Improving Software Effort Estimation Using an Expert-Centred Approach

Emilia Mendes

Improving Software Effort Estimation Using an Expert-Centred Approach

Emilia Mendes

School of Computing, Blekinge Institute of Technology, SE-371 79 Karlskrona, Sweden
Emilia.Mendes@bth.se

Abstract. A cornerstone of software project management is effort estimation, the process by which effort is forecasted and used as basis to predict costs and allocate resources effectively, so enabling projects to be delivered on time and within budget. Effort estimation is a very complex domain where the relationship between factors is non-deterministic and has an inherently uncertain nature, and where corresponding decisions and predictions require reasoning with uncertainty. Most studies in this field, however, have to date investigated ways to improve software effort estimation by proposing and comparing techniques to build effort prediction models where such models are built solely from data on past software projects - data-driven models. The drawback with such approach is threefold: first, it ignores the explicit inclusion of uncertainty, which is inherent to the effort estimation domain, into such models; second, it ignores the explicit representation of causal relationships between factors; third, it relies solely on the variables being part of the dataset used for model building, under the assumption that those variables represent the fundamental factors within the context of software effort prediction. Recently, as part of a New Zealand and later on Brazilian government-funded projects, we investigated the use of an expert-centred approach in combination with a technique that enables the explicit inclusion of uncertainty and causal relationships as means to improve software effort estimation. This paper will first provide an overview of the effort estimation process, followed by the discussion of how an expert-centred approach to improving such process can be advantageous to software companies. In addition, we also detail our experience building and validating six different expert-based effort estimation models for ICT companies in New Zealand and Brazil. Post-mortem interviews with the participating companies showed that they found the entire process extremely beneficial and worthwhile, and that all the models created remained in use by those companies. Finally, the methodology focus of this paper, which focuses on expert knowledge elicitation and participation, can be employed not only to improve a software effort estimation process, but also to improve other project management-related activities.

Keywords: Software Effort Estimation, Expert-centred Approach, Process Improvement, Cost Estimation, Project Management.

1 Introduction

The purpose of estimating effort is to predict the amount of effort (person/time) required to develop an application (and possibly also a service within the Web context), often based on knowledge of 'similar' applications/services previously developed. The accuracy of an effort estimate can affect significantly whether projects will be delivered on time and within budget; therefore effort estimation is taken as one of the main foundations of a sound project management. However, because effort estimation is a complex domain where corresponding decisions and predictions require reasoning with uncertainty, there are countless examples of companies that underestimate effort. Jørgensen and Grimstad [1] reported that such estimation error can be of 30%-40% on average, thus leading to serious project management problems.

Fig. 1 provides a general overview of an effort estimation process [2]. Estimated characteristics of the new application/service to be developed, and its context (project), are the input, and effort is the output we wish to predict. For example, a given software company may find that to predict the effort necessary to implement a new e-commerce Web application, it will need to estimate early on in the development project the following characteristics:

- *Estimated number of new Web pages.*
- *The number of functions/features (e.g. shopping cart, on-line forum) to be offered by the new Web application.*
- *Total number of developers who will help develop the new Web application*
- *Developers' average number of years of experience with the development tools employed.*
- *The choice of main programming language used.*

Of these variables, *estimated number of new Web pages* and *the number of functions/features to be offered by the new Web application* characterise the **size** of the new Web application; the other three, *total number of developers who will help develop the new Web application*, *developers' average number of years of experience with the development tools employed*, and *main programming language used*, characterise the project - the context for the development of the new application, and are also believed to influence the amount of effort necessary to develop this new application. The project-related characteristics are co-jointly named 'cost drivers'.

No matter what type of development it is (of an application or of a service), in general the one consistent input found to have the strongest effect on the amount of effort needed to develop an application or service is size (i.e. the total number of server side scripts, the total number of Web pages), with cost drivers also playing an influential role.

Evidence also shows that for most part of existing software & service projects, effort estimation is based on past experience, where knowledge or data from past finished applications & projects are used to estimate effort for new applications & projects not yet initiated [1]. The assumption here is that previous projects are similar to the new projects to be developed, and therefore knowledge and/or data from past projects can be useful in estimating effort for future projects. This process is also

illustrated in Fig. 1. Those steps (some or all) can be executed more than once throughout a given software development cycle, depending on the process model adopted by the company. For example, if the process model adopted by a company complies with a waterfall model this means that most probably there will be an initial effort estimate for the project, which will remain unchanged throughout the project. If a company's process model complies with the spiral model, this means that for each cycle within the spiral process a new/updated effort estimate is obtained, and used to update the current project's plan and effort estimate. If a company uses an agile process model, an effort estimate is likely to be obtained for each of a project's iterations (e.g. sprints). In summary, a project's process model drives at what stage(s) an effort estimate(s) is/are obtained, and whether or not these estimates are revisited at some point throughout a project's development life cycle.

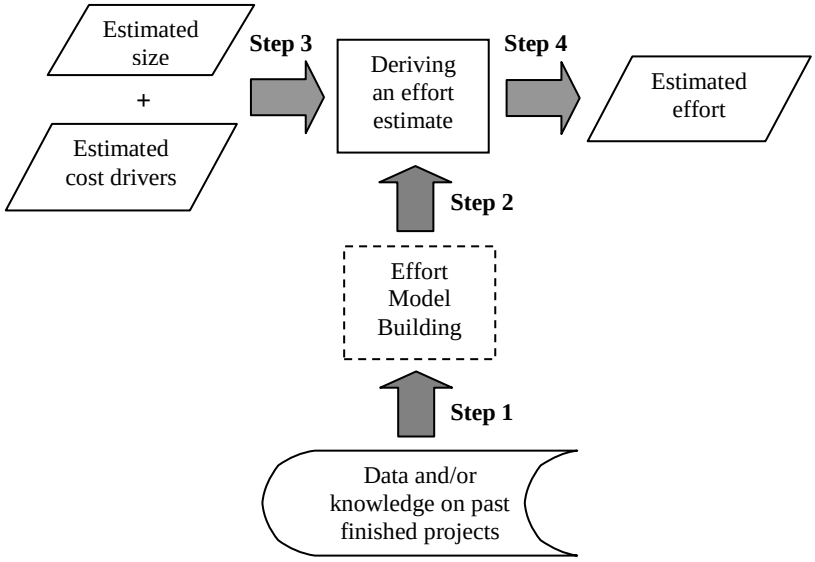

Fig. 1. Effort Estimation process [2]

Note that cost and effort are often used interchangeably within the context of most effort estimation literature since effort is taken as the main component of project costs. However, given that project costs also take into account other factors such as contingency and profit [3] we will use the word "effort" and not "cost" throughout this paper.

The remaining of this paper is organised as follows: the next Section provides a motivation for using an expert-centred approach to improving software effort estimation, followed by another two Sections where first the approach we propose is briefly introduced and second explained in more detail using the experience from eliciting six different expert-centred real models. Finally, our last two Sections

discuss our experience building six expert-centred models to improve effort estimation, and lessons learnt, respectively.

2 Motivation to Employing an Expert-Centred Approach

Most research in software & Web effort estimation has to date focused on solving companies' inaccurate effort predictions via investigating techniques that are used to build formal effort estimation models, in the hope that such formalisation will improve the accuracy of estimates.

They do so by assessing, and often also comparing, the prediction accuracy obtained from applying numerous statistical and artificial intelligence techniques to datasets of completed software/Web projects developed by industry, and sometimes also developed by students. Recent literature reviews of software and Web effort estimation studies are given respectively in [4] and [5].

The variables characterising such datasets are determined in different ways, such as via surveys [6], interviews with experts [7], expertise from companies [8], a combination of research findings [9], or even a researcher's own consulting experience [10]. In all of these instances, once variables are defined, a data gathering exercise takes place, obtaining data (ideally) from industrial projects volunteered by companies. Except when using research findings to inform variables' identification, invariably the mechanism employed to determine variables relies on experts' recalling, where the subjective measure of an expert's certainty is often their amount of experience estimating effort.

However, in addition to eliciting the important effort predictors (and optionally also their relationships), such mechanism does not provide the means to also quantify the uncertainty associated with these relationships and to validate the knowledge obtained. Why should these be important?

Our experience developing and validating single-company expert-centred Software and Web effort prediction models that incorporate the uncertainty inherent in this domain [11] showed that the use of an expert-centred structured iterative process in which factors and relationships are identified, quantified and validated [12][13][14] leads the participating companies to a much more thorough and deep understanding of their mental processes and their decisions when estimating effort, when compared to just the recalling of factors and their relationships. The iterative process we use employs Bayesian inference, which is one of the techniques employed in root cause analysis [15]; therefore it aims at a detailed analysis and understanding of a particular phenomenon of interest.

In all the case studies we conducted, the original set of factors and relationships initially elicited was always modified as the model evolved; this occurred as a result of applying a root cause analysis approach comprising a Bayesian inference mechanism and feedback into the analysis process via a model validation. In addition, post-mortem interviews with the participating companies showed that the understanding they gained by being actively engaged in building those models led to both improved estimates and estimation processes [11][12][13][14].

We therefore contend that the use of an expert-centred structured iterative process provides the means to elicit a more robust set of predictors and relationships, when compared to other means of elicitation. We argue that the recalling mechanism used in surveys and interviews to elicit the important factors (and also occasionally their relationships) when estimating effort does not provide any means for experts to understand thoroughly their own decision making processes via the quantification of the uncertainty part of that decision process, and the validation of the factors they suggested during the elicitation. This means that the list of factors elicited is most likely based on a superficial process.

3 An Overview of the Expert-Centred Approach

3.1 Technique Used

The expert-centred approach proposed herein is based on a technique called Bayesian Networks (BN). This technique and corresponding process are briefly introduced in this Section, and further details are given in [14].

A BN is a model that enables the characterisation of a knowledge domain in terms of its factors, their relationships, and the uncertainty inherent to that domain. It has two parts [15]. The first part, known as the BN's qualitative part, results in a graphical structure comprising the factors and causal relationships identified as fundamental in the domain being modelled. This structure is depicted by a Directed Acyclic Graph (DAG) (see Fig. 2(a)). In addition to identifying factors and relationships, this part also includes the identification of the states (values) that each factor should take (e.g. Small (1 to 5), Medium (6 to 15), or Large (16+) in Fig. 2(a)).

The second part, known as the quantitative part, represents the relationships identified in the qualitative part, and their quantification, done probabilistically. This quantification represents the uncertainty in the domain being modelled, and in order for it to be accomplished, a Conditional Probability Table (CPT) is associated to each node in the graph. A parent node's CPT describes the relative probability of each state (value) (Fig. 2(b) CPTs for nodes 'Total Number of Web pages' and 'Total Number of Images'); a child node's CPT describes the relative probability of each state conditional on every combination of states of its parents (Fig. 2(b) CPT for node 'Total Effort'). Each row in a CPT represents a conditional probability distribution and therefore its values sum up to one [8]. Such probabilities can be attained via expert elicitation, automatically from data, from existing literature, or using a combination of these. However, within the context of this research all probabilities were obtained via expert elicitation. Once both qualitative and quantitative parts are specified, the BN is validated using data on past finished projects, where one project at a time is entered as evidence (see Fig. 2(d)) and used to check whether the BN provides the highest probability to a value (range of values) that includes the real actual effort for that project, which is known. If not, then the BN is re-calibrated.

The building of a BN model is an iterative process where one can move between the three different steps of this process – building the BN's structure, or qualitative

part, building the CPTs, or the quantitative part, and validating the model. Once a BN is validated (see Fig. 1(c), evidence (e.g. values) can be entered into any node, and probabilities for the remaining nodes automatically calculated using Bayes' rule [8] (see Figs. 2(d) and 1(e). This was the validation method employed herein.

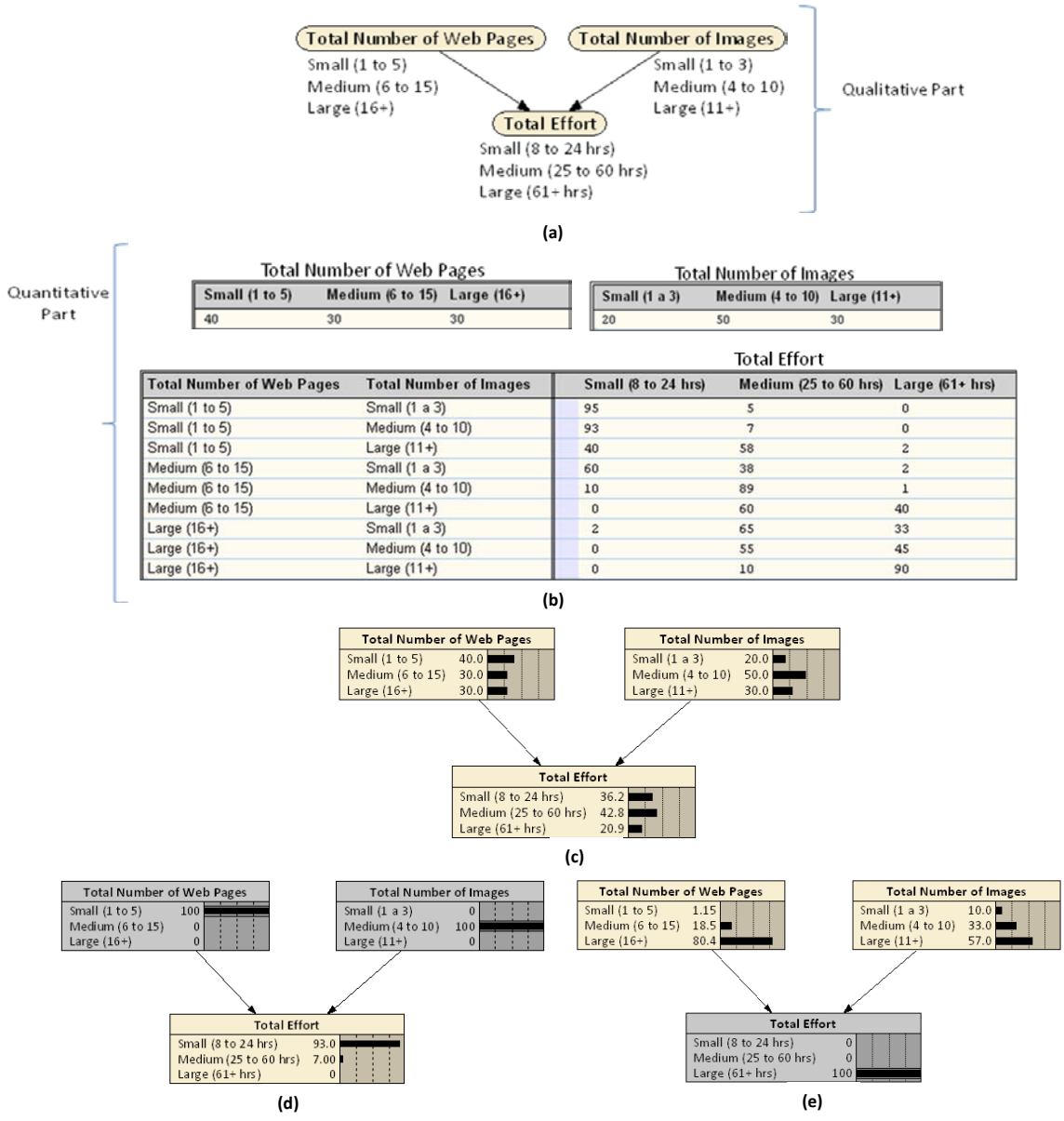

Fig. 2. Parts of a Bayesian Network an Types of Reasoning

In summary, BNs can be used for different types of reasoning, such as predictive (see Fig. 2(d)), diagnostic (see Fig. 2(e)), and "what-if" analyses to investigate the impact that changes on some nodes have on others [15].

3.2 Process We Employed to Building the Expert-Centred Models (BNs)

The process that was used to build and validate the expert-centred models focus of this research is an adaptation of the Knowledge Engineering of Bayesian Networks (KEBN) process proposed in [16] (see Fig. 3). As shown in Fig. 3, this process iterates over three steps - Structural Development, Parameter Estimation, and Model Validation, until a complete BN is built and validated. Each of the steps is detailed next:

Structural Development: This step represents the qualitative component of a BN, which results in a graphical structure comprised of, in our case, the factors (nodes, variables) and causal relationships identified as fundamental for effort estimation of software & Web projects. In addition to identifying variables, their types (e.g. query variable, evidence variable) and causal relationships, this step also comprises the identification of the states (values) that each variable should take, and if they are discrete or continuous. In practice, currently available BN tools require that continuous variables be discretised by converting them into multinomial variables, also the case with the BN software used in this study. The BN's structure is refined through an iterative process where existing literature in the field can also be used as input to the process. This structure construction process has been validated in previous studies (e.g. [16][17]) and uses the principles of problem solving employed in data modelling and software development [18]. Throughout this step the Kowledge Engineer (responsible for eliciting the knowledge from the Domain Expert(s) (Des)) also evaluates the structure of the BN, checking whether variables and their values have a clear meaning; all relevant variables have been included; variables are named conveniently; all states are appropriate (exhaustive and exclusive); a check for any states that can be combined. Once the BN structure is assumed to be close to final the KE may still need to optimise this structure to reduce the number of probabilities that need to be elicited or learnt for the network. If optimisation is needed, techniques that change the causal structure (e.g. divorcing [19]) are employed.

Parameter Estimation: This step represents the quantitative component of a BN, where conditional probabilities corresponding to the quantification of the relationships between variables [19] are obtained. Such probabilities can be attained via Expert Elicitation, automatically from data, from existing literature, or using a combination of these. When probabilities are elicited from scratch, or even if they only need to be revisited, this step can be very time consuming. In order to minimise the number of probabilities to be elicited some techniques have been proposed in the literature (e.g. [16][17]).

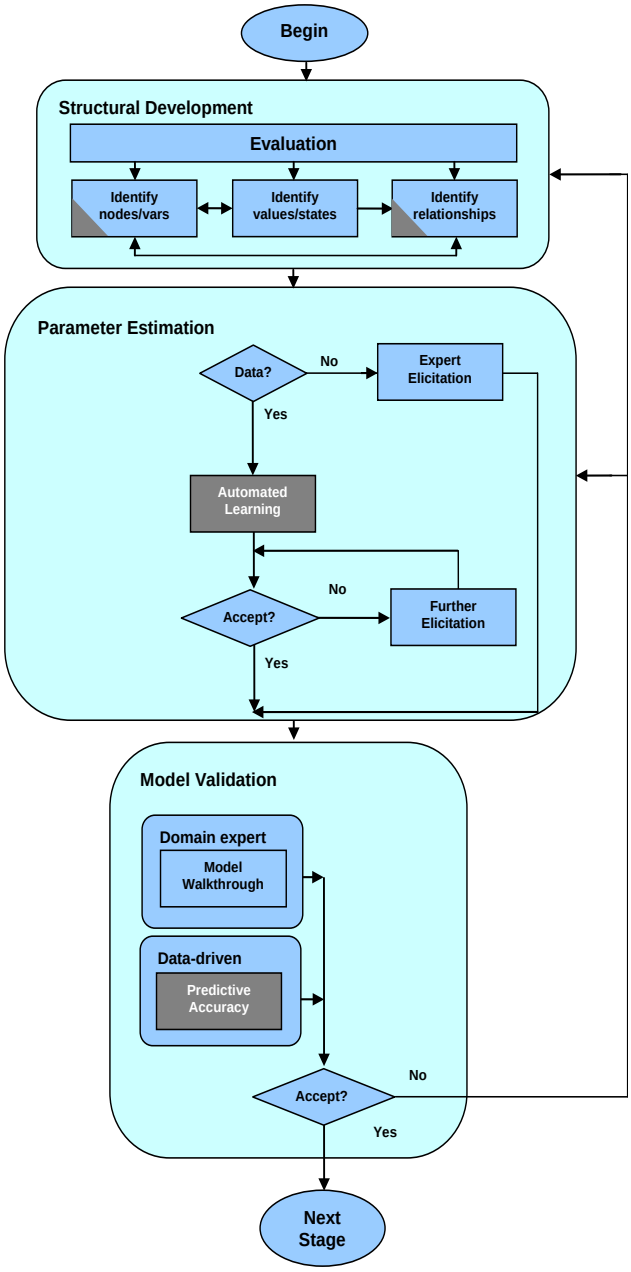

Fig. 3. KEBN, adapted from [16]

Model Validation: This step validates the BN resulting from the two previous steps, and determines whether it is necessary to re-visit any of those steps. Two different validation methods are generally used - Model Walkthrough and Predictive

Accuracy. Model walkthrough represents the use of real case scenarios that are prepared and used by DEs to assess if the predictions provided by the model correspond to the predictions experts would have chosen based on their own expertise. Success is measured as the frequency with which the model's predicted value for a target variable (e.g. quality, effort) that has the highest probability corresponds to the experts' own assessment.

Predictive Accuracy uses past data (e.g. past project data), rather than scenarios, to obtain predictions. Data (evidence) is entered on the model (see example in Fig. 2(d)), and success is measured as the frequency with which the model's predicted value for a target variable (e.g. quality, effort) showing the highest probability corresponds to the actual value from past data.

4 Detailing the Expert-Centred Approach

This Section revisits the adapted KEBN process (see Fig. 3), detailing the tasks carried out for each of the three main steps part of that process within the context of six expert-centred effort estimation models that were separately elicited by the author with the participation of Domain Experts (Des) from six different Companies (five in New Zealand and one in Brazil).

In all cases, prior to eliciting the expert-centred effort models, the DEs from all participating companies were presented with an overview of the technique that was going to be used, and examples of "what-if" scenarios, using a made-up BN model. This, we believe, facilitated the entire process as the use of an example, and the brief explanation of each of the steps in the KEBN process, provided a concrete understanding of what to expect. We also made it clear that the KE was a facilitator of the process, and that the companies' commitment was paramount for the success of the collaboration. The effort required by each company to have their Expert-centred models created and the characteristics of each model are detailed in Table 1.

Table 1. Expert-Centred Models' Characteristics and DEs

	Companies in New Zealand					Co. Brazil
Characteristics	A	B	C	D	E	F
Number of DEs	1	1	2	2	7/2	1
Number of Employees	~5	~5	~20	~30	~100	~30
Number of 3-hours elicitation sessions	12	6	8	12	12/12	20
Total hours to elicit & validate model	36	18	24	36	98	60
Effort to elicit & validate model (person/hours)	72	36	72	108	324	120
Number of factors	14	13	34	33	38	19
Number of relationships	18	12	41	60	50	37
Number of past projects used as validation set	22	8	11	22	22	9

The DEs who took part in the case studies were all project managers of well-established companies in either Auckland (New Zealand), or Rio de Janeiro (Brazil), each with at least 10 years of experience in project management. These companies varied in their size, measured as the total number of employees. In addition, all six companies were consulting companies and as such, developed a wide range of applications, from conventional software (only company E), and static & multimedia-

like to very large e-commerce solutions. All six companies employed a wide range of technologies, mostly focusing on the development of Web 1.0, 2.0 and Web 3.0 applications. Finally, when approached, they were all looking at improving their current effort estimates, and agreed to participate for two main reasons: i) because the models being created were expert-centred single-company models geared towards their specific needs; ii) and also because their expertise and participation were acknowledged as essential to eliciting those models.

Detailed Structural Development and Parameter Estimation: In order to identify the fundamental factors that the DEs took into account when preparing a project quote we used the set of variables from the Tukutuku dataset [6] as a starting point (see Table 2). We first sketched them out on a white board, each one inside an oval shape, and then explained what each one meant within the context of the Tukutuku project. Our previous experience eliciting expert-centred models in other domains (e.g. ecology, resource estimation) suggested that it was best to start with a few factors (even if they were not to be reused by the DE), rather than to use a "blank canvas" as a starting point [16].

Table 2. Tukutuku Variables

	Variable Name	Description
Project Data	*TypeProj*	Type of project (new or enhancement).
	nLang	Number of different development languages used
	DocProc	If project followed defined and documented process.
	ProImpr	If project team involved in a process improvement programme.
	Metrics	If project team part of a software metrics programme.
	DevTeam	Size of a project's development team.
	TeamExp	Average team experience with the development language(s) employed.
Application	*TotWP*	Total number of Web pages (new and reused).
	NewWP	Total number of new Web pages.
	TotImg	Total number of images (new and reused).
	NewImg	Total number of new images created.
	Num_Fots	Number of features reused without any adaptation.
	HFotsA	Number of reused high-effort features/functions adapted.
	Hnew	Number of new high-effort features/functions.
	TotHigh	Total number of high-effort features/functions
	Num_FotsA	Number of reused low-effort features adapted.
	New	Number of new low-effort features/functions.
	TotNHigh	Total number of low-effort features/functions

Within the context of the Tukutuku project, based on collected data, a new high-effort feature/function and a high-effort adapted feature/function require respectively at least 15 and 4 hours to be developed by one experienced developer.

Once the Tukutuku variables had been sketched out and explained, the next step was to remove all variables that were not relevant for the DEs, followed by adding to the white board any additional variables (factors) suggested by them. This entire process was documented using digital voice recorders and also text editors. We also documented descriptions and rationale for each factor proposed by the DEs. The factors proposed were indeed influenced by DEs' hunches and insights; however DEs decisions and choices were also very much influenced by their solid previous experience managing Web projects, and estimating development effort.

Next, we identified the possible states that each factor would take. All states were discrete. Whenever a factor represented a measure of effort (e.g. Total effort), we also

documented the effort range corresponding to each state, to avoid any future ambiguity. For example, to one of the participating Web companies, 'very low' Total effort corresponded to 4+ to 10 person hours, etc. Once all states were identified and thoroughly documented, it was time to elicit the cause and effect relationships. As a starting point to this task we used a simple medical example from [19] (see Fig. 4).

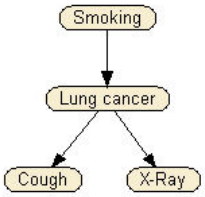

Fig. 4. An example of a cause and effect relationship

This example clearly introduces one of the most important points to consider when identifying cause and effect relationships – timeline of events. If smoking is to be a cause of lung cancer, it is important that the cause precedes the effect. This may sound obvious with regard to the example used; however, it is our view that the use of this simple example significantly helped the DEs understand the notion of cause and effect, and how this related to Web effort estimation and the BNs being elicited. Once the cause and effect relationships were identified, we worked on the elicitation of probabilities to quantify each of the cause and effect relationships previously identified. In all four cases, there was an iterative process between the structural development and parameter elicitation steps.

Detailed Model Validation: Both Model walkthrough and Predictive accuracy were used to validate all six expert-centred models, where the former was the first type of validation to be employed in all cases. DEs used different scenarios to check whether the node Total_effort would provide the highest probability to the effort state that corresponded to the DE's own suggestion. However, it was also necessary to use data from past projects, for which total effort was known, in order to check the model's calibration. Table 1 details the number of projects used by each company as validation set. In all cases, DEs were asked to use as validation set a range of projects presenting different sizes and levels of complexity, and being representative of the types of projects developed by their companies.

For each project in a validation set, evidence was entered in the expert-centred model, and the effort range corresponding to the highest probability provided for 'Total Effort' was compared to that project's actual effort. Whenever actual effort did not fall within the effort range associated with the category with the highest probability, there was a mismatch; this meant that some probabilities needed to be adjusted. In order to know which nodes to target first we used a Sensitivity Analysis report, which provided the effect of each parent node upon a given query node. Within our context, the query node was 'Total Effort'. Whenever probabilities were adjusted, we re-entered the evidence for each of the projects in the validation set that had already been used in the validation step to ensure that the calibration already carried out had not been affected. This was done to ensure that each calibration would always be an improvement upon the previous one. Once all projects were used to calibrate a model, the DE(s) assumed that the Validation step was complete.

Each of the five New Zealand expert-centred models has been in production for at least 18 months, and the Brazilian model has been in production since May 2011. Due to shortage of space we cannot show the models; however, more details about these six expert-centred models are given in [11].

5 Further Gains from Eliciting Expert-Centred Models

Except for Company E, where the expert-centred model is also used to estimate effort for non-Web-based projects, all the five remaining models represent solely the knowledge elicited from domain experts relating to their previous experience estimating effort for Web development projects; therefore we believe that by aggregating their knowledge we can obtain a wider understanding on the fundamental factors affecting effort estimation (mainly Web effort estimation) and their causal relationships.

The type of aggregation mechanism we are suggesting herein applies solely to the qualitative parts of the expert-centred models elicited (factors and relationships only), as aggregating their quantitative parts and also the different categories used to measure each factor would prove to be a herculean (if not impossible) task; in addition, our goal is not to obtain a cross-company expert-centred model.

Some of the advantages of such type of aggregation are as follows [20]:

- The aggregation from different experts of knowledge relating to the same phenomenon has the obvious advantage of providing an opportunity to amplify and broaden our overall understanding of that phenomenon.
- The graph (map) resulting from the aggregation mechanism uses as input factors and relationships from models that were built and validated using a process based on a root cause analysis technique [15]; such technique, by requiring a thorough and deep understanding of experts' mental processes and their decisions when estimating effort, provides the means to truly portray the phenomenon focus of this research, i.e. effort estimation.
- The introduction of more structure into the effort estimation process, as such map can be used as a checklist to help improve judgment-based effort estimates [1].
- Anecdotal evidence we obtained throughout the elicitation process and post-mortem meetings with several companies revealed that the use of a map aggregating companies' expert knowledge with regard to factors and relationships relevant for Web effort estimation would be extremely useful to help them elicit factors and relationships when building their own Web effort prediction models; therefore they would like to use such aggregated map at the start of their elicitation process.
- Our aggregated map shows graphically not only the set of factors and relationships from the input models, but also a way to identify visually the most common factors and relationships resulting from the aggregation. Such knowledge may also be useful to project managers to revisit the factors they consider when estimating effort for new projects.

- The aggregated map can be used to provide companies with a starting point to building a single-company expert-based Web effort estimation model. This is also an approach suggested in [1][21].

Herein we will just present the main patterns observed from aggregating those six expert-centred models [11]; however, further details on the aggregation mechanism employed and an example of the sort of aggregated map we are focusing herein is given in [20].

The main patterns observed from aggregating the six expert-centred effort estimation models are presented below.

Apart from total effort, which was identified by all participating companies, there were three factors that were chosen by five of the six Companies:

- Average Project Team Experience with Technology
- Effort to Program Features
- Project Management Effort

The next set of factors selected by four Companies were the following:

- Adaptation Effort of Features off the shelf
- Development Effort of New Features
- Effort to Develop User Interface
- Project Risk Factor
- Effort Production testing

Except for Project Risk Factor and Average Project Team Experience with Technology, all the remaining factors were related to the effort to accomplish certain tasks, such as adapting or developing a new feature, testing and interface design. Note that these factors are very much related to more dynamic Web applications, which offer a large set of features (this requiring more detailed testing). It is also interesting that the effort to develop the user interface was also chosen by four of the six companies. Nowadays, with the plethora of Web technologies and possibilities available, good interface design and usability can also add very much so to a company's competitive advantage on the global market.

All six expert-centred models presented several effort-related factors as predictors of Total effort. Note that these factors represent tasks that are part of an effort estimation process and therefore, their relationship with Total effort has an associative nature, however not of type cause&effect.

6 Lessons Learnt

The elicitation of expert-centred models, and their aggregation has provided numerous lessons, as follows:

First: engaging with industry. At the start of this research, in order to reach out to industry, we invited the local NZ IT industry to attend a seminar about software & Web effort estimation and how to improve their estimates. The seminar provided an introduction to using expert-centred models, their value as estimation tools, and their capability for running "what-if" scenarios. Many of the participating companies saw

the immediate value in such an approach, in particular because it enabled the very close and fundamental participation of in-house domain experts while building and validating the company-specific model. Several companies sign up to collaborate.

Second: Time constraints. Depending on the complexity of the expert-centred model, the elicitation of probabilities can be very time consuming. Motivated by this issue, we have also made a preliminary attempt at investigating mechanisms to enable the automatic generation of probability tables. An attempt was devised and used with two NZ companies that participated more recently in this research. This solution comprised the comparison between different probability generation algorithms and expert-driven probabilities. A tool was implemented as a result of this work [22]; however, further work is needed to validate the proposed solution.

Third: Value for a company. Except for the company in Brazil, the other participating companies were contacted for post-mortem interviews. The main points highlighted were the following:

- The elicitation process enabled experts to think deeply about their effort estimation process and the factors taken into account during that process, which in itself was considered advantageous to the companies. This has been pointed out to us by all the DEs interviewed.
- Once a company's expert-centred model was validated, DEs started to use their model not only for obtaining better estimates than the ones previously prepared by subjective means, but also sometimes as means to guide their requirements elicitation meetings with prospective clients. They focused their questions targeting at obtaining evidence to be entered in the model as the requirements meetings took place; by doing so they basically had effort estimates that were practically ready to use for costing the projects, even when meetings with clients had short durations. Such change in approach proved to be extremely beneficial to the companies given that all estimates provided using the models turned out to be more accurate on average than the ones previously obtained by subjective means.
- Clients were not presented the models due to their complexity; however by entering evidence while a requirements elicitation meeting took place enabled the DEs to optimise their elicitation process by being focused and factor-driven.
- One of the participating companies, the largest company in total number of employees, and also the one that built the largest BN model, provided the following feedback: The DEs who participated in the causal structure and probabilities' elicitation changed completely their approach to estimating effort as follows: they presented the BN model to all of their development teams, and asked that from that point onwards every estimate for any task should be based on the factors that had been elicited. This means that an entire team started to use the factors that have been elicited, as well as the BN model, as basis for their effort & risk estimation sessions. In addition, the DEs presented the model at a meeting with other company branches, so to detail how the Auckland branch was estimating effort and risk for their healthcare projects. The other branches were so impressed, in particular the one from the US, that they increased the number of Healthcare software projects outsourced to the NZ Branch, as they recognised the benefits of using a model that represented factors and uncertainties. Overall, such change in approach provided extremely beneficial to the company.

All the companies remained positive and very satisfied with the results. We believe that the successful development of these six expert-centred models was greatly influenced by the commitment of the participating companies, and also by the DEs' experience estimating effort.

Acknowledgments. We thank all the Web companies who participated in this research. This work was sponsored by the Royal Society of New Zealand (Marsden research grant 06-UOA-201), and by CAPES/PVE (Brazil).

References

1. Jørgensen, M. and Grimstad, S. Software Development Effort Estimation: Demystifying and Improving Expert Estimation, In: Simula Research Laboratory - by thinking constantly about it, ed. by Aslak Tveito, Are Magnus Bruaset, Olav Lysne. Springer, Heidelberg, chap. 26, pp. 381-404. (ISBN: 978-3642011559) (2009)
2. Mendes, E. Cost Estimation Techniques for Web Projects. IGI Global Publishers. (2007)
3. Kitchenham, B.A., Pickard, L.M., Linkman, S., & Jones, P. Modelling Software Bidding Risks. IEEE Transactions on Software Engineering. June, 29(6), 542-554. (2003).
4. Jørgensen M., Shepperd, M. J., A Systematic Review of Software Development Cost Estimation Studies. IEEE Transactions Software Engeneering 33(1), 33-53.(2007)
5. Azhar, D., Mendes, E., and Riddle, P. A Systematic Review of Web Resource Estimation. Proceedings of PROMISE'12 (accepted for publication). (2012)
6. Mendes, E., Mosley, N., and Counsell, S. Investigating Web Size Metrics for Early Web Cost Estimation, Journal of Systems and Software, Volume 77, Issue 2, August 2005, pp. 157-172, DOI: 10.1016/j.jss.2004.08.034. (2005)
7. Ruhe, M., Jeffery, R. and Wieczorek, I. Cost estimation for Web applications, Proceedings ICSE 2003, 285-294, (2003)
8. Ferrucci, F., Gravino, C., Martino, S. Di: A Case Study Using Web Objects and COSMIC for Effort Estimation of Web Applications.EUROMICRO-SEAA, p. 441-448. (2008)
9. Mendes, E. Mosley, N., and Counsell, S. Web metrics - Metrics for estimating effort to design and author Web applications. IEEE MultiMedia, January-March, 50-57. (2001)
10. Reifer, D.J. Web Development: Estimating Quick-to-Market Software, IEEE Software, Nov.-Dec., 57-64. (2000).
11. Mendes, E. Using Knowledge Elicitation to Improve Web Effort Estimation: Lessons from Six Industrial Case Studies, Proceedings of the International Conference on Software Engineering (ICSE' 2012), track SE in Practice pp. 1112-1121. (2012).
12. Mendes, E.. Knowledge Representation using Bayesian Networks A Case Study in Web Effort Estimation, Proceedings of the World Congress on information and Communication Technologies (WICT 2011), pp. 310-315. (2011)
13. Mendes, E.. Building a Web Effort Estimation Model through Knowledge Elicitation, Proceedings of the International Conference on Enterprise Information Systems (ICEIS), pp. 128-135. (2011).
14. Mendes, E., Polino, C., and Mosley, N. Building an Expert-based Web Effort Estimation Model using Bayesian Networks, 13th International Conference on Evaluation & Assessment in Software Engineering. (2009).
15. Ammerman, M., The Root Cause Analysis Handbook: A Simplified Approach to Identifying, Correcting, and Reporting Workplace Errors. (1998).
16. Woodberry, O., Nicholson, A., Korb, K., and Pollino, C. Parameterising Bayesian Networks, in Australian Conference on Artificial Intelligence, pp. 1101-1107.(2004)

17. Druzdzel, M.J., & van der Gaag, L.C. Building Probabilistic Networks: Where Do the Numbers Come From?. IEEE Trans. on Knowledge and Data Engineering, 12(4), 481-486. (2000)
18. Tang, Z., & McCabe, B. Developing Complete Conditional Probability Tables from Fractional Data for Bayesian Belief Networks, Journal of Computing in Civil Engineering, 21(4), 265-276. (2007)
19. Jensen, F. V. (1996). An introduction to Bayesian networks. UCL Press, London.
20. Baker, S., Mendes, E. Aggregating Expert-driven Causal Maps for Web Effort Estimation, Proceedings of the Advances in Software Engineering (ASEA) Conference: Communications in Computer and Information Science, Volume 117, pp. 264-282, DOI: 10.1007/978-3-642-17578-7_27 (2010).
21. R. Montironi, W. F. Whimster, Y. Collan, P. W. Hamilton, D. Thompson, and P. H. Bartels, How to develop and use a Bayesian Belief Network, Journal of clinical pathology, vol. 49, p. 194. (1996)
22. Baker, S., Mendes, E. Evaluating the Weighted Sum Algorithm for Estimating Conditional Probabilities in Bayesian Networks, Proceedings of the Software Engineering and Knowledge Engineering Conference (SEKE 2010), pp. 319-324. (2010)