Earned Value Management Meets Big Data

Glen B. Alleman
Niwot Ridge, LLC
glen.alleman@niwotridge.com
+1 303 241 9633

Thomas J. Coonce
Thomas J. Coonce, LLC
tom.coonce@verizon.net
+1 703 362 2568

ABSTRACT

The Earned Value Management System (EVMS) maintains period–by–period data in its underlying databases. The contents of the Earned Value repository can be considered BIG DATA, characterized by three attributes – 1) Volume: Large amounts of data; 2) Variety: data comes from different sources, including traditional data bases, documents, and complex records; 3) Velocity: the content is continually being updated by absorbing other data collections, through previously archived data, and through streamed data from external sources. [1]

With this *time series* information in the repository, analysis of trends, cost and schedule forecasts, and confidence levels of these performance estimates can be calculated using statistical analysis techniques enabled by the Autoregressive Integrated Moving Average (ARIMA) algorithm provided by the **R** programming system. ARIMA provides a statistically informed Estimate At Completion (EAC) and Estimate to Complete (ETC) to the program in ways not available using standard EVM calculations. Using ARIMA reveals underlying trends not available through standard EVM reporting calculations.

With ARIMA in place and additional data from risk, technical performance and the Work Breakdown Structure, Principal Component Analysis can be used to identify the *drivers* of unanticipated EAC.

INTRODUCTION

This is not a paper about Earned Value Management per se. It is not about the details of statistical forecasting using tools. It is not about the underlying mathematics of how these forecasts are made or the mathematics of how the "principal components" that drive the Estimate at Completion are applied.

This paper is about how to overcome the limitations of the traditional methods to identify and predict problems and Estimates at Completion. The literature is replete with nature of cost growth and how our "standard practices" of using EVM data fail to enlighten stakeholders in time to make corrective action. [RAND12], [RAND11], [RAND13], [IDA11], [IDA11a], and [IDA10]. Some of these reasons:

- The Earned Value data is not based on the statistical behavior of the underlying work being performed.
- The assessment of Budgeted Cost of Work Performed (BCWP) is not adjusted for the technical performance of the resulting work. EAI–748–C states, *Earned Value is a direct measurement of the quantity of work accomplished. The quality and technical content of worked performed is controlled by other processes.* [2]
- The calculation of the EAC using the standard formula does not consider future risk, compliance with quality or technical content, and most important, the statistical nature of future performance based on the statistical nature of past performance.
- The use of cumulative measures for SPI and CPI wipes out the variances of the past, preventing the use of this information for forecasting future performance in any statistically credible manner.

[1] There are many definitions of BIG DATA, for the purposes of our work, we'll use these. This definition is derived from *The Four V's of Big Data*, Volume, Velocity, Variety, and Veracity. We've dropped the Veracity, since the IPMR defines the submission formats for Earned Value data in the DOD. http://goo.gl/ObXYTO

[2] TechAmerica, Earned Value Management Systems, EIA–748–C, §3.8, page 19.

With this background, we present two possible solutions, making use of existing data repositories, to increase confidence in the EAC using statistical processes. These processes include applying the ARIMA to the past Earned Value Management data submitted on a monthly basis and PCA to identify the drivers of unanticipated growth to the EAC.

The ARIMA process hs been applied in multiple settings for some time – it is not new. [3] What is needed is to apply this method to the repository for the existing data to produce a credible EAC. The PCA efforts will require additional data not currently in most repositories, including technical performance measures, risk reduction, and other assessment of other *…illities*" called out in the Systems Engineering Management Plan (SEMP) of any large, complex, high risk program.

DATA ANALYTICS

There are three types of data analytics that can be applied to data held in an Earned Value Management repository.

- **Descriptive** – looking in the past we can learn what happened, but it is too late to take corrective action. EV descriptive analytics,
  - Condenses big data into smaller, useful nuggets of information.
  - Most raw Earned Value data is not suitable for human consumption since WBS reports it without the connectivity to the product or programmatic topology.
  - The data summarizes what happened in the past, many times 45 days in the past.
  - No correlations between WBS elements nor correlations between risk, technical performance or Systems Engineering attributes – MOE, MOP, KPP [4].
- **Predictive** – using past performance we can answer the question what will happen if we do the same as we've done in the past.
  - This is the role played by the current EAC formula
- **Prescriptive** – past performance data used to make predictions and suggest decision options to take advantage of the predictions. EV data prescriptive analytics,
  - Can be used when we need to prescribe an action so leadership can take the data and act.
  - Where predictive analytics doesn't predict one future outcome – but multiple outcomes based on the decision maker's actions.
  - Requires a predictive model with two additional components:
    - Actionable data.
    - Feedback system that tracks the outcome produced by the action taken.

It is the prescriptive analytics we are after. Prescriptive analytics not only anticipates what will happen and when it will happen, but why it will happen. The current descriptive analytics of the Earned Value Management reporting process does not point to possible sources of problems, other than manually looking through each reported data element. This is a major complaint of Program Managers.

Nice EV data report you got there, but you did not tell me what to do about the problems it shows.

[3] *Probability Methods for Cost Uncertainty Analysis: A Systems Engineering Perspective*, CRC Press, 2000. This book is a must read on the shelf of any cost, schedule, or technical performance manager.

[4] The *Defense Acquisition Guide* defines how to define and apply Measures of Effectiveness, Measures of Performance, Technical Performance Measures, and Key Performance Parameters to assess program performance.

Earned Value Management in a Nutshell

Government programs larger than $20M make use of an Earned Value Management System defined in EAI–748–C. For Department of Defense, Department of Energy, and NASA programs a monthly submittal (DI–MGMT–81865 and NASA 533M) containing information about the Earned Value performance of the program. This information includes the Budgeted Cost of Work Scheduled (BCWS), Budgeted Cost of Work Performed (BCWP), and Actual Cost of Work Performed (ACWP). With these three data items a Cost Performance Index (CPI) and Schedule Performance Index (SPI) can be computed using familiar formulas:

$$CPI = BCWP / ACWP \tag{1}$$

$$SPI = BCWP / BCWS \tag{2}$$

These formulas represent the *performance* of the program in units of dollars. They describe the efficacy of the money budgeted (BCWS), the money spent (ACWP), and the monetary value of the money *Earned* (BCWP). The CPI and SPI can then be used to forecast the Estimate At Completion for the remaining work in the program using,

$$EAC_{CPI} = ACWP_{cum} + \left[\left(BAC - BCWP_{cum} \right) / CPI_{cum} \right] \tag{3}$$

Equation (3) is one of several available for calculating *EAC* from the Earned Value Data.

Drivers of Program Variance

We all know programs operate with wide ranges of variance. From tabloid reports to the RAND and IDA Nunn McCurdy Breach Root Cause Analyses, the drivers of these variances are well known at the macro level shown in **Figure 1**. [5]

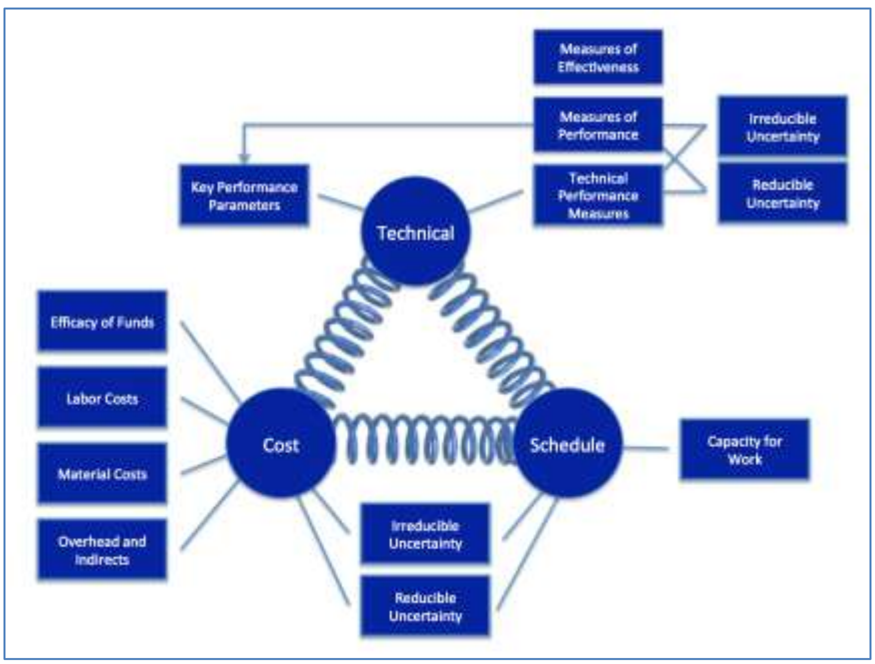

Figure 1 – all project and programs are driven by three interconnected random processes – Cost, Schedule, and Technical Performance. These three variables behave as stochastic processes with random couplings, themselves driven by other random variables. Each second tier driver may be coupled to another second tier driver to form a network of stochastic processes.

[5] RAND and IDA Nunn McCurdy reports can be found at www.rand.org and www.ida.org

The First Step Toward Better Estimate At Completion

Using the data in a repository, the first step is to treat each monthly report of CPI and SPI for each Work Breakdown Structure element (Format 1 of the IPMR) as a data sample in a time series. This time series of past performance is the source of ARIMA forecasting of EAC. With the time series and the ARIMA, the analyst can produce "forecasts" of the possible values of the EAC based on past statistical behaviors of CPI and SPI at the lowest reporting levels – the lowest submitted SPI and CPI per WBS element. Many times these WBS elements go to the physical components being delivered by the program. Insight into the forecasts of Configuration Items or End Item Deliverables are available in the current repository.

In this section, we'll describe how the ARIMA works and how it can be applied to the data in the repository to produce an EAC. Use of ARIMA will allow us to answer the question: "What is this thing going to cost us when we are done? And ARIMA will do this to a known confidence level that is statistically sound. We need to remind ourselves that the current formulas for calculating EAC use linear, non-statistical, non-risked adjust arithmetic on the cumulative values of SPI and CI that have had their past variances "whipped out." So not visibility into the past statistical excursion is available for to computed one of the most critical numbers on any program.

Autoregressive Integrated Moving Average (ARIMA) in a Nutshell

Autoregressive Integrated Moving Average (ARIMA) models are a class of forecasting models using time series data to better understand the contents of this data or predict future data points in the time series. ARIMA models are also referred to as Box–Jenkins [BOX70]. Let us look at how ARIMA works and confirm it can be used to forecast he Estimate at Completion using the repository data from he elements of the program, including CPI and SPI at the lowest Work Breakdown Structure available.

ARIMA models have three parts:

1. ***Autoregressive*** – assumes the time series is an observed value that depends on some linear combination of previous observed values up to a maximum defined lag (denoted by *p*), plus a random error term ε.
2. ***Integrated*** – the integrated part is applicable where the data shows evidence of non–stationarity, and the initial differencing step, the *I*, can be applied to remove the non–stationarity aspect of the time series. The *I* portion is denoted by *d*. A stationary process is a stochastic process whose joint probability distribution does not change when shifted in time. A non-stationary stochastic process will have a time-varying variance or time-varying mean or both.
3. ***Moving Average*** – assumes the observed value is a random variable plus some linear combination of previous random error terms up to a defined maximum lag (denoted by *q*).

A core assumption of the Autoregressive part of ARIMA is that the time series observations are independently identifiable. Autocorrelation refers to the correlation of a time series with its own past and future values. Autocorrelation is also sometimes called "lagged correlation" or "serial correlation", which refers to the correlation between members of a series of numbers arranged in time. Positive autocorrelation might be considered a specific form of "persistence", a tendency for a system to remain in the same state from one observation to the next. Autocorrelation can be exploited for predictions: an autocorrelated time series is predictable, probabilistically, because future values depend on current and past values. This means there is no autocorrelation in the series and the series will have a zero mean (after normalization). For this to be met, all the trending and seasonal components must be removed so we are left with only the noise. The result is only the irregular components of the time series are modeled, not the seasonal or trend components.

To summarize the ARIMA parameters in preparation for the use:

ARIMA(*p*,*d*,*q*) where

- *p* is the number of autoregressive terms,
- *d* is the number of nonseasonal differences,
- *q* is the number of lagged forecast errors in the prediction equation.

The model used in forecasting EAC starts with ARIMA(0,1,1) – simple exponential smoothing, where,

$$\hat{Y}(t) = Y(t-1) - qe(t-1) \tag{4}$$

where $e(t-1)$ denotes the error at period $t-1$. We can add a constant to the ARIMA(0,1,1) forecast with,

$$\hat{Y}(t) = m + Y(t-1) - qe(t-1) \tag{5}$$

ARIMA Used to Forecast EAC In This Paper

With the ARIMA algorithm and the time series of Earned Value data from a repository, we can construct forecasts of the EAC based on statistical data from each period of performance, rather than the cumulative data and the current period of performance as reported in the IPMR. Our first serious problem is how to select the ARIMA parameters. It is beyond this short paper to delve into this problem, so we will take a short cut using our R tool [6] and apply `auto.arima` and have the tool figure out which is best for our time series.

We will skip all the background and theory and go straight to the outcomes.

```
R> CPITS=ts(CPI)                    #Turn the raw data into a time series
R> CPIFIT=auto.arima(CPITS)         #apply auto fitting of the time series CPI data
R> CPIFCAST=forecast(CPIFIT)        #make a forecast of the future values of CPI
R> plot(CPIFCAST)                   #plot the CPI data and the forecast
```

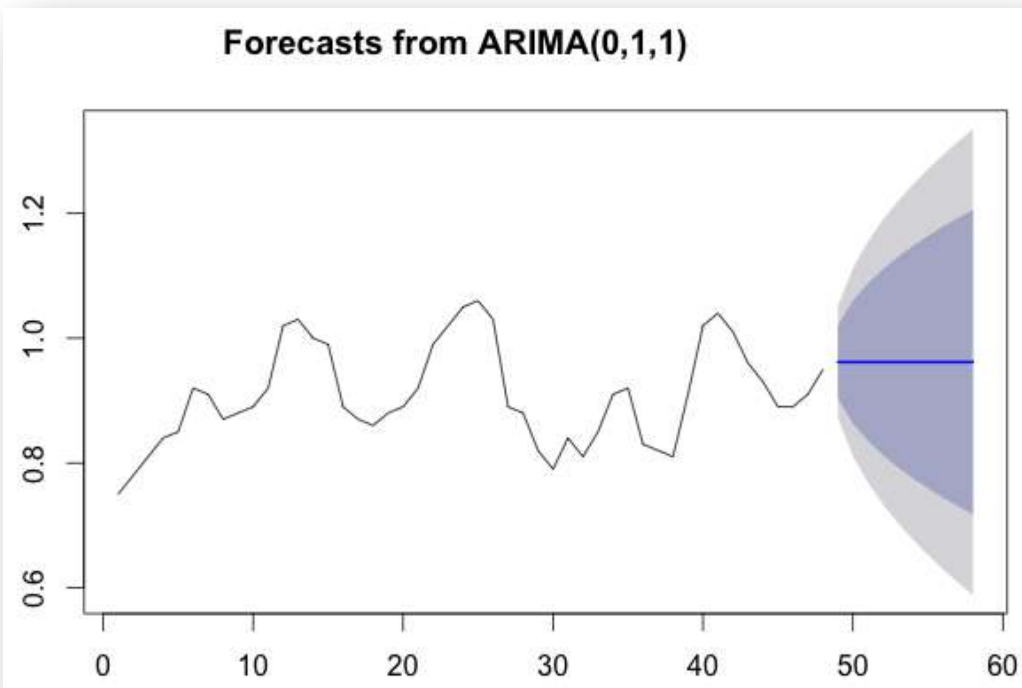

Figure 2 – a plot of CPI past performance using `auto.arima` to show the possible future values of CPI from the past performance. This specific example is from LRIP data, so the CPI values stabilize around period 25. The future values

[6] R is an Open Source programming language for statistics. This tool can be downloaded from [http://www.r–project.org/](http://www.r-project.org/) for both Windows and OS–X. Since it is free and can be applied with little effort it is a logic step in the forecasting of EAC using data contained in Earned Value Management repositories. As well the tool provide the basis for the Principal Component Analysis described in the next section of this paper. The final benefit is any college study majoring in economics, biology or chemistry, or pre–med has applied the tool to their homework problems.

are reflective of the dramatic swings in CPI early in the program and must be adjusted if we are to get actual forecasts. This diagram is for illustrative purposes only, to show the ease of use of the R tool and how it can be applied to "real" data in a repository.

What Can We Do With This Information

Now that we have a mechanism for using the repository Earned Value Management data to forecast future performance of that data, what can we do with it? First we need to establish the data architecture for the repository contents. This starts with normalizing the Work Breakdown Structure topology. One approach is to use MIL–STD–881C. In the appendencies there are "notional" structures for the WBS of products. These appendices are not inclusive, and in some case they are not all that helpful. But they can be a notional start to a standardized approach to capturing and recording data.

The beneficial outcomes of applying Time Series Forecasting using the data from the Earned Value repository includes:

Action	Outcome	Benefits
Normalization of EV data	Consistent data content needed for time series analysis	Setting the baseline for further analysis without having to adjust units of measure, remove spurious data, or fill in the gaps for missing data.
EAC forecasts by WBS element	Statistically corrected Estimate At Completion	Replace the non–statistical, linear EACs used today to improve confidence in forecast.
Correlation between time series	Correlation assessments to make visible coupling at WBS level of CPI/SPI	Connecting the dots at the work performance level of potential drivers of variance.

Table 1 – three steps needed to start using Earned Value Management repository data in forecasting the Estimate At Completion using statistically sound processes. The current approach in **Equation 3**, using linear non–statistically adjusted, non–risk adjusted arithmetic is used to forecast future performance with a straight–line project for the current CPI/SPI numbers. These values also wipe out any statistical variance information from the past and assume no statistical variance will occur in the future.

The Second Step Toward Better Estimate At Completion

With a forecast of EAC based on past performance of CPI and SPI using time series analysis and ARIMA, we can ask about improving that forecast by adding other measures of performance that should already be applied to programs through the Systems Engineering processes and described in the Systems Engineering Plan (SEP) and the Systems Engineering Management Plan (SEMP). The contents of these two documents are shown in **Table 2.**

Systems Engineering Plan (SEP)	Systems Engineering Management Plan (SEMP)
▪ What needs to happen from the customer perspective	▪ A formal response to the SEP and contract language
▪ Approaches to key areas: – Requirements – Technical staffing and Organizational Planning – Technical Baseline Management – Technical Review Planning – Integration with Program Management	▪ Defines contractor technical planning processes ▪ Articulates details of: – Processes – Tools – Organization
▪ Provides contact guidance for Systems Engineering applied to program	▪ Describes activities involved in transforming requirements into deliverables

Table 2 – the Systems Engineering Plan (SEP) and Systems Engineering Management Plan (SEMP) are guidance for creating the measurement needed to provide assessments of "Physical Percent Complete" for the program deliverables. These measures when connected with Earned Value Management measures (SPI and CPI) are the foundation of Predictive analytics and the first steps to Prescriptive Analytics needed to identify risks to the program's EAC and their corrective actions.

From this systems engineering guidance we can extract other measures of future behavior for the program:

- Technical Performance Measures (TPM) – current research connects TPMs with measures of BCWP
- Risk Values from risk management tool – adjusted EAC forecasts for future risk impacts
- Measures of Effectiveness (MOE)
- Measures of Performance (MOP)
- Key Performance Parameters (KPP)

These data elements are arranged in a matrix in preparation for the next step, Principal Component Analysis, where each column $Attribute_i$ contains the data from that period, for example the (1) WBS element, along with (2) CPI, (3) SPI, (4) TPM, (5) Risk, (6) MOE, (7) MOP, (8) KPP, (9) Staffing values reported for that period in the row elements labeled $X_{i=1\cdots 8,\, j=1\cdots n}$.

This general data structure is then used to find the Principal Components that are the primary variance generators.

$$
\begin{array}{cccccc}
X_{i=1\cdots 8,\, j=1\cdots n} & t_1 & t_2 & t_3 & \cdots & t_n \\
Attribute_1 & X_{1,1} & X_{2,1} & X_{3,1} & \cdots & X_{n,1} \\
Attribute_2 & X_{1,2} & X_{2,2} & X_{3,2} & \cdots & X_{n,2} \\
Attribute_3 & X_{1,3} & X_{2,3} & X_{3,3} & \cdots & X_{n,3} \\
Attribute_4 & X_{1,4} & X_{2,4} & X_{3,4} & \cdots & X_{n,4} \\
Attribute_5 & X_{1,5} & X_{2,5} & X_{3,5} & \cdots & X_{n.5} \\
Attribute_6 & X_{1,6} & X_{2,6} & X_{3,6} & \cdots & X_{n,6} \\
Attribute_7 & X_{1,7} & X_{2,7} & X_{3,7} & \cdots & X_{n,7} \\
Attribute_8 & X_{1,8} & X_{2,8} & X_{3,8} & \cdots & X_{n.8} \\
Attribute_9 & X_{1,9} & X_{2,9} & X_{3,9} & \cdots & X_{n,9}
\end{array}
$$

Figure 3 – the matrix of data extracted from a repository in preparation for Principal Component Analysis (PCA) to reveal the *drivers* of unfavorable variances and connect with the WBS elements to guide the program manager to the location to take corrective actions.

Principal Component Analysis in a Nutshell [7]

If x is a random vector of dimension p with finite $p \times p$ variance-covariance matrix $V[x] = \Sigma$, then the *principal component analysis* finds the directions of the greatest variance of the linear combinations of *x*'s. In other words, it seeks the orthonormal set of coefficient vectors $a_1, \dots, a_k$, such that,

$$
\begin{aligned}
a_1 &= \arg\max_{\mathbf{a}:\|a\|=1} V[a'x], \\
&\cdots \\
a_k &= \arg\max_{\substack{\mathbf{a}:\|a\|=1 \\ \mathbf{a}\perp a_1,\dots,a_{k-1}}} V[a'x],
\end{aligned}
\tag{6}
$$

The linear combination $a'_k x$ is referred to as the k^{th} *principal component*. [JOLL01], [EVER11]. This of course is a mathematical description of *Principal Component Analysis* and completely ***un*****_actionable_** for a Program Manager. So the *actionable* description is simple.

Tell me which of the variables in my program, represented by the time series of their respective values, is driving the variances I see in the selected variables – the SPI and CPI – so I can go find a corrective action to keep my program GREEN.

7 Of course this *nutshell* is likely a very large object, well beyond this paper, this presentation, or any discussion outside of a few weeks of effort, data processes, and experimenting. But this is the basis of modern data analysis in marketing, social sciences, and hard sciences.

THIRD STEP TOWARD BETTER ESTIMATE AT COMPLETION

With improved forecasting tools, the Third step is to make visible the connections between each measure to reveal the *drivers* of EAC. This means identifying of the connections between the technical performance, risk, and other variables on the project including core Earned Value data.

This approach provides the Program Manager with insight to the dynamic behavior of the program in ways not available from Descriptive Analytics. The standard analysis using CPI and SPI, using EAC calculations only states the value of the estimated cost. It does not reveal what is driving that cost and what the contribution of those drivers are to the total cost.

When we add more variables to the collection of program performance data, we create a new problem. We will need to identify which of these data items are the *Principal* ones that can provide *actionable* information to the Program Manager. We will apply Principal Component Analysis (PCA) to identify patterns in the data and express this data in a way to highlight similarities and differences.

Using Principal Component Analysis To Discover Drivers of Probabilistic EAC

Principal Component Analysis (PCA) decomposes a number of correlated variables within a dataset into a number of uncorrelated Principal Components. The result is a reduced dimensionality of a data set consisting of a large number of interrelated variables, while retaining as much as possible of the variation present in the data set [JOLL02]. This is achieved by transforming to a new set of variables, the Principal Components, which are uncorrelated, and which are order so the first few retain most of the variation present in all of the original variables.

The extracted Principal Components are estimated as the projections on the eigenvectors of the covariance or correlation matrix of this dataset. The variance of a dataset is an indicator of how spread out the data is. The larger the deviation, the more information is included. In practice, if 80–90% of the total variance in a multivariate dataset can be accounted for by the first few PCs, corresponding to the largest eigenvalues of the covariance matrix, then the remaining components can be rejected without much loss of information. [SAMA99]

CALL TO ACTION

After this very quick overview of the problem and two proposed solutions to increasing visibility to unanticipated growth in EAC, here is are actionable steps, known to address the issue:

0	Normalize data in the Central Repository in preparation for analysis	EV, WBS, Timeline units normalized with No missing items
1	Apply ARIMA to normalized data to forecast CPI, SPI, and Calculated EAC	Product EAC now based on statistically sound forecast
2	Adjust ARIMA parameters using past performance compliance	Tune ARIMA for program phase sensitivity
3	Integrate external data with repository EV data to build correlations to EAC	Add risk register and SEMP measures to ARIMA
5	Apply Principal Component Analysis (PCA) to identify correlated *drivers* of EAC growth	Research the Use multivariate forecasting for EAC

Normalize Data In Repository

For a “Big Data” repository to function properly the data needs to be normalized. This means the periods of performance are consistent, the data scaled are consistent, the primary data keys – the WBS numbers – are consistent, and the values of the data have defined ranges consistent with their represented elements. Using the WBS structure from MIL–STD–881C as a start, no matter what the structure turns out to be important attributes are needed. The WBS must be well formed, that is it must possess transitive closure as a minimal attribute so navigation of the tree structure is consistent across all programs held in the repository. [8]

Apply ARIMA

With a consistent set of data, known data elements with normalized values, no missing data – or if it is missing it is identified as missing – and a well structure decomposition of the WBS, the time series analysis can take place.

This time series analysis can be one of many choices, simple ARIMA(*p*,*d*,*q*) is a start. Holt–Winters is another popular approach, but others are available. Research will be needed to determine to appropriate approach. This can start with a long times series of data, apply the forecasting algorithm to reveal an intermediate value, confirm that the forecast value matches the actual values. With the variances from the result, adjust the parameters to improve the forecasting ability.

Adjust ARIMA

With actual programs, tested for intermediate forecasting, and comparisons of actual data to forecast data, sensitivity analysis of tuning parameters can be observed. This is the basis of all control closed loop tuning. Adaptive tuning, dynamic characterization, *feed forward* tuning are possibilities for early detections of unanticipated growth of EAC [KEAT11].

Integrate External Data

Earned Value data alone provides Descriptive assessment of past performance – assuming no cumulative data being used. With period of performance data, ARIMA can be used to forecast future values of CPI and SPI.

But our research has shown – and is spoken about in a parallel presentation by the same authors – that connecting Technical Performance Measures with the assessment of Physical Percent Complete is a critical factor in crating a credible BCWP.

This external data starts with the contents of the Systems Engineering Management Plan (SEMP). This is where Measures of Effectiveness, Measures of Performance, Technical Performance Measures, and Key Performance Parameters are identified.

- **Measures of Effectiveness** –operational measures of success, closely related to the achievements of the mission or operational objectives evaluated in the operational environment, under a specific set of conditions.
- **Measures of Performance** – characterize physical or functional attributes relating to the system operation, measured or estimated under specific conditions.
- **Technical Performance Measures** –determine how well a system or system element is satisfying or expected to satisfy a technical requirement or goal.
- **Key Performance Parameters** – capabilities and characteristics so significant that failure to meet them can be cause for reevaluation, reassessing, or termination of the program.

[8] A Set of all children aggregated as an activity $\in$A forms a transitive closure

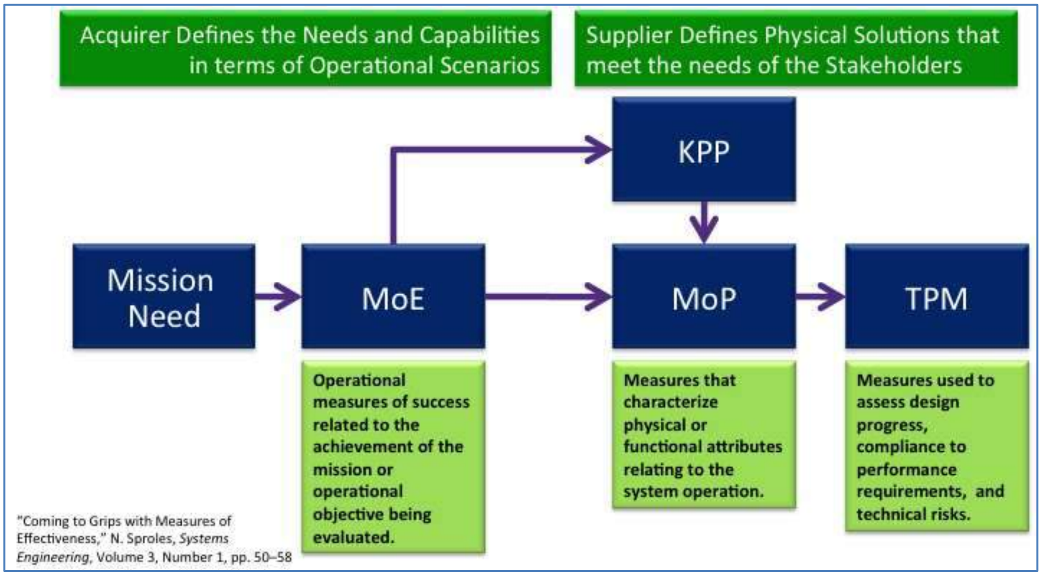

Figure 4 – connectivity between MOE, MOP, KPP, and TPM forms the basis of assessing physical percent complete and adjusting the BCWP in the Earned Value Management database. By applying these measures with EV data a credible forecast of Estimate At Completion cannot be developed from the integrated data.

Apply Principal Components Analysis

With the baseline data from the Earned Value reporting process, augmented with the data in **Figure 4**, Principal Components can be found that *drive* the programs variance. This variance is the source of increases in the Estimate At Completion. This alone is not sufficient to provide the Program Manager with actionable information.

Connecting these variances with each Work Breakdown Structure element is the source of this actionable information. By connecting variance with WBS elements, the source of variance can be revealed. Then corrective actions can be taken.

CHALLENGES TO OUR CALL TO ACTION

We have presented a possible solution, making use of existing data repositories, to increase confidence in the EAC using statistical processes. These processes include applying the ARIMA to the past Earned Value Management data submitted on a monthly basis and PCA to identify the drivers of unanticipated growth to the EAC. These techniques better descriptive and prescriptive qualities for forecasting EAC.

For these benefits to be realized, the Performance Assessment community must make progress on the following:

- Make the IMP mandatory to collect MOE, MOP, KPP, and TPM – Without the IMP there is no traceability to these measures to the work performed in the IMS.
- Ensure the SEMP is flowed down to the contractor and data reported back to the government – the SEMP defines the performance measures
- Establish uniform data quality business rules for the CR
- Make a concerted effort to integrate EVM CR data with the SEMP
- Make use of ARIMA and PCA in the PARCA/Performance Assessment tool set using all available data

References

[EVER11] *An Introduction to Applied Multivariate Analysis with R (Use R!)*, Brian Everitt and Torsten Hothorn, Springer, 2011.

[JOLL02] *Principal Component Analysis*, Second Edition, I. T. Jolliffe, Springer–Verlag, 2002

[KEAT11] "Using Earned Value Data to Detect Potential Problems in Acquisition Contracts," C. Grant Keaton, Second Lieutenant, USAF, Air Force Institute of Technology, March 2011.

[BOX70] *Time Series Analysis: Forecasting and Control*, G. Box and G. Jenkins, Holden–Day, 1970.

[RAND13] Management Perspectives Pertaining to Root Cause Analyses of Nunn–McCurdy Breaches, Volume 4 Program Manager Tenure, Oversight of Acquisition Category II Programs, and Framing Assumptions, RAND Corporation, 2013

[RAND12] Root Cause Analyses of Nunn–McCurdy Breaches, Volume 2 Excalibur Artillery Projectile and the Navy Enterprise Resource Planning Program, with an Approach to Analyzing Complexity and Risk, RAND Corporation, 2012.

[RAND11] "Root Cause Analyses of Nunn–McCurdy Breaches, Volume 1 Zumwalt–Class Destroyer, Joint Strike Fighter, Longbow Apache, and Wideband Global Satellite," RAND Corporation, 2011.

[IDA11] Expeditionary Combat Support System: Root Cause Analysis, Institute for Defense Analyses, 2011

[IDA11a] Global Hawk: Root Cause Analysis of Projected Unit Cost Growth, Institute for Defense Analyses, 2011.

[IDA10] "Root Cause Analysis for the ATIRCM/CMWS Program," Institute for Defense Analyses, 2010.

[SAMA99] "Experimental comparison of data transformation procedures for analysis of principal components," Samal, M., Kárny, M., Benali, H., Backfrieder, W., Todd–Pokropek, A., & Bergmann, H. (1999). *Physics in Medicine and Biology*, 44(11), 2821–2834.