INTERNATIONAL CONFERENCE ON ENGINEERING DESIGN
ICED 05 MELBOURNE, AUGUST 15-18, 2005

# INTRODUCING GRID-BASED, SEMI-AUTONOMOUS EVOLUTIONARY DESIGN SYSTEMS

I.C. Parmee, J. Abraham, M. Shackelford, D. Spilling, O. F. Rana, A. Shaikhali

## Abstract

The paper describes the initial development of the data modelling and search, exploration and optimisation processes (SEO) of a Grid-enabled problem solving environment (PSE). This environment will enable a client to access coupled computational components sited at different "centres of expertise". Each centre offers a data generation and analysis approach that aids a better understanding of the design domain whilst providing a route to the identification of appropriate high-performance design solutions. The intention is to support satisfactory, remote problem definition that leads to the selection and application of appropriate design search, exploration and optimisation techniques. This should occur seamlessly so that the client is unaware that these processes are to be undertaken at different sites. The intention is that the system will support clients with extensive knowledge of their design domain but little expertise in state-of-the-art data modelling and SEO processes.

*Keywords: Conceptual Design, Data Modelling, Evolutionary Computing, Grid Computing*

## 1. Introduction

The paper introduces the initial development of a problem solving environment (PSE) which integrates data modelling and search, exploration and optimisation (SEO) services located at geographically distributed centres of expertise. Integration is supported through Grid technologies [1] which provide access to modelling and SEO software components via a Grid-workflow engine. The components will provide high quality design information and solutions that support design decision-making. These have been integrated with a Grid computing infrastructure to provide a seamless (standards-based) connectivity between the centres and the secure access to computational resources needed to execute the modelling and optimisation services that are available. The paper concentrates upon the development of the Modelling and SEO components.

The intention of the system (DIPSO – Distributed Problem-solving Environment) is to provide a service for those clients who either do not possess the necessary in-house expertise relating to data modelling and SEO or who need to access an off-site capability to supplement their in-house resource. A description of the functionality of each component of the distributed system architecture follows. **The Modeller** component can be accessed by clients who require a **data modelling** service. The component comprises a number of ancillary data processing techniques plus neural network and statistical modelling software for the generation of models from incoming data sets passed into the system by the client. **The Interrogator / Optimiser** extracts information relating to design space characteristics either from such data models generated within the Modeller or directly from a **parametric model** that resides with a client.

The **Interrogator** prototype comprises a number of space-sampling techniques to identify well-distributed points within the design space and standard hill-climbers which search from these points. A clustering algorithm applied to the hill-climber output can provide an indication of the number and distribution of local optima to the client. The clustering output is passed through a simple rule-set which determines which, if any, further optimisation technique should be applied to the problem. **The Optimiser** currently comprises three search and optimization algorithms: a genetic algorithm (GA), a simulated annealing algorithm (SA) and a tabu search algorithm (TS) [2]. The intention here is to provide a library of local and global search and optimization procedures that can be combined to achieve robust, high-performing SEO systems. **The Knowledge Repository** provides a storage capability for background information submitted by the clients and emerging information appertaining to problem characteristics and solutions from the Interrogator / Optimizer.

The initial distributed architecture is presented diagrammatically in figure 1. The Modeller would be the property of one centre of expertise whereas the Interrogator / Optimiser would be owned by another. Such centres may be geographically remote and several centres may offer similar services giving the client the option to select and link varying service providers using criteria relating to cost and reputation.

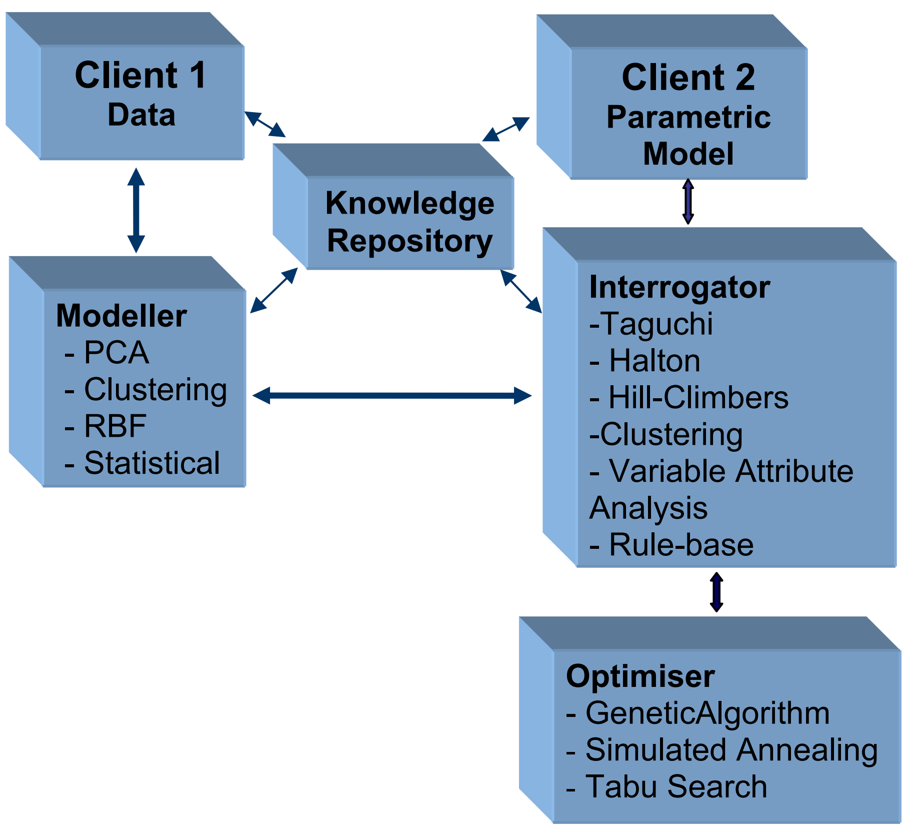



Figure 1. Initial Architecture

The following sections describe the various components in more detail and illustrate their capabilities through the presentation of results from simple test functions and an engineering application relating to the conceptual design of an autonomous undersea vehicle.

# 2. The Interrogator / Optimiser

To investigate the two possible configurations of the Interrogator (i.e. coupled either directly to a client's parametric model or to a client via the Modeller component) let us first assume that a client possessing a machine-based parametric model of a design problem wishes to access SEO processes. A data exchange capability between the client's web-enabled parametric model and the Interrogator is established. Currently the experimental system deals only with single objective, unconstrained problems therefore the client, via an on-line pro-forma, must define model characteristics relating to number and type of variables (real or integer), the upper and lower bounds of each variable range and the interval between each variable value.

## 2.1 Sampling the Design Space

The initial sampling of the design space is achieved either by introducing Taguchi [3] or Halton [4] techniques. Sampling serves the following two purposes:

**i) Sampling tests the robustness and the fidelity of the client's model by presenting the model with variable sets from diverse points across the design space**.

The Interrogator provides the client with relatively simple on-line guidance and related options such as:

"Firstly, we suggest that you run an inspection on your parametric model to verify that your model generates the correct values. The following options are available:

- Quick Inspection – this will provide a good spread of inspection points.
- In Depth – this will take a more detailed view via a greater number of inspection points but will take longer to process."

Established space sampling techniques are then introduced. The second option relates to the introduction of Halton sequences which provide a far denser sampling of the space than the first option, Taguchi, but would be more expensive both computationally and financially. The client's confidence in the fidelity and robustness of the model would therefore play a role in the sampling choice. The Interrogator component then runs either Halton or Taguchi utilizing the information previously defined by the client relating to number of variables and variable resolution. The resulting sample solutions (i.e. variable sets) are then passed to the client's parametric model which returns the calculated objective value for each solution. A complete list of the solution variable sets and their objective values is then returned to the client. The client then has three options:

- If some or all of the calculated solution objective values are erroneous then abandon the process, review and modify the model and re-present for further sampling and testing.
- If all the objective values appear sensible and one or more prove to satisfy the client's requirements in terms of a sufficiently high-performance solution then accept these solutions and terminate the process.

- Otherwise continue with the SEO process with the objective of discovering better performing solutions.

**ii) Sampling provides diverse starting points for exploratory hill-climbing processes.**

If the client wishes to proceed then further choices are available. A Simplex hill climber can be instigated either from:

- the most fit solution point in the sampled set;
- the best 10% of the solution points in the sampled set;
- all the solution points in the sampled set.

Assuming the third option is chosen then a hill climber is instigated from each of the sample points and the variable set and best result of each hill-climber are then passed to the client. If any solution identified by the hill climbers satisfies the client's requirements then the process can be terminated. Indeed, if all the hill climbers converge to very similar solutions in terms of their variable and objective function values then the client is advised that it is probable that little further improvement is possible i.e. it is likely that the solution space is unimodal / monotonic and the optimal solution has been identified.

## 2.2 Clustering and Stochastic Search

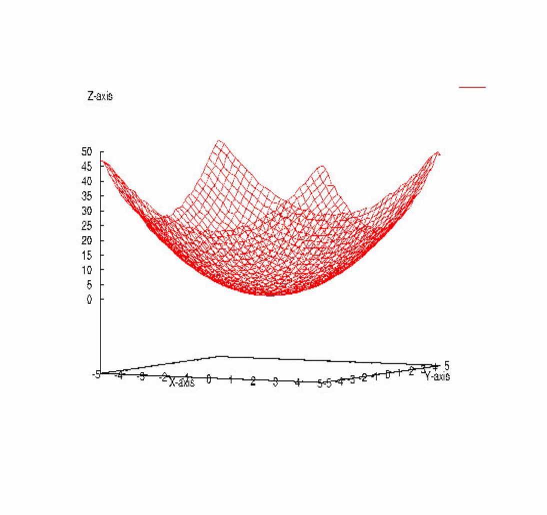
If the client wishes to continue search for possible better solutions then a near-neighbours clustering algorithm [5] is introduced to the set of 'best' hill-climber solutions. A series of Interrogator-based rules then determine which further optimizer is likely to be most appropriate. For instance:

- If just one cluster of solutions is identified but the Euclidean distance between each solution in the cluster is significant then either a simulated annealing or tabu search algorithm is initiated within the cluster region i.e. it is assumed that this is a high performance region containing a number of local optima.
- If more than one high performance cluster is identified and they are in diverse parts of the overall design space then it is considered likely that the overall search space is multi-modal i.e. many local optima may exist across the space. It is then assumed that a more global search process is required and a genetic algorithm is introduced.

In either case the best 10% of all solutions identified by the optimization are returned to the client.

## 2.3 Illustration of the Above Processes

The output from some of the above processes is illustrated in figures 2 and 3 using two standard test functions. This is followed by results from interrogation and optimization relating to the conceptual design of an autonomous undersea vehicle which provide greater detail of the processes. Given the overall complexity of what we are trying to achieve, i.e. the identification of the primary characteristics of multi-dimensional surfaces, it is very necessary to develop and test the various approaches in low-dimensional spaces so that potential problems can be identified. The scaleability of techniques that prove robust at low dimension can then be ascertained through application to more complex higher dimensional problems and subsequent analysis. Current work is addressing these scaleability issues through application to various design domains.



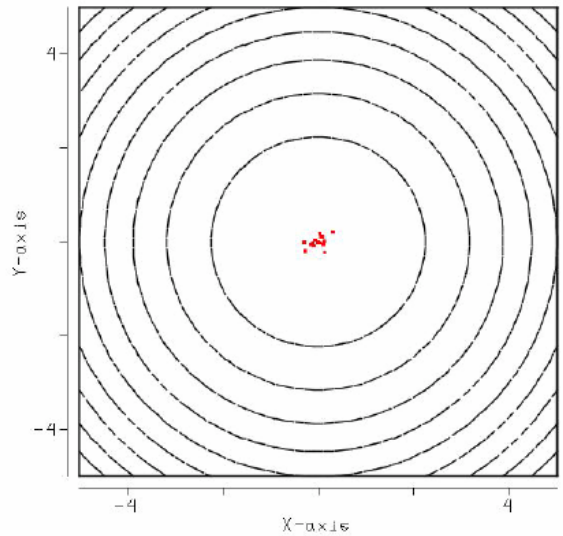
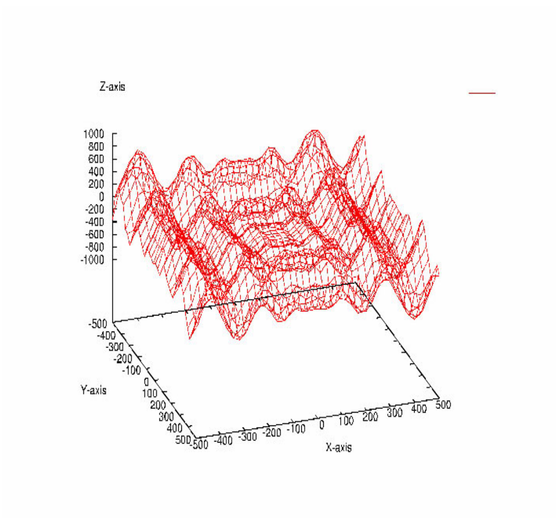




Figure 2a. Unimodal Test Function

2b. Tight cluster of best-performing hill climber solutions



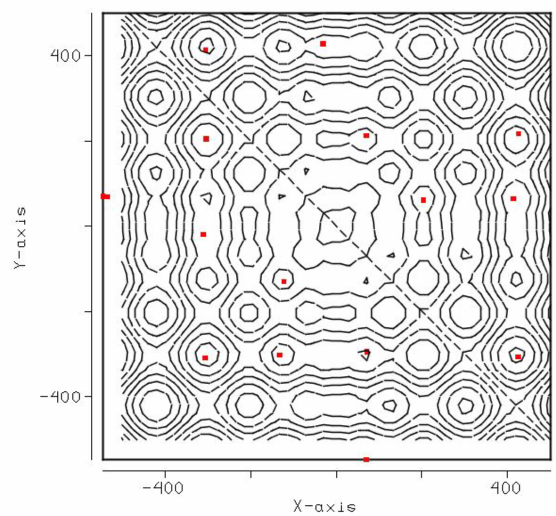




Figure 3a. Multimodal Schweffel function

3b. Best performing hill-climber solutions distributed across the design space

Relating to the test functions figure 2b shows the final output from simplex hill-climbers that have commenced from Halton-generated sample points. As would be expected in a unimodal (i.e. single peak / trough) space they have all converged to one tight cluster. As stated in the previous section, the low Euclidean distance between solutions in the cluster prompts an advisory message to the client suggesting that further improvement is unlikely. Figure 3b shows the final output of simplex hill-climbers again commencing from Halton sampling points well-distributed across the multi-modal landscape of the Schweffel function. As can be seen, the hillclimbers have ascended those peaks closest to their starting points. Subsequent clustering results in the identification of eleven clusters indicating high modality and leading to the introduction of a genetic algorithm which eventually converges upon the optimal solution.

## 2.4 Application to the Conceptual Design of a Subsea Vehicle

Systems Engineering & Assessment (SEA – http://www.sea.co.uk) is one of two main industrial collaborators providing either data or parametric design models for the testing and development of the Dipso System. The SEA parametric model represents the conceptual design of a remote-operated undersea vehicle. The model comprises eighteen variable parameters relating to vehicle dimensions, air-breathing, propulsion and power characteristics and structural and buoyancy considerations. The design objective is to identify a vehicle configuration that best satisfies a range of missions that the vehicle would be expected to perform. Currently, only single missions are being utilized although as multi-objective evolutionary approaches are introduced to the Dipso system, several missions will be concurrently processed. A simple combination of three criteria to determine solution performance is currently being utilized. These three criteria comprise a Mission Metric which determines to what degree the mission was successful; a Power Metric which determines to what degree a solution performs in terms of available power and a Design Metric which determines to what degree a solution is under or over designed. Each of these metrics varies from zero to one with an ideal vehicle scoring a total of 3.0. Mission and Power scores totaling over 2.0 indicate that the mission was successful and sufficient power was available. Design metric scores between 2.0 and 3.0 give an indication of under / over design.

Due to confidentiality issues it is not possible to define the mission utilized or to show the variable vectors for each solution in the following results.

**Step 1:** The design space defined by the eighteen variable parameters is sampled at 12 points using Halton sequences and the variable vectors relating to these points are passed to the client's model for evaluation. The evaluated solutions are shown in Table 1.

Table 1. Fitness of each sampled solution.

| | |
|---|---|
| 1. 0.070 | 7. 0.083 |
| 2. 0.069 | 8. 0.054 |
| 3. 0.069 | 9. 0.045 |
| 4. 0.049 | 10. 0.064 |
| 5. 0.130 | 11. 0.073 |
| 6. 0.071 | 12. 0.040 |

This output is returned to the client along with the variable vectors relating to each solution so that the integrity of the parametric model can be assessed. If the client considers the output to be sensible and it can therefore be assumed that the model is functioning correctly then the SEO process can be continued to ascertain whether better performance is possible.

**Step 2.** A Simplex hill climber is then initiated from each sample point.

After 30 Iterations of Hill Climbing from each point, the results are:

Table 2. Fitness of each final hill climbing solution

| | |
|---|---|
| 1. 0.337 | 7. 2.078 |
| 2. 0.866 | 8. 2.043 |
| 3. 1.187 | 9. 2.061 |
| 4. 0.278 | 10. 2.023 |
| 5. 1.625 | 11. 2.086 |
| 6. 0.964 | 12. 1.809 |

These results plus the associated variable vectors are returned to the client for inspection which reveals a diverse set of solutions both in terms of their relative fitness and in terms of the parameter values of the variable vectors. The client may therefore accept one of the better solutions (ie solutions 7 to11) or, as some diversity is apparent, decide to continue. Assuming that the client wishes to continue:

**Step 3:** Run Clustering Algorithm on hill climber results ie cluster the solutions in terms of similarity (Euclidean distance) between variable vectors.

The clustering results show five clusters in diverse parts of the design space which suggests that the overall space is multi-modal and high-performance solutions exist in diverse areas. This leads to the Interrogator's rule base selecting a final **genetic algorithm** search.

**Step 4:** Optimise with a GA

The initial population (100 individuals) of the GA is seeded with the hill climber results. After 200 generations the GA identifies an optimal solution with an overall fitness value of **2.999**. Both mission and power metrics have been fully satisfied i.e. they both achieved a value of 1.0.

The best 10% of the solutions from the final GA generation are passed to the client.

It is now intended to utilize the tripartite fitness function of the SEA model to investigate and implement evolutionary constraint handling techniques (i.e. various penalty functions) and evolutionary multi-objective (EMO) approaches within the Optimizer. EMO approaches will also be tested via the concurrent introduction of several missions.

# 3. The Modeller

The second configuration of the system relates to the linking of the Modeller and the Interrogator / Optimiser to a client who wishes to develop a data model and then identify high-performance solutions from this model utilizing the Interrogator / Optimiser capabilities. The Modeller component comprises a number of data pre-processing techniques including Principal Component Analysis (PCA) and Near-neighbours Clustering. Currently underlying these pre-processes lies a Radial Basis Function (RBF) Neural Network [6]and both Partial Least Squares and Principal Component Regression techniques [7].

A client can download a dataset to the Modeller which will pre-process the data whilst informing the client of the outcome of such processing. The client, once satisfied that the processed data is a satisfactory representation, then informs the Modeller of an acceptable level of error in terms of the required model. The appropriate modelling technique is then selected on the basis of the data type, density and distribution. In terms of the RBF, a subset of the data is used to train the network until the degree of fit to the remaining data does not exceed the error criteria. A graphical representation of the improving degree of fit can also be made available to the client to give an indication of the manner in which the modelling of the data is progressing (figure 4c). The client is then informed that the model is now ready to be subjected to Interrogation / Optimisation. If the client is happy for the process to proceed the Modeller establishes a connection with the Interrogator and a data transfer capability is established that is very similar to that previously described between the Interrogator and a client possessing a parametric model.

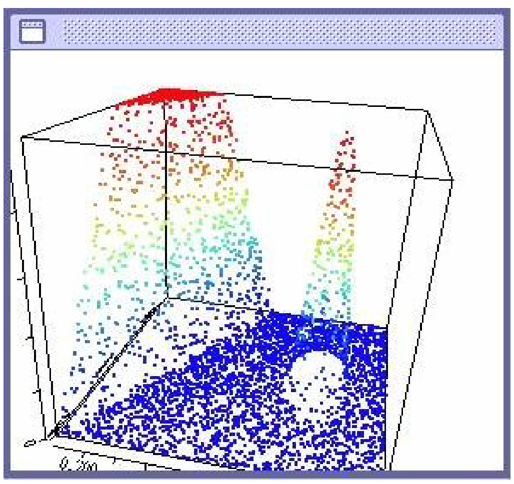
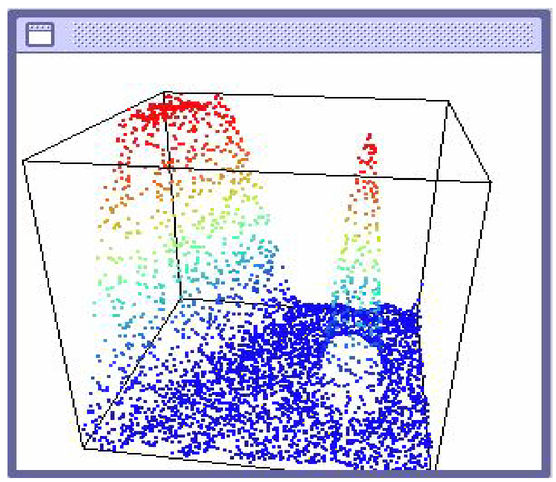
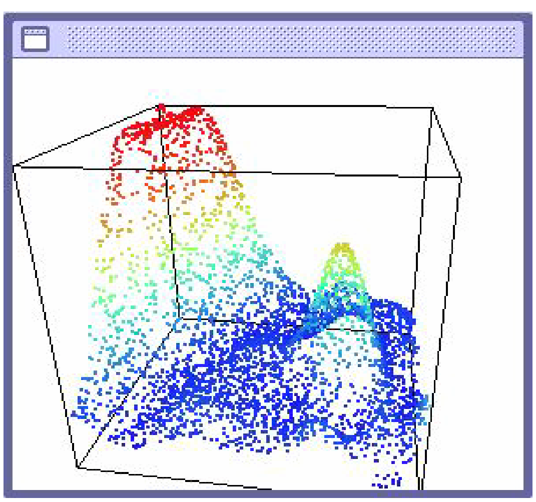


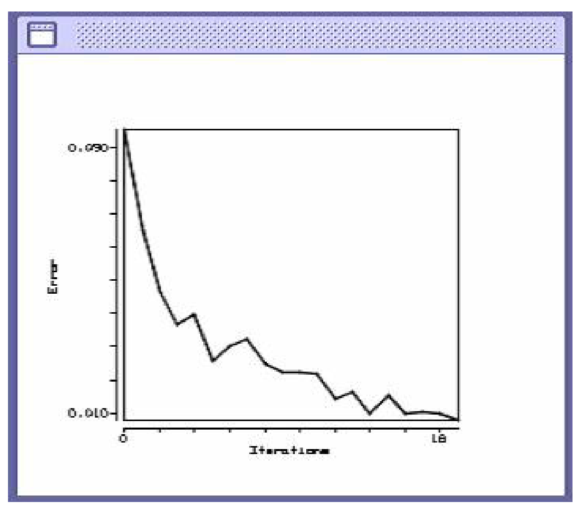
Figure 4(a). The test function training data



b) RBF surface with max error of 0.009



c) RBF surface with max error of 0.035




d) Error graphic

Figure 4. Example output from the Modelling component

The Interrogation / Optimisation process then follows the steps identified in the previous section i.e. the Interrogator now considers the RBF or statistical representation to be a parametric model. However, the model testing procedures (i.e. robustness and fidelity) are not relevant when dealing with these representations. Figure 4 shows the Modeller's RBF generation capability when applied to a standard statistical modeling test function [8] with differing levels of allowable error.

# 4. Current and future work

As previously stated the paper has concentrated upon the development of the system components and the above results have been generated from a stand-alone demonstrator sited within the ACDDM Lab at UWE, Bristol. The demonstrator has provided a deal of insight relating to the manner in which we can present high-quality information back to the client at various stages of the modelling and interrogation procedures. Further sampling techniques that will provide information relating directly to the nature of the problem space represented by the parametric model are currently being integrated. It is intended that such information will be accompanied by further recommendations as to what course of action the client may now wish to take. The Interrogator rule-base which determines choice of stochastic optimizer is being refined based upon problems identified from the demonstrator and this refinement will continue as search space sampling techniques provide more definitive information.

Further techniques for extracting and displaying high-quality information will be essential as we move from single objective problems to problems where several objectives that are conflicting to a varying degree are involved. Standard evolutionary multi-objective approaches such as Pareto will be included in the Optimizer and, if time permits, a simple Cluster-oriented GA will be introduced as a further sampling technique. This will support multi-objective data visualization as presented in Abraham and Parmee [9] and Parmee and Abraham [10]. The inclusion of various penalty function approaches to evolutionary optimization will also be required as constrained design problems are introduced. The objective of the current DTI funded project is to provide a proof-of-concept and it is expected that we will be able to achieve this in terms of relatively basic single, multi-objective and constrained design problems within the project time-scale.

The Modeller, the Interrogator / Optimiser and the SEA model have now been web-enabled at SEA; UWE, Bristol and at WESC, Cardiff University and the first distributed trials of the system are now underway. These trials involve two industrial collaborators, SEA and Evotec OAI who are enabling access to parametric models and data respectively. Results from these initial trials are fully supporting those from the stand-alone system and will be presented at the conference.

We are also discussing to what extent the client may wish to establish communication with the Modeller and the Interrogator. The following scenarios are envisaged:

1. A 'data' client has confidentiality agreements with both the Interrogator and the Modeller services. In this case the client's data can be passed to the Modeller and the generated model can be passed to the Interrogator. Both the Modeller and the Interrogator report back to the client. In this scenario a minimum of data passing is required as the RBF representation will reside in the same space as either of the two service providers.

2. A'data' client has a confidentiality agreement with the Modeller service only. In this case the client's data can be passed to the Modeller but the generated model must remain in the Modeller's space. In this scenario data must be passed between the Interrogator and the model residing in the Modeller's space. Information generated from the sampling, clustering and optimization can be passed from the Interrogator back to the client. via the Modeller or directly to the client.
3. A'data' client has no confidentiality agreement with either of the services. In this case the Modeller uploads an instance of its service to the client and the model is generated within the client's space. Once this is achieved the instance self-destructs leaving the generated model in the client's space. Data must then be passed between the Interrogator and the model residing in the client's space. Information generated from both services are passed directly to the client.

Similar scenarios would exist with the 'parametric model' where the model may remain with the client or be passed to the Interrogator. Once the web enablement is complete these options can be further investigated.

# 5. Conclusions

As previously stated the objective of this current two year project is to provide proof-of-concept relating to the establishment of a distributed problem solving environment involving a coupled data modeling component and an optimization component. An initial architecture involving both of these services has been developed and results so far support the feasibility and potential utility of this distributed problem solving approach. A stand-alone demonstrator has provided an indication of further system requirements during the initial development stage. Such requirements include:

- Further development of the Interrogator sampling and hill climbing techniques to minimize computational expense.
- Refinement and significant expansion of the Interrogator rule sets which define the appropriate choice of design search and exploration algorithms.
- Better definition and improved utility of the data pre-processors within the Modeller.
- Increased efficiency of the RBF network and statistical modeling techniques.

All of these areas are being constantly improved as whole system performance becomes more evident through application to increasingly complex problems.

The entire system is now web-enabled and we are achieving similar results from the Grid-based system. We do expect initial problems relating to data transfer and translation costs which will be addressed as they arise. The development of appropriate user-friendly interfaces to the two components is essential and these will be implemented shortly and tested by the industrial collaborators (our 'test clients'). Results from this implementation and from the constantly improving Interrogator/ Optimiser and the Modeller components will be presented at the conference.

## Acknowledgements

The research and development described in the paper is supported by the UK DTI and EPSRC. The authors' wish to thank these organizations for their continuing support.

**References**

1. Rana O. F., Pouchard L. (2003) Agent-based Semantic Grids: Research Issues and Challenges. *Journal of Parallel and Distributed Computing Practices.*

2. Parmee I. C. (2001) "*Evolutionary and Adaptive Computing in Engineering Design"*. Springer Verlag, London.

3. Taguchi G. (1987) *Systems of Experimental Design.* Kraus International Publications.

4. Sobol, I. M. (1967) On the Distribution of Points in a Cube and Approximate Evaluation of Integrals. *Computational Mathematics and Mathematical Physics*, **7**, (4), pp 784-802.

5. Jarvis R. A., Patrick E. A. (1973) Clustering using a Similarity Measure Based on Shared Near-neighbours. *IEEE Transactions on Computers; 22[11].*

6. Haykin S. (1999) *Neural Networks: A Comprehensive Foundation.* Second Edition, Macmillan, New York.

7. Jackson J. E., (1991) *A User's Guide to Principal Components.* John Wiley and Sons, New York.

8. Lancaster P., Salkauskas K. (1986) *"Curve and Surface Fitting".* Academic Press

9. Abraham J. R., Parmee I. C. (2004) "User-centric Evolutionary Design Systems – the Visualisation of Emerging Multi-Objective Design Information". $X^{th}$ International Conference on Computing in Civil and Building Engineering, Weimar, Germany.

10. Parmee I. C, Abraham J. R. (2004) Supporting Implicit Learning via the Visualisation of COGA Multi-objective Data. IEEE Congress on Evolutionary Computation, Portland, USA; pp 395 – 402.

I.C. Parmee
ACDDM Lab, CEMS, University of the West of England, Bristol, BS16 1QY, UK
ian.parmee@uwe.ac.uk
Phone: ++44 (0)117 328 3137
Fax: ++44 (0) 117 328 2587
http://www.ad-comtech.co.uk/ACDDM_Group.htm