Improving the NICM Software CER

Jairus Hihn
Leora Juster
Mike Dinicola
Jet Propulsion Laboratory, California Institute of Technology

2015 NASA Cost Symposium
August 26, 2015

Purpose

- Show the new CER and explain why its better
- Continue to emphasize the use of better methods for evaluating models
  - Use of MRE as a statistic
  - Leave one out cross-validation
- Make you aware that we will be knocking on some doors to get better technical data
  - There either is (1) none, (2) it makes no sense, (3) or the little that exists does not produce a usable regression

Background

- There is a fundamental issue with the V6 NICM SW CER
  - Current model is that Software Cost is 12% of Instrument sensor cost if instrument has “Intensive Software Development” and 4% otherwise
  - At the time this was the only CER that could be fit to the data as everything else failed so it was the best that could be done.
  - Issues are that
    - This input is totally subjective and not tied to any objective information even notionally.
      - Intention was to eventually fix this
    - Results are hard wired to the current set
- The NICM software CER has not been updated in over six years while the available data has doubled, it is clearly time for an updated CER and one that **uses only objective inputs**
- This work is a joint activity by the NASA Software Cost Modeling Task and the NICM team

Topics

- Data Summary
- Methodology
  - Magnitude of Relative Error (MRE) statistics and bootstrap cross-validation
  - Demonstrate MRE is an important supplemental metric to the traditional regression statistics
- Overview of Final Results
- NICM SW CER
- Future directions

Summary of Existing NICM SW Cost data

NICM Software Cost Data Summary							
Instrument Type	**Number of Records**	**Avg. SW Cost ($K FY2015)**	**Median SW Cost ($K FY2015)**	**Range SW Cost ($K FY2015)**	**Avg. SW Cost/ Sensor Cost**	**Median SW Cost/ Sensor Cost**	**Range SW Cost/ Sensor Cost**
Fields	**3**	**425**	**372**	**102 - 801**	**0.09**	**0.11**	**0.03 - 0.14**
Particles	**16**	**1,458**	**877**	**399 - 6802**	**0.09**	**0.08**	**0.03 - 0.20**
Optical	**37**	**2,376**	**1,329**	**83 - 10284**	**0.06**	**0.05**	**0.01 - 0.17**
In Situ	**10**	**930**	**594**	**90 - 4397**	**0.06**	**0.05**	**0.02 - 0.17**
Passive Microwave	**4**	**3,319**	**2,907**	**601 - 6859**	**0.07**	**0.03**	**0.02 - 0.21**

- Summary shown includes all available data
- No FSW data for Active Microwave Instruments

Ranges are very large

Approach and Methodology

- Evaluated various options
  - *Ln linear vs linear*
  - *With and without older records*
  - *Single Software CER versus CERs by Instrument type*
    - *Optical vs optical and passive microwave*
      - *There are only 4 microwave records*
  - *Independent Variables considered*
    - *Sensor cost (total and non-SW subsystem total)*
    - *Sensor mass, sensor data rate, others*
- *Evaluated based on*
  - *F-test, t-test,* R^2
  - *MRE stats*
    - *Median, Mean, Inter-quartile range, MIN-MAX AND MRE pareto curve*
  - *Do parameters make sense*
- *Some data points removed from analysis*
  - *AIRS and MLS were removed as outliers*
  - Oldest data points – Pre-1990

General Instrument Software CER

Generic vs LN Generic CER

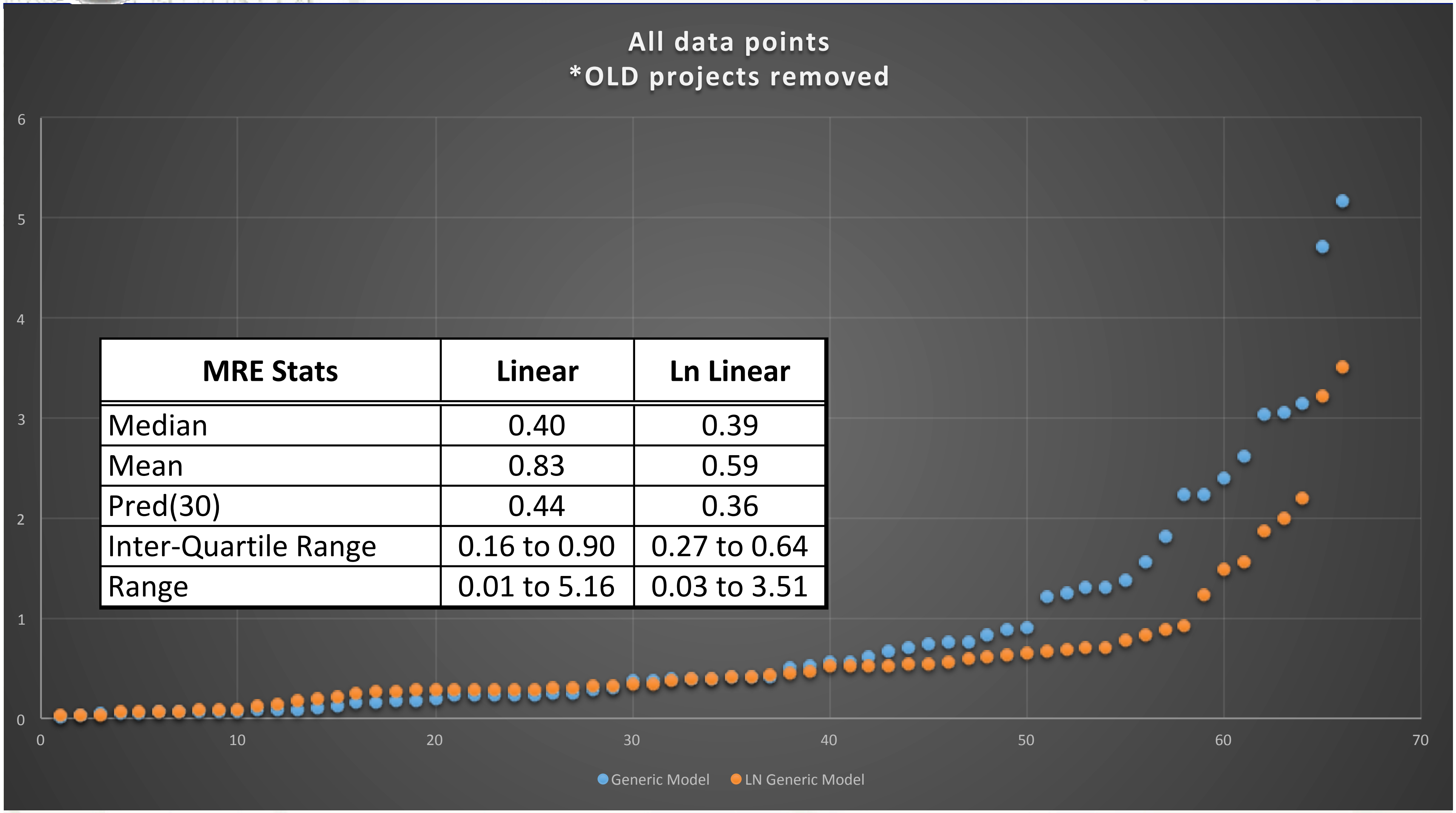
SW_Cost = 185 + 0.06*Sensor_Cost

SW_Cost = $0.18*Sensor\_Cost^{0.88}$

Basic Stats	Generic	LN Generic
Sample Size:	64	64
R-Squared:	0.8	0.7
F-test:	254.7	153
Slope (significance):	< 0.0001	< 0.0001
Intercept (significance):	< 0.275	< 0.018

- MRE results on next page
- Results are mixed between the two functional forms
  - *Ln-Linear is less effected by extreme cases*
  - *Linear model is easier to interpret*

Generic Instrument Software CER 2



MRE Stats	Linear	Ln Linear
Median	0.40	0.39
Mean	0.83	0.59
Pred(30)	0.44	0.36
Inter-Quartile Range	0.16 to 0.90	0.27 to 0.64
Range	0.01 to 5.16	0.03 to 3.51

Particles and Fields Software CERs

Particles and Fields SW CER

SW_Cost = 170+ 0.01*Sensor_Cost

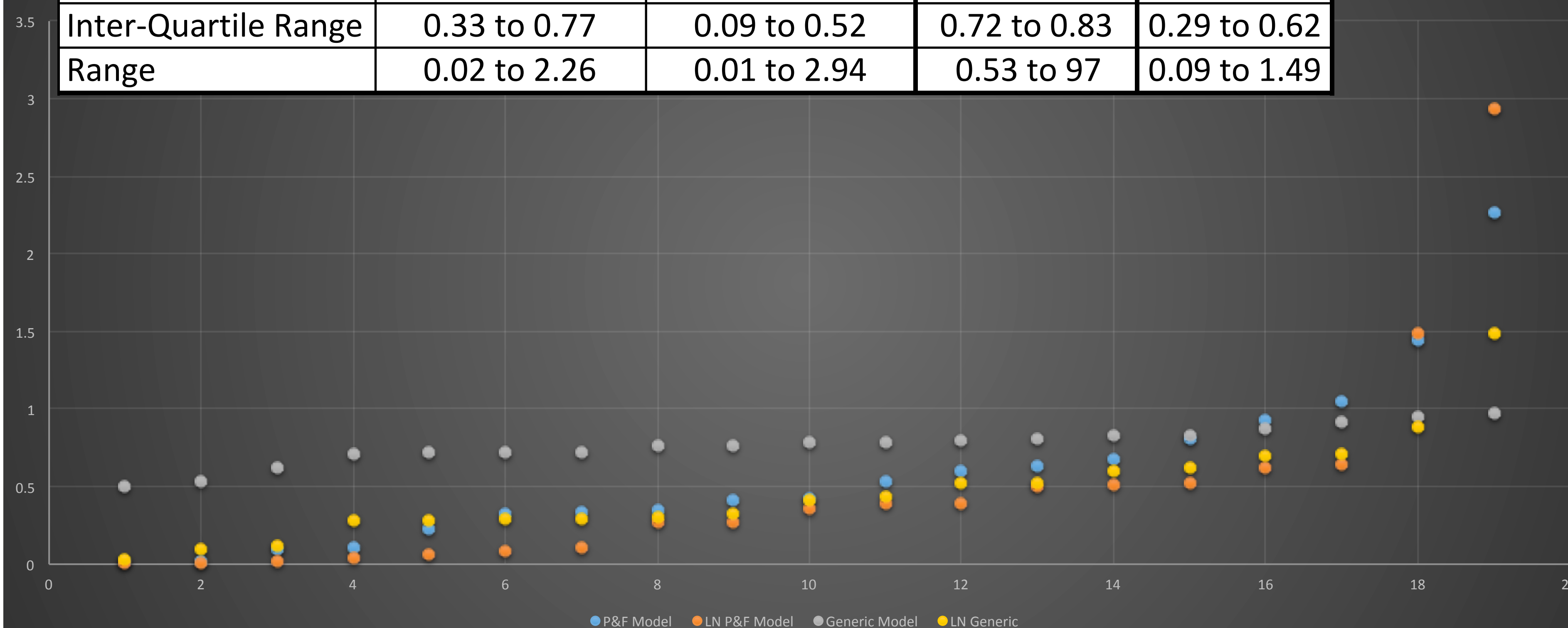
SW_Cost = 0.65*Sensor_Cost^(.78)

Basic Stats	**Linear**	**LN Linear**
Sample Size:	19	19
R-Squared:	0.55	0.56
F-test:	23	24
Slope (significance):	< 0.0001	< 0.0001
Intercept (significance):	0.67	0.77

- MRE results on next page
- Weakest over all results and no difference in model performance based on traditional stats
- Ln-Linear model performs better based on MRE results

Particles and Fields Software CER 2

MRE Stats	Linear Particles&Fields	Ln Linear Particles&Fields	Linear Generic	Ln Linear Generic
Median	0.48	0.37	0.79	0.42
Mean	0.62	0.51	0.78	0.49
Pred(30)	0.11	0.42	0.00	0.37
Inter-Quartile Range	0.33 to 0.77	0.09 to 0.52	0.72 to 0.83	0.29 to 0.62
Range	0.02 to 2.26	0.01 to 2.94	0.53 to 97	0.09 to 1.49



Peer Review

- Conducted multiple reviews
  - NICM development team
  - Payload engineers including Team X instrument chair and Team I lead facilitator
  - Instrument software CogE (separately)

Bootstrap Cross Validation

- Conducted multiple reviews
  - NICM development team
  - Payload engineers including Team X instrument chair and Team I lead facilitator
  - Instrument software CogE (separately)

Model Performance

Jet Propulsion Laboratory

- Bootstrap Cross Validation
  - *Bootstrap*: Process for generating more meaningful statistics
    - *Example based on derivation of mean*
  - *Cross Validation*: Partitioning of data set into training and testing sets. Out-of-sample validation.
    - Models should be tested with data that were not used to fit the model. If you have enough data, it is best to hold back a random portion of the data to use for testing.
    - Cross validation is a trick to get out-of-sample tests but still use all the data by doing multiple fits, each time leaving out a different portion of the data

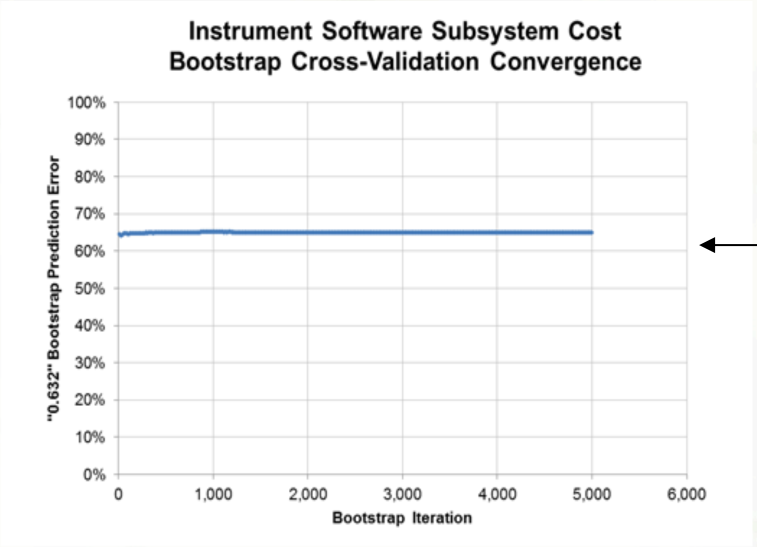
Model Performance continued…

- Bootstrap Cross-Validation:
  - 1) Use ALL original data → make MODEL
    - 1 model error measure σ^2(app)
  - 2) Use 999 random samples using SOME of the original data → make 999 sample MODELS
    - Find average of 999 model error measures σ^2(BCV)

$$VAR_{bs} = \sum_{i}^{N} \left[\frac{\sum_{m}^{|M(i)|} \left(EC_{im} - C_i\right)^2}{|M(i)|} \right] * \frac{1}{N}$$

Model Performance continued…

- Validation is done in log-log space
- Validated statistical techniques.
  - All Excel workbook algorithms for Cluster Analysis, PCA and Bootstrap Cross Validation have been verified by comparison with S-Plus and Matlab functions



VERY stable error results

REesult

Basic Regression before boot strapping

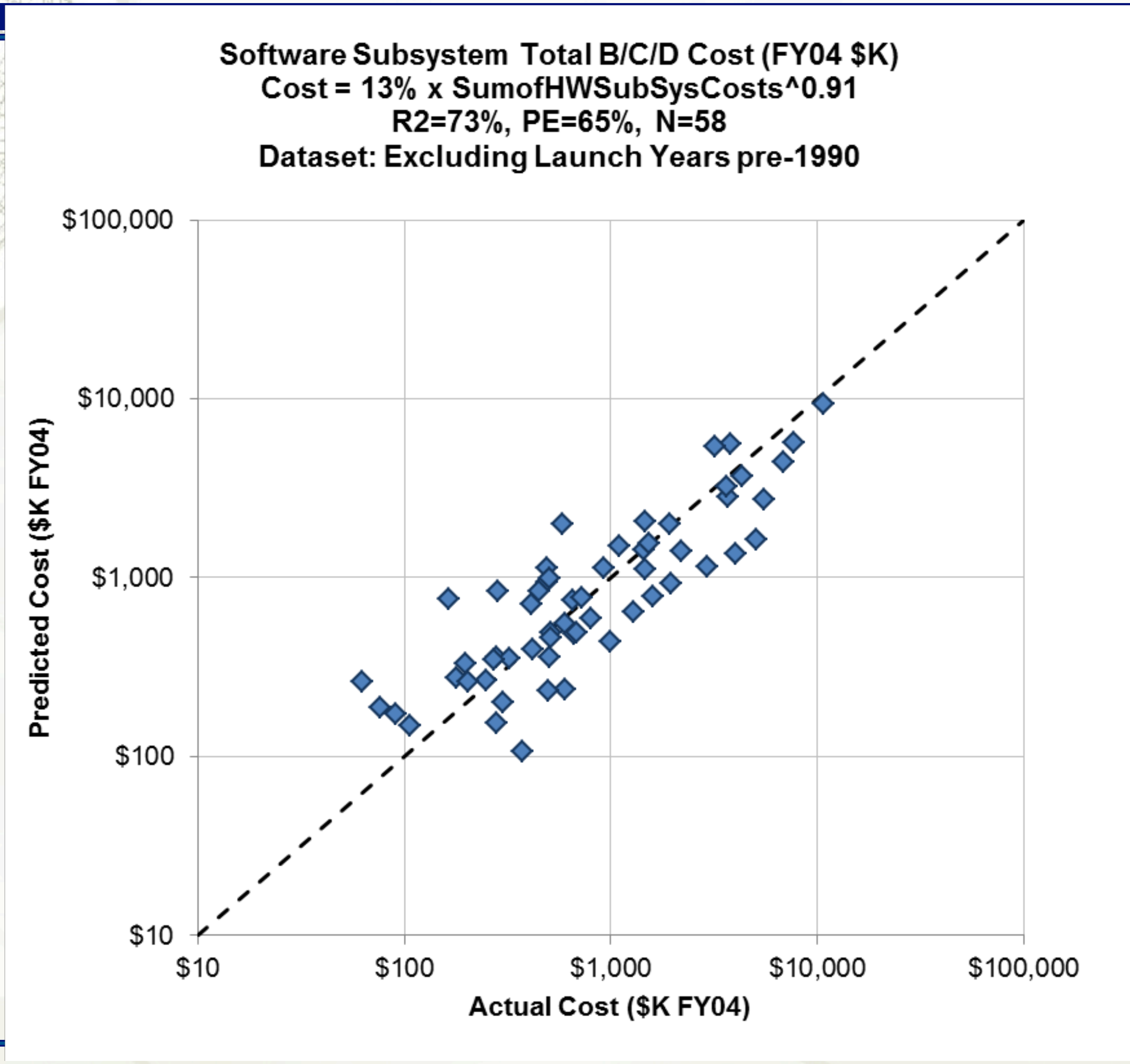
$$SW\_Cost = 0.18*Sensor\_Cost^{0.88}$$

Basic Regression after boot strapping

$$SW\_Cost = 0.13*Sensor\_Cost^{0.91}$$

Final NICM SW CER

Laboratory



Conclusion and Recommendations

Recommendation

- Use the LN-Linear form of the Generic Model

Reason

- There is no dominant CER
  - While some Instrument type models do perform better based on some measures no specific CER is best by all statistics
  - Only Optical and Particles & Fields Instruments have enough observations to build a "proper" model
- Always in ball park of best performing CER (it is either second and occasionally first
- A consistent CER that is "good enough" is desirable over a number of sub-models with complicated selection rules
- The only way to improve the CER is to identify technical parameters that should and expand data set

Parameters being Contemplated

- SW Effort
- LOC
- Deployables
- Actuators
- Complexity
  - Science
  - Mechanical control
  - Electronics
- Heritage
- Reuse/Inheritance
- Pointing accuracy
- Number of and complexity of Interfaces
- Developer