Modelling the Evolution of Computer Aided Design Models: Investigating the Potential for Supporting Engineering Project Management

James A. Gopsill[1], Chris Snider[1], Lei Shi[2] & Ben J. Hicks[1]

[1]Dept. of Mechanical Engineering, University of Bristol, UK
J.A.Gopsill@bristol.ac.uk
Chris.Snider@bristol.ac.uk
Ben.Hicks@bristol.ac.uk

[2]Dept. of Mechanical Engineering, University of Bath, UK
Lei.Shi@bath.ac.uk

Abstract. The development of Computer Aided Design (CAD) models is a fundamental and distinct feature of Engineering Projects. CAD models can be considered to be the digital embodiment of the products' design and are used to support a wide variety of tasks that span the embodiment, detail, manufacture and commissioning phases of a project. With this in mind, it is proposed that the monitoring and modelling of the edit trace behaviour of CAD files may provide additional understanding and evidence that supplements current approaches to monitor and manage engineering projects.

To explore this proposition, this paper presents some results from an exploratory study that seeks to model the edit trace behaviour of CAD files based upon their meta-data attributes (for example, file size, date modified & date accessed). The edit trace behaviour has been mapped to a sigmoid function in order to be able to describe and potentially predict future behaviour. The potential impact this information may have on engineering project management is also discussed.

Keywords: Computer Aided Design, Sigmoid, Prediction, Evolution, Edit Trace, Engineering Project Management

1   Introduction

Within less than half a century, Computer Aided Design (CAD) software has developed to become an integral tool that supports engineers across many of their core tasks. This is further reinforced by the fact that the development of CAD skills is a core feature of engineering course syllabi and is increasingly being taught at a secondary school education level. In addition, the CAD industry has recently been estimated to be worth $7 billion U.S. dollars with revenues being distributed 37%, 38%, 21% and 4% for the Americas, Europe, Middle East and Africa (EMEA), Asia and the Rest of the World (ROW) respectively [1]. Further evidence of the ubiquity and importance of CAD is that of 2014 there are an estimated 19 million users.

An important factor in the success and uptake of CAD is the significant increase in the capabilities of CAD software, which has enabled CAD to support a vast array of engineering activities. From the initial objective of improving the accuracy and speed of 2-dimenional engineering drawings [2], CAD software is now more commonly associated with the development and handling of 3-dimensional geometry. Its fundamentality has been extended to handle the assembly of components, detection of interface issues, automatic generation of supporting documentation (for example, Bill-of-Materials), generation of standard parts, analysis of engineering systems, and support for meetings through the provision of models for to support collaborative discussions [3, 4]. Furthermore, there exists a wealth of software that integrates and/or utilises the models created by CAD software (for example, Finite Element Analysis, Dynamics Analysis and Computational Fluid Dynamics). Further, with the increasing interoperability of Product Data/Lifecycle Management (PDM/PLM) systems, it is argued that the increase in capability and ubiquity of CAD is set to continue.

It has also been acknowledged that the advances in CAD software have been a key enabler in the development and production of more complex products. Argyres [5] discusses how the development of the B2-bomber could not have been achieved without CAD tools to support the engineering project. More recently, Briggs [6] revealed that the development of the Boeing 787 Dreamliner generated approximately 300,000 parts being modelled in CAD and the associated PDM system typically saw between 75,000-100,000 accesses per week.

In addition to the increased product complexity, engineering projects have also increased in their complexity, which has been driven by the ICT, globalisation, and the need for specific skills. As a consequence, the management of engineering projects is becoming increasingly challenging. This is supported by a number of case studies highlighting that many large, multi-disciplinary and distributed engineering projects continue to overspend and overrun. For example, the development of the Airbus A380 initially saw a shortfall of €4.8 billion errors due to project overruns and the Eurotunnel was originally estimated at €2.8 billion but came in at €5.6 billion [7, 8].

While there are substantial bodies of work associated with improving project management via organisational management and improving product complexity management there are few – if any – approaches that bridge these two interrelated strands [9, 10]. It is proposed that due to the increasing reliance upon CAD as the primary digital embodiment of the product and its persistence across the majority of engineering activities, there exists a unique opportunity in being able to monitor engineering activities and the progress being made via the evolution of CAD files.

To investigate this opportunity, this paper presents the results from an exploratory study into modelling the evolutionary behaviour of CAD models. This paper first summarises the CAD dataset that has been analysed and then continues by discussing the analysis performed, whereby the fitting of a sigmoid function has been used in order to characterise the CAD file behaviour. This is followed by a discussion of the results where the key findings of the common characteristics and the predictive nature of the curve fits are described. The paper then concludes by discussing the potential impact this may have on the management of engineering projects and the ability to predict time to completion.

2   Study Context & Dataset Overview

The CAD dataset to be analysed has been captured from a Formula Student team at the University of *(omitted)*. Formula Student is a motor-sport educational programme whereby teams of students from competing universities create a single-seat race car that then competes in various challenges set-out by the competition organisers (Figure 1). The competitions are held worldwide including the UK, US, Australia and Europe.


Fig. 1. *omitted*

The creation of a Formula Student race car is highly multi-disciplinary involving students undertaking a range of engineering courses including: automotive, aerospace, electrical, manufacturing and mechanical. In the case of the *omitted*, the team consisted of 33 engineering students.

During the engineering project, a complete CAD model of the Formula Student car is generated. In order to manage this process, the *omitted* team utilise a custom designed lightweight CAD management tool that manages the naming convention, relationships and organisation of the CAD files. The CAD files are stored on a shared network drive that can be accessed by the teams' workstations.

To monitor the evolution of the CAD files, a Raspberry Pi – connected to the network – was used to monitor the status of the shared network drive at 20-minute intervals. More specifically, the folder structure alongside the meta-data attributes of the files stored where captured. This included file size, date accessed and date modified. The data capture was performed over a thirteen-week period and during this time, 892 CAD files were created and 8,264 updates were made to these files. Table 1 provides a summary of the dataset and also highlights the breakdown of the CAD file into their respective sub-systems.

Table 1. CAD Dataset Summary
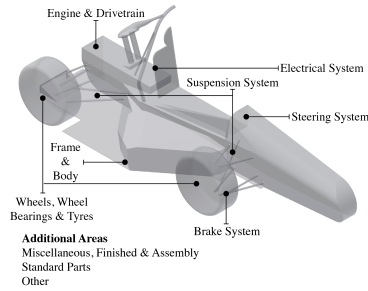
Description	Value	
Number of Weeks of Data Capture	13	
Total Number of CAD Files Created	892	
Brake System		17
Suspension System		188
Frame & Body		235
Engine & Drivetrain		237
Electrical		11
Steering System		71
Miscellaneous, Finish & Assembly		2
Wheels, Wheel Bearings & Tyres		35
Standard Parts		90
Other		6
Number of CAD File Updates	8,264	

Figure 2a shows the contribution of the CAD files to the total number of edits observed in the dataset. It is apparent that a relatively small proportion of CAD files represent a large proportion of the total number of edits. More specifically and of consideration in this analysis, are the 117 (20%) number of files that form 60% of all the edits. It is argued that these files would be of most interest for monitoring engineering activity due to the high number of edits made to them.

A summary of the CAD files of interest and the sub-system they pertain to is presented in Table 2 and it can be seen that the files of interest cover the entirety of the sub-systems involved in the development of the car. Therefore, it can be argued that the analysis of the subset of CAD files does not compromise on the coverage of activity occurring across the project.

Table 2. Distribution of CAD Files of Interest

Sub-System	Value
Brake System	6
Suspension System	42
Frame & Body	21
Engine & Drivetrain	29
Electrical	2
Steering System	5
Miscellaneous, Finish & Assembly	2
Wheels, Wheel Bearings & Tyres	3
Standard Parts	3
Other	5
Total	117

Figure 2b shows the subset of files selected for the analysis in relation to their CAD file life in days. It can be seen that the subset of CAD files to be analysed will encompass the CAD files with the total days in existence. These could be considered the most critical files to monitor as they likely form the assemblies where key areas of integration of components occur and files that transition across multi-disciplinary boundaries. For example CAD files that form the bodywork could also be utilised in the Computational Fluid Dynamics of the race car.
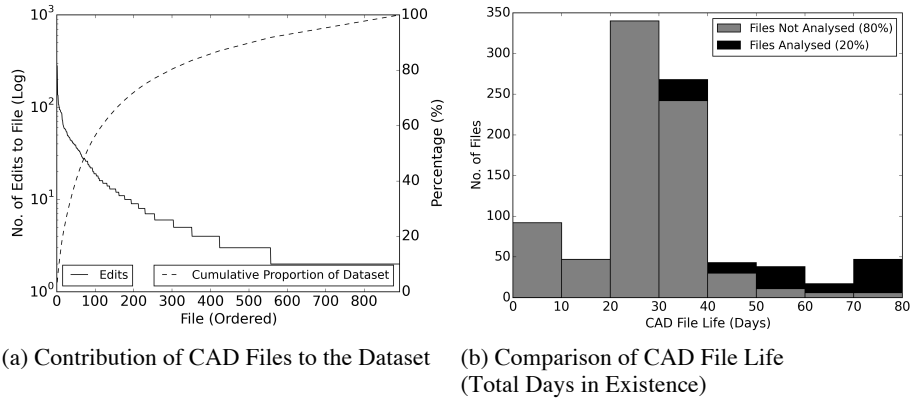
(a) Contribution of CAD Files to the Dataset

(b) Comparison of CAD File Life
(Total Days in Existence)

Fig. 2. Characteristics of the CAD File Dataset

Taking a closer look at the CAD files of interest, Figure 3 shows the cumulative frequency of edits. It can be seen that one file (i) clearly distinguishes itself from the others due to the total number of changes that has been made to it. On analysis of the file name, it had been indicated that this file is of the general assembly of the entire race car. The cumulative frequency plot also suggests the sigmoid like evolution of the CAD files and hence the proposition of using a sigmoid function for the curve fit. It appears a common trait that the CAD files are initially generated with few changes and then the activity ramps up to constant level of heightened activity before plateauing to a relatively stable final condition. As this is consistent for the majority of the CAD files, it could be considered the 'normal' profile of a CAD file and if the profile does not appear to reflect this then it may be an indicator of an anomaly.

Given the observation of the sigmoid-like evolution of the CAD edit traces, the paper presents the results of curve fitting using a sigmoid function to the evolution of the CAD file edit traces.
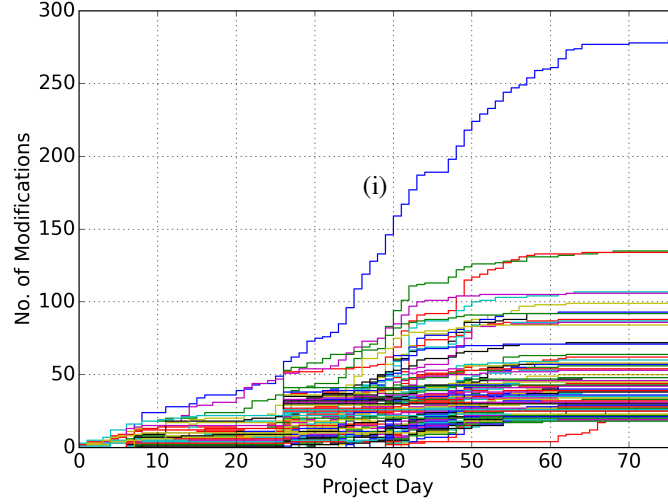
Fig. 3. Raw evolution traces of the top modified (117) CAD files within the Formula Student Dataset

3 Modelling the Evolution of CAD Files

In order to characterise the evolutionary behaviour of CAD files, this paper proposes the fitting of a curve based upon the sigmoid function (Equation 1).

$$y = \frac{a}{b + e^{-xc}}$$

(1)

As the CAD files were generated on different dates, a process of shifting the curves to the same datum position has to be undertaken. The results of this are shown in Figure 4a. This then enables the fitting of the sigmoid function to each of the CAD file data points using the least mean squares method for a curve of best fit (Figure 4b). An average curve R^2 value of 0.73 has been attained and 71% (82 files) of the CAD files of interest had an $R^2 > 0.90$. The high R^2 value provides confidence in the use of the sigmoid function as a lifecycle model for the majority of CAD files involving a large number of edits. It can also be seen that the erroneous curve fits in Figure 4b (i) are clearly out of scope of the likely progression of project given the rest of the curve fit population. Therefore, it is argued that it would relatively easy to determine whether a curve fit is likely to provide a suitable lifecycle model for a given CAD file.
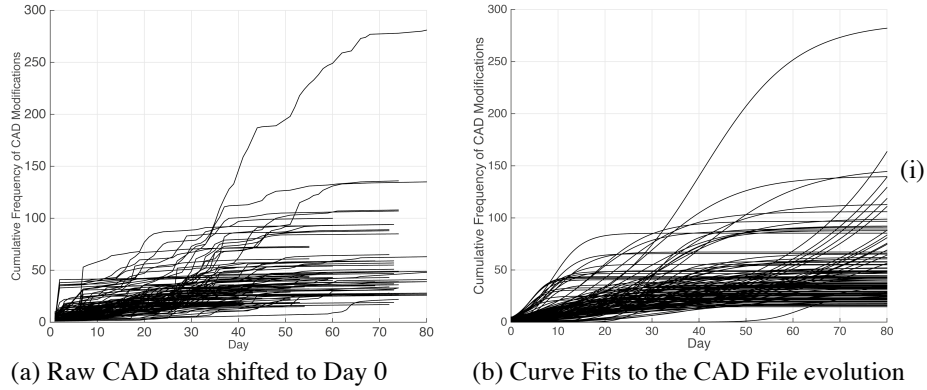
(a) Raw CAD data shifted to Day 0   (b) Curve Fits to the CAD File evolution

Fig. 4. Fitting curves to the evolution of the CAD file.

Figure 5 presents box plot distributions of the coefficients attained from the fitting of the sigmoid functions to the CAD files. It is apparent that the greatest variability lies within the (a) coefficient of the sigmoid function, whilst (b) & (c) have little variability in comparison. Although, there appears to be a long tail in the (c) coefficient. As the CAD files have not be assessed for their 'normality' in their generation, it may be that the (c) coefficient may be an key indicator of unusual CAD edit trace behaviour as the algorithm attempts to compensate for an edit trace that does not fit the lifecycle model.
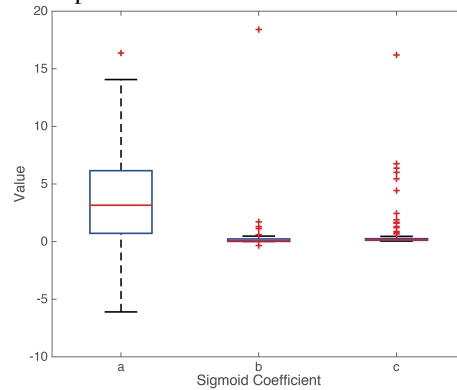

Fig. 5. Distribution of Curve Fit Coefficients

Given the range of coefficients typically seen in the evolution of CAD files, one can limit the range of possible options when performing a curve fit. Using the max-min range of $(5, 0)$, $(0.5, 0)$ & $(0.5, 0)$ for a, b & c respectively, the analysis continued into the assessment of the potential predictive power of a sigmoid curve lifecycle model to predict the future edit trace behaviour and time to completion of a CAD file.

Figure 6a reveals the accuracy of the prediction of a CAD file being completed in relation to the number of days prior to completion. It can be seen that the accuracy of the prediction is initially very poor at the early stages of the CAD files lifecycle although the accuracy quickly improves over time (Figure 6a, i). This can be attributed to the lack of data available as well as the fact the CAD file has yet to ramp up in update

activity. A key finding is that although initially inaccurate and erratic, as the CAD files reach halfway to completion (30 days prior to the final completion date) the prediction becomes highly accurate and consistent (Figure 6a, ii). This highlights that an indication of a completion date could be made significantly ahead of time and may be potentially useful information for project management.
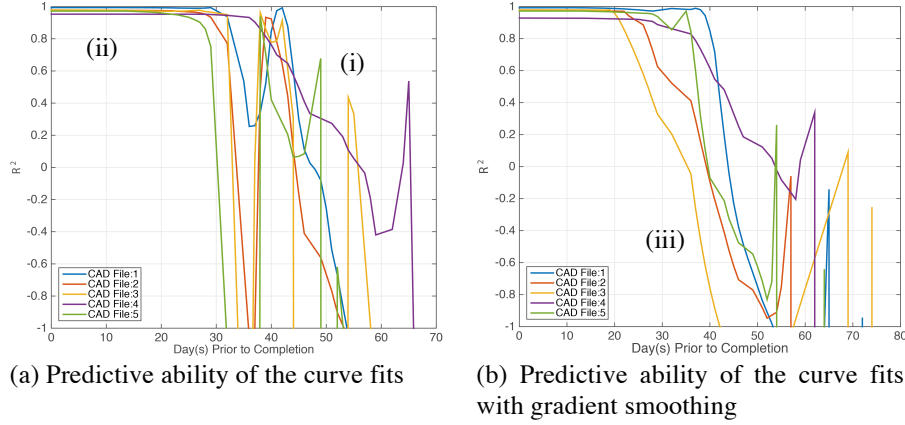

(a) Predictive ability of the curve fits

(b) Predictive ability of the curve fits with gradient smoothing

Fig. 6. Curve fitting results

In order to combat the sudden variation in the curve fits, Figure 6b shows the results from the introduction of a permissible margin of change from one curve fit prediction to the next. In this case, the margins were set to $0.05, 0.01$ & 0.01 for a, b & c coefficient respectively. Using the margins of change, it can be seen through the comparison of Figures 6a & 6b that the sudden drop of in predictive power of the curve fit is eliminated and a more consistent prediction is produced (Figure 6b, iii). However, this appears to be at a detriment of the predictive power of the curve fit in the early stages of the CAD file edit trace. It is also important to note that this analysis is not only assessing the accuracy of the final predicted time to completion but also for the CAD files entire edit trace. Thus, it can be used to monitor whether the CAD file is evolving along the expected path.

4   Discussion and Future Work

From the results of this exploratory study, it has been shown that the majority of CAD files follow a sigmoid curve of evolution whereby the file is initially instantiated, which is then followed by a period of high activity that finally plateaus to the final version of the file. Given this identification of a potentially 'normal' evolutionary routine, it is proposed that real-time monitoring solutions to assess file evolution are possible. Further it is suggested that these could provide indications of key project events/issues to project management in a more responsive and immediate manner.

Continuing to the element of the prediction of CAD file evolution, it has been demonstrated that there is potential in the ability to generate predictions. It has been shown that reasonably accurate predictions ($R^2 > 0.9$) of the edit trace path and time to

completion can be made up to 30 days in advance. The relative high level of conformance of the edit traces of the CAD files might suggest that conformance to the sigmoid function could be a useful indicator of normality. Thus, the testing of conformance through the fitting of a sigmoid function could potentially detect anomalies or issues that may require managerial attention.

These insights could have a profound effect on the management of engineering projects and their ability to monitor progression (Figure 7). With this prediction alongside expert opinion & discretion of project managers, it is contended that this could provide evidence to support project managerial decisions and interventions.
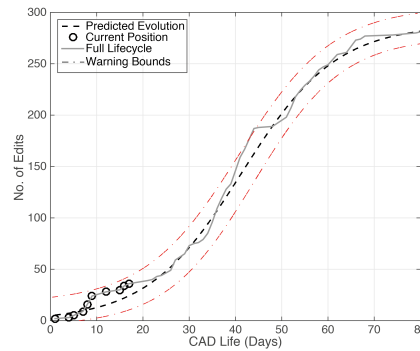

Fig. 7. CAD file prediction with potential warning bounds

In addition, the initial fitting of the sigmoid function to the emerging edit traces of the CAD files revealed considerable fluctuations (low stability) in the prediction of the future trace. This was mitigated through the addition of permissible margins of change of the sigmoid coefficients from the current prediction iteration to the next. The strategy improved the stability of the prediction although this has been at the detriment of the accuracy of the early edit trace prediction. It is argued that future work could seek to address this through a dynamically changing permissible margin given the current stage in the lifecycle of the part. In the early stages the margin could be set to be wider and then to slowly converge as the CAD file continues to develop.

Finally, it is key to note that such analysis has been performed on the meta-data of the CAD files and is significant in the fact that the analysis could be applied to any PLM/PDM system. Future studies into this area could benefit from a study whose CAD files are coded by their relative 'normality' in generation as determined by the engineers. In addition, future analysis could also consider the content of the CAD files, which may provide further and more detailed insights into their evolution and as a consequence, the state of an engineering project.

5 Conclusion

Computer Aided Design files are a fundamental feature of engineering projects and are the digital embodiment of a products' design. With CAD files being used to support a wide variety of engineering tasks, this paper sought to investigate whether their

evolution – in terms of their edit traces – could be characterised and predicted, and in turn be used to support project management.

From the analysis of 892 CAD files generated from a Formula Student project, it has been shown that 60% of all the edits come from 20% (117) of CAD file corpus. Taking these as the CAD files of interest, it has been shown that >70% can be characterised by a sigmoid function with an $R^2 > 0.9$. Thus, it is argued that sigmoid functions can be used as a lifecycle model for highly edited CAD files.

The prediction of the curve fits has also been investigated and revealed that accurate predictions of the time to completion and the expected edit trace can be made up to 30 days prior to their completion. The stability of this prediction has also been improved by the introduction of a permissible margin of change between iterations of the prediction.

Being able to provide this information alongside expert opinion & discretion of project managers, it is contended that this could provide evidence to support project managerial decisions and interventions.

References

1. John Peddie Researcrh, Jon Peddie Research releases the Worldwide CAD Market Report 2012, Retrieved from http://jonpeddie.com/press-releases/details/worldwide-cad-market-report-2012/
2. Ivan E. Sutherland. Sketch pad a man-machine graphical communication system. In Proceedings of the SHARE Design Automation Workshop, DAC '64, pages 6.329–6.346, New York, NY, USA, 1964. ACM. doi: 10.1145/800265.810742
3. J.A. Gopsill, H. McAlpine, and B. J. Hicks. Learning from the lifecycle: The capabilities and limitations of current product lifecycle practice and systems. In International Conference on Engineering Design ICED'11, 2011.
4. David Robertson and Thomas J. Allen. CAD system use and engineering performance. IEEE Transactions on Engineering Management. 40(3):274-282, 1993
5. Nicholas S. Argyres. The impact of information technology on coordination: Evidence from the b-2 stealth bomber. Organization Science, 10(2):162–180, 1999. doi: 10.1287/orsc.10.2.162. URL http://orgsci.journal.informs.org/content/10/2/162.abstract.
6. David Briggs. Establish digital product development (DPD) Low End Viewer (LEV) and archival standard for 787 project. Collaboration & Interoperability Congress (CIC), 2012
7. Nicola Clarke, Airbus Faces More Delays for A380 Jet (The New York Times), Retrieved from: http://www.nytimes.com/2006/10/04/business/worldbusiness/04airbus.html?_r=0, Oct 4th, 2006
8. Wallis Shani, Channel Tunnel hindsight, Retrieved from: http://www.tunneltalk.com/Channel-Tunnel-Feb93-Hindsight-view.php , Feb 1993
9. Söderlund, J. (2004). Building theories of project management: past research, questions for the future. International journal of project management, 22(3), 183-191.
10. Kedar, A.P.; Lakhe, R.R.; Deshpande, V.S.; Washimkar, P.V.; Wakhare, M.V., "A Comparative Review of TQM, TPM and Related Organisational Performance Improvement Programs," Emerging Trends in Engineering and Technology, 2008. ICETET '08. First International Conference on , vol., no., pp.725,730, 16-18 July 2008 doi: 10.1109/ICETET.2008.133