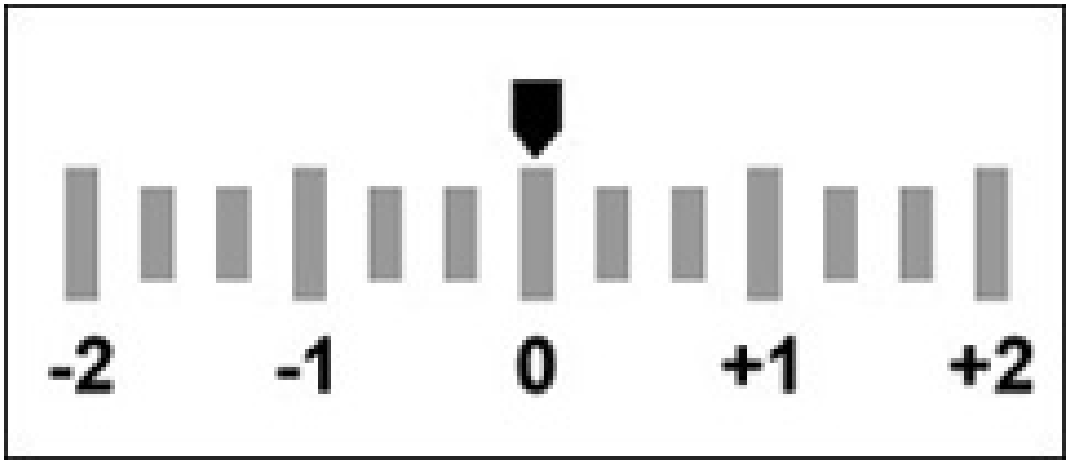
The Zone System of Uncertainty Analysis

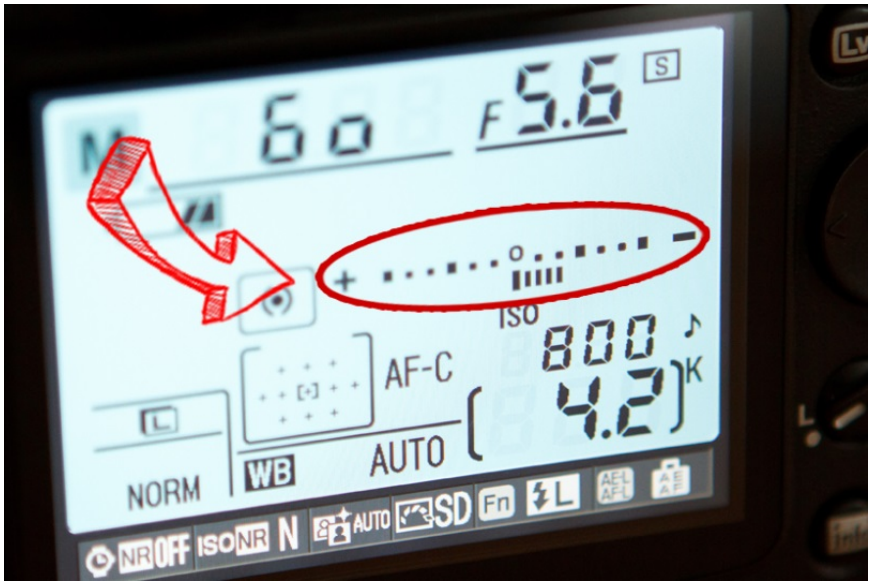
Jeff McDowell, CCEA
ICEAA Professional Development & Training Workshop
June 2016

PRT 207

- Zone Systems
  - Zone System For Photography
  - Zone System for Uncertainty Analysis
- Empirical Values for Zone System
  - CRUAMM Data Analysis
  - Orders of Dispersion
- Result


-2
-1
0
+1
+2

-2
-1
0
+1
+2
Fn

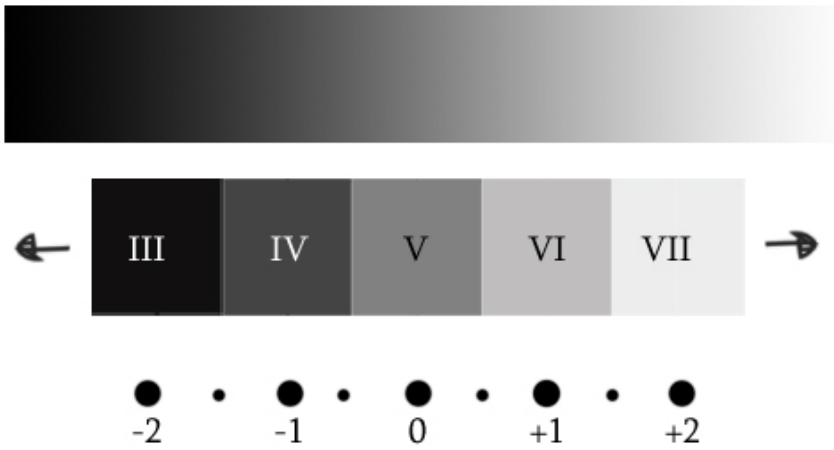

F5.6
S
ISO
800
AF-C
AUTO
4.2
K
NORM
WB
AUTO
NR OFF
ISO NR
N
SD
Fn

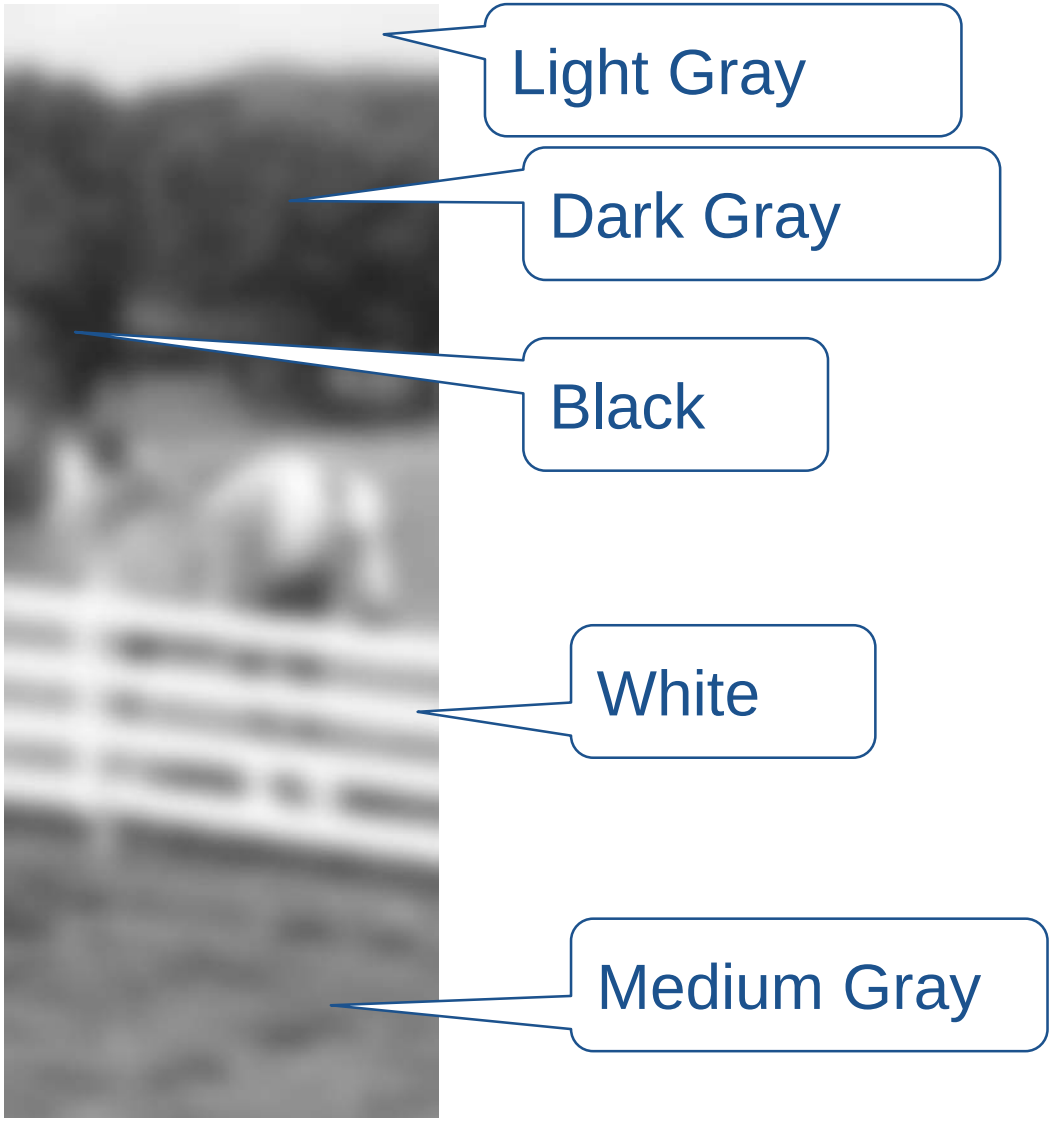
The Zone System's Key Concepts

"The zone system divides a scene into 10 zones on the tonal scale (though there are variations of 9 and 11 zones). Every tonal range is assigned a zone. Every zone differs from the one before it by 1 stop, and from the one following it by 1 stop. So every zone change equals 1 stop difference. Zones are identified by roman numbers, with the middle tone (with 18% reflectance) being a zone V which is zone 5.

"If you point your camera at an area with average reflectance and obtain the correct meter readings (a zero on the light meter), that area would be rendered as average. If you open up your lens or slow down your shutter speed by one stop, that area will become over-exposed by one stop. If you close down your lens or increase your shutter speed by one stop, that area will become under-exposed by one stop.

"Now, we've agreed that an average tone is naturally placed into zone V. If you over-expose it by one stop, you'll be placing it in zone VI (zone 6), causing it to render brighter than it actually is. If you under-expose it by one stop, you'll be placing it in zone IV (zone 4) causing it to render darker than it actually is."




Light Gray
Dark Gray
Black
White
Medium Gray

-2 -1 0 +1 +2

-2 -1 0 +1 +2

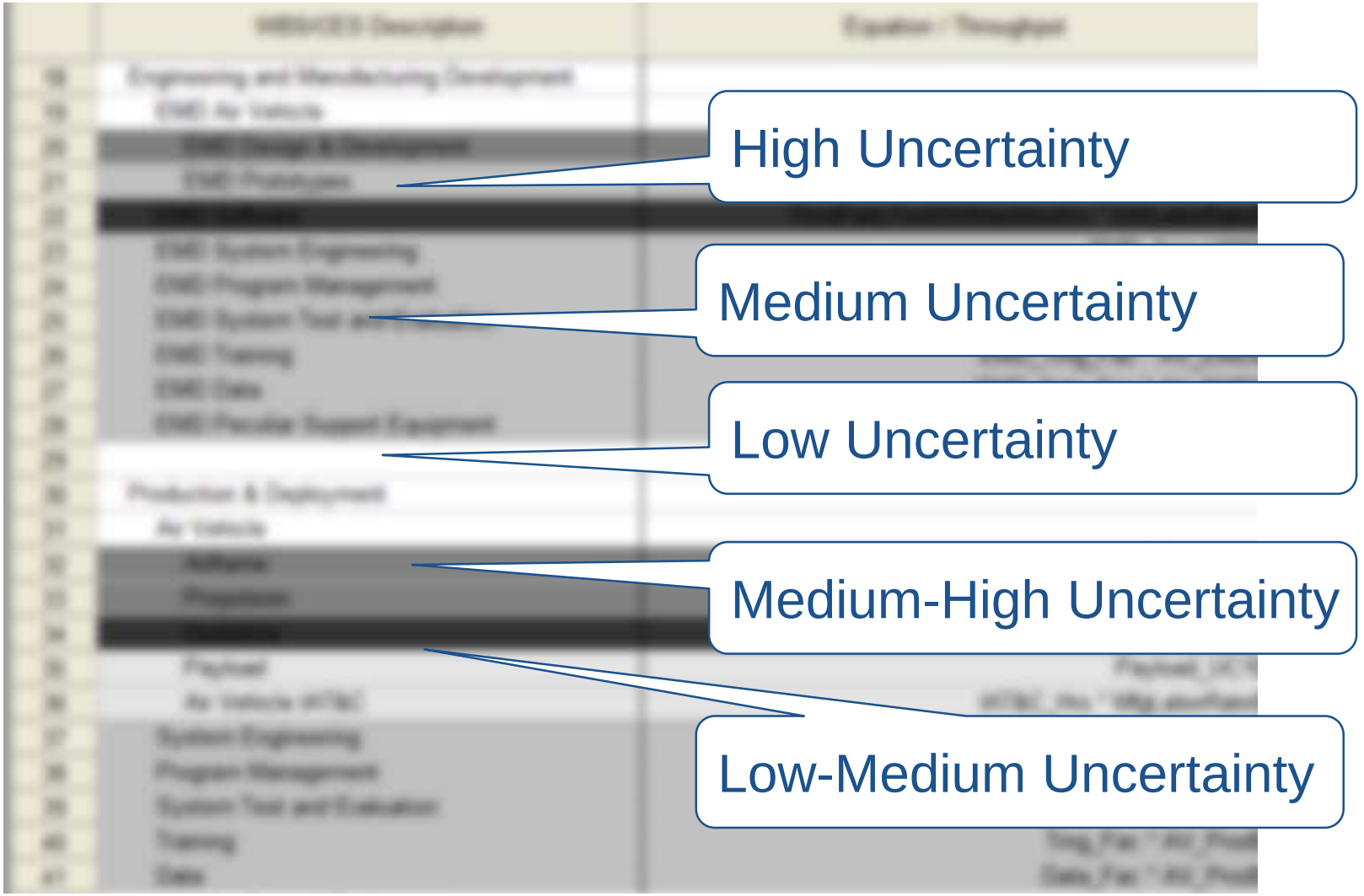
-2 -1 0 +1 +2

-2 -1 0 +1 +2

-2 -1 0 +1 +2

	WBS/CES Description	Equation / Throughput
17	Missile System	
18	Engineering and Manufacturing Development	
19	EMD Air Vehicle	
20	EMD Design & Development	DesignDev$ * EMD_Mths
21	EMD Prototypes	ProdT1 * ProdToEMD_Step
22	EMD Software	ThirdPartyToolSWManMonths * SWLaborRate$
23	EMD System Engineering	EMD_AnnualSE$
24	EMD Program Management	EMD_AnnualPM$
25	EMD System Test and Evaluation	EMD_STE_Fac * Proto_EMD$
26	EMD Training	EMD_Trng_Fac * AV_EMD$
27	EMD Data	EMD_Data_Fac * AV_EMD$
28	EMD Peculiar Support Equipment	EMD_SptEquip_Fac * Proto_EMD$
29		
30	Production & Deployment	
31	Air Vehicle	
32	Airframe	Airframe_UC1$
33	Propulsion	Prop_UC1$
34	Guidance	Guid_UC1$
35	Payload	Payload_UC1$
36	Air Vehicle IAT&C	IAT&C_Hrs * MfgLaborRate$
37	System Engineering	IF(AV_Prod$ <> 0, Prod_AnnualSE$, 0)
38	Program Management	IF(AV_Prod$ <> 0, Prod_AnnualPM$, 0)
39	System Test and Evaluation	IF(ProdQty > 0, STE$, 0)
40	Training	Trng_Fac * AV_Prod$

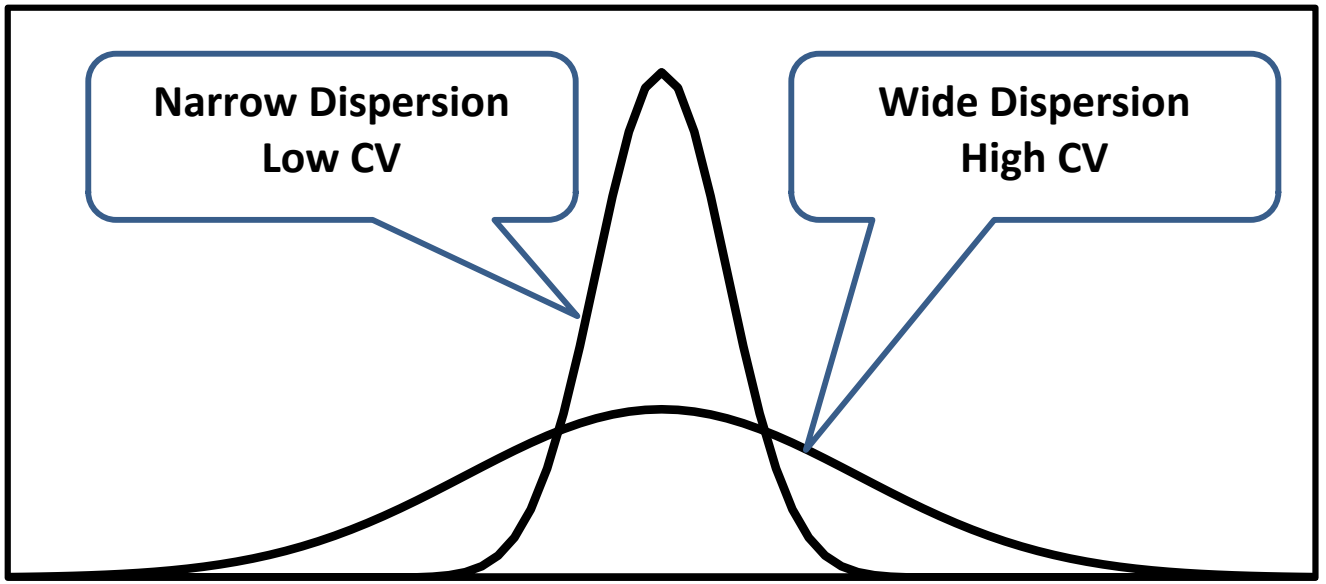
	WBS/ICE3 Description	Equation / Throughput
18	Engineering and Manufacturing Development	
19	EMD Air Vehicle	
20	EMD Design & Development	
21	EMD Prototypes	
22		
23	EMD System Engineering	
24	EMD Program Management	
25	EMD System Test and Evaluation	
26	EMD Training	
27	EMD Data	
28	EMD Peculiar Support Equipment	
29		
30	Production & Deployment	
31	Air Vehicle	
32	Airframe	
33	Propulsion	
34		
35	Payload	
36	Air Vehicle AITbC	
37	System Engineering	
38	Program Management	
39	System Test and Evaluation	
40	Training	
41	Data	


High Uncertainty
Medium Uncertainty
Low Uncertainty
Medium-High Uncertainty
Low-Medium Uncertainty

- An analyst familiar with a program and its life-cycle cost model could step-back from their detailed knowledge of the program and the cost model's inner workings and consider from a bird's eye view the relative uncertainty of each model element:
  - A few elements could be deemed much more uncertain than the other elements.
  - Likewise relative judgment made on a few elements could deem them much more certain than the other elements.
  - An even greater number of elements could be deemed as having medium uncertainty.
  - Inevitably some elements would be deemed to have a degree-of-uncertainty that is not medium nor at either extreme:
    - medium-high or
    - low-medium.
- Thus five gradations are possible for each cost element as depicted using five shades of gray.

Low Uncertainty	Low-Medium Uncertainty	Medium Uncertainty	Medium-High Uncertainty	High Uncertainty

- A metric used throughout this presentation is the fitted Coefficient of Variation (CV). CV is a normalized relative measure of overall dispersion expressed as standard deviation divided by the mean.
  - A low CV is associated with a narrow distribution (low uncertainty)
  - A high CV is associated with a wide distribution (high uncertainty) as depicted.



- The Zone System concept with notional CVs
  - (for illustration only)

	Low Uncertainty	Low-Medium Uncertainty	Medium Uncertainty	Medium-High Uncertainty	High Uncertainty
Dispersion (Notional CV)	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**

- From this construct, an analyst could choose their preferred distribution, determine the parameters corresponding to each of the five CVs and readily assign them to each model element based on their bird's eye assessment, and rapidly assemble a functioning Monte Carlo simulation.

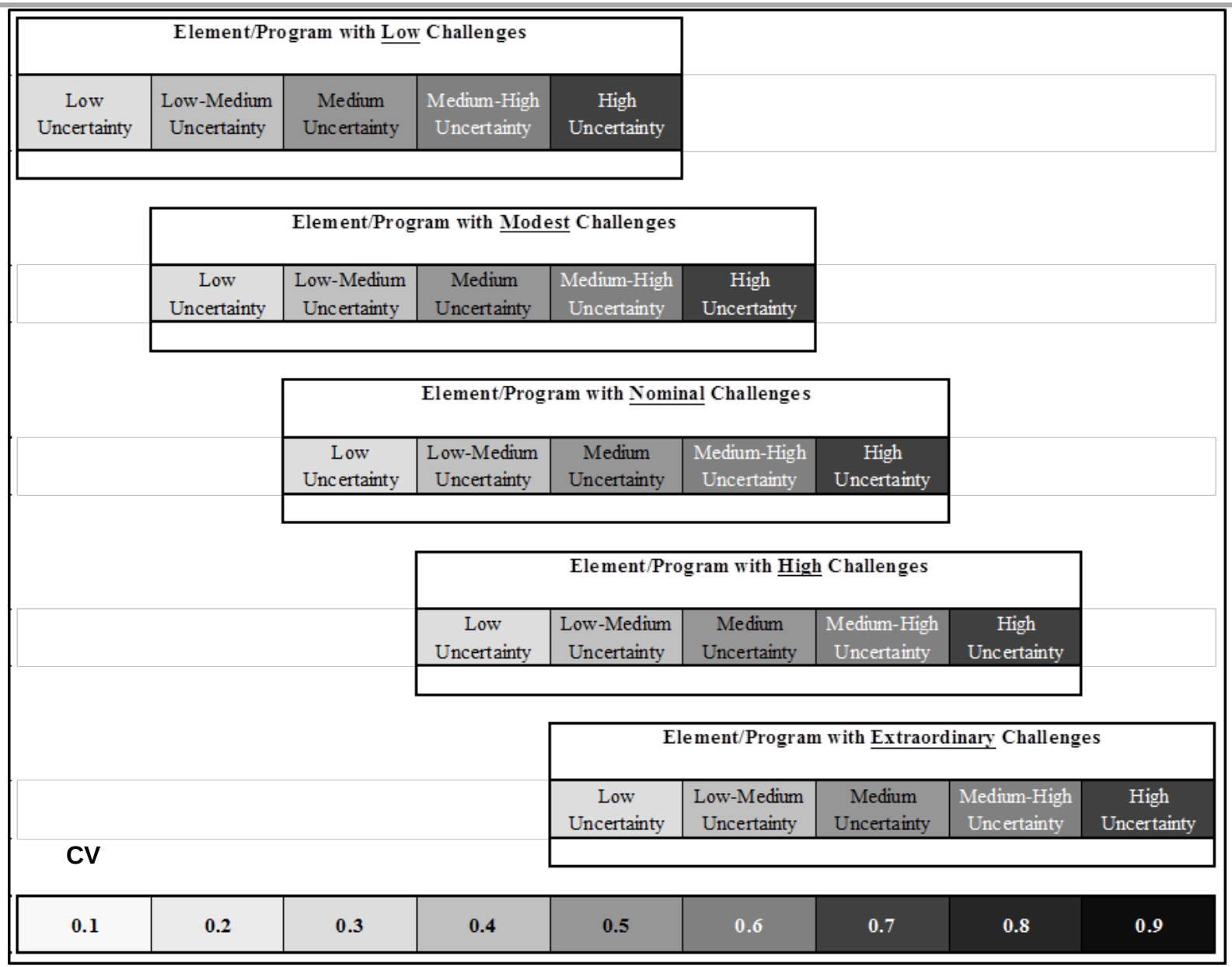
Program Challenges

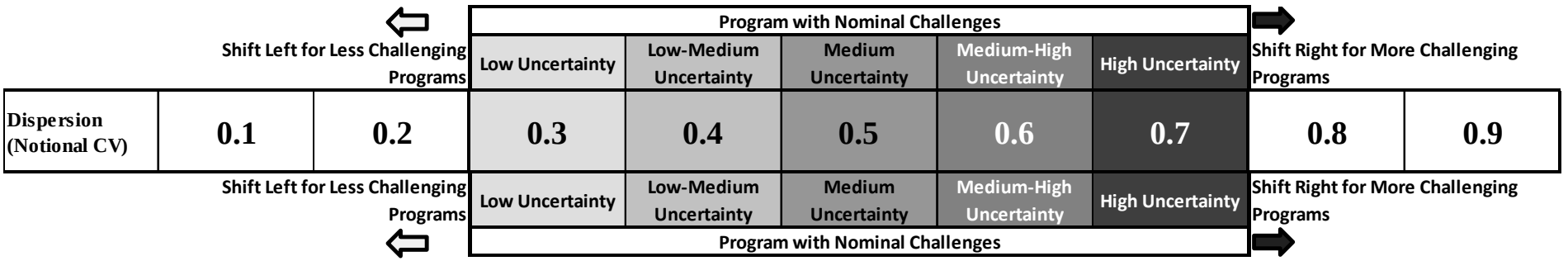
- Consider a routine program.

	Element/Program with Nominal Challenges				
	Low Uncertainty	Low-Medium Uncertainty	Medium Uncertainty	Medium-High Uncertainty	High Uncertainty
Dispersion (Notional CV)	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**

- Such as one with low TRLs, utilizing exotic materials, pursuing an ambitious schedule, with a nebulous mission, with a multi-Service user base, or engaged in a complex acquisition approach.
- Medium uncertain items on this program may be judged to be as widely dispersed as the most highly dispersed items in the first program (CV=0.6).

	Element/Program with High Challenges				
	Low Uncertainty	Low-Medium Uncertainty	Medium Uncertainty	Medium-High Uncertainty	High Uncertainty
Dispersion (Notional CV)	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**


Element/Program with Low Challenges
Low Uncertainty
Low-Medium Uncertainty
Medium Uncertainty
Medium-High Uncertainty
High Uncertainty
Element/Program with Modest Challenges
Low Uncertainty
Low-Medium Uncertainty
Medium Uncertainty
Medium-High Uncertainty
High Uncertainty
Element/Program with Nominal Challenges
Low Uncertainty
Low-Medium Uncertainty
Medium Uncertainty
Medium-High Uncertainty
High Uncertainty
Element/Program with High Challenges
Low Uncertainty
Low-Medium Uncertainty
Medium Uncertainty
Medium-High Uncertainty
High Uncertainty
Element/Program with Extraordinary Challenges
Low Uncertainty
Low-Medium Uncertainty
Medium Uncertainty
Medium-High Uncertainty
High Uncertainty
CV
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9


Shift Left for Less Challenging Programs
Program with Nominal Challenges
Low Uncertainty
Low-Medium Uncertainty
Medium Uncertainty
Medium-High Uncertainty
High Uncertainty
Shift Right for More Challenging Programs
Dispersion (Notional CV)
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Shift Left for Less Challenging Programs
Low Uncertainty
Low-Medium Uncertainty
Medium Uncertainty
Medium-High Uncertainty
High Uncertainty
Shift Right for More Challenging Programs
Program with Nominal Challenges

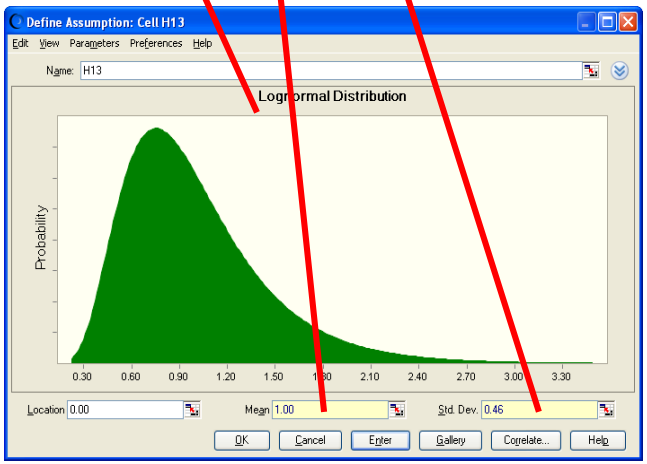
- Replace notional CVs with empirical-based CVs
- Solution: Use observed AFCAA Cost Risk and Uncertainty Analysis Metrics Manual (CRUAMM) CVs to replace the notional CVs
- Better solution: Use observed CRUAMM fitted distributions to select distribution shapes
- CRUAMM References
  - CR-1501/1 Cost Risk and Uncertainty Analysis Metrics Manual (CRUAMM), Alf Smith, Jeff McDowell, Dr. Lew Fichter, Bryan Blevins, Nick DeTore. Tecolote Research Prepared for Air Force Cost Analysis Agency (AFCAA), November 2011. https://www.ncca.navy.mil/tools/csruh/References/080_2012.pdf
  - AFCAA Cost Risk and Uncertainly Analysis Metrics Manual, ICEAA Workshop Presentation, Wilson Rosa AFCAA, Alfred Smith Tecolote Research, Dr Lew Fichter Tecolote Research, Jeff McDowell Tecolote Research, 08 June 2010.
  - Real Data, Real Uncertainty, 45th Annual Department of Defense Cost Analysis Symposium on Cost Analysis and the Downturn February 14-17, 2012, Alfred Smith, Jeff McDowell, Dr. Lew Fichter Tecolote Research.
  - Real Data, Real Uncertainty, Dr. Wilson Rosa AFCAA, Alfred Smith Tecolote Research, Jeff McDowell Tecolote Research, Dr. Lew Fichter Tecolote Research, 27 June 2012.

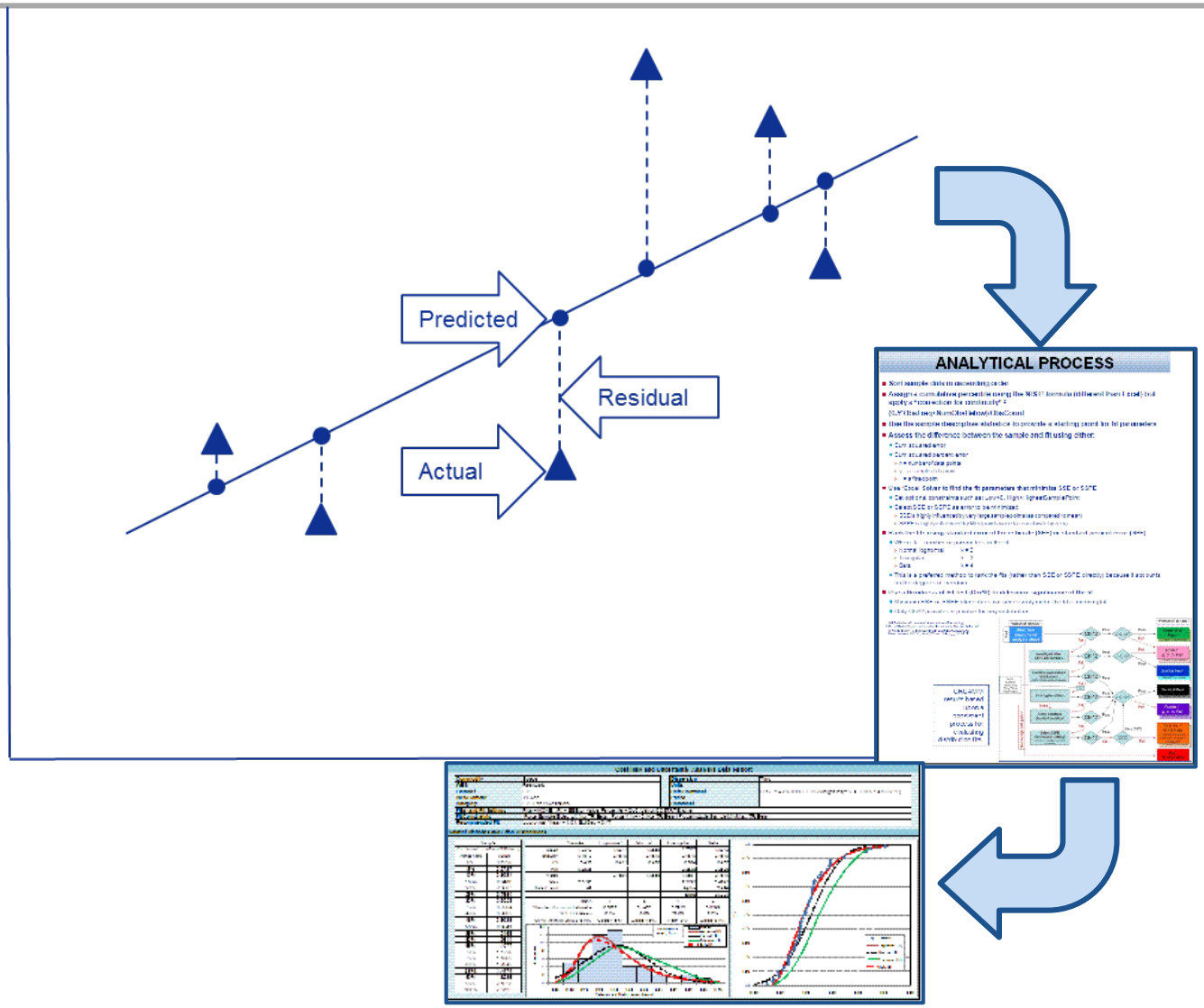
- CRUAMM provides guidance on commonly used distribution shapes and their parameters.
- 1400 distributions fitted for cost data, factors, and CER residuals.
- The distributions are all unitized so that they may be modeled as multipliers of point estimates:

*Cost Element Uncertainty = Your Methodology * Unitized Distribution*

Parameter	Mean	Median	Mode
T1 CER [1]	Lognormal (1.0000, 0.4512)	Lognormal (1.0971, 0.4950)	Lognormal (1.3204, 0.5958)

Sample CRUAMM Table





CER Distributions = f(data homogeneity, use of a cost driver, error minimization)

- Orders of Dispersion is a categorization of residuals based on the types of methodology :
  - CER Methods
  - Factor Methods
  - Central Tendency Methods
- Four distributions addressed:
  - Normal
  - Triangular
  - Lognormal
  - Beta

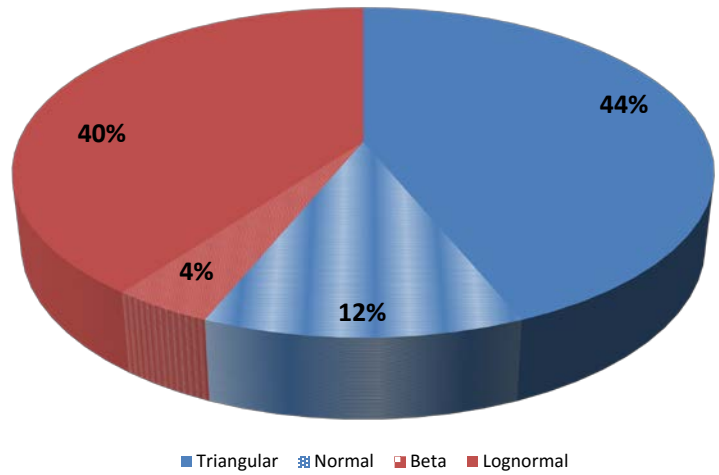
"tidy" - Normal distributions are symmetric and triangular distributions are bounded.

"untidy" - Lognormal distributions have an infinite right tail and beta distributions can have varied shapes.

- As an example consider the case of the one category, CERs.
  - 153 fitted distributions comprise the subset.
  - The fitted CVs range from a low of 0.048 to a high of 1.149.
  - The triangular distribution was the most common having been selected 67 times and representing 44% of the fits. The average CV over those 67 triangular fits is 0.341.

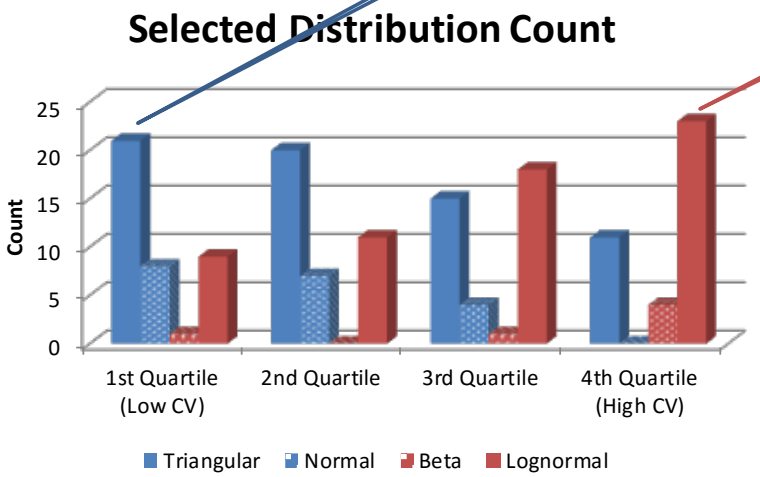
Number of Fitted Distributions	153
Range of Fitted CVs	0.048 - 1.149

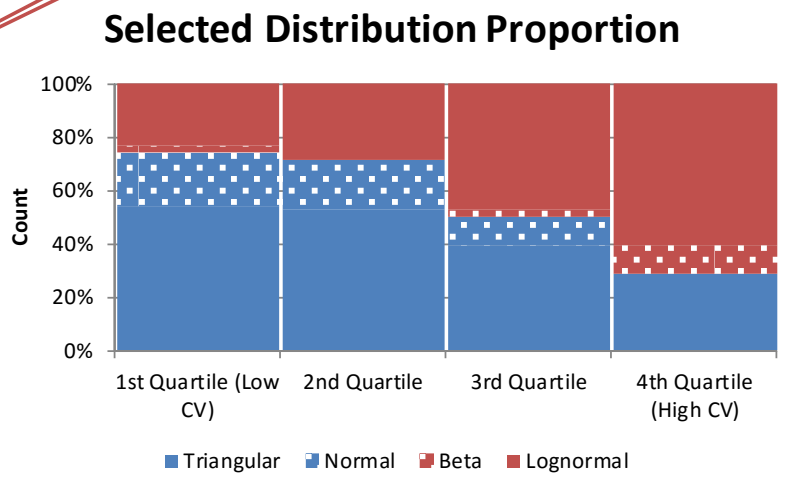
	Selection Count	Average Fitted CV
Triangular	67	0.341
Normal	19	0.270
Beta	6	0.687
Lognormal	61	0.488

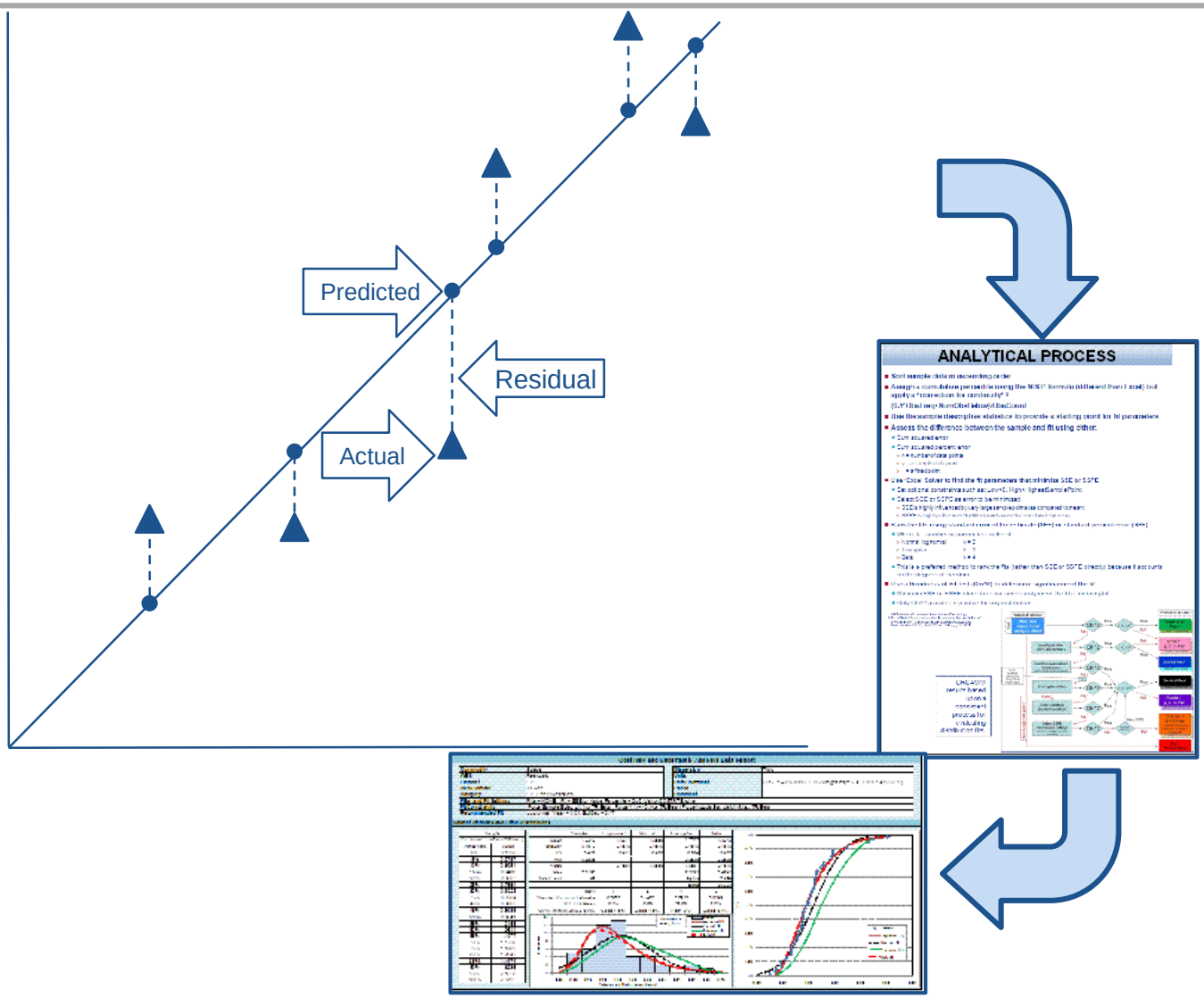


- Results by CV quartile where the first quartile is the lowest dispersion and the fourth quartile is the highest dispersion.
  - In the lowest quartile (where CVs ranged from 0.048 – 0.160) the triangular was the most common fitted distribution being selected 21 times. In the highest quartile (where CVs ranged from 0.516 to 1.149) lognormal was the most common, being selected 23 times.

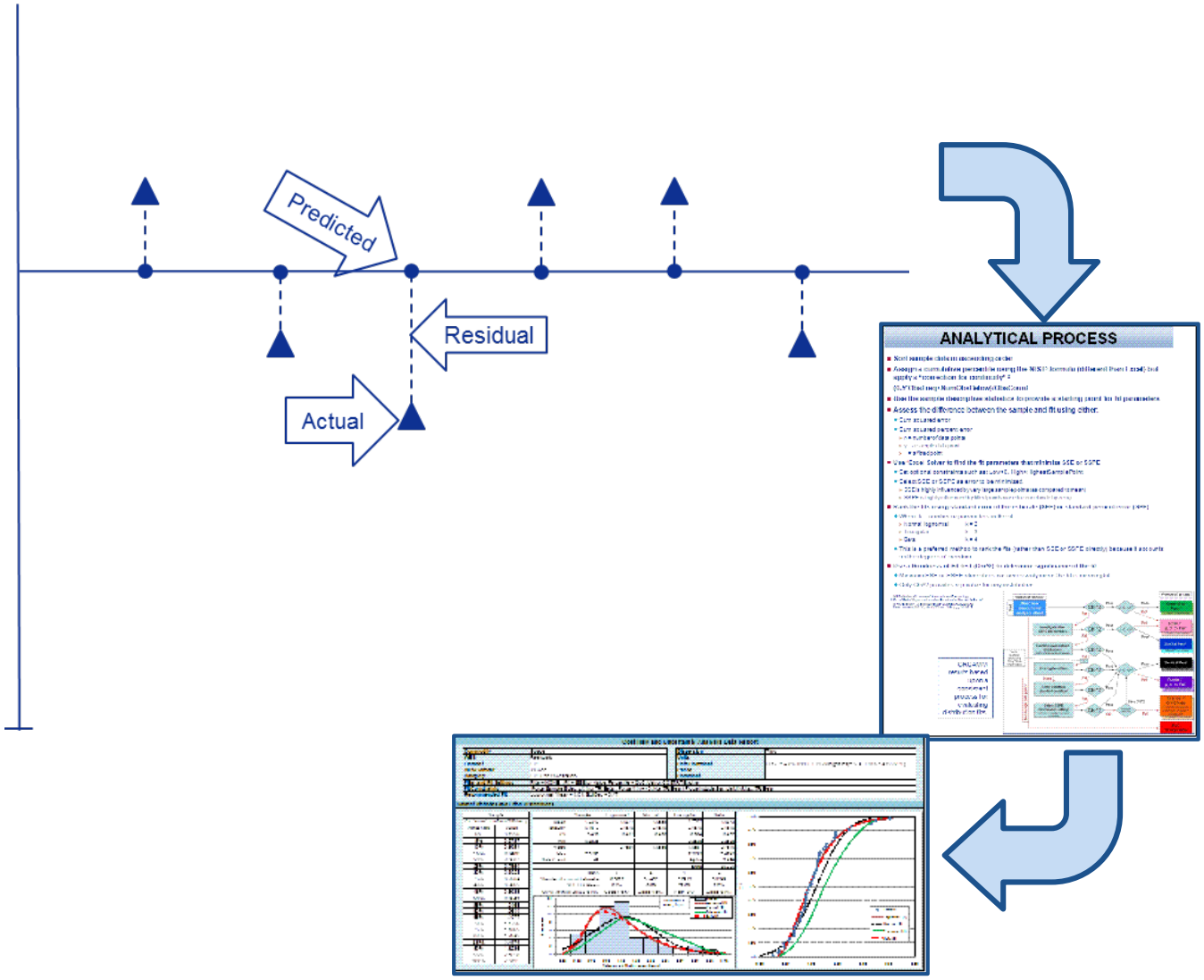
	Selection Count			
	1st Quartile (Low CV)	2nd Quartile	3rd Quartile	4th Quartile (High CV)
CV Min	0.048	0.163	0.390	0.516
CV Max	0.160	0.390	0.512	1.149
Triangular	21	20	15	11
Normal	8	7	4	0
Beta	1	0	1	4
Lognormal	9	11	18	23







Factor Distributions = f(data homogeneity, use of a cost driver)



Data-only Distributions = f(data homogeneity)

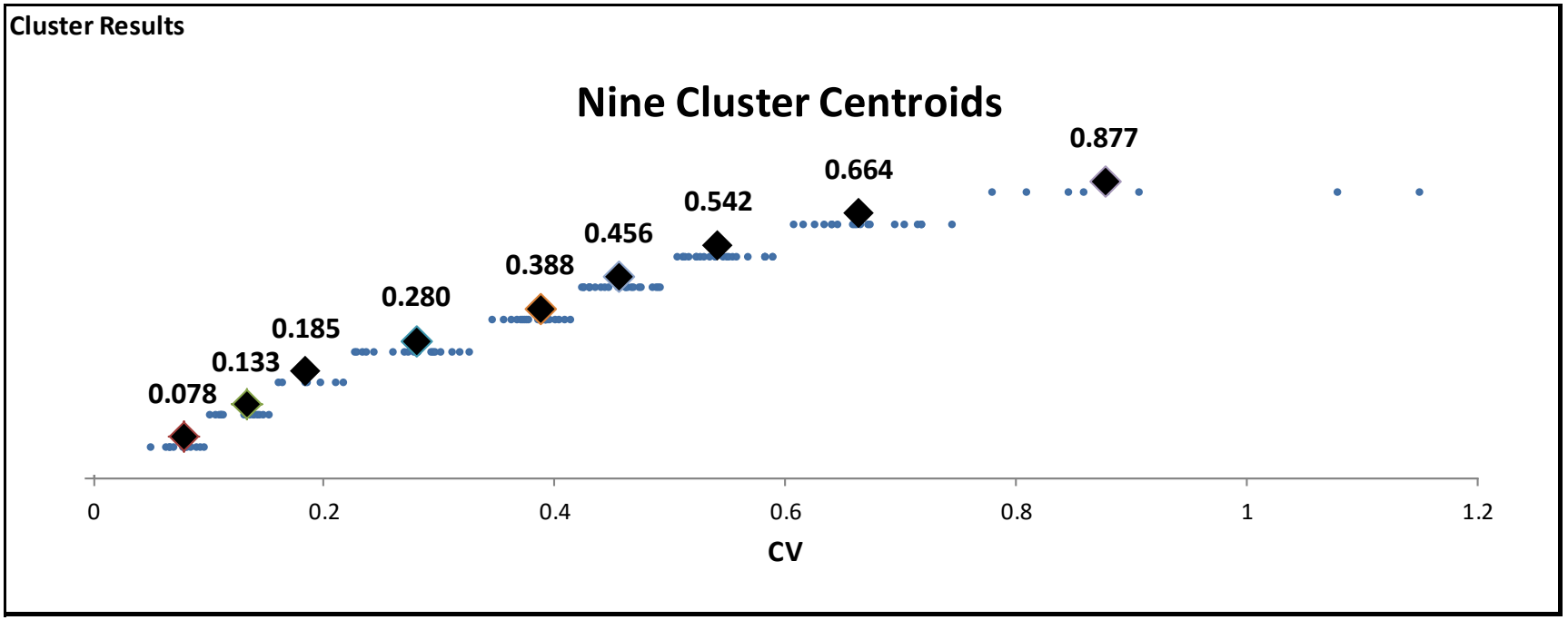
- Given the relative amount of information in each of these three settings, it is reasonable to expect tighter dispersions would emanate from the settings utilizing the greatest information. Therefore, one would expect.
  - CER Dispersion < Factor Dispersion < Data-only Dispersion

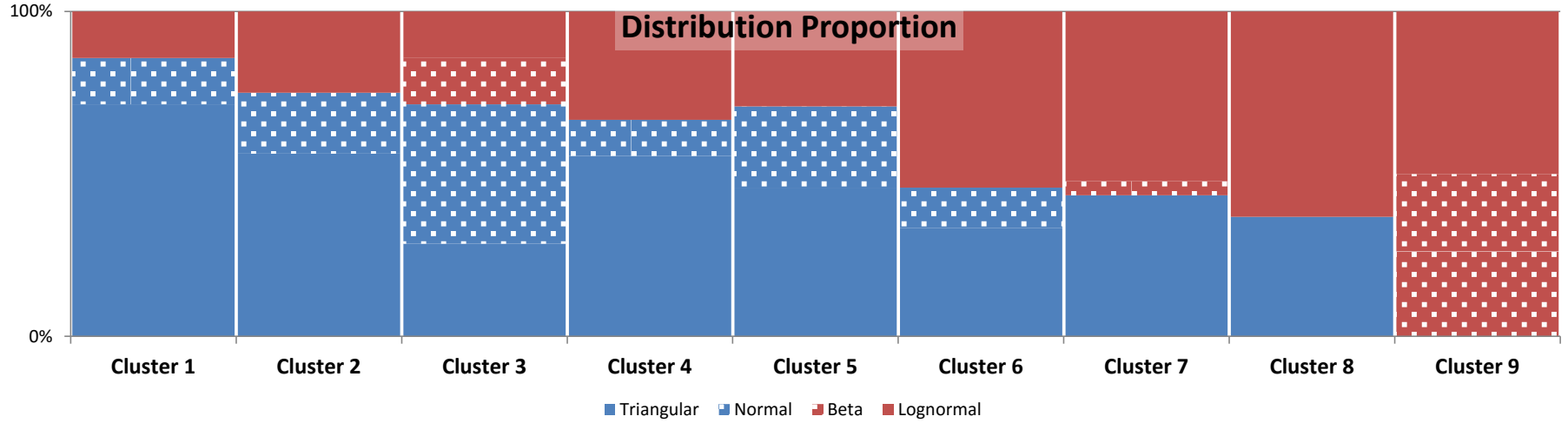
- The lexicon for describing these three settings is:
  - First Order Dispersion (Data-only)
  - Second Order Dispersion (Factor)
  - Third Order Dispersion (CER)

- Clustering is a process of slicing a set of data into a set of meaningful subclasses called clusters. There are numerous clustering techniques which fall into two primary categories:
  - **Hierarchical** clustering algorithms determine not only the cluster content but determine the number of final bins. Since the problem at hand has already been structured as a predetermined number of bins, nine, a hierarchical technique is not necessary.
  - A **Partional** technique, k-means, was selected for use on this task. K-means is a centroid-based method defined as an optimization problem: find the cluster centers and assign the objects to the nearest cluster center, such that the squared distances from the cluster are minimized.
- The k-means algorithm is:
  1. Select K points as initial centroids. The nine initial (or "prototype") centroids are the percentile CV values at 10%, 20% … 90%. Occasionally, these initial values will not lead to a converged solution, in which case the range of CVs would be divided into nine equally-spaced values to obtain alternate initial values.
  2. Form K clusters by assigning each point to its closest centroid. Each fitted distribution's CV is assigned to a cluster.
  3. Recompute the centroid of each cluster.
  4. Repeat 2 and 3 until the centroids do not change (i.e. a converged solution).

- Cluster Results, presents the cluster results with the nine large markers labeled with the centroid value. The smaller markers represent the fitted distributions' CVs in the dataset belonging to each cluster.

  Note each partition is separated with vertical distance for readability; the height of each partition has no meaning other than providing visual separation.





Number of Times Distribution was Selected									
	Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Cluster 5**	**Cluster 6**	**Cluster 7**	**Cluster 8**	**Cluster 9**
Triangular	10	9	2	10	11	8	10	7	0
Normal	2	3	3	2	6	3	0	0	0
Beta	0	0	1	0	0	0	1	0	4
Lognormal	2	4	1	6	7	13	12	12	4
Typical Distribution	**Triangular**	**Triangular**	**Triangular**	**Triangular**	**Triangular**	**Lognormal**	**Lognormal**	**Lognormal**	**Lognormal**

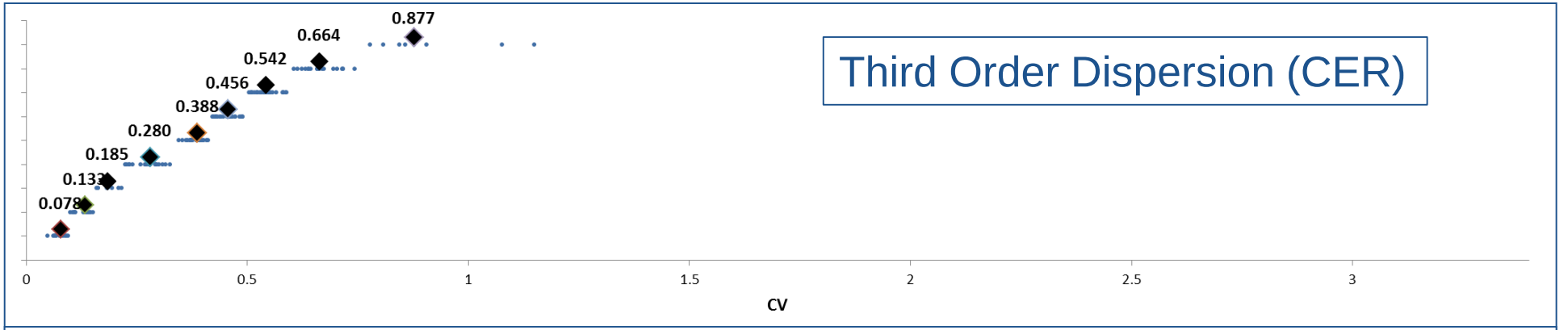
- Typical Distribution (last row), narrows the results to two distribution types:
  - Triangular when either tidy distribution dominates
  - Lognormal when either untidy distribution dominates

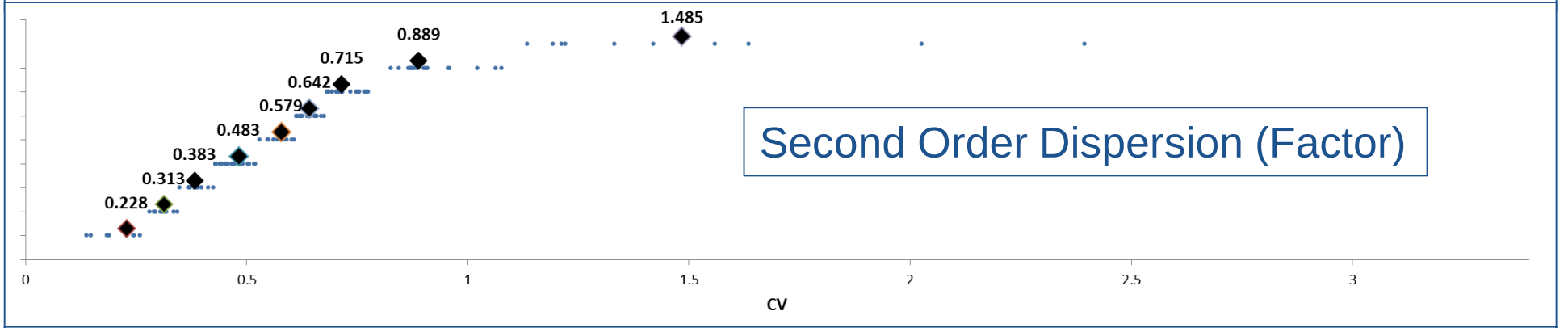
- The first row repeats the CV values from the previous panels.
- The second row presents the skew.
- The third row presents a selected distribution which, in most cases, is the most common distribution within that cluster.
- The content of the remaining rows on this table vary depending on the selected distribution.

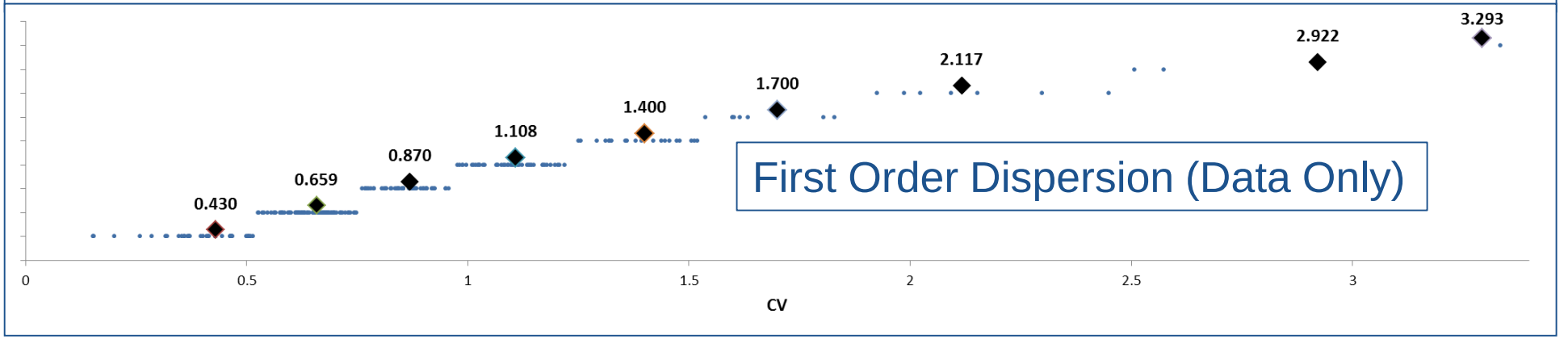
Final Results

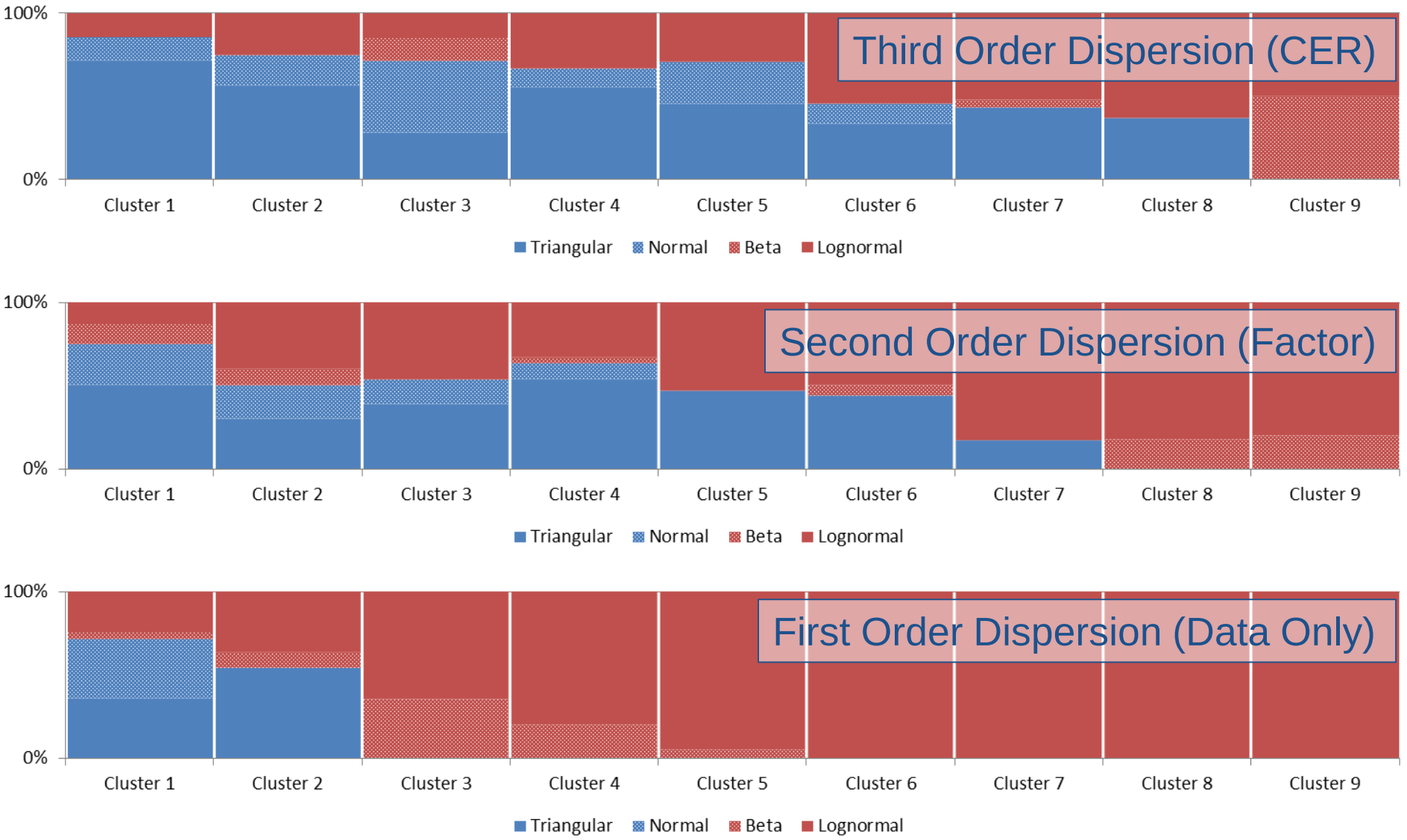
	Methodology Parameters of Interest								
	Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Cluster 5**	**Cluster 6**	**Cluster 7**	**Cluster 8**	**Cluster 9**
CV	0.08	0.13	0.18	0.28	0.39	0.46	0.54	0.66	0.88
Skew[1]	0.46	0.45	0.44	0.25	0.27	0.17	0.10	0.07	0.33
Select Distribution	Triangular	Triangular	Triangular	Triangular	Triangular	Lognormal	Lognormal	Lognormal	Lognormal
Triangular Min	0.813	0.654	0.517	0.518	0.329				
Triangular Max	1.160	1.264	1.465	1.768	2.092				
Lognormal Mean						1.101	1.140	1.210	1.399
Lognormal StdDev						0.502	0.618	0.803	1.227

1 - Skew expressed as CDF at the mode


0.877
0.664
0.542
0.456
0.388
0.280
0.185
0.133
0.078
0
0.5
1
1.5
2
2.5
3
CV
Third Order Dispersion (CER)


1.485
0.889
0.715
0.642
0.579
0.483
0.383
0.313
0.228
0
0.5
1
1.5
2
2.5
3
CV
Second Order Dispersion (Factor)


3.293
2.922
2.117
1.700
1.400
1.108
0.870
0.659
0.430
0
0.5
1
1.5
2
2.5
3
CV
First Order Dispersion (Data Only)


Third Order Dispersion (CER)
100%
0%
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
Cluster 7
Cluster 8
Cluster 9
Triangular
Normal
Beta
Lognormal
Second Order Dispersion (Factor)
100%
0%
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
Cluster 7
Cluster 8
Cluster 9
Triangular
Normal
Beta
Lognormal
First Order Dispersion (Data Only)
100%
0%
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
Cluster 7
Cluster 8
Cluster 9
Triangular
Normal
Beta
Lognormal

Third Order Dispersion (CER)

Methodology Parameters of Interest									
	Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Cluster 5**	**Cluster 6**	**Cluster 7**	**Cluster 8**	**Cluster 9**
CV	0.08	0.13	0.18	0.28	0.39	0.46	0.54	0.66	0.88
Skew[1]	0.46	0.45	0.44	0.25	0.27	0.17	0.10	0.07	0.33
Select Distribution	Triangular	Triangular	Triangular	Triangular	Triangular	Lognormal	Lognormal	Lognormal	Lognormal
Triangular Min	0.813	0.654	0.517	0.518	0.329				
Triangular Max	1.160	1.264	1.465	1.768	2.092				
Lognormal Mean						1.101	1.140	1.210	1.399
Lognormal StdDev						0.502	0.618	0.803	1.227

Second Order Dispersion (Factor)

Methodology Parameters of Interest									
	Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Cluster 5**	**Cluster 6**	**Cluster 7**	**Cluster 8**	**Cluster 9**
CV	0.23	0.31	0.38	0.48	0.58	0.64	0.72	0.89	1.48
Skew[1]	0.41	0.19	0.17	0.18	0.06	0.09	0.07	0.20	0.30
Select Distribution	Triangular	Triangular	Triangular	Triangular	Lognormal	Lognormal	Lognormal	Lognormal	Lognormal
Triangular Min	0.701	0.718	0.435	0.273					
Triangular Max	1.517	1.981	2.131	2.510					
Lognormal Mean					1.153	1.188	1.238	1.363	1.869
Lognormal StdDev					0.667	0.763	0.885	1.212	2.775

First Order Dispersion (Data Only)

Methodology Parameters of Interest									
	Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Cluster 5**	**Cluster 6**	**Cluster 7**	**Cluster 8**	**Cluster 9**
CV	0.43	0.66	0.87	1.11	1.40	1.70	2.12	2.92	3.29
Skew[1]	0.34	0.10	0.26	0.24	0.25	0.27	0.32	0.35	0.38
Select Distribution	Triangular	Triangular	Lognormal	Lognormal	Lognormal	Lognormal	Lognormal	Lognormal	Lognormal
Triangular Min	0.181	0.049							
Triangular Max	2.138	3.161							
Lognormal Mean			1.315	1.480	1.714	1.932	2.356	2.731	3.550
Lognormal StdDev			1.144	1.640	2.399	3.285	4.989	7.978	11.692

	Shift Left for Less Challenging Programs		Program with Nominal Challenges					Shift Right for More Challenging Programs	
			Low Uncertainty	Low-Medium Uncertainty	Medium Uncertainty	Medium-High Uncertainty	High Uncertainty		
CER Methods	TRIANGULAR (0.813, 1, 1.160)	TRIANGULAR (0.654, 1, 1.264)	TRIANGULAR (0.517, 1, 1.465)	TRIANGULAR (0.518, 1, 1.768)	TRIANGULAR (0.329, 1, 2.092)	LOGNORMAL (1.101, 0.502)	LOGNORMAL (1.140, 0.618)	LOGNORMAL (1.210, 0.803)	LOGNORMAL (1.399, 1.227)
Factor Methods	TRIANGULAR (0.701, 1, 1.517)	TRIANGULAR (0.718, 1, 1.981)	TRIANGULAR (0.435, 1, 2.131)	TRIANGULAR (0.273, 1, 2.510)	LOGNORMAL (1.153, 0.667)	LOGNORMAL (1.188, 0.763)	LOGNORMAL (1.238, 0.885)	LOGNORMAL (1.363, 1.212)	LOGNORMAL (1.869, 2.775)
Central Trendency Methods	TRIANGULAR (0.181, 1, 2.138)	TRIANGULAR (0.049, 1, 3.161)	LOGNORMAL (1.315, 1.144)	LOGNORMAL (1.480, 1.640)	LOGNORMAL (1.714, 2.399)	LOGNORMAL (1.932, 3.285)	LOGNORMAL (2.356, 4.989)	LOGNORMAL (2.731, 7.978)	LOGNORMAL (3.550, 11.692)
	Shift Left for Less Challenging Programs		Low Uncertainty	Low-Medium Uncertainty	Medium Uncertainty	Medium-High Uncertainty	High Uncertainty	Shift Right for More Challenging Programs	
			Program with Nominal Challenges						

Legend
TRIANGULAR(a, b, c) = Triangular(minimum, most-likely, maximum)
LOGNORMAL(a, b) = Lognormal(mean, standard-deviation)

- Use this table to select a distribution shape and distribution parameters via these four steps:
  1. Determine the degree of program challenge. Use judgment or use a commodity-specific scoring matrix or complexity calculator.
  2. Identify the category row that most closely matches each element.
  3. Determine the relative uncertainty of this cost model element compared to the rest of the cost model.
  4. Use the distribution shape and parameters shown in the cell intersected by the first three choices.

- Presentation Summary
  - The Zone System framework was presented
  - The Zone System was populated with distributions drawn from an examination of the CRUAMM body of work
  - The concept of Orders of Dispersion was illustrated
- The Zone System is an outputs-based method for performing cost uncertainty analysis.
- While the Zone System may be useful as a quick-turn method, or to "gut-check" another's estimate, or as a source for relative uncertainty bounds, it is not intended to supplant the current best practices of detailed inputs-based uncertainty analysis using program-specific data.