Integrating Excel, R, and ACEIT for a Comprehensive Analysis Package

Karen Mourikas SD&E - Experimentation

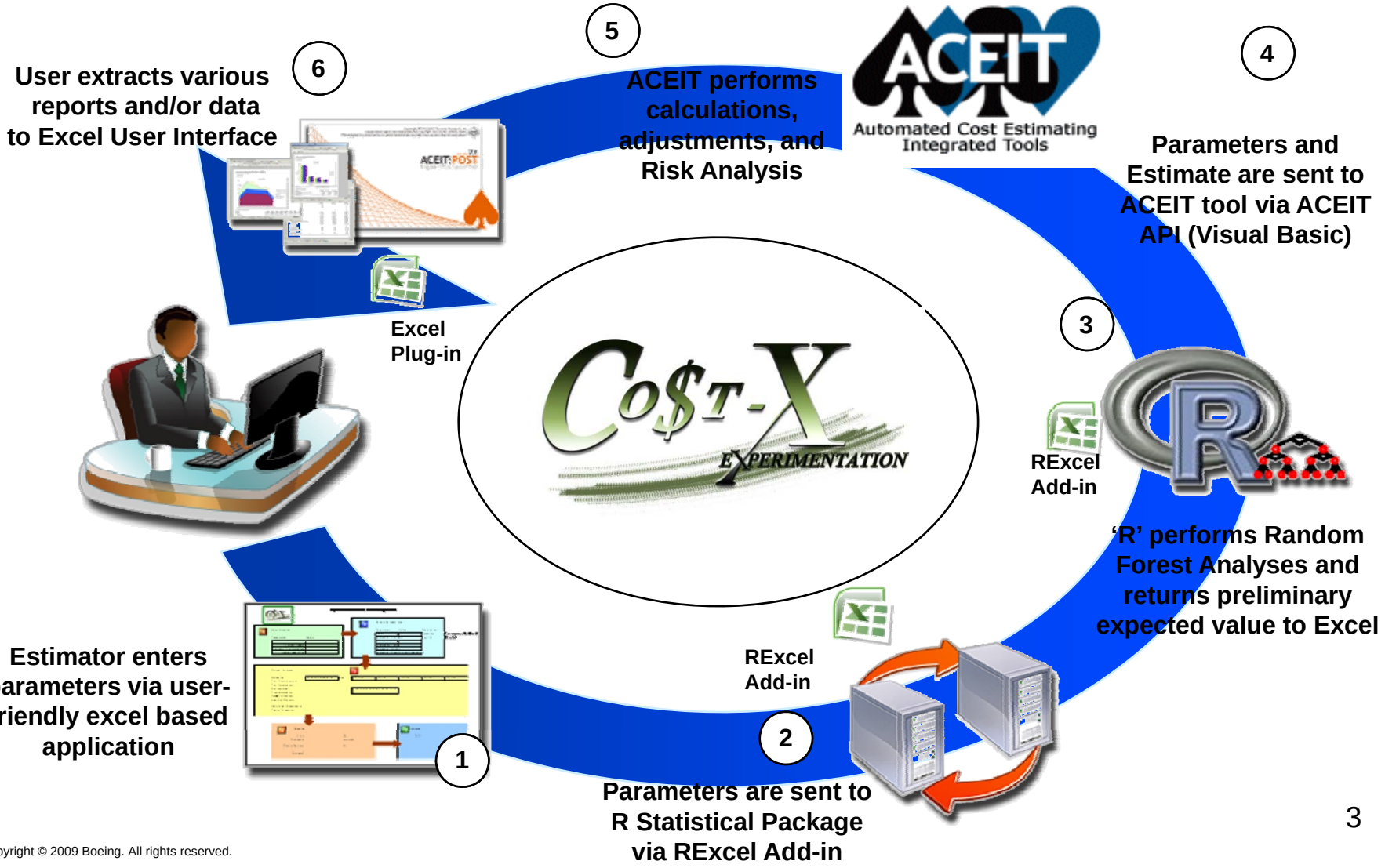
Denise Nelson BR&T – Affordability

January 2011

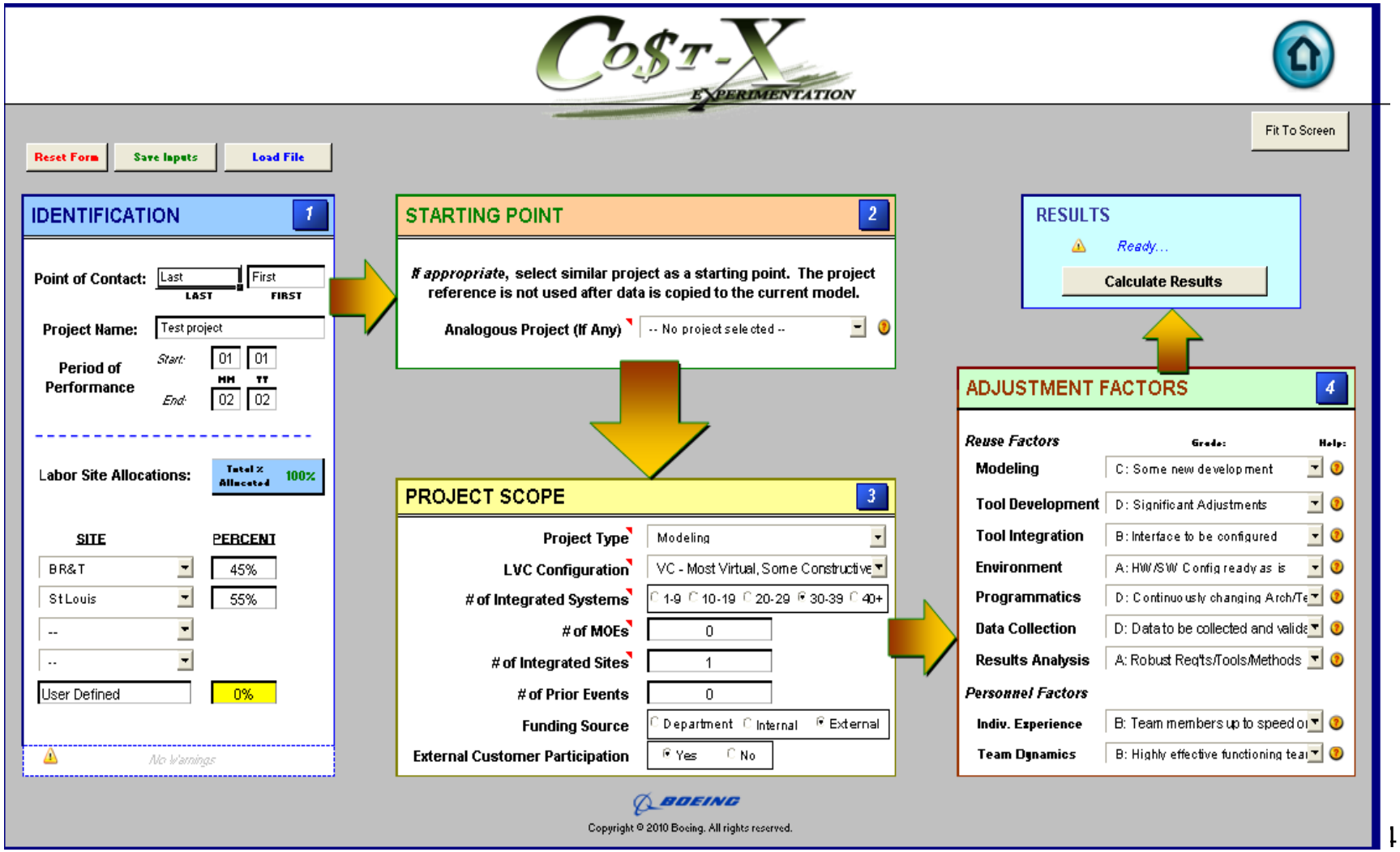
Overview of Project

- Objective
  - Estimate the costs of simulation-based experimentation
  - Develop a standard process and toolkit
  - Be accurate, consistent, easy-to-use, and value-added
- Customer Constraints
  - Use standard cost estimating tools
  - Incorporate standard processes
  - Base on historical data
- Our recommendations
  - Develop toolkit using ACEIT suite of tools as backbone
  - Incorporate other analytical tools as needed

Co$t-X Process



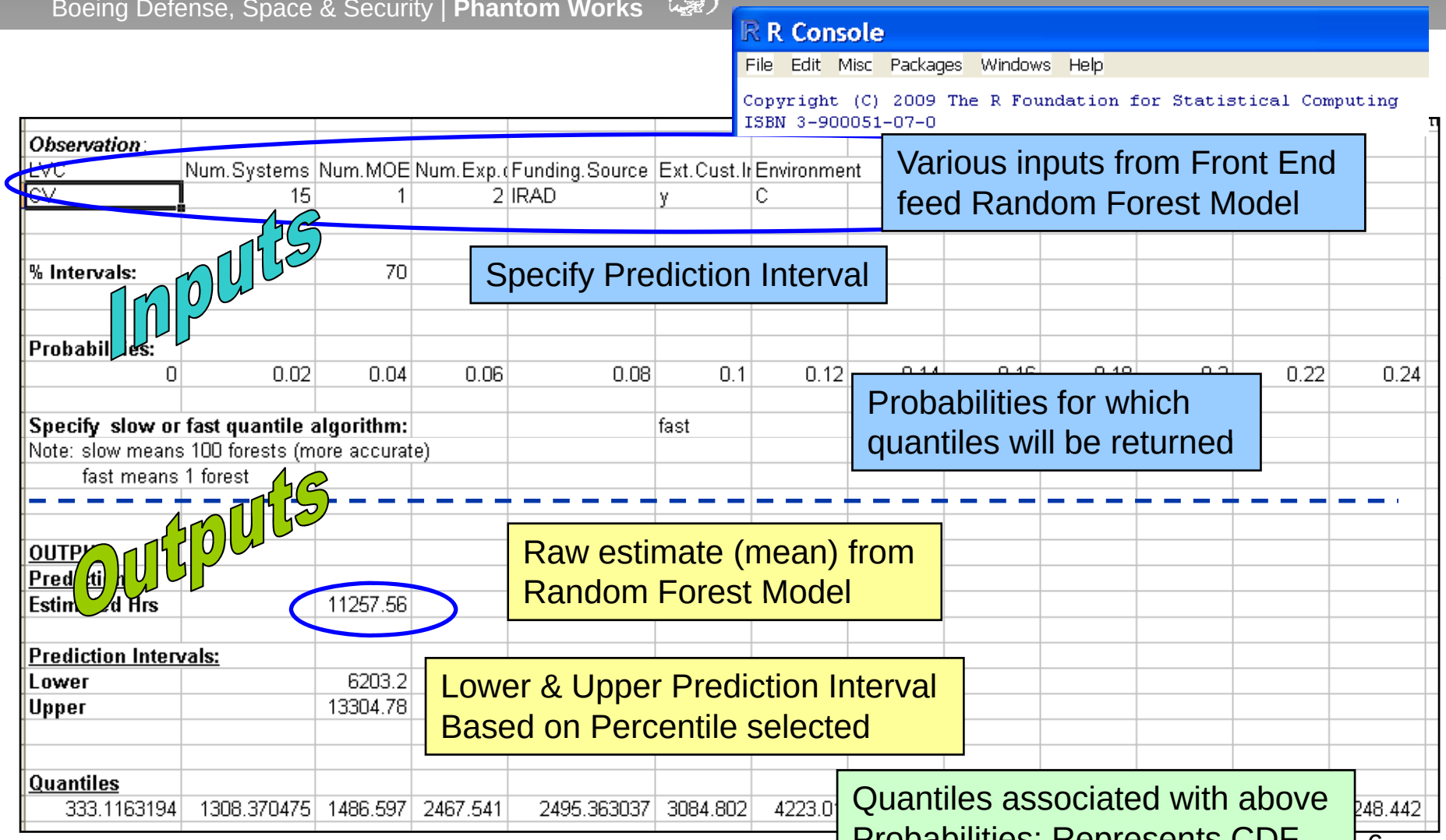
Co$t-X User Interface



When “Calculate Results” button is selected

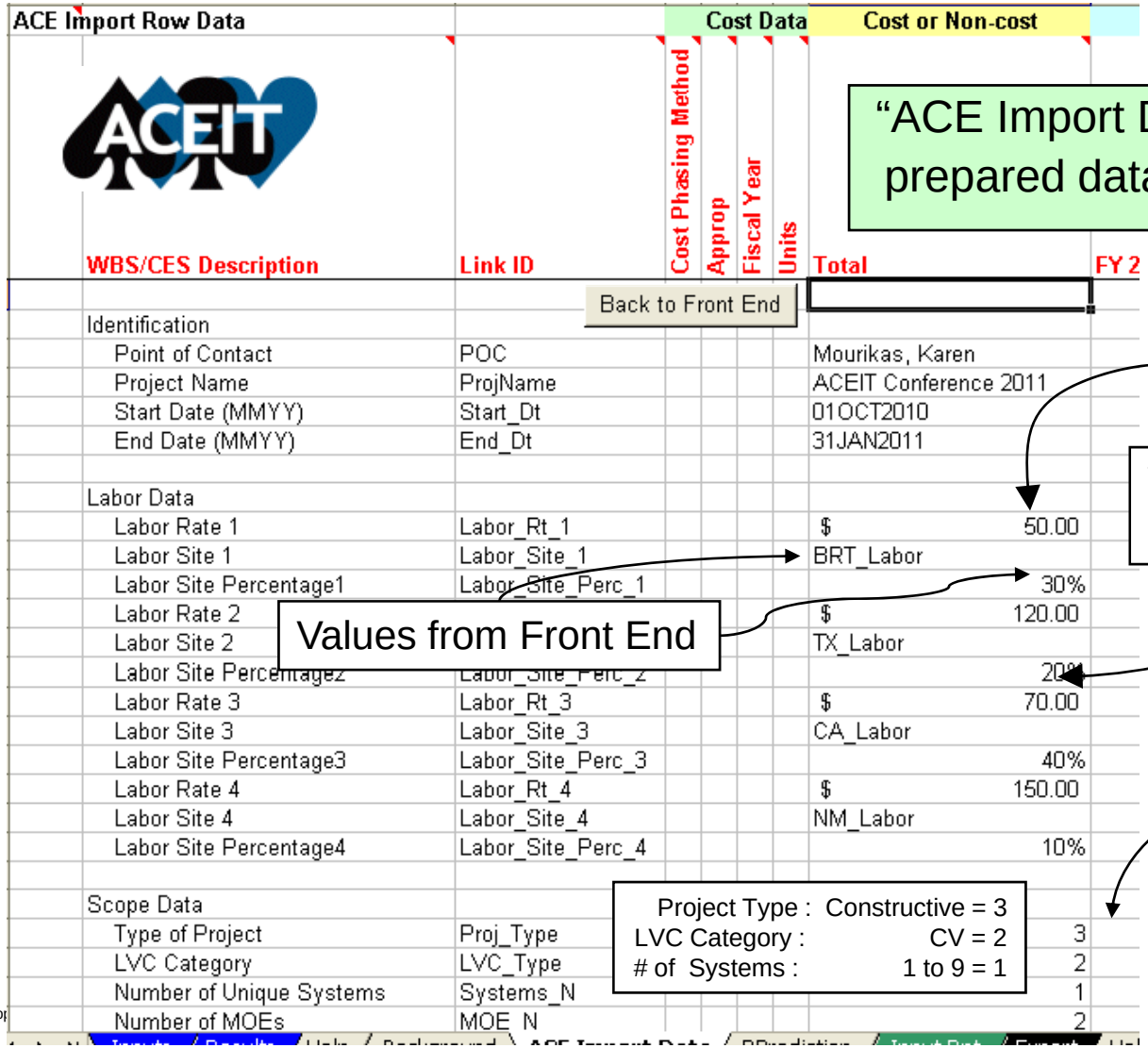
- Random Forest model is executed via call to R statistical package
  - Returns preliminary costs estimate, prediction intervals, and quantiles
- Visual Basic macros transfers data to/from ACE model
  - Formats the data for ACE (“Create_ACE_Import_Data”)
  - Opens the ACE application
  - Loads the indicated ACE model
  - Transfers required data to the ACE model (“Export_to_ACE AceSession”)
- ACE model is executed (via Visual Basic macro)
  - ACE performs additional calculations and modifications
  - Cost Risk/Uncertainty Analysis is performed
  - Results passed back into Excel (“Get_ACE_results”)
- Summary page populated with data
  - Options to print / view additional reports

R Excel Interface
Call to R Random Forest Model



Observation:											
LVC	Num.Systems	Num.MOE	Num.Exp.(	Funding.Source	Ext.Cust.Ir	Environment					
CV	15	1	2	IRAD	y	C					
% Intervals:		70									
Probabilities:											
0	0.02	0.04	0.06	0.08	0.1	0.12	0.14	0.16	0.18	0.2	0.22	0.24
Specify slow or fast quantile algorithm:					fast						
Note: slow means 100 forests (more accurate)											
fast means 1 forest											
OUTPUTS:											
Predicted Hrs		11257.56									
Prediction Intervals:											
Lower		6203.2									
Upper		13304.78									
Quantiles											
333.1163194	1308.370475	1486.597	2467.541	2495.363037	3084.802	4223.0					248.442

ACE Import Data Sheet used to transfer data via macros (1 of 2)



ACE Import Row Data		Cost Data				Cost or Non-cost	
WBS/CES Description	Link ID	Cost Phasing Method	Approp	Fiscal Year	Units	Total	FY 2
		Back to Front End					
Identification							
Point of Contact	POC					Mourikas, Karen	
Project Name	ProjName					ACEIT Conference 2011	
Start Date (MMYY)	Start_Dt					01OCT2010	
End Date (MMYY)	End_Dt					31JAN2011	
Labor Data							
Labor Rate 1	Labor_Rt_1					$ 50.00	
Labor Site 1	Labor_Site_1					BRT_Labor	
Labor Site Percentage1	Labor_Site_Perc_1					30%	
Labor Rate 2						$ 120.00	
Labor Site 2						TX_Labor	
Labor Site Percentage2	Labor_Site_Perc_2					20%	
Labor Rate 3	Labor_Rt_3					$ 70.00	
Labor Site 3	Labor_Site_3					CA_Labor	
Labor Site Percentage3	Labor_Site_Perc_3					40%	
Labor Rate 4	Labor_Rt_4					$ 150.00	
Labor Site 4	Labor_Site_4					NM_Labor	
Labor Site Percentage4	Labor_Site_Perc_4					10%	
Scope Data							
Type of Project	Proj_Type						3
LVC Category	LVC_Type						2
Number of Unique Systems	Systems_N						1
Number of MOEs	MOE_N						2

"ACE Import Data" sheet contains prepared data from Excel Front End

Values from hidden linked excel pages

Values from Front End

Project Type : Constructive = 3
LVC Category : CV = 2
# of Systems : 1 to 9 = 1

Translated values From Front End

ACE Import Data Sheet (2 of 2)

WBS/CES Description	Link ID		FY
Number of Sites During Execution	Sites_N	1	
Number of Related Past Experimer	Rel_Exp_N	2	
Funding Source	Cust_Fund_I	2	
External Customer Participation	Cust_Inv_I	1	
ReUse Factors			
Modeling	Model_ReF	D	
Tool Development	Tool_ReF	C	
Tool Integration	Int_ReF	D	
Environment	Env_ReF	D	
Programmatics	PM_ReF	C	
Data Collection	Data_ReF	C	
Analysis Results	Analysis_ReF	C	
Individual Experience	Exp_ReF	C	
Team Dynamics	Team_ReF	C	
R Results			
RF Estimated Hours	RF_Hrs	11257.56316	
RF Hrs Prediction Interval Low	RF_Hrs_PIL	6203.199988	
RF Hrs Prediction Interval High	RF_Hrs_PIH	13304.77674	

Scope Parameters from Front End

Translated values

Funding Source : Customer = 2
Customer Partifcipation : Yes = 1

ReUse & Learning Parameters from Front End

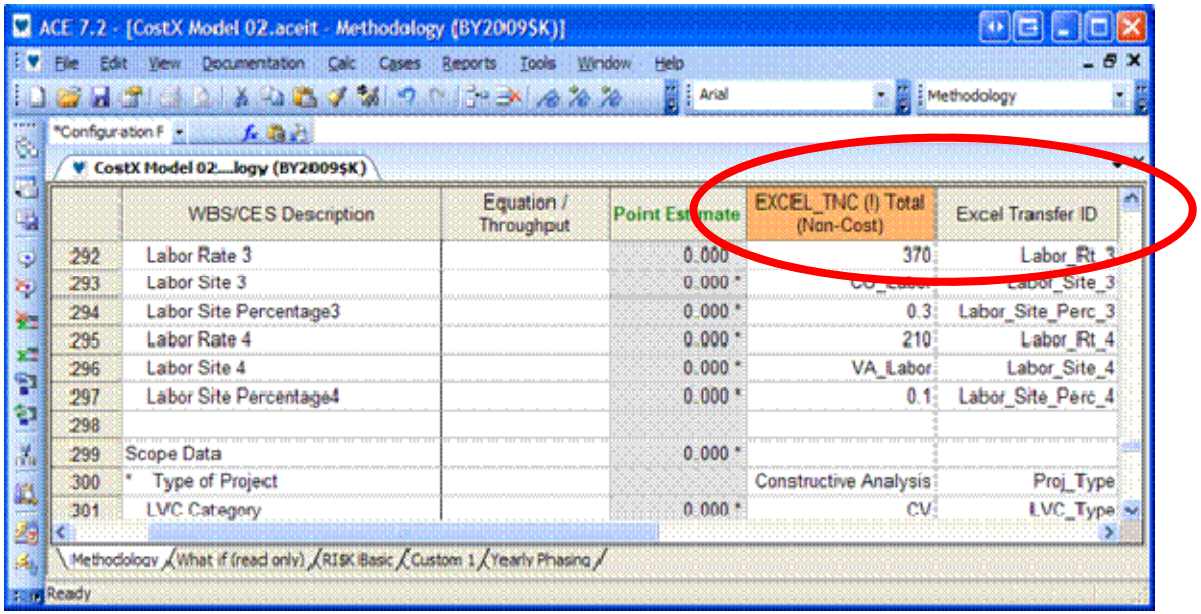
Random Forest values

RF Estimate

Prediction Interval based on selected percentile (70%)

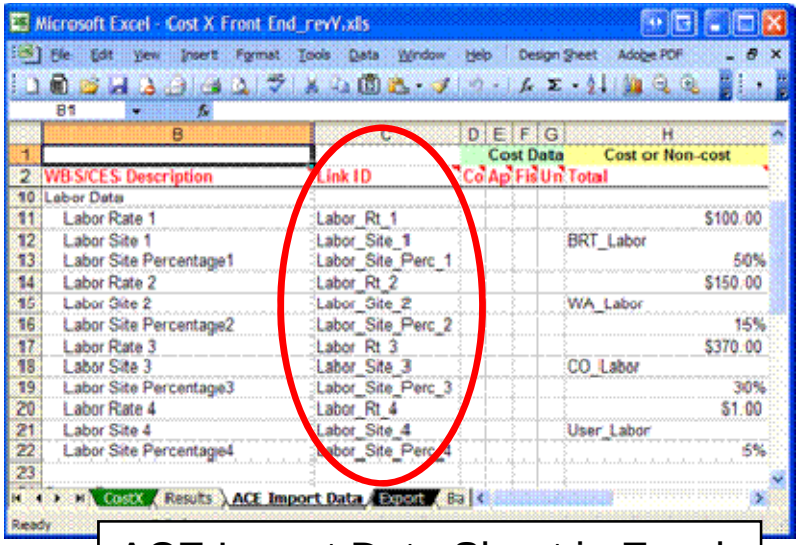
ACE Model accepts imported data from Excel

- ACE model updated to accept input from Excel via two columns of data
  - Excel Transfer ID Column
    - Contains ID strings to identify data in Excel model
  - EXCEL_TNC (!) Total (Non-Cost) Column
    - Receives values generated in Excel and makes available to the ACE model

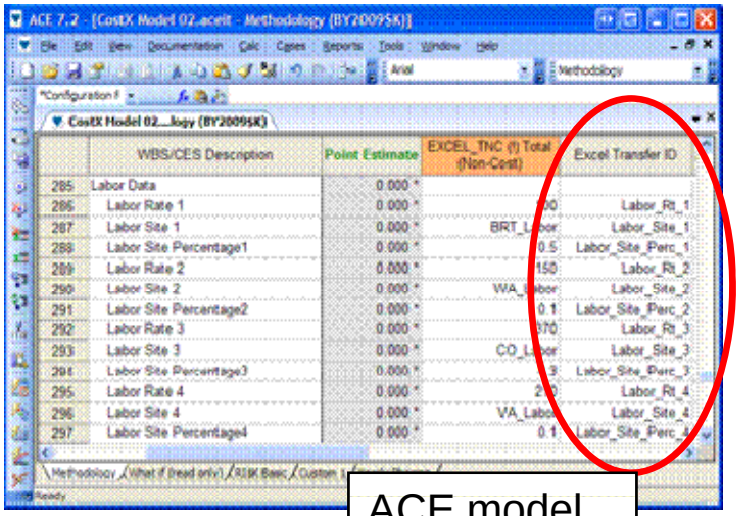


ACE Excel linkages

- Excel's "ACE Import Data" worksheet talks to ACE model
  - Excel!'ACE Import Data'!'Link ID' column must be identical to the "Excel Transfer ID" column in the ACE model



ACE Import Data Sheet in Excel



ACE model

Processing within ACE model

- Data input to ACE model
  - Random Forest estimate and prediction data
  - Probabilities and Quantile data (CDF) translated as multipliers
  - All Input parameters, excluding identification text fields
    - Categorical (text) data transferred into numerical values
- Parametric adjustments to raw estimate
  - Random Forest model calibration
  - Regression adjustments based on additional parameters
  - Final estimate
- Risk calculations
  - Custom Cumulative Distribution Function (CDF)
  - 1000 iterations
  - Risk defined for Random Forest estimate
    - Applies to final estimate

Data transferred into ACE model

CostX Model 0...y (BY2010$K)

	WBS/CES Description	Unique ID	EXCEL_TNC (I) Total (Non-Cost)	Excel Transfer ID
445	Identification			
448	Start Date (M			
449	End Date (MMYY)			
450				
451	Labor Data			
452	Labor Rate 1			
453	Labor Site 1			
454	Labor Site P			
455	Labor Rate 2			
456	Labor Site 2	XL		
457	Labor Site Percentage2	Labo		
458	Labor Rate 3			
459	Labor Site 3	XL		
460	Labor Site Percentage3	Labo		
461	Labor Rate 4			
462	Labor Site 4	XL		
463	Labor Site Percentage4	Labo		
464				
465	Scope Data			
466	Type of Projec			
467	LVC Category			
468	Number of Unique Systems			
469	Number of MOEs	XL_MOE_N	2	MOE_N
470	Number of Sites During Execu	XL_Sites_N	1	Sites_N
471	Number of Prior Events	XL_Rel_Exp_N	2	Rel_Exp_N

ID data

Labor data

Scope da

	WBS/CES Description	Unique ID	EXCEL_TNC (I) Total	Excel Transfer ID	Approp
475	Project Starting Poin			Ref_Proj	
476					
477	ReUse Factors				
478	Modeling	XL_Model_ReF	B	Model_ReF	
479	Tool Developme		C	Tool_ReF	
480	Tool Integration		B	Int_ReF	
481	Environment	XL_Env_ReF	B	Env_ReF	
482	Programmatics	XL_PM_ReF	B	PM_ReF	
483	Data Collection	XL_Data_ReF	B	Data_ReF	
484	Analysis Results	XL_Analysis_ReF	B	Analysis_ReF	
485	* Team Factors				
486	Individual Experience	XL_Exp_ReF	C	Exp_ReF	
487	Team Dynamics	XL_Team_ReF	B	Team_ReF	
488					
489	* Results from Random Fores				
490	RF Estimated Hours	XL_RF_Hrs	5228.09051707939	RF_Hrs	
491	RF Hrs Prediction Interval Low	XL_RF_Hrs_PIL	1197.63633885797	RF_Hrs_PIL	
492	RF Hrs Prediction Interval High	XL_RF_Hrs_PIH	10268.9914094023	RF_Hrs_PIH	
493					

Unique IDs start with “XL”
to identify as passed in via Excel

Reuse / Learning

RF Estimate/Prediction data

Estimate manipulations

WBS/CES Description	Unique ID	Equation / Throughput	Point Estimate
* RF Calculate / Pre Adjusted			
*Project Estimate in $			
Proj Est $ Total	Proj_Total_$		916,360. (53%) *
Proj Est $ Site 1	Site1_$	Hrs_Site1*Labor_Rate_1	365,326.20 (53%) *
Proj Est $ Site 2	Site2_$	Hrs_Site2*Labor_Rate_2	140,042. (53%) *
Proj Est $ Site 3	Site3_$	Hrs_Site3*Labor_Rate_3	340,971. (53%) *
Proj Est $ Site 4	Site4_$	Hrs_Site4*Labor_Rate_4	70,021. (53%) *
*Project Estimate in Hrs: Calcula			
Proj Est Hrs Total	Proj_Total_Hrs	XL_RF_Hrs.EXCEL_TNC	6,089. (53%) *
Site Point Estimate in Hrs	Site_Total_Hrs		
Proj Est Hr Site 1	Hrs_Site1	Proj_Total_Hrs*Labor_Percent_1	1,827. (53%) *
Proj Est Hr Site 2	Hrs_Site2	Proj_Total_Hrs*Labor_Percent_2	1,218. (53%) *
Proj Est Hr Site 3	Hrs_Site3	Proj_Total_Hrs*Labor_Percent_3	2,436. (53%) *
Proj Est Hr Site 4	Hrs_Site4	Proj_Total_Hrs*Labor_Percent_4	609. (53%) *

Translated to $ per site

Random Forest Estimate – in hours

Hours alllocated to sites

- Raw estimate from Random Forest (in hours) to be adjusted in ACE model incorporating additional parameters from CostX Front End

Model Calibration & Adjustments

- Random Forest model calibrated based on residual values
- Additional parametric equations applied to Random Forest raw estimate
  - Utilizing additional parameters from front end
  - Adjustments determined via regression and analysis

WBS/CES Description	Unique ID	Equation / Throughput	Point Estimate
***Ccalibration Adjustment**			
*Generally, overestimates low cost, underestimat			
*Linear I			
*-0.3342			
*Do not			
Regress			
Regress			
Calibrati			

WBS/CES Description	Unique ID	Equation / Throughput	Point Estimate
***Num Site Adjustment**			
*Adjust if more than 1 site involved			
* Based on Linear Regression of Hrs (Y) per Num			
Num Site Regression Coefficient A	Site_Regr_Coeff_A	2143.6	2,144. *
Num Site R			
Num Site =			
Num Sites A			
Num Site A			
Num Site Max Adjustment	Site_Adj_Max	Site_Adj_Amt+CommEffects_Adj_Est	4,403. (61%) *
Num Site Adjusted Estimate	Site_Adj_Est	If (Site_Adj_Amt- CommEffects_Adj_Est > 0, CommEffects_Adj_Est+Site_Adj_Amt/2,	4,403. (61%) *

WBS/CES Description	Unique ID	Equation / Throughput	Point Estimate
* Enhancement of Previous Effort			
Upgrade of Previous effort	Upg_I	XL_SP_Upg.EXCEL_TNC	0. *
Upgrade Reduction %	Upg_Pct	If(Cont_NoDB_I = 1, 0.1, -.25)	-0.25 *
Upgrade Adj Estimate	Upg_Adj_Est	BU_Supt_Adj_Est * (1+Upg_Pct*Upg_I)	3,962. (61%) *

Final Adjusted Point Estimate

WBS/CES Description	Unique ID	Equation / Throughput	Point Estimate
*Final Adjusted Total Hours			
Final Point Estimate Hours	Final_PE_Hrs	AOA_Adj_Est	3,425. (53%)
PE Site Estimates Hrs			3,425. (53%)
Proj Est Hr Site 1	Adj_Hrs_Site1	Final_PE_Hrs*Labor_Percent_1	1,027. (53%)
Proj Est Hr Site 2	Adj_Hrs_Site2	Final_PE_Hrs*Labor_Percent_2	685. (53%
Proj Est Hr Site 3	Adj_Hrs_Site3	Final_PE_Hrs*Labor_Percent_3	1,370. (53%
Proj Est Hr Site 4	Adj_Hrs_Site4	Final_PE_Hrs*Labor_Percent_4	342. (53%)
* Final Adjusted Total $			
Proj Est $ Total	Final_PE_$		515,452. (53%)
Proj Est $ Site 1	Adj_Site1_$	Adj_Hrs_Site1*Labor_Rate_1	205,495.99 (53%)
Proj Est $ Site 2	Adj_Site2_$	Adj_Hrs_Site2*Labor_Rate_2	78,773. (53%)
Proj Est $ Site 3	Adj_Site3_$	Adj_Hrs_Site3*Labor_Rate_3	191,796. (53%)
Proj Est $ Site 4	Adj_Site4_$	Adj_Hrs_Site4*Labor_Rate_4	39,387. (53%)

Hours allocated to sites

Translated to $ per site

- Final Point Estimate derived from raw estimate from Random Forest

Custom Cumulative Distribution Function from Random Forest (CDF)

- Random Forest returns Estimated Hrs
  - Median value
  - Average prediction from collection of trees
  - As a “Raw Estimate”

OUTPUT	
Prediction:	
Estimated Hrs	11257.56

- Quantiles = estimates associated with the Probability Distribution
  - Output from R Random Forest model
- Multiplier used in ACE Custom CDF definition
  - Quantile / Estimated Hrs
  - Multiplier of 1 refers to the Raw estimate

Prob	Quantiles	Multiplier
0	333.1163194	0.02959
0.02	1308.370475	0.116221
0.04	1486.59731	0.132053
[illegible]	[illegible]62	0.21919
[illegible]	[illegible]	[illegible]
0.1	3084.80245[illegible]	[illegible]
0.12	4223.00999[illegible]	[illegible]
0.14	5784.56351[illegible]	[illegible]
0.16	6285.5	0.558336
0.18	8399.198341	0.746094
0.2	9248.442368	0.821531
0.22	9248.442368	0.821531
0.24	9248.442368	0.821531

Output from RF

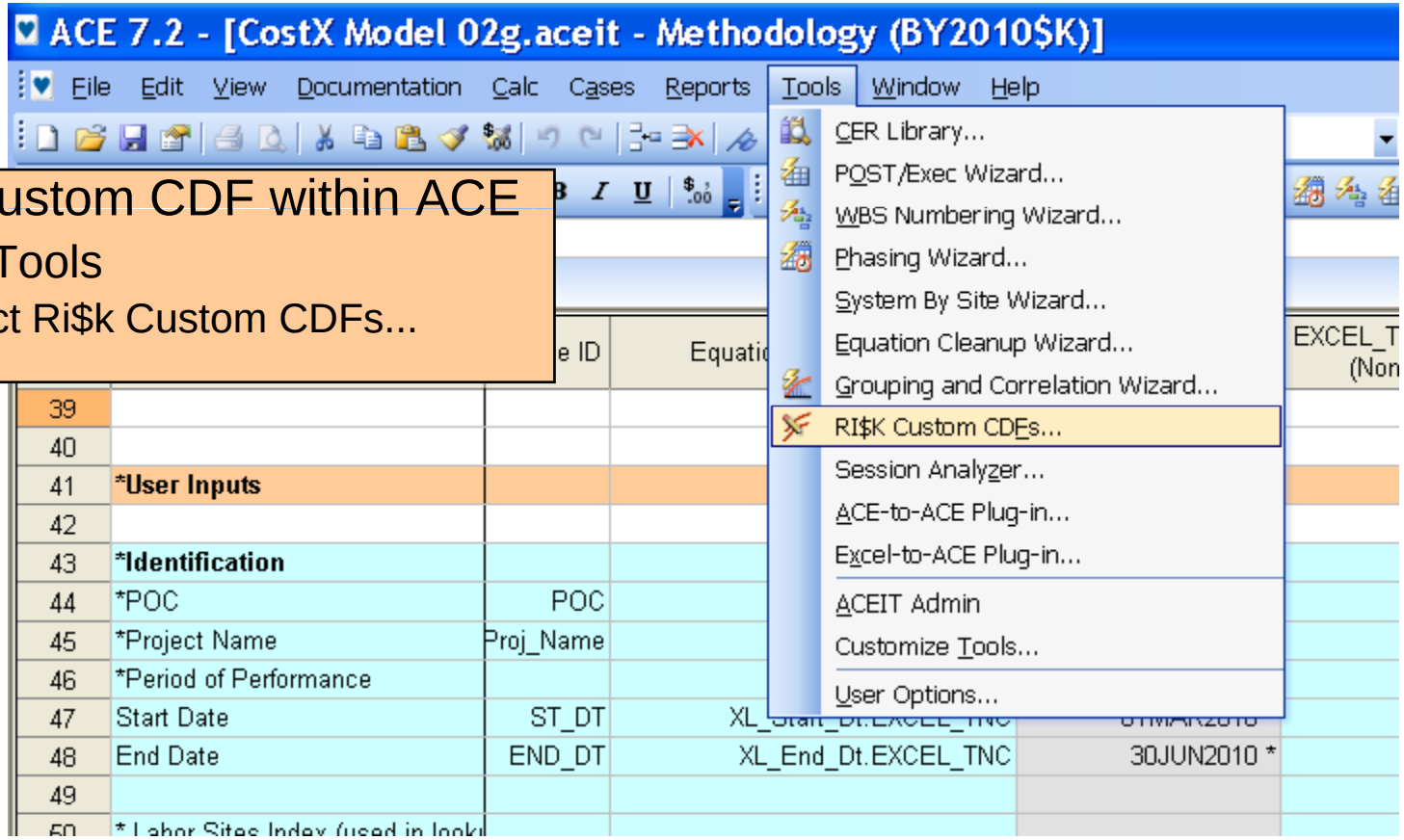
Input to ACE

- RQuantiles = name of CDF for ACE model

ACE RPred CDF name	RQuantiles

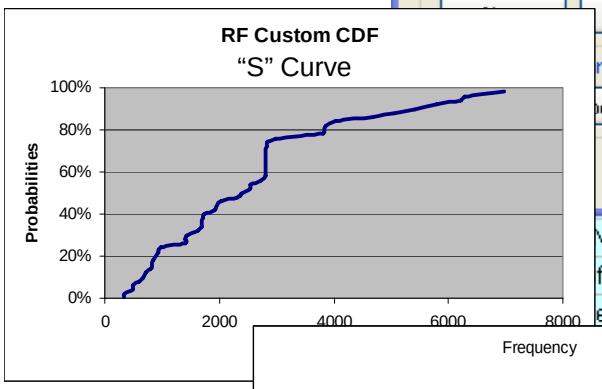
Custom Cumulative Distribution Function in ACE

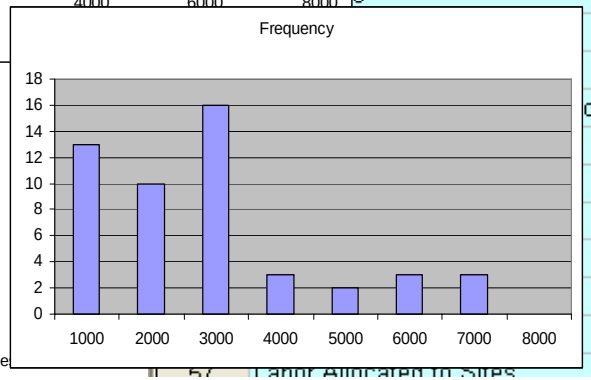
- To view Custom CDF within ACE
  - Select Tools
    - Select Ri$k Custom CDFs...



ACE Custom Cumulative Distribution Function

Name of CDF from Excel





Probability Confidence Levels and Multipliers passed in from Excel

Cost Risk Applied to Point Estimate

WBS/CES Description	Unique ID	Point Estimate	Distribution Form	PE Position in Distribution	CDF Keyword	Random Seed
*Project Estimate in Hrs: Calculated from RF						
Proj Est Hrs Total	Proj_Total_Hrs	2,649. (55%) *	CDF	Undefined	RQuantiles	1687442

CDF applied to Raw estimate

WBS/CES Description	Unique ID	Equation / Throughput	Point Estimate
*Final Adjusted Total Hours			
Final Point Estimate Hours	Final_PE_Hrs	AOA_Adj_Est	1,950. (55%) *
PE Site Estimates Hrs			1,950. (55%) *
Proj Est Hr Site 1	Adj_Hrs_Site1	Final_PE_Hrs*Labor_Percent_1	97. (55%) *
Proj Est Hr Site 2	Adj_Hrs_Site2	Final_PE_Hrs*Labor_Percent_2	487. (55%) *
Proj Est Hr Site 3	Adj_Hrs_Site3	Final_PE_Hrs*Labor_Percent_3	1,267. (55%) *
Proj Est Hr Site 4	Adj_Hrs_Site4	Final_PE_Hrs*Labor_Percent_4	97. (55%) *

Final Estimate

Final Estimate allocated to sites

Specify Cost Risk Percentile

* Risk Calculations			
* Risk Quantiles			
Low Risk Percentile	RiskConf_Low	50. *	50
Middle Risk Percentile	RiskConf_Mid	70. *	70
High Risk Percentile	RiskConf_High	85. *	85

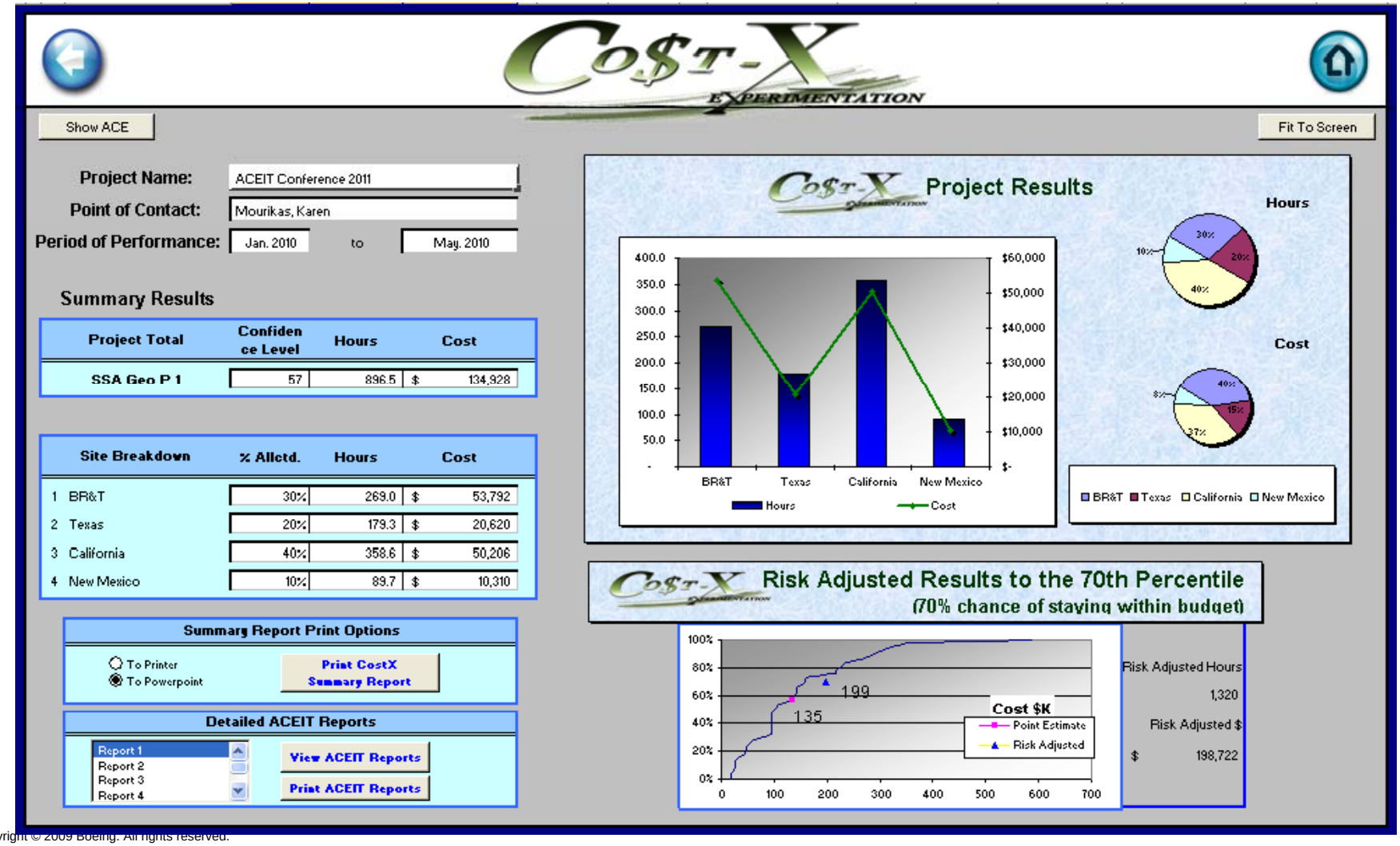
Three levels of risk for flexibility

Can be easily changed

WBS/CES Description	Unique ID	Point Estimate
*Final Adjusted Total Hours		
Final Point Estimate Hours	Final_PE_Hrs	1,950 (55%) *

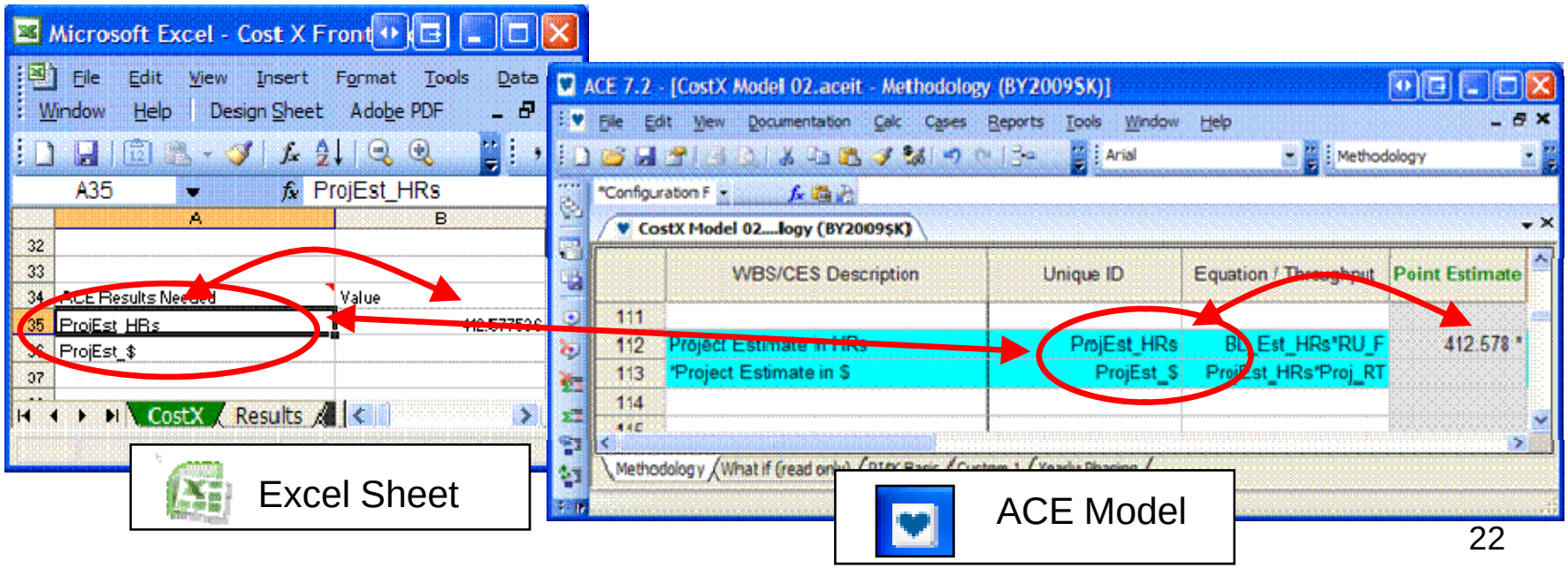
RiskConf_PE		RiskConf(@Final_PE_Hrs, Final_PE_Hrs,1)		55. *
***Final Adjusted RA Hours**				
Final RA Low Hours	Final_RA_Low_Hrs	RiskCost(@Final_PE_Hrs,RiskConf_Low)	50%	1,853. *
Final RA MidHours	Final_RA_Mid_Hrs	RiskCost(@Final_PE_Hrs,RiskConf_Mid)	70%	2,409. *
Final RA HighHours	Final_RA_Hi_Hrs	RiskCost(@Final_PE_Hrs,RiskConf_High)	85%	3,285. *

Co$t-X Summary Results



Results returned from ACE to Excel

- Values from ACE placed in corresponding columns in Excel
  - Excel's column A contains the "Unique ID" names from ACE
    - i.e. "ProjEst_HRs"
  - The "Point Estimate" column in ACE contains the value associated with the Unique ID and passed to column B in excel



Labor Rates Lookup Tables

- Look up tables contain labor rates per year
  - MatVal(@Hrly_RatesM, Labor_Site_1, Labor_Rate_Col)
  - Project Start Year determines which labor rate to use

WBS/CES Description	Unique ID	FY 2009	FY 2010	FY 2011	FY 2012
***Rates Per Site Inflated by 4% from 2009**		2009	2010	2011	
Hourly_Rates Matrix	**Hrly_RatesM**				
Alabama		100.0	104.0	108.2	
Arizona		110.0	114.4	119.0	
BR&T		120.0	124.8	129.8	
California					
Colorado					
Kansas					
New Mexico					
Philly					
St Louis					
Texas					
Virginia					
Washington					
UserDefined		220.0	228.8	238.0	

DATEYR(ST_DT)-2008

Row 47 {01JAN2010}
Start Date

stX Model 02...ogy (BY2010$K)

BY2010$K)

WBS/CES Description	Unique ID	Equation / Throughput	Point Estimate
First Year of Rates in DB	FIRST_YR	2008	2,008. *
Project Start Year Column in DB	Start_Yr_Col	DATEYR(ST_DT)-2008	2. *
Labor Rate Column in DB	Labor_Rate_Col	If(Start_Yr_Col > 3, 3, Start_Yr_Col)	2. *

ReUse Lookup Tables

WBS/CES Description	Unique ID	FY 2009	FY 2010	FY 2011	FY 2012	FY 2013
*** ReUse Factor Matrix Headers**		**A**	**B**	**C**	**D**	**E**
Reuse Factors Matrix	**FactorM**					
Modeling		0.25	0.5	1		
Tool Development		0.8	1	1.5		
Tool Integration		0.5	0.8	1		
Environm		0.3	0.7	1		
Programm		0.5	0.7	1		
Data Collection		0.6	0.8	1		
Analysis Results		0.75	0.85	1		
Team Experience (individual)		0.8	0.9	1	1.25	1.4
Team Dynamics		0.7	0.85	1	1.25	1.4

ReUse Factor Table

- Reuse data determines cost savings/penalties
  - → A ~ Cost Savings
  - → E ~ Cost Penalties

WBS/CES Description	EXCEL_TNC (Total
ReUse Factors	
Modeling	D
Tool D	B
T gration	C
Environment	B
Programmatics	C
Data Collection	C
Analysis Results	C

ReUse Factor Values

WBS/CES Description	Unique ID	Equation / Throughput	Point Estimate
*Matrix Lookups and Functions			
*Factor Lookup			
Modeling	Mod_F	MatVal(@FactorM, 1, Mod_RU)	1.25 *
Tool Development	Tool_F	MatVal(@FactorM, 2, Tool_RU)	1.00 *
Tool Integration	Int_F	MatVal(@FactorM, 3, Int_RU)	1.00 *
Environment	Env_F	MatVal(@FactorM, 4, Env_RU)	0.70 *
Programmatics	Prog_F	MatVal(@FactorM, 5, Prog_RU)	1.00 *
Data Collection	Data_F	MatVal(@FactorM, 6, Data_RU)	1.00 *
Analysis Results	Ana_F	MatVal(@FactorM, 7, Ana_RU)	1.00 *

Matrix Lookup

Next Steps for 2011

- Data Collection/Analysis
  - Collect new data (on-going)
  - Automatic database update (in-work)
  - Analyze additional data (on-going)
- Model
  - Investigate ACEIT 7.2 features
    - Cat(), RLookup(), Random Number generator from ACEIT
  - Create additional reports (in-work)
    - Cost Risk Assessment Report
    - Detailed Baseline Report