Uncertainty Analysis with COCOMO

Dr. Karen W. Pullen
Presented at the Third Annual COCOMO Users' Group Meeting
November 4-5, 1987

MITRE
Background

- Cost analyst often asked for:
  - Confidence in estimate
  - Confidence interval around estimate
  - An estimate with a given likelihood of occurring for both effort and schedule

MITRE
Problem Statement

What is the Cumulative Probability Distribution of

Effort

and

Schedule

Given:

(1) COCOMO Effort Estimate

(2) Some additional uncertainty information about that estimate

MITRE
Why Is This New?

Boehm Gives confidence limits:

Basic COCOMO
"Within a factor of 2, 60% of the time"

Intermediate COCOMO
"Within 20% of the actuals, 68% of the time"

But those results refer to statistical model uncertainty on COCOMO database with perfect knowledge of inputs

COST ANALYSTS HAVE NO CERTAINTY REGARDING INPUTS

MITRE
Scope of Presentation

- Effort uncertainty
  - Appropriate models
    Triangular distribution
    Normal distribution
  - Uncertainty Parameters

- Schedule uncertainty
  - Appropriate model
  - Monte Carlo Analysis
  - Normalizing

- Examples

MITRE
Effort Uncertainty - First Model

Early in Development
Pre-Specification

EVOLUTION OF SYSTEM DEFINITION

TRIANGULAR DISTRIBUTION OF EFFORT

MITRE
Uncertainty Model When Effort is Triangular

If \( \hat{E} \) is the Point Estimate and \( a \) is the Uncertainty Multiplier

\[
F_E(x) = \begin{cases} 
\frac{(x - \hat{E}/a)^2}{(a\hat{E} - \hat{E}/a)(\hat{E} - \hat{E}/a)} & \text{if } \hat{E}/a < x < \hat{E} \\
1 - \frac{(x - a\hat{E})^2}{(a\hat{E} - \hat{E}/a)(a\hat{E} - \hat{E})} & \text{if } \hat{E} < x < a\hat{E}
\end{cases}
\]

\( F_E(x) \) is the probability that actual effort \( \leq x \)

MITRE
Example: Cumulative Distribution of Effort

MITRE
Effort Uncertainty - Second Model

- More sophisticated analysis using
  - Size probability distribution
  - Cost driver attribute uncertainty
  - Model uncertainty


- Yields:

  Normal Distribution of Effort with mean $\bar{E}$ and variance $\sigma^2$
Example: Effort is Normally Distributed

\[ D = \frac{\sigma}{E} \quad \text{(Coefficient of Dispersion)} \]

MITRE
Schedule Uncertainty

- More complicated analytically since

  \[ \hat{S}|E = 2.5 \ E^{32} \]
  (embedded-mode schedule model)

- E is a random variable

- Schedule model has its own uncertainty
Problem Statement

- Find distribution of $S$, the actual schedule if all we know is:

  $\hat{E}$ and $a$ if Effort has a triangular distribution

  or

  $E$ and $D$ if Effort is normally distributed
Monte Carlo Approach

- Use computer program to
  - Generate random "sample" of 50000 efforts according to effort distribution
  - For each effort, calculate \( \hat{S}|E \)
  - Then generate a random \( S \) according to distribution of \( S|E \)
2 Kinds of Results for Schedule Distribution

1. If effort has a triangular distribution
   family of curves, one for each value of
   \( \alpha = 1.5, 2, 3, 4 \)

2. If effort is normally distributed
   family of curves, one for each value of
   \( D = .083, .167, .25, .3 \)

MITRE
Probability Density of Schedule when Effort is Triangular

OBSERVED CUMULATIVE PROBABILITY FROM MONTE CARLO

GIVEN \( \frac{500}{a} < E < 500a \)

MITRE
Probability Density of Schedule when Effort is Normal

Observed cumulative probability from Monte Carlo

Schedule given $E = 500$

$D = 0.3$

$D = 0.25$

$D = 0.167$

$D = 0.083$

MITRE
Normalize S So Results Can Be Generalized

Define $SNORM = \frac{S - S_{min}}{S_{max} - S_{min}}$

Where

$S_{min}$ is: $S_{max}$ is:

Normal Effort $(.7)(2.5)(\bar{E}(1-3D))^{32}$ $(2)(2.5)(\bar{E}(1+3D))^{32}$

Triangular Effort $(.7)(2.5)(\hat{E}/a)^{32}$ $(2)(2.5)(a\hat{E})^{32}$
Example
Given $\hat{E} = 500$ and $a = 2$

What are the 80% Confidence Limits for Effort and Schedule?

In General:
Example (continued)

- What are the 80% confidence limits for effort?

Solve

\[ .9 = F(e_2) \quad \text{for } e_1, e_2 \]
\[ .1 = F(e_1) \]

\[ .1 = \frac{(e_1 - 250)^2}{(1000 - 250)(500 - 250)} \Rightarrow e_1 = 387 \]

\[ .9 = 1 - \frac{(e_2 - 1000)^2}{(1000 - 250)(1000 - 500)} \Rightarrow e_2 = 806 \]

(don't forget that 250 < e_1 < e_2 < 1000!)

MITRE
Example (concluded)

What are the 80% Confidence Limits for Schedule?

Solve \[ .9 = F(S_{\text{NORM}2}) \] for \( S_{\text{NORM}_1}, S_{\text{NORM}_2} \) 
\[ .1 = F(S_{\text{NORM}_1}) \]

Then convert to \( s_1, s_2 \) using 
\[ S_{\text{min}} = .7(2.5)(250)^{.32} = 10.2 \]
\[ S_{\text{max}} = 2(2.5)(1000)^{.32} = 45.6 \]

From Table
\[ S_{\text{NORM}_1} = .175 \implies s_1 = S_{\text{NORM}_1} (S_{\text{max}} - S_{\text{min}}) + S_{\text{min}} = 16.4 \text{ months} \]
\[ S_{\text{NORM}_2} = .60 \implies s_2 = S_{\text{NORM}_2} (S_{\text{max}} - S_{\text{min}}) + S_{\text{min}} = 31.4 \text{ months} \]

MITRE
Conclusions

- Specifying probabilities more useful than single point estimates

- Since decision-maker can then position the estimate with respect to risk

MITRE
Cumulative distribution of storm

When effort is triangularly distributed

Results from Monte Carlo analysis

\[ A = 1.5 \]

\[ A = 2.0 \]

\[ A = 3.0 \]

\[ A = 4.0 \]

\[ 100 \text{ SOWM} \]
RESULTS FROM MONTE CARLO ANALYSIS FOR NORMALLY DISTRIBUTED EFFORT

CUMULATIVE DISTRIBUTION OF 100 * SNORM

<table>
<thead>
<tr>
<th>100*SNORM</th>
<th>D = 0.083</th>
<th>D = 0.167</th>
<th>D = 0.25</th>
<th>D = 0.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>2</td>
<td>0.0001</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>3</td>
<td>0.0004</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>4</td>
<td>0.0008</td>
<td>0.0002</td>
<td>0.0001</td>
<td>0.0000</td>
</tr>
<tr>
<td>5</td>
<td>0.0020</td>
<td>0.0005</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>6</td>
<td>0.0041</td>
<td>0.0010</td>
<td>0.0003</td>
<td>0.0002</td>
</tr>
<tr>
<td>7</td>
<td>0.0072</td>
<td>0.0017</td>
<td>0.0009</td>
<td>0.0003</td>
</tr>
<tr>
<td>8</td>
<td>0.0130</td>
<td>0.0032</td>
<td>0.0016</td>
<td>0.0007</td>
</tr>
<tr>
<td>9</td>
<td>0.0196</td>
<td>0.0057</td>
<td>0.0022</td>
<td>0.0010</td>
</tr>
<tr>
<td>10</td>
<td>0.0273</td>
<td>0.0100</td>
<td>0.0028</td>
<td>0.0020</td>
</tr>
<tr>
<td>11</td>
<td>0.0344</td>
<td>0.0146</td>
<td>0.0049</td>
<td>0.0027</td>
</tr>
<tr>
<td>12</td>
<td>0.0424</td>
<td>0.0220</td>
<td>0.0078</td>
<td>0.0033</td>
</tr>
<tr>
<td>13</td>
<td>0.0519</td>
<td>0.0297</td>
<td>0.0111</td>
<td>0.0045</td>
</tr>
<tr>
<td>14</td>
<td>0.0635</td>
<td>0.0393</td>
<td>0.0150</td>
<td>0.0067</td>
</tr>
<tr>
<td>15</td>
<td>0.0771</td>
<td>0.0482</td>
<td>0.0212</td>
<td>0.0088</td>
</tr>
<tr>
<td>16</td>
<td>0.0913</td>
<td>0.0608</td>
<td>0.0294</td>
<td>0.0120</td>
</tr>
<tr>
<td>17</td>
<td>0.1057</td>
<td>0.0732</td>
<td>0.0394</td>
<td>0.0162</td>
</tr>
<tr>
<td>18</td>
<td>0.1226</td>
<td>0.0888</td>
<td>0.0498</td>
<td>0.0223</td>
</tr>
<tr>
<td>19</td>
<td>0.1394</td>
<td>0.1078</td>
<td>0.0629</td>
<td>0.0304</td>
</tr>
<tr>
<td>20</td>
<td>0.1602</td>
<td>0.1250</td>
<td>0.0788</td>
<td>0.0400</td>
</tr>
<tr>
<td>21</td>
<td>0.1802</td>
<td>0.1467</td>
<td>0.0951</td>
<td>0.0505</td>
</tr>
<tr>
<td>22</td>
<td>0.2015</td>
<td>0.1682</td>
<td>0.1136</td>
<td>0.0640</td>
</tr>
<tr>
<td>23</td>
<td>0.2219</td>
<td>0.1895</td>
<td>0.1327</td>
<td>0.0793</td>
</tr>
<tr>
<td>24</td>
<td>0.2417</td>
<td>0.2127</td>
<td>0.1555</td>
<td>0.0973</td>
</tr>
<tr>
<td>25</td>
<td>0.2645</td>
<td>0.2383</td>
<td>0.1788</td>
<td>0.1166</td>
</tr>
<tr>
<td>26</td>
<td>0.2884</td>
<td>0.2599</td>
<td>0.2052</td>
<td>0.1353</td>
</tr>
<tr>
<td>27</td>
<td>0.3102</td>
<td>0.2837</td>
<td>0.2295</td>
<td>0.1584</td>
</tr>
<tr>
<td>28</td>
<td>0.3310</td>
<td>0.3080</td>
<td>0.2541</td>
<td>0.1830</td>
</tr>
<tr>
<td>29</td>
<td>0.3538</td>
<td>0.3341</td>
<td>0.2804</td>
<td>0.2099</td>
</tr>
<tr>
<td>30</td>
<td>0.3746</td>
<td>0.3571</td>
<td>0.3059</td>
<td>0.2383</td>
</tr>
<tr>
<td>31</td>
<td>0.3952</td>
<td>0.3811</td>
<td>0.3345</td>
<td>0.2651</td>
</tr>
<tr>
<td>32</td>
<td>0.4174</td>
<td>0.4060</td>
<td>0.3622</td>
<td>0.2927</td>
</tr>
<tr>
<td>33</td>
<td>0.4378</td>
<td>0.4296</td>
<td>0.3918</td>
<td>0.3194</td>
</tr>
<tr>
<td>34</td>
<td>0.4592</td>
<td>0.4541</td>
<td>0.4169</td>
<td>0.3515</td>
</tr>
<tr>
<td>35</td>
<td>0.4774</td>
<td>0.4775</td>
<td>0.4412</td>
<td>0.3808</td>
</tr>
<tr>
<td>36</td>
<td>0.4966</td>
<td>0.4981</td>
<td>0.4658</td>
<td>0.4101</td>
</tr>
<tr>
<td>37</td>
<td>0.5155</td>
<td>0.5191</td>
<td>0.4922</td>
<td>0.4361</td>
</tr>
<tr>
<td>38</td>
<td>0.5330</td>
<td>0.5390</td>
<td>0.5179</td>
<td>0.4626</td>
</tr>
<tr>
<td>39</td>
<td>0.5520</td>
<td>0.5595</td>
<td>0.5415</td>
<td>0.4908</td>
</tr>
<tr>
<td>40</td>
<td>0.5694</td>
<td>0.5821</td>
<td>0.5653</td>
<td>0.5188</td>
</tr>
<tr>
<td>41</td>
<td>0.5877</td>
<td>0.6012</td>
<td>0.5880</td>
<td>0.5447</td>
</tr>
<tr>
<td>42</td>
<td>0.6058</td>
<td>0.6202</td>
<td>0.6088</td>
<td>0.5714</td>
</tr>
<tr>
<td>43</td>
<td>0.6229</td>
<td>0.6399</td>
<td>0.6310</td>
<td>0.5967</td>
</tr>
<tr>
<td>44</td>
<td>0.6385</td>
<td>0.6588</td>
<td>0.6526</td>
<td>0.6175</td>
</tr>
<tr>
<td>45</td>
<td>0.6562</td>
<td>0.6758</td>
<td>0.6728</td>
<td>0.6393</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>46</td>
<td>0.6712</td>
<td>0.6931</td>
<td>0.6914</td>
<td>0.6618</td>
</tr>
<tr>
<td>47</td>
<td>0.6867</td>
<td>0.7096</td>
<td>0.7090</td>
<td>0.6844</td>
</tr>
<tr>
<td>48</td>
<td>0.7000</td>
<td>0.7259</td>
<td>0.7286</td>
<td>0.7053</td>
</tr>
<tr>
<td>49</td>
<td>0.7143</td>
<td>0.7402</td>
<td>0.7444</td>
<td>0.7233</td>
</tr>
<tr>
<td>50</td>
<td>0.7281</td>
<td>0.7558</td>
<td>0.7600</td>
<td>0.7419</td>
</tr>
<tr>
<td>51</td>
<td>0.7437</td>
<td>0.7701</td>
<td>0.7774</td>
<td>0.7614</td>
</tr>
<tr>
<td>52</td>
<td>0.7572</td>
<td>0.7845</td>
<td>0.7921</td>
<td>0.7772</td>
</tr>
<tr>
<td>53</td>
<td>0.7687</td>
<td>0.7975</td>
<td>0.8075</td>
<td>0.7937</td>
</tr>
<tr>
<td>54</td>
<td>0.7808</td>
<td>0.8105</td>
<td>0.8212</td>
<td>0.8091</td>
</tr>
<tr>
<td>55</td>
<td>0.7921</td>
<td>0.8233</td>
<td>0.8342</td>
<td>0.8243</td>
</tr>
<tr>
<td>56</td>
<td>0.8049</td>
<td>0.8347</td>
<td>0.8470</td>
<td>0.8387</td>
</tr>
<tr>
<td>57</td>
<td>0.8159</td>
<td>0.8468</td>
<td>0.8592</td>
<td>0.8504</td>
</tr>
<tr>
<td>58</td>
<td>0.8273</td>
<td>0.8587</td>
<td>0.8709</td>
<td>0.8637</td>
</tr>
<tr>
<td>59</td>
<td>0.8384</td>
<td>0.8692</td>
<td>0.8825</td>
<td>0.8756</td>
</tr>
<tr>
<td>60</td>
<td>0.8487</td>
<td>0.8789</td>
<td>0.8939</td>
<td>0.8876</td>
</tr>
<tr>
<td>61</td>
<td>0.8591</td>
<td>0.8895</td>
<td>0.9023</td>
<td>0.8975</td>
</tr>
<tr>
<td>62</td>
<td>0.8685</td>
<td>0.8995</td>
<td>0.9109</td>
<td>0.9065</td>
</tr>
<tr>
<td>63</td>
<td>0.8768</td>
<td>0.9072</td>
<td>0.9190</td>
<td>0.9161</td>
</tr>
<tr>
<td>64</td>
<td>0.8861</td>
<td>0.9161</td>
<td>0.9270</td>
<td>0.9246</td>
</tr>
<tr>
<td>65</td>
<td>0.8946</td>
<td>0.9240</td>
<td>0.9337</td>
<td>0.9315</td>
</tr>
<tr>
<td>66</td>
<td>0.9027</td>
<td>0.9303</td>
<td>0.9397</td>
<td>0.9388</td>
</tr>
<tr>
<td>67</td>
<td>0.9108</td>
<td>0.9360</td>
<td>0.9468</td>
<td>0.9457</td>
</tr>
<tr>
<td>68</td>
<td>0.9189</td>
<td>0.9429</td>
<td>0.9541</td>
<td>0.9529</td>
</tr>
<tr>
<td>69</td>
<td>0.9265</td>
<td>0.9502</td>
<td>0.9605</td>
<td>0.9595</td>
</tr>
<tr>
<td>70</td>
<td>0.9332</td>
<td>0.9563</td>
<td>0.9651</td>
<td>0.9646</td>
</tr>
<tr>
<td>71</td>
<td>0.9386</td>
<td>0.9619</td>
<td>0.9696</td>
<td>0.9688</td>
</tr>
<tr>
<td>72</td>
<td>0.9454</td>
<td>0.9668</td>
<td>0.9740</td>
<td>0.9732</td>
</tr>
<tr>
<td>73</td>
<td>0.9510</td>
<td>0.9716</td>
<td>0.9778</td>
<td>0.9774</td>
</tr>
<tr>
<td>74</td>
<td>0.9565</td>
<td>0.9763</td>
<td>0.9812</td>
<td>0.9811</td>
</tr>
<tr>
<td>75</td>
<td>0.9622</td>
<td>0.9793</td>
<td>0.9841</td>
<td>0.9836</td>
</tr>
<tr>
<td>76</td>
<td>0.9674</td>
<td>0.9823</td>
<td>0.9873</td>
<td>0.9872</td>
</tr>
<tr>
<td>77</td>
<td>0.9709</td>
<td>0.9856</td>
<td>0.9899</td>
<td>0.9899</td>
</tr>
<tr>
<td>78</td>
<td>0.9748</td>
<td>0.9880</td>
<td>0.9925</td>
<td>0.9922</td>
</tr>
<tr>
<td>79</td>
<td>0.9791</td>
<td>0.9910</td>
<td>0.9941</td>
<td>0.9942</td>
</tr>
<tr>
<td>80</td>
<td>0.9829</td>
<td>0.9931</td>
<td>0.9949</td>
<td>0.9950</td>
</tr>
<tr>
<td>81</td>
<td>0.9861</td>
<td>0.9943</td>
<td>0.9959</td>
<td>0.9958</td>
</tr>
<tr>
<td>82</td>
<td>0.9884</td>
<td>0.9955</td>
<td>0.9964</td>
<td>0.9964</td>
</tr>
<tr>
<td>83</td>
<td>0.9906</td>
<td>0.9964</td>
<td>0.9970</td>
<td>0.9971</td>
</tr>
<tr>
<td>84</td>
<td>0.9925</td>
<td>0.9970</td>
<td>0.9979</td>
<td>0.9981</td>
</tr>
<tr>
<td>85</td>
<td>0.9942</td>
<td>0.9977</td>
<td>0.9987</td>
<td>0.9987</td>
</tr>
<tr>
<td>86</td>
<td>0.9960</td>
<td>0.9985</td>
<td>0.9991</td>
<td>0.9992</td>
</tr>
<tr>
<td>87</td>
<td>0.9972</td>
<td>0.9992</td>
<td>0.9996</td>
<td>0.9996</td>
</tr>
<tr>
<td>88</td>
<td>0.9979</td>
<td>0.9996</td>
<td>0.9998</td>
<td>0.9998</td>
</tr>
<tr>
<td>89</td>
<td>0.9987</td>
<td>0.9998</td>
<td>0.9998</td>
<td>0.9999</td>
</tr>
<tr>
<td>90</td>
<td>0.9993</td>
<td>0.9999</td>
<td>0.9999</td>
<td>0.9999</td>
</tr>
<tr>
<td>91</td>
<td>0.9997</td>
<td>0.9999</td>
<td>0.9999</td>
<td>0.9999</td>
</tr>
<tr>
<td>92</td>
<td>0.9999</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>93</td>
<td>0.9999</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>94</td>
<td>0.9999</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>95</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>96</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>97</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
Calibrating COCOMO for Ada

Barbara E. Wolfinger

Presented at the Third Annual COCOMO User's Group Meeting
4-5 November 1987

MITRE
Data Collection Process

RESEARCH LITERATURE — IDENTIFY 200 PROJECTS

LOCATE CONTACTS FOR 52 PROJECTS

CONTACTS ON 26 PROJECTS WILLING TO GIVE INFO

RECEIVE MATERIAL FROM 14 PROJECTS

INTERVIEW DEVELOPERS OF 12 PROJECTS AT THEIR FACILITIES

RECEIVE ESD QUESTIONNAIRE FOR 1 PROJECT

ASSIMILATE DATA FROM 13 QUESTIONNAIRES

BUILD DATA BASE OF PARAMETRIC DATA ON 13 Ada PROJECTS

ELIMINATE 3 ATYPICAL PROJECTS

PERFORM DATA ANALYSES ON 10 DoD-LIKE PROJECTS

PUBLISH RECOMMENDATIONS FOR ESTIMATING COSTS OF Ada DEVELOPMENTS

MITRE
Ada Data Base

- 10 projects, 8 different developers
- Size: 2 - 95 KDSI
- Effort: 11 - 640 SM
- Timeframe: 1984 - 1987
- Commercial and military applications

Typical features:
- First Ada effort
- Low-risk, familiar application
- Stable requirements
- Organic mode developments

MITRE
Calibrated Basic Effort Model

- $SM = 2.8 \times (KDSI)^{1.17}$
- Estimated within 20% of development effort of projects in data base, 80% of the time.
- Nonlinear regression techniques.
- $R^2 = .98$
COCOMO and Ada Basic Effort Equations

EFFORT (SM)

SIZE (KDSI)

COCOMO EMB
ADA DOD
COCOMO SD
COCOMO ORG

MITRE
Ada Vs. COCOMO Data

EFFORT (SM)

COCOMO SD
COCOMO ORG
Ada EFFORT

SIZE (KDSI)

MITRE
Calibrated Intermediate Effort Model

- SM = EAF x 2.2 (KDSI)^1.17
- R^2 = .49
- Estimated within 20% of development effort of projects in the database, 60% of the time.
- DEM correlation

<table>
<thead>
<tr>
<th>Attribute</th>
<th>DEMs vs. IEMs</th>
<th>DEMs vs. Ideal Ratios</th>
</tr>
</thead>
<tbody>
<tr>
<td>TIME</td>
<td>.01</td>
<td>-.19</td>
</tr>
<tr>
<td>TURN</td>
<td>.03</td>
<td>-.25</td>
</tr>
<tr>
<td>AEXP</td>
<td>-.07</td>
<td>.07</td>
</tr>
<tr>
<td>VEXP</td>
<td>.09</td>
<td>.12</td>
</tr>
<tr>
<td>MODP</td>
<td>.44</td>
<td>.36</td>
</tr>
<tr>
<td>TOOL</td>
<td>.18</td>
<td>.30</td>
</tr>
<tr>
<td>SCED</td>
<td>.13</td>
<td>-.27</td>
</tr>
</tbody>
</table>

MITRE
Recalibrated Intermediate Effort Model

- $SM = EAF \times 3.1 \ (KDSI)^{1.17}$
- $EAF = MODP \times TOOL$
- $R^2 = .99$
- Estimated within 20% of development effort of projects in the data base, 60% of the time.
Calibrated Schedule Model

- TDEV $= 5.8 \times (SM)^{23}$
- Estimated within 20% of development schedule of projects in data base, 100% of the time.
- Nonlinear regression techniques.
- $R^2 = .82$
Conclusions and Recommendations

- Ada overall productivity similar to non-Ada.

- Effort and schedule equations require calibration for Ada DOD projects.

- Basic calibrated effort model preferred over intermediate model for use in near future.

- Continue collecting data to determine:
  - Appropriate cost driver attributes for Ada.
  - Appropriate DEMs for Ada.
Expert System Cost Model

(ESCOMO)

Prototype

Virginia C. Day

Presented at the third Annual COCOMO User's Group Meeting
November 4-5, 1987

MITRE
Scope

- Motivation
- Objectives
- ESCOMO
- Lessons learned
- Future extensions
- Value of this approach

MITRE
Motivation

- Attribute setting criteria are often applied incorrectly.
- Cost attributes are not independent.
- Estimates are error prone.
- Estimates are difficult to check.
- Some development situations are unrealistic.
Objectives

- Promote use of standardized criteria to set cost attribute ratings
- Teach methodology
- Discover attribute relationships
- Check estimates for
  - Consistency
  - Reasonableness of attribute settings
What ESCOMO IS

- Implementation of Intermediate COCOMO
- PC/AT with TOPSI
- Includes
  - Interactive cost attribute rate setting with user help
  - Consistency checking
- Prototype status

MITRE
40 Consistency Checks

- Covering
  - Workforce size
  - Mode
  - Language
  - Schedule
  - Database size
  - Personnel

- Sources
  - Boehm's Software Engineering Economics
  - MITRE Cost Analysis Technical Center
Estimate Example with Inconsistencies

Input

root component: P1, mode: semidetached
  virt: high vexp: high tool: high acap: high
  pcap: high lexp: high

subcomponent: Data1
  data: high acap: high pcap: high lexp: high kedsi: 65

subcomponent: Msg-pr
  rely: high virt: low stor: high tool: very-high pcap: low kedsi: 40

available schedule: 17 months

estimate timeframe: prior to PDR

Output

schedule: 22 months
effort: 475 staff months
average workforce: 22 people
productivity: 221 DSI/SM

MITRE
Consistency Checks for Estimate Example

- SCED should be set to LOW because the available months are < 85% of the estimated schedule.

- For component P1:
  - If VIRT is greater than NOMINAL, VEXP should not be greater than NOMINAL.

- For component Data1:
  - warning: DATA has been set at HIGH or greater; have the database design functions been included in the equivalent LOC estimate?

MITRE
Consistency Checks for Estimate Example (Cont.)

• For component MSG-PR
  - Early in development, STOR should not be greater than NOMINAL. It is almost always advisable to purchase additional processing power.
  - If PCAP is less than NOMINAL, it is not likely that the programmers will make full use of greater than NOMINAL TOOLS.
  - PCAP should not be less than nominal if RELY is greater than NOMINAL.
Use/Evolution

- System could be used for
  - Training
  - Estimate preparation
  - Checking
  - Collect review information
  - Detect expert review errors
  - Point out developmental problems

- ESCOMO will evolve through use of prototype
  - Training → deficiencies in help
  - Comparison of output and expert review → new / corrected rules

MITRE
Lessons Learned

- Experts don't always agree on the relative importance of consistency checks
- Difficult to verify consistency checks
- Expert system languages for PC are immature
Possible Extensions

- Tailor cost attribute information to application
  - Example: C3I projects - RELY usually high or very-high
- Include ideas from other models
  - Example: Jensen - multiple site development
    ESD Quick SET - application, function categories
- Add to consistency checks
Value of this Approach

- Reduce software costing errors
- Flag problematic developments
- Facilitate and improve the review process
ESCOMO includes consistency checking which can be invoked at any time during (or after) estimate preparation. This part of the system traps user errors (or gives warnings) whenever cost attribute settings and other parameters violate the rules governing reasonable combinations of values for these parameters. Although most of the rules are concerned with settings within individual components, some relate to system-wide attributes such as the project schedule constraint, the size of the workforce involved in the development, or the development mode.

An attempt has been made to separate the consistency checks into categories for clarity. However, since most of these rules involve several parameters, it is difficult to determine which group a particular rule should be placed in. Most checks, which are of the form:

condition relating to attribute
implication relating to other attributes,

are categorized by the first attribute.

Workforce Size

When a project team is large (≥ 60), it is not generally possible to have better than average capability among project personnel. This is so because good people are hard to find and workforce quality tends to average out. In addition, a project utilizing a large development team experiences losses due to communications overhead.
PCAP—ACAP—wkforce

If the workforce is large,
then
PCAP and ACAP should probably be set to NOMINAL.
Reason: Good programmers and analysts are hard to find and, in a
large workforce, personnel quality tends to average out.
A large workforce is considered to be one consisting of
60 or more staff persons.

Project Mode

The mode checks try to insure that 1) project mode is correctly
set, and 2) that other attributes are consistent with the chosen
mode.

If the software project requires very high reliability and is
highly complex, then the mode should be "embedded" and vice versa.
In such projects, it is also true that the virtual machine should not
be volatile (i.e., in the process of evolution), and that the
schedule should not be constrained. Although the schedule is
generally tight for such projects, it should not be too constrained
or the system is likely to be less reliable than it should be.

The reverse of most of these rules is true for the "organic"
mode. However, the virtual machine should also not be volatile and
the schedule should not be constrained. By definition, organic mode
projects are developed in "familiar, in-house" environment and have
few constraints. In addition, the project staff should be
experienced with the application in development, the virtual machine,
and the language.

The only consistency check for semidetached mode project checks
that applications experience is not less than average. This again is
the result of Boehm's definition. The reason that there are no more
checks for this mode is that it is very difficult to characterize semidetached projects because they are in an intermediate category where neither organic nor embedded projects fit. Most of these projects have a mixture of properties of the two extreme modes.

mode-check-1

If both RELY and CPLX are greater than HIGH, then
MODE should be embedded.
Reason: COCOMO definition for embedded mode software development. (See also reason for MODE-checks-2 and 3.)

MODE-check-2 and -3

If MODE is embedded then
neither RELY nor CPLX should be less than HIGH.
Reason: Boehm’s definition of embedded mode software states that the system under development is tightly constrained. The software is embedded in a “strongly coupled complex of hardware, software, regulations, and operational systems.”

MODE-check-4

If MODE is EMBEDDED, then
VIRT should not be less than NOMINAL.
Reason: See mode-check-2 and -3. The virtual machine on which an embedded system project is developed is using new and changing technology and is, therefore, not very stable.

MODE-check-5

If MODE is EMBEDDED, then
it is not likely that SCED will be greater than NOMINAL.
Reason: Embedded mode COCOMO definition. There is generally a high premium on early completion of these projects.
MODE-check-6
If mode is SEMIDETACHED
then
AEXP should be at least NOMINAL.
Reason: Boehm states that personnel have considerable experience
in working with related software applications in
semidetached mode development.

MODE-check-7
If both RELY and CPLX are NOMINAL or lower
then
MODE should be ORGANIC
Reason: COCOMO definition for organic mode projects. This type
of project has flexible requirements and is not
excessively complex.

MODE-check-8
If MODE is ORGANIC,
then
VIRT should not be greater than NOMINAL.
Reason: The virtual machine should be very stable for organic
mode projects.

MODE-check-9 and -10
If MODE is organic,
then
AEXP and VEXP should be greater than NOMINAL.
Reason: Organic mode projects are developed in a stable and
familiar environment. Personnel have experience with
both the type of application and the virtual machine.

MODE-check-11 and -12
If MODE is ORGANIC
then
neither RELY nor (overall) CPLX should be greater than NOMINAL.
Reason: Organic mode projects, by definition, have flexible
requirements and are not overly complex.
MODE-check-13

If MODE is ORGANIC
then
SCED should not be less than NOMINAL.
Reason: Organic mode projects, by definition, do not operate under tight constraints of any kind.

Language

The language checks deal with 1) the fact that non-procedural languages are not covered by COCOMO, and 2) that for Ada projects with recently validated compilers, virtual machine volatility cannot be less than NOMINAL.

lang-check-2

Warning: Non-procedural languages are not covered by COCOMO. Special adjustments must be to estimates involving these languages.
Reason: The COCOMO database does not include any projects developed in non-procedural languages or Ada. The Ada parameter file developed by George Huff at MITRE allows adjustment to COCOMO for the use of Ada.

ADA-VIRT-check

For Ada projects developed in the far term.
VIRT should not be set less than NOMINAL.
Reason: Recently developed compilers (and other parts of the virtual machine) require frequent modification.

Schedule

The next five consistency checks deal with the ratio of the available to the estimated schedule. If this ratio is $\leq 85$ or $\geq 1.3$, then the project effort and schedule will be increased. When a development has too little time available, both the cost and duration
are higher because of overtime, increased errors, and low morale. A penalty is also enforced if the schedule is too long because people with too much time tend not to work as hard.

The remaining schedule checks are enabled for projects with a schedule constraint. For these projects, good use of software tools is important as are a stable virtual machine, reasonable turnaround time, and a capable staff.

VL-SCED-check
If the ratio of available to optimal schedule is less than or equal to .75,
then
SCED should be set to VERY-LOW.
Reason: COCOMO definition.

L-SCED-check
If the ratio of available to optimal schedule is greater than .75 and less than or equal to .85,
then
SCED should be set to LOW.
Reason: COCOMO definition.

N-SCED-check
If the ratio of available to optimal schedule is between .85 and 1.3,
then
SCED should be set to NOMINAL.
Reason: COCOMO definition.

H-SCED-check
If the ratio of available to optimal schedule is greater than or equal to 1.3 and less than 1.6,
then
SCED should be set to VERY-HIGH.
Reason: COCOMO definition.
VH-SCED-check
If the ratio of available to optimal schedule is greater than or equal to 1.6, then
SCED should be set to VERY-HIGH.
Reason: COCOMO definition.

SCED-TOOL-check
TOOL should not be less than NOMINAL if SCED is less than NOMINAL.
Reason: If the program schedule is constrained, the use of software tools should be at least average. Software tools make programming more efficient and should be used for all projects with tight schedules.

SCED-staff-check
If SCED is less than NOMINAL, then
both PCAP and ACAP should be at least NOMINAL or the project will probably not be finished on time.
Reason: Staff of less than average capability will probably not be able to successfully complete a project which is under a tight schedule constraint.

SCED-TURN-check
If SCED is less than NOMINAL, then
TURN should not be greater than NOMINAL.
Reason: If programmers have to wait excessive amounts of time to get their results, performance will suffer and programs with tight schedules will probably not be completed on time.

RELY-VIRT-check
Reliability and Complexity

These checks state that when reliability requirements are high, the virtual machine should be stable and that project schedule should not be constrained if both reliability and complexity are high.

If RELY is greater than NOMINAL, then
VIRT should not be greater than NOMINAL.
Reason: Systems requiring high reliability should not be developed in immature environments because of possibility of failure due to some change in the development system. Sometimes this situation occurs, but if it does, it is likely that system reliability will be compromised.

This check indicates that if both required reliability and project complexity are very high, the development schedule should not be constrained.

RELY-CPLX-SCED

If both RELY and CPLX are HIGH or greater, then
SCED should not be less than NOMINAL.
Reason: If a schedule constraint is placed on the development of a complex system the reliability is usually compromised.

The next set of consistency checks are warnings that if the data base size is very large, the developer must cope with this added complexity by not imposing additional difficulties on the project. In such developments, the personnel capability should be average or better and modern programming practices should be used. The developer must also be aware that the effort associated with data base design and management is often omitted from estimates for projects using large DBMS.
Large Database

HIGH-DATA-check
Warning: DATA has been set at HIGH or greater; has the effort associated with data management (if applicable) been included in the line of code estimate?
Reason: Data has been set at high or greater. Have the database design functions been included in the equivalent LOC estimate?

DATA-ACAP-check
Warning: PCAP and ACAP should not be less than NOMINAL if DATA is greater than NOMINAL (in most cases).
Reason: Most systems with a very large database have added complexity due to data management. Average programmer and analyst capability must be high enough to efficiently handle the associated problems.

DATA-MODP-check
MODP and TOOL should not be less than NOMINAL if DATA is greater than NOMINAL.
Reason: Most systems with a very large database have added complexity due to data management. Use of modern programming practices and software tools must be high enough to efficiently handle the associated problems.

Computer Related
Computer related constraints impact virtual machine experience and software tool use. If the virtual machine is volatile, it is not possible that project personnel will have had extensive experience with it. It is also true that the use of software tools requires an interactive environment and a stable virtual machine. Finally, early in any development, ample spare computer storage, speed, and turnaround time should be allowed, as projects almost always end up being much larger and more difficult than anticipated.
VIRT-VEXP-check

If VIRT is greater than NOMINAL, it is not likely that VEXP will be greater than NOMINAL.

Reason: If the virtual machine is continually undergoing change, it is not possible that the development staff will have, effectively, more than one year of experience with it.

TURN-TOOL-check

TURN is greater than NOMINAL, it is unlikely that TOOL use will be greater than NOMINAL.

Reason: The use of sophisticated software tools requires an interactive environment.

TIME-STOR-TURN

Early in development, TIME, STOR, and TURN should be NOMINAL or less.

Reason: It is almost always advisable to have the developers purchase enough additional processing power so that ample execution time and storage is available.

Software Tools

Use of software tools is important to the timely production of reliable software, but effective tool use requires capable personnel with virtual machine experience. Another check assumes that the use of modern programming practices implies the use of tools.

TOOL-VIRT-check

If VIRT is greater than NOMINAL, then it is not possible for TOOL use to be greater than NOMINAL.

Reason: If the virtual machine is not stable, then staff will not be able to use software tools effectively.
available software tools are probably also being used.

Reason: If modern programming practices are in use, then

\[ \text{IF MODP IS NOMINAL or GREATER}, \text{IT IS NOT LIKELY that TOOL USE} \]

\[ \text{TOOL-MODP-CHECK} \]

machine (effectively). Will not be able to use the tools (which are part of the virtual

Reason: Start with little experience with the virtual machine

\[ \text{IF VEXP IS LESS than NOMINAL}, \text{IT IS NOT POSSIBLE for TOOL USE to be GREATER than NOMINAL} \]

\[ \text{TOOL-VEXP-CHECK} \]

better than average capability.

Reason: Sophisticated software tool use requires programmers of

Programmers will make full use of GREATER than NOMINAL TOOLS.

\[ \text{IF PAP IS LESS than NOMINAL}, \text{IT IS NOT LIKELY that these} \]

\[ \text{TOOL-PAP-CHECK} \]

Sophisticated software tools.

average capability will probably not be able to make use of

The only personal related rule states that escort of LESS than
Recalibration of Basic and Nominal COCOMO Equations to Recent Air Force Acquisitions

Dr. Paul G. Funch
The MITRE Corporation
Bedford, Massachusetts

COCOMO is a primary tool used in preparing cost estimates for Air Force/Electronic Systems Division software acquisitions. These acquisitions are largely for embedded, human-machine interface programs that are part of command, control, communications, and/or radar systems. Historical software development data has been collected on 19 Air Force and 7 similar acquisitions, comprising a total of 110 Computer Software Configuration Items (CSCI). A great deal of effort was expended to verify the data, which resulted in a subset of projects being used to recalibrate five of the Basic and Nominal COCOMO equations. Our analyses indicate that, for this subset, significant alterations are required in the Basic and Nominal effort equations for the embedded and semidetached modes and in the schedule equation for the embedded mode.

The recalibration described here is based on determinations of the actual number of total equivalent delivered source instructions (TEDSI), the actual software development schedule (months), the actual effort data (staff months), and retrospective cost driver attribute ratings. The set of data used for the recalibrations was obtained from 14 Air Force projects and seven similar projects, comprising a total of 74 CSCIs which have been completed since 1976. Based on the complexity of the CSCI-CSCI interfaces and the programmer/analyst intercommunications required, an assessment was made of the extent to which each CSCI, or group of CSCIs, could be treated as independent software developments (subsystems). It was determined that the 21 projects and 74 CSCIs could be analyzed as 17 embedded mode, 11 semidetached mode, and 3 organic mode subsystems with completion dates since 1981.

Two recalibrations were performed on the estimating equations: 1) only the coefficient; and 2) both the exponent and the coefficient. Both recalibrations involved performing least squares linear regressions of the estimating equations on the actual data, but the second required an initial logarithmic transformation. Three statistical criteria were used to judge the accuracy of the estimating equations: 1) the standard error of the slope and the coefficient of determination ($r^2$) of the regression lines described above; 2) a paired t-test of the estimates vs. the actuals to assess whether the residuals are normally distributed with a mean of zero; and 3) a test of the hypothesis that the slope of the regression line of estimates vs. actuals is unity. All three criteria indicated that the recalibrated coefficient equations were the best matches to the actuals for all five estimating equations. The recalibrated equations are significantly different from the published Boehm equations and/or provide significantly better estimates to the ESD/MITRE data.

Productivity was found to be inversely proportional to the effort adjustment factor, which supports the general formulation of the Intermediate COCOMO model. Furthermore, it was found that the variance of the residuals was moderately decreased for the Nominal effort equations relative to the Basic effort equations. The Intermediate model thus provides an improved estimate over the Basic model. Further research is required to determine if the estimates could be further improved by adjusting the development effort multipliers, or by eliminating some of the cost driver attributes.
Recalibration of Basic and Nominal COCOMO Equations to Recent Air Force Acquisitions

Dr. Paul G. Funch

Presented at the Third Annual COCOMO User's Group Meeting
4-5 November 1987

MITRE
Introduction

• Objectives.
  - Unified, standard source of historical data on ESD-type acquisitions.
  - Recalibrate COCOMO equations to the ESD environment.

• Sources of data.
  - ESD/ACCR.
  - MITRE technical staff and management.
  - Contractors.

• Data verification.

MITRE
**Data Base Summary - I**

<table>
<thead>
<tr>
<th>Mode</th>
<th>Semidetached</th>
<th>Embedded</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of completed projects</td>
<td>3</td>
<td>18</td>
</tr>
<tr>
<td>Number of completed CSCIs</td>
<td>14</td>
<td>48</td>
</tr>
<tr>
<td>Number of recent subsystems*</td>
<td>11</td>
<td>17</td>
</tr>
<tr>
<td>Size range of subsystems</td>
<td>5 - 166 KDSI</td>
<td>6 - 264 KDSI</td>
</tr>
<tr>
<td>Effort range of subsystems</td>
<td>28 - 704 SM</td>
<td>56 - 6500 SM</td>
</tr>
<tr>
<td>Schedule of projects **</td>
<td>---</td>
<td>13 - 75 MOS.</td>
</tr>
<tr>
<td>Years of completion</td>
<td>1981 - 86</td>
<td>1981 - 86</td>
</tr>
</tbody>
</table>

* Subsystem is a group of (one or more) interdependent CSCIs.

**Only 12 projects provided schedule data.**

MITRE
Data Base Summary - II

- Applications represented in ESD/MITRE data base.
  - Command systems.
  - Communication systems.
  - Radar systems.
  - Simulations.
  - Training systems.

- Types of data collected.
  - Size (new, adapted, total equivalent DSI).
  - Effort, by phase.
  - Schedule, by milestones.
  - COCOMO, System-3, PRICE-S, and SLIM inputs.
Scope of Today's Talk

- Statistical methods.
- Example of techniques for recalibration and evaluation.
- Summary of five equation recalibrations:
  - Schedule, embedded mode: \( S = aE^b \).
  - Basic effort, semidetached and embedded modes: \( E = cI^d \).
  - Nominal effort, semidetached and embedded modes: \( (E/EAF) = c_nI^d \).
Statistical Evaluations of Recalibrations

- Results from linear regressions used in recalibration.
  - Coefficient only, and exponent plus coefficient.
  - Slope, 95% c.i. of slope, $r^2$.
- Paired t-test of estimates vs. actuals.
  - Standard deviation (sd) of residuals, Student t-statistic.
  - Test of $H_0$: residuals are normally distributed with mean of zero. Probability, $p$.
- Linear regression of estimates vs. actuals ($X_{est} = sX_{act}$).
  - Slope, 95% c.i. of slope, $r^2$.
  - Test of $H_1$: slope is unity.
Statistical Evaluations Summary

- For each of the five COCOMO estimating equations, the performance of the coefficient only recalibration was superior to that of either the Boehm equation or the exponent plus coefficient recalibration.
Boehm Equation Estimates vs. Actuals

Basic Effort, Embedded Mode

MITRE
Recalibrated Equation Estimates vs. Actuals

Basic Effort, Embedded Mode

MITRE
Actual Effort vs. Size

Basic Effort, Embedded Mode

MITRE
Statistical Results Summary

Basic Effort Estimating Equations, Embedded Mode

<table>
<thead>
<tr>
<th></th>
<th>Boehm Equation</th>
<th>Recalibrated Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Recalibration linear regression</strong></td>
<td>( E = 3.6(l)^{1.20} )</td>
<td>( E = 6.5(l)^{1.20} )</td>
</tr>
<tr>
<td><strong>95% c.i.</strong></td>
<td>1.09</td>
<td>0.88</td>
</tr>
<tr>
<td><strong>( r^2 )</strong></td>
<td>0.85</td>
<td>0.85</td>
</tr>
<tr>
<td><strong>df</strong></td>
<td>16</td>
<td>16</td>
</tr>
</tbody>
</table>

|                               | sd = 1065      | sd = 642               |
| **Paired t-tests**            | \( t = -1.50 \) | \( t = 1.02 \)         |
| (estimates vs. actuals)       | \( p = 0.15 \) | \( p = 0.32 \)         |
| **df** | 16             | 16                    |

|                               | Eest = 0.51(Eact) | Eest = 0.91(Eact) |
| **Linear regression**         | 95% c.i. = 0.09   | 95% c.i. = 0.15    |
| (Eest = s Eact)               | \( r^2 = 0.85 \)  | \( r^2 = 0.85 \)   |

MITRE
Productivity vs. 1/EAF

Productivity = \frac{TEDSI}{SM} \propto \frac{TEDSI}{EAF(TEDSI)^b} \approx \frac{1}{EAF}

MITRE
Recalibration Summary

- Basic effort
  - Embedded mode
  - Semidetached mode
- Nominal effort
  - Embedded mode
  - Semidetached mode
- Schedule, embedded mode

<table>
<thead>
<tr>
<th>ESD/MITRE</th>
<th>BOEHM</th>
</tr>
</thead>
<tbody>
<tr>
<td>E=6.5(I)^{1.20}</td>
<td>E=3.6(I)^{1.20}</td>
</tr>
<tr>
<td>E=2.4(I)^{1.12}</td>
<td>E=3.0(I)^{1.12}</td>
</tr>
<tr>
<td>E/EAF=3.3(I)^{1.20}</td>
<td>E/EAF=2.8(I)^{1.20}</td>
</tr>
<tr>
<td>E/EAF=3.1(I)^{1.12}</td>
<td>E/EAF=3.0(I)^{1.12}</td>
</tr>
<tr>
<td>S=3.8(E)^{0.32}</td>
<td>S=2.5(E)^{0.32}</td>
</tr>
</tbody>
</table>

MITRE
Conclusions

- New equations calibrated to ESD applications environment.

- Recalibration of coefficient only is superior to combined exponent/coefficient recalibration.

- Nominal equation estimates result in moderately decreased residual variance. Therefore Intermediate model provides improved estimates over Basic model.

- Further research required.
  - Evaluations of each development effort multiplier (DEM).
  - Additional project/subsystem data.