Language Factors in SEER for Software

The View From 2019

GALORATH

Summary

SEER for Software uses a set of language factors for a large list of programming languages and development frameworks. The recent Agile Planner release includes a major update to these language factors. This presentation will provide an overview of the language factors and how they are developed and updated and trends emerging from the analysis.

Presentation Overview

SEER for Software = SEER-SEM

How SEER-SEM Uses Language Factors

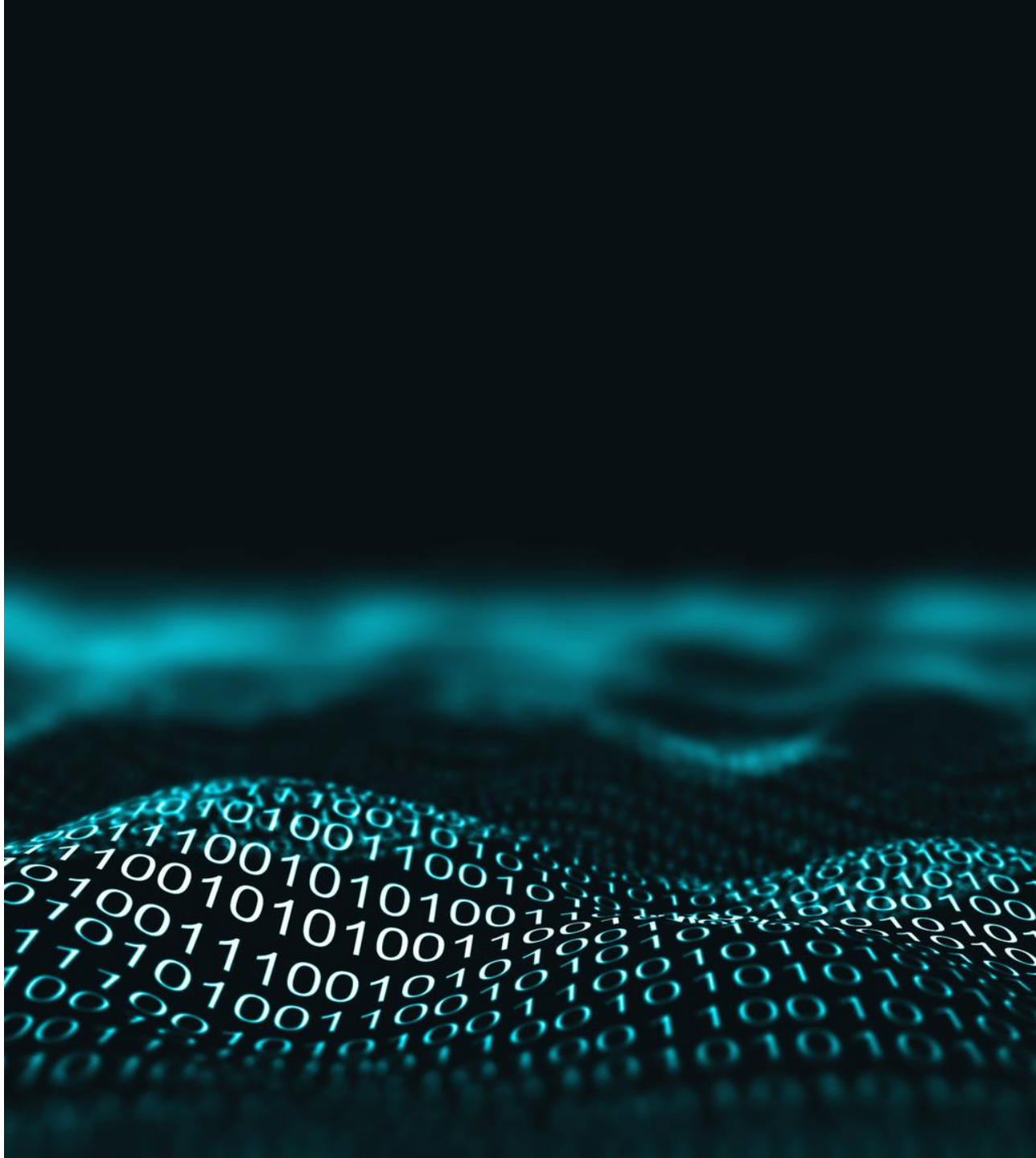
Language factors are key metrics in generating a SEER-SEM estimate, especially when using functional size metrics.

How Language Factors are Developed

Analysis and derivation of language factors follows a process that is specific to SEER-SEM.

What We Have Observed

Trends observed in this update follow general trends of what we have seen in the past.



GALORATH

Language Factors In SEER-SEM

Karen McRitchie
SoCal ISPA Chapter Meeting – El Segundo CA
March 16, 2006

1

SEER™ www.galorath.com

Relating back to the language factor

- Recall the function based size formula:

$$S_e = L_x \times (AdjFactor \times UFP)^{(Entropy/1.2)}$$

$$\hat{S}_e = \hat{L}_x \times (AdjFactor \times UFP)^{(Entropy/1.2)}$$

- In both cases, $AdjFactor$, UFP and $Entropy$ are the same. So the above ratio can be reduced to:

$$EffAdj = \left(\frac{\hat{S}_e}{S_e}\right)^{1.2} \Rightarrow EffAdj = \left(\frac{\hat{L}_x}{L_x}\right)^{1.2}$$

- Solving for $\hat{L}_x$ gives:

$$\hat{L}_x = (EffAdj)^{\frac{1}{1.2}} L_x$$

21

SEER™ www.galorath.com

The language expansion factor

- The SEER-SEM.ini file, has a list of commonly used languages with several factors that apply to estimates

```
[LANGUAGES]
3rd Generation Languages =  51.54  51.54  1.
4th Generation Languages =  23.32  28.88  0.
ABAP =  21.45  25.35  0.88  3.51  2.20  19
```

- The first is the "language expansion factor"
- The others relate to defect estimates (that's another paper entirely)

8

A Recurring Topic

New languages and new data require that we review the factors built into SEER-SEM. New languages are added at each release along with minor changes to existing languages. Major updates lead to major review of the language factors and a table for each major version is included with the application.

Functional Size Metrics

Key ideas about FSM

SIZE METRIC

Size Metric Description	Fast Function Points		
NEW			
External Inputs (EI)	20	20	30
External Outputs (EO)	50	55	60
External Inquiries (EQ)			
External Interface			
Internal Logical File			

SIZE METRIC

Size Metric Description	Mark II Function Point Analysis		
NEW			
Input Data Elements (IT)	100	110	120
Entity References (ER)	18	20	22
Output Data Elements (OT)	30	30	36

SIZE METRIC

Size Metric Description	Cosmic Function Points		
NEW			
CFP - Entries	20	20	22
CFP - Exits	20	20	24
CFP - Reads	100	100	100
CFP - Writes	80	80	100

	FiSMA		
	3	3	5
IEU Input (i)	4	4	6
NIEU Output (o)	4	8	8
Interface to Other Apps (t)	3	3	5
Interface from Other Apps (f)	2	2	4
Data Storage (d)	10	14	16
Algorithmic (a)	22	25	30

SIZE METRIC

Size Metric Description	NESMA Indicative		
NEW			
Internal Data (ILF)	3	4	5
External Data (EIF)	10	12	12

Measures Functionality

The **Function Point** is a normalized measure of software functionality and can be used to measure the functional requirements.

Technology Independent

FSM measures what is being delivered, not how. Therefore it can be used to compare projects across different platforms and languages.

ISO Standard Metrics

ISO/IEC has five recognized functional size metrics. There are many variations, but the 5 standard ones are widely used.

GALORATH

Relating Size to Estimated Effort and Schedule

$$Effort = D^p \left(\frac{S_e}{C_{te}}\right)^q$$

$$Schedule = D^r \left(\frac{S_e}{C_{te}}\right)^s$$

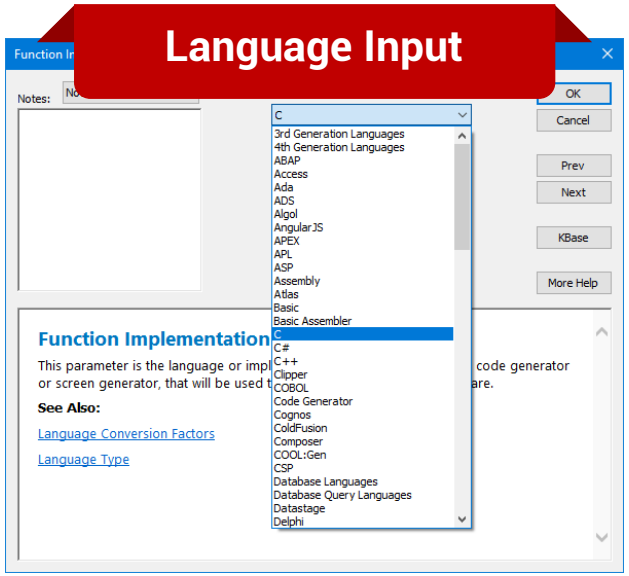
S_e is the effective size

Formulas shown are a simplification, but do represent the basic form. S_e is normalized effective size. D represents the staffing complexity. C_{te} is the effective technology which is the quantification of the many parameters in SEER-SEM that describe people, process, tools and other factors impactive the overall productivity in the development process.

GALORATH

Use of Language in Calculations

Function Implementation Mechanism is SEER-SEM for Language



Language Factors

$$S_e = L_x \times (AdjFactor \times UFP)^r$$

Function Implementation Mechanism is the language input in SEER-SEM. When using functional base size metrics, this is a key input for estimating effort. For Lines of Code, this gives an indication of relative volume, thus indirectly impacting effort.

Language Factors L_x, are used to compute a normalized effective size from a functional size metric. AdjFactor represent an adjustments for platform, application and phase at estimate. That normalized size is in turn used to estimate effort.

GALORATH

The Process of Updating Factors

Recipe for Making Changes

01

Gather Data

Galorath maintains a private repository of nearly 20K projects.

02

Model Using Current Factors

Aided by scripting, generate estimates for all data points

03

Measure Errors

Generate estimate ratios, compute errors and PRED figures

04

Compute New Factors

Updated language factors can be computed based on analysis of estimate ratios.

05

Review Changes

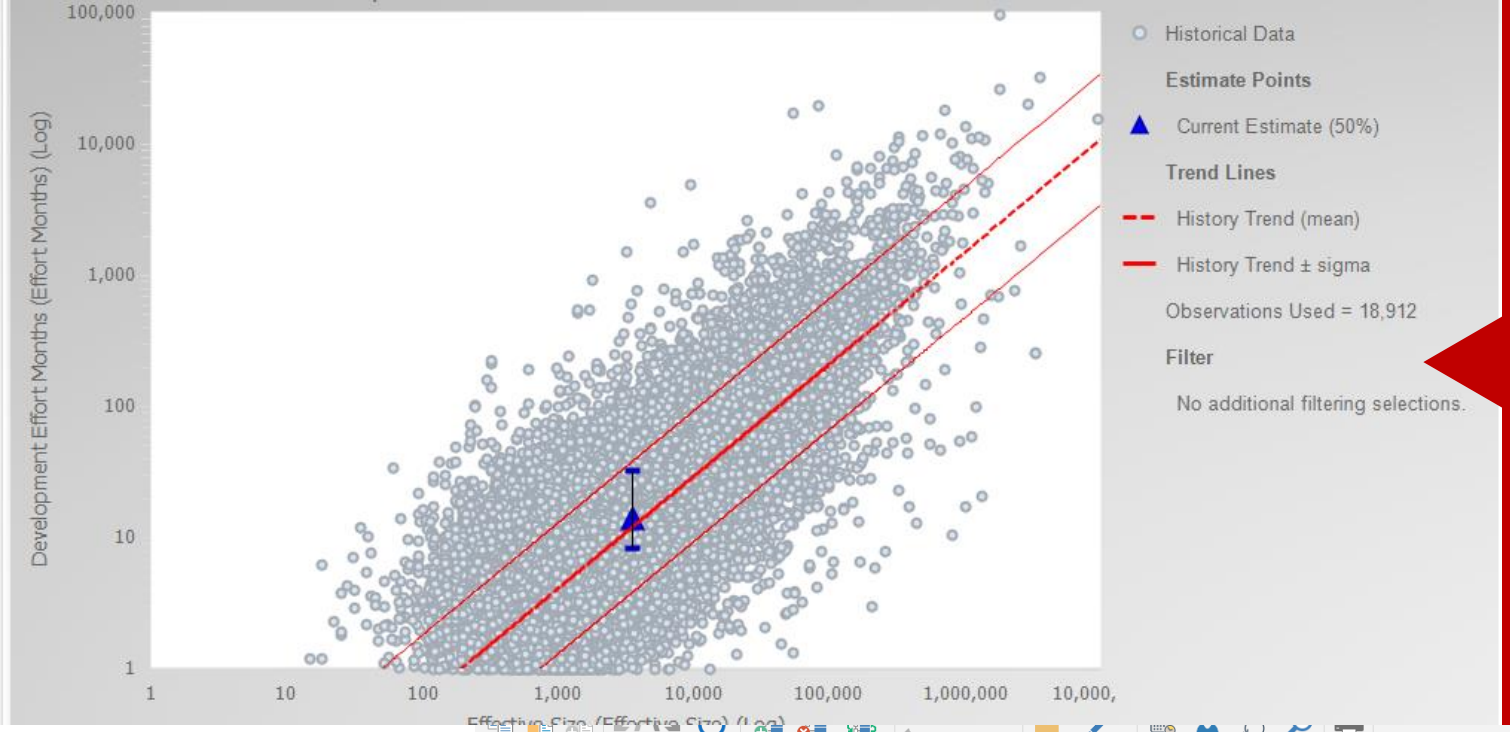
Reasonableness checks, judgement review.

06

Implement the Table

Compile figures for use in estimating.

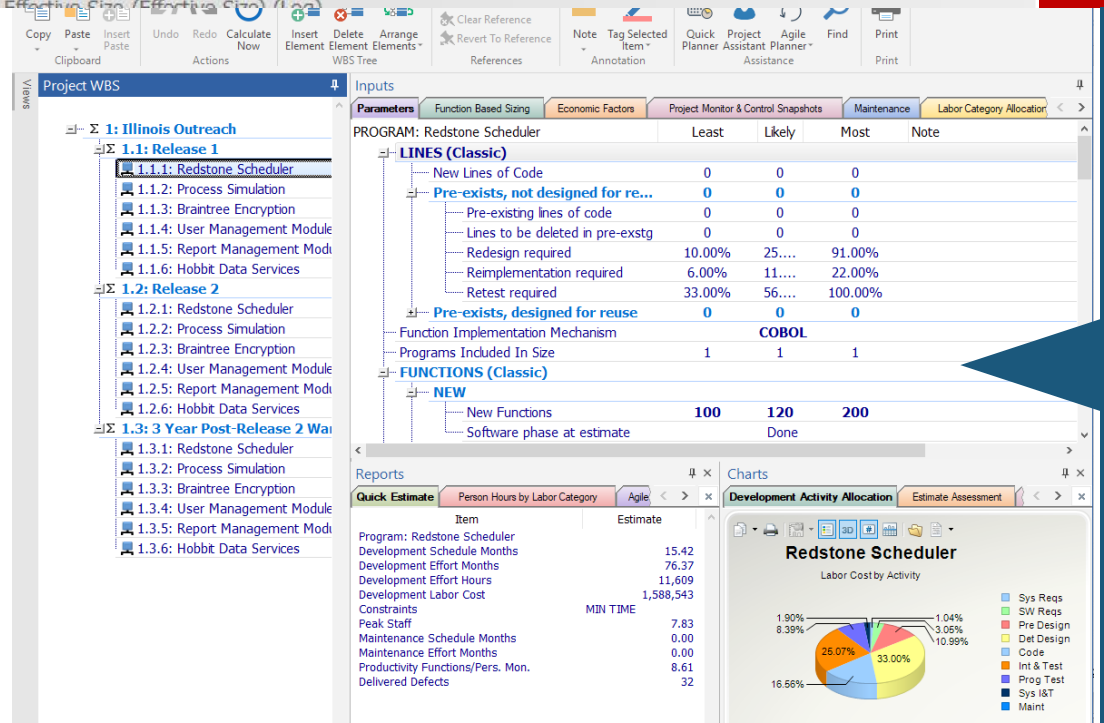
GALORATH



01

Data Roundup

Historical Project Data in SEER-HD format is an important part of the process. This data has size, effort schedule metrics along with descriptive information such as language, platform and application. Other metrics relating to staffing and defects are also included.



02

Model Data Points

Modeling in SEER-SEM is accomplished through use of automated scripting. This will be used to make before & after comparisons.

GALORATH

03

Measure Errors

Estimate Ratios are computed for each observation.

Aggregate factors for each language grouping such as PRED(25%) are also computed.

ration Summary

	Uncal. Estimate	Actual/Required	Difference
ALIBRATION ACTUALS			
Development Actual Effort (hours)	11,608.58	12,000.00	-3%
Development Actual Schedule (months)	15.42	13.00	19%
Maintenance Actual Effort (hours)	0.00	0.00	0%
Other Metrics			
OMPUTED CALIBRATION FACTORS			
Computed Calibration Development Ef...	1.03		
Computec Calibration Development Sc...	0.84		
Computed Calibration Maintenance Eff...	0.00		
Other Computed Metrics			
Computed Calibration Technology Adj...	0.00		
Computed Calibration Complexity Adj...	0.00		

3	75%	1.98
4	50%	0.95
5	25%	0.48
6	10%	0.27
7		Use Macro control-w to copy rightwards
8	Language	-
10	Count	2,153
11	Pred(25%)	20%
12	Language Factor	52.57
14	NewLang_Med	50.44
16	New Pre(x)	20%
17	Change in Avg (Abs)	0.07
18	Change in Avg (%)	4%
19	Change in Pred(x) (Abs)	0%
20	Change in Pred(x) (%)	1%
21	Outliner1	FALSE
22	Outlier2	FALSE

04

Compute New Factors

Proposed New Language Factors for each language are computed based on the median estimate ratio.

$$\hat{L}_x = L_x(EffAdj)^{\frac{1}{q}}$$

GALORATH

Math Behind New Language Factor

How Estimate Ratios Can Be Applied

Before

After

Estimated Effort

$$E = D^p \left(\frac{S_e}{C_{te}}\right)^q$$

$$\hat{E} = D^p \left(\frac{\hat{S}_e}{C_{te}}\right)^q$$

$$S_e = L_x \times (AdjFactor \times UFP)^r$$

$$\hat{S}_e = \hat{L}_x \times (AdjFactor \times UFP)^r$$

Estimate Ratio Compares The Before & After

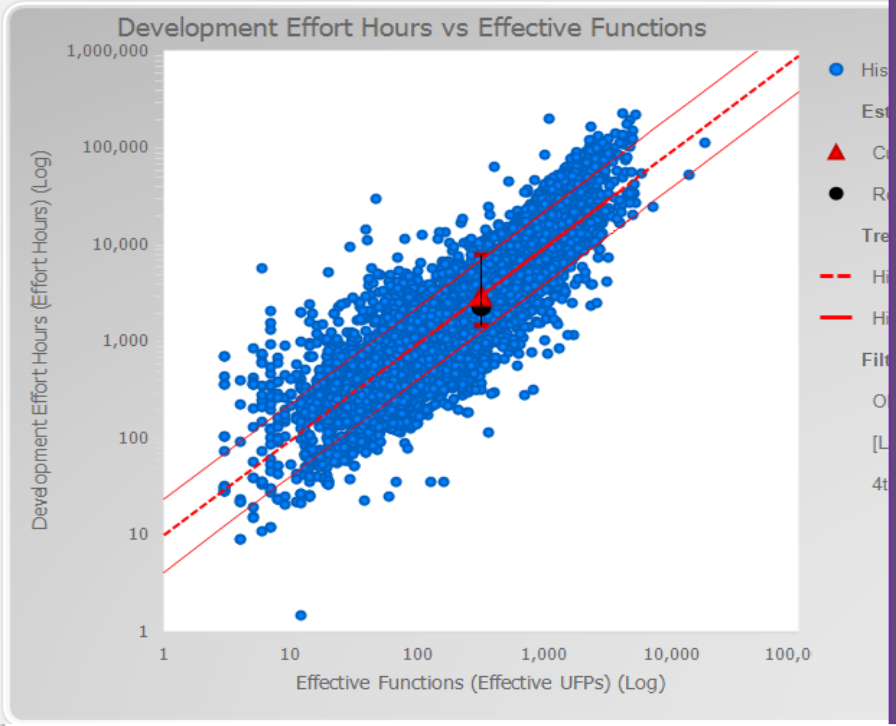
$$EffAdj = \frac{\hat{E}}{E} = \left(\frac{\hat{S}_e}{S_e}\right)^q = \left(\frac{\hat{L}_x}{L_x}\right)^q$$

Most constants cancel out, making the new **Language Factor** a function of the estimate ratio and the effort entropy

New Language Factor

$$\hat{L}_x = L_x (EffAdj)^{1/q}$$

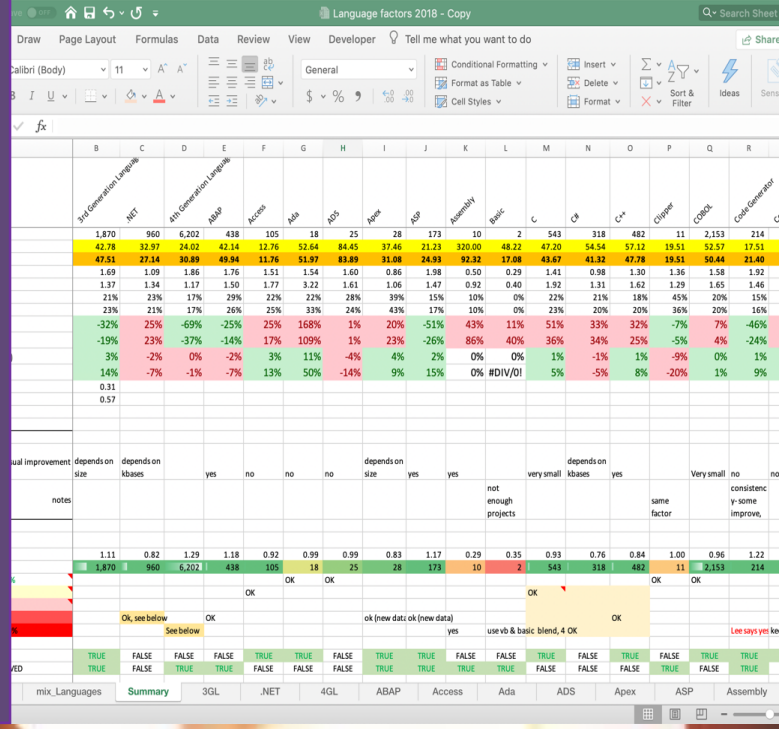
GALORATH



05

Review Changes

Deep Dive into the results to understand the change. This involved looking different cuts of the data including super and subsets. Lots of pivot tables filtering on data set sub groups. Visual checks against scatterplot data were also examined. When results are compiled, a team review of each language change was conducted.




```
;[Table Name as appears on the choice list]
;Language Name =  c1   c2   c3   c4   c5   c6   c7   c8   c9
;
; Language name as it is to appear on the list
; c1 - language expansion factor used with internal calculations against a
;      Must be a positive number.
; c2 - not used
;
; c3 - c9 are used as part of the delivered defect computation
;      If you are a adding a new language, select factors from a similar l
; c3 - Language Defect Difficulty
; c4 - OperatorsPerLOC
; c5 - OperandsPerLOC
; c6 - MaxVocabulary
; c7 - Mean%ofMaxVocabularyUsed
; c8 - UniqueOperandScalingFactor
; c9 - UniqueOperandExponentialFactor

[SEM 8.3]
3rd Generation Languages   =   39.19   51.54   1   3.98   2.09   170.47
4th Generation Languages   =   29.61   28.88   0.85   3.51   1.93   194.54
ABAP   =   49.94   25.35   0.88   3.51   2.2   194.54   0.63   6.89   0.51
Access   =   11.76   14.8   0.91   3.51   2.47   194.54   0.63   6.89   0.
Ada   =   51.97   48.22   1.02   3.98   2.09   170.47   0.63   6.89   0.51
```

06

Implement the Table

SEM 8.3 contains the new table, which is automatically used for estimation

Estimate Continuity

Options for Using Language Factors

New Estimates

SEM 8.3 Language Factors will be used when you create a new estimate.

Legacy Estimates

Earlier Language Factors will be used when you open a legacy estimate. You can change it to use the new table, but it wont happen automatically.

You Can Choose

Language Table Selector is found in the project parameters. You can choose which table to use for any estimate.

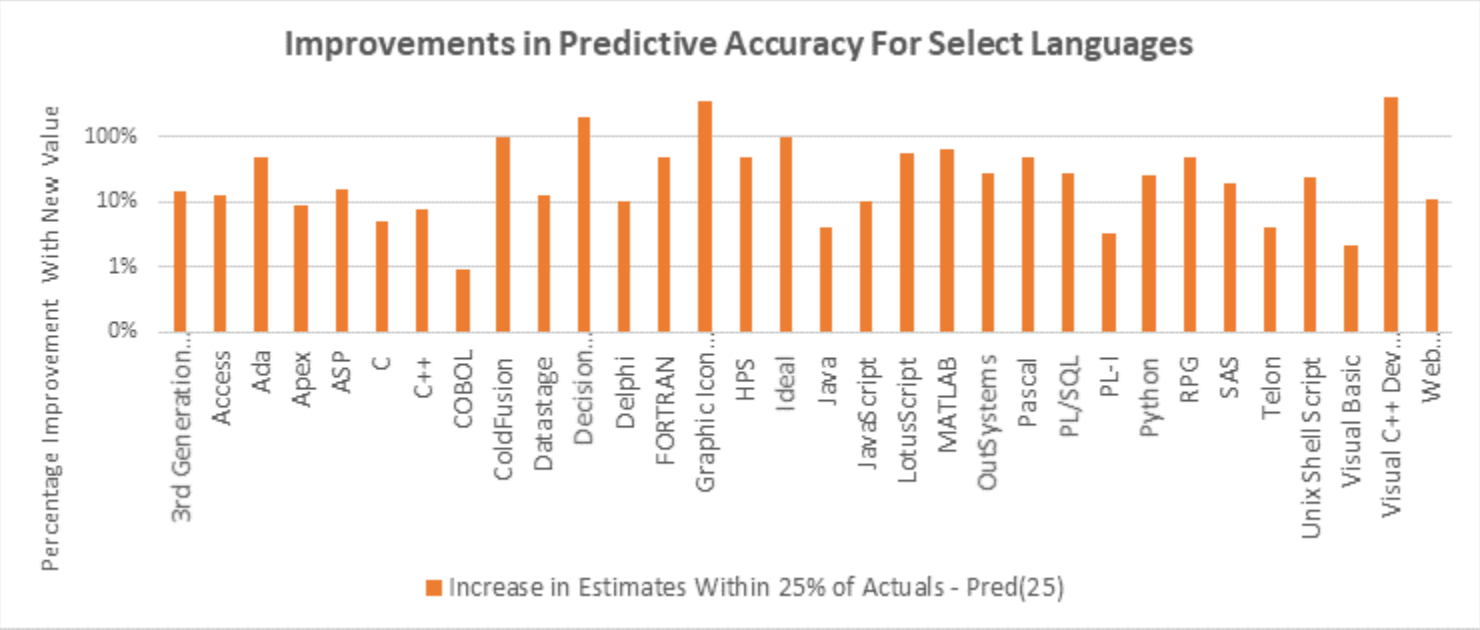
We always encourage customers to use the latest version of the factors and knowledge bases offered. However there are cases and situations where maintaining continuity for an estimate is important. The system is set up to use the new factors for new estimates, but user has control over what is used. In some cases, customers have tailored the language factor list.

GALORATH

Performance Improvement With Language Updates

SEER-SEM captures the latest trends in relative productivity based on programming language

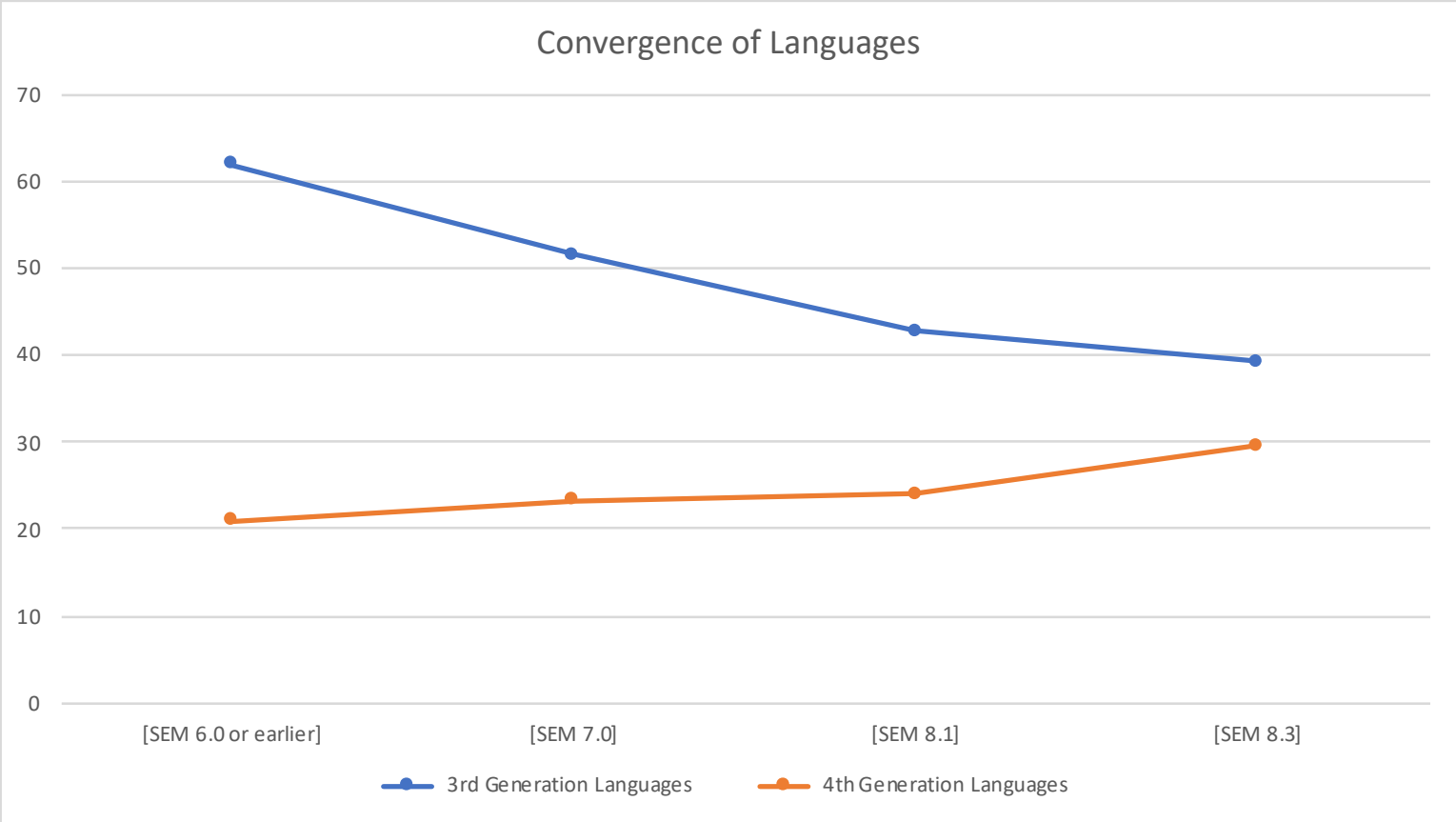
- Several new languages and many more updated in a comprehensive update and review based on over 16,000 records.
- New languages include AngularJS, Cognos, Groovy, Informatica, Jquery, MSBI, Oracle, Oracle Data Integrator, R, SAP Business Objects, TIBCO and Xquery
- Data-derived updates were judged for their ability to improve statistical measures of estimate performance.
- Updates were reviewed to understand the nature of the change, and to ensure that single sources were not introducing specific bias.



GALORATH

3GL and 4GLs

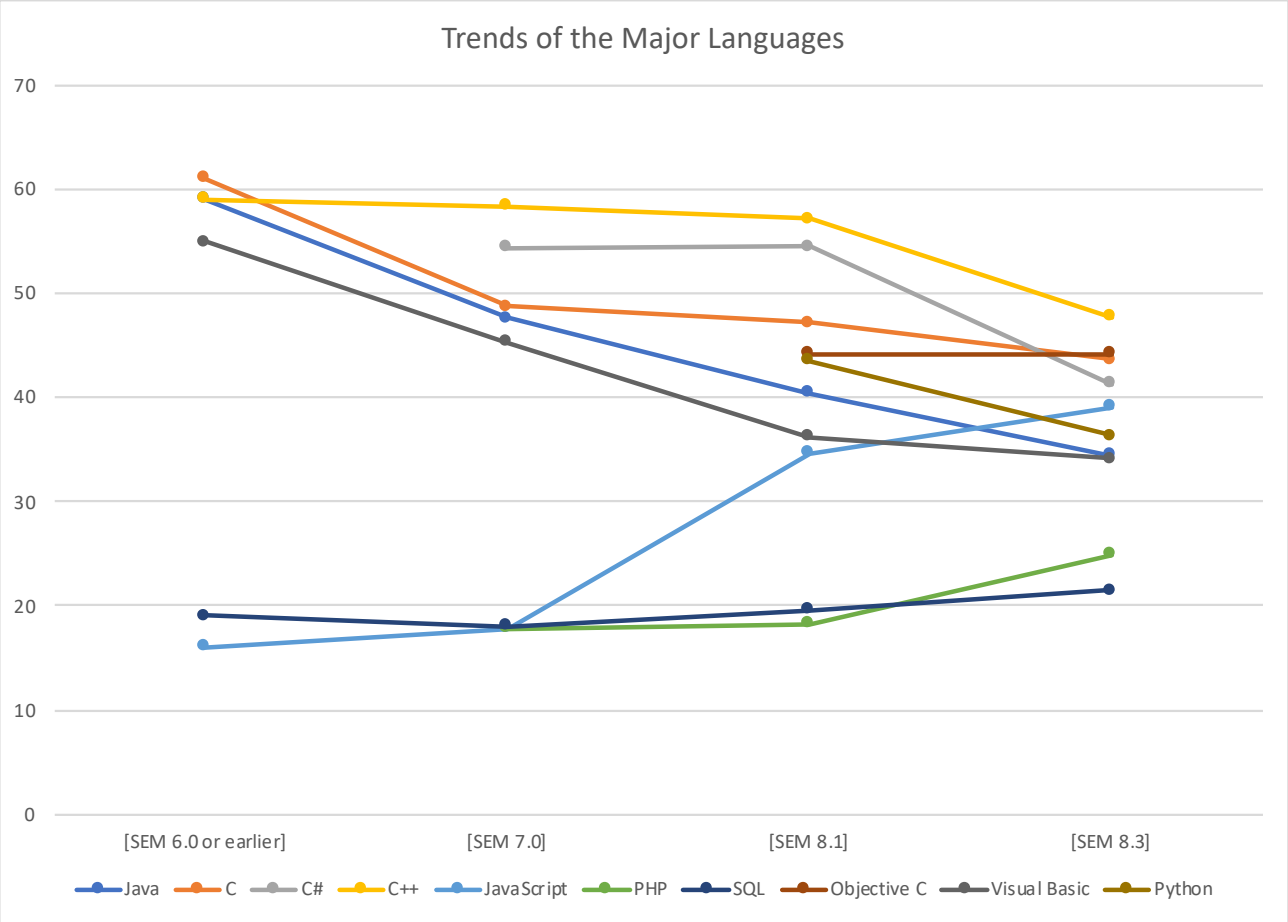
A trend of convergence?



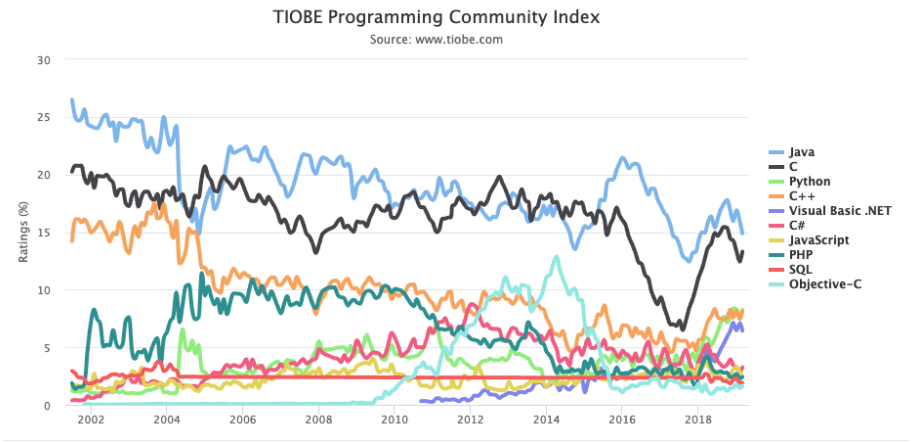
GALORATH

Looking at the Popular Languages

https://www.tiobe.com/tiobe-index/



Looking at 10 most popular languages as indicated by the Tiobe Community Index..



GALORATH

The Study Continues

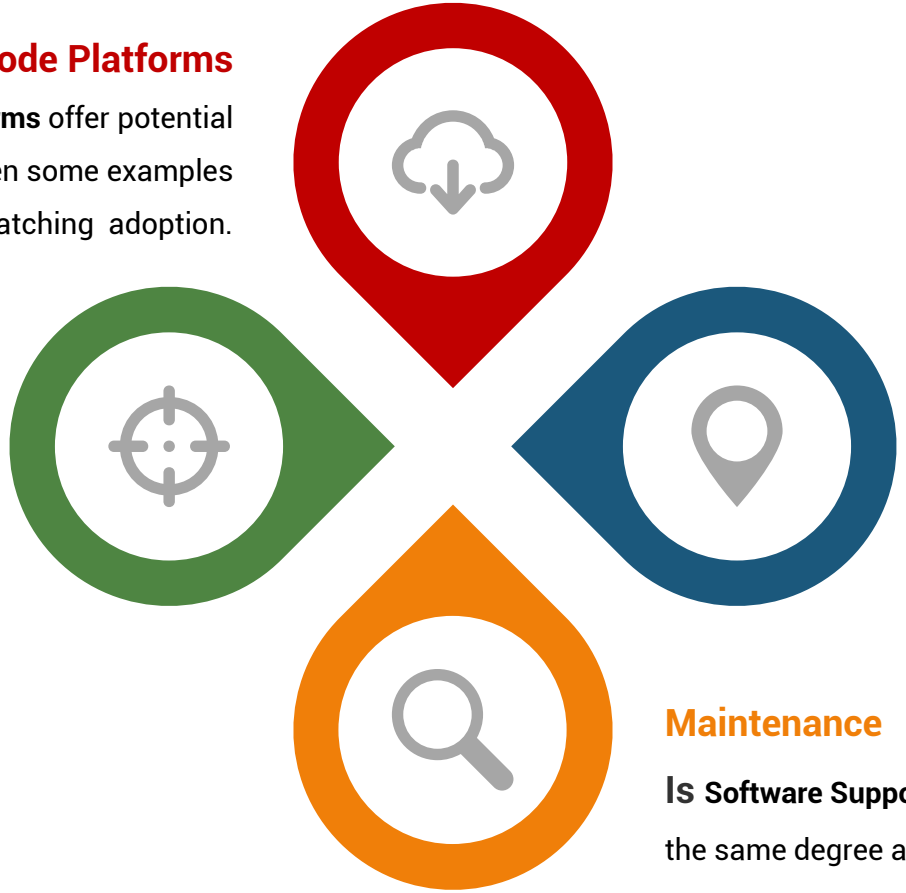
What we are watching

Low Code Platforms

Low Code Development Platforms offer potential productivity boosts. We have seen some examples of this (Outsystems) and are watching adoption.

Frameworks/Language

Frameworks offer efficiency and play a role in productivity. Frameworks combined with different languages are being looked at.



Stabile Workhorses

Minimal Change Detected for Cobol, Ada despite new data.

Maintenance

Is **Software Support** impacted by language to the same degree as it is for development.

GALORATH

The SEER for Software Team

Who Did This

Steve Acelar
Manager Software/IT Products

Alton Ng
Application Analyst

Karen McRitchie
VP of Product Development

Lee Fischman
Senior Director of Product Development

Svetlana Johnson
Senior Systems Engineer