BAYESIAN PROJECT MONITORING

Peter C Matthews and Alex D M Philip
School of Engineering and Computing Sciences, Durham University

ABSTRACT
This paper studies how subtle signals that can be observed from the execution order of a project with several tasks can be used to diagnose potential problems that will hinder the project. Specifically, by representing the workflow of the project as a Markov Chain and observing how long the project takes to arrive at its first gateway, this research shows that there is the potential to infer the nature of any problems with the project. This diagnosis is achieved through using Bayesian methods, and provides a ranked list of candidate causes, along with the probability for each cause. Two examples are used to illustrate how this approach works.

Keywords: Project Management, Uncertainty, Markov Chains, Monte Carlo Simulation

1 INTRODUCTION
Building construction projects are mostly linear processes, but can iterate when faults occur. These iterations add time and cost to the construction process, especially if the iteration was not anticipated [1]. There are several reasons for unanticipated iteration in a construction project, including poor brief, political considerations, insufficient budget, and so on. Frequently, the disruption to the project occurs later on in the execution of the project [2]. In these events, it would be beneficial to be provided with early warning that a type of problem was anticipated. If some signal could be observed early on in the process, the project manager could be made aware of the potential problem and take some suitable mitigating action.
This research seeks to exploit Bayesian methods to interpret signals based on the progress of a project to generate predictions of potential problems. The Bayesian method will compute the set of probabilities that certain problems exist. It will then be for the project manager to interpret these results in the broader context of the project execution as to what action would be appropriate.

2 BACKGROUND
The construction industry follows a (mostly) linear process. This process starts with the project inception and definition and is expected to terminate with the completion of the building. The process can terminate in other ways, however these represent project failures, as the project has come to some conclusion other than completing the building. Within this process there are also gateways [3]. These gateways ensure that the project has reached sufficient maturity and quality that it may proceed to the next phase of the project. In the event that a gateway blocks the project, two outcomes are then possible: either the process must return to the start of the phase or the project is terminated (i.e. it fails).
For the purposes of this work, the Royal Institute of British Architects (RIBA) construction process is adopted [4]. Figure 1 contains the earliest part of that process, represented as a flowchart. This flowchart starts with the project inception and terminates at the granting of building permission. The process then continues with the construction of the building, but for the purposes of this work it is sufficient to only consider the earliest phases.
This flowchart can be thought of as a Markov Chain [5, 6]. Each element of the flowchart can be represented by a node. From each node, there are a number of nodes the process could progress on to which are represented by the directed arcs exiting the node. For example, from the ‘Identify Site’ node, the process could move on to any of: ‘Outline Objectives’, ‘Determine Budget’ or ‘Project Definition’. When simulating the process, the node that the process moves on to is determined stochastically. Specifically, each arc has a probability of being followed. Based on these probabilities, an arc is selected at random thereby taking the process on to the next step. For the simulation model, these probabilities were estimated through a combination of literature and discussions with domain experts.

2.1 Construction Design Process

The construction design process is a variant of generic product development, as described by for example [8]. The fundamental aspects of this process are that is mostly linear, but divided into major sections (phases) that are delimited by stage gates [3]. These stage gates provide the opportunity to review the progress of the project and determine if it should go ahead, require further work within the current stage or be terminated. This ensures that 'bad' designs are not taken through to further downstream phases thereby wasting resource or risking failure [9]. The construction design process first key steps before any physical construction occurs are:

Firstly, the client proposes an idea for a building project and it is then up to the architect to transform this idea into a practicable building solution. To achieve this successfully, the design must satisfy the client's requirement for both the functionality and aesthetics of the building.

Secondly, the financial considerations of a building project are common. Typically there is a fixed overall budget with very little leeway for the total cost. The client will have expectations of what can be achieved for their budget and it is the architect's responsibility to provide solutions that can maximise what can be achieved for that budget.

Thirdly, the consideration and adherence to legal issues is common. There are significant numbers of legal regulations that buildings must adhere to. Although there will be different specialised legislation for different types of building projects, the overall process of adhering to legislation remains common.

In the UK, the building process is governed by a set of rules detailed by RIBA. These rules divide the construction process into several stages. The main stages are defined as: (1) Preparation, (2) Design, (3) Pre-construction, (4) Construction, and (5) Use. Each main stage is then further broken down into smaller work stages. This work will focus on the first two of the RIBA main stages: Preparation and Design. These stages occur before significant resource has been invested into the project, and therefore it is during these stages where it is easiest (most cost-effective) to change the design. The reasons for changing design could be for any number of reasons, including due to misinterpretation, change in budget, or change of opinion.

Clough, et al. [10] note that the 'Planning and Definition' stages of the project must define the requirements and (budgetary) constraints. The project definition must include "establishing broad project characteristics such as location, performance criteria, layout, equipment, services and other owner requirements needed to establish the general aspect of the project". The design phase involves completing the architectural and engineering design of the entire project. This results in the production of the final working drawings and specification of the total construction programme.

Ritz [11] states that the most critical stages of the pre-construction phase are the planning for construction execution and resource usage (time, money, equipment). In projects where these aspects are neglected, there is a greater risk in the project of failure at a later stage due to overruns of time and/or money.

2.2 Uncertainty within Construction

Construction projects frequently overrun, measured in terms of either time or financial resources. Given that the budgets are set before any (physical) work is done, this is perhaps not too surprising. Essentially, the early phases of the construction process are to formulate an executive plan of work. The resources required for the various tasks involved are based on estimates and assumptions of how well the work will progress. These estimates and assumptions form the first source of uncertainty.

The construction process is complex, and contains a number of actors (e.g. client, architect, builder, planners, etc.). Although the interaction between these actors is defined in terms of when they should occur and how they should proceed, there is no guarantee that this will happen. Further, the quality of the interaction is determined by the abilities of each actor. Ineffective or poor actions taken by certain actors will generate unacceptable work or fail to meet predetermined deadlines. Unacceptable work will require rework, which in turn will cause delays [1]. These events occur seemingly randomly (although will be biased by the capabilities of the various actors), and hence represent the second source of uncertainty.

Finally, the construction process occurs outdoors. In this context there are external events that are beyond the control of any of the actors within the construction process, such as extreme weather. This introduces the third and final source of uncertainty.

2.3 Conclusion

The general construction process can be reasonably modelled as a Markov Chain, with probabilistic transitions from one activity to the next. Although for any specific project there will have been deterministic reasons for moving from one activity to the next, this is not important when considering a large set of projects which will appear to randomly move the process. Further, given that the exact nature of uncertainty in a *current* project is unknown, the best estimate that can be made for the transition probabilities at the outset are given by prior probabilities. Specifically, when thinking about the future possible directions for a project that is about to start, a stochastic approach is suitable. As further information is gathered, this can be used to refine the understanding of the nature of project under execution.

3 PROCESS MODELLING METHODOLOGY

The process modelling methodology is based on a Markov Chain [12–14]. A Markov Chain consists of a set of nodes which are linked by directed arcs. The temporal domain is modelled discretely, namely that one action happens per discrete time step. The nodes represent activities and the arcs then represent the possible subsequent nodes the process could move to in the next time step, potentially including a 'loop-back' arc, which models the process remaining in the same state at the next time step.

One aim of this work has been to make this approach as broadly applicable within the construction industry. In order to achieve this, the key generic process stages in the early construction process were identified, based upon the authoritative RIBA framework. These stages are represented as the nodes in the Markov Chain (Figure 1). The nodes are:

A Project inception and definition: the 'official' start of the project; formulating the core ideas of the project
B Identify site: selection of the project location
C Outline project objectives: specification of the fundamental design ideas
D Determine budget: identify the total funding for the project
E Appoint design team and Project Manager: identify the team of individuals who will execute the construction project
F Outline project programme and risks: define the various stages within the construction project, and associated expected time and cost
G Develop ideas and preliminary sketches: creating the first designs, including architectural drawings

At the outset of a project, it is assumed that none of these tasks have been started. Each node contains a completion status, and hence at the outset of the project these are all set to 'false' (incomplete). This is an extension to the pure Markov Chain which is 'memoryless'. The addition of 'node-memory' through a status variable provides a more intuitive model for the construction process where a number of tasks can happen in parallel and must all have been completed successfully. Therefore, the node status can be used at various stage gates to ensure that all relevant tasks have been completed and that the project may proceed. To implement this, it was also necessary to develop 'gateway' nodes that were able to verify the completion status of prior nodes. These gateway nodes were inserted at key stages within the construction process. The key stages are where there is a significant transition in the project. Within the RIBA framework, this occurs when moving into and out of RIBA Stage C, and is labelled GR1 in Figure 1. This gateway represents a review of the project status, and if the project is not in a satisfactory state it sends the project back to the start. The transitions from the gateway node are again modelled stochastically, however should the gateway review fail, the process returns to the earlier state with status aspects intact (for example the site will remain identified). The result of this is that a different set of transition probabilities come into effect. This different set of transition probabilities represents the fact that some work has been done on the project and this will impact how the process is likely to flow through the nodes second time around.

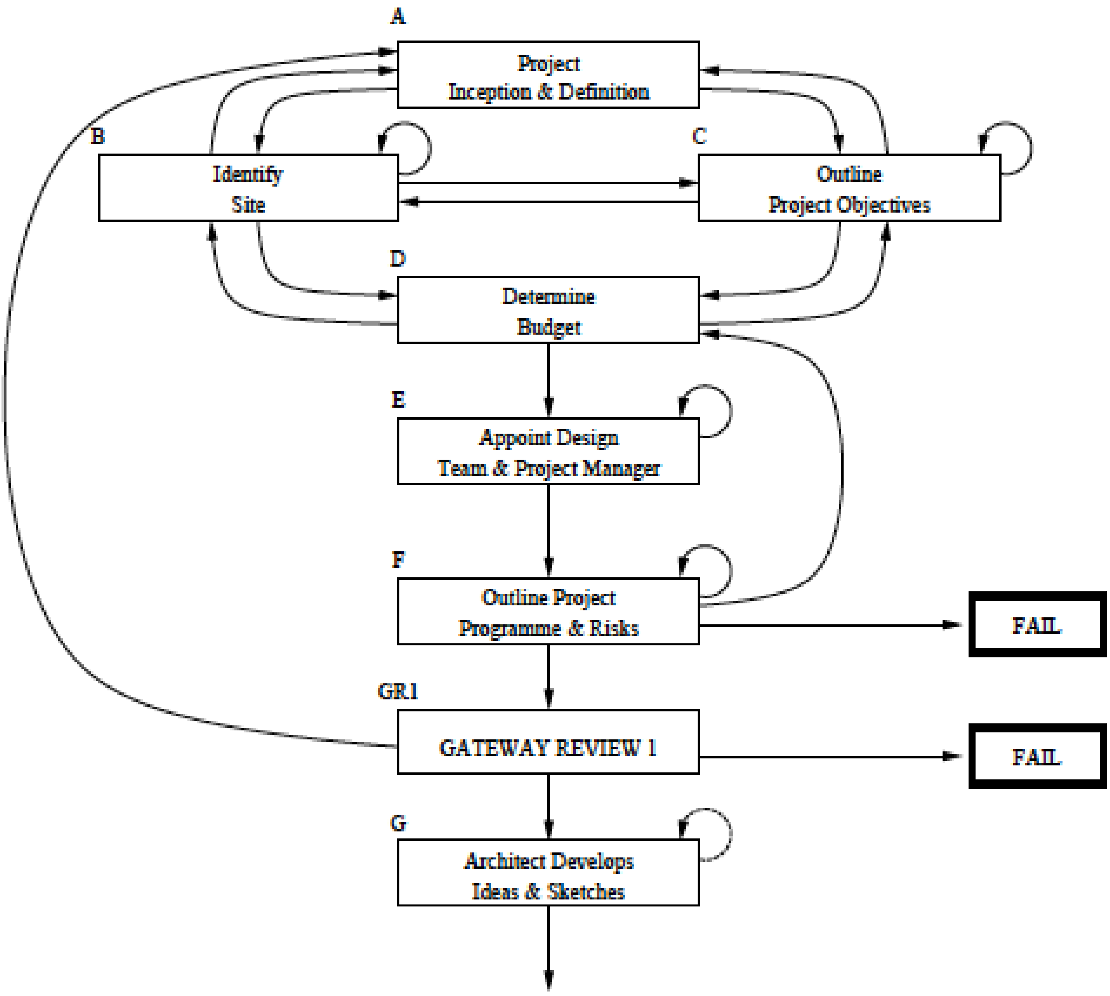

Figure 1: Markov Chain implementation of the RIBA process for this study.

In addition to the activity nodes, it is also necessary to include terminal nodes. These are nodes that from a Markov simulation perspective can be entered but can never be left. When the simulation process enters one of these nodes, that simulation run terminates. For this simulation, there are two types of terminal nodes: success and failure. The success terminal node is entered when all tasks in the construction process have been successfully completed and in this case represents that planning permission has been granted. Upon granting planning permission, the construction project is able to proceed with the physical construction. The failure nodes represent where the project is cancelled. There are a number of points within the construction process where it is possible to cancel the project, and hence there are a number of different failure terminal nodes. It is assumed that once a project enters a failure node, there is no possible remedial action to be taken and that therefore the project is completely abandoned. It is also theoretically possible due to the cycles that exist in the Markov chain that a process takes can take an arbitrary number of hops to reach a terminal node. Therefore, if a process takes more than a given number of hops (60 in this case), the project is considered to have failed.

Once the nodes have been determined, the next step is to connect the nodes with directed arcs. The ‘normal’ progression of the project, as determined by literature (e.g. RIBA), is straightforward to implement. These arcs determine how the project would progress if all went well: all activities are successful and there is no need for any rework. However, for a complete simulation framework, it is also necessary to represent how the process progresses when activities are not successful. This is modelled by arcs linking a node to a ‘previous’ node, or possibly as a loop-back to the original node. It is worth noting that feedback loops do not feedback any further than the gateway ‘above’ them. This is as it is assumed that once the project has passed a gateway, all the tasks prior to that gateway have been successfully completed and therefore do not need revisiting.

3.1 Parameterisation of the Markov Chain transition probabilities

The construction process simulation model (Figure 1) contains a set of activity nodes and arcs connecting these nodes. Each node is connected to a minimum of two and a maximum of five other nodes. These connections represent the outcomes possible from each activity node. The parameterisation of the model is the association of a probability distribution for each node that represents the probabilities for following each arc. These probability values determine how the simulation proceeds through the construction process model [6]. Therefore, it is important that the probability values are as realistic as possible. Specifically, if the probability of a task being successfully completed is too high, then the simulation will register that is takes fewer attempts to achieve this task than is reality. Conversely, setting a probability that a task is completed too low will result in the simulation reporting that the overall process time is higher than in reality.

The estimates for the probability values in this work were arrived at through a combination of literature survey (performed earlier by Mason [16]) and expert estimation/refinement. This estimation process was first undertaken for the 'Nominal' scenario (where no significant negative influences exist). Using the Nominal scenario probabilities, the probability transition tables were estimated for the remaining scenarios. These other scenarios represent departures from the Nominal scenario, and using the scenario characteristics detailed below, revised transition probabilities were estimated. This is expanded on in the following section.

3.2 Scenario development

The primary aim of this work is to seek for warning signals given by construction projects that are at risk of performing poorly. A signal is defined as an observation that could be linked to a potential future problem. Observing these signals will provide the opportunity to take preventative action to ensure that the problem does not impact the project in the future. The signals that will be observed in this work are temporal: specifically how long the project takes to enter certain nodes. For example, consider the following scenario: a construction project has progressed to the point where it had detailed the building material and identified the source of for the material. However, due to unforseen circumstances, the preferred building material supply company goes out of business. The builder must now identify a new preferred supplier and potentially review the design if certain requested materials are no longer available. This adds to the time it takes the project to progress to the next gateway review, and it is this longer than expected time that is observed as the signal.

The basic scenario that is considered is where all tasks have nominal, or 'good', progression probabilities. This represents a realistic ideal case, where there might be some rework necessary, however this is not the result of some fundamental problem within the project. In addition to this nominal scenario, a set of common problems were identified and used as a basis for developing problematic scenarios. Below are the details a set of problematic scenarios with their associated design issues:

Over Complex Design: Underground conditions may be an issue; Regulation needs are more complex; Unrealistic time demands; Poor management due to complexity; Project complexity is increased

Small Budget: Supplier problems; Client requirements change; Cost cutting reduces quality; Over ambitious quantity estimations; Poor management: attempting to achieve results with insufficient resource

Difficult Planners: Political considerations make progress more difficult; Increasing Health & Safety needs; Regulation needs result in harder to complete tasks; Disputes may cause delays

Poor Brief: Client requirement changes; Brief continuously being modified; Poor client visualisation due to lack of clarity; Errors by the construction team; Designer decisions that the client does not like

Tight Schedule: Unrealistic time demands; Mistakes due to rushed work; Poor workmanship; Poor management

For each scenario the nominal transition probability table was modified to represent the changes in how the process would proceed. It should be noted that only the transition probabilities were modified, not the structure of the Markov model.

4 BAYESIAN BASED SCENARIO IDENTIFICATION

The premise of using the process model as a means of identifying what potential scenario is being played out is based on Bayesian theory. For example, let the signal be how long it takes the project to pass the first gateway and denote this λ, which is the integer value representing the number of hops from the start of the project. Using Bayes, it is then possible to compute the probability for being in each scenario (S_i) given this signal [15]:

$$P(S_i \mid \lambda) = \frac{P(\lambda \mid S_i)P(S_i)}{\sum_j P(\lambda \mid S_j)P(S_j)} \tag{1}$$

This equation can then be used to determine the most likely scenario once the signal (λ) has been observed. The conditional probability $P(S_i|\lambda)$ is computed for all possible scenarios S_1, S_2, …, S_N. These scenarios can then be ranked according to their associated conditional probability score. This ranked list can then be used by a decision maker, for example the project manager, to further investigate the root cause of any difficulties.

To be able to successfully use Bayes' theory, as expressed in Equation 1, there is also a need for further probabilistic information. Specifically, there is the need to know the prior probability of each scenario, $P(S_i)$, and the conditional probability distribution of λ for each scenario, $P(\lambda|S_i)$. The method for obtaining this information is detailed in the following sections.

4.1 Model Calibration

The Markov process model is used to simulate the early construction process. By repeating a number of such simulations using a Monte Carlo approach, a distribution of time taken to pass the gateway review can be generated (see Figure 2). For each scenario, a different set of transition probabilities are used. It is desired to obtain an analytical model of the underlying conditional distribution, and hence it is necessary to calibrate the Monte Carlo simulation results with the analytical model. The calibration of the model is in effect determining which conditional distribution function $P(\lambda|S_i)$ best fits for each possible scenario. It is assumed that the underlying model for this conditional distribution is either a Poisson distribution or a Normal distribution. This is reasonable: the Poisson distribution models the time taken for an event to occur and is parameterised by a single value representing the expected duration, whereas the Normal distribution is a good distribution where averages are taken over larger event samples. For the construction Markov Chain, it can be thought that successfully passing a gateway is in effect waiting for an event to occur and hence suitable for being represented by a Poisson distribution. However, due to the potential of having several cycles in the Markov process, this represents a more general category and the frequency distribution here might be better represented by the Normal distribution.

Therefore, model calibration in this case is simply a matter of identifying which distribution best fits the simulation data. For each distribution, the best fit is determined by the simulation data's mean and standard deviation. The best fit is determined by measuring the χ^2 statistic.

4.2 Application of Model

Using Equation 1 with the appropriate probability distribution, the model estimates which scenario is most likely to be occurring based on the single observed value of how long it takes to pass the first gateway, denoted by λ. This is done for each scenario, each time using the appropriately parameterised scenario model. As there is only one observed value, λ, it is possible to generate a lookup table for a range of values of λ and then rank the possible scenarios. These lookup tables can then be used by the project manager as a guide for diagnosing a construction process based on the time taken to pass the first gateway node.

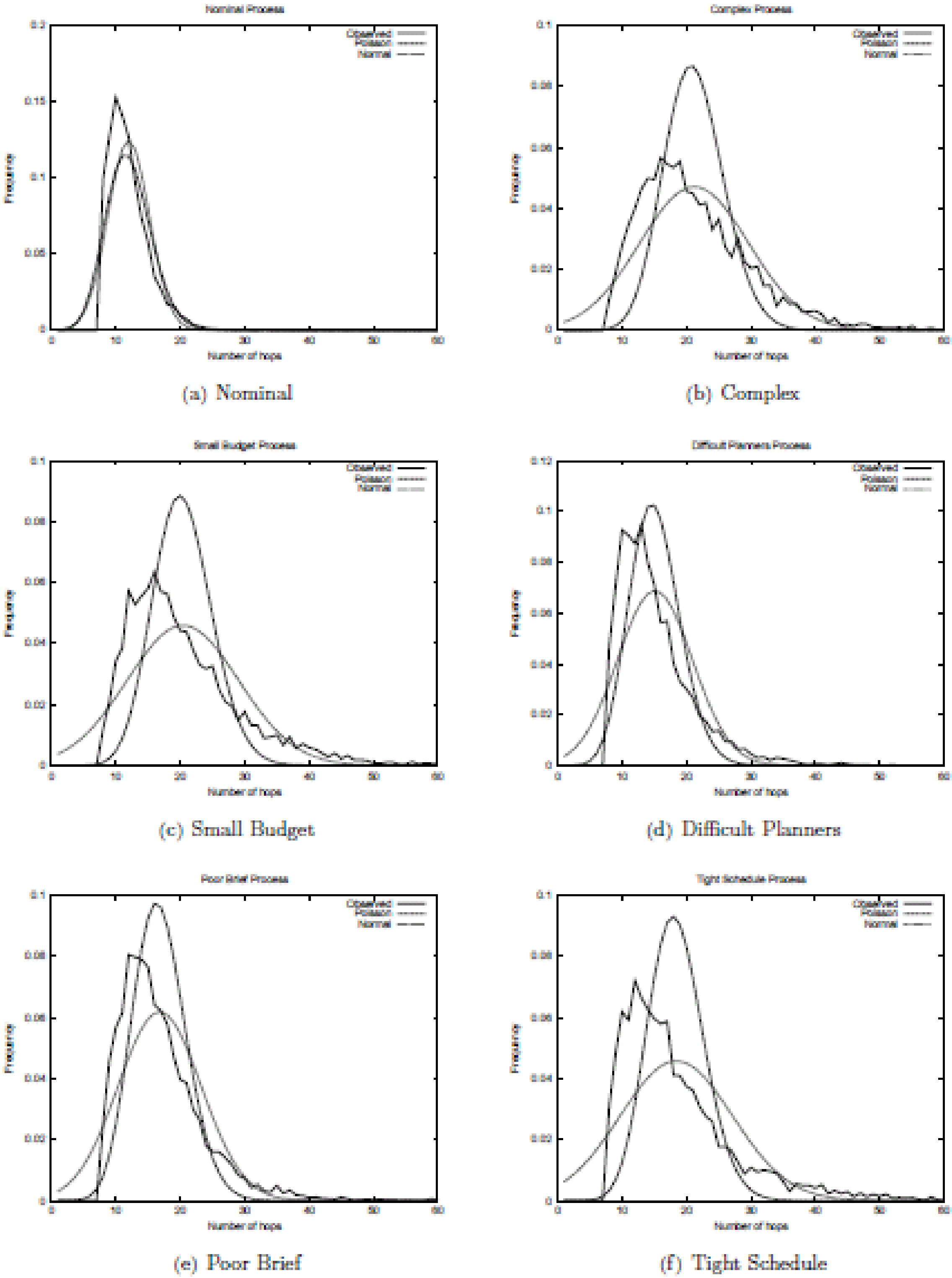

Figure 2. Calibration of the model: for each scenario, the simulation data is plotted along with the best estimate for the Poisson and Normal distributions.

5 ILLUSTRATION

To demonstrate how the project monitoring systems is used, two illustrations are given. These are based on fictitious cases, primarily focusing on how long the construction process takes to pass the first gateway review node. The first illustration is a case where the project has swiftly moved through the first gateway. The second illustration is a case where there have been more delays and the process has taken longer to pass through the first gateway. In both cases, the project manager is unsure if there are underlying problems and if so what they might be. In each case, the aim is to illustrate how the system can be used to provide a ranked list of potential problem sources to the project manager. It then remains to the project manager to decide how to use this information.

5.1 Nominal Scenario

In the first scenario, the construction process progresses swiftly through to the first gateway. Specifically, in this scenario it is observed that the process passes the first gateway review after 11 'hops', i.e. that in this case λ=11. Table 1 provides the conditional probabilities for each scenario.

Table 1. Ranked list of predicted scenarios for the case λ=11 and associated conditional probabilities.

Rank	Scenario	$P(S_i \vert \lambda=11)$
1	Nominal	0.36
2	Poor Brief	0.31
3	Tight Schedule	0.17
4	Planning	0.06
5	Small Budget	0.06
6	Complex	0.06

From Table 1 the project manager can determine that this project is most likely running without any significant problems (the Nominal case). If the project manager believes that this might not be the case, the system suggests that the next most likely scenario is that the brief is poor, followed by a tight schedule. The project manager can use the associated probabilities to provide guidance as to the relative likelihood of these scenarios. In this case, it can be noted that the poor brief is more than twice as likely as the tight schedule scenario. Therefore, if the project manager suspects that the project does have problems, the brief in this case would be the most likely scenario.

5.2 Problem Scenario

The second scenario is one where the construction process has more iterations, and therefore takes longer to pass the first gateway. In this scenario it is observed that the process passes the first gateway after 26 hops, i.e. that λ=26. Again, the ranked order can be read and Table 2 provides this ranking with the associated conditional probabilities.

From Table 2 it can be seen that the top ranked scenario is that the schedule is too tight. However, it is worth noting that the second most likely scenario, a poor brief, has only a slightly lower probability of occurring. With this additional information regarding the closeness of scenario probabilities, a project manager should investigate both scenarios. The third ranked scenario, Complex Project, is sufficiently distant that a project manager need only investigate this should the top two scenarios prove not to be the case.

Table 2. Ranked list of predicted scenarios for the case λ=26 and associated conditional probabilities.

Rank	Scenario	$P(S_i \vert \lambda=26)$
1	Tight Schedule	0.3262
2	Poor Brief	0.3260
3	Complex Project	0.1669
4	Small Budget	0.1554
5	Planning	0.0246
6	Nominal	0.0011

6 DISCUSSION

The Bayesian project monitoring support system assumes that the underlying distributions for the time taken to arrive at the gateway for each scenario can be represented by either a Poisson or Normal distribution. The simulation process is used to estimate the parameters of those models. For most scenarios, the χ^2 test suggested that this assumption was reasonable, as determined by sufficiently low χ^2 values. However, for both the Difficult Planner and Poor Brief, the χ^2 test statistic for both distribution models was very high. This is not too critical for the Difficult planner scenario, as this occurs relatively rarely. However, for the Poor Brief, which is the most common scenario, this potentially poses a problem.

7 CONCLUSION

This study has shown that by using a Bayesian approach, there is the potential to estimate the scenario that a process is running in. Using this information, a project manager is able to take pre-emptive action to mitigate the impacts that are likely due to that scenario.

There remain challenges with this work. The key challenge lies with the parameterising of the underlying conditional probability distributions. In this study this was achieved by modelling the process as a Markov Chain, and then using a Monte Carlo approach to generate a sample distribution that the models could be fitted against. Within this approach the key challenge lies within the ability to suitably estimate the transition probabilities. These probabilities will be different for each construction company and for each type of project. The probabilities for this study were estimated through a combination of literature survey and discussions with practising construction project managers.

A drawback of the approach taken is that estimates of which scenario the project lies in are only available once the first gateway has been passed. This might well be too late for any mitigating actions to have significant impact. Solutions to this drawback could include observing different types of signals or to use time taken to other nodes. The first approach, identifying other signals, would require further work on the nature of the scenarios with a focus on characterising the scenarios and thereby identifying the observable signals. The second approach would be to apply a similar methodology as presented in this paper, but extending it to all nodes in the process. With both these approaches, the Bayesian method presented here can still be applied.

Overall, the Bayesian approach provides promising results. The key aspect is that relatively simple and subtle signals, such as the time taken to pass the first gateway review, can be used to estimate the global conditions in which the project finds itself. This study shows that there is potential for considering incorporating other signals to achieve better estimates of the scenario that a project is operating within. Further, although beyond the scope of this work, there is a need to develop suitable mitigating actions that would provide the means for recovering a project that is suffering from being in a hindering scenario.

REFERENCES

[1] P Mitropoulos and G A Howell. Renovation projects: Design process problems and improvement mechanisms. Journal of Management in Engineering, 18(4):179–185, 2002.

[2] M Chester and C Hendrickson. Cost impacts, scheduling impacts, and the claims process during construction. Journal of Construction Engineering and Management – ASCE, 131(1):102– 107, 2005.

[3] L Soibelman, L Y Liu, J G Kirby, E W East, C H Caldas, and K Y Lin. Design review checking system with corporate lessons learned. Journal of Construction Engineering and Management – ASCE, 129(5):475–484, 2003.

[4] RIBA Plan of Work. Royal Institute of British Architects, London, 2007.

[5] Hsin-Hung Wu and Jiunn-I Shieh. Using a Markov chain model in quality function deployment to analyse customer requirements. International Journal of Advanced Manufacturing Technology, 30(1-2):141–146, 2006.

[6] H.A. Taha. Operations research: an introduction. Pearson/Prentice Hall, 2007.

[7] G Pahl and W Beitz. Engineering Design: A Systematic Approach. Springer-Verlag London, second edition, 1996.

[8] N Cross. Engineering Design Methods: Strategies for Product Design. JohnWiley, Chichester, UK, 2000.

[9] B Von Stamm. Managing innovation, design and creativity. Wiley, Chichester, 2008.

[10] R H Clough, G A Sears, and S K Sears. Construction Project Management. Wiley, New York, 2000.

[11] G J Ritz. Total Construction Project Management. McGraw-Hill, 1994.

[12] D G Pandelis. Markov decision processes with multidimensional action spaces. European Journal of Operational Research, 200:625–628, 2010.
[13] C Eckert, P J Clarkson, and W Zanker. Change and customisation in complex engineering domains. Research in Engineering Design, 15(1):1–21, 2004.
[14] T Flanagan, C M Eckert, and P J Clarkson. Extemalizing tacit overview knowledge: A model-based approach to supporting design teams. Artificial Intelligence for Engineering Design Analysis and Manufacturing, 21(3):227–242, 2007.
[15] J. Pearl. Causality: models, reasoning, and inference. Cambridge University Press, 2000.
[16] P Mason A Design Change Warning System for Improved Change Management in Construction Projects, Master's dissertation, Durham University, 2009.

Dr Peter Matthews
School of Engineering and Computing Sciences
Durham University
South Road
Durham DH1 3LE
United Kingdom
Tel: +44 191 334 2538
Email: p.c.matthews@durham.ac.uk
URL: http://www.dur.ac.uk/ecs/

Peter Matthews is a Lecturer in Design Informatics at the School of Engineering and Computing Sciences. His core interests lie in supporting the conceptual design process. The goal of Dr Matthews' research is to seek and develop methods for providing objective and quantitative feedback to the designer during the conceptual stage.