# A Bayesian Network for IT Governance Performance Prediction

Mårten Simonsson
Department of Industrial Information and Control Systems,
KTH, Royal Institute of Technology,
Stockholm, Sweden
+46 8 790 68 29
ms101@ics.kth.se

Robert Lagerström
Department of Industrial Information and Control Systems,
KTH, Royal Institute of Technology,
Stockholm, Sweden
+46 8 790 68 66
robertl@ics.kth.se

Pontus Johnson
Department of Industrial Information and Control Systems,
KTH, Royal Institute of Technology,
Stockholm, Sweden
+46 8 790 68 25
pj101@ics.kth.se

## ABSTRACT
The goal of IT governance is not only to achieve internal efficiency in an IT organization, but also to support IT's role as a business enabler. The latter is here denoted IT governance performance. IT management cannot control the IT governance performance directly. Instead, their realm of control includes several IT governance maturity indicators such as the existence of different IT activities, documents, metrics and roles. Current IT governance frameworks are suitable for describing IT governance, IT-systems, and business processes, but lack the ability to predict how changes to the IT governance maturity indicators affect IT governance performance. Bayesian networks are widely used for goal modeling and prediction in several research fields. This paper presents an application of Bayesian networks for IT governance performance prediction. Data from 35 case studies conducted in a variety of organizations has been used to determine the behavior of the network. An assumption on linearity is introduced in order to compensate for the limited amount of data, and the network learns using the proposed Linear Conditional Probability Matrix Generator. The resulting Bayesian network for IT governance performance prediction can be used to support IT governance decision-making.

## Categories and Subject Descriptors
K.6.0 [**Management of Computing and Information Systems**]

## General Terms
Management, Measurement

## Keywords
IT Governance, Bayesian networks



*10th Int. Conf. on Electronic Commerce (ICEC) '08 Innsbruck, Austria*


## 1. INTRODUCTION
Although there has been a need to provide guidance on the use of IT since the early days of computing, the concept of IT governance did not emerge until the nineties. Hendersen & Venkatraman [8] and Loh & Venkatraman [18] used the term to describe the complex array of interfirm relationships involved in achieving strategic alignment of business and IT. Effective IT governance provides mechanisms that enable IS/IT management to develop integrated business and IT plans, allocate responsibilities, and prioritize IT initiatives [16][26][33].

It is important to ensure that the IT governance is not only designed to achieve internal efficiency in the IT organization, such as deploying good IT processes and making sure that the means and goals are documented. The final goal of good IT governance is to provide the business with the best support needed in order to conduct business in a good manner. The IT governance mechanisms should be chosen so that the impact on the business is maximized. There are many activities in the IT organization that can be changed, but clearly, not all changes affect the business in a positive way. From an IT manager's point of view, it would be of great interest to know what impact each change made to the IT organization would have on the business, in order to choose the most beneficial way to govern IT.

Several frameworks for IT governance exist. Weill & Ross have developed a high level IT governance framework that can be used to assign responsibilities for IT decision making [34], but their work gives no further guidance on how the IT organization should actually transform theory into practice. The IT Infrastructure Library (ITIL) details establishment and maintenance of service level agreements (SLA) and aids the creation of processes related to delivery and support [11][22]. Though ITIL has traditionally provided little support for strategic IT concerns, this has been improved in recent ITIL v3 publications [23]. ITIL v3 will probably gain market share in the future, but currently the Control Objectives for Information and related Technology (COBIT) is most well-known known framework for IT governance improvement, risk mitigation, IT value delivery and strategic alignment maturity assessments [4][7][9][25][30][32].

In this paper, IT governance performance describes the goodness of an enterprise's IT organization from a business point of view. Current IT governance frameworks, as presented in the paragraph above, are mainly of descriptive nature, i.e. they describe the state of an IT organization according to best practice on IT governance. None of them has however the ability to foresee how the IT governance performance is linked to the maturity of the IT organization in terms of its activities, level of documentation, etc. This paper proposes a method for prediction of IT governance performance. In particular, by using such method, it is possible to compare the current state of IT governance performance with future scenarios. For instance, if the decision-making authority for acquisition of commodity software is moved from business unit level to IT operations level, how would that affect the IT governance performance? Making such predictions also enables prescription, i.e. not only evaluating different scenarios, but also to chose rationally between them.

A method for IT governance performance prediction should be based on previous experience, which could be gathered in several ways. For instance, knowledge from experts in the field could be collected in a best-practice based library of actions that could be taken to improve the business. Another option would be to study a variety of organizations and base the method on the findings. The organization enjoys good IT governance performance if business stakeholders are satisfied with the support that IT delivers to the business. By aggregating information from many organizations, a model for the correlation between IT governance performance and the maturity of the IT organization could be created. The research described in this paper is based on the findings from 35 organizations. Figure 1 shows a conceptual view of the model for IT governance performance prediction. On the left side is IT governance performance, as seen from the business point of view. The aim of any organization would be to improve the IT governance performance to increase stakeholder satisfaction and make sure that business runs as smoothly as possible.

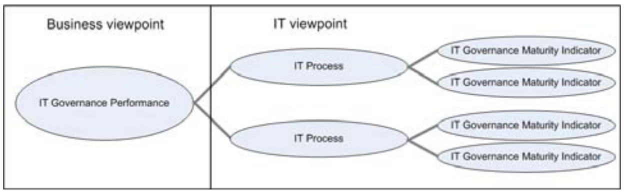



**Figure 1. The conceptual model for IT governance performance prediction. The IT governance performance, as seen from the business viewpoint, is not directly controllable. Within the realms of control for IT management are IT processes and IT governancematurity indicators.**

IT governance performance is not directly controllable by IT management, but IT processes for e.g. hardware acquisition, IT project management and IT strategy are in the realm of control. Even though such IT processes are difficult to measure directly, they comprise numerous and measureable IT governance maturity indicators such as maturity of individual IT related activities, level of monitoring, level of documentation and level of role assignment. It is reasonable to believe that some of the IT governance maturity indicators are correlated with IT governance performance. A model for prediction of IT governance performance would need to take into account and define the impact of each one of the IT governance performance indicators.

A method appropriate for modeling IT governance performance prediction needs to be able to represent a number of things. Firstly, basic-decision making support requires that the method is able to represent the decision maker's domain of control, his goals, and the causal relations between that which can be controlled and the desired goals. Secondly, the method should express causal uncertainty, i.e. uncertainty with respect to how concepts affect each other. Thirdly, the method should be able to learn from previous experience. Several prediction methods are used in the research community today, including Dempster-Shafer, Bayesian networks, neural networks, and multivariate analysis. Of the above presented methods, the Bayesian networks best fulfill all the requirements [15]. Another key feature for Bayesian networks is the existence of readily available graphical tools for decision support on the market, which makes modeling easy and straightforward. Based on the above discussion, Bayesian networks are used for IT governance performance prediction in this paper.

What is novel in this work is the use of empirical data from a large number of case studies as a base for decision-support in IT governance. The vehicle chosen for realization of the decision-support tool is a Bayesian network, whose mathematical apparatus is described. The value from understanding the novel aspect of this work is to gain detailed knowledge of the strength of the correlation between different IT governance maturity indicators, and IT governance performance. This allows IT managers to focus their effort on improving the IT processes that really have an impact on IT governance performance. The Bayesian network used to visualize the correlations can be used to predict IT governance in different change scenarios.

# 2. IT GOVERNANCE

It was decided to base the IT viewpoint of the Bayesian network, cf. Figure 1, on COBIT as this framework provides the most relevant and detailed support for IT governance [12]. The business viewpoint is based on the IT governance performance measure developed by Weill and Ross [34].

## 2.1 COBIT

COBIT [12] is the most well-known framework for IT governance maturity assessments [4][7][9][25][30][32]. It was first issued by the IT Governance Institute, ITGI, in 1998 and has been constantly evolving ever since. COBIT features a maturity model for IT governance, which follows the same principles as the Software Engineering Institute's Capability Maturity Model [10]. The framework also provides a definition of IT governance as consisting of four domains: Plan & Organize, Acquire & Implement, Deliver & Support, and Monitor & Evaluate. The domains define 34 processes, each process containing a number of IT governance maturity indicators, such as activities, documents, metrics and support for role and responsibility assignment. The domains and processes in COBIT are described briefly in the following subsections.

The process domain Plan and Organize (PO) covers ten IT processes of strategy and tactics concerning the identification of the way IT can best contribute to the achievement of the business objectives. The realization of the strategic vision needs to be planned, communicated and managed for different perspectives. A proper organization as well as technological infrastructure should be put in place.

To realize the IT strategy, IT solutions need to be identified, developed or acquired, as well as implemented and integrated into the business process. In addition, changes in and maintenance of existing systems are covered by this domain to make sure the solutions continue to meet business objectives. The process domain Acquire and Implement (AI) covers the seven processes that concern acquisition and implementation.

The process domain Deliver and Support (DS) is concerned with the actual delivery of required services, which includes service delivery, management of security and continuity, service support for users, and management of data and operational facilities.

All IT processes need to be regularly assessed over time for their quality and compliance with control requirements. The process domain Monitor and Evaluate (ME) addresses performance management, monitoring of internal control, regulatory compliance and governance.

## 2.2 Weill and Ross

IT governance performance is the quality of the services that the IT organization delivers, as seen from a business viewpoint [34]. In the early 2000's, MIT researchers Weill & Ross conducted a large set of case studies on IT governance performance of financially top-performing organizations. Their research method and the findings from more than 250 organizations were published in a book that is perhaps the most widely cited work in the field today [34]. Weill & Ross defined IT governance performance as the effectiveness of IT governance in delivering four objectives weighted by their importance to the enterprise, c.f. Table 2. Though simple and sometimes criticized therefore [29], Weill & Ross' IT governance performance definition is straightforward, easy to use and well-recognized. It has therefore been widely used by researchers and practitioners for benchmarking purposes [1][3][9][30].

# 3. EMPIRICAL DATA

The IT governance performance predictions discussed in this paper is based on empirical data from 35 organizations. Due to the nature of the study, organizations could not be randomly selected, but instead had to be chosen based on their willingness to participate and on their geographic location. However, significant effort was put into finding organizations from a wide range of industries and of different sizes to make the results as general as possible.

In total, 13 case study investigators conducted 158 interviews and collected results from 60 surveys in 37 organizations throughout the study, c.f. Table 8. Two organizations decided to drop the study without finishing, resulting in 35 useable cases. About 50% of the organizations were multinational with offices in several countries, all having headquarters in Europe. The number of employees in the investigated organizations spans from 10 to over 50000 with a median value of 2000. The revenue ranges from 1 M€ to 480 B€, with a median of 0.46 B€. All studies were carried out in 2007 and 2008, except for one of the pilot case studies that was made in 2006. The findings were validated in 15 interviews with IT governance experts.

# 4. BAYESIAN NETWORKS

Friedman describes a Bayesian network, $B=(G, P)$, as a representation of a joint probability distribution, where $G=(V, E)$ is a directed acyclic graph consisting of vertices, $V$, and edges, $E$ [6]. The vertices denote a domain of random variables $X_1,\ldots, X_n$, also denoted chance nodes. Each chance node, $X_i$, may take on a value $x_i$ from the finite domain $Val(X_i)$. The edges denote causal dependencies between the nodes, i.e. how the nodes relate to each other. The second component, $P$, of the network $B$, describes a conditional probability distribution for each chance node, $P(X_i)$, given its parents $Pa(X_i)$ in $G$. It is possible to write the joint probability distribution of the domain $X_1,\ldots, X_n$ using the chain rule of probability, in the product form:

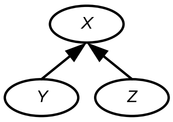
$$P(X_1,\ldots,X_n)=\prod_{i=1}^{n} P(X_i \mid Pa(X_i))$$

In order to specify the joint distribution, the respective conditional probabilities that appear in the product form must be found. The second component $P$ describes distributions $P(x_i \mid pa(X_i))$ for each possible value $x_i$ of $X_i$, and $pa(X_i)$ of $Pa(X_i)$, where $pa(X_i)$ is the set of values of $Pa(x_i)$. These conditional probabilities are represented in matrices, here on called conditional probability matrices (CPMs).Using a Bayesian network, it is possible to answer questions such as what is the probability of $X = x_1$ given that $Y = y_2$ and $Z = z_1$. An example of a Bayesian network and a CPM representing the chance nodes $X$, $Y$, and $Z$ and how these relate is shown in Figure 2. The CPM next to the network answers the question $P(x_1 \mid y_2, z_1)$ stated above. More comprehensive treatment on Bayesian networks can be found in e.g. [14][27][24][21].

The generic process for constructing Bayesian networks consists of the three steps for defining the nodes, the relations and the conditional probability matrices. These are described in the context of IT governance performance prediction in the following sections.

# 5. DEFINING THE NODES

The concept of IT governance as consisting of processes, activities, roles, documents and metrics was adopted from COBIT. The use of *processes* to describe an IT organization is commonly employed in many frameworks, e.g. COBIT and ITIL, and is also used in the herein proposed approach. Further, each process contains one or more *activities*, which represent the actual content of the work performed within the IT organization. The *documents* correspond to process inputs and outputs as stated in COBIT. *Metrics* are used to monitor the execution of each process, and a representation for metrics monitoring is also incorporated.




| Z | | $z_1$ | | | $z_2$ | | | $z_3$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Y | | $y_1$ | $y_2$ | $y_3$ | $y_1$ | $y_2$ | $y_3$ | $y_1$ | $y_2$ | $y_3$ |
| X | $x_1$ | $P(x_1 \mid y_1,z_1)$ | $P(x_1 \mid y_2,z_1)$ | $P(x_1 \mid y_3,z_1)$ | $P(x_1 \mid y_1,z_2)$ | $P(x_1 \mid y_2,z_2)$ | $P(x_1 \mid y_3,z_2)$ | $P(x_1 \mid y_1,z_3)$ | $P(x_1 \mid y_2,z_3)$ | $P(x_1 \mid y_3,z_3)$ |
| | $x_2$ | $P(x_2 \mid y_1,z_1)$ | $P(x_2 \mid y_2,z_1)$ | $P(x_2 \mid y_3,z_1)$ | $P(x_2 \mid y_1,z_2)$ | $P(x_2 \mid y_2,z_2)$ | $P(x_2 \mid y_3,z_2)$ | $P(x_2 \mid y_1,z_3)$ | $P(x_2 \mid y_2,z_3)$ | $P(x_2 \mid y_3,z_3)$ |
| | $x_3$ | $P(x_3 \mid y_1,z_1)$ | $P(x_3 \mid y_2,z_1)$ | $P(x_3 \mid y_3,z_1)$ | $P(x_3 \mid y_1,z_2)$ | $P(x_3 \mid y_2,z_2)$ | $P(x_3 \mid y_3,z_2)$ | $P(x_3 \mid y_1,z_3)$ | $P(x_3 \mid y_2,z_3)$ | $P(x_3 \mid y_3,z_3)$ |

**Figure 2. A Bayesian network and the conditional probability matrix for the chance node X given Y and Z.**

**Table 1. Four node types describe each IT process in the Bayesian network. Each node type can be assigned maturity levels ml0-ml5 according to the legend.**

| *Maturity Level (ml)* | *0* | *1* | *2* | *3* | *4* | *5* |
|---|---|---|---|---|---|---|
| ***Node type A: Activity execution*** | *No awareness of the importance of issues related to the activity. No monitoring is performed. No documentation exists. No activity improvement actions take place* | *Some awareness of the importance of issues related to the activity. No monitoring is performed. No documentation exists. No activity improvement actions take place.* | *Individuals have knowledge about issues related to the activity and take actions accordingly. No monitoring is performed. No documentation exists. No activity improvement actions take place.* | *Affected personnel are trained in the means and goals of the activity. No monitoring is performed. Documentation is present. No activity improvement actions take place.* | *Affected personnel are trained in the means and goals of the activity. Monitoring is performed. Documentation is present. The activity is under constant improvement. Automated tools are employed in a limited and fragmented way* | *Affected personnel are trained in the means and goals of the activity. Monitoring is performed. Documentation is present. Automated tools are employed in an integrated way, to improve quality and effectiveness of the activity* |
| ***Node type M: Metrics monitoring*** | *0 %* | *20 %* | *40 %* | *60 %* | *80 %* | *100 %* |
| ***Node type D: Documents in place*** | *0 %* | *20 %* | *40 %* | *60 %* | *80 %* | *100 %* |
| ***Node type R: Responsibility assignment*** | *No relations exist* | *Responsible or Accountable relations exist.* | *Responsible or Accountable relations exist. At least 40 % or relations in line with COBIT.* | *Responsible or Accountable relations exist. At least 40 % or relations in line with COBIT.* | *Responsible or Accountable relations exist. At least 80 % of relations in line with COBIT.* | *Responsible or Accountable relations exist. 100 % of relations in line with COBIT.* |

The concept of *roles* being responsible, accountable, consulted or informed on the execution of different activities is also incorporated. The role representation features the distinction between executives, business and IT as stated by Weill & Ross and in the Val IT framework [34][13], but also employs IT operations and audit roles taken from COBIT [12]. Indicators for IT governance maturity, as seen from IT's viewpoint, can be obtained by gathering information on the above mentioned entities for each IT process. Then, the activity execution (A), metrics monitoring (M), documents in place (D), and the responsibility assignment (R) can be evaluated [28]. The latter are represented as chance nodes in the Bayesian network for IT governance performance prediction. The nodes are assigned maturity levels according to Table 1.For instance, the Bayesian network representation of the COBIT process AI6, "Manage changes", consists of the nodes AI6_A, AI6_M, AI6_D, and AI6_R. The maturity of each node is determined according to the scale presented in Table 1. If, for process AI6, more than 20% of the documents mentioned in COBIT are in place, the node AI6_D is given the maturity level ml1. The entire IT organization is represented by means of 136 different nodes, i.e. four IT governance maturity indicators for each of COBIT's 34 processes.

**Table 2. Weill & Ross' IT governance performance objectives are used in the Bayesian network [34].**

| | |
|---|---|
| Q1. How important are the following outcomes of your IT governance, on a scale from 1 (not important) to 5 (very important)?<br><br>Cost effective use of IT<br>Effective use of IT for growth<br>Effective use of IT for asset utilization<br>Effective use of IT for business flexibility | Q2. What is the influence of IT governance in your business on the following measures of success, on a scale from 1 (not successful) to 5 (very successful)?<br><br>Cost effective use of IT<br>Effective use of IT for growth<br>Effective use of IT for asset utilization<br>Effective use of IT for business flexibility |

As mentioned earlier the purpose of the Bayesian network is not solely to study the IT organization in terms of controllable maturity indicators. The main goal is to predict the uncontrollable business perception of IT governance performance by studying the controllable IT governance maturity indicators, cf. Figure 1.

Weill & Ross have previously determined IT governance performance in 250 top performing organizations by letting senior management rate the importance of a few objectives, cf. Table 2 [34]. The same objectives have been used in the research presented here. Weill & Ross' objectives were aggregated and transformed into performance levels pl0-pl5 for the node ITG_Performance according to Formula 1. Please also refer to Table 2.

$$ITG\_Performance = 1.25\left(\frac{\sum_{n=1}^{4} Q1_n * Q2_n}{\sum_{n=1}^{4} Q1_n} - 1\right) \quad \mathbf{1}$$

## 6. DEFINING THE RELATIONS

In order to use a Bayesian network for predictions, not only knowledge about different nodes and their maturity levels or intrinsic performance levels is needed. It is also necessary to define how the nodes are related. According to previous research, the processes in COBIT well delimit the scope of the work performed by the IT organization [28]. The IT processes are controllable by IT management, and it is reasonable to believe that the maturity of an IT process is causally linked to the IT governance performance. The IT processes can be measured indirectly through analyzing the controllable activities, documents, metrics, roles and responsibilities, i.e. IT governance maturity indicators. There is a causal relation between the IT governance maturity indicators, and the IT processes. By determining only the maturity of one or more of the indicators as discussed in Section 4 (e.g. the nodes representing documents in place, PO1_D,…, ME4_D), the network can provide a prediction of IT governance performance. This is the idea behind the Bayesian network for prediction of IT governance performance, c.f. Figure 3. As a matter of fact, two separate IT governance maturity indicators or two IT processes are not necessarily independent. They may well co vary, hence the direction of the arrows in Figure 3. The strength of the causal relations in the network is mathematically described as CPMs.

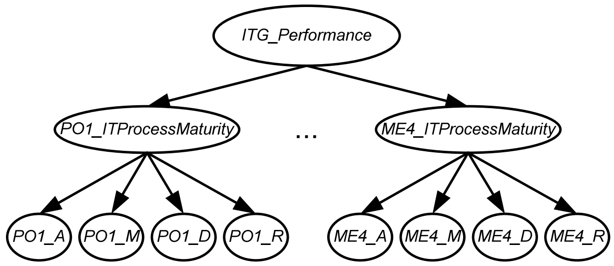



**Figure 3. The Bayesian network for IT governance maturity prediction features 34 IT Process Maturity nodes ranging from PO1 to ME4 [12].**

# 7. DEFINING THE CONDITIONAL PROBABILITY MATRICES

The conditional probability matrices define the chance nodes in the Bayesian network for IT governance performance prediction. In order to do so, the network must be learned, i.e. the parameters in the matrices need to be determined. The basic approach for this is to collect empirical data for the nodes by conducting case studies. Then, Bayesian network learning algorithms can e used to assign the parameters to the matrices.

## 7.1 Requirements on methods for learning Bayesian networks

There are several methods for learning Bayesian networks. This subsection briefly presents requirements for evaluation of four such methods.

The amount of data sets that the network can learn from is often limited. Therefore, the learning method must be able to obtain representable conditional probabilities based on a fairly small number of datasets. As shown in Figure 3, the structure of the network has already been determined and a learning method should not change it. In other words, the user herself should be able to choose network structure. The desired ability to learn conditional probabilities without changing the structure of the network is denoted parameter learning. Finally, it is deemed that the output of the method should be a conditional probability matrix.

## 7.2 Evaluation of methods for learning Bayesian networks

Four methods were evaluated, including the Expectation Maximization (EM), B-Course, Path Condition (PC), and the Necessary Path Condition (NPC) algorithms. The evaluation is focused on the requirements presented in the previous subsection, namely support for learning from a limited number of data sets, parameter learning, user choice of structure, and method outcome.

The most common method for learning Bayesian networks with statistical data is called the EM algorithm [5][17]. The main disadvantage with EM learning is that, when using only a small number of datasets, the learning will result in conditional probability matrices with zero entries. This means that if a set of values has not appeared in any of the learning cases, the set cannot be used for prediction.

B-course is a web-based online data analysis method that allows the user to analyze data for multivariate probabilistic dependencies [20]. The outcome of the method is a Bayesian network structure with learned conditional probability matrices. The main drawback with this method is that it learns the structure from data. It is not possible to force an already set structure upon B-course and only learn the parameters of the variables. B-course also requires large number of datasets to provide useable conditional probability matrices.

The PC algorithm is a constraint-based learning algorithm. This means that the algorithm uses statistical tests to derive a set of conditional independent and dependent statements, and learns the structure of a Bayesian network. The NPC algorithm is an enhancement of the PC algorithm which intends to bridge the latter's deficiencies in learning from small number of data sets. Both algorithms have the disadvantage that their outcomes are structures and not CPMs [19].

To summarize, the main requirement is that the method should be able to learn parameters based on collected data, which excludes the PC and NPC algorithms. Since data collection in the case of IT governance performance prediction is made through case studies, a key requirement is the limited amount of data sets available. Neither the EM algorithm nor B-course addresses this issue. The result of the evaluation of the methods is visualized in Table 3. Linear regression is a commonly used method for prediction of the outcome of one variable based on the information of other variables [2][31]. It may thus also be appropriate for learning conditional probability matrices in Bayesian networks. The main weakness of this approach is that the outcome of a linear regression is not a conditional probability matrix, but rather an equation y=ax+b. However, if the outcome of linear regression could be translated into a CPM, the approach would be appropriate for our purposes.

**Table 3. A comparison of different learning approaches for Bayesian networks.**

| Method / Requirement | EM Algo rithm | B-Course | PC Algo rithm | NPC Algo tithm | LCPMG |
|---|---|---|---|---|---|
| Support for learning from a limited number of data sets | | | | √ | √ |
| Parameter learning | √ | √ | | | √ |
| User choice of structure | √ | | | | √ |
| Method outcome | CPM | CPM & Structu re | Struc ture | Struc ture | CPM |

## 7.3 The Linear Conditional Probability Matrix Generator

Unfortunately, no linear learning approach with conditional probability matrices as outcome exists in the readily available tools for Bayesian statistics. Therefore, the Linear Conditional

Probability Matrix Generator (LCPMG) was developed. In general, LCPMG takes into account gathered observation data, processes it, and returns a conditional probability matrix made with an assumption of linearity in the input data. The generator works according the following steps:

Observations on a quantitative scale of measurement are made and a structure is decided upon. Assume 20 simultaneous observations of the variable $X$ and $Y$. The nodes $X$ and $Y$ are causally related to one another in the network, $X$ affecting $Y$. The observed values are on a continuous scale $x_i \in [0...5]$ and $y_i \in [0...5]$.

A linear regression on the observations is performed according to standard procedures [2][31]. The result is an equation $Y_{estimate} = aX + b$, where *a* and *b* are scalar constants. The residuals constitute the difference between the linear approximation that is fitted to the observations $(X_i, Y_i)$, and the actual observations $R = Y - Y_{estimate}$. The standard deviation S of the residuals, an approximation of the certainty with which the linear approximation is made, is calculated. The purpose of the LCPMG is to generate a discrete CPM from a continuous linear approximation $Y_{estimate}$. In order to do that, six different ranges for $y_i$ are created, given that $x_i \in [0,1,2,3,4,5]$, and $-0.5 \le y_0 < 0.5 \le y_1 < 1.5 \le y_2 < 2.5 \le y_3 < 3.5 \le y_4 < 4.5 \le y_5 < 5.5$. Based on the linear approximation and the standard deviation S, the probability mass $P(y_i \mid x_i)$ in each cell of the CPM is calculated, c.f. Table 4. The total probability $\sum_{y_0}^{y_5} P(y_i \mid x) = 1$ for each $x \in [0,1,2,3,4,5]$. As an example, if the linear approximation is $y_{estimate} = 0.5x + 1$, and the standard deviation for the residual vectors equals $0.5$, this corresponds to $P(y_2 \mid x_2) = 68.27$ % [2][31]. In summary, the LPCMG has now transformed two arrays with observations on a continuous scale, to a CPM describing the causal relation between two nodes in a Bayesian network.

**Table 4. The CPM is created from a continuous linear approximation. Each cell holds the probability mass for P(Y | X).**

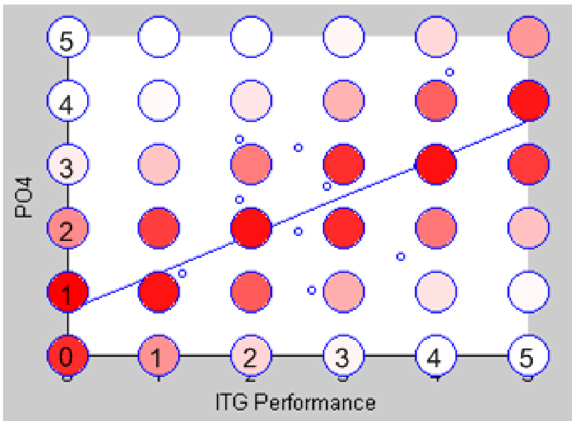
| X | | $x_1$ | $x_2$ | $x_3$ |
|---|---|---|---|---|
| **Y** | $y_3$ | $P(y_3 \mid x_1)$ | $P(y_3 \mid x_2)$ | $P(y_3 \mid x_3)$ |
| | $y_2$ | $P(y_2 \mid x_1)$ | $P(y_2 \mid x_2)$ | $P(y_2 \mid x_3)$ |
| | $y_1$ | $P(y_1 \mid x_1)$ | $P(y_1 \mid x_2)$ | $P(y_1 \mid x_3)$ |

Returning to the requirements on methods for learning Bayesian networks and comparing the LCPMG to other already available methods one finds that the LCPMG fulfills all four requirements. Table 3 shows a final comparison of some common learning approaches for Bayesian networks and the here proposed LCPMG.

## 7.4 Using LCPMG for creation of a Bayesian network for IT governance performance prediction

The LCPMG is suitable for generating the CPMs of nodes that are linearly related to one another. In the case of IT governance maturity prediction, the Bayesian network has got three hierarchical levels, c.f. Figure 3. The first level contains the measureable, yet not controllable IT governance performance node. The second level contains 34 IT process nodes that are controllable, but not measureable. On the third level, 136 measureable and controllable IT governance maturity indicator nodes reside, cf. Table 1. The CPMs of all nodes at all levels must be defined, and the LCPMG can be applied stepwise in order for the network to learn the CPMs.

The first step is to calculate the regressions for all IT governance maturity indicator nodes and the IT governance performance node. The regressions are then used to assign normalized weights $w_i$ to each of the four node types; activities, metrics, documents and responsibilities.




**CPM for PO4 and ITG_Performance (rounded)**

| | pl0 | pl1 | pl2 | pl3 | pl4 | pl5 |
|---|---|---|---|---|---|---|
| **ml5** | 0,01 | 0,01 | 0,01 | 0,02 | 0,06 | 0,17 |
| **ml4** | 0,01 | 0,01 | 0,04 | 0,13 | 0,26 | 0,39 |
| **ml3** | 0,03 | 0,10 | 0,22 | 0,35 | 0,40 | 0,33 |
| **ml2** | 0,19 | 0,32 | 0,40 | 0,35 | 0,23 | 0,10 |
| **ml1** | 0,42 | 0,39 | 0,27 | 0,14 | 0,05 | 0,01 |
| **ml0** | 0,35 | 0,18 | 0,07 | 0,02 | 0,01 | 0,01 |

ITG_Performance = 0.60*PO4 + 0.69, S = 0.96

**Figure 4. Calculations of IT process PO4's maturity and observations of ITG_Performance (small dots), the linear approximation of the relation between them, and how these fit into the CPM (colored bubbles). The resulting CPM is shown to the right. S denotes the standard deviation of the residuals [2][31].**

The maturity for an IT process, $m_p$, can then be calculated as $m_p = w_a * ml\_a_p + w_m * ml\_m_p + w_d * ml\_d_p + w_r * ml\_r_p$ for each of the *N**34 IT processes, *N* representing the number of different observations made. LCPMG is then used to determine the CPMs for each of the 34 IT process nodes, based on the $m_p$:s and the ITG_Performance node. Then, the LCPMG is used to determine the CPMs for each of the 136 maturity indicator nodes, based on maturity levels for the maturity indicators, and the $m_p$:s. Finally, the prior of the ITG_Performance node is set by analyzing the occurrence of each one of the possible levels pl0-pl5. If only small datasets have been used in order for the Bayesian network to learn, all levels of ITG_Performance have perhaps not been observed. This can be corrected for by using

Laplace's estimation, i.e. add 1 to the number of observations assigned to each state [21]. In this way, no zeros will be present in the resulting CPM and it is thus resulting in a better and more smoothly predicting Bayesian network.

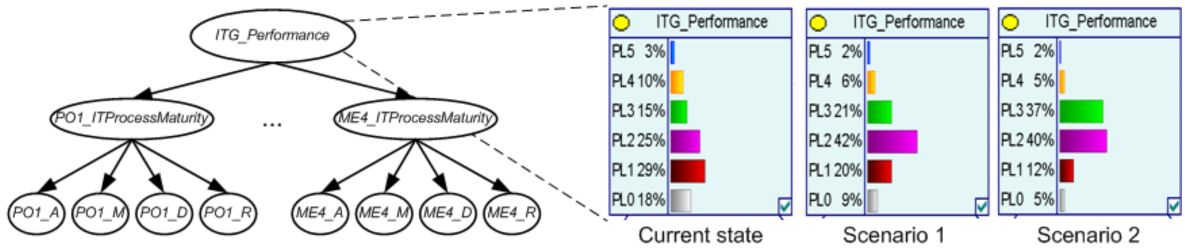



**Figure 5. IT governance performance predictions for the current organization and the two change scenarios. The bar charts, showing the probability of occurrence for each state of ITG_Performance, are created using the GeNIe tool, http://genie.sis.pitt.edu/.**

Figure 4 shows observations for $Y$ = PO4 (Define the IT processes, Organization and Relationships) and $X$ = ITG_Performance, the linear approximation and a graphic representation of the probability mass for each cell in the CPM. The darker red the color of a bubble, the higher the probability mass $P(y_i|x_i)$ of the corresponding cell in the CPM.

## 8. APPLICATION OF THE BAYESIAN NETWORK – AN EXAMPLE

This section illustrates how the Bayesian network for IT governance performance prediction can be used in a decision-making situation for an IT manager. Let us assume a municipality. Even though the IT department within the municipality feels comfortable about providing good services, the head of municipality is dissatisfied with the IT department's level of flexibility towards change. The municipality CIO faces the problem of improving the IT governance performance, which lies beyond his domain of control, but the values of the nodes representing IT governance maturity indicators can be changed, cf. Figure 1. Two different change scenarios that might impact the IT governance performance in a positive way are created:

- Improve the level of documentation in the entire IT organization, i.e. raising the documentation maturity level one step for all IT processes PO1_D, PO2_D,…, ME4_D.
- Focus improvements on a few critical IT processes, i.e. raising the maturity of activities, metrics, documentation, and relations two levels for the selected IT processes AI2, AI3, AI6, and DS10.

The Bayesian network for IT governance performance prediction has already learned from several, similar organizations before. By interviewing key IT stakeholders in the municipality, such as IT managers, IT operations personnel and IT auditors, the current IT governance performance can be represented in the Bayesian network. Changes according to each of the scenarios are then inserted into the network. The resulting predictions for IT governance performance, showing that the second change scenario has a larger impact on the IT governance performance, are displayed in Figure 5. Such graphs can serve as decision making support on IT governance performance issues.

## 9. DISCUSSION & CONCLUSIONS

As of June 2008, 158 interviews have been conducted within in 35 different organizations in order to collect data from a variety of industries, including banks, the public sector, telecommunications, electric utilities and manufacturing. The LCPMG has been applied upon the collected data in order for the Bayesian network for IT Governance Performance to learn. In spite of the variety of empirical data, correlations between IT governance performance and IT governance maturity indicators are clearly visible, and the Bayesian network is usable for making predictions. Based on the current sets of data it seems that the maturity indicators that most strongly correlate with IT governance performance do not differ among industries.

In summary, this paper has been discussing the use of Bayesian networks for prediction of IT governance performance. Such information can be used to make well-informed decisions on IT governance, e.g. selecting between two different change scenarios. The Linear Conditional Probability Matrix Generator, LCPMG, is proposed as a way for Bayesian networks to learn from small datasets.

Last but not least, the authors would like to thank Professor Stefan Arnborg, Royal Institute of Technology, Stockholm, Sweden, for his valuable input on the nature of Bayesian statistics.

# References

[1] Clementi, S. and Carvalho, T.C.M.B. 2007 Methodology for IT governance Assessment and Design. In Proceedings of the 6th IFIP International Conference on e-Commerce, e-Business, and e-Government (Turku, Finland 2006).

[2] Cohen, J. and Cohen, P. 1983 Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences. Lawrence Erlbaum Associates.

[3] Dahlberg, T. and Kivijärvi, H. 2006 An Integrated Framework for IT Governance and the Development and Validation of an Assessment Instrument. In Proceedings of the 39th Hawaii International Conference on System Sciences. (Hawaii, January 2006).

[4] Debraceny, R.S. 2006 Re-engineering IT Internal Controls - Applying capability Maturity Models to the Evaluation of IT Controls. In Proceedings of the 39th Hawaii International Conference on System Sciences. (Hawaii, January 2006). IEEE. DOI= http://doi.ieeecomputersociety.org/10.1109/HICSS.2006.407

[5] Friedman, N. 1998. The Bayesian Structural EM Algorithm. In Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence, UAI '98. (San Francisco, United States, 1998).

[6] Friedman, N., Linial, M., Nachman, I. and Pe'er, D. 2000 Using Bayesian Networks to Analyze Expression Data. Journal of Computational Biology, 7,3-4 (2000), 601-620. Mary Ann Liebert, Inc., publishers.

[7] Guldentops, E. 2004 Governing Information Technology through COBIT. In Strategies for Information Technology Governance. Van Grembergen, W., Ed. Idea Group Publishing.

[8] Hendersen, J.C. and N. Venkatraman 1993 Strategic Alignment: Leveraging Information Technology for Transforming Organizations. IBM Systems Journal 32,1. (1993), 472-485.

[9] Holm Larsen, M., Kühn Pedersen M., and Viborg Andersen, K. 2006 IT Governance – Reviewing 17 IT Governance Tools and Analyzing the Case of Novozymes A/S. In Proceedings of the 39th Hawaii International Conference on System Sciences. (Hawaii, January 2006). IEEE. DOI = http://doi.ieeecomputersociety.org/10.1109/HICSS.2006.234

[10] Humphrey, W.S. 1989 Managing the Software Process. Addison-Wesley Professional, Boston MA, USA.

[11] International Organization for Standardization 2005. ISO/IEC 20000-1 & ISO/IEC 20000-2. International Organization for Standardization.

[12] IT Governance Institute 2005. Control Objectives for Information and Related Technology, 4th Edition. http://www.itgi.org.

[13] IT Governance Institute 2007 ValIT Framework. http://www.itgi.org.

[14] Jensen, F. 2001 Bayesian Networks and Decision Graphs. Springer-Verlag, New York, USA.

[15] Johnson, P., Lagerström, R., Närman, P. and Simonsson, M. 2007 Enterprise Architecture Analysis with Extended Influence Diagrams. Information Systems Frontiers. 9,2 (2007), 163-180.

[16] Korac-Kakabadse, N. and Kakabadse, A. 2001 IS/IT Governance: Need For an Integrated Model. Corporate Governance 4,1 (2001).

[17] Lauritzen, S. 1995 The EM algorithm for graphical association models with missing data. Computational Statistics and Data Analysis, 19 (1995), 191-201.

[18] Loh, L. and Venkatraman, N. 1992 Diffusion of Information Technology Outsourcing: Influence Sources and the Kodak Effect. Information Systems Research 3,4 (1992), 334-359.

[19] Madsen, A., Lang, M., Kjaerulff, U. and Jensen, F. 2004 The Hugin Tool for Learning Bayesian Networks. Symbolic and Quantitative Approaches to Reasoning with Uncertainty, Springer Berlin / Heidelberg, 2711 (2004), 594-605.

[20] Myllymäki, P., Silander, T., Tirri, H. and Uronen, P. 2002 B-Course: A Web-Based Tool for Bayesian and Causal Data Analysis. International Journal on Artificial Intelligence Tools, 11,3 (2002), 369-387.

[21] Neapolitan, R. 2003 Learning Bayesian Networks. Prentice-Hall.

[22] Office of Government Commerce. 2003 IT Infrastructure Library Service Delivery. The Stationery Office.

[23] Office of Government Commerce. 2007 Service Strategy Book. The Stationery Office.

[24] Pearl, J. 1988 Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann.

[25] Ridley, G., Young, J. and Carroll, P. 2004 COBIT and its utilization - A framework from the literature. In Proceedings of the 37th Hawaii International Conference on System Sciences. (Hawaii, January 2005). IEEE. DOI= http://doi.ieeecomputersociety.org/10.1109/HICSS.2004.1265566

[26] Ross, J.W. 2003 Creating a Strategic IT Architecture Competency - Learning in Stages. MIT Sloan School of Management Working Paper. 4314-03. MIT Sloan School of Management.

[27] Shachter, R. 1988 Probabilistic inference and influence diagrams. Operations Research, 36,4 (July-August 1988), 589-604.

[28] Simonsson, M., Johnson, P. 2008 The IT organization modeling and assessment tool: Correlating IT governance maturity with the effect of IT. In Proceedings of the 41st Hawaii International Conference on System Sciences. (Hawaii, January 2008). IEEE. DOI= http://doi.ieeecomputersociety.org/10.1109/HICSS.2008.447

[29] Tu, W. 2007 Identifying Incentive Factors in IT Governance: An Exploratory Study. In Proceedings of the 2007 International Conference on Convergence Information Technology. (Gyeongju, Korea, November 2007).

[30] Van Grembergen, W. and De Haes, S. 2008 Implementing Information Technology Governance: Models, Practices and Cases. IGI Publishing. Hershey.

[31] Walsh, C. 1990 Statistics for the Social Sciences with Computer Applications. Harper & Row Publishers.

[32] Warland, C. and Ridley, G. 2005 Awareness of IT Control Frameworks in an Australian State Government: A Qualitative Case Study. In Proceedings of the 38th Hawaii International Conference on System Sciences. (Hawaii, January 2005). IEEE. DOI= http://doi.ieeecomputersociety.org/10.1109/HICSS.2005.116

[33] Weill, P. and Woodham , R. 2002. Dont Just Lead, Govern: Implementing Effective IT Governance. MIT Sloan School of Management Working Paper, 4237-02. MIT Sloan School of Management.

[34] Weill, P., Ross, J. 2004 IT Governance – How top performers manage IT decision rights for superior results. Harvard Business School Press.