Adaptive Curve Fitting: An Algorithm in a Sea of Models

Michael L. Schiavoni

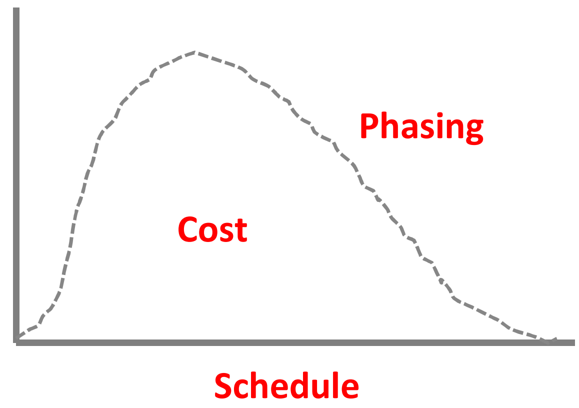
Presented at the ICEAA Professional Development & Training Workshop
May 14-17, 2019
Tampa, FL

All data appearing directly in this presentation has been simulated to mimic real observed programs. No proprietary data is shown.

TECOLOTE RESEARCH

In Short

- Adaptive Curve Fitting (ACF) is an automated procedure for analyzing time series data to predict cost, schedule, and phasing
  - Integrated approach ensures consistency between each of these three components
- In development since May 2017
  - Funded by SMC, AFCAA, and Tecolote
- Utilizes multiple new methods
  - Algorithm… NOT a model
  - Code written in R… NOT a widely deployable tool yet
- Current status: applying ACF to recent and ongoing estimates

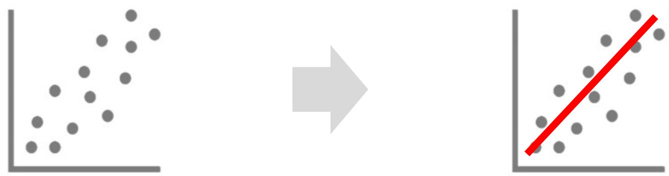


Benefits

- Crosschecking
- What-if scenarios
- Early detection of slips/overruns
- Inform budgeting

Models vs. Algorithms

Cost estimators typically work with **models**.



Historical data is collected

An equation that describes the data is applied to new observations

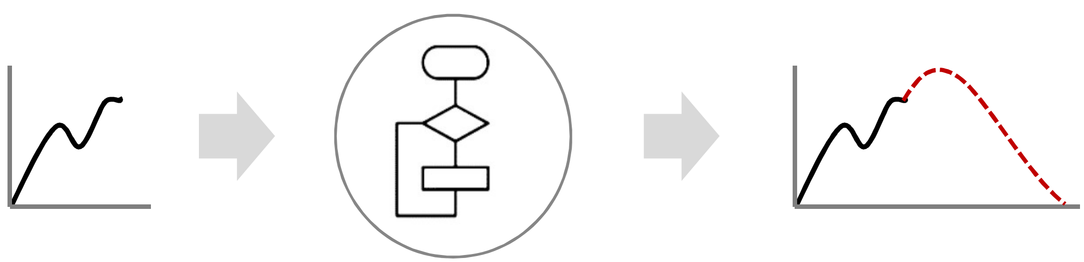
Things to note

Cost models utilize independent variables that are typically:

- technical or programmatic
- stable

They capture historical risk but don't adjust to the target system's performance.

Algorithms behave differently.



Input data

Defined sequence of calculations

Output

Algorithms are <u>rule-based</u>

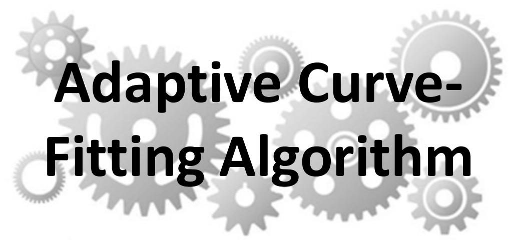
- Rules are typically based on historical data or SME insight
- Can be visualized as a flowchart

examples: Netflix recommendations, fraud detection, self-driving cars, Facebook newsfeed, spam filters, etc.

Adaptive Curve Fitting: What It Does

Inputs from in-progress program

- monthly expenditures (required)
- range for total cost (optional)
- range for total duration (optional)

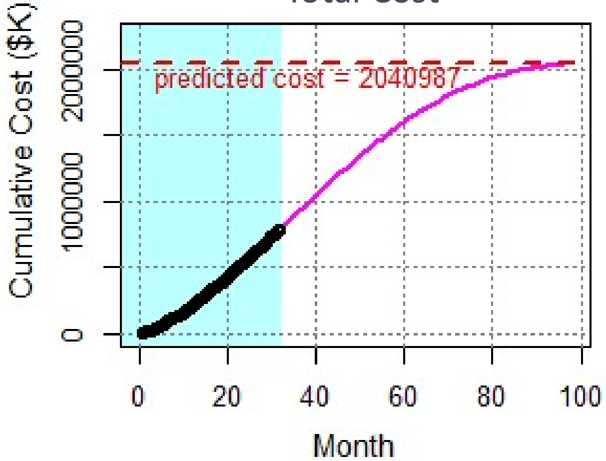


Employs a variety of mathematical techniques, including:

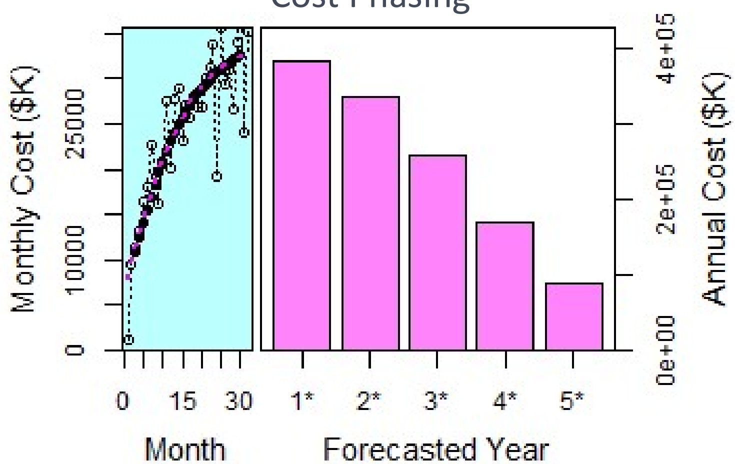
- Smoothing filters
- Calculus-based curve projection
- Nonlinear least-squares optimization
- Rayleigh/Weibull/Beta/Normal distributions

Outputs

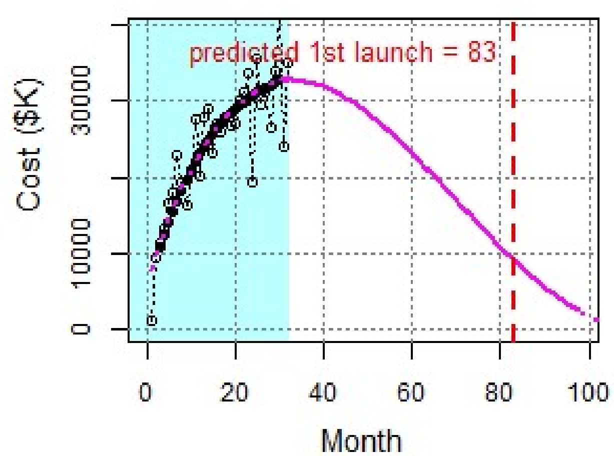
Total Cost



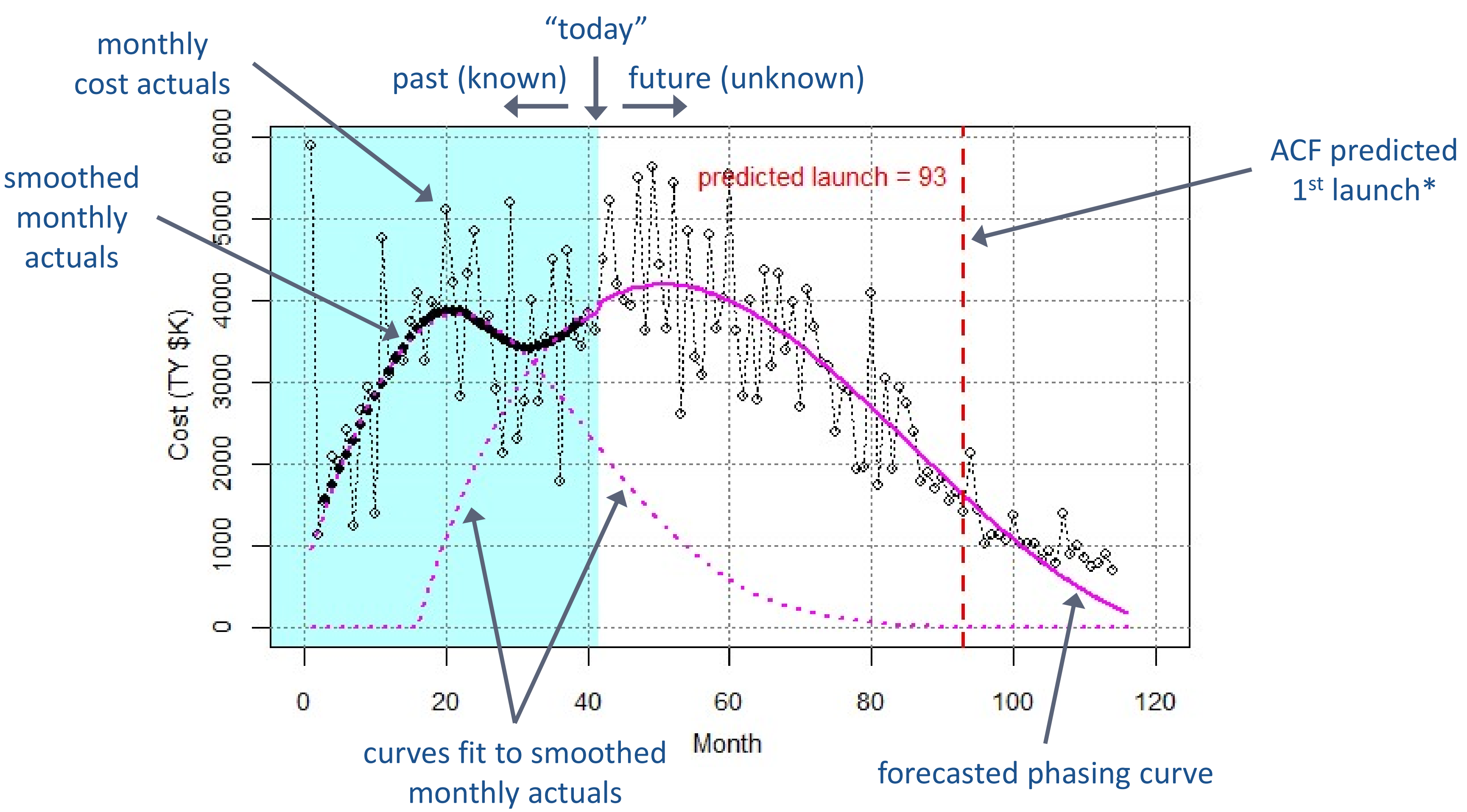
Cost Phasing



Duration to Launch

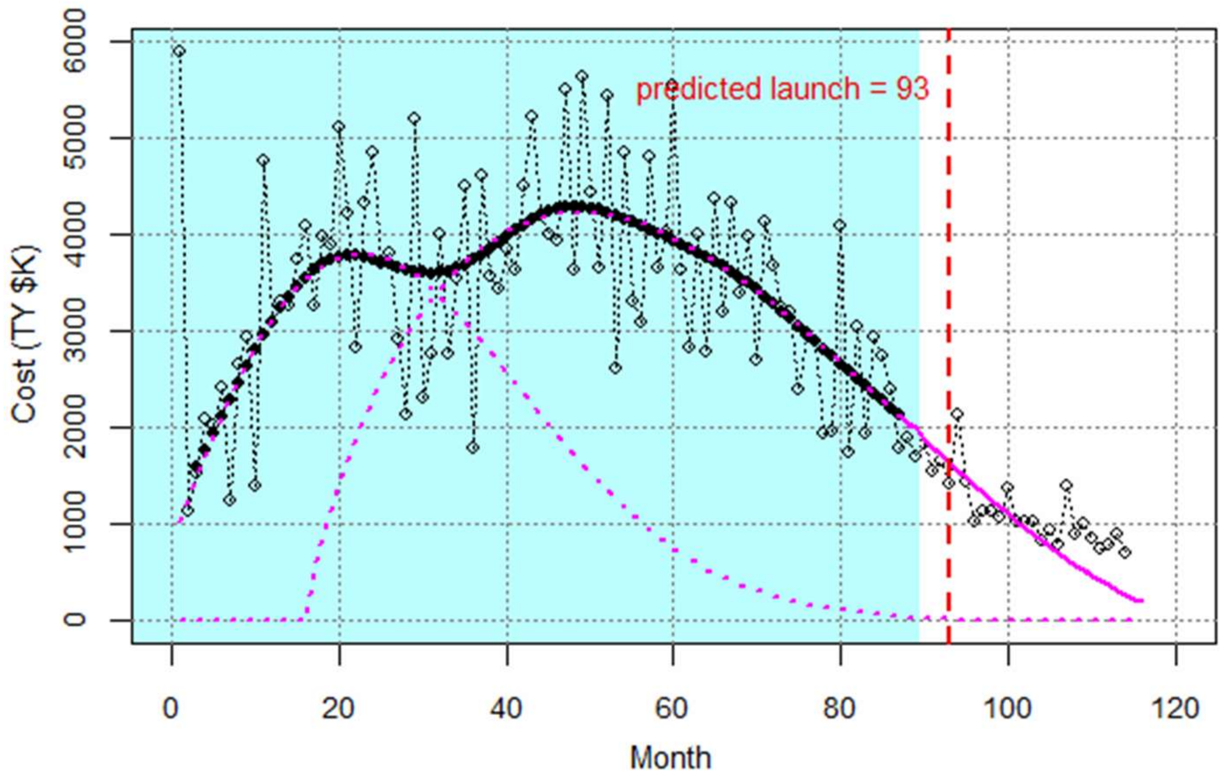


ACF Example



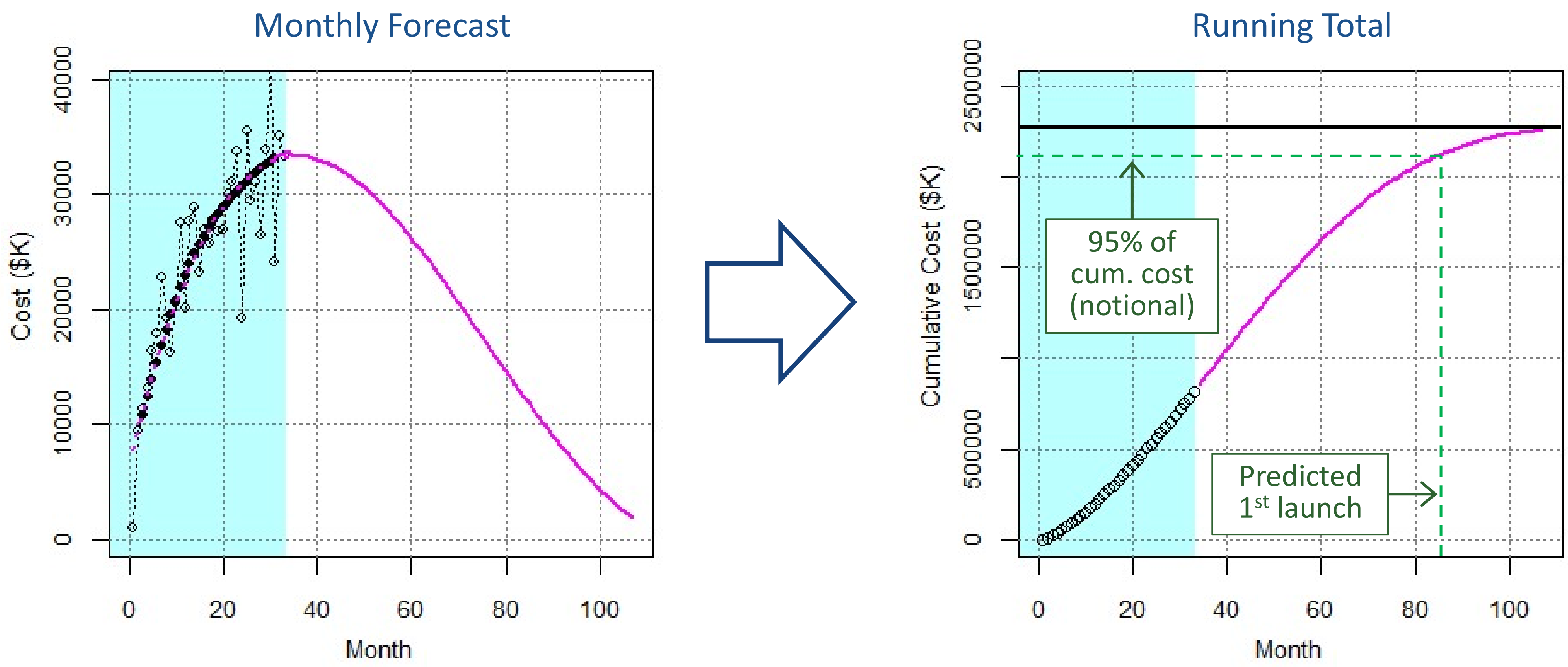
*Launch is predicted by determining the month along the forecasted phasing curve at which a %-spent metric is met.

Animated Example over Time



[view in slide show mode for animation]

Predicting Launch (or other major milestone)



ACF can predict duration to milestones by applying a %-spent metric to the forecasted cumulative curve. The metric can come from historical averages, contractor plan, SME input, etc.

Presented at the 2019 ICEAA Professional Development & Training Workshop - www.iceaaonline.com

Constraining the Forecast

- Constraining the forecast with limits on total cost or duration is an optional feature
- It is an iterative process
  - At each iteration, the forecasted portion of the model is stretched/shrunk horizontally and expanded/contracted vertically
- Examples of constraints:

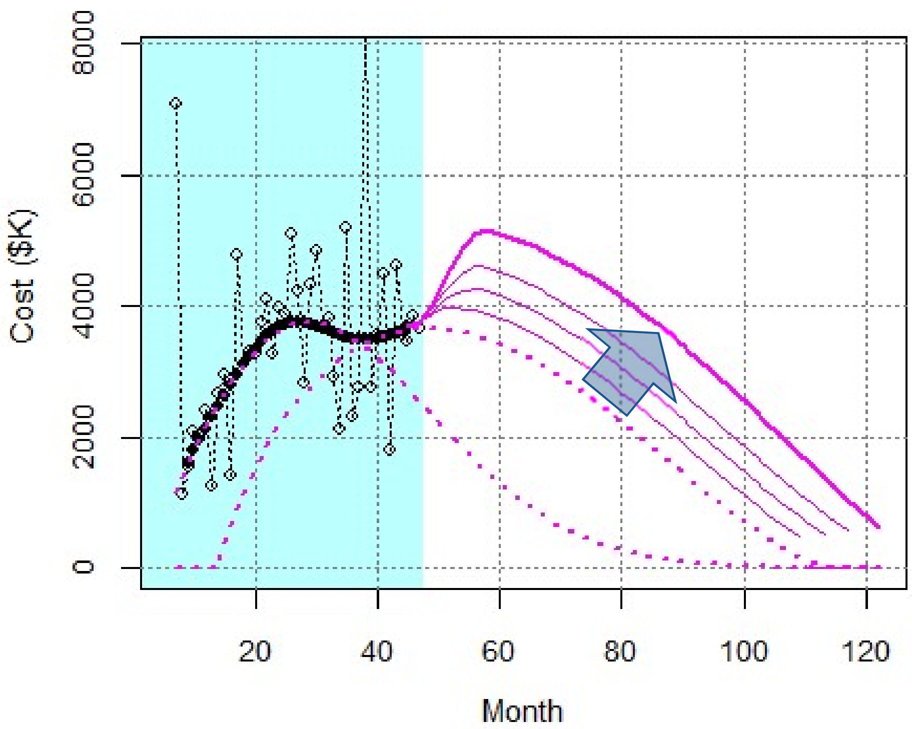
$$cost \geq EAC_{ktr}$$

$$(GEAC - 5\%) \leq cost \leq (GEAC + 10\%)$$

$$cost = SBE$$

$$duration \geq IMS$$

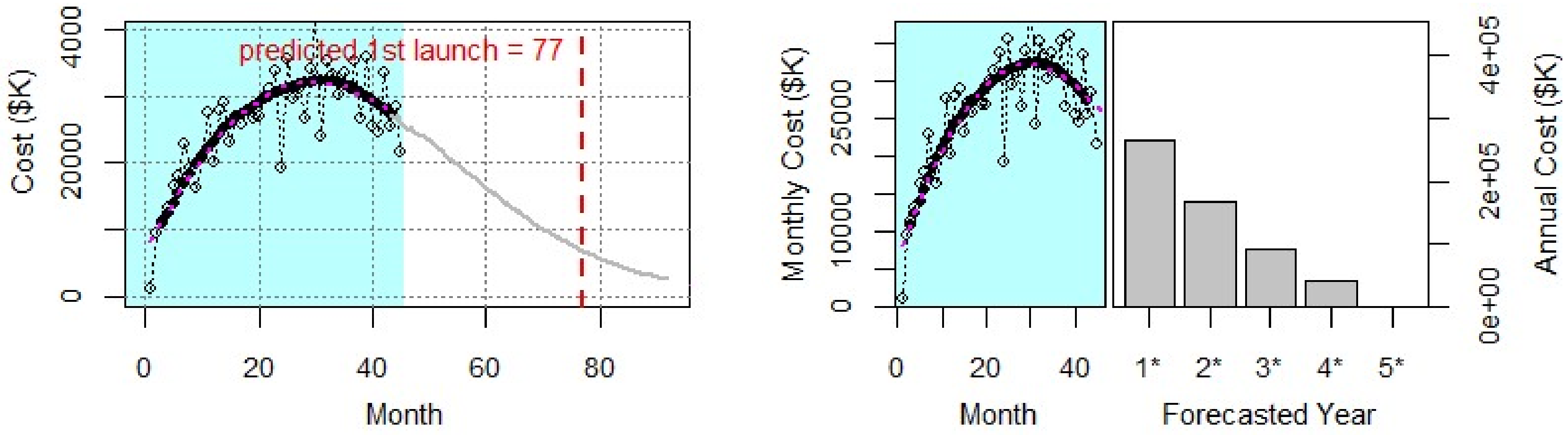
$$(SRA - 3\ mo) \leq duration \leq (SRA + 7\ mo)$$



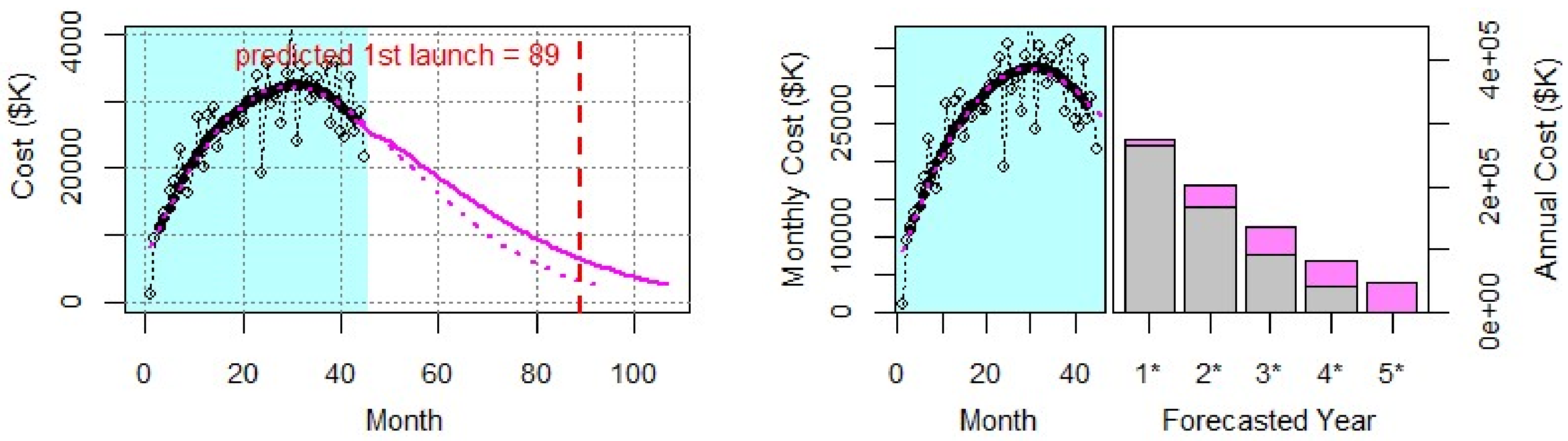
This feature enables rapid exploration of "What-if" scenarios.

"What if" Scenario 1: Schedule Delay

Current plan:

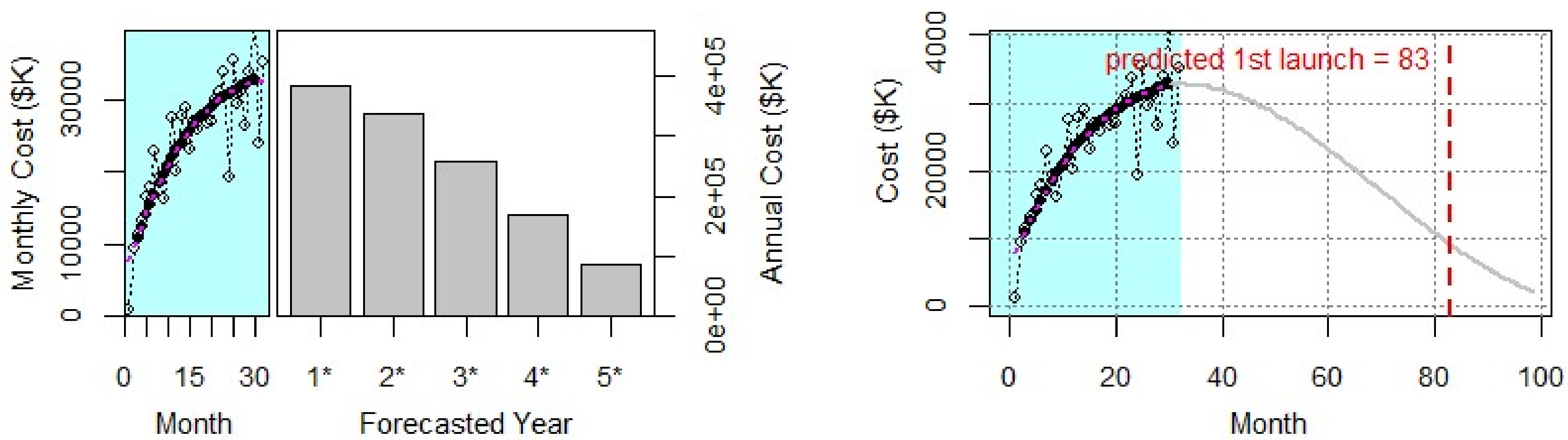


Likely effect of 12-month launch delay:

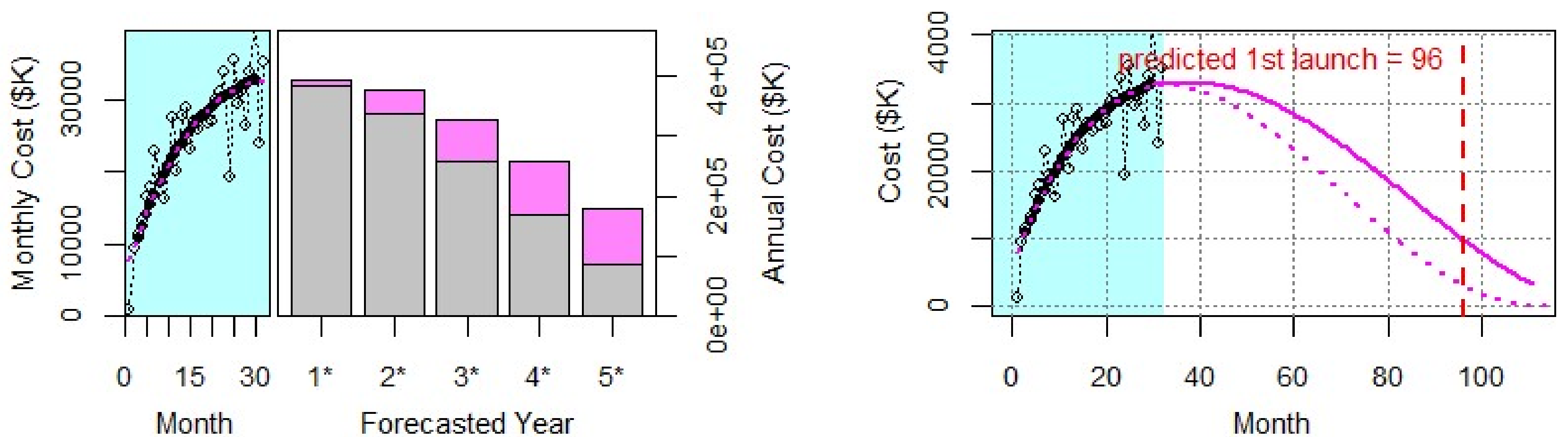


"What if" Scenario 2: Cost Overrun

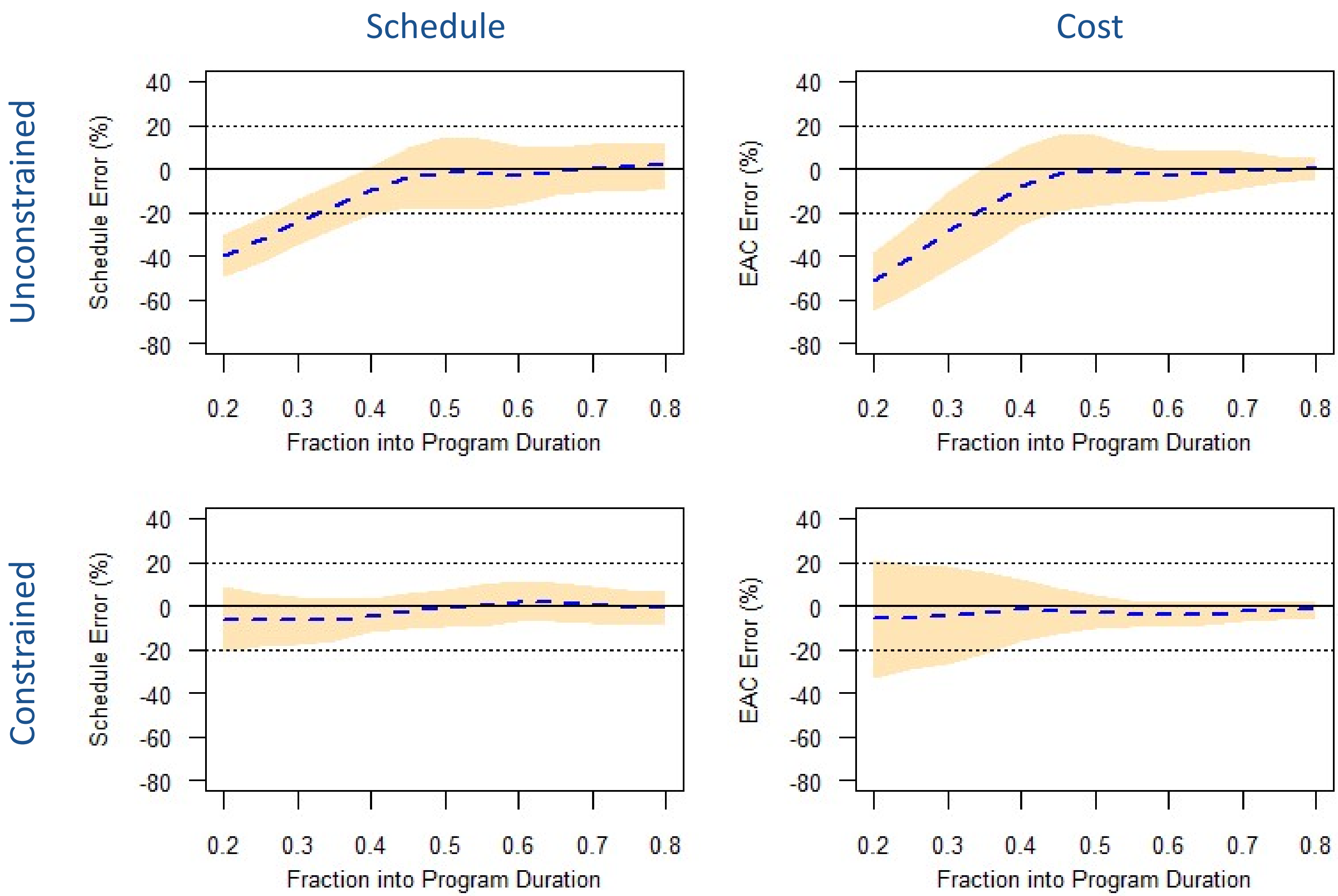
Current plan:



Likely effect of 20% cost overrun:



ACF has been validated using a sample of historical programs with known actuals



Sample

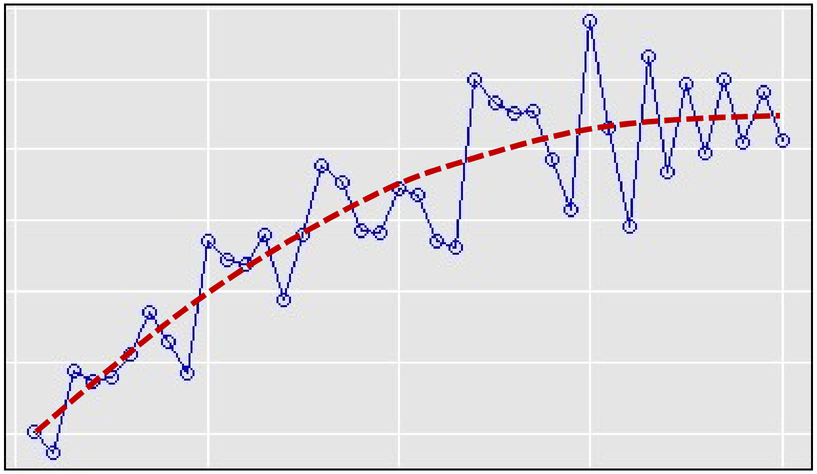
- Size: 20 programs
- Agencies: military & NASA
- Domains: space & ground
- Contract types: Dev. & follow-on
- Missions: sensing & communications

Presented at the 2019 ICEAA Professional Development & Training Workshop - www.iceaaonline.com

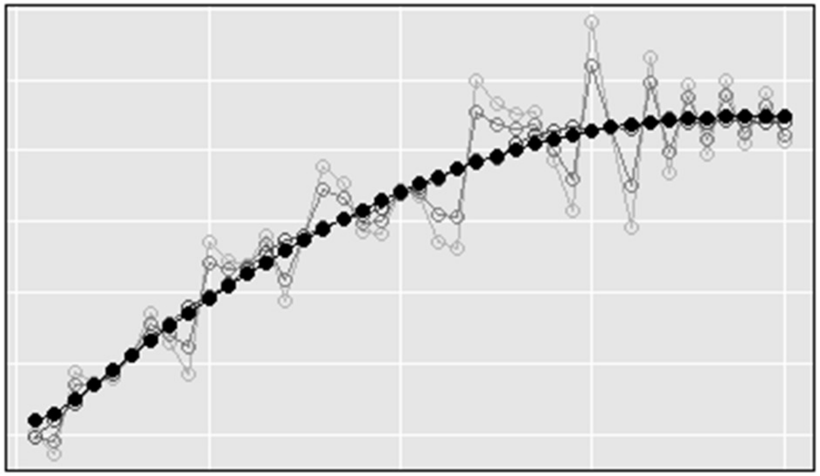
How it works

Smoothing

- Can you spot the trend?
  - Yes!



- Can the trend be extracted in an automated and objective fashion?
  - Yes! – by filtering out the noise
  - ACF utilizes a multi-step smoothing procedure, including <u>lo</u>cal regr<u>ess</u>ion (LOESS) and an iterative moving average



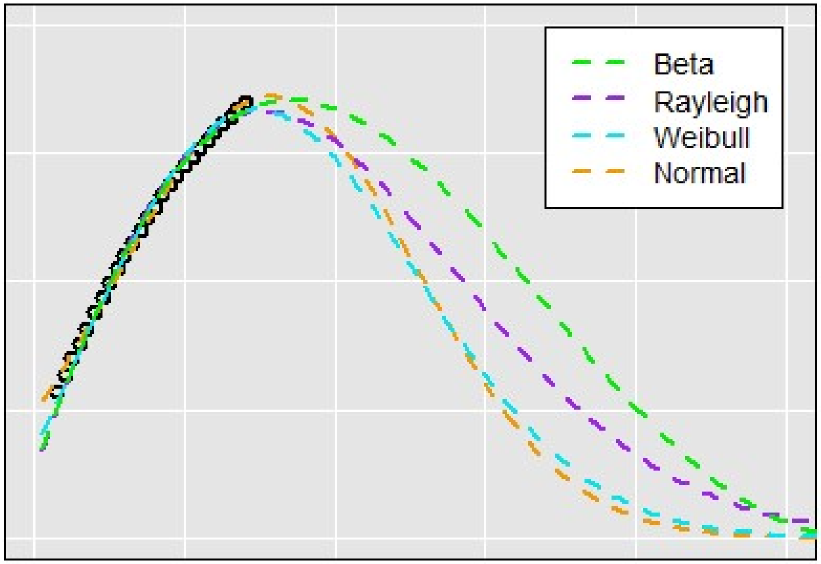
[view in slide show mode for animation]

Curve Fitting

- Can this data sequence be described?
  - Yes!
- A straight line could be fit
  - Poor fit; unlikely to extrapolate well
- A quadratic or higher-order polynomial could be fit
  - Better fit; still unlikely to extrapolate well
- Alternatively, we can fit known resource phasing forms
  - Empirical and theoretical foundation
  - Best fit; highest chance of extrapolating accurately
  - ACF fits the following forms:
    - Rayleigh $y = (x/\lambda^2)\exp\left(-x^2 / (2\lambda^2)\right)$
    - Weibull $y = (k/\lambda)(x/\lambda)^{k-1}\exp\left(-(x / \lambda)^k\right)$
    - Beta $y = cx^{\alpha-1}(1-x)^{\beta-1}$
    - Normal $y = \left(1/\sigma\sqrt{2\pi}\right)\exp\left(-(x-\mu)^2 / (2\sigma^2)\right)$

[view in slide show mode for animation]

$y = a + bx$ $y = a + bx + cx^2$



Essentially, ACF runs four nonlinear least-squares regressions for each curve segment. Then the one with minimum SSE is selected to forecast the future.

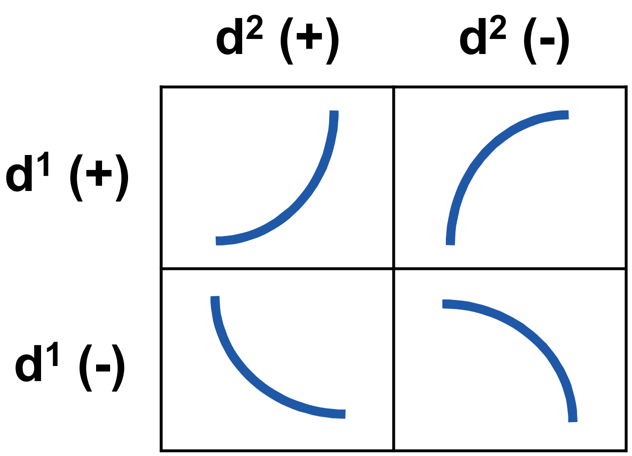
Why does ACF use the curve forms that it does?

- Project management theory and empirical research indicates that these are nominal resource phasing curves
  - P.V. Norden, *Useful Tools for Project Management*, Management of Production (1970)
  - L.H. Putnam, *A General Empirical Solution to the Macro Software Sizing and Estimating Problem*, IEEE Transactions on Software Engineering (1978)
  - H. Watkins, *An Application of Rayleigh Curve Theory to Contract Cost Estimation and Control*, NPS Thesis (1982)
  - T. Abernethy, *An Application of the Rayleigh Distribution to Contract Cost Data*, NPS Thesis, 1984
  - D.A. Lee, *Time Histories of Expenditures for Defense Acquisition Programs in the Development Phase*, ISPA (1993)
  - M. Gallagher, *Final-Cost Estimates for R&D Programs Conditioned on Realized Costs*, OSD Report (1995)
  - J. Dukovich, *The Rayleigh Analyzer Volume I – Theory and Applications*, LMI report prepared for DoD (1999)
  - E.J. Unger, *Relating Initial Budget to Program Growth with Rayleigh and Weibull Models*, AFIT Thesis (2001)
  - H.F. Chelson, *Rayleigh Curves – A Tutorial*, SCEA-ISPA Conference (2004)
  - E.L. Burgess, *R&D Budget Profiles and Metrics*, Journal of Parametrics (2006)
  - D. Davis, *Using the Rayleigh Model to Assess Future Acquisition Contract Performance and Overall Contract Risk*, CAN report prepared for the Department of the Navy (2009)
  - A.R. Jones, *Project Team Sizing and Cost Forecasting using Norden-Rayleigh Curves*, ACostE Conf. (2011)
  - A. Sokri, *Weibull-based Time-phasing of Budget Expenditures*, Defence R&D Canada (2012)
  - E.L. Burgess, *Weibull Analysis Method*, ICEAA Workshop (2014)
  - G.E. Brown, *Time Phasing Aircraft R&D Using the Weibull and Beta Distributions*, JCAP (2015)

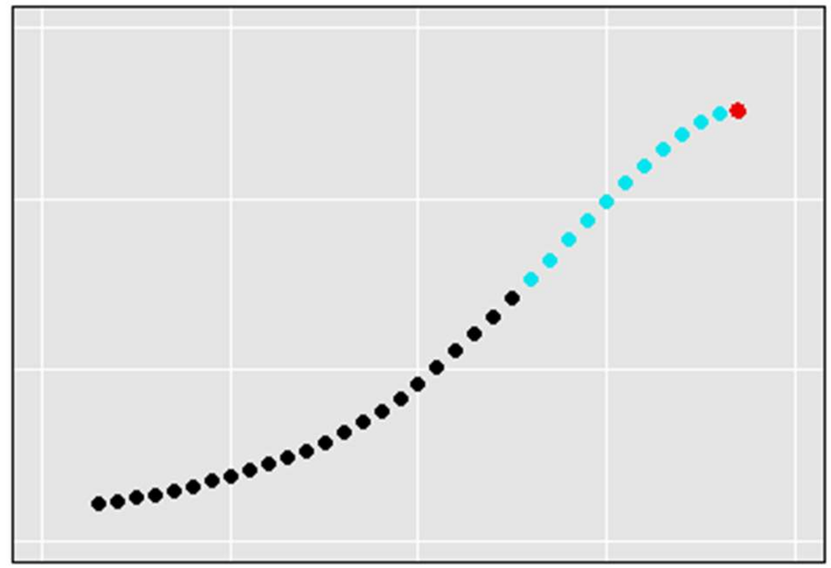
**** This list is not exhaustive! ****

Curve Projection

- Can we forecast when this curve will peak?
  - Yes!
- Recall from calculus:
  - 1st derivative (d1) = rate of change of curve (a.k.a. slope or velocity)
  - 2nd derivative (d2) = rate of change of d1 (a.k.a. concavity or acceleration)
  - 3rd derivative (d3) = rate of change of d2



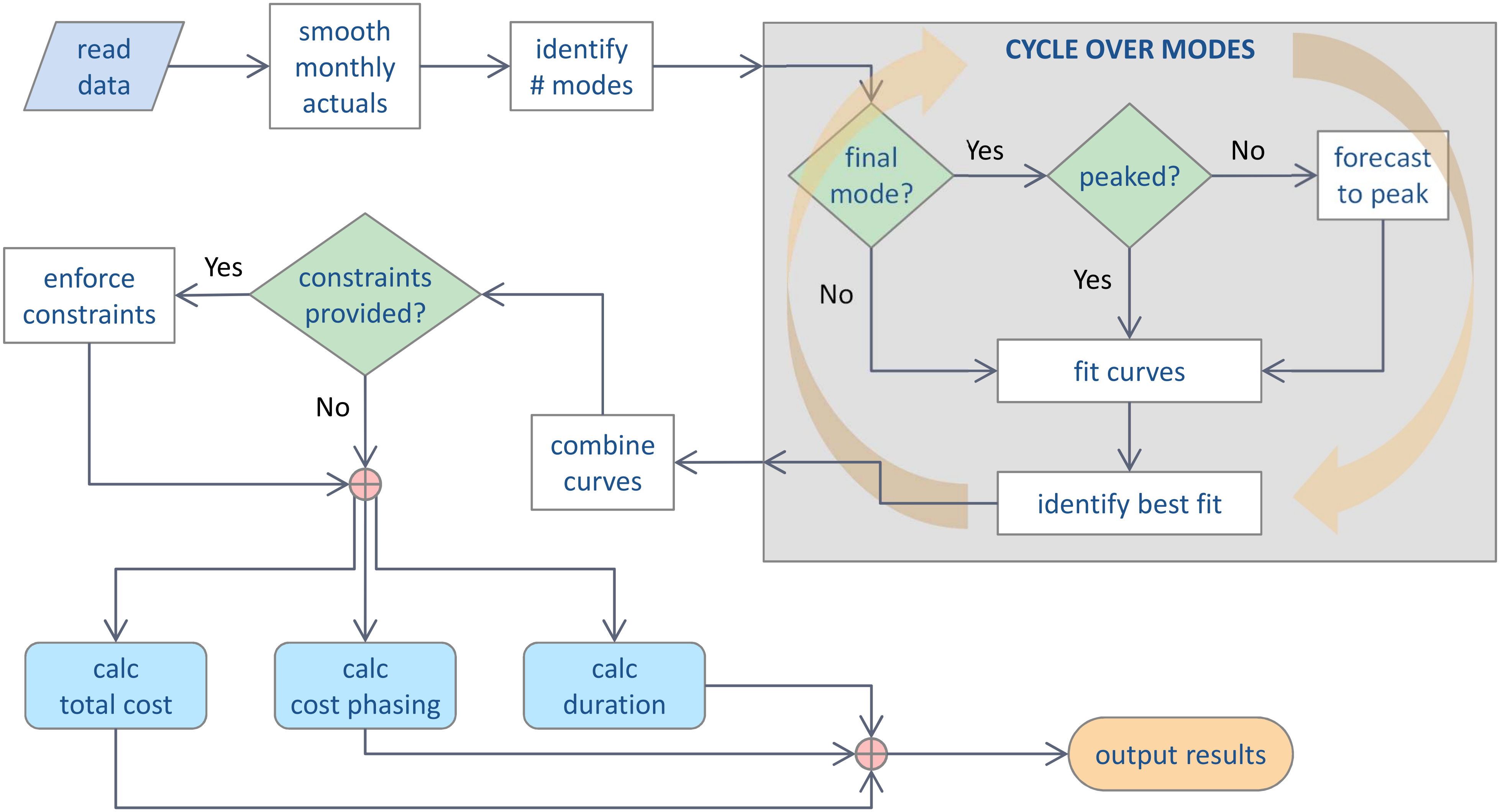
[view in slide show mode for animation]



```
repeat{  use d³ to update d²
         use d² to update d¹
         use d¹ to update value }
```

It's similar to a projectile motion problem, in that the future path is predicted based on the most recent known trajectory.

ACF Algorithm



¡Important Point!

- Nothing about this technique is specific to expenditures data, or any one commodity
  - It can theoretically be applied to <u>any</u> time series data with these two characteristics:
    - is finite, i.e. has an ending point (so not applicable to a stock market index)
    - can be modeled as one or more known probability density functions
- Other potential data streams:
  - Earned value (BCWP)
  - Labor hours/heads
  - Software effort/ESLOC/DRs
  - # of concurrent schedule tasks
  - ... etc.
- Probably <u>not</u> applicable to:
  - Sustainment contracts and other constant level of effort tasks
  - Production contracts with many units in assembly line process
  - Agile software development and other rolling wave efforts

Conclusion

- ACF intelligently fits known resource phasing curve forms to monthly time series data in order to forecast the future
  - Adjusts to current program performance
  - Objective and repeatable
  - Based in theory and empirical research
  - Validated on historical programs with known actuals
- Applications:
  - Analyst: crosscheck existing estimates, rapidly explore excursions
  - Program Manager: early detection of cost overruns, schedule slips
  - Organization: inform budgeting decisions across programs within a portfolio

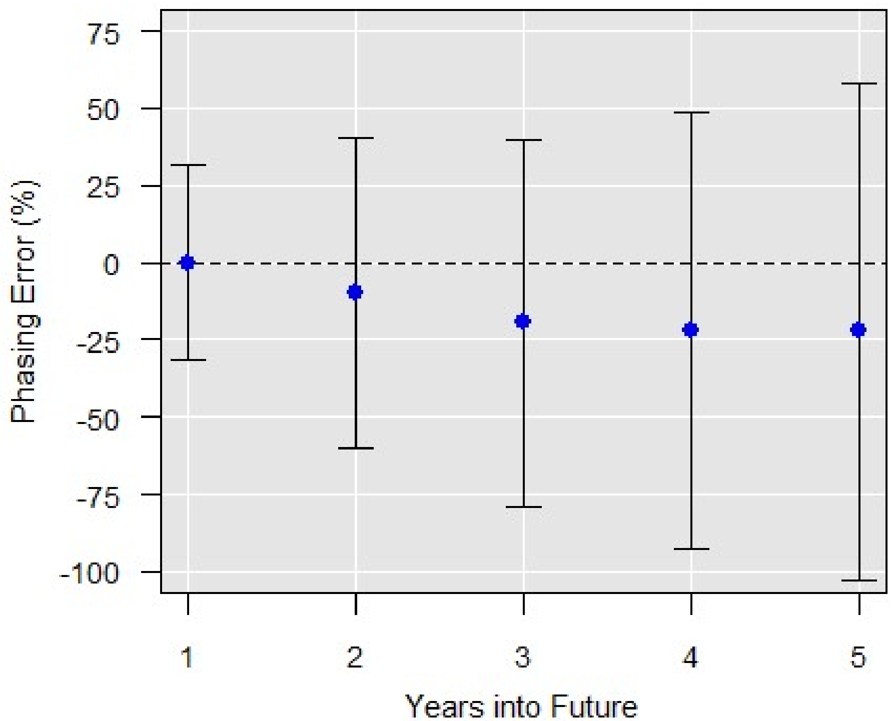
As the problems we face continue to grow in complexity & difficulty, we expect to increasingly rely on non-traditional techniques including algorithms, semi- & non-parametric modeling, and other machine learning methods.

Presented at the 2019 ICEAA Professional Development & Training Workshop - www.iceaaonline.com

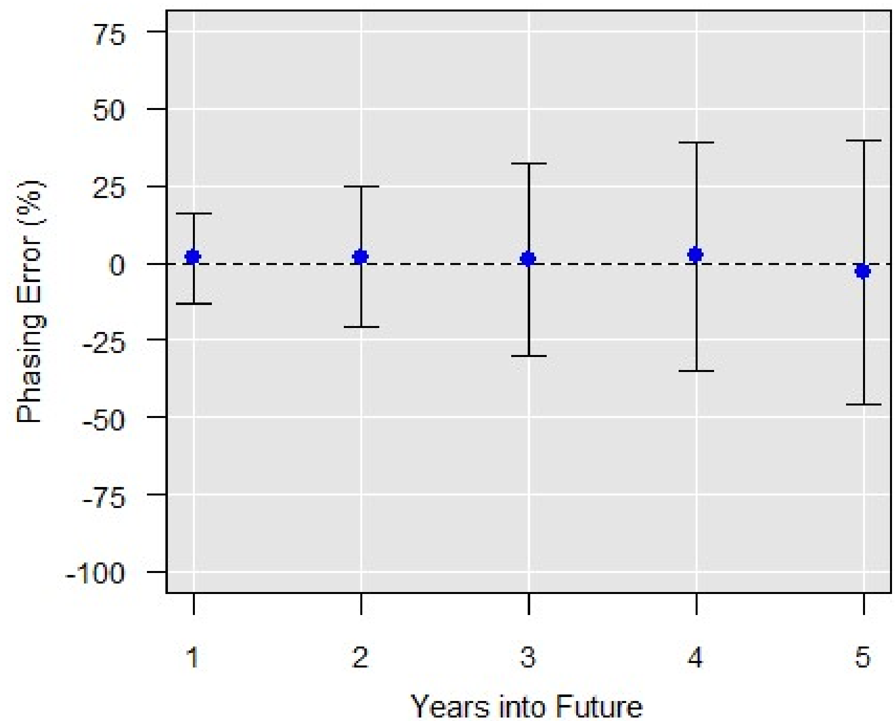
Contact:
mschiavoni@tecolote.com

Phasing Prediction Errors (mean +/– 1σ)

Unconstrained



Constrained to True Cost and Duration



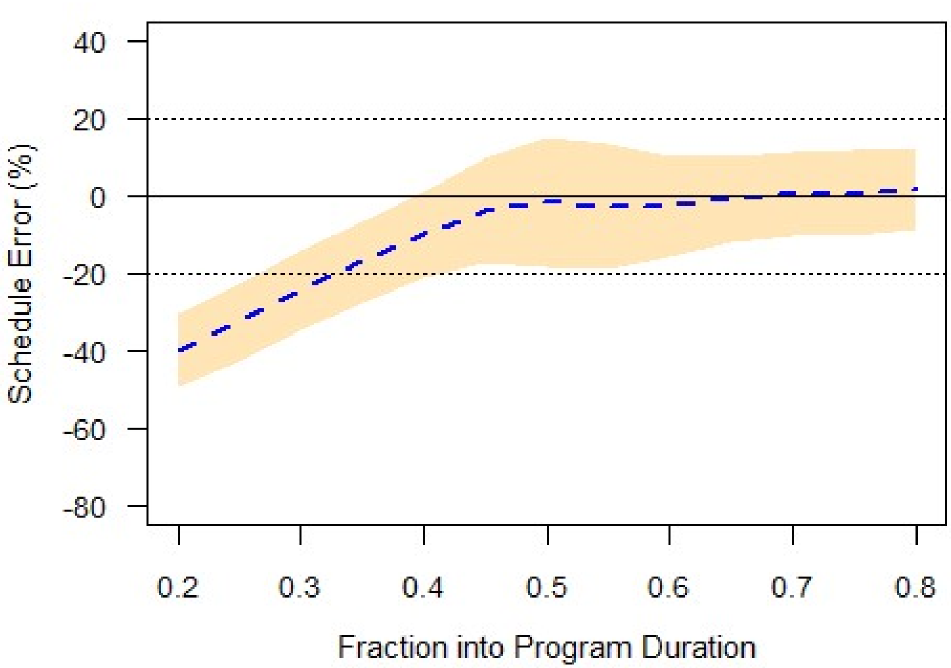
Shows how accurate ACF is at forecasting phasing when no information is available about total cost or duration.

Shows how accurate ACF is at generating a phasing profile that is consistent with a given cost and schedule estimate.

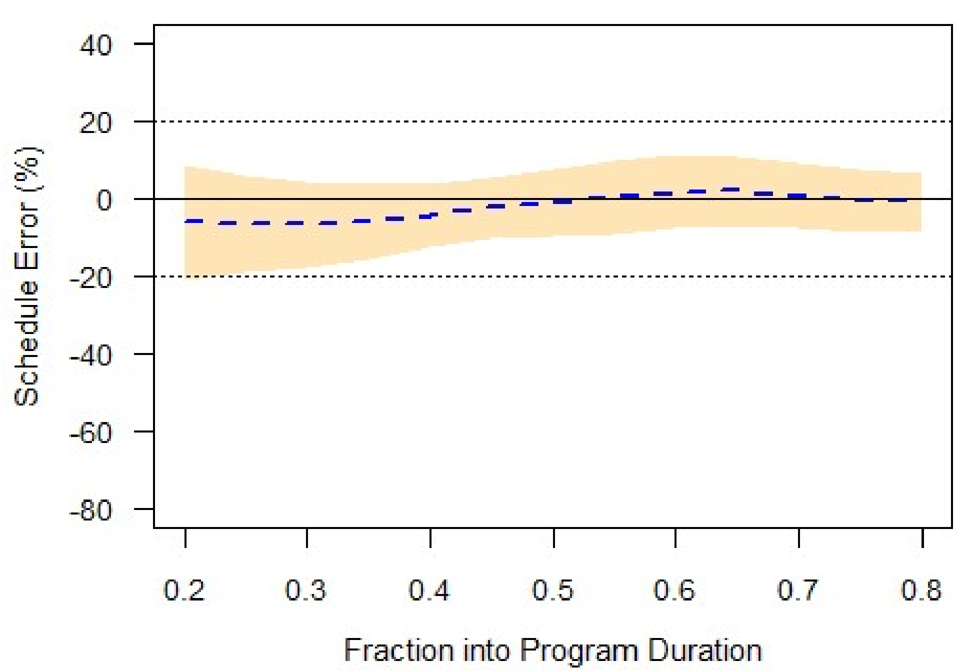
** Real world performance should be between these extremes **

Schedule Prediction Errors (mean +/– 1σ)

Unconstrained



Constrained to True Cost



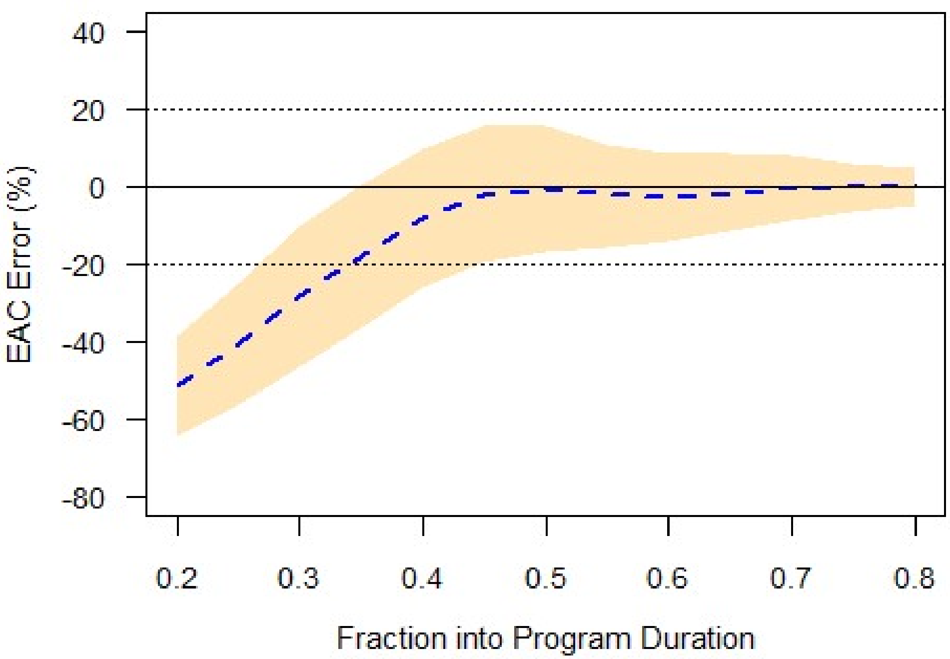
Shows how accurate ACF is at forecasting duration to launch when no information is available about total cost.

Shows how accurate ACF is at generating a schedule estimate that is consistent with a given cost estimate.

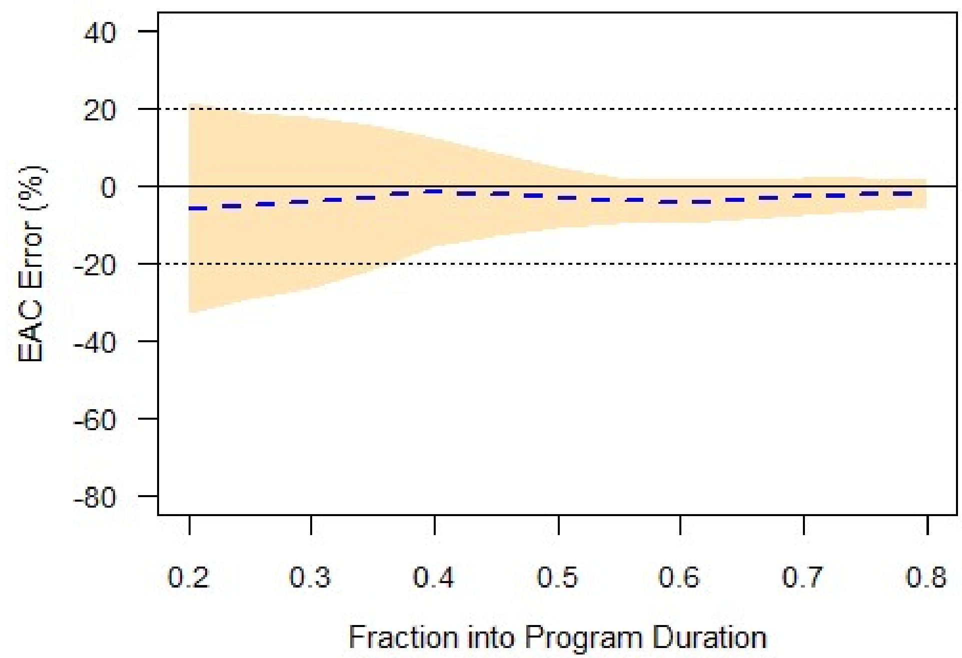
** Real world performance should be between these extremes **

Cost Prediction Errors (mean +/– 1σ)

Unconstrained



Constrained to True Duration



Shows how accurate ACF is at forecasting total cost when no information is available about duration.

Shows how accurate ACF is at generating a cost estimate that is consistent with a given duration estimate.

** Real world performance should be between these extremes **