Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 55 (2016) 29 – 34

www.elsevier.com/locate/procedia

5th CIRP Global Web Conference Research and Innovation for Future Production

Short interval control for the cost estimate baseline of novel high value manufacturing products – a complexity based approach

Oliver Schwabe[a*], Essam Shehab[a] and John Erkoyuncu[a]

[a]*Cranfield University, Building 50, Cranfield, Bedford MK43 0AL, U.K.*

* Corresponding author. Tel.: +49 (0) 170 9053671;. *E-mail address:* o.schwabe@cranfield.ac.uk

Abstract

Novel high value manufacturing products by default lack the minimum a priori data needed for forecasting cost variance over of time using regression based techniques. Forecasts which attempt to achieve this therefore suffer from significant variance which in turn places significant strain on budgetary assumptions and financial planning. The authors argue that for novel high value manufacturing products short interval control through continuous revision is necessary until the context of the baseline estimate stabilises sufficiently for extending the time intervals for revision. Case study data from the United States Department of Defence Scheduled Annual Summary Reports (1986-2013) is used to exemplify the approach. In this respect it must be remembered that the context of a baseline cost estimate is subject to a large number of assumptions regarding future plausible scenarios, the probability of such scenarios, and various requirements related to such. These assumptions change over time and the degree of their change is indicated by the extent that cost variance follows a forecast propagation curve that has been defined in advance. The presented approach determines the stability of this context by calculating the effort required to identify a propagation pattern for cost variance using the principles of Kolmogorov complexity. Only when that effort remains stable over a sufficient period of time can the revision periods for the cost estimate baseline be changed from continuous to discrete time intervals. The practical implication of the presented approach for novel high value manufacturing products is that attention is shifted from the bottom up or parametric estimation activity to the continuous management of the context for that cost estimate itself. This in turn enables a faster and more sustainable stabilisation of the estimating context which then creates the conditions for reducing cost estimate uncertainty in an actionable and timely manner.

Peer-review under responsibility of the scientific committee of the 5th CIRP Global Web Conference Research and Innovation for Future Production

Keywords: Context complexity; Cost estimate baseline; Cost estimate uncertainty

1. Introduction

Novel high value manufacturing products [1] by default lack the minimum a priori data [2,3] needed for forecasting cost variance over of time using regression based techniques [4,5,6]. The reference point for determining cost variance is the cost estimate baseline [7]. The degree of cost variance is the trigger for selecting actions to reduce such in order to maintain it within the limits of an initially determined budget. The implicit assumption being that this cost variance is an unplanned event. The earlier the actions for reducing cost variance can be triggered the lower their inherent cost and the lower the probability that budgetary limits are exceeded over time. The problem addressed by the authors is thus determining what intensity of cost variance monitoring is in fact required to identify such changes as early as possible while at the same time understanding when a forecast of cost variance can be performed in the first place. The monitoring intensity is defined by the length of time intervals between discrete monitoring intervals thus ranging from being continuous (therefore every registered change is immediately used to update the baseline estimate akin to real-time monitoring) through to the typical annual update intervals encountered in annual organizational financial reporting. The

Peer-review under responsibility of the scientific committee of the 5th CIRP Global Web Conference Research and Innovation for Future Production
doi:10.1016/j.procir.2016.08.016

authors hereby propose that short interval (continuous) control of the cost estimate baseline of novel high value manufacturing products should be used as long as the computational complexity of cost variance as measured by Kolmogorov principles is not steady between two measurements. The presented approach evaluates the complexity of cost variance over discrete time in order to identify when the patterns of cost variance change. This allows for earlier determination of when previously applied assumptions to the cost model will fail to accurately represent the propagation of cost estimate uncertainty, permit more timely adjustment of relevant budgets and enable more proactive management of factors leading to cost variance. In this respect it must be remembered that financial baselines are typically only revisited upon significant and repeated cost variance above commonly applied default contingency values to the financial baseline and thus represent reactive versus proactive cost management behavior [8, 9]. The presented method shifts the existing hindsight focus (therefore managing exceptions that have occurred) to one of foresight (therefore forecasting when exceptions will probably occur) in parallel to an observation and learning paradigm versus one of command and control which is by default unsuitable for the open complex system represented by the whole product life cycle of novel high value manufacturing products [10].

Section 2 introduces the use of Kolmogorov complexity for describing patterns of cost estimate uncertainty in the whole product life cycle. Section 3 describes how complexity patterns can be identified and Section 4 explains the process applied for determining the control mechanism of relevance, therefore when continuous (short interval) or discrete cost estimate uncertainty monitoring should be applied and whether or not a forecast is feasible. Section 5 contains conclusions and recommendations for further research.

2. Patterns of complexity in the whole product life cycle

Forecasting depends on identifying repeatable patterns and determining where in that pattern a context currently is [3, 11]. If such patterns do not exist then the control interval must be set at the shortest feasible intervals for cost management. The first step in identifying these patterns is determining where these patterns begin and where they end. The underlying assumption is that these patterns exist due to the open complex systems nature of the whole product life cycle [10]. This assumption is illustrated in a generic manner in Figure 1. The generic principle is then applied to case study data in the following section.

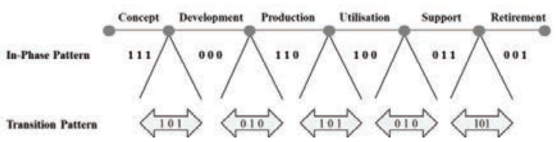

Fig. 1. Exemplary pattern structure

Figure 1 shows the phases of the whole product life cycle (therefore concept, development, production, utilization, support and retirement) [12] with exemplary cost variance patterns both in the phases and between the phases. The propagation behaviour of cost estimate uncertainty across discrete time intervals is described by a “0” for the lack or decrease of cost estimate uncertainty since the previous time step and a “1” for an increase in this cost estimate uncertainty. Figure 1 illustrates in an exemplary manner that each phase of the whole product life cycle consists of at least one unique pattern, for example “111” or “000”. The transition itself may then also have unique patterns such as “101”or “010”.

Determining when a pattern begins and when it ends is a process of inductive inference regarding the randomness of a sequence. Due the conditions of small data in the context the existence of a universal prior probability distribution is discounted thus leading to the use of the fundamental metric of Kolmogorov complexity as a starting point for examination of the generated bit string [13]. The metric of Kolmogorov complexity signifies the degree of compression a binary string can be subject to whereby compression is understood as the process of converting the sequence of bits into the description of the pattern represented by that bit sequence. The bit sequence is hence transformed into a program that can generate exactly that bit sequence. The program consists of a descriptor language which explains how a sequence of instructions is applied by a Turing Machine in order to generate the bit string.

When estimating and forecasting early in the whole product life cycle the small amount of data (if any) available will always result in the length of the descriptor and string exceeding the length of the bit strings themselves. This then means that the specific patterns cannot be identified although a change in these can be determined when applying the concept of complexity group changes.

A simple example is the bit string “1111111111”. The Kolmogorov complexity score of this string is approximately equal to 22.78 (see also http://www.complexitycalculator.com). The Kolmogorov complexity score does not tell us what the specific pattern is; it tells us how long the shortest program describing that pattern will be. Compression can thus also be considered a sufficient test of non-randomness [13].

Single and double bit sequences each share the same Kolmogorov complexity. Three bit sequences are the first bit strings can be structured into different groups of identical complexity. The researchers assume that at that point in time where this complexity changes the addition of the last data point marks the transition to the next pattern. Exemplary transitions are described in Table 1.

Table 1: Kolmogorov complexity groups [based on 14]

Sequence	Complexity	Group
111	5.40	1
000	5.40	1
110	5.45	2
100	5.45	2
011	5.45	2
001	5.45	2
101	5.51	3
010	5.51	3

3. Identification of complexity patterns

The earlier in the whole product life cycle of a novel high value manufacturing product an estimate is to be completed in the more the cost estimator is tasked to prepare an estimate without minimum prior information [14, 15]. In order to reduce this challenge the authors apply a "sliding window" approach in order to identify when a complexity group (pattern) ends as illustrated in Figure 2. The complexity is measured by Kolmogorov complexity (*Km*) for the binary string in brackets (i.e. "011) and calculated using the online tool www.complexitycalculator.com [13].

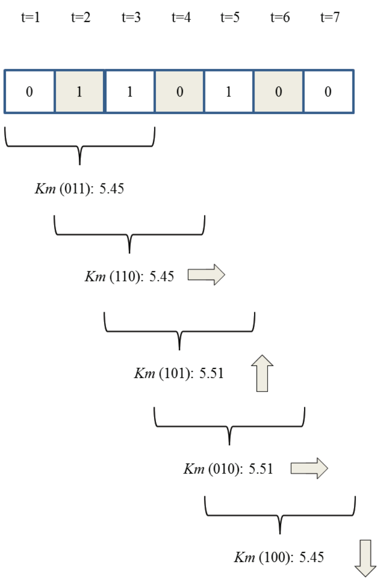

Fig. 2. Exemplary pattern separation

The change in complexity group can thus be interpreted as a "pulse" [16] as visualised by Figure 3.

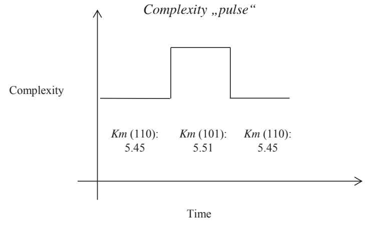

Fig. 3. Exemplary complexity "pulse"

The emphasis is hence moved from identifying the pattern itself to identifying that point in time where the pattern changes. The limitation is that at least three time periods of information need to be available before the method can be applied. The sliding window consists of overlapping three bit strings for which the Kolmogorov complexity is calculated and assigned to a complexity group. The interval in which a complexity group changes and then returns to its original group is considered to be the transitions phase between two signatures. In Figure 2 the string "0110100" is hence decomposed into overlapping three bit strings and the Kolmogorov complexity calculated for each. The initial complexity is 5.45 (three bit complexity group II) and rises to complexity 5.51 (three bit complexity group III) to then return to a value of 5.45 (three bit complexity group II). This change can be visualised as a "pulse" [16] indicating where the complexity change reached a threshold value to the next higher state, remained in the higher state for two sliding window periods and then again reach a threshold state where the next lower stage was passed to. Table 2 applies the method to case study data from the United States Department of Defence Scheduled Annual Summary Reports for the time period 1986-2013 [17]. The cost variance listed is the sum of absolute variance across the cost variance factors reported on for this period therefore that cost variance due to changes in quantity, schedule, engineering, estimating, other and support.

Table 2: Time, symmetry and complexity group trending

Time increment	Absolute cost variance (USD$ mil.)	Cost variance trend ("1" = increase; "0"=decrease or unchanged)	Complexity string (sliding window size 3)	Complexity group
1	112 733	N/A	N/A	N/A
2	85 882	0	0	N/A
3	115 081	1	01	N/A
4	92 968	0	010	3
5	84 783	0	100	2
6	90 068	1	001	2
7	55 148	0	010	3
8	64 580	1	101	3
9	45 418	0	010	3
10	52 484	1	101	3
11	63 285	1	011	2

12	85 939	1	111	1
13	101 016	1	111	1
14	117 376	1	111	1
15	127 229	1	111	1
16	162 505	1	111	1
17	177 869	1	111	1
18	201 927	1	111	1
19	245 456	1	111	1
20	159 672	0	110	2
21	263 012	1	101	3
22	257 726	0	010	3
23	264 185	1	101	3
24	290 521	1	011	2
25	289 536	0	110	2
26	242 056	0	100	2
27	142 301	0	000	1
28	81 752	0	000	1

4. Process for determining control mechanism

The process applied for determining whether continuous or discrete cost monitoring should be performed, including when forecasts can be made with a degree of subjective confidence can thus be described as follows:

1. Determine cost variance captured in continuous monitoring
2. Update last figure of binary string depending on change in cost variance
3. Update and evaluate the three interval binary string for its complexity group
4. Compare the complexity group to the complexity groups of the two previous time increments
    a. If the complexity group has changed then continuous monitoring is maintained and no forecast is possible
    b. If the complexity group has not changed then discrete monitoring can be commenced and a forecast is possible. The longer the period of constant complexity group the more stable the forecast can be considered to be.

Figure 4 illustrates the results of applying the data to the case study data:

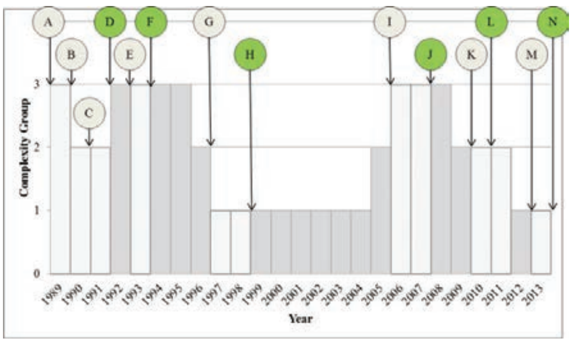

Fig. 4. Exemplary process application

Figure 4 illustrates the complexity group of the sliding window for cost variance data from 1989 to 2013. The figure is interpreted as follows:

- Point A: At the beginning of 1989 the first two time increments of data are available (not shown on the graph) and the sliding window approach can begin to be applied. Continuous monitoring commences since by default no patterns over at least two discrete time intervals are known at this point in time. A forecast is not possible.
- Point B: The end of 1989 the first complexity group is known. Continuous monitoring continues. A forecast is still not possible.
- Point C: At the end of 1990 the second complexity group is known and has changed from the previous year. Continuous monitoring therefore continues and a forecast is still not possible.
- Point D: At the end of 1991 the complexity group has remained unchanged for two years. Discrete monitoring can be commenced and a first forecast can be made.
- Point E: At the end of 1992 it is determined that the complexity group has changed. Continuous monitoring is resumed and a forecast cannot be made. The forecast made at the beginning of 1992 was valid for only a single year.
- Point F: At the end of 1993 the results of the continuous monitoring indicate that the complexity group is unchanged from the previous year. Discrete monitoring resumes and a new forecast can be made. The forecast continuous until the end of 1996 is arrived at.
- Point G: At the end of 1996 the complexity group changes again. Continuous monitoring is resumed and a forecast can no longer be made.
- Point H: At the end of 1998 two periods of stable complexity are again identified. Discrete monitoring is resumed and a forecast can be made.
- Point I: At the end of 2005 the complexity group again changes. Continuous monitoring is resumed and a forecast can no longer be made. The previous forecast remained valid for seven time increments.
- Point J: At the end of 2007 the complexity group has remained stable for two periods. Discrete monitoring is resumed and a forecast can again be made.
- Point K: At the end of 2009 the complexity group has again changed. Discrete monitoring is replaced by continuous monitoring and a forecast can no longer be made.
- Point L: At the end of 2010 two periods of stable complexity group are again determined and discrete monitoring is re-commenced. A forecast is again possible.
- Point M: At the end of 2012 a change in complexity group is again identified. Continuous monitoring resumes and no forecast can be made.
- Point N: At the end of 2013 two periods of stable complexity group are again determined and discrete monitoring is re-commenced. A forecast is again possible.

Green shaded circles therefore indicate when discrete monitoring begins and grey shaded columns indicate when forecasts can be applied.

Particular insights gained additionally are:

1. The complexity of the examined context exhibits a volatility that only seldom enables a transition from continuous to discrete cost monitoring and therefore forecasts are only seldom possible.
2. While two time increments with an identical complexity group are required before discrete monitoring might be considered applicable, only one time period is required to determine that a forecast is no longer valid and a return to continuous monitoring is warranted.
3. Complexity fluctuates only in single complexity group intervals, therefore these changes appear to be gradual versus disruptive.

Of importance to note is that the case study data aggregates the cost variance behaviour of many individual projects which have all already exceeded budgeted cost variances and are therefore being subjected to additional controls. While these in general remain on an annual time interval the monitoring process underlying the United States Department of Defence Selected Annual Reports does mandate quarterly reporting for projects that exhibit an even higher degree of cost variance [17]. The authors do suggest that continuous monitoring is the preferred method of introducing even more effective short interval control but accept that the ensuing resource demand and complexity may be insurmountable in practice.

5. Conclusions and recommendations for further research

This paper presented an approach for determining whether short interval (continuous) control of the cost estimate baseline of novel high value manufacturing products based on the principles of computation complexity should be used versus monitoring in discrete time intervals. The presented approach provides a clear and quantitatively derived answer to when patterns of data behaviour begin/end, when continuous versus discrete cost monitoring is required and when forecasts can be made (regardless of the level of confidence involved).

In order to generate a forecast minimum a priori data must hence be available at the time of the forecast. This means at least four discrete time intervals of data which confirm that a complexity group has achieved stability over two “sliding” time windows of three discrete time intervals.

The implications of the research for industrial practice are significant and currently suited for enhancing established cost estimation techniques in order to determine their utility in an applied context. Of greatest significance perhaps is that support is given for an estimator to declare that a forecast of cost estimate propagation is not feasible in the first place. This lack of feasibility is given by the lack of minimum prior data to apply any forecasting technique. This then leads to the position that the emphasis of cost estimation must be on helping to create a context of sufficient stability for such forecasting to occur in the first place. This “permission to doubt” or even “permission to refuse” a cost estimate is then expected to support a different paradigm in the way cost variance is treated in business decision making contexts in the first place. Additionally the variation of applicable cost estimation techniques introduced based on the Kolmogorov complexity of emerging cost variance bit strings places the cost estimation efforts central versus peripheral to the complete whole product life cycle. The greatest concern raised in this respect however is that established cost estimation practice in most organisations does not rely on the monitoring of cost variance reasons and their relevant mitigation but rather on a non-investigative accounting for variance in its place.

In the presented case study the underlying intervals of cost control remain constant (annual basis) while the investigation suggests that the intervals of relevance for effective cost control should fluctuate to a degree highlighted in Figure 4. Whether such an adaptation would result in overall lower cost variance volatility is unclear and warrants closer investigation in a controlled experimental environment.

The presented approach was developed based on the use of specific case study data related to the manufacturing sector. A wider application to other sectors may be relevant but warrants further research into the axioms needing introduction to achieve such a generalisation.

In respect to the patterns of interest it must also be remembered that the higher the Kolmogorov complexity score for a binary string the less redundant information is contained in such, hence the less visible the actual patterns of data behaviour are and the lower the expected fidelity of any forecast generated for it. The lower the Kolmogorov complexity score the less information is required to describe the binary string due to repeating patterns of data found in it. The more repeating patterns that are found the more simple the system can be interpreted to be. The simpler the system the greater the possible fidelity of a forecast is then assumed to be. The fidelity of a forecast can hereby also be understood as the degree to which we understand cause-and-effect relationships which lead to parametric cost-estimating-relationship models which can be relied upon for forecasting purposes. The challenge for stakeholders is hence to reduce the complexity of the context represented by the binary string in order to create conditions of stability that enable accurate forecasting in the first step.

Since the presented approach does not attempt to answer the question how to actually determine propagation patterns of relevance and thus how to forecast with acceptable accuracy, further research is also recommended into the fundamental question of how to determine a suitable length of forecast windows and what metrics of cost estimate uncertainty are most suited for forecasting purposes. An additional direction for future research is what approaches might be available for forecasting when cost variance data is available only for one to three discrete time intervals.

Acknowledgements

The authors wish to thank the members of the LinkedIn Group “Cost Risk and Uncertainty” [18] for their continued feedback and support and the anonymous referees for their constructive comments and helpful suggestions for improving this paper.

Key terms and definitions

This article is based on a number of important terms whose definition is provided in Table 3 for the sake of clarity.

Table 3. Key terms and definitions.

Term	Definition
Assumptions	The agreed state of the context the cost estimate is being performed in and for.
Baseline estimate	The agreed cost of producing a unit or delivering agreed support services. This cost consists of costed technical line items (often called the technical baseline estimate) and a risk contingency.
Complexity	As defined by Kolmogorov this metric quantifies the length of the shortest computer program that reproduces a specific binary string.
Complexity group	The Kolmogorov complexity shared by different binary strings of equal length.
Cost uncertainty	Manifested and unintended future cost variance with a probability of 100% and an unknown quantity.
Cost variance	Deviations from the baseline estimate.
Deep uncertainty	A decision-making situation where Knightian uncertainty, conflicting divergent paradigms and emergent decision making are relevant.
Forecast	Predictions of the future development of the baseline estimate.
High value manufacturing product	Products, production processes, and associated services which have the potential to create sustainable growth and high economic value.
Minimum a priori data	The historical cost variance known in advance of estimation which suffices for the application of standard regression techniques.
Novelty	A product attribute which exists when no verified cost estimates are available.
Open complex system	A group of dependent variables that form a purposeful whole, interacts with its environment and exhibits unpredictable behavior.
Pattern	Recurring behaviour of data as it propagates over time.
Prior information	The probability distribution function applied to a data set before the identification of relevant evidence.
Scenario	A future use case for a product or service for which a business model has been created.
Small data	Data sets which are significantly smaller than those encountered in daily practice and arise from a context of few measurement points, little prior experience, little to no known history, low quality and conditions of deep uncertainty.
Stability	The consistency of the complexity group over time.

References

[1] Technology Strategy Board (2012) “High Value Manufacturing Strategy”.
[2] Caron, F. (2013) “Managing the Continuum: Certainty, Uncertainty, Unpredictability in Large Engineering Projects”, Springer Briefs in Applied Sciences, ISBN 978-88-470-5244-4.
[3] Garthwaite, P.H., Kadane, J.B., O´Hagan, A. (2005) “Statistical Methods for Eliciting Probability Distributions”, Carnegie Mellon University, Research Showcase.
[4] Haskins, C. ed. (2010) “INCOSE Systems Engineering Handbook v. 3.2”, International Council on Systems Engineering, INCOSE-TP-2003-002-03.2.
[5] Schwabe, O., Shehab, E., Erkoyuncu, J.A. (2015) “Uncertainty Quantification Metrics for Whole Product Life Cycle Cost Estimates in Aerospace Innovation”, Journal Progress in Aerospace Sciences, pp. 1-24, DOI: 10.1016/j.paerosci.2015.06.002.
[6] Wierman, M.J. (2010) “An Introduction to the Mathematics of Uncertainty”, Center for the Mathematics of Uncertainty, Creighton University.
[7] United States Naval Center for Cost Analysis (2014) “Joint Agency Cost Schedule Risk and Uncertainty Hand Book”.
[8] Lempert, R.J., Warren, D., Henry, R., Button, R.W., Klenk, J., Giglio, K. (2016) “Defence Resource Planning Under Uncertainty”, RAND report.
[9] Makridakis, S., Bakas, N. (2015) “Forecasting and uncertainty: A survey”, Risk and Decision Analysis, Vol. 6, pp. 37-64 DOI 10.3233/RDA-150114.
[10] Mitleton-Kelly, E. (2003) “Complex Systems and evolutionary perspectives on organisations: The application of complexity theory to organisations”, Elsevier ISBN: 0-08-043957-8.
[11] Schwabe, O., Erkoyuncu, J.A., Shehab, E. (2014) “Long tail uncertainty distributions in novel risk probability classification”, Procedia CIRP, Vol. 28, pp. 191-196, 3rd CIRP Global Web Conference - Production Engineering Research — Advancement beyond state of the art .
[12] International Standards Organisation (2008) “Systems and software engineering - System life cycle processes”, ISO/IEC 15288, 2nd ed., IEEE Computer Society, Software & Systems Engineering Standards Committee.
[13] Soler-Toscano, F., Zenil, H., Delahaye, J-P., Gauvrit, N. (2014) “Calculating Kolmogorov Complexity from the Frequency Output Distributions of Small Turing Machines”, Preprint submitted to the Journal of Theoretical Computer Science, March 2015.
[14] RAND Corporation (2013) “Making Good Decisions Without Predictions. Robust Decision Making for Planning Under Deep Uncertainty”, Research Brief, RB-9701.
[15] Scales, J.A., Tenorio, L. (2001) “Prior information and uncertainty in inverse problems”, Geophysics, Vol. 66, No. 2, pp. 389–397.
[16] Shannon, C.E. (1948) “A Mathematical Theory of Communication”, The Bell System Technical Journal, Vol. 27, pp. 379–423, 623–656.
[17] United States Department of Defence (2015) “Selected Acquisition Reports (SAR) Summary Tables”.
[18] Cost Risk and Uncertainty (2016) LinkedIn Group.