Managing Product Maturity

Problem presented by

Richard Hadji

BAE Systems, Warton

Problem statement

Major defence engineering contracts involve highly innovative product development. For these products to be managed effectively, BAE Systems has implemented a novel management system to chart progression through the product lifecycle, thereby allowing early identification and resolution of problems. Considering the product hierarchically from the "bottom up", an elaborate monthly scoring system is in place to measure the progress of component parts.

BAE Systems asked the study group to analyse data gleaned as if from this process to determine whether past performance is a guide to future progress. This included understanding the impact of interdependencies and the effect of aggregation up to the product level. BAE Systems also asked for comment on the nature of the maturity modelling approach currently used and advice on how improvements can be made, including on how the maturity value is calculated and whether tasks should be weighted. They also wished to know what relationships exist between dependent parameters, and how assignment of issues moves between them.

In response, the Study Group analysed the data; they suggested that "value at risk" models might be useful in this context and considered how incentives might be designed to encourage accurate reporting.

Study Group contributors

John Billingham (University of Nottingham)
Dave Hewett (University of Oxford)
Marcin Jaroszewski (University of Nottingham)
Ricardo Pachon (University of Oxford)
Dave Szotten (University of Manchester)
Eddie Wilson (University of Bristol)
Tim Gardener (Smith Institute)

Report prepared by

Tim Gardener (Smith Institute)

FINAL VERSION

1 Background

BAE Systems is a global company that manages and delivers large and complex innovative engineering products. This complexity entails multi-national partners and suppliers, changing requirements, cutting-edge technology together with long timescales and multi-billion pound investments. After many years of experience in the delivery of major defence projects, BAE Systems has initiated an innovative product management process to chart progression through the product lifecycle, thereby allowing early identification and resolution of problems.

1.1 Maturity described

How exactly can one chart the progression through the product lifecycle to allow early identification and resolution of problems? This question is decomposed into four quantifiable questions; the first linked to the contractual specification at a product level, and the subsequent three to issues in the lower level specifications on what is actually required to make the product work:

- (M1) What progress is there against customer specification?
- (M2) What progress is there against changes to lower level specifications that have been identified as important?
- (M3) What progress is there against interface issues currently under negotiation with stakeholders (internal or external customers, suppliers or other partners)?
- (M4) What other possible changes might we envisage impacting on progress?

The first of these questions (M1) relates to standard project management practice. The remaining three are BAE Systems' current means to manage the difference between contractual theory and engineering practicality. At one end of the scale, (M2) is recognised and agreed by all parties as necessary changes to the lower level specifications; slightly worse, (M3) is the uncertainty caused by negotiations between stakeholders; and worst of all, (M4) is the amount of potential change as a result of issues we do not yet fully grasp.

1.2 Product complexity

A product (or system-of-systems) is broken down into systems, sub-systems and equipments, together with integrating elements at each level. This can be understood in terms of a *product tree.* (See Figure 1)

At the lowest level, for each branch in the tree the element is ascribed a *maturity score.* The maturity score M for an element is currently calculated by multiplying together the percentage scores responding to the questions (M1)-(M4) as previously described.

$$M = M1 \times M2 \times M3 \times M4$$

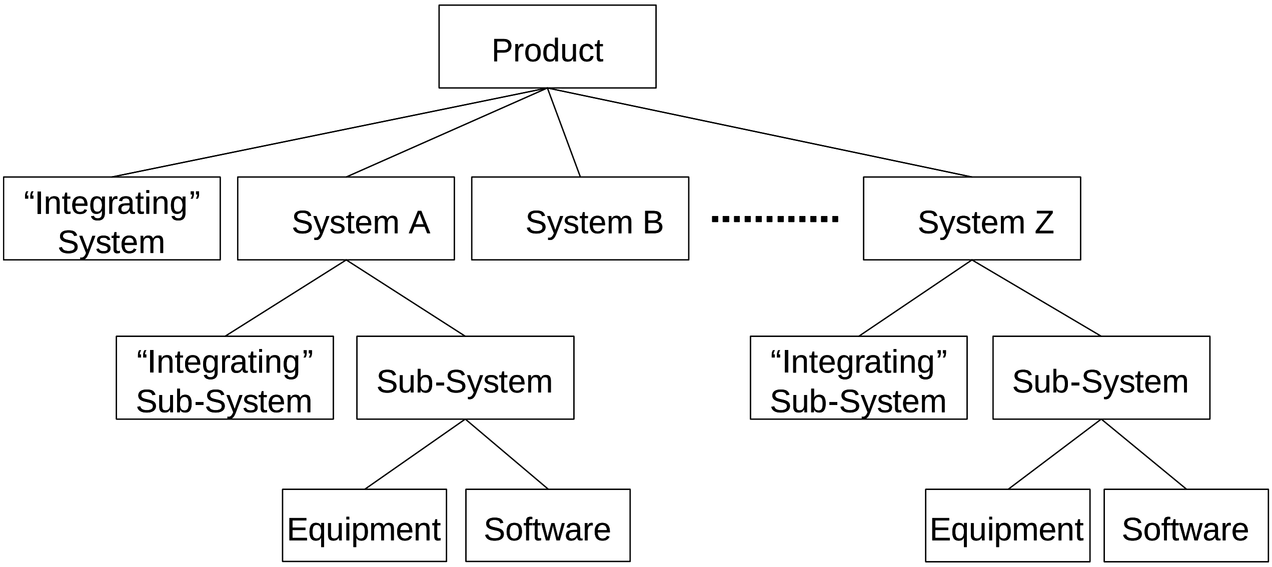

Figure 1: Product Tree

Maturity scores from each lowest level element of the product are recorded monthly so that over time the evolution of the product maturity can be traced. The maturity scores at lower levels of the product breakdown are aggregated to give a representation of maturity at the higher levels. When all maturity scores are aggregated then one gains an assessment of the maturity state of the entire product. These scores are used in planning since the project managers want to be able to understand when and where to allocate resources most efficiently.

Aggregation entails taking a simple average across all elements of interest. For example, in subsystem A, the average over all elements in the product tree below subsystem A would give a maturity score for that subsystem.

2 Data Analysis

2.1 Caveat

The product data given to the study group was assumed to be dummy data, though representative of genuine data amassed over a number of months of product development. Consequently the analysis of the data given is "dummy analysis"; that is to say, inferences taken from the dummy data may not apply to the genuine data. Nevertheless, techniques of data analysis employed at the study group should be illuminating for BAE Systems for "in-house" analysis of their real data.

2.2 Data Extraction

The data given to the study group comprised a number of Excel spreadsheets:

In total there were 7 products; for each of these products there were six systems; for each system there were between 8 and 10 subsystems (50 subsystems in total). Each subsystem had up to 10 lifecycle phases; each lifecycle phase there were 4 maturity scores per time period.

This amounts to a large data set. However, in the Excel format it is not straightforward (indeed not possible) to analyse trends in particular scores across all elements over all time periods. Such analysis is essential to see how the scoring system is working.

The first part of the study week was spent extracting data from the Excel spreadsheets for analysis within MATLAB. The MATLAB package is much better suited to this sort of data analysis than Excel. Extracting data from Excel was a moderately painful process, but once achieved it enabled some insights. For this reason it is recommended that BAE Systems may wish to consider a tool such as MATLAB for data analysis of this sort.

2.2.1 First Observation: not much action

The histogram of month on month changes to the M1 scores (Figure 2) indicates that most of the time there is no change at all. Of all the data points available, over 5200 indicate no change between months; the next most frequent change is a 5% increase, but there less than 80 of these events recorded over the period covered by the data.

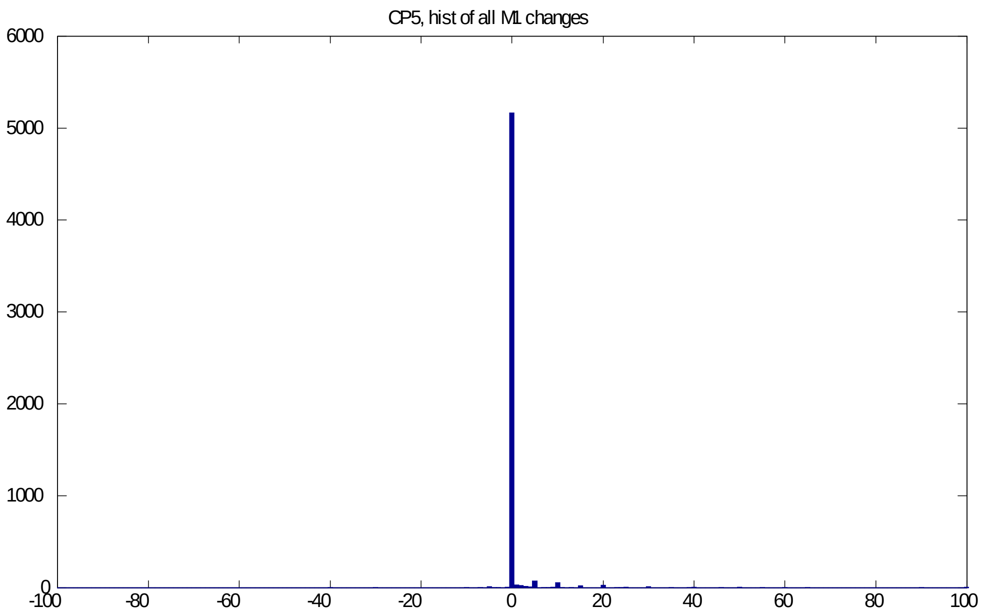

Figure 2: Month on month change in the M1 score

Magnifying this graph by a factor of 50 (Figure 3) one can look in more detail at the subtle changes in the M1 variable month on month. (Remember that the point at zero is off the scale!)

Now this histogram indicates that scores tend to shift very rarely, but when they do shift it is in multiples of 5%. Most shifts are less than ±20, but there are some notable exceptions which shift from 0 to 100, and one egregious exception which shifts from 100 to 0!

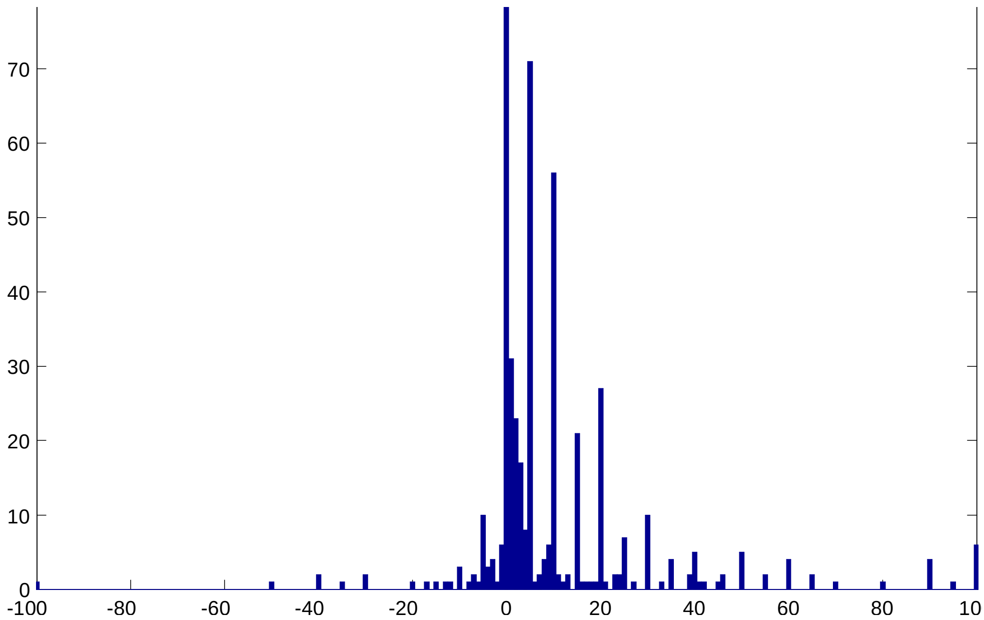

Figure 3: Month on month change in M1 score, multiplied by a factor of 50

2.2.2 Second Observation: correlated values.

BAE Systems asked the study group to determine any relationships which might exist between the variables M1, M2, M3 and M4. Once more, looking at the data as a whole, one can plot scatter diagrams of these value pairwise (Figures 4, 5). These diagrams have been adapted (by adding a small element of random noise) so that the density of points clustered on a single score becomes clear.

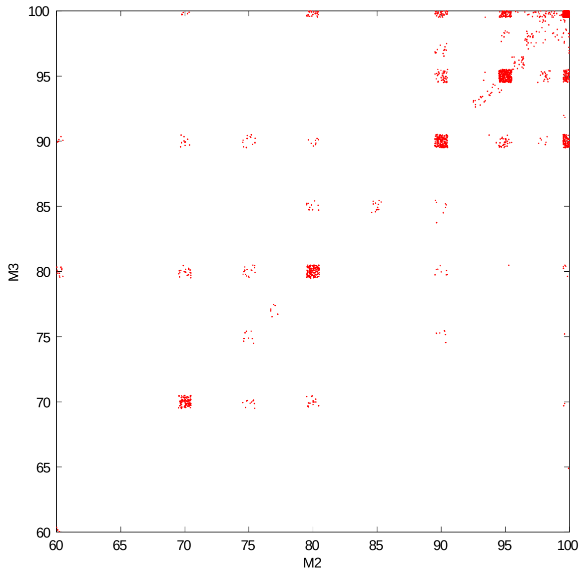

Figure 4: M2 values plotted against M3 values

Looking at the diagram of M2 values plotted against M3 values, there is a very strong correlation evident. In fact, the scores are usually on the leading diagonal, in effect these measures are mostly the same. That is to say, the M3 score tells you virtually nothing that is not already captured by the M2 score; there is no point having two separate scores. Above the 90% threshold the correlation is slightly weaker, though it is still very obvious. The fact that scores cluster round multiples of 5% is also reinforced by this data.

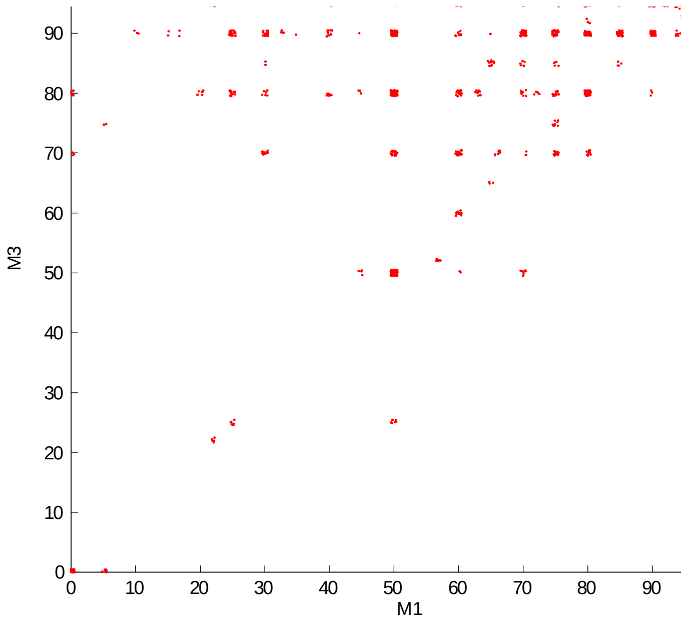

Figure 5: M1 values plotted against M3 values

Now considering the values of M1 plotted against M3, the correlation is less. In general, M3 scores are kept high (mostly above 70%) while M1 scores range more widely. This may be due to the phase of production to which the data refers. But it may be also that while M1 is readily understandable as a score, the definition of M3 is less clear. As a consequence, people score more confidently against the readliy understandable scale; where there is doubt in the score the range is likely to be smaller.

2.2.3 Third Observation: uncertainty flow?

Noting the definitions of M2 and M3, it seemed reasonable to formulate the hypothesis that there should be evidence of "uncertainty flow" from M3 to M2. That is to say, if one way of progressing "issues under negotiation" is for them to become "changes to lower level specification that have been agreed" then from time to time the M3 score should go up in the same month that the M2 score decreases. To examine this, the study group plotted the change in M2 scores against the change in M3 scores.

The first point to note is that most months there is no movement in either score. However what is much more remarkable about this chart is that there are no data points at all in

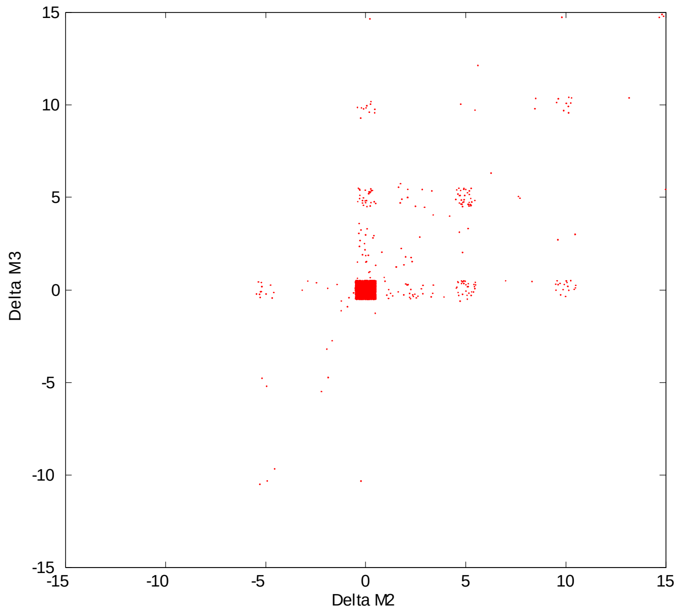

Figure 6: Flow between M2 and M3?

the top left quadrant or the lower right quadrant. This indicates that there is no flow from M3 to M2 or from M2 to M3 ever. So either the study group has misunderstood the expected relationship between M2 and M3, or the dummy data indicates that M2 and M3 are not being used as intended.

2.3 Issues arising from the Data Analysis

2.3.1 Issue 1

Since most M1 scores do not move month by month, one should ask to what extent is M1 measuring monthly progress? Should one expect a more linear increase month on month, reflecting a steady work rate on each item? If not, why not?

2.3.2 Issue 2

Also from looking at the values in the scatter diagrams together with the movement in the M1 variable, it seems likely that scorers think in terms of intervals of 5%. This may be unimportant, but there is some possibly relevant research that looks at the psychology of scoring. ("*Questions and Answers in Attitude Surveys*", Howard Schuman and Stanley Presser, Sage Publications, 1996.)

2.3.3 Issue 3

One might ask why M1 scores range more than M3 scores? From the numerical data alone this is impossible to deduce. However it is not unreasonable to speculate that scorers, being aware that the product of the M values is what matters, take the obvious step to control this product; namely keep three variables high and vary only the first one. The study group was in no position to assert that this phenomena was definitely occurring, but it is one obvious possibility that could not be discounted from the data alone.

If the geometric mean were used rather than the product, that is take

$$M = (M1 \times M2 \times M3 \times M4)^{\frac{1}{4}}$$

then altering M1 alone would have much less effect; one would perhaps be more inclined to vary the other three scores.

2.3.4 Issue 4

A further possibility is that the definitions of the M2, M3 and M4 variables are not intuitively obvious and consequently not easy to score. Since M1 is a familiar project management quantitative metric, scorers may be more comfortable with this as a measure of performance.

2.3.5 Issue 5

Systems that require human scoring are notoriously difficult to calibrate. It is essential to rely on human expertise but at the same time there is always the possibility that the scorers are unintentionally given incentives which skew accurate reporting. The study group began to explore this general possibility, and look at the question of how to design a reward mechanism that encourages accurate forecasts of stretching targets.

2.3.6 Issue 6

One final remark on the low level data: it may be that looking so closely at the low level data is not a wholly fair thing to do; only by stepping back and aggregating to the product level does the measurement process gain requisite clarity.

3 Aggregation of the Data

Aggegating the data is a difficult problem. Unless one keeps track of the four numbers separately, multiplying all four numbers together loses a lot of information. Four numbers yield four dimensions of information, multiplying them together and retaining only the product gives only one dimension of information. Tracking the scores separately for longer will retain more information and allow a more sensitive assessment of the project progress.

The current BAE Systems way of using the scores is firstly to multiply them together and then to aggregate by taking an average over the multiplied scores. Because multiplying the scores together like this loses information, a better alternative may be to aggregate

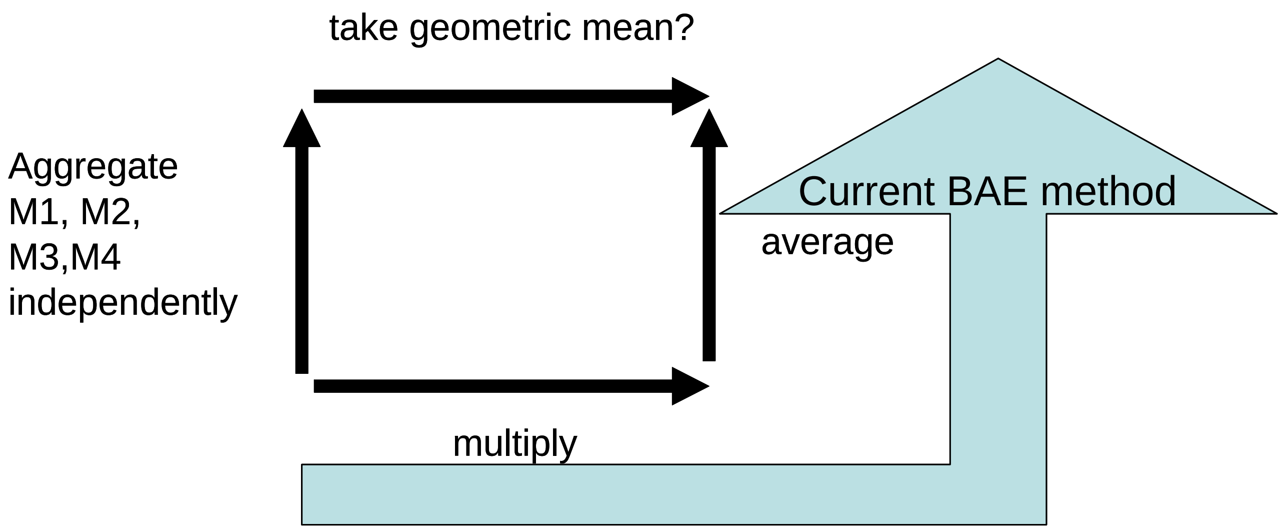

Figure 7: Aggregation

M1,M2,M3 and M4 independently and then decide how to combine the scores. The results would be markedly different, possibly giving a more accurate picture of progress. Of course one should do both for comparison. This simple idea warrants further investigation. During the week of the study group there was not sufficient time to explore this further.

One possible direction to take, having aggregated the scores separately, is to determine how best to model the uncertainty. With their current definitions, BAE Systems use M3 and M4 to gauge the uncertainty of a project with M1 and M2 measuring the status and rate of progress. It may be illuminating to think of maturity in terms of expected rate of completion together with a short term estimate of uncertainty. Short term estimates are much easier to give than longer term estimates. The point of thinking this way is that the methods of stochastic processes may then be useful; one can then work by analogy, considering the expected rate of progress as a "drift" and the short term estimates of uncertainty as "volatility". For this it is essential to look at the data aggregated, for as has already been described, the majority of subsystem level scores move so infrequently. But aggregating more sensitively may make the problem amenable to an approach through stochastic processes.

3.1 A risk-analysis approach to the maturity problem

3.1.1 Value at Risk in financial markets

In economics and finance, value at risk (VaR) is a measure of how the market value of a portfolio of assets is likely to change over a given time period under certain conditions. It is typically used by security houses or investment banks to measure the market risk of their asset portfolios (market value at risk), but is actually a very general concept that has broad application. The Study Group has recognised a similarity between financial risks analysis and maturity scoring, and we propose a model inspired by VaR techniques that might increase BAE Systems' capability to measure, analyze and understand the progress of highly complex projects.

3.1.2 A Time-at-Risk model for the maturity problem

Our model is defined on a parent node and its child nodes in the heirarchical product tree of Figure 1. Using the current maturity of each of the child nodes, together with historical data from the progress made so far, the model determines a probability density function for the finishing time of the project associated to the parent node. Since the product tree is made of parent-child blocks, we can "roll up" the outputs from our model, from the lowest level to the highest, and obtain in this way a probability density function for the whole project's finishing time.

The model that we propose is probabilistic in nature instead of BAE Systems' current deterministic model. We believe that there are two principal advantages with this approach: clarity and transparency. The probabilistic elements can be used directly to obtain a clear definition of maturity with a precise mathematical formulation. It is also straightforward to understand. So while the model has a solid mathematical basis on which many improvements can be made, the final output will still have a simple and precise meaning. It is easy to explain to non-experts, and is appropriate to take decisions for the project's development.

Suppose that a particular parent node P has associated n child nodes $\{c_1, \ldots, c_n\}$, i.e., the development of a particular system is divided in n subsystems, each of them having so far a particular progress. For the simple model that we present in this report, we assume that the task associated to P is made up of the individual contributions of each c_i, so the project finishes only once the tasks of all child nodes are completed.

We associate to each c_i the finishing time T_i of all its tasks. We consider T_i as a random variable. The finishing time of P is modelled with the random variable T_p. Our goal is to find a relation between the T_p and the T_i's that we can use to compute the former from the later. To obtain this, we have three steps:

- If $\{c_i\}$ are nodes at the lowest level, we use historical data of the progress of maturity to obtain approximate density functions for each T_i. We refer to the methods used to determine the density functions T_i as *inference procedures*.
- From the particular relations between $\{c_i\}$, and between $\{c_i\}$ and P, we construct a function θ that maps the random variables T_i to the random variable T_p so $T_p = \theta(T_1, \ldots, T_n)$. The purpose of this function is to aggregate the contributions of $\{c_i\}$ in such a way that we can infer a realistic value for T_p, or at least to approximate it. This function θ crucially accounts for the integration issues between the contributions. For a simple situation we could propose a particular function, but more complex relations can be modelled by introducing other terms. We refer to the design of the particular function θ as the *mapping design*.
- We use the density functions of the T_i's to compute from the function θ the distribution (or approximate distribution) of T_p. We refer to this computation as the *transformation procedures*.

3.1.3 An Example of the Time-at-Risk model

To develop the ideas that we have presented above, we present in this section a simple model that can be used as a starting point.

Consider the child nodes $\{c_i\}$ of a parent node P at the lowest level in the hierarchical tree. The parent node could be, for example, a subsystem called "A" and its child nodes are the life cycles: Requirements Capture, Design, Build, Qualification, Certification, Flight Test, Verification Claim, Acceptance, and Service Support. We develop a model that will have as input the current completeness of each child node and historical data, not only of these nodes, but perhaps of the whole project, i.e., progress of other subsystems and even at higher levels, and the output is a measure of how delayed is the progress of P. We begin by explaining the interpretation of this output and then elaborate the ideas of the three stages of the model that we mentioned in the previous section.

3.1.4 Inference procedures

If we are considering a parent-children block that is not in the lowest level, we use the T_i for each c_i computed from its child nodes. However, if the child nodes are at the lowest level, we need to propose for them some suitable distributions.

We believe that the work that BAE Systems have done for maturity scoring may be used for this purpose. The metrics M2, M3 and M4 (for stability, consistency and uncertainty) can be thought as the parameters that model the uncertainty of the project for the future, however they need to be modified for our purposes.

The right choice of a density function for T_i is very important and some expert knowledge might help to identify the most appropriate ones. For example, if at the beginning of the project we estimate that the finishing time for P is μ, a typical gaussian distribution with mean μ does not seem to be appropriate, since it is very unlikely that we expect such a symmetry in the finishing times around μ (it is very unlikely that, if we expect the project will end in July, we expect with the same probability that the project finishes in February or in December!).

Distributions skewed to lesser values seem to be an option. A chi-squared distribution with three or more degrees of freedom might be used as a first option. A beta distribution with $\alpha = 2, \beta = 5$ might be used as well.

Historical data can also be used to identify a good choice for the density function of T_i. The idea is the following: From the progress observed in different projects of BAE Systems, a certain pattern of the progress of a particular node might be associated with a particular behavior in the future. For example, projects with a slower beginning than expected might be observed to develop long stagnation times. A simple ranking for the nodes could be implemented using standard time-series analysis tools and at different time steps, each child node is tagged with different parameters or with an entirely different density function.

3.1.5 Mapping design

Suppose that we have the density functions $f_1, \ldots, f_n$ for each child node $c_1, \ldots, c_n$. We propose a function θ that takes the completion times $\{T_i\}$ and produces the variable T_p. The simplest function

$$\theta(T_1, \ldots T_n) := \max(T_1, \ldots, T_n)$$

so that the subsystem is complete only when the final contributing subsubsystem is complete.

We can model more complex relations by changing this function. For example, dependance of several T_i's on certain events can be introduced by taking their product, or if the composition of P depends of different percentages of their child nodes, a weighted average could be used.

We believe that the mapping design can give a lot of flexibility to BAE Systems to model the expected progress of its projects. Decisions taken at a higher level to modify the progress could be studied and analyzed by designing a suitable function that incorporates particular changes at lower levels in the hierarchical tree.

3.1.6 Transformation procedures

From θ together with the density functions $\{f_i\}$ one then calculates the associated density function f_p. For simple θ and independent variables T_i this may be straightforward, but for more complicated θ or with crucial dependencies, it may be necessary to use approximation methods. When θ is simple, i.e., linear or quadratic, elemental probability theory can be used to give an exact solution of f_p. However in the real world cases likely to arise that is not true. For example, for the simple function that we presented above we cannot estimate its variance.

Several numerical techniques can be used to approximate f_p. The one that we recommend to BAE Systems at this point is Monte Carlo simulation. Monte Carlo simulation can be computationally expensive for very big problems, however the magnitude of the problems given by the typical hierarchical tree that BAE Systems presented to the Study Group makes us believe that this will not be an issue. For a simple implementation of the Monte Carlo method, all that is needed is a reliable random number generator that could take the values specified for the density functions $\{f_i\}$. The construction of f_p is straightforward, and from this, the TaR measure. Advanced software is not even required; good estimates can be obtained with `Excel` or `Crystal Ball`

3.1.7 A Predictive Element: The time-at-risk measure

Suppose that we have already obtained a density function for T_p. How can we use it to obtain a meaningful measure of the expected progress of P in the future? Replicating the ideas of VaR, we compute the minimum time that exceeds some specified quantile of T_p. We define this value as the **time-at-risk metric** (TaR) and it tells us, with a given confidence level, which is the maximum finishing time that we should expect for P. For example, if the TaR metric that we find for subsystem A, specifying a .95 quantile is eleven months from now, we expect with a 95% of confidence that subsystem A will be finished at most in eleven months.

The TaR metric is an extreme value that measures the worst case scenario but still gives a realistic picture of it. It is easy to understand and is comparable with other projects. However other measures can be taken from the computed density function T_p and can be used to describe the situation of the project, i.e., values for central tendency, dispersion and skewness. In particular, it could be valuable to establish at the beginning of the project for P an expected density function $\tilde{T}_p$ of its finishing time. Then, at each update the deviation between T_p and $\tilde{T}_p$ could give deeper knowledge of the progress made so far, and what is expected in the future.

3.2 Future work and recommendations

The approach using risk-analysis is an exciting one which the study group could only begin to explore. It has a number of advantages. There is a solid mathematical theory which is reasonably well developed already but one would need to adapt it to support this style of question relevant to BAE Systems. Although the underpinning mathematics may appear to be complicated, the outputs should be clear and understandable to the non-expert. Graphs indicating the time-to-complete likelihoods for component parts are readily accessible. The underpinning mathematical functions are based on the data gathered through expert scoring and the aggregation from the low level data is sufficiently rich to give a more accurate picture of maturity status. However applying these ideas to the BAE context would be highly innovative.

4 Incentivising Accurate Reporting

The study group gave some thought to the problem of incentivising accurate reporting. The senior manager wishes to have as accurate a picture as possible, though he also wishes to encourage hard work and ambitious targets. The reporting expert has a much better (though still uncertain) grasp of what the reality of the situation is; he is guarded about disclosing the full facts to the senior manager since this could be misunderstood, or worse later used as evidence for even harder target setting.

The data analysis exercise suggests that managers may not be using the reporting system in the desired manner. In particular, for any one low-level task, the measures M_{2-4} change infrequently and tend to be highly correlated with each other with values close to 100% for most of the time. This last feature indicates an element of game-playing where the low-level manager attempts to control his overall score $M_1 M_2 M_3 M_4$ by forcing it to follow the well-defined measure M_1. Thus we are concerned here with the design of alternative reporting methods which encourage accurate reporting, motivate good performance, and penalise game-playing.

Our focus is a simplified analysis of a single task managed by a single junior manager who reports to a single senior manager. The junior manager will be assumed to be entirely self-serving, in that he will behave so as to maximise a reward function which is handed down by his senior manager, whose task is to design the reward function so that the junior manager behaves in the required manner.

Our feeling is that in practice, measures M_{2-4} represent aspects of uncertainty in the final date of completion of the task. We therefore propose a more direct and simplfied system where the junior manager is asked to forecast the date of completion of his task. Because of uncertainty in future progress, this forecast should essentially take the form of a probability distribution for completion time. However, this formulation would be too complex and would impose unrealistic demands upon the junior manager.

Hence we propose a simpler scheme, where the junior manager is asked to forecast a completion time, e.g. T_{90}, by when he is 90% certain that the task will be completed. Here we consider only the simplest 'one-shot' approach, where the forecast is made just once at the commencement of the task. However, it is relatively straightforward to generalise this set-up to one where the forecast is repeated at regular intervals, thus measuring progress.

When a task is handed down, the junior manager will rapidly evaluate it and form an impression of its likely rate of progress and the points at which problems may occur. We therefore suppose that the junior manager has in essence constructed a probability distribution for the completion time of the task, although he will not be conscious of it in these terms. The senior manager does not have this level of understanding, so asks the junior to report (e.g.) T_{90}. However, since the junior manager is self-serving, he does not report the true T_{90} but instead reports a value T_{90}^* which is chosen to maximise his expected reward.

We now consider the design of the reward function R. The key points are as follows.

(1) The senior manager's first goal is to design R so that $T_{90}^* = T_{90}$, i.e., so that the junior manager reports the likely completion date accurately. In particular, if the task eventually takes longer than T_{90}^*, the junior manager should be penalised, although the reward function should not drop off too dramatically since it is necessary to incentivise the completion of tasks which have run over their forecast completion date.

(2) It is necessary to incentivise the junior manager to produce as early a forecast as possible. This feature is required to discourage extremely pessimistic forecasts which can always be met since they do not require full effort.

(3) Finally, it is not clear whether very early completion should be incentivised too. Is it better to complete a task very early rather than just before the forecast completion time? To do so may be viewed either as good (because it demonstrates hard work, commitment to the task in hand) or bad (because it demonstrates that the original forecast was unnecessarily pessimistic).

We now model the above discussion in mathematical terms. We denote the probability density function for the completion time $T > 0$ by $f(T;\lambda)$ and we suppose that whereas the functional form is known to both managers (for example, it could be derived from historic data), the parameter λ is known only to the junior. Furthermore, the reward function will be denoted $R(T, T_p^*)$, where T is the actual time of completion and T_p^* is the time at which the junior manager forecasts the task to be complete with probability p. Thus we have generalised the above discussion which chose $p = 90\%$. Note that in general we should expect R to depend on both f and p also.

From this discussion, we obtain

$$\int_0^{T_p} f(T;\lambda)\,\mathrm{d}T = p, \tag{1}$$

and moreover the expected reward E is given by

$$E(T_p^*) = \int_0^\infty f(T;\lambda) R(T, T_p^*)\,\mathrm{d}T, \tag{2}$$

which the junior manager attempts to maximise by choosing T_p^* to solve $E'(T_p^*) = 0$, which yields

$$\int_0^\infty f(T;\lambda) \frac{\partial}{\partial T_p^*} R(T, T_p^*)\,\mathrm{d}T = 0. \tag{3}$$

Here we suppose that the reward and probability density functions are sufficiently smooth and that there is a unique interior turning point given by this formula which is a local and global maximum.

We now model the design requirements 1.-3. for R. The key observation is that if R is chosen correctly, then the junior's self-serving solution T_p^* of (3) must yield

$$\int_0^{T_p^*} f(T;\lambda)\,\mathrm{d}T = p, \tag{4}$$

for any value of the parameter λ, c.f. eq. (1), so that the correct T_p is always reported irrespective of the details of the task (property 1).

Since f is non-negative, eq. (3) shows that the reward function $R(T, T_p^*)$ itself must have a turning point in T_p^* (since otherwise no root would be possible, which would force the endpoint maximum case where the junior reports $T_p^* = \infty$). Furthermore, note that

$$E''(T_p^*) = \int_0^\infty f(T;\lambda) \frac{\partial^2}{\partial T_p^{*2}} R(T, T_p^*)\,\mathrm{d}T, \tag{5}$$

so that $\partial^2 R/\partial T_p^{*2} < 0$ for all arguments guarantees a local maximum if eq. (3) is satisfied. The desirable properties are thus

$$\frac{\partial R}{\partial T_p^*} > 0 \quad \text{for small } T_p^*, \qquad \text{and} \qquad \frac{\partial R}{\partial T_p^*} < 0 \quad \text{for large } T_p^*. \tag{6}$$

These requirements encapsulate property 2 since the reward decreases as T_p^* is increased past some critical point, which thus discourages the reporting of 'lazy' T_p^*. A key design requirement which needs further investigation concerns the point at which $\partial R/\partial T_p^*$ changes sign, and in particular how it is related to the actual completion time T. For instance should the reward be maximised for completion at $T = T_p^*$?

Finally, if early completion is always considered to be a good thing (property 3), we require

$$\frac{\partial R}{\partial T} < 0. \tag{7}$$

For our scheme to be implemented in practice, some knowledge of the density function $f(T;\lambda)$ would be required. Here we give an example of how the calculations might work in practice where we assume the simple exponential model

$$f(T;\lambda) := \begin{cases} 0, & \text{for } T < 1 \\ \lambda \exp\{-\lambda(T-1)\} & \text{for } T \geq 1, \end{cases} \tag{8}$$

where $T = 1$ is the normalised minimum completion time. Since the model is entirely autonomous in T, we may shift time and compute with

$$f(T;\lambda) := \lambda \exp(-\lambda T) \quad \text{for } T > 0. \tag{9}$$

Furthermore, informed by the above discussion, we propose the prototype reward function

$$R(T, T_p^*) := \frac{T_p^* - kT}{T_p^{*2}}, \tag{10}$$

where k is a parameter which must be found. Note that simple calculations verify requirements (6,7), with $\partial R/\partial T_p^*$ changing sign at $T_p^* = 2kT$. Therefore we need only solve eq. (3) for T_p^*, which gives

$$T_p^* = \frac{2k}{\lambda}, \tag{11}$$

for the junior manager's optimum strategy. We must now check that eq. (4) is satisfied, or equivalently that $T_p = T_p^*$. Since eqs. (1,9) give $T_p = -\log(1-p)/\lambda$, property 1 (correct reporting) may be incentivised if we set

$$k = -\frac{1}{2}\log(1-p). \tag{12}$$

It should be emphasised that the resulting reward function (10) is entirely independent of the density function parameter λ, and this is a key requirement since λ is what in effect the senior manager is trying to discover.

5 Conclusions

BAE Systems is pioneering in its development of tools to measure product maturity. The Study Group was pleased to consider this work and to make suggestions for improvement.

Firstly, BAE Systems should consider using a program such as MATLAB for its data analysis. Although Excel is convenient for most purposes, there is an additional functionality required to analyse data across the whole programme and this is where MATLAB excels Excel.

Secondly, BAE Systems should explore the techniques of Value at Risk in order to make full use of the data collected. Current BAE Systems techniques merge information too early (by multiplying scores together) and aggregation is a simple average. The VaR methods also lend themselves more naturally to a forecasting requirement. A probabilistic approach should yield more than the current static appraoch.

Finally, capturing expert judgement is essential but experts are clever and some care must be given to ensure that disincentives for accurate scoring are not unintentionally built into the reporting system.