David P. Dupplaw
School of Electronics and Computer Science,
University of Southampton,
Southampton SO17 1BJ,
England

David Brunson
Jaguar Cars Limited,
Whitley Engineering Center,
Coventry CV3 4LF,
England

Anna-Jane E. Vine
Statistical Sciences Research Institute,
School of Mathematics,
University of Southampton,
Southampton SO17 1BJ,
England

Colin P. Please
School of Mathematics,
University of Southampton,
Southampton SO17 1BJ,
England

Susan M. Lewis
Email: S.M.Lewis@maths.soton.ac.uk
Statistical Sciences Research Institute,
School of Mathematics,
University of Southampton,
Southampton SO17 1BJ,
England

Angela M. Dean
Department of Statistics,
Ohio State University,
Columbus, OH 43210-1247,
USA

Andy J. Keane
School of Engineering Sciences,
University of Southampton,
Southampton SO17 1BJ,
England

Marcus J. Tindall
Department of Mathemtics,
University of Oxford,
Oxford OX1 3LB,
England

A Web-Based Knowledge Elicitation System (GISEL) for Planning and Assessing Group Screening Experiments for Product Development

When planning experiments to examine how product performance depends on the design, manufacture and environment of use, there are invariably too few resources to enable a complete investigation of all possible variables (factors). We have developed new algorithms for generating and assessing efficient two-stage group screening strategies which are implemented through a web-based system called GISEL. This system elicits company knowledge which is used to guide the formulation of competing two-stage strategies and, via the algorithms, to provide quantitative assessment of their efficiencies. The two-stage group screening method investigates the effect of a large number of factors by grouping them in a first stage experiment whose results identify factors to be further investigated in a second stage. Central to the success of the procedure is ensuring that the factors considered, and their grouping, are based on the best available knowledge of the product. The web-based software system allows information and ideas to be contributed by engineers at different sites and allows the experiment organizer to use these expert opinions to guide decisions on the planning of group screening experiments. The new group screening algorithms implemented within the software give probability distributions and indications of the total resource needed for the experiment. In addition, the algorithms simulate results from the experiment and estimate the percentage of important or active main effects and interactions that fail to be detected. The approach is illustrated through the planning of an experiment on engine cold start optimization at Jaguar Cars.
[DOI: 10.1115/1.1778192]

Introduction

In multi-national companies where expertise is distributed across many centers, it is important to be able to access and pool existing knowledge and experience in order to guide the effective and efficient design of new and existing products and manufacturing processes. Design improvements can be identified by the use of planned experiments (design of experiments or DOE) to determine the crucial factors in the design. A first step in such problems is therefore to identify a list of possible factors that might influence product performance. Often, in practice, a very large number of contending factors is identified for possible investigation and this makes direct application of conventional DOE methods impractical. The existing knowledge-pool within the company concerning these factors needs to be accessed and interrogated in order to optimize the competing risks of using an overly large experiment that would waste resources, against using an experiment that is too small and may produce confusing results. The aims are to eliminate from consideration those factors whose influence is generally believed to be negligible and to guide decisions on how to investigate the remaining factors most efficiently. In this paper we present a web-based software system that (1) implements a methodology to identify factors and extract knowledge about their action by use of a dynamic questionnaire which

Contributed by the Computer Aided Product Development (CAPD) Committee for publication in the JOURNAL OF COMPUTING AND INFORMATION SCIENCE IN ENGINEERING. Manuscript received July 31, 2003; Revised May 29, 2004. Associate Editor: S. Szykman.

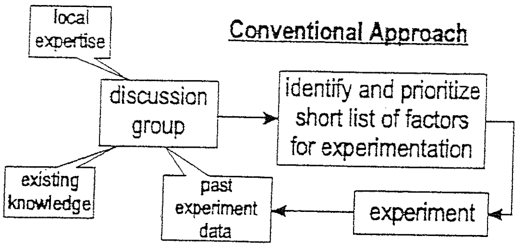


Fig. 1 A common approach to gathering information for industrial experiments

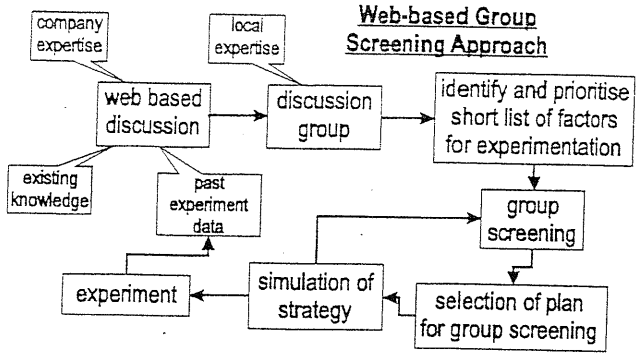


Fig. 2 Elicited expert data-led approach to information gathering

can be used iteratively, (2) allows users to summarize and explore the knowledge gathered in order to inform and suggest various possible strategies for planning efficient screening experiments and (3) calculates numerical comparisons between possible strategies so that sound decisions can be made. The system is called GISEL (Grouping In Screening with Elicitation). The paper illustrates the use of this system to plan experiments for engine cold start optimization at Jaguar Cars.

Background

The most common approach to gathering information to inform the planning of industrial experiments is through the use of "brain-storming" (see Fig. 1). In this approach a set of people with relevant expertise is identified and brought together for a focused meeting run by an organizer or moderator. The purpose of the meeting is for ideas to be put forward and a consensus reached on the factors for investigation and the role that the factors might play in the performance of the product. The outcomes from such a meeting would be (i) a list of factors to be varied during the experiment because they are believed to be important for product performance, and (ii) a list of unimportant factors to be kept constant in the experiment so that any performance variability that they might cause will not mask the impact of the other factors. In addition, any available knowledge and opinions would be pooled on joint action (interaction) between, typically, pairs of factors that need to be investigated or taken into account in the planning of experiments. There are many disadvantages of such unstructured brainstorming sessions including the fact that the dynamics of the meeting, such as the personalities or status of the participants, may affect the identified lists of factors. A discussion of bias and practical guidance on knowledge elicitation is available in [1].

Several other elicitation methods are in use that avoid or reduce this type of bias caused by interactions between experts. The most labor-intensive and time-consuming is the individual interview of each expert. More economically, a conventional questionnaire can be used to gather data, for example in surveys, and to inform the use of Bayesian methods [2,3] of data analysis. An alternative approach to eliciting expert opinions for experiments is given by [4]. From our experiences with collaborating companies in the aeronautical, agricultural, electromechanical and automotive fields, where planned experiments can take many different forms, we have found that the fixed structure of a conventional questionnaire exerts too much control on the format of the answers and, by its nature, excludes the possibility of generating dialog. Our approach uses a questionnaire to collect available knowledge about factors and exploits web-based technology to allow the questionnaire to evolve as new ideas are contributed by the experts.

A variety of knowledge elicitation systems have been developed arising from research in the fields of judgment and decision making, human factors, cognitive science and expert systems, see [1,5,6] for reviews and analysis. A significant advance on the conventional use of questionnaires was the development by the Rand corporation of the Delphi method see, for example, [7]. This method has been used widely in a variety of areas, such as those described in [8], for eliciting and forming a group judgment from experts' opinions. The elicitation method within GISEL has several features in common with the Delphi method, such as the iterative nature of the information gathering and the opportunity for the organizer to feedback unattributed judgments to the experts. It overcomes several disadvantages of the Delphi method through allowing short sessions, scheduled and controlled by the participants, email communication to enable the organizer to initiate and maintain active participation, the facility for experts to incorporate new ideas for additional factors, and on-line access to the results of both the elicitation process and subsequent experiments. The main feedback mechanisms between the organiser and the experts are the rapidly evolving questionnaire and the collected data. The system allows the organizer to assign a level of access to the feedback to address bias and commercial confidentiality. In addition, all experts remain anonymous to all except the organizer. Analysis of the data can be undertaken at any time and by experts with appropriate access. The interactive development of the questionnaire contrasts with the more sequential transmission and testing of questionnaires used in the traditional Delphi method.

Our experience is that when product improvement is primarily in the domain of a small group of people then important factors may be overlooked. This can be due to the group of people bringing very similar experience, knowledge and scientific training to the problem and can be overcome by combining the views of diverse groups of people. Most importantly, experts' opinions from both the design and manufacturing areas within a company need to be accessed (see Fig. 2). A further possible source of bias in elicitation arises from the selection of the experts by the organizer. When the organizer relies on his or her personal knowledge, or that of contacts within the company, then bias may result. More objective methods, such as searching the literature and company reports, may reduce this source of bias, see, for example, [9] which describes recent developments of agent-based methods.

The purpose of the information acquisition in GISEL is to guide the organizer on how to plan the experiment in an effective and economical manner. The aim is not to use the information to give a fixed prescription for an experiment plan, but to allow the organizer to explore a variety of possible plans in light of their interpretation of the information gathered and any additional economic and practical constraints associated with the project. In addition, the information gathered may also be useful when the data from the experiments are to be analyzed using Bayesian methods [10].

In this paper we present a web-based system that includes knowledge elicitation and also two novel algorithms for the assessment of strategies for two-stage group screening experiments, [11,12]. Such experiments may be used as a preliminary stage of product improvement and we describe a specific application in the automotive industry. The aim is to find those factors whose main effects and interactions ([13], p. 87) are sufficiently large to produce a substantive improvement (in engineering terms) in the mean and variation of the product performance. The minimum size of this worthwhile improvement is judged by engineers on the basis of scientific and cost returns from the improved performance.

In screening experiments, factors are typically investigated at two values or levels, called the "high" level (believed likely to produce the highest performance or response across the factor range to be explored) and a "low" level (where the lowest response is anticipated). In group screening, these individual factors are first put in groups and the factors within each group are varied *together* between these two levels. This enables new "grouped factors" to be defined, each of which represents a group of factors.

This type of screening involves two stages of experiment. At the first stage, an observation is made on each of the various different combinations of the levels of the grouped factors according to an experiment plan, for example a fractional factorial plan [13]. From an analysis of the data (such as estimation of main effects and interactions or a Bayesian analysis), a decision is made on which of the groups are important. The individual factors within these groups are then taken forward to a second stage experiment whose aim is to identify the key individual factors that affect product performance.

Two different two-stage group screening strategies are available in the software system described here; for further technical details see [11,12]. In one strategy, only main effects are investigated at the first stage (*classical group screening*) whilst, in the second, interactions between pairs of factors as well as main effects are examined at the first stage (*interaction group screening*). In product development experiments, the factors are usually of two types: control or design factors, whose values can be set during the manufacturing process, and noise factors whose values vary as they arise from the environment where the product is used, or from variation in the manufacturing process. Noise factors can often be set, or mimicked, in an experiment. The technique takes account of the fact that some interactions, particularly those between control and noise factors, are of greater interest and value than others in product improvement. In forming groups of factors, only factors of one type (control or noise) are put in a group in order to allow examination of interactions between grouped control and noise factors.

A key question when an experiment is being planned is whether the resource requirements are economically feasible. In two-stage group screening, the total experiment size cannot be determined with certainty at the planning stage because of the two-stage nature of the study. Information about the likely importance of the factors, elicited from experienced engineers, can be used to derive the probability distribution of the predicted total number, S, of individual main effects and interactions to be estimated. This number serves as a surrogate for the total experiment size.

In order to decide on the strategy, that is, the number of groups of factors and how many factors to assign to each group, the following practical criteria may be used:

(a) minimize the expected total experiment size

(b) minimize the probability of exceeding a budgeted target for experiment size

(c) maximize the probability of detecting the important design factors, especially those which can be set to make product performance less sensitive to varying noise factors.

In the software system described here, criteria (a) and (b) can be implemented by using the calculated distribution of S, and criterion (c) by simulation of two-stage group screening experiments.

Methodology

The process of setting up a plan for an experiment on physical products is often a recursive process. Hence a system to plan such experiments needs to be flexible and interactive. In addition, the steps in the process need to be documented so that information that leads to new understanding can subsequently be properly accessed and examined. The strategy employed is to use as much information as possible to guide the grouping of factors in a two-stage experiment. The procedure is not meant to be algorithmic but rather to act as a tool for informing decisions on possible grouping strategies for the factors involved in the particular product and its manufacturing process.

1. **Setup**—The organizer of the experiment determines the aspect(s) of the product to be investigated. In particular, the primary measure of "performance" of the product, together with an initial list of those factors within the product that may affect the performance. Each factor is identified as a control or noise factor and, for each factor, a range of values is suggested. For the particular experiment being considered the organizer must supply (i) a target for the total number of individual observations on the experimental products that can be used in the experiment based on available time and resources, (ii) an estimate of the minimum change in performance measure that would be considered to be of practical value, (iii) an estimate of the likely error in the measured observations which are available from previous experiments or pilot studies. This initial information may be determined by the organizer or created in a small local brainstorming session.

In making these lists it is important that the organizer seeks to minimize the scope for misinterpretation of the performance of the product, the features represented by the factor names and the meaning of the ranges of the factors. We have found that by piloting the methodology on a small group of people, followed by discussion of the interpretations of terms and the results, terms can be refined and many of the potential difficulties of interpretation eliminated. Following [1], Chaps. 4 and 9, we recommend that all experts be provided with as much background information and training as possible, making use of email and supporting websites.

2. **Acquisition of Information**—In this phase a wider range of views is elicited from other experts. The organizer is required to identify experts, distribute user names and assign levels of access to the acquired information.

The experts' opinions are sought on the importance of each factor in affecting the product performance and on the anticipated form of this influence as the factor level changes over the range proposed by the organizer. This information is gathered through direct questions and also through experts expressing their views in a Comments box. The immediate use of this information is restricted to screening experiments; for example, factors are limited to having only two levels. However, the information elicited may be wider and useful in other design procedures and future experiments. Information in both quantitative and qualitative form is gathered, together with an indication of each responder's confidence in their views. It is also important that the experts have the opportunity to identify, and incorporate into the elicitation process, any factors that have been overlooked. When such new factors are identified, opinions on the importance and form of their effects and the existence of any likely interactions need to be gathered. The system allows those completing the questions to add new factors which may, in turn, be commented on by others in an iterative fashion. The organizer may make decisions about the experiment based on the information elicited via the web or, as we have done, use a smaller local discussion group to further inform the final decisions. Although the latter approach may introduce additional bias into the decision-making process, we have found that it has the advantage of allowing local operational concerns in the running of the experiment to be fully explored and resolved.

A potential disadvantage of elicitation using an evolving questionnaire is that it may take so long to complete that participants' motivation, and the resulting quality of information, may become poor. Such issues can be addressed at the pilot stage, see also [1], Chap. 4. Our experience so far is that the questionnaire has not been prohibitively long and that the option of several short elicitation sessions at each expert's convenience has encouraged its completion.

3. **Summary of Importance**—The information gathered from the questionnaire is summarized in a number of different forms. This allows the organizer to examine the information in order to assign a simple subjective probability to the main effect of each factor being active. (Such probabilities are needed for criteria (a) and (b) of the previous section). Some factors will be expected to have little effect on the performance, some to have a substantial effect and for others there will be much less certainty and consensus about their influence on performance. The organizer may decide not to include factors in the experiment that are generally judged to be of little importance in order to reduce the experiment size. He/she is likely to investigate factors for which there is no consensus about their importance so that more technical information and experience can be gained from the experiment. This stage of the process uses both the quantitative information, in the form of spreadsheet data, and qualitative information in order to draw up the final list of control and noise factors for investigation.

4. **Choice of Groupings and Strategy**—Different strategies (classical or interaction group screening) and group sizes for the two stage experiment are assessed and compared. For a particular strategy, a choice for the number of groups and the numbers of factors in each group, and the probabilities from phase 3 are used to assess the total expected experimental effort required to examine (i) the effects of the grouped factors and (ii) the effects of the individual factors in the second stage of the experiment. The novel algorithms developed for this purpose are general in that they allow any groupings of the factors to be investigated under each of the group screening strategies and for any choice of probabilities for the individual main effects and interactions. By considering a number of different groupings, ranging from a small number of large groups to a large number of small groups, a short list of economical choices can be formed using criteria (a) and (b) above for each of the strategies. If the experiment cannot be carried out within the available resources, then the factors for inclusion should be reassessed and the procedure iterated.

5. **Simulation**—Having identified a short list of groupings and strategies that can be expected to fit within the available resources, following the use of criteria (a) and (b), the organizer (or other user) then makes an informed selection from the short list by considering the proportions of active effects that are likely to be missed (criterion (c)) calculated from simulations of the group screening experiment. The simulation enables two important practical concerns to be addressed: (i) where the direction of effects is unknown or incorrectly assessed, it is possible that a grouping may inadvertently be made in such a way that the contributions to product performance from the factors within a group cancel each other out, and (ii) within a group, several very small contributions from factors acting in the same direction may combine to mask the effect of an active factor that acts in a different direction. The simulation software calculates the proportions of important main effects and interactions that fail to be detected due to the grouping in the simulated experiments and the user can make an informed judgment using criterion (c) of the previous section. Note that, as simulating experiments for all possible groupings would be computationally intensive, only the best groupings and strategies identified under criteria (a) and (b) are normally considered. However, the system has the facility to simulate any specified grouping for either strategy.

Finally, the organizer should decide if one of the strategies and groupings considered lies within available resources and is likely to have acceptably small proportions of undetected active main effects and interactions. If a suitable strategy and grouping has not been found, then an increase in resources for the experiment or a reduction in the number of factors investigated should be considered and the process iterated.

The above five steps guide the planning of a two stage group screening experiment.

Software Implementation

The above steps have been incorporated into a software system which allows the necessary information to be collected and organized in an interactive manner. The design of the flow of the software system reflects the flow of the experiment planning process with the possibility of return to various steps at any time being an integral part of the procedure. The system resides on a central server where all the information is stored and calculations performed. Currently the system has been tested on both Linux and Windows XP operating systems. The system has been carefully constructed to provide transparent access to users. All interaction with the system is through a web browser. The browsers that have been tested include Internet Explorer, Netscape Communicator and Opera. The system has a number of different levels of user, namely administrators—who can initiate and analyze investigations into new product problems, users—who can input information and analyze various strategies, and guests—who can only input information to particular problems.

The system is based on open source software available over the internet. This provides a cost effective way to provide much of the functionality required. Installation of the software has to be carried out by a network administrator or support staff. It requires the installation of four pieces of software:

Apache web server—to provide the web-based interactive environment that allows the users to be physically dispersed.

MySQL database server—to provide the database structures needed to control the system and to collect and store the information within the system.

PHP Tools and code—to provide the functionality and flexibility of the system.

Group screening and simulation code—to perform the necessary calculations needed in the assessment procedure.

All these tools are provided with the software but require root or administrator privileges to install.

Central to the software is the interactive questionnaire that elicits from users and guests their opinions on the importance of factors and their possible influence on product performance. Traditionally this would be done by a paper-based method, but this software allows the information to be gathered via the world-wide web, avoiding duplication errors, the labor of aggregating results, and printing and administration costs of the paper questionnaires.

The user is presented with a split screen which, on the left, displays a list of the possible factors in an experiment and, on the right, a panel where questions relating to each factor will be shown. When the user clicks on the name of a factor, a questionnaire is presented to the user in the right-hand panel. The questionnaire tries to help the user to quantify his/her understanding of the influence of the factor on the product performance by asking the questions shown in Fig. 3.

The possible responses to the questions shown in the figures contain only approximate ideas of the response. If the user indicates a belief that varying the value of this factor will alter the performance of the product, the questionnaire asks for an indication of the expected trend of this response. Using "A" as the low value for the factor and "B" as the high value, the user indicates whether the performance of the product is expected to greatly increase, greatly decrease, slightly increase, or slightly decrease as the factor changes from "A" to "B". In the current implementation, such trends are only shown as linear. If a user considers the response to be a non-linear function of the factor values across the defined range, then the unstructured comment section allows a written explanation of this opinion and the opportunity to relate views on the functional form to the choice of factor levels. For each question, users select from the radio buttons the answer which matches their opinion. The questionnaire responses are stored in a database, and are recalled when a user logs back into the system. This means that users can complete as much of the questionnaire as convenient and then come back at another appro-

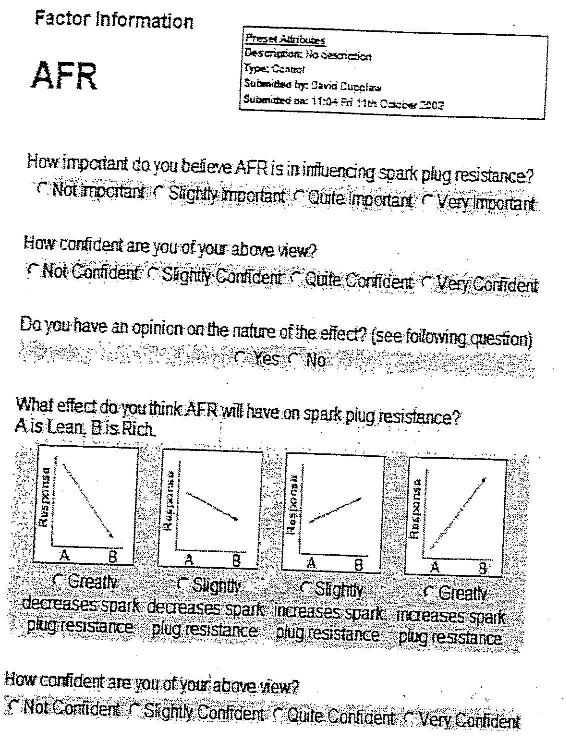


Fig. 3 Page in the questionnaire on the response trend of a factor

priate time to complete more. They may also alter their previously submitted opinions on a factor by reselecting a factor from the factor list.

If a user addressing the questionnaire believes that a factor that is not listed may be important to the performance of the product, then the user may add the factor through the "Add My Own" button (not shown) at the bottom of the factor list. When this is done, a screen is presented into which the factor name, a brief description of the factor and the possible levels for consideration may be entered. On submission of this new factor, the factor list is updated and the questionnaire can be immediately completed for this new factor. The additional factor will be visible to any new, or returning, users. As the user submits opinions with the "Submit Opinion" button, the factor list is updated on the left-hand side of the page. If the user has correctly filled in the fields, then a small tick is placed next to the completed factor (see Fig. 4). If a mistake was made during the completion of the questionnaire (for example, not all mandatory fields were completed), then a small

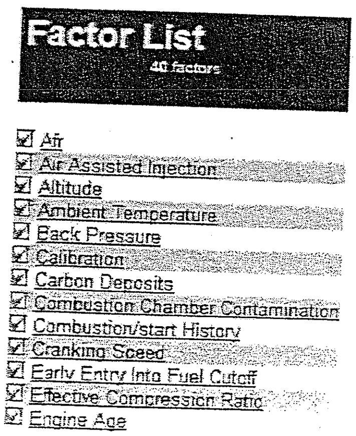
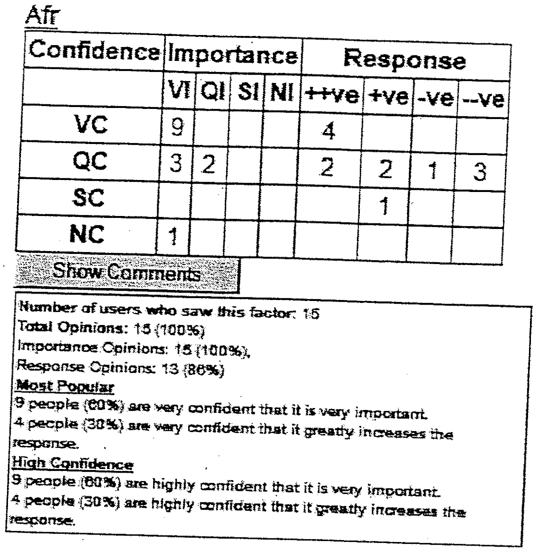


Fig. 4 Page to indicate some factors under consideration



Afr

Confidence	Importance				Response			
	VI	QI	SI	NI	++ve	+ve	-ve	--ve
VC	9				4			
QC	3	2			2	2	1	3
SC						1		
NC	1							

Show Comments

Number of users who saw this factor: 15
Total Opinions: 15 (100%)
Importance Opinions: 15 (100%),
Response Opinions: 13 (86%)
Most Popular
9 people (60%) are very confident that it is very important.
4 people (30%) are very confident that it greatly increases the response.
High Confidence
9 people (60%) are highly confident that it is very important.
4 people (30%) are highly confident that it greatly increases the response.

Fig. 5 Page to summarize questionnaire data on a particular factor

cross will be visible next to that factor. Factors where the current user has given no opinion will have no visible marking next to the factor name.

The summary of the data collected through the questionnaires is presented in several forms. At any time, users can view all opinions including written comments. However, to aid understanding of the quantitative data from the questionnaire, the data are summarized in small tables. These show not only the number of opinions given for each possible response direction, but also the confidence that the users expressed in their opinions. An example of such a table is given in Fig. 5. In the example, most responders were confident that the factor in question was very important. This display helps to give the administrator a clear view of the importance of each factor and hence allows an allocation of a probability to the factor being important for the later grouping stages. Since there is seldom sufficient information to allow any precise probabilities to be defined, the factors are categorized into one of the three categories (i) Very likely to affect performance (ii) Less likely to affect performance (iii) Negligible effect on performance. Probabilities for these categories are then requested, including probabilities that indicate the importance anticipated of the interactions between pairs of factors. These can either be entered manually or can be calculated using the approach of [3].

The software then allows the user to assess both classical group screening and interaction group screening strategies over two stages of experiment. Our experience of these methods indicates that, although interaction group screening usually requires more experimentation than classical, the risk of failing to detect active interactions can be far less, particularly for the important interactions between control and noise factors.

For a given grouping of the factors, specified by the user as in Fig. 6, the software allows the user to calculate a measure of the probability of exceeding the available experimental resource for both strategies for any specified set of groupings. In order to simulate the two stage experiments, the software requires plans for the first stage of experiment on the grouped factors. Currently, efficient fractional factorial plans [14] that are tailored to prioritize the investigation of main effects and interactions between control and noise factors are provided by the software by default, with the option of a user-supplied plan being available. The results of the simulations are stored and, at any time, a user can review all the assessments of any particular grouping and strategy. (The calculations can be fairly computationally intensive, so it is preferable to store results rather than re-run them). There is no automatic ranking of these results and the user is given several ways of

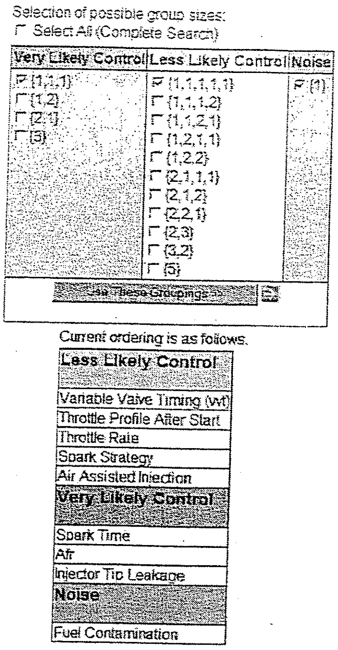


Fig. 6 Page to indicate possible group sizes for assessment by software

viewing the data to aid in making a decision. Figure 7 shows an example of a comparison of two different groupings with the two different strategies using criterion (b) using superimposed graphs of the probability of exceeding a target experiment size. Once a particular choice of strategy, number and sizes of groups is identified as good, the simulation software is then run repeatedly to calculate the proportion of experiments in which important effects are missed.

Case Study

The software described above has been used to plan a fairly large experiment at Jaguar Cars. This experiment concerned the optimization of cold start performance using spark plug resistance measurements on a particular engine type. Design of experiment methods are commonly used by Jaguar Cars for product improvement and the methodology described in this paper fitted in well

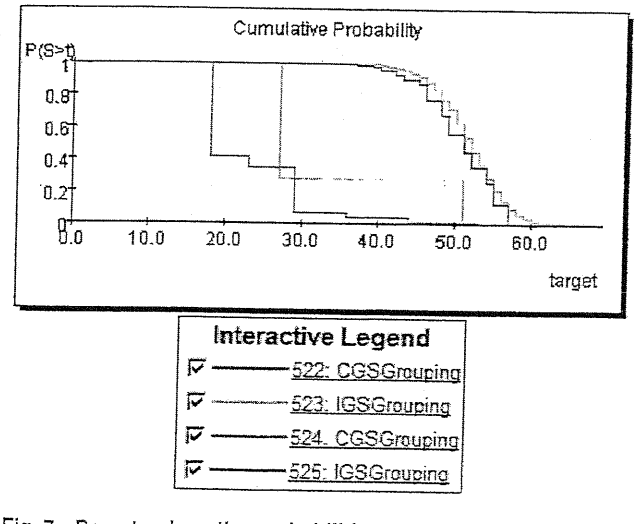


Fig. 7 Page to show the probabilities of a specified target size t being exceeded for two groupings with IGS and CGS strategies

with their regular DOE working practices. The main difference here was the use of the web-based software both to enhance their usual brainstorming sessions in guiding factor selection and to exploit the group screening methodology to allow more factors to be considered than would be possible by conventional DOE methods.

1. **Setup**—After suitable local discussions, the performance measure to be used to assess the quality of the product was identified as the percentage of a given number of engine cycles with a "low" spark plug resistance. The expected resource available for the total experiment was anticipated to be in the region of one hundred observations on different engine configurations. A two percent change in performance measure was perceived as the minimum worthwhile change and a standard deviation of the performance measurements of about 0.8 percent was obtained from a pilot study. An initial list of factors for consideration for the experiment was drawn up based upon previous, smaller scale investigations and local knowledge. Each of these factors was also identified as being either a control or a noise factor.

2. **Information gathering**—User names for access to the web-based software questionnaire were distributed to relevant Jaguar, Land Rover and Ford centers in the UK and USA. Two iterations of the questionnaire were used with a local discussion occurring after receipt of the initial returns. This discussion was particularly focused on the number of new factors that had been proposed. Subsequently, responders were asked to comment again particularly on these new factors. Of special value was the general comment section of the questionnaire as this led to the need to clarify the definition and levels of certain of the factors.

3. **Summary of importance**—A total of fifteen people contributed to the questionnaire which gave a greater range of input than was obtained from the usual local elicitation methods. The data were summarized into the tables discussed earlier. In addition the data helped to guide subsequent local meetings by the identification of certain factors as consistently judged important by the experts, and other factors as consistently viewed as irrelevant. The face-to-face sessions could therefore concentrate on the more controversial factors identified through the questionnaire and, in particular, on whether they should be included in the experiment. It was decided that, from the full list of factors, there were twelve control factors and two noise factors that needed to be investigated in the experiment. Of these, five control factors were believed to be very likely to be active with confident information on the expected direction of their effects. Each of these factors had a value of 0.7 assigned to the probability that the main effect is active. Our software investigations have shown that the assessments based on the expected size of the experiments (see criteria (a) and (b)) are insensitive to this precise value with similar results occurring when the probability value was increased up to 1.0. There was much less consistency of opinion amongst the experts on the remaining control factors but, overall, there was some support for the view that they were likely to be active. However, little consistent information was available on the direction of their influence. It was anticipated that perhaps only one such factor would actually be active so each of them was assigned a probability of 0.14 (=1/7). Finally, it was anticipated that at least one of the two noise factors would be active so a probability of 0.5 was assigned to each. The interaction probabilities were assigned using the approach of [3].

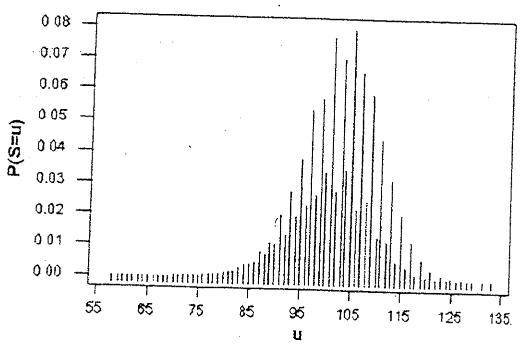
4. **Grouping**—Many possible strategies for grouping the factors are available and were considered. It was found that using groups of similar size produced experiments that were expected to be smaller than those with very disparate group sizes. Two approaches to grouping the noise factors were considered: a single group, and two separate groups. As an illustration, Fig. 8 shows the results of the grouping software for the case where the "very likely to be active" control factors are in two groups of sizes 2 and 3, "less likely to be active" control factors are in four groups of sizes 2, 2, 2, 1, and the noise factors are in a single group. The

Table 1 List of factors for experiment from the questionnaire and discussions

Control factors	Noise factors
AFR	Injector tip leakage
Spark Time	Ambient Temp./Humidity
Calibration	
Engine off timing	
Idle Speed	
Plug Type	
Injection timing	
Spark Advance	
Transient fuel with calibration	
Plug gap	
Variable valve timing	
Injector spray angle/direction	

figure displays the distribution of the predicted size of the experiment and the resulting risk of exceeding various targets. It was found that the experiment would fit within the available resources and, as expected, that for a given grouping, classical group screening (CGS) would need considerably less resource than interaction group screening (IGS). Two economic groupings, for each method, were taken forward to be assessed further by simulation. A summary of some of the properties of the two selected IGS groupings is given in Table 2.

5. **Simulation**—As expected from earlier results, the simulation showed that the groupings with CGS risked missing a far higher percentage of active interactions between control and noise factors than the groupings with IGS. As these interactions were of particular interest, the IGS method was adopted. For deciding between one or two groups for the noise factors, the results showed that there was little difference in the expected total experiment size, but the proportion of missed active interactions between control and noise factors was much greater when the noise factors were in a single group. Hence two noise groups were used each composed of a single noise factor. The fractional factorial plan used to simulate the experiment then provided the basis of the plan for use in the actual experiment. The final plan required randomization before use and this had to accommodate the practical constraint that four of the factors had levels that were very time-consuming to change.



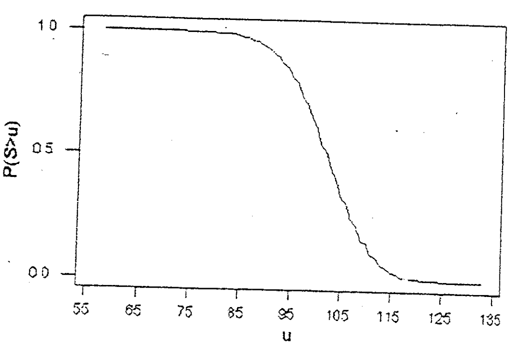


Fig. 8 Information for assessment of a grouping for an experiment under interaction group screening

Table 2 Results from two possible noise factor groupings using IGS, including the mean $E(S)$ and the standard deviation $\sigma(S)$.

5 very likely control factors	7 likely control factors	2 noise factors				
Grouping			$E(S)$	$\sigma(S)$	Prob. $S>110$	Prob. $S>120$
2,3	2,2,2,1	2	102	7.57	0.15	0.01
2,3	2,2,2,1	1,1	106	7.20	0.26	0.01

Conclusions and Extensions to Functionality

The use of a web-based system was found to have considerable advantages over more common approaches to information gathering for planned experiments. It enabled an international group of experts to pool their ideas and knowledge in order to plan an efficient and effective experiment to examine factors that are relevant to product improvement. A further advantage was that the software provided a record of the information underlying any new decisions made in changing a product design, as well as an archive of the new knowledge found from the experiments. The software also made accessible new research developments in the area of group screening that allow investigation of a larger number of factors than can be accommodated through conventional approaches. Interaction group screening results in a larger experiment than conventional group screening. However, through investigation of a larger number of factors and interactions between them, the experiments enable more accurate identification of the important factors and a decrease in the probability of missing key interactions. Hence these methods are economically more efficient and reduce the chance of the results of the experiment being inconclusive.

In order to ensure that the software is adequately responsive for general use, improvements have been implemented in the speed of the algorithms. Also, extra functionality has been integrated to allow further elicitation of information and greater user feedback. In particular, the information gathering sub-system has been improved (i) by saving all changes made by respondents to previous opinions thereby indicating where new information has entered the system and (ii) by current development of an XML based generic questionnaire that allows for greater tailoring to the specific problem. In addition, the history function has been enhanced to allow any previously explored grouping strategy to be considered in the simulation sub-system.

The installation of the system into any organization would be straightforward and cost effective and require only simple software installation. The two stages of experiment outlined in this paper are scheduled to be performed at Jaguar Cars. Those wishing to try the system as beta-testers are welcome to contact Professor Susan Lewis for details.

Acknowledgment

This research was supported by the Engineering and Physical Science Research Council (Grant GR/N16754). Thanks are due to Jaguar Cars, Hosiden Besson and Goodrich Engine Control Systems for their support and collaboration on this work. We are grateful for the comments of a referee which greatly improved the paper.

References

[1] Meyer, M. A., and Booker, J. M., 2001, *Eliciting and Analyzing Expert Judgment*, ASA-SIAM, Philadelphia.

[2] Box, G. E. P., and Meyer, R. D., 1993, "Finding the Active Factors in Fractionated Screening Experiments," J. Quality Technol., 25, pp. 94–105.

[3] Chipman, H., 1996, "Bayesian Variable Selection With Related Predictors," The Canadian Journal of Statistics, 24, pp. 17–36.

[4] Grimshaw, S. D., Collings, B. J., Larsen, W. A., and Hurt, C. R., 2001, "Eliciting Factor Importance in a Designed Experiment," Technometrics, 43, pp. 133–146.

[5] Cooke, R. M., 1991, *Experts in Uncertainty: Opinion and Subjective Probability in Science*, Oxford University Press, New York.

[6] Hoffman, R. R., Shadbolt, N. R., Burton, A. M., and Klein, G., 1995, "Eliciting Knowledge from Experts: A Methodological Analysis," Organisational Behavior and Human Decision Processes, 62, pp. 129–158.

[7] Helmer, O., 1968, "Analysis of the Future: The Delphi Method," *Technological Forecasting for Industry and Government*, Bright, J., ed., Prentice-Hall, Englewood Cliffs, N.J., pp. 116–134.

[8] Adler, M., and Ziglio, E., eds., 1996, *Gazing into the Oracle: The Delphi Method and Its Application to Social Policy and Public Health*, Jessica Kingsley, London, UK.

[9] Crowder, R., Hughes, G., and Hall, W., 2003, "An Agent Based Approach to Finding Expertise in the Engineering Design Environment," *Proc. 14th International Conference on Engineering Design*, A. Folkeson et al., eds.

[10] Chipman, H., Hamada, M., and Wu, C. F. J., 1997, "A Bayesian Variable-Selection Approach for Analysing Designed Experiments With Complex Aliasing," Technometrics, 39, pp. 372–381.

[11] Lewis, S. M., and Dean, A. M., 2001, "Detection of Interactions in Experiments on Large Numbers of Factors" (with discussion), J. R. Stat. Soc. Ser. B. Methodol., 63, pp. 633–672.

[12] Dean, A. M., and Lewis, S. M., 2002, "Comparison of Group Screening Strategies for Factorial Experiments," Comput. Stat. Data Anal., 39, pp. 287–297.

[13] Myers, R. H., and Montgomery, D. C., 2002, *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, Wiley, New York.

[14] Russell, K. G., Lewis, S. M., and Dean, A. M., 2004, "Fractional Factorial Designs for the Detection of Interactions Between Design and Noise Factors," Journal of Applied Statistics, 31, 545–552.