The MUCM case studies: current status and proposals

MUCM INTERNAL REPORT

John Paul Gosling

31st January, 2008

Abstract

This internal report serves to explain the current position with regards to models and potential case studies for the MUCM project. It also provides some discussion points on the direction of MUCM's three major case studies, which are a main part of the project's outputs. This is the second version of this report: it has been revised in the light of the team meeting in November and the comments made over December and January.

Contents

List of Figures

1 What is a MUCM case study?

The purpose of the case studies is to provide exemplars of how the MUCM methods can be applied in non-trivial problems. They will illustrate combinations of techniques and will address serious practical problems (although not necessarily offering complete solutions). Three case studies will be completed and published. The case studies will not be toy problems: in each of them, we will address questions of current and pressing concern, guided by experts on our Advisory Panel. The production of the case studies will necessarily interact with all the other aspects of the project by providing test beds for specific techniques and pulling these together into major works that are of interest to the wider scientific community. In order to achieve this, each case study will be driven by a research question that is formulated by the relevant model owner.

A key aim of the MUCM project is to develop a basic technology that can be utilised wherever complex computer-intensive models are employed. Models are used throughout science, whether in academia or industry, to investigate all kinds of phenomenon. It is therefore important for the case studies to cover as many application areas as possible. Broadly speaking, the MUCM team has strongest links to the following three application areas: biology, environment and engineering. Although many of the challenges in terms of handling complex models are similar in each of those areas, it is important that we demonstrate our methodology has application across as much of science as possible.

Whilst the primary objective is to produce three case studies, the specific outputs of the case study workpackage will be varied and will certainly not be just three items. It is envisaged that each case study should be produced as a single, quite large document, and these will probably be made available on the MUCM website. A brief overview of the structure of the case study documents is given in Figure 1. We may find that the three finished documents might be worth publishing as a book. This could be coupled to the output of the toolkit

Example case study contents

1. Overview of the application area and motivation for the case study.
2. An extensive description of the computer model including challenges and past analyses.
3. An overview of MUCM's analysis.
4. Fitting an emulator including a full description of the choices we have to make when using emulator technology:
   (a) experimental design,
   (b) the main case study method,
   (c) validation of the emulator.
5. The results and reality.
6. Scientific conclusions and future directions.

Figure 1: Example case study contents.

workpackage to provide an extensive overview of dealing with uncertainty in complex models. In addition, there will be a number of journal articles that will be published in both the statistics and application area literatures: inevitably, we will face challenges in the production of the case studies that will lead to methodological advances, and we must show the practical value of our methods to non-statisticians.

A challenge that we have when planning the case studies is getting the timing right: we cannot demonstrate methods if they have not been fully developed. Therefore, it is essential that each workpackage has realistic milestones that will be reached at least around the time they have been specified. With a set of well defined methodological advances and their respective timings, we can confidently find an application that demonstrates our advances at that particular time. However, we will not be simply choosing an application area for which we know our current technology will solve all its problems. We want applications that will push our technology to new extremes and will uncover new challenges.

In § 2, we will review the techniques being developed in MUCM and when they are expected to be ready. § 3 details the computer models that we have access to so far and the challenges each model poses. In § 4, the management of the case studies is discussed along with an explanation of how everybody fits into the case study workpackage. Finally, in § 5 to 7, the proposals for the three case studies are given. Throughout this document, we will be referring to workpackages and their numbers; these are listed for reference in the appendix at the end of this report.

2 Methodology and case studies

As stated in § 1, the case studies will illustrate combinations of techniques developed across the MUCM group. Hence, it is crucial that the case studies researcher has knowledge about what methodologies are being developed by whom and when they are going to be ready for application. As part of our planning, each researcher is required to create milestones for the development of their particular methodological area. Figure 2 shows the milestones set down for each researcher and when they are expected to be reached.

The short-term milestones within MUCM are associated with investigating current best practice for dealing with uncertainty in computer modelling and laying the foundations for more sophisticated methodology. These broad reviews can be seen in milestone 1.1.1 for screening methods and 3.1.3 for experimental design amongst others in Figure 2. The first case study will therefore be based on current best practice and the recent advances the Aston group have made in handling high dimension models (through dimension reduction and sparse Gaussian process approximations). Computer models typically have very many

#	Milestone	Month of completion
2.2.1	Simple suite of diagnostics for emulator validation	May 2007
2.3.1	Dynamic emulation	June 2007
1.1.1	Review of traditional screening methods	August 2007
1.3.2	Simple multiscale emulation	August 2007
1.1.2	Screening methods in emulation	September 2007
1.2.2	Traditional dimension reduction techniques	October 2007
3.1.3	Library of designs	November 2007
3.2.6	Review of emulator use in climate	December 2007
1.1.3	Screening for stochastic models	January 2008
1.3.3	Multilevel emulation	January 2008
2.2.3	Methods for validating emulators	January 2008
3.2b.2	Using derivative information	January 2008
3.2.8	Correlation function effect	February 2008
1.2.3	Sparse sequential dimension reduction	February 2008
1.3.5	Spatio-temporal properties in emulation	February 2008
2.1.3	Exchangeable models applied to crop and climate	February 2008
3.1.4	Review of low discrepancy sequences in design	February 2008
2.3.2	UA and SA for dynamic emulators	March 2008
3.1.5	Constraints on ASCM and space filling designs	March 2008
2.3.3	Choices for model discrepancy functions	May 2008
1.1.4	Coupled emulation via screening	June 2008
3.2.9	Value of information analysis	July 2008
1.2.4	Dimension reduction in dynamic emulation	August 2008
3.1.8	Sequential experimental design for Haar kriging	September 2008
3.2.3	Input distribution effect	September 2008
2.2.4	Validating GP model discrepancy	October 2008
1.1.5	Sequential screening methods for sparse GPs	December 2008
1.2.5	Modular emulation	February 2009
3.2b.5	Review of Gaussian process derivative theory	March 2009
2.2.5	Validation of emulator and discrepancy	May 2009
3.1.13	ASCM for design	May 2009
1.3.7	Multilevel technology complete	June 2009
1.2.6	Modular emulators with dim red	August 2009
2.2.6	Model failure targeting via diagnostics	October 2009
2.3.4	Calibration for dynamic models and DA	December 2009
1.3.8	Software complete for high dimensional data	April 2010
1.3.8	Optimised resolution in multiscaled models	May 2010
2.1.7	Multiple simulators and links to reality	May 2010

Figure 2: A list of milestones in order of their (expected) completion date

inputs and produce many outputs. Both inputs and outputs can comprise values over large spatial domains. Outputs are often both space- and time-varying. The Gaussian process methodology becomes itself more computer-intensive as the dimensionality of the input space increases; hence, there is a need to develop efficient methods of dimension reduction or ways to exploit simple local structure within globally complex models.

The second case study will be based around developments in the emulation of dynamic computer models to coincide with milestones reached during the summer of 2008. Many computer models are dynamic: they model a system that is evolving over time and they operate iteratively over fixed time steps. The computer model requires the current value of a state vector as an input, and the updated value of the state vector becomes an output. It may have other inputs that in the context of a many-step run of the simulator can be classified as model-parameters and forcing inputs. Model-parameters have fixed values for all the time steps of a simulator run; they describe either fundamental parameters of the mathematical model or enduring characteristics of the specific system being simulated by that run. Forcing inputs vary from one time step to the next and represent external influences on the system. This type of computer model is common in many areas, and an efficient approach to handling the uncertainty in those models would be widely applicable. Although there is an obvious temporal structure in these types of models, we often just fix time and consider model output after a certain number of steps.

As time passes, more sophisticated methods will be available for the case studies, and, to link in with our expected achievements during 2010, the third case study should be dealing with multiscale models and have a genuine appreciation of the model's link to reality. As this work is being developed in Durham, we expect that this case study will be analysed using Bayes linear methods (an extensive overview is given in Goldstein and Wooff, 2007). By using Bayes linear methods, we may be able to handle more complicated problems than if we used a probability-based specification. There is also scope for making our methodology more efficient.

The original research on computer models considered experimental design as an integral part of the area (see Sacks et al., 1989). Each case study will have a substantial portion dedicated to experimental design. It is envisaged that the first two case studies will involve the careful selection of currently available designs; the review of experimental designs in milestone 3.1.3 will become important when we are setting up each case study. In time for the third case study, we should have developed an Adaptive Sampler for Complex Models (ASCM) that will provide an novel and efficient way to optimally design a computer experiment.

Another strand that is important for each case study is the validation of the emulators (workpackage 2.2). Emulators are often built and used without a check that they are actually close to matching the model; more often, their link to reality is completely ignored. Workpackage 2.2 will first produce methods for criticising and validating emulators as predictors of the underlying simulator. Later, diagnostics will be created to validate simulator performance against observed model runs and observed real-life data. Currently, through milestone 2.2.1, we have access to novel methods for validating an emulator's predictive performance. This will be used in the first case study, and validation will remain an integral part of the final two case studies.

As we have already stated, the case studies will encapsulate the current knowledge of the group at the time of their compilation. This means that the case studies researcher must have a solid understanding of all the workpackages. It also means that, although we will be presenting screening and dimension reduction techniques for the first time in the first case study, we will be using the same techniques on the subsequent case studies. Techniques that have been developed and presented in earlier case studies will be carried forward to make each case study a reflection of our cutting-edge technology at that particular time. In addition to this, there may be scope for us to revisit earlier case studies (while completing case study three or during the 'writing-up' period) in order to update them will relevant methodology advances.

3 Models available to MUCM

Within MUCM, the researchers have access to a number of computer models. Many of these are simply toy models that are constructed to test out approximation and emulation methods or have become technologically outdated within their own application area. We also have access to models that are on the cutting edge of scientific research that are both computationally expensive (and perhaps outside the reach of MUCM's budget) and have few or no real data associated with them. The availability of data becomes important if we are looking to illustrate calibration or data assimilation techniques. In this section, we will review the models that have been investigated so far within the MUCM group.

One toy example that has been used by the group to sharpen our understanding of emulation techniques is the Daisyworld model, which is described in Watson and Lovelock (1983). Daisyworld models a zero-dimensional planet of unit area that is illuminated by a sun and supports only two simple forms of vegetation: black daisies and white daisies. Its outputs are parameters that define the proportion of each flower and current climatic conditions on the planet. Although developed to test the hypothesis that the earth is a self-regulating system in which the individual members of the biosphere can exert collective control over the system itself, the model is computationally inexpensive and has no grounding in reality; hence, it is treated as a toy model used solely for testing purposes. Another model that has similar function, is the time-stepping advective/diffusive surface layer meridional energy balance model (surfEBM) that is a simple planetary climate model that outputs temperature at a given number of latitude bands. It is of use to MUCM researchers because it is simple to run through MatLab code and we can increase the number of inputs and outputs to whatever we like by adding extra latitude bands.

Through our involvement at SAMSI's year-long program on "Development, Assessment and Utilization of Complex Computer Models", we were introduced to two more models that have been used in the past to test statistical techniques. We have access to a rainfall-runoff model, which is described in Kuczera et al. (2006), that has a simple dynamic structure and an extensive data set with it. The model itself is simple as it has only three compartments and the flow of water between these compartments is governed by a small set of differential equations. The model's output is river discharge and this is affected by evapotranspiration potential and rainfall in the river's catchment. The data we have alongside this model consists of daily values of rainfall, evapotranspiration potential and discharge in the Abercrombie river catchment in Australia, over about 700 days. This makes this model a candidate for testing dynamic emulation and data assimilation techniques (workpackage 2.3).

Also through SAMSI, Prof. David Steinberg introduced us to a residual radioactivity (RESRAD) model that was developed by the US government (details are available online at `http://web.ead.anl.gov/resrad/home2/resrad.cfm`). The computer model simulates the effect of storing nuclear waste in different ways on drinking water. There are 27 inputs into the model that outputs a measure of increased radioactivity of drinking water in the surrounding area after a period of time. For many input configurations, the model returns an output of zero that means the storage method is working as it should. To emulate this type of model well, we may need to break the input space up in regions where there is positive and zero output. Many models have areas of their input space where the output is fixed value or the code fails to return a value. We may need to develop technology within MUCM to deal with this and this model would provide a suitable test.

We can place the more substantial models into the three application areas that we gave in § 1: biology, environment and engineering. For biology models, we have a metabolic reaction model and a rotavirus model both of which provide substantial challenges to current emulator technology. The metabolic reaction model is based on a compartmental representation of the glycolysis pathway, which is described in Pritchard and Kell (2002). The model is available using the Systems Biology Markup Language (SMBL); tools for implementing the model are available from `http://www.SBML.org` and the language is reviewed in Hucka et al. (2003). The model itself is computationally inexpensive: it takes seconds to run; the challenges come with the high dimensional input and output space due to large number of compartments and interactions between the compartments. This model is a prime candidate for developing both screening and dimension reduction techniques. The rotavirus model that will be supplied to us through GSK has many similar features and is reviewed later in § 5.

We have access to two models that can be classed as environmental models apart from the rainfall-runoff model that was described earlier. The first is a dynamic vegetation model designed to calculate the carbon fluxes and pools in the biosphere and, specifically, their variations under changing climate. The Sheffield Dynamic Global Vegetation Model (SDGVM) is described in Woodward et al. (1995) and Woodward and Lomas (2004). The model represents the growth and decay of generic plant functional types (PFTs), rather than individual species, in discrete grid-cells that together can cover large areas up to the whole land surface. SDGVM has already been used with standard emulation techniques in Kennedy et al. (2007), and more sophisticated techniques are being developed for the model through the NERC Centre of Excellence in Terrestrial Carbon Dynamics that is based at the University of Sheffield.

On a desktop PC, SDGVM will simulate 10 years per second. One usual run of the model will involve a 500 year spin-up run where the carbon pools stabilise and 100 years of a proper run using real climate data. Hence, one full run will take about a minute (this computation time depends on the type of vegetation being simulated). The model produces 28 outputs that can be viewed as a time series on different temporal scales: daily, monthly or weekly. This kind of model could provide a test for our dynamic emulation methods; however, it would not be possible to have real data for each output in order to test a data assimilation scheme.

The second environmental model that we have access to is a runoff and chemical tracer model that is similar in principle to the rainfall-runoff model described earlier. The model is a three compartment model with parallel transfer; the three compartment part refers to the modelling of three locations in which the rainwater can be stored effectively, and parallel transfer describes two mechanisms by which water may discharge. The mechanisms are characterised by speed, either fast or slow, and each operate in parallel to contribute to the final estimated water discharge. Iorgulescu et al. (2005) gives a detailed description of the model. The model is fast to run through R, and Monte Carlo analyses could easily be carried out on it. However, the dynamic structure of the model and the availability of real data make this a perfect test model for the techniques developed in workpackage 2.3.

In addition to these two environmental models, we also have access to a set of climate models through our links to the Met Office and the NOC. These models are considered in § 6.

We have several models that fall under the broad heading of engineering models. One model that is similar to the RESRAD model is a model made available to us through Prof. Martin Dove. The model simulates the containment of radioactive material and is available in several forms ranging from an approximation of the associated force equations to an extremely complex model that computes quantum mechanics equations. The models are described in Todorov et al. (2006a) and Todorov et al. (2006b). The model based on quantum mechanics can take one day to run on a 1000 processor supercomputer, which costs £1000 per hour to run. The most expensive model is definitely outside the budget of MUCM. However, we could use multiscale emulation techniques to help improve the selection of best input parameters as there is currently little appreciation of uncertainty in this field at the moment.

Rolls Royce have provided us with a model of a 24-blade fan assembly. The model has 24 inputs that relate to the resonant frequency for each blade and one input that controls damping. An ensemble of 100 model runs has been available to MUCM since the start of the project. It is clear that conventional emulation techniques are not appropriate for this type of model: emulator validation techniques that have been developed as part of workpackage 2.2 have shown how poor the fitted emulator is at capturing the model's actual behaviour. To help us understand the behaviour of the model, we have now got access to 4- and 8-blade versions of the model. All the models can be run locally at the LSE as they have the necessary software to run the model. We envisage that this model will provide a challenging test for our more sophisticated screening and multiscale methods.

The engineering models that we have access to in MUCM tend to have multiple degrees of complexity and we also have access to real data that could be used in our analyses. A hydrocarbon reservoir model is described in § 7 that can be set-up for various spatial resolutions and there are data available to help us create a link to reality.

Through our association with the Central Science Laboratory (CSL), we have access to two more types of model. One is a model that assesses acute risks to various species from chemicals; this model can be run online as it is part of the WEBFRAM development (see `http://www.webfram.com/`) that is described in Central Science Laboratory (2007). The models are essentially instantaneous to run; however, they might still be of use to MUCM as real data is available for certain species and chemicals. For instance, we have been introduced to a model that assesses acute risks to skylarks for pesticides used on UK cereals during the summer.

The second model that CSL have provided us with is a raccoon-fox-dog model that simulates the population dynamics between species. This model has many challenging features: it has about 140 inputs, it has a dynamic temporal structure and its outputs are stochastic. Stochastic models have not yet been discussed in this report as all the models so far have been deterministic. Stochastic models present us with a new source of uncertainty that makes accurate emulation more difficult. We have access to a substantial stochastic model through our researchers in Durham. This is a cosmology model that is being analysed in a separate project by Dr Ian Vernon at Durham and as part of workpackage 2.1. Also, in Sheffield, we will soon have access to a diabetes model that has a stochastic nature; details of this model will be produced on the wiki in the coming months.

As the project develops, we will come across more computer models that we can use to test our new methods on. In terms of the case studies, we must be flexible in that a perfect application for our methods that has not yet been introduced to the group may appear. The MUCM Wiki has an important role to play in informing the whole group about the models that we have available to us and the challenges associated with each one of those models. It is therefore vital that all models are reviewed on the Wiki even if the researcher who first had access to the model believed it is of no use for their techniques. That way we can make an informed decision when choosing applications to demonstrate our techniques.

4 Organisation of the case study workload

The key to the case studies achieving all the aims we have set out thus far is to have everybody involved in the production of the case studies. Within the MUCM investigators, postdoctoral researchers and PhD students, there is a lot of experience of using emulator technology across a wide range of application areas. It is also essential that the relevant Advisory Panel members are fully on board with the development so that we can turn an academic exercise into a case study that is of interest to the wider scientific community.

For each of the three case studies, there is a simple way of creating a core group to oversee development. Every postdoctoral researcher is linked to several models either through their advisory panel contact or through other contacts at their institution. The researcher that is responsible for the model will be the primary contact for the case studies researcher; this could be for explanation of the model or for actual model runs. Also, through the work packages each researcher has a strand of methodological development they are responsible for. As described in § 2, each case study will be based on one strand of methodological development. The researcher who is responsible for the particular area of research that is being showcased in the case study will become the main guide for computation for the case studies researcher.

For each case study, the case studies researcher along with the postdoctoral researchers responsible for the main method and the model will form a core team that will oversee the development of the case study. This team of three researchers will meet on a regular basis to discuss how each strand of the case study will fit together. In addition to these three, it would be good to have a principal investigator (PI) involved in a consultancy role. The appropriate PI will be chosen based on their research interests and experience in the relevant application area. Communication between the group will be essential. Regular meetings and teleconferences will be scheduled; alongside these, the wiki and repository will help us to share thoughts and files.

As suggested earlier by its inclusion in Figure 1, experimental design will form a key part of each case study's analysis. Normally, when designing computer experiments, we take the easy option and use an off-the-shelf design tool like a maximin Latin hypercube (see Morris and Mitchell, 1995) or a design stemming from a Sobol sequence (an overview is given in Saltelli et al., 2000). When developing the case studies, we will recognise the importance of good experimental design, and, in order to achieve the best results we can as efficiently as possible, we will consult with MUCM researchers at the London School of Economics who are working on workpackage 3.1. Initially, for each case study, this consultation will consist of a focus day to be held between both the core team and the LSE researchers. Time will be spent familiarising this whole group with the application area and the methods that we are proposing to use. We can then use our pooled expertise to develop designs that are relevant for the results we are trying to achieve. This is why experimental design will form a chapter in each of the three case studies. It is envisaged that by the third case study a smart design algorithm will have been developed that will change the way we all design computer experiments.

The ultimate responsibility for the production of the case studies lies with the case studies researcher. Inevitably, the case studies workpackage interacts with all the other workpackages; hence, there is a great opportunity to stimulate group communication through the development of the case studies. Although, each case study has a central method that it is based around, methodological advances from other workpackages will be incorporated where possible. Consultation between the core team and other members of the group (this includes PhD students) is essential for the case studies to represent cutting edge methodology. To facilitate their production, the following programme for each case study is suggested:

1. initial identification of objectives for the case study,
2. set-up of core team and identification of plausible methods,
3. codification of case study actions with associated deadlines,
4. implementation of plan from previous step,
5. evaluation and publication of results.

Point 1 in this programme is discussed in § 5 to 7, and the objectives for the first case study will be defined before the end of February 2008. The programme will be released onto our wiki site for each case study along with deadlines and researcher's responsibilities. The dates for the points 1 and 2 will be set by the case study researcher. The dates for the subsequent points will be decided by the core team under the supervision of the case study PI; in this stage, a rigourous programme will be set for the case study. We envisage that each case study will have approximately eight months for points 1 to 4 and the fifth point will be completed when appropriate (overlap is expected and some publications may happen during the analysis stage of the case study).

The workload associated with being involved in the development and production of case studies will not interfere with a researcher's own workpackage responsibilities. In fact, they will complement them well by providing a substantial and challenging test for their techniques. The PhD researchers may find that the use of their methods on the case studies provides them with the application chapter for their theses. The responsibility for the running of models and our analyses will be decided on a case by case basis. It would be preferable for the case studies researcher to do most of this, but constraints on the models or availability of code will make that difficult. In terms of writing the final case study documents, the case study researcher will be in a position to write everything; we want to avoid a disjointed text that can come from having many contributors.

5 Case study I: biology

The first case study provides us with an opportunity to focus our current expertise on a substantial application area. We are in a position where we have recently developed methodology for screening inputs and reducing the output dimension of computer models taken from machine learning. However, the the methods being reviewed in workpackages 1.1 and 1.2 have yet to be tested on a serious application. It is clear that high dimensionality is common in computer modelling as model builders tend to increase the number of compartments (grid cells, pools or time-steps etc.) of their models until their computers struggle to process the associated equations and each compartment may have a number of inputs associated with it. High dimensional inputs and outputs also causes computational problems for our emulation techniques; the work of Conti and O'Hagan (2007) and Rougier (2007a) has moved us closer to a method that is computationally affordable. However, it may be that the employment of these methods may be for nothing: for instance, many of the inputs may not be influential on the outputs of interest or many of the outputs could be representable by a single number. Our experience with models is that only a handful of inputs are influential for the output of interest; an example of this is found in the analysis of SDGVM, which was mentioned in § 3, in Kennedy et al. (2007).

There has already been a lot of review work done on screening and dimension reduction within the MUCM group. There are technical reports available to the group that detail both traditional and newer (machine-learning-based) approaches for handling models with both high dimensional inputs and outputs. Dimension reduction alongside a sparse Gaussian process approximation will be the main methodological theme in the first case study. Workpackage 3.1 will play a strong role in producing designs that will aid screening and dimension reduction. In particular, a design that maintains its space-filling properties when projected into a lower dimensional subspace would be useful for our screening methodology. Also, the validation techniques of workpackage 2.2 will make up a substantial chapter in the first case study to show how we must check our emulator's performance when the results are being used in real applications.

We propose using a biology model that will be provided by GSK for the first case study. The GSK model captures current knowledge about natural history of rotavirus infection and its transmission in some population; this type of model is common in epidemiology and is of real medical importance. The model in question is a deterministic compartmental model with 672 compartments (16 disease stages x 42 age classes). Figure 3 shows a compartmental model for one

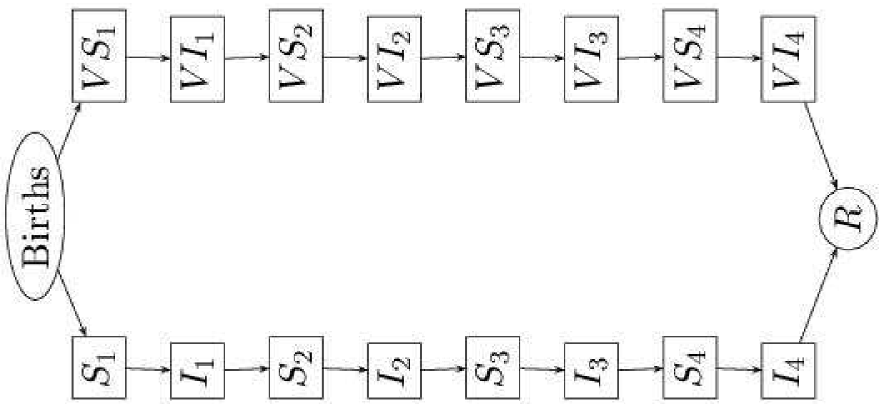

Figure 3: A rotavirus model

age class: disease states are represented by compartments and flows between states are represented by arrows. This is an extension of the classical SIR and SIS compartmental models (S denotes susceptible, I denotes infectious and R denotes recovered) that accounts for the progressive build-up of immunity against infection, disease and infectiousness (if infected) with increasing number of prior infections. The compartments on the lower row correspond to non-vaccinated states, while on the upper row are vaccinated states (V denotes vaccinated). This model is important to GSK as they have developed a vaccine for the rotavirus infection and rotavirus claims hundreds of thousands of lives every year.

One version of this is a dynamic state analysis of the compartmental model where, for every run of the model, the set of differential equations that governs the compartmental interactions is solved for a specified time horizon. The length of time it takes to run this model is dependent on the specified time horizon: when the horizon is short (10 years say), the model takes approximately three seconds to run, and, for a longer horizon (100-150 years say), each run takes over 30 seconds. This makes a Monte Carlo analysis infeasible for a 100 year time horizon.

Currently the GSK scientists are using a simpler version of the model with a cruder differential equation solver to make inferences about vaccine impact. A single run of the basic model takes approximately 3 seconds to complete. Also, they limit their sensitivity analysis to nine input variables and one output, and, using Monte Carlo techniques, they use 11,540 simulations to produce the sensitivity analysis results (this takes between 9 and 10 hours). We propose to use the computationally expensive version of the rotavirus model and all of its inputs and outputs in order to demonstrate the efficacy of our screening and dimension reduction. We should be able to produce results that are comparable to those already achieved in a fraction of the time whilst considering many more sources of uncertainty. This case study has the potential to have a major impact on the way these infection models are analysed and will help our collaborators at GSK to gain a greater understanding of their models.

6 Case study II: environment

The second case study coincides with the development of dynamic emulation as a tool for data assimilation. Therefore, a dynamic computer model (this is defined in § 2) that can be linked to real life data is desirable for this case study. Environmental models typically have a temporal aspect to them. Also, there are usually many real life observations available for them: for climate models, a wealth of weather observations are available.

Climate models are familiar to a number of the investigators within MUCM through personal research interests and through association with the Probability and Uncertainty in Climate Models project. Climate change is an important topic in science at the moment and it would be beneficial to MUCM if we could help advance the understanding of uncertainties associated with the models.

Climate models are typically computationally expensive, have a spatio-temporal structure, have many thousands of inputs and outputs and can be ran at different resolutions. The model's adjoint is often produced alongside the model so derivative information could also be available; this may allow us to use the technology being developed in workpackage 3.2b. All of these features make climate models interesting from a MUCM point of view.

Climate models also pose challenges quite unlike any model we have discussed so far in this report. A state of the art climate model can easily take months to complete one single run. The time taken for this type of model to be run makes standard emulation techniques expensive and Monte Carlo techniques almost impossible. Typically, analyses of these models are based on ensembles of model runs (see Rougier, 2007b; Rougier and Sexton, 2007). The input configurations for the ensemble are chosen by the institute who develop and run the model. Dependent on the output we are able to access, we may be able to employ the dynamic emulation techniques of workpackage 2.3 on a substantial climate model like HadCM3.

A faster type of climate model is available to us in GENIE-1, which is described in Marsh et al. (2002). Although this model is still relatively computationally expensive, we may be able to run the model locally at Southampton. We also know that there will be an adjoint version available for this model. As this model is quicker to run and we could have control over the input designs, this model is more attractive for the case study to employ techniques from workpackages 2.3 and 3.1. A thorough analysis of GENIE-1 would be of interest to the climate science community as it could help other scientists understand the workings of their climate models.

The format of this case study is still far from decided. It is clear that we have two routes that we could follow: we could use a small ensemble of runs to try to capture the behaviour of a massive climate model like HadCM3 or we could concentrate on doing a thorough analysis of a more accessible model like GENIE-1. Unfortunately, we will not be in a position to handle both at the time of compiling this particular case study as the technology we are developing to link different models will still be in its infancy. We should also keep in mind the dynamic emulation methodology and derivative theory that we wish to demonstrate in this case study.

7 Case study III: engineering

Our final case study will provide us with an opportunity to demonstrate the full scope of our advances in analysing complex computer models. By the time the third case study is being produced, we should have a greater understanding of how we can deal with the discrepancy between models of varying complexity and the gap between models and reality. Methodology built up in workpackages 1.3 and 2.1 will be the main focus of this case study. Therefore, the application in this case study needs a model that can be run at different levels of complexity and has some data that can be used to help quantify the model-to-reality discrepancy.

A hydrocarbon reservoir is a three-dimensional region of porous rock within which hydrocarbons (that is oil and gas) can be found. This region of porous rock is capped with a layer of impermeable rock which traps the hydrocarbons in place, thus creating the reservoir. The hydrocarbons are usually found with water in the reservoir. The complex geology of the reservoir and the location and operating conditions of any wells are two fundamental components in simulating reservoir performance. Hydrocarbon reservoir simulation is not a new subject and is described in Thomas (1982).

The model we have access to through Energy SciTech simulates a real hydrocarbon reservoir and is computationally expensive. Work has been done by the Durham group on reservoir models in the past with some success (see Craig et al., 1996, 1997, 2001). A diagrammatic representation of a similar reservoir model's grid is shown in Figure 4 where the arrows represent the wells which produce our outputs of interest. In MUCM, we would like to increase the number of inputs to the model by freeing more of the model parameters in each grid cell. Previous work on the reservoirs has used 62400 grid cells where some model-parameters are fixed at the same value for most of the cells and they have focused on a

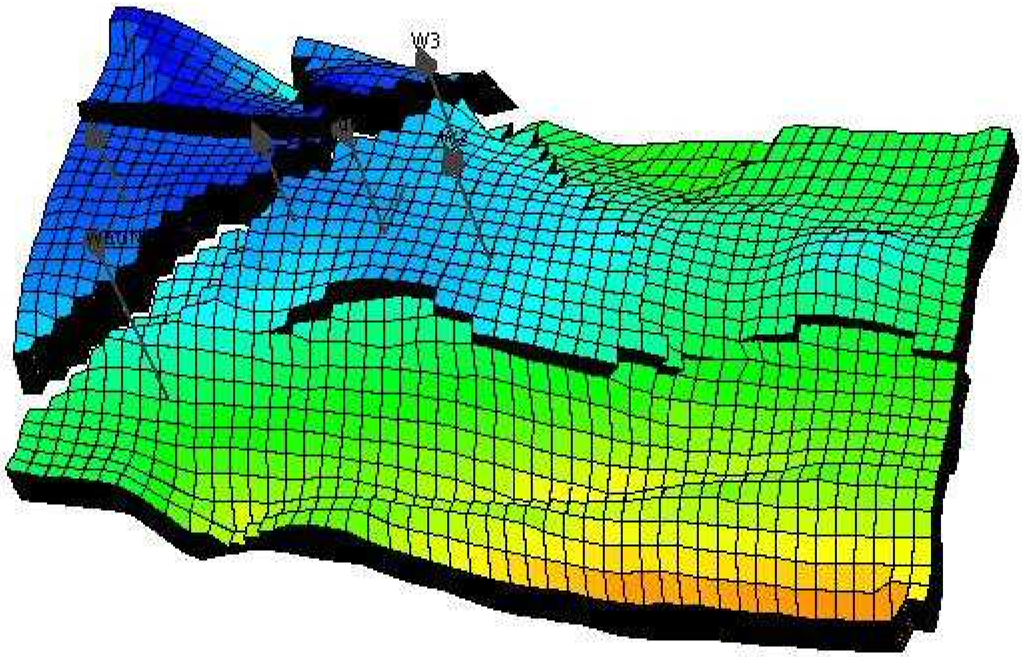

Figure 4: A toy hydrocarbon reservoir model: Daisy

small subset of outputs. Using our new technology for screening and dimension reduction, we will able to free up more inputs in our real reservoir model that has many more grid cells.

The actual number of grid cells the model has depends on the resolution that we choose to run the model at. The finer the resolution, the longer the model takes to run. As we have the model running locally at Durham, we can control the resolution of the model and all its inputs. This gives this model the desired flexibility to demonstrate our multilevel emulation techniques that will have developed through workpackage 1.3.

Also, for our real hydrocarbon reservoir, we may have access to data regarding the geology of the reservoir (to inform us about model inputs) and data on the model outputs. Hence, this model is perfect for us to employ the advances of workpackages 1.3 and 2.1. We can also use this case study to demonstrate our progress in terms of experimental design and validation, which, as stressed throughout, are important components of all the case studies. In particular, the ASCM suite of design algorithms (from workpackage 3.1) should be ready for use.

8 Discussion and timescale

This document sets out the proposed content of the case studies and how they will be organised. To supplement this, as part of the planning for each case study, a separate proposal document will be produced that discusses the models and techniques to be used in depth. In those documents, a detailed timetable for the project management of the case study, including the timing of its experiments and analyses, will be given.

The toolkit of workpackage 3.2 is to be utilised in each of the case studies. In principle, this may make the toolkit more accessible to scientists in each of the application areas. We envisage that the link between the case study and toolkit will be two-way: the case studies will not just be produced in line with the state of the art methodology documented in the toolkit, they will help to inform what is actually state of the art in terms of analysing complex computer models.

The case studies will stretch our techniques and enhance our understanding of the analysis of computer models. We must be realistic in our expectations as we cannot spend three years of research on getting nothing by setting impossible targets: the case studies are a major deliverable of the project. The three case studies currently have the deadlines for completion of August 2008, June 2009 and January 2010 respectively. We expect the experiments and analyses to have been completed by these dates and the writing up of the case study documents to overlap. In the final few months of the project, the case studies researcher will work to turn the three case studies into one major work that will take the form of a book. If the case studies have been successful in achieving their aims as set out in § 1, then the final document will encompass the work of the whole MUCM team and could have a major role in how computer models are analysed.

Appendix: the MUCM workpackages

Figure 5 is a table that lists the main methods being developed in the workpackages and the associated postdoctoral researcher or PhD student.

Workpackage no	Method strand being developed	Researcher
	High dimensionality	
1.1	Screening	Alexis Boukouvalas
1.2	Sparsity and projection	Dharmesh Maniyar
1.3	Multiscale models	Jonathan Cumming
	Using observational data	
2.1	Linking models to reality	Leanna House
2.2	Diagnostics and validation	Leo Bastos
2.3	Calibration and data assimilation	Richard Wilkinson
	Realising the potential	
3.1	Experimental design	Hugo Maruri-Aguilar
		Noha Youssef
3.2	Building the toolkit	Jim Gattiker
3.2b	Derivative information	Gemma Stephenson
3.3	Case studies	John Paul Gosling

Figure 5: Workpackages: numbering and methods.

References

Central Science Laboratory (2007). "CSL Science Review 2007." Technical report, Central Science Laboratory, York, UK.

Conti, S. and O'Hagan, A. (2007). "Bayesian emulation of complex multi-output and dynamic computer models." Technical report, *Research report 569/07*, Department of Probability and Statistics, University of Sheffield. Submitted to *Journal of Statistical Planning and Inference*.

Craig, P. S., Goldstein, M., Rougier, J. C., and Seheult, A. H. (2001). "Bayesian Forecasting for Complex Systems Using Computer Simulators." *JASA*, 96: 717–729.

Craig, P. S., Goldstein, M., Seheult, A. H., and Smith, J. A. (1996). "Bayes Linear Strategies for Matching Hydrocarbon Reservoir History." In *Bayesian Statistics 5*, 69–95.

— (1997). *Case Studies in Bayesian Statistics III*, chapter Pressure matching for hydrocarbon reservoirs: A case study in the use of Bayes Linear strategies for large computer experiments. Springer: New York.

Goldstein, M. and Wooff, D. (2007). *Bayes Linear Statistics, Theory and Methods*. Chichester: Wiley.

Hucka, M., Finney, A., Sauro, H. M., Bolouri, H., Doyle, J. C., Kitano, H., Arkin, A. P., Bornstein, B. J., Bray, D., Cornish-Bowden, A., Cuellar, A. A., Dronov, S., Gilles, E. D., Ginkel, M., Gor, V., Goryanin, I. I., Hedley, W. J., Hodgman, T. C., Hofmeyr, J.-H., Hunter, P. J., Juty, N. S., J. L. Kasberger, U. K. N. L. N., A. Kremling, Loew, L. M., Lucio, D., Mendes, P., Minch, E., Mjolsness, E. D., Nakayama, Y., Nelson, M. R., Nielsen, P. F., Sakurada, T., Schaff, J. C., Shapiro, B. E., Shimizu, T. S., Spence, H. D., Stelling, J., Takahashi, K., Tomita, M., Wagner, J., and Wang, J. (2003). "The Systems Biology Markup Language (SBML): A Medium for Representation and Exchange of Biochemical Network Models." *Bioinformatics*, 19: 524–531.

Iorgulescu, I., Beven, K., and Musy, A. (2005). "Data-based modelling of runoff and chemical tracer concentrations in the Haute-Mentue research catchment (Switzerland)." *Hydrological Processes*, 19: 2557–2573.

Kennedy, M., Anderson, C., O'Hagan, A., Lomas, M., Woodward, F., Heinemeyer, A., and Gosling, J. (2007). "Quantifying uncertainty in the biospheric carbon flux for England and Wales." To appear in *J. R. Statist. Soc. Ser. A*.

Kuczera, G., Kavetski, D., Franks, S., and Thyer, M. (2006). "Towards a Bayesian total error analysis of conceptual rainfall-runoff models: characterising model error using storm-dependent parameters." *Journal of Hydrology*, 331: 161–177.

Marsh, R., Edwards, N., and Shepherd, J. (2002). "Development of a fast climate model (C-GOLDSTEIN) for Earth System Science." Technical report, Southampton Oceanography Centre, Document No. 83.

Morris, M. and Mitchell, T. (1995). "Exploratory designs for computer experiments." *J. Statist. Planning and Inference*, 43: 381–402.

Pritchard, L. and Kell, D. (2002). "Schemes of flux control in a model of Saccharomyces cerevisiae glycolysis." *Eur. J. Biochem.*, 269: 3894–3904.

Rougier, J. (2007a). "Efficient Emulators for Multivariate Deterministic Functions." Technical report, Department of Mathematics, Bristol University.

— (2007b). “Probabilistic inference for future climate using an ensemble of climate model evaluations.” *Climatic Change*, 81: 247–264.

Rougier, J. and Sexton, D. (2007). “Inference in ensemble experiments.” *Philosophical Transactions of the Royal Society, Series A*, 365: 2133–2143.

Sacks, J., Welch, W., Mitchell, T., and Wynn, H. (1989). “Design and analysis of computer experiments.” *Statistical Science*, 4: 409–423.

Saltelli, A., Chan, K., and Scott, E. (eds.) (2000). *Sensitivity Analysis*. New York: Wiley.

Thomas, G. W. (1982). *Principles of Hydrocarbon Reservoir Simulation*. IHRDC: Boston.

Todorov, I., Purton, J., Allan, N., and Dove, M. (2006a). “Simulation of radiation damage in gadolinium pyrochlores.” *Journal of Physics: Condensed Matter*, 18: 2217–2234.

Todorov, I., Smith, W., Trachenko, K., and Dove, M. (2006b). “DL POLY 3: new dimensions in molecular dynamics simulations via massive parallelism.” *Journal of Materials Chemistry*, 16: 1911–1918.

Watson, A. J. and Lovelock, J. E. (1983). “Biological homeostasis of the global environment: the parable of Daisyworld.” *Tellus*, 35B: 286–289.

Woodward, F. and Lomas, M. (2004). “Vegetation dynamics - simulating responses to climatic change.” *Biol. Rev.*, 79: 643–670.

Woodward, F., Smith, T., and Emanuel, W. (1995). “A global primary productivity and phytogeography model.” *Global Biogeochemical Cycles*, 9: 471–490.