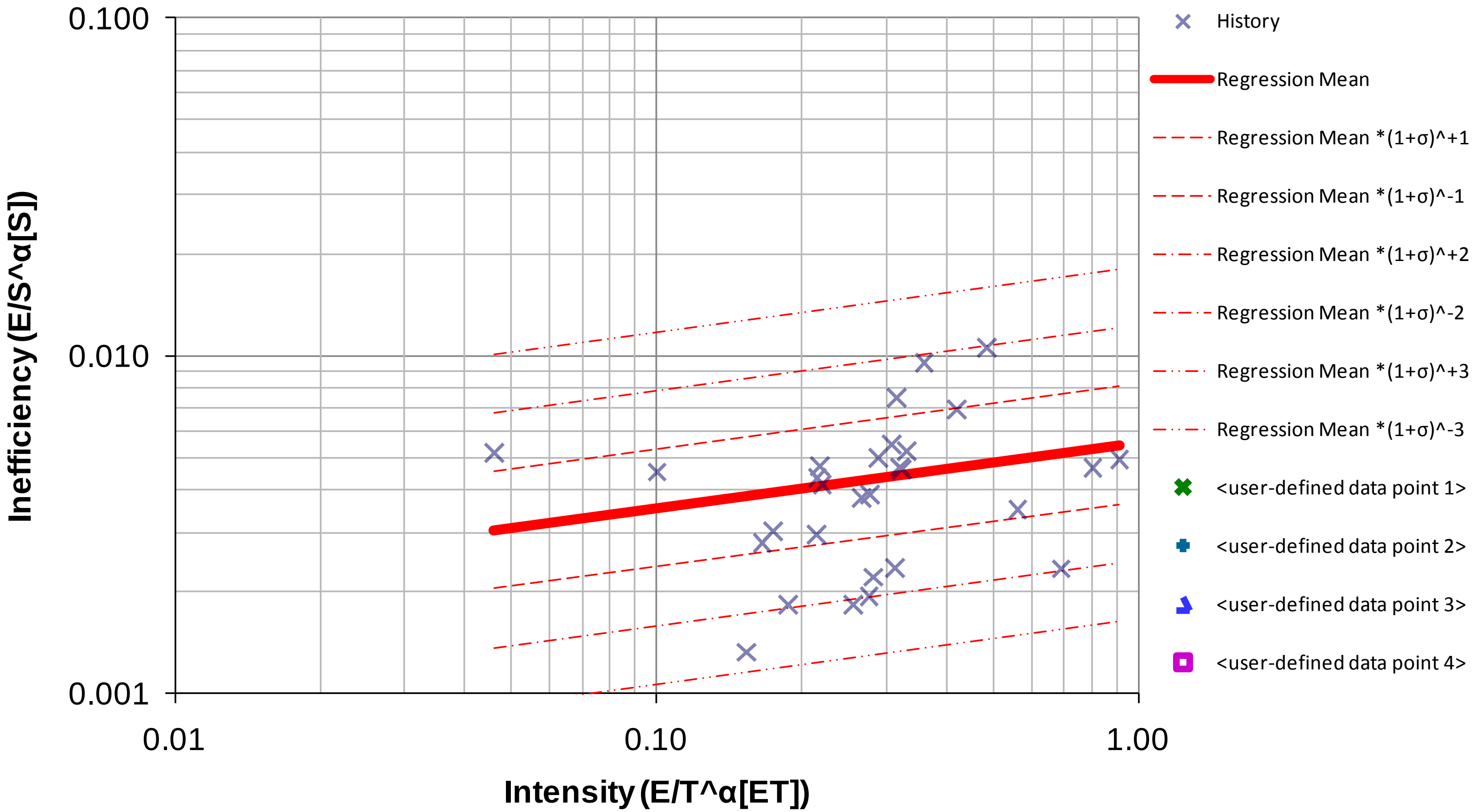


Difficulty Regression

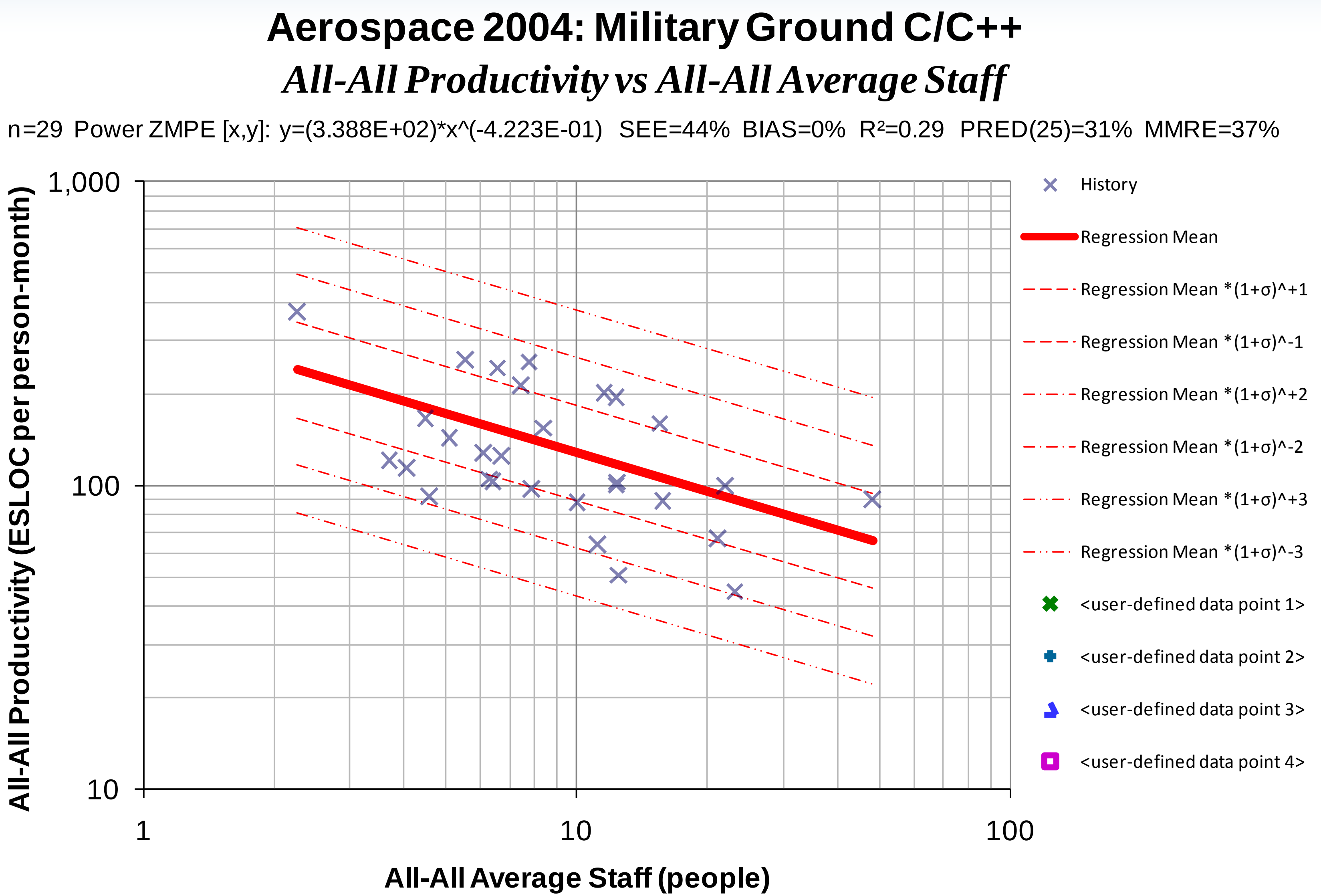
Aerospace 2004: Military Ground C/C++

Inefficiency vs Intensity

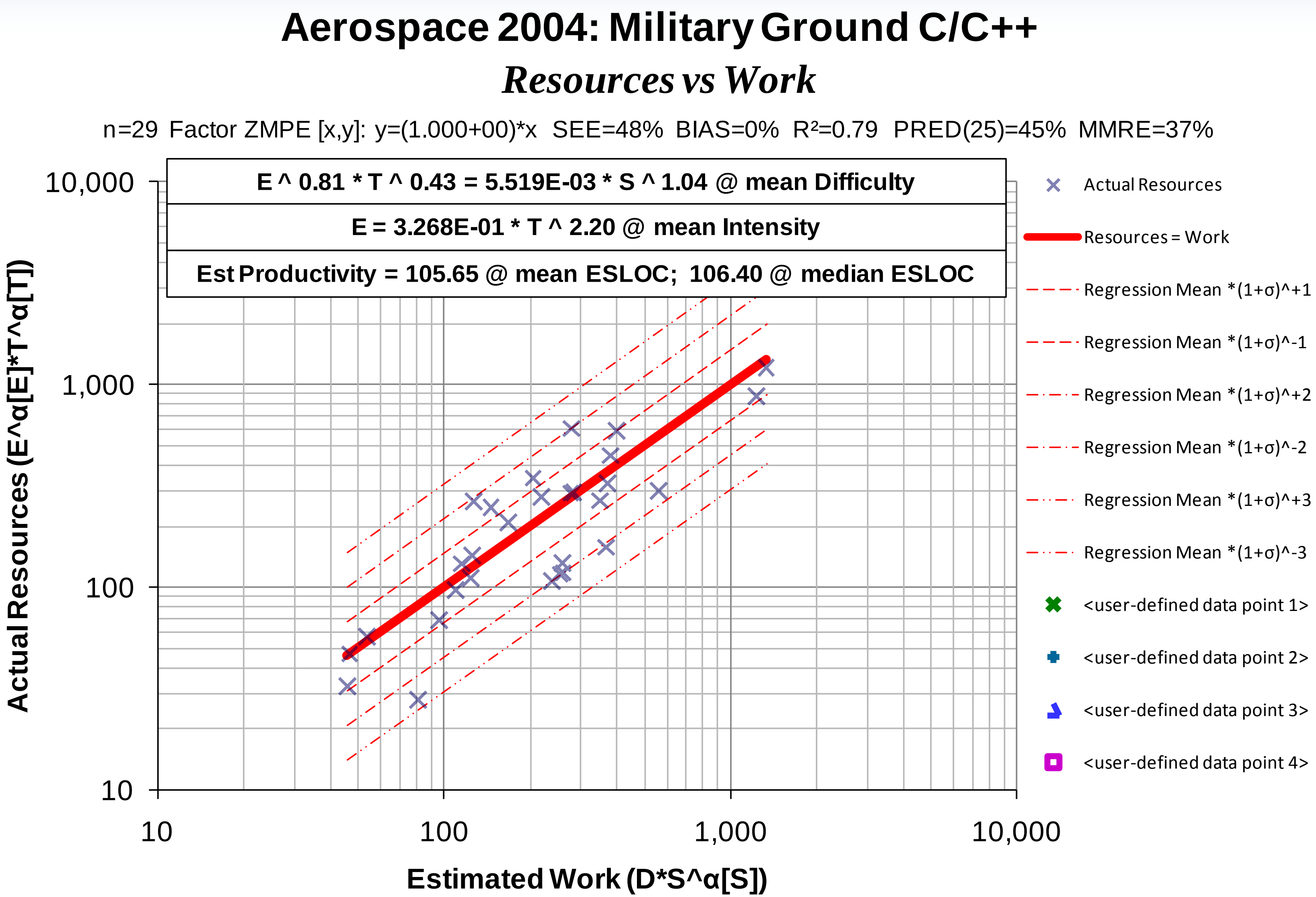
n=29 Power ZMPE [x,y]: y=(5.519E-03)*x^(1.932E-01) SEE=49% BIAS=0% R^2=0.05 PRED(25)=45% MMRE=37%



Productivity Trends Downward as Team Size Increases



Example CDER Regression Results



Single-point Solution CDER

Work Equation (solved for Effort)

$$E = \bar{\Psi}^{1/\alpha_E} \left(T/\mathcal{T}_{af}\right)^{(-1)\alpha_T/\alpha_E} \mathcal{E}_{af}$$

$$\left[E = \left((0.005519)(127,571)^{1.04} \right)^{1/0.81} \left(T/1.37\right)^{(-1)0.43/0.81} 1.44 \right]_{\dagger}$$

Intensity Equation (solved for Effort)

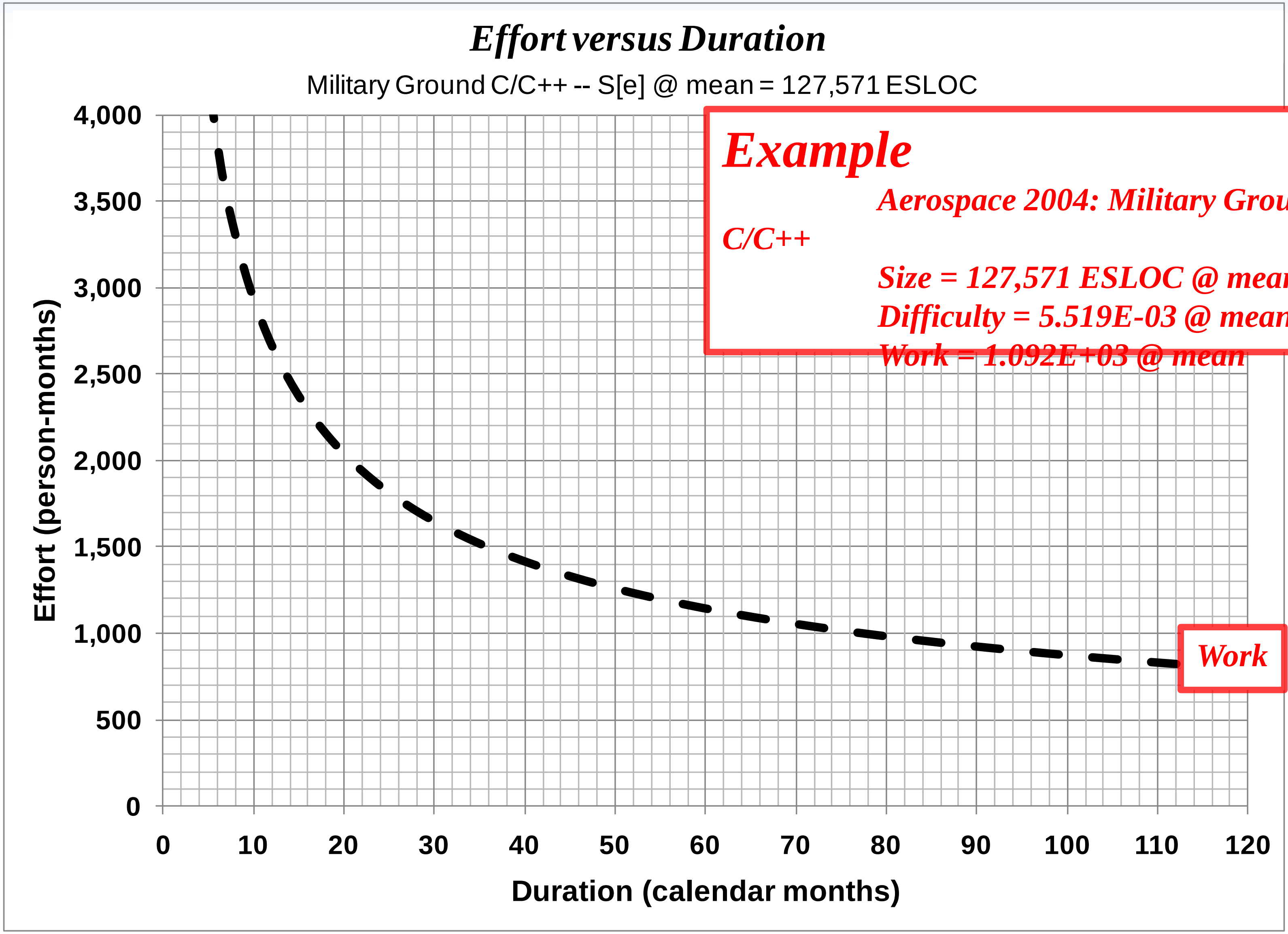
$$E = \bar{I} \left(T/\mathcal{T}_{af}\right)^{\alpha_{ET}} \mathcal{E}_{af}$$

$$\left[E = 0.3268 \left(T/1.37\right)^{2.20} 1.44 \right]_{\dagger}$$

† Aerospace 2004: Military Ground C/C++ data set

r2™

Work Function at a Particular Size and Difficulty



Example

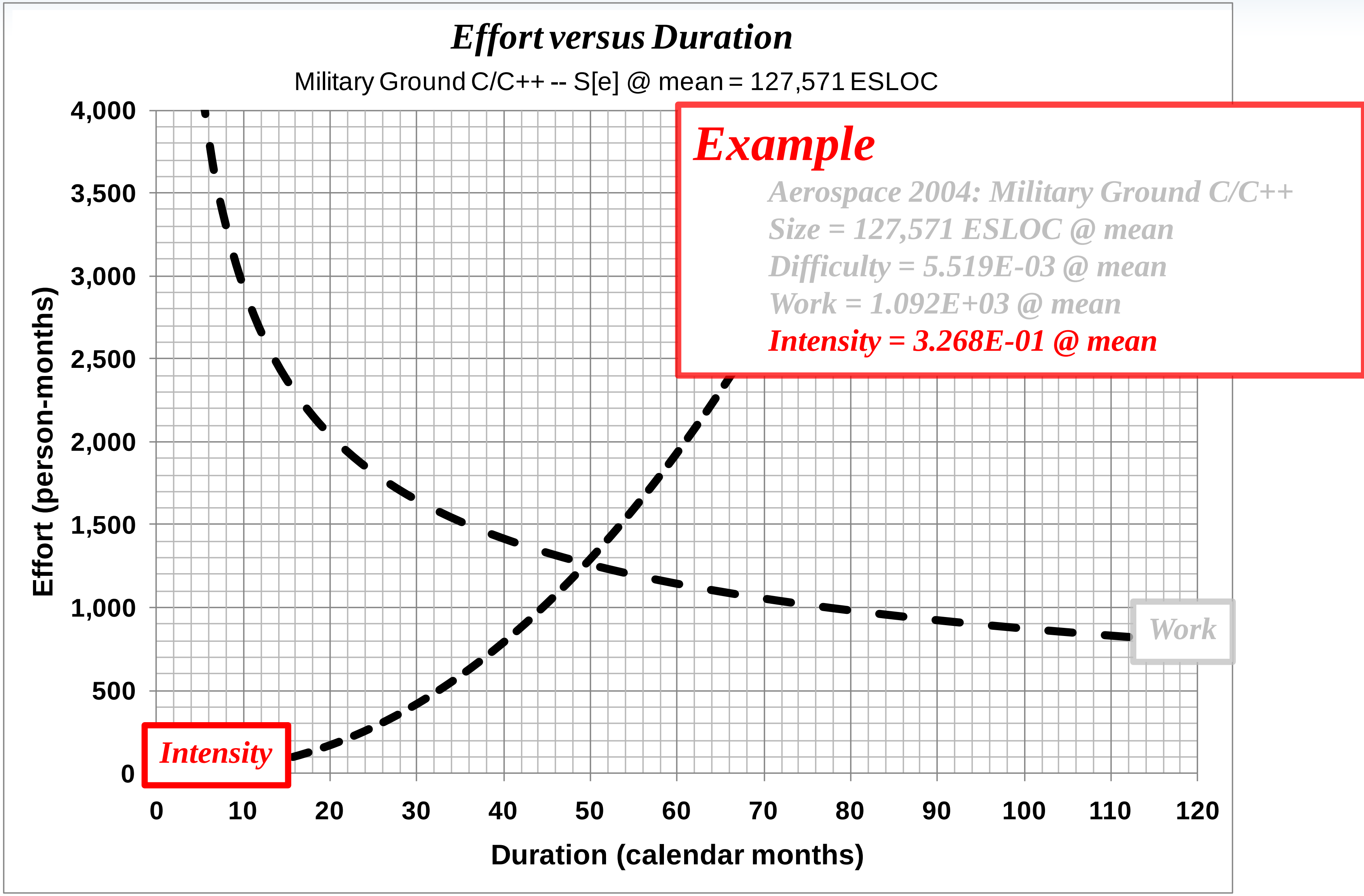
Aerospace 2004: Military Ground C/C++

Size = 127,571 ESLOC @ mean

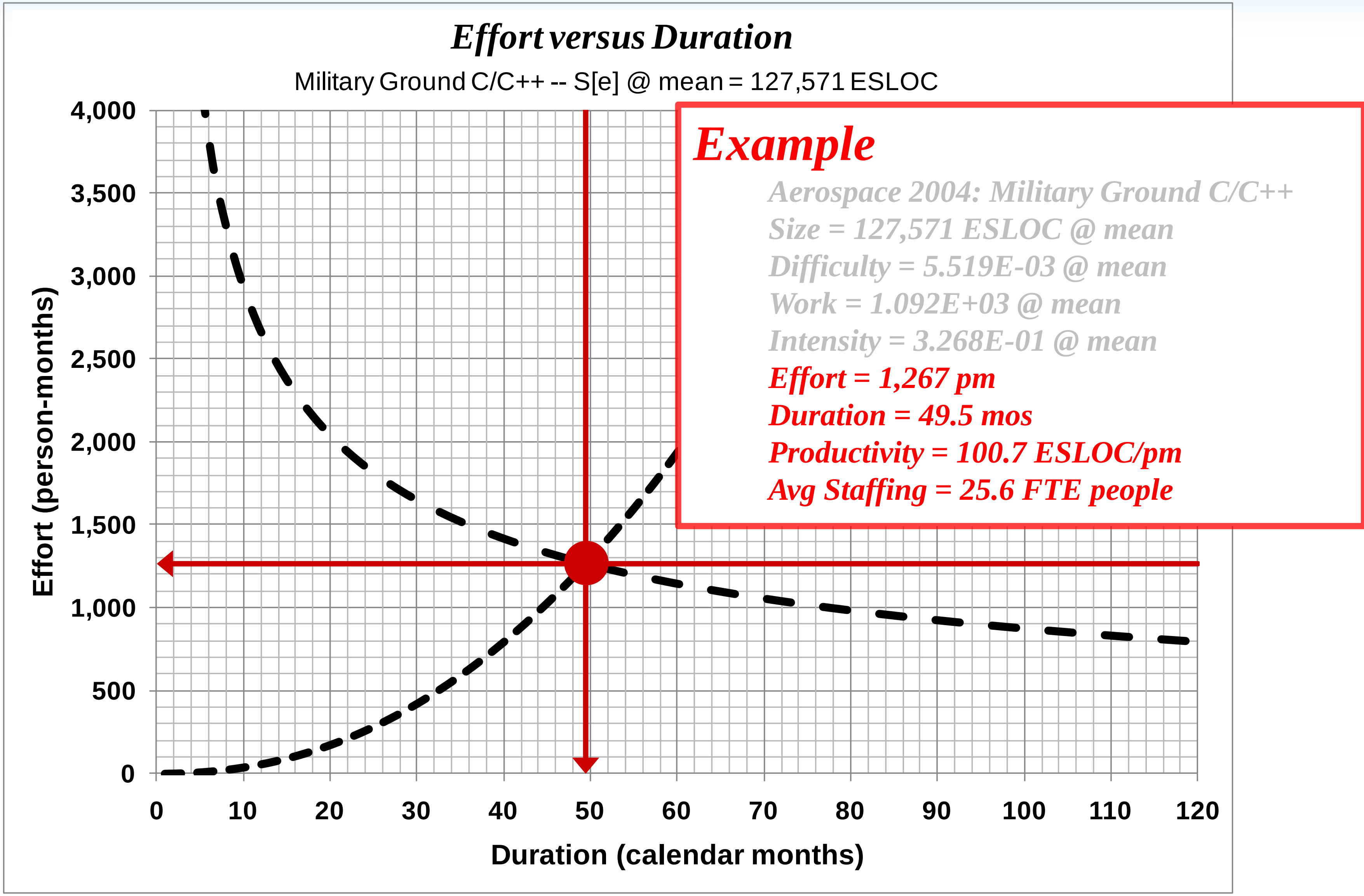
Difficulty = 5.519E-03 @ mean

Work = 1.092E+03 @ mean

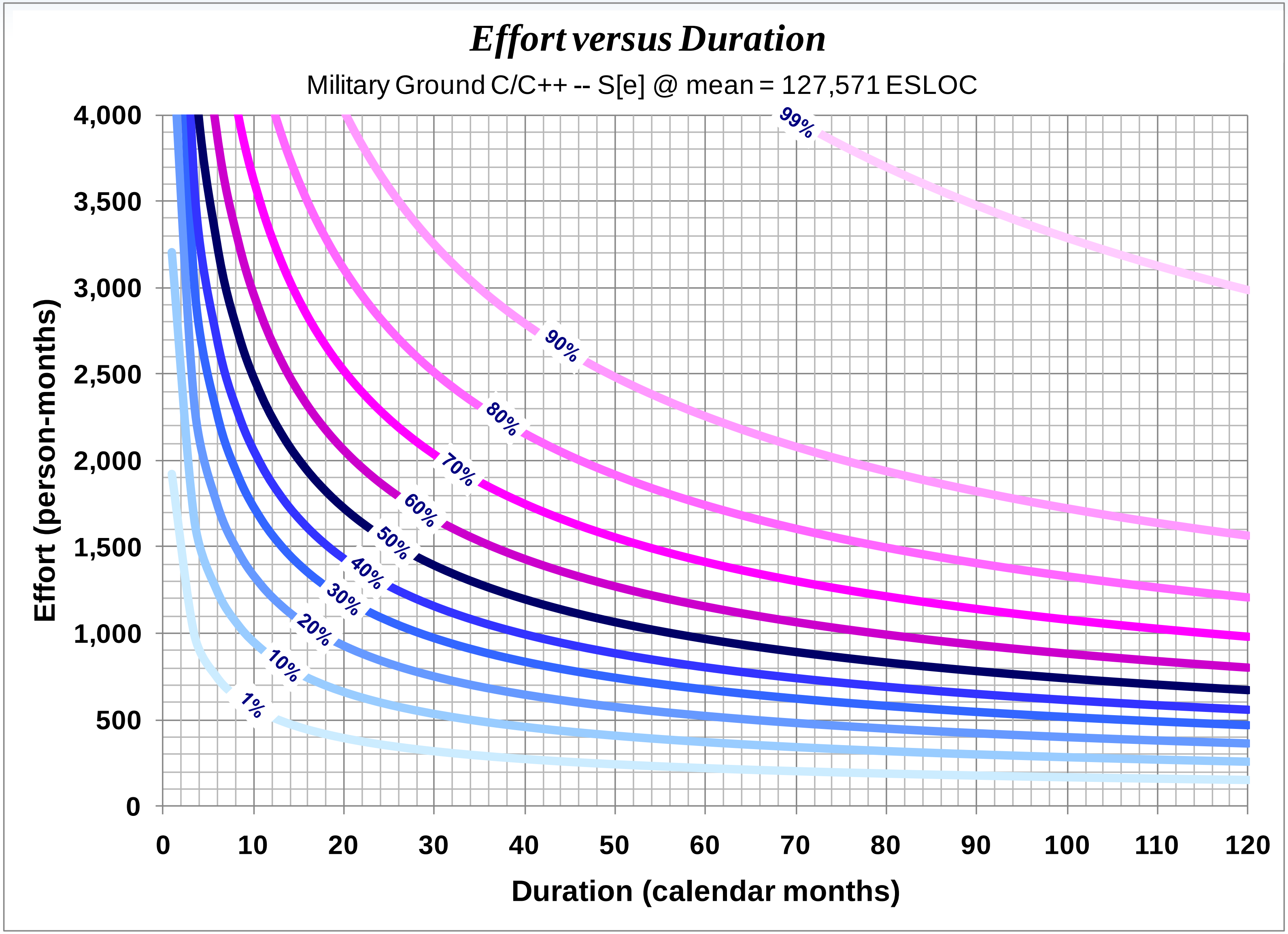
Intensity Function at a Particular Intensity



Work Function (CDER) at a Particular Intensity

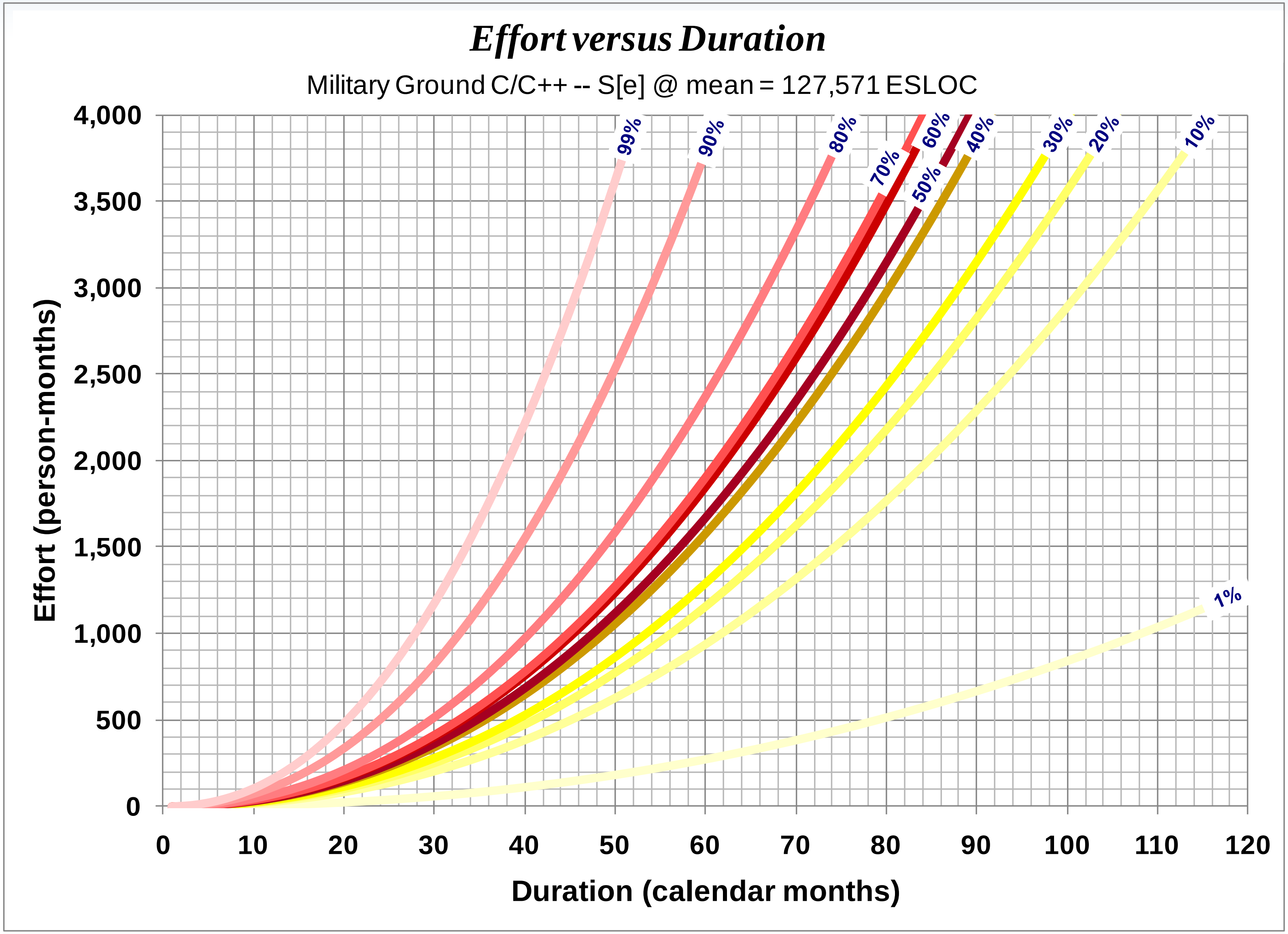


Work Function as a Probability Field



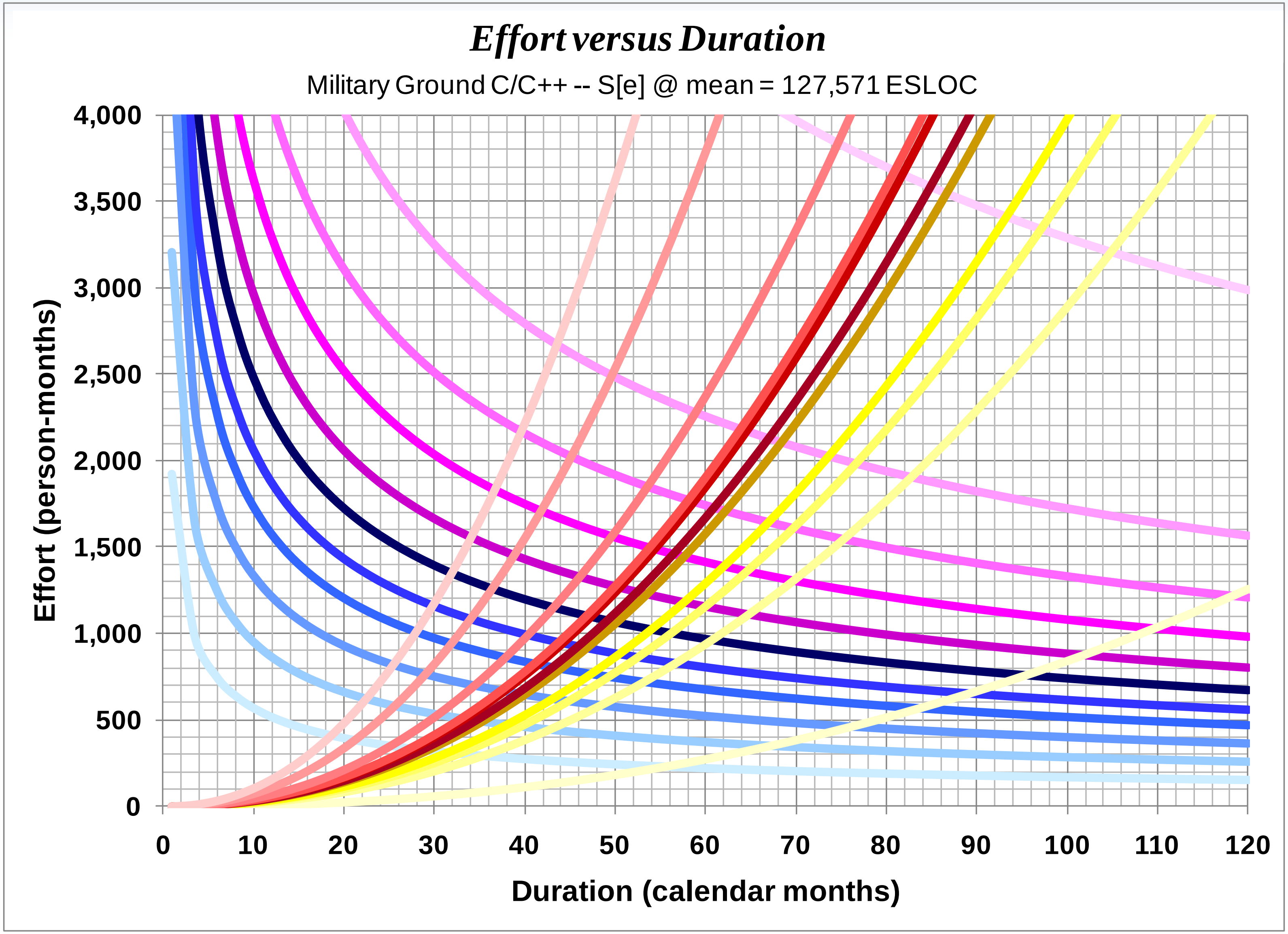
r2™

Intensity Function as a Probability Field



r2 ™

Work and Intensity Probability Fields



r2™

r2SEF CDERs – Solving for the Work

Cost (Effort) Estimating Relationship (CER)

$$Work = \Psi D S^{\alpha_S} E^{(\alpha_{ET}\alpha_E+\alpha_T)/\alpha_{ET}} I \left(\; -1\right)^{\alpha_T/\alpha_{ET}}$$

Schedule (Duration) Estimating Relationship (SER)

$$Work = \Psi D S^{\alpha_S} T^{\alpha_{ET}\alpha_E+\alpha_T} I \; \alpha_E$$

(Ross, 2008a), (Ross, 2008b)

r2SEF CDERs – Solving for Cost and Schedule

Cost (Effort) Estimating Relationship (CER)

$$E \Psi I^{\alpha_T/(\alpha_{ET}\alpha_E+\alpha_T)} \quad \alpha_{ET}/(\alpha_{ET}\alpha_E+\alpha_T)$$

Schedule (Duration) Estimating Relationship (SER)

$$T \Psi \left(I^{-1}\right)^{\alpha_E/(\alpha_{ET}\alpha_E+\alpha_T)} \quad 1/(\alpha_{ET}\alpha_E+\alpha_T)$$

(Ross, 2008a), (Ross, 2008b)

r2 TM

r2SEF CDERs – Solving for Intensity

Cost (Effort) Estimating Relationship (CER)

$$I\Psi \left(\quad -1 \right)^{\alpha_{ET}/\alpha_T} E^{\left(\alpha_{ET}\alpha_E + \alpha_T\right)/\alpha_T}$$

Schedule (Duration) Estimating Relationship (SER)

$$I\Psi \quad {}^{1/\alpha_E} T \left(\quad -1 \right)^{\left(\alpha_{ET}\alpha_E + \alpha_T\right)/\alpha_E}$$

(Ross, 2008a), (Ross, 2008b)

r2SEF CDERs – Effort vs. Duration Graphing Functions

Work Function (solved for Effort)

$$E = \Psi^{1/\alpha_E}\left(T^{-1}\right)^{\alpha_T/\alpha_E}$$

Intensity Function (solved for Effort)

$$E = IT^{\alpha_{ET}}$$

(Ross, 2008a), (Ross, 2008b)

Joint and Conditional Probability Defined

Joint

- The probability that effort will be less than or equal to some given (goal) value and duration will be less than or equal to some given (goal) value

$$\boldsymbol{P}\left(Effort \le x_1 \wedge Duration \le x_2\right)$$

Conditional

- The probability that effort will be less than or equal to some given (goal) value based on some point estimate of duration being equal to some given (goal) value or vice versa

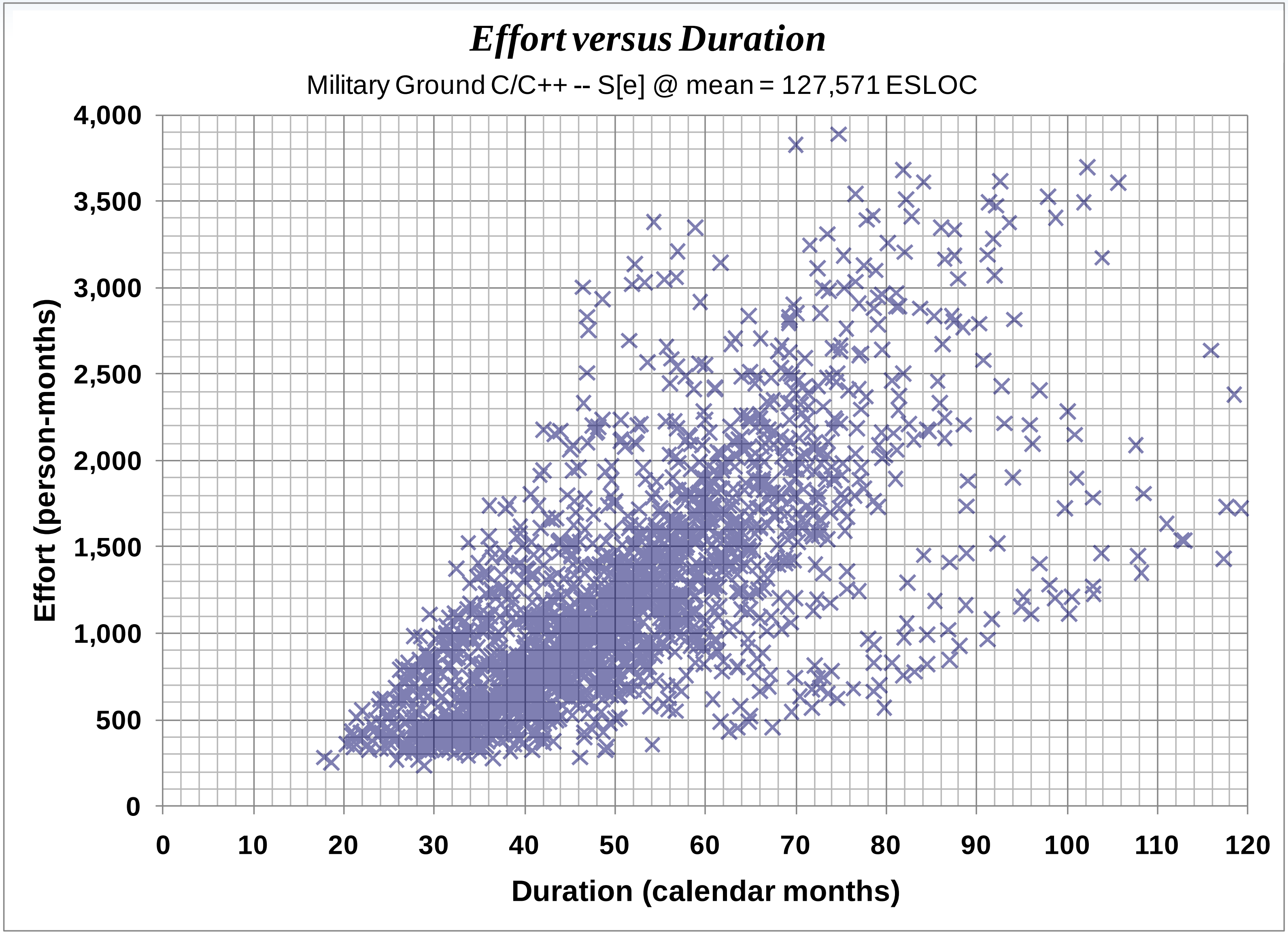
$$\boldsymbol{P}\left(Effort \le x_1 \mid Duration = x_2\right)$$

or

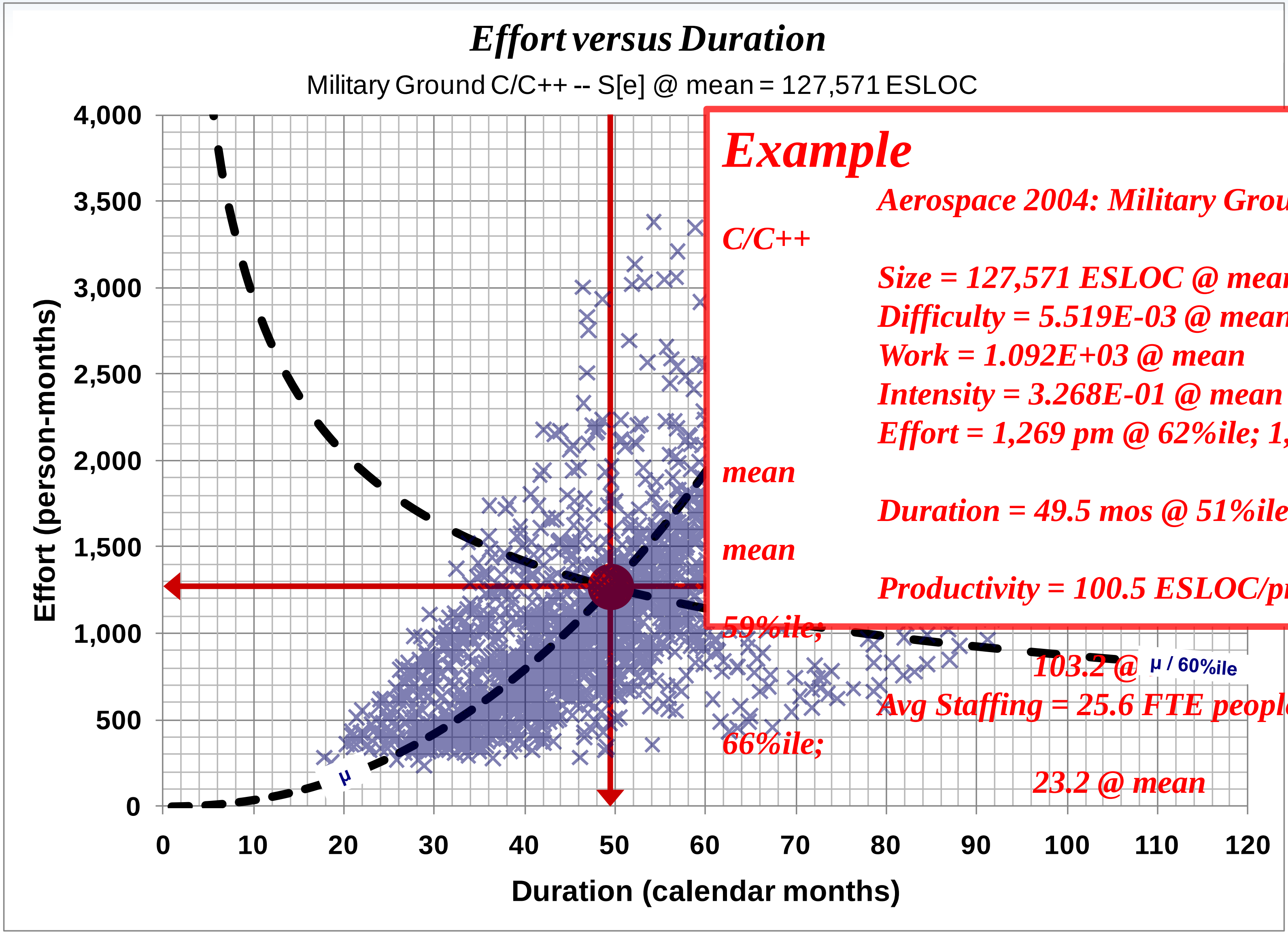
$$\boldsymbol{P}\left(Duration \le x_1 \mid Effort = x_2\right)$$

(Garvey, 2000)

Probabilistic (Monte Carlo) Solution – 2,001 Samples of the Example Problem



Probabilistic (Monte Carlo) Solution – Mean Size, Difficulty, and Intensity



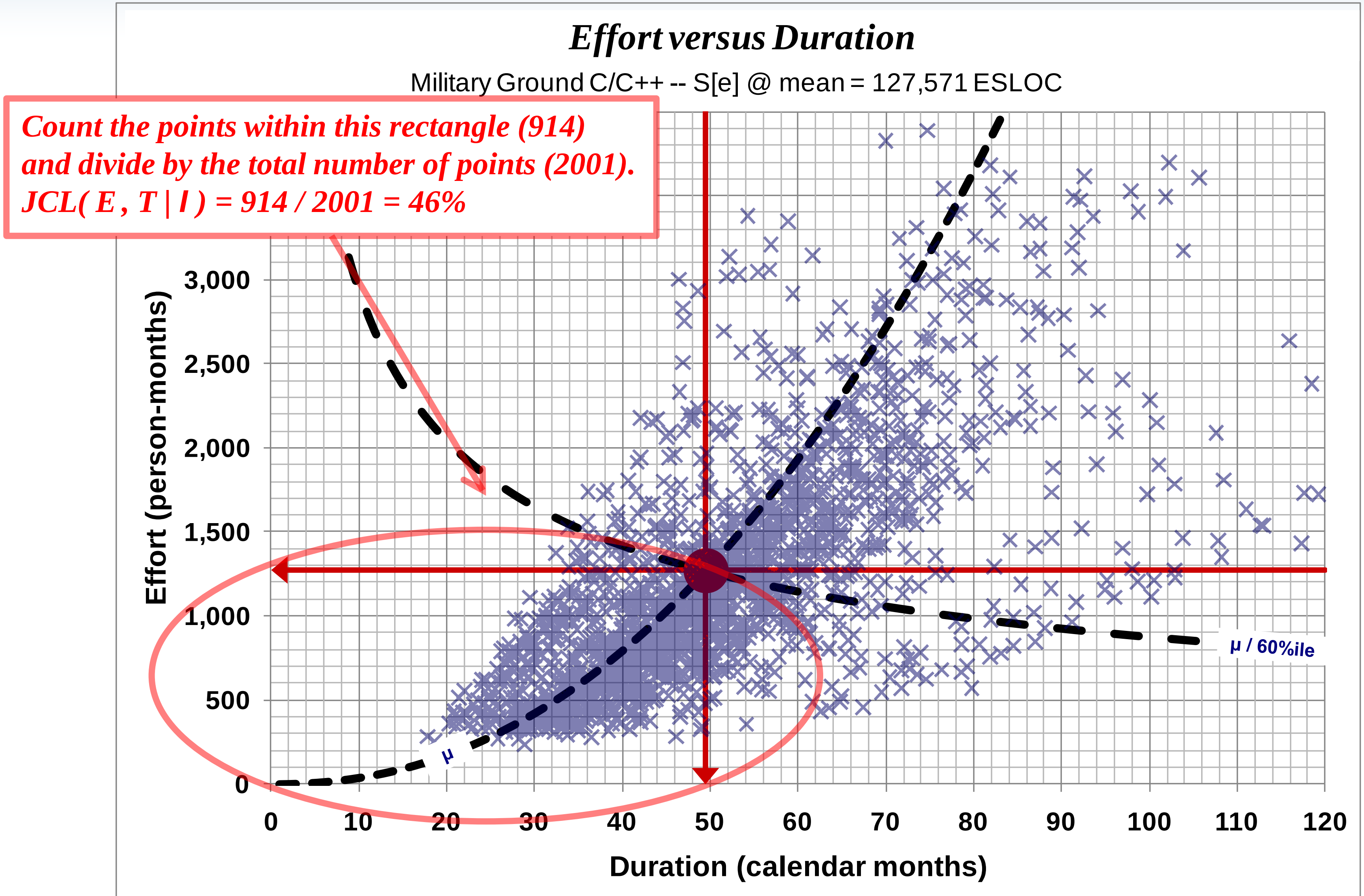
Example

Aerospace 2004: Military Ground C/C++

Size = 127,571 ESLOC @ mean
Difficulty = 5.519E-03 @ mean
Work = 1.092E+03 @ mean
Intensity = 3.268E-01 @ mean
Effort = 1,269 pm @ 62%ile; 1,240 @ mean
Duration = 49.5 mos @ 51%ile; 51.9 @ mean
Productivity = 100.5 ESLOC/pm @ 59%ile; 103.2 @ mean
Avg Staffing = 25.6 FTE people @ 66%ile; 23.2 @ mean

r2™

Joint Confidence Level (JCL) of Effort and Duration – Scatter Plot Approach



r2 ™

JCL of Effort and Duration – Boolean Random Variable Approach

$$\boldsymbol{J} \equiv \left(\boldsymbol{E} \le \hat{E}\right) \wedge \left(\boldsymbol{T} \le \hat{T}\right) \rightarrow \boldsymbol{J} \cong \mathbf{J}_i = \left(\mathbf{E}_i \le \hat{E}\right) \wedge \left(\mathbf{T}_i \le \hat{T}\right)\Big|_{i=1}^{N}$$

$$\text{(CDF of } \boldsymbol{J}\text{)} \qquad \boldsymbol{F}_{\boldsymbol{J}}\left(x\right) \cong \boldsymbol{F}_{\mathbf{J}}\left(x\right) = \begin{cases} 0 & x = 0 \\ \frac{1}{N}\sum_{i=1}^{N} \neg \mathbf{J}_i & 0 < x < 1 \\ \frac{1}{N}\sum_{i=1}^{N} \left(\neg \mathbf{J}_i + \mathbf{J}_i\right) & x = 1 \end{cases}$$

$$\text{JCL} \equiv \frac{1}{N}\sum_{i=1}^{N} \mathbf{J}_i \quad \text{(\% of the } \mathbf{J} \text{ elements that evaluate to TRUE)}$$

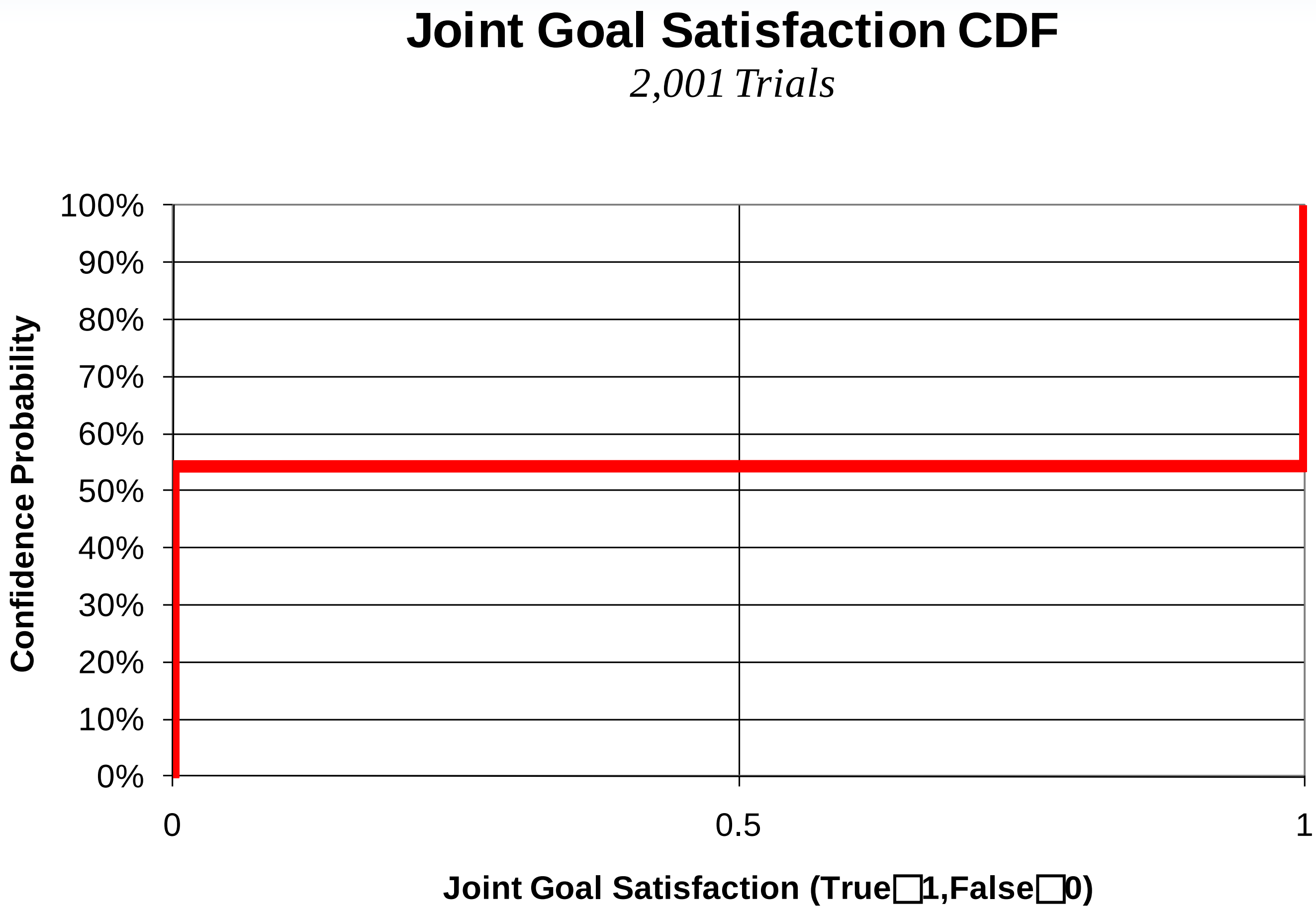
$$\frac{1}{N}\sum_{i=1}^{N} \left(\neg \mathbf{J}_i + \mathbf{J}_i\right) \equiv 100\% \quad \text{(a } \mathbf{J} \text{ element must evaluate to either TRUE or FALSE)}$$

$$\therefore \frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{J}_i\right) = 100\% - \frac{1}{N}\sum_{i=1}^{N}\left(\neg \mathbf{J}_i\right) \text{ and by substitution JCL} = 100\% - \frac{1}{N}\sum_{i=1}^{N}\left(\neg \mathbf{J}_i\right)$$

$$\boldsymbol{F}_{\mathbf{J}}\left(0 < x < 1\right) = \frac{1}{N}\sum_{i=1}^{N} \neg \mathbf{J}_i = 1097/2001 = 54\%$$

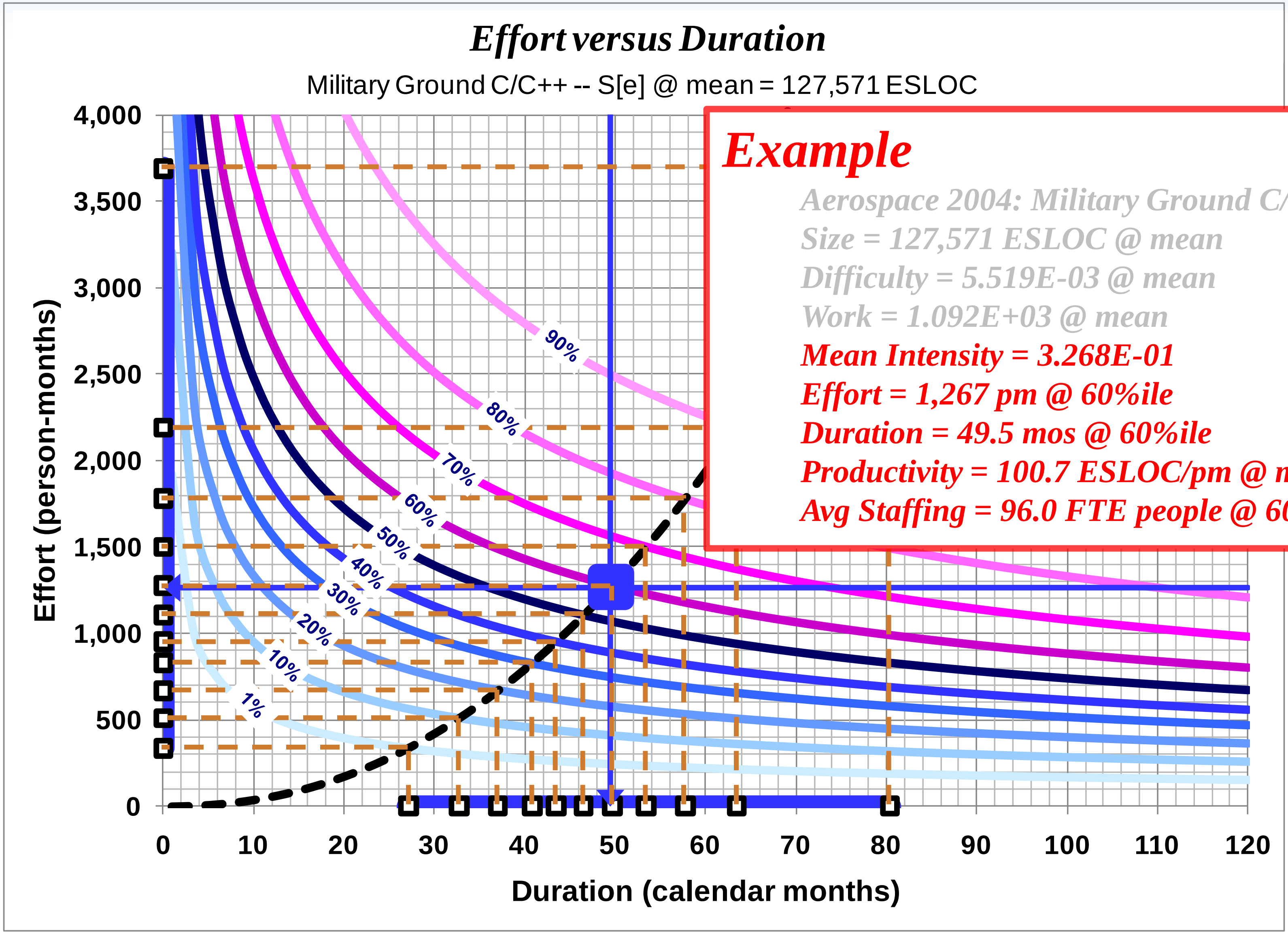
$$\text{JCL} = 100\% - \frac{1}{N}\sum_{i=1}^{N}\left(\neg \mathbf{J}_i\right) = 100\% - 54\% = 46\%$$

JCL of Effort and Duration – Boolean Random Variable Approach



JCL = 100% − 54% = 46%

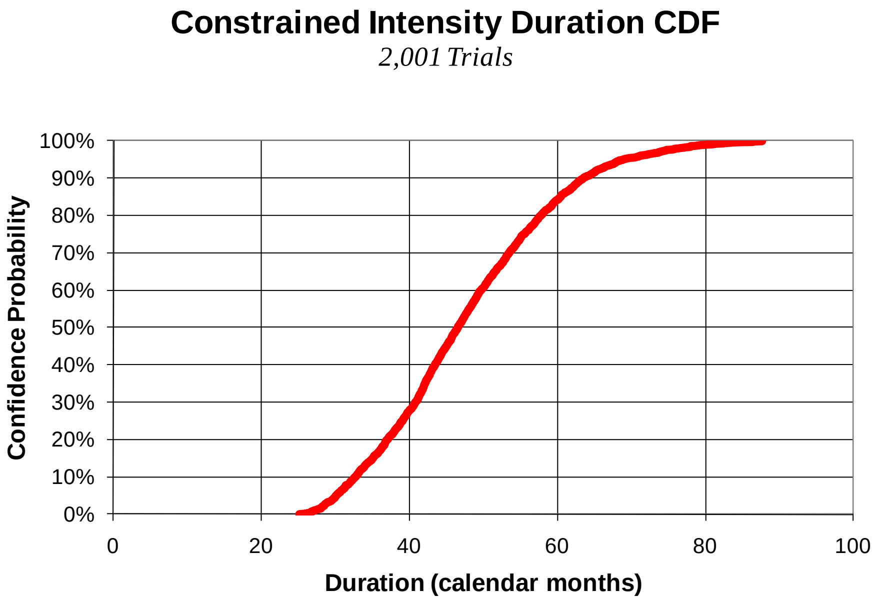
Conditional Confidence Level (CCL) of Effort and Duration – Intensity Gradient

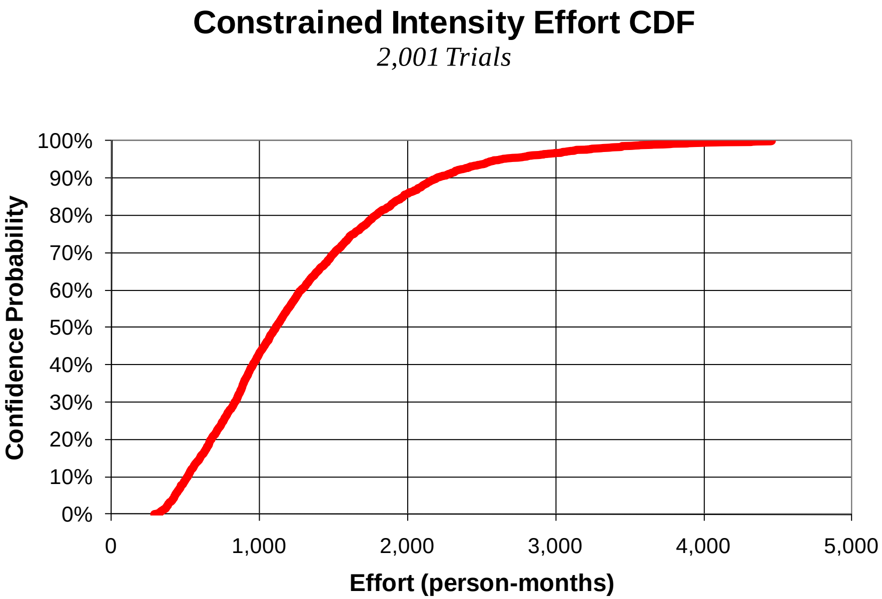


Example

Aerospace 2004: Military Ground C/C++
Size = 127,571 ESLOC @ mean
Difficulty = 5.519E-03 @ mean
Work = 1.092E+03 @ mean
Mean Intensity = 3.268E-01
Effort = 1,267 pm @ 60%ile
Duration = 49.5 mos @ 60%ile
Productivity = 100.7 ESLOC/pm @ mean
Avg Staffing = 96.0 FTE people @ 60%ile

Duration and Effort CDFs





(CCL) of Effort and Duration – Intensity from Effort and Time Constraints

Finding the Intensity that Satisfies an Effort Goal

$$\left[I_{effort\_constraint@p\%} = \left(\frac{E_{effort\_constraint@p\%}^{\alpha_{ET}\alpha_E+\alpha_T}}{\boldsymbol{F}_{\boldsymbol{\Psi}}^{\boldsymbol{-1}}(p)^{\alpha_{ET}}} \right)^{1/\alpha_T} \right]_{\textbf{<dataset name>}}$$

Finding the Intensity that Satisfies a Duration (Time) Goal

$$\left[I_{duration\_constraint@p\%} = \left(\frac{\boldsymbol{F}_{\boldsymbol{\Psi}}^{\boldsymbol{-1}}(p)}{T_{duration\_constraint@p\%}^{\alpha_{ET}\alpha_E+\alpha_T}} \right)^{1/\alpha_E} \right]_{\textbf{<dataset name>}}$$

Note: $\boldsymbol{\Psi} \equiv \text{size} \times \text{difficulty list (work)}$

$= \text{convolved (Monte Carlo) ratio of outcome lists } \mathbf{S} \text{ (size) and } \mathbf{D} \text{ (difficulty)}$

Note: $\boldsymbol{F}_{\boldsymbol{\Psi}}^{\boldsymbol{-1}}(p) \equiv \text{inverse CDF of outcome list } \boldsymbol{\Psi} \text{ at probability } p$

(Ross, 2008a)

(CCL) of Effort and Duration – Intensity from Schedule Compression

Finding the Intensity that Satisfies a Desired Schedule Compression or Stretch - out

$$\left[\%T = \left(\frac{\bar{I}}{I_{\%T}} \right)^{\alpha_E / (\alpha_{ET} \alpha_E + \alpha_T)} \right]_{\textbf{<dataset name>}}$$

$$\therefore \left[I_{\%T} = \bar{I} \left(\frac{1}{\%T} \right)^{(\alpha_{ET} \alpha_E + \alpha_T) / \alpha_E} \right]_{\textbf{<dataset name>}}$$

(Ross, 2008a), (Ross, 2008b)

(CCL) of Effort and Duration – Intensity from a Staffing Constraint

Finding the Intensity that Satisfies a Desired Average Staff Level (People)

$$\bar{\Omega} \equiv \frac{E}{T}$$

$$E = I_{staffing\_constraint@p\%}^{\alpha_T/(\alpha_{ET}\alpha_E+\alpha_T)} \boldsymbol{F}_{\boldsymbol{\Psi}}^{\boldsymbol{-1}}(p)^{\alpha_{ET}/(\alpha_{ET}\alpha_E+\alpha_T)} \mathcal{E}af$$

$$T = \left(\frac{1}{I_{staffing\_constraint@p\%}}\right)^{\alpha_E/(\alpha_{ET}\alpha_E+\alpha_T)} \boldsymbol{F}_{\boldsymbol{\Psi}}^{\boldsymbol{-1}}(p)^{1/(\alpha_{ET}\alpha_E+\alpha_T)} \mathcal{D}af$$

$$\therefore I_{staffing\_constraint@p\%} = \frac{\left(\frac{\mathcal{D}af}{\mathcal{E}af}\bar{\Omega}\right)^{(\alpha_{ET}\alpha_E+\alpha_T)/(\alpha_E+\alpha_T)}}{\boldsymbol{F}_{\boldsymbol{\Psi}}^{\boldsymbol{-1}}(p)^{(\alpha_{ET}-1)/(\alpha_E+\alpha_T)}}$$

(Ross, 2008a), (Ross, 2008b)

Some Possible Solution Scenarios that Are Supported by this Methodology

- **Minimum time or minimum effort**
  - What is the shortest amount of time or what is the cheapest I can develop a certain amount of code with some given confidence (probability of attainment)?
- **Schedule Compression**
  - What is the effort (and hence cost) impact if I compress or stretch out the schedule by some percentage?
- **Duration (Time) As an independent Variable (TAIV); Effort As an Independent Variable (EAIV); Cost As an Independent Variable (CAIV), average Staffing As an Independent Variable (SAIV)**
  - What happens if I constrain one or more of time, effort, cost, and staffing with associated confidence(s)?
- **Size as a Dependent Variable**
  - How much code can I develop within a certain budget and schedule and with some given confidence?
- **Productivity as an Independent Variable**
  - Given a certain schedule, how much money can I save if I can increase productivity by say 5 ESLOC/pm?
- **Confidence Analysis**
  - What is the impact of budgeting at the mean versus budgeting at the 60^{th} percentile?

Use algebra to isolate the CDER variable you wish to solve for and apply any necessary assumptions / constraints to the other CDER variables

QUESTIONS?

References

Aerospace 2004. Gayek, Jonathan E. , Long, Lutrell G., Bell, Kim D., et. al., "Software Cost and Productivity Model", Aerospace Report No. ATR-2004(8311)-1, Aerospace Corporation, El Segundo, CA.

Garvey, Paul 2000. ***Probability Methods for Cost Uncertainty Analysis: A Systems Engineering Perspective*****, Chapman-Hall/CRC-Press, Boca Raton, London, New York, NY.**

Ross, Michael A. 2008a. "Next Generation Software Estimating Framework: 25 Years and Thousands of Projects Later. [ed.] Stephen A. Book and Edward White III. ***Journal of Cost Analysis and Parametrics.*** **s.l. : Society of Cost Estimating and Analysis - International Society of Parametric Analysts, Fall 2008. Vol. 1, 2, pp. 7-30. ISSN 1941-658X.**

Ross, Michael A. 2008b. "Uncertainty to Probability for Cost with Duration Estimating Relationships (CDERs)." ***Proceedings, 2008 SSCAG/SCAF/EACE Meeting*****, Noordwijk, The Netherlands.**

r2™