Bayesian uncertainty analysis: linking simulators and climate

Michael Goldstein
Department of Mathematical Sciences,
Durham University *

* Thanks to EPSRC, NERC for funding, and to Jonathan Cumming, Ian Vernon, Danny Williamson

Climate simulators

Much of climate science falls into the general area of the analysis of complex physical systems by use of computer simulators. Here is a framing quote for the area, from Jon Lockley, manager of the Oxford Supercomputing Centre [from the BBC web-site]

“Whenever possible, everything is done in a supercomputer.

Look at Formula One - it’s getting rid of all of its wind tunnels and replacing them with supercomputers. It’s the same in the aerospace industry as well.

It means you can all the modelling in the supercomputer and then do just one real world test.”

Current state of the art

To use complex simulators to make statements about physical systems (like climate), we need to quantify the uncertainty involved in moving from the model to the system.

Current state of the art

To use complex simulators to make statements about physical systems (like climate), we need to quantify the uncertainty involved in moving from the model to the system.

The issues are both

practical/methodological (how can we work out what climate is likely to be?)

Current state of the art

To use complex simulators to make statements about physical systems (like climate), we need to quantify the uncertainty involved in moving from the model to the system.

The issues are both

practical/methodological (how can we work out what climate is likely to be?)

and

foundational (why should our methods work and what do our answers mean?)

Current state of the art

To use complex simulators to make statements about physical systems (like climate), we need to quantify the uncertainty involved in moving from the model to the system.

The issues are both

practical/methodological (how can we work out what climate is likely to be?)

and

foundational (why should our methods work and what do our answers mean?)

Many people work on different aspects of these uncertainty analyses

Great resource: the Managing Uncertainty in Complex Models web-site http://www.mucm.ac.uk/ (for references, papers, toolkit, etc.)

[MUCM is a consortium of U. of Aston, Durham, LSE, Sheffield, Southampton - with Basic Technology funding. Now mutating into MUCM community.]

Current state of the art

To use complex simulators to make statements about physical systems (like climate), we need to quantify the uncertainty involved in moving from the model to the system.

The issues are both

practical/methodological (how can we work out what climate is likely to be?)

and

foundational (why should our methods work and what do our answers mean?)

Many people work on different aspects of these uncertainty analyses

Great resource: the Managing Uncertainty in Complex Models web-site http://www.mucm.ac.uk/ (for references, papers, toolkit, etc.)

[MUCM is a consortium of U. of Aston, Durham, LSE, Sheffield, Southampton - with Basic Technology funding. Now mutating into MUCM community.]

This talk is concerned with a general overview of some key issues in this area.

Examples

Climate change Large scale climate simulators are constructed to assess likely effects of human intervention upon future climate behaviour.
Aims are both scientific - to understand the large scale interactions which determine climate - and also very practical, as we are only going to carry out one real world test for climate.

Examples

Climate change Large scale climate simulators are constructed to assess likely effects of human intervention upon future climate behaviour.

Aims are both scientific - to understand the large scale interactions which determine climate - and also very practical, as we are only going to carry out one real world test for climate.

Oil reservoirs An oil reservoir simulator is used to manage assets associated with the reservoir.

The aim is commercial: to develop efficient production schedules, determine whether and where to sink new wells, and so forth.

Examples

Climate change Large scale climate simulators are constructed to assess likely effects of human intervention upon future climate behaviour.
Aims are both scientific - to understand the large scale interactions which determine climate - and also very practical, as we are only going to carry out one real world test for climate.

Oil reservoirs An oil reservoir simulator is used to manage assets associated with the reservoir.
The aim is commercial: to develop efficient production schedules, determine whether and where to sink new wells, and so forth.

Galaxy formation
The study of the development of the Universe is carried out by using a Galaxy formation simulator.
The aim is scientific - to gain information about the physical processes underlying the Universe.

Sources of Uncertainty

Sources of Uncertainty

(i) parametric uncertainty (each model requires a, typically high dimensional, parametric specification)

Sources of Uncertainty

(i) parametric uncertainty (each model requires a, typically high dimensional, parametric specification)

(ii) condition uncertainty (uncertainty as to boundary conditions, initial conditions, and forcing functions),

Sources of Uncertainty

(i) parametric uncertainty (each model requires a, typically high dimensional, parametric specification)

(ii) condition uncertainty (uncertainty as to boundary conditions, initial conditions, and forcing functions),

(iii) functional uncertainty (model evaluations take a long time, so the function is unknown almost everywhere)

Sources of Uncertainty

(i) parametric uncertainty (each model requires a, typically high dimensional, parametric specification)

(ii) condition uncertainty (uncertainty as to boundary conditions, initial conditions, and forcing functions),

(iii) functional uncertainty (model evaluations take a long time, so the function is unknown almost everywhere)

(iv) stochastic uncertainty (either the model is stochastic, or it should be),

Sources of Uncertainty

(i) parametric uncertainty (each model requires a, typically high dimensional, parametric specification)

(ii) condition uncertainty (uncertainty as to boundary conditions, initial conditions, and forcing functions),

(iii) functional uncertainty (model evaluations take a long time, so the function is unknown almost everywhere)

(iv) stochastic uncertainty (either the model is stochastic, or it should be),

(v) solution uncertainty (as the system equations can only be solved to some necessary level of approximation).

Sources of Uncertainty

(i) parametric uncertainty (each model requires a, typically high dimensional, parametric specification)

(ii) condition uncertainty (uncertainty as to boundary conditions, initial conditions, and forcing functions),

(iii) functional uncertainty (model evaluations take a long time, so the function is unknown almost everywhere)

(iv) stochastic uncertainty (either the model is stochastic, or it should be),

(v) solution uncertainty (as the system equations can only be solved to some necessary level of approximation).

(vi) structural uncertainty (the model only approximates the physical system),

Sources of Uncertainty

(i) parametric uncertainty (each model requires a, typically high dimensional, parametric specification)
(ii) condition uncertainty (uncertainty as to boundary conditions, initial conditions, and forcing functions),
(iii) functional uncertainty (model evaluations take a long time, so the function is unknown almost everywhere)
(iv) stochastic uncertainty (either the model is stochastic, or it should be),
(v) solution uncertainty (as the system equations can only be solved to some necessary level of approximation).
(vi) structural uncertainty (the model only approximates the physical system),
(vii) measurement uncertainty (as the model is calibrated against system data all of which is measured with error),

Sources of Uncertainty

(i) parametric uncertainty (each model requires a, typically high dimensional, parametric specification)

(ii) condition uncertainty (uncertainty as to boundary conditions, initial conditions, and forcing functions),

(iii) functional uncertainty (model evaluations take a long time, so the function is unknown almost everywhere)

(iv) stochastic uncertainty (either the model is stochastic, or it should be),

(v) solution uncertainty (as the system equations can only be solved to some necessary level of approximation).

(vi) structural uncertainty (the model only approximates the physical system),

(vii) measurement uncertainty (as the model is calibrated against system data all of which is measured with error),

(viii) multi-model uncertainty (usually we have not one but many models related to the physical system)

Sources of Uncertainty

(i) parametric uncertainty (each model requires a, typically high dimensional, parametric specification)
(ii) condition uncertainty (uncertainty as to boundary conditions, initial conditions, and forcing functions),
(iii) functional uncertainty (model evaluations take a long time, so the function is unknown almost everywhere)
(iv) stochastic uncertainty (either the model is stochastic, or it should be),
(v) solution uncertainty (as the system equations can only be solved to some necessary level of approximation).
(vi) structural uncertainty (the model only approximates the physical system),
(vii) measurement uncertainty (as the model is calibrated against system data all of which is measured with error),
(viii) multi-model uncertainty (usually we have not one but many models related to the physical system)
(ix) decision uncertainty (to use the model to influence real world outcomes, we need to relate things in the world that we can influence to inputs to the simulator and through outputs to actual impacts. These links are uncertain.)

RAPID-WATCH (a NERC funding call)

Durham University

What are the implications of RAPID-WATCH observing system data and other recent observations for estimates of the risk due to rapid change in the MOC? In this context risk is taken to mean the probability of rapid change in the MOC and the consequent impact on climate (affecting temperatures, precipitation, sea level, for example). This project must:

* contribute to the MOC observing system assessment in 2011;
* investigate how observations of the MOC can be used to constrain estimates of the probability of rapid MOC change, including magnitude and rate of change;
* make sound statistical inferences about the real climate system from model simulations and observations;
* investigate the dependence of model uncertainty on such factors as changes of resolution;
* assess model uncertainty in climate impacts and characterise impacts that have received less attention (eg frequency of extremes).

The project must also demonstrate close partnership with the Hadley Centre.

RAPID-WATCH (a NERC funding call)

Durham University

What are the implications of RAPID-WATCH observing system data and other recent observations for estimates of the risk due to rapid change in the MOC? In this context risk is taken to mean the probability of rapid change in the MOC and the consequent impact on climate (affecting temperatures, precipitation, sea level, for example). This project must:

* contribute to the MOC observing system assessment in 2011;
* investigate how observations of the MOC can be used to constrain estimates of the probability of rapid MOC change, including magnitude and rate of change;
* make sound statistical inferences about the real climate system from model simulations and observations;
* investigate the dependence of model uncertainty on such factors as changes of resolution;
* assess model uncertainty in climate impacts and characterise impacts that have received less attention (eg frequency of extremes).

The project must also demonstrate close partnership with the Hadley Centre.

Subjectivist Bayes

In the subjectivist Bayes view, a probability statement is the uncertainty judgement of a specified individual, expressed on the scale of probability. This interpretation has a well-defined and operational meaning.

Subjectivist Bayes

In the subjectivist Bayes view, a probability statement is the uncertainty judgement of a specified individual, expressed on the scale of probability. This interpretation has a well-defined and operational meaning.

In this interpretation, any probability statement is the judgement of a named individual, so we should speak not of “the probability of rapid change in the MOC”, but instead of “Anne’s probability” or “Bob’s probability of rapid change in the MOC” and so forth.

Subjectivist Bayes

In the subjectivist Bayes view, a probability statement is the uncertainty judgement of a specified individual, expressed on the scale of probability. This interpretation has a well-defined and operational meaning.

In this interpretation, any probability statement is the judgement of a named individual, so we should speak not of “the probability of rapid change in the MOC”, but instead of “Anne’s probability” or “Bob’s probability of rapid change in the MOC” and so forth.

Most people expect something more authoritative and objective than a probability which is one person’s judgement.

Subjectivist Bayes

In the subjectivist Bayes view, a probability statement is the uncertainty judgement of a specified individual, expressed on the scale of probability. This interpretation has a well-defined and operational meaning.

In this interpretation, any probability statement is the judgement of a named individual, so we should speak not of “the probability of rapid change in the MOC”, but instead of “Anne’s probability” or “Bob’s probability of rapid change in the MOC” and so forth.

Most people expect something more authoritative and objective than a probability which is one person’s judgement.

However, the truly disappointing thing is that, almost always, stated scientific probabilities are not even the judgements of any individual.

Subjectivist Bayes

In the subjectivist Bayes view, a probability statement is the uncertainty judgement of a specified individual, expressed on the scale of probability. This interpretation has a well-defined and operational meaning.

In this interpretation, any probability statement is the judgement of a named individual, so we should speak not of “the probability of rapid change in the MOC”, but instead of “Anne’s probability” or “Bob’s probability of rapid change in the MOC” and so forth.

Most people expect something more authoritative and objective than a probability which is one person’s judgement.

However, the truly disappointing thing is that, almost always, stated scientific probabilities are not even the judgements of any individual.

Nor are these probabilities given any other meaning
(apart from as quantities in idealised models).

Scientific Bayesian analysis

A natural objective of a serious scientific uncertainty analysis, which aims to make statements about the real world, should be assessment of probabilities which are

Scientific Bayesian analysis

A natural objective of a serious scientific uncertainty analysis, which aims to make statements about the real world, should be assessment of probabilities which are
asserted by at least one person, expert in the area
[more would be better!]

Scientific Bayesian analysis

A natural objective of a serious scientific uncertainty analysis, which aims to make statements about the real world, should be assessment of probabilities which are

asserted by at least one person, expert in the area

[more would be better!]

for reasons that are open to outside scrutiny

Scientific Bayesian analysis

A natural objective of a serious scientific uncertainty analysis, which aims to make statements about the real world, should be assessment of probabilities which are

asserted by at least one person, expert in the area

[more would be better!]

for reasons that are open to outside scrutiny

including consideration of the range of alternative probability judgements which could plausibly be reached by other experts in the area

Scientific Bayesian analysis

A natural objective of a serious scientific uncertainty analysis, which aims to make statements about the real world, should be assessment of probabilities which are

asserted by at least one person, expert in the area

[more would be better!]

for reasons that are open to outside scrutiny

including consideration of the range of alternative probability judgements which could plausibly be reached by other experts in the area

This may be hard (but not harder than modelling climate itself), but it is necessary, if we intend to use our analysis to argue for change in real world behaviour.

Scientific Bayesian analysis

A natural objective of a serious scientific uncertainty analysis, which aims to make statements about the real world, should be assessment of probabilities which are
asserted by at least one person, expert in the area
[more would be better!]
for reasons that are open to outside scrutiny
including consideration of the range of alternative probability judgements which could plausibly be reached by other experts in the area
This may be hard (but not harder than modelling climate itself), but it is necessary, if we intend to use our analysis to argue for change in real world behaviour.
We shall describe the Bayesian approach to these problems, in which all uncertainties are treated probabilistically and combined by rules of Bayesian analysis.

Durham University

Within the Bayesian approach, we have two choices.

(i) Full Bayes analysis: complete joint probabilistic specification for all quantities

Bayes linear analysis

Within the Bayesian approach, we have two choices.

(i) Full Bayes analysis: complete joint probabilistic specification for all quantities

(ii) Bayes linear analysis, based only on expectation as a primitive, involving prior specification of means, variances and covariances

Bayes linear analysis

Within the Bayesian approach, we have two choices.

(i) Full Bayes analysis: complete joint probabilistic specification for all quantities

(ii) Bayes linear analysis, based only on expectation as a primitive, involving prior specification of means, variances and covariances

Probability is the most common choice, but there are advantages in working with expectations - the uncertainty specification is simpler, the analysis (particularly for experimental design) is much faster and more straightforward and there are careful foundations.

Bayes linear analysis

Within the Bayesian approach, we have two choices.

(i) Full Bayes analysis: complete joint probabilistic specification for all quantities

(ii) Bayes linear analysis, based only on expectation as a primitive, involving prior specification of means, variances and covariances

Probability is the most common choice, but there are advantages in working with expectations - the uncertainty specification is simpler, the analysis (particularly for experimental design) is much faster and more straightforward and there are careful foundations.

The statistical approach based around expectation is termed Bayes linear analysis, based around these updating equations for mean and variance:

$$\mathsf{E}_z[y] = \mathrm{E}(y) + \mathrm{Cov}(y,z)\mathrm{Var}(z)^{-1}(z - \mathrm{E}(z)),$$
$$\mathsf{Var}_z[y] = \mathrm{Var}(y) - \mathrm{Cov}(y,z)\mathrm{Var}(z)^{-1}\mathrm{Cov}(z,y)$$

Bayes linear analysis

Within the Bayesian approach, we have two choices.

(i) Full Bayes analysis: complete joint probabilistic specification for all quantities

(ii) Bayes linear analysis, based only on expectation as a primitive, involving prior specification of means, variances and covariances

Probability is the most common choice, but there are advantages in working with expectations - the uncertainty specification is simpler, the analysis (particularly for experimental design) is much faster and more straightforward and there are careful foundations.

The statistical approach based around expectation is termed Bayes linear analysis, based around these updating equations for mean and variance:

$$\mathsf{E}_z[y] = \mathrm{E}(y) + \mathrm{Cov}(y,z)\mathrm{Var}(z)^{-1}(z - \mathrm{E}(z)),$$
$$\mathsf{Var}_z[y] = \mathrm{Var}(y) - \mathrm{Cov}(y,z)\mathrm{Var}(z)^{-1}\mathrm{Cov}(z,y)$$

The examples that we will describe use Bayes linear methods.

(There are natural (but much more complicated) probabilistic counterparts.)

Function emulation

Our simulator can be conceived as a function $f(x)$, where

x: input vector, representing unknown properties of the physical system;

$f(x)$: output vector representing system behaviour.

Function emulation

Our simulator can be conceived as a function $f(x)$, where

x: input vector, representing unknown properties of the physical system;

$f(x)$: output vector representing system behaviour.

Uncertainty analysis, for high dimensional problems, is particularly challenging if $f(x)$ is expensive, in time and computational resources, to evaluate for any choice of x. [For example, large climate models.]

In such cases, f must be treated as uncertain for all input choices except the small subset for which an actual evaluation has been made.

Function emulation

Our simulator can be conceived as a function $f(x)$, where
x: input vector, representing unknown properties of the physical system;
$f(x)$: output vector representing system behaviour.
Uncertainty analysis, for high dimensional problems, is particularly challenging if $f(x)$ is expensive, in time and computational resources, to evaluate for any choice of x. [For example, large climate models.]

In such cases, f must be treated as uncertain for all input choices except the small subset for which an actual evaluation has been made.
Therefore, we must construct a description of the uncertainty about the value of $f(x)$ for each x.
Such a representation is often termed an emulator of the function - the emulator both suggests an approximation to the function and also contains an assessment of the likely magnitude of the error of the approximation.

Function emulation

Our simulator can be conceived as a function $f(x)$, where
x: input vector, representing unknown properties of the physical system;
$f(x)$: output vector representing system behaviour.
Uncertainty analysis, for high dimensional problems, is particularly challenging if $f(x)$ is expensive, in time and computational resources, to evaluate for any choice of x. [For example, large climate models.]

In such cases, f must be treated as uncertain for all input choices except the small subset for which an actual evaluation has been made.
Therefore, we must construct a description of the uncertainty about the value of $f(x)$ for each x.
Such a representation is often termed an emulator of the function - the emulator both suggests an approximation to the function and also contains an assessment of the likely magnitude of the error of the approximation.
We use the emulator either to provide a full joint probabilistic description of all of the function values (full Bayes) or to assess expectations variances and covariances for pairs of function values (Bayes linear).

Form of the emulator

We may represent beliefs about component f_i of f, using an emulator:

$$f_i(x) = \sum_j \beta_{ij} g_{ij}(x) \oplus u_i(x)$$

Form of the emulator

We may represent beliefs about component f_i of f, using an emulator:

$$f_i(x) = \sum_j \beta_{ij} g_{ij}(x) \oplus u_i(x)$$

$Bg(x)$ expresses global variation in f: $B = \{\beta_{ij}\}$ are unknown scalars, g_{ij} are known deterministic functions of x (for example, polynomials)

Form of the emulator

We may represent beliefs about component f_i of f, using an emulator:

$$f_i(x) = \sum_j \beta_{ij} g_{ij}(x) \oplus u_i(x)$$

$Bg(x)$ expresses global variation in f: $B = \{\beta_{ij}\}$ are unknown scalars, g_{ij} are known deterministic functions of x (for example, polynomials)
$u(x)$ expresses local variation in f: $u_i(x)$ is a weakly second order stationary stochastic process, with (for example) correlation function

$$\mathrm{Corr}(u_i(x), u_i(x')) = \exp(-(\tfrac{\|x-x'\|}{\theta_i})^2)$$

Form of the emulator

We may represent beliefs about component f_i of f, using an emulator:

$$f_i(x) = \sum_j \beta_{ij} g_{ij}(x) \oplus u_i(x)$$

$Bg(x)$ expresses global variation in f: $B = \{\beta_{ij}\}$ are unknown scalars, g_{ij} are known deterministic functions of x (for example, polynomials)
$u(x)$ expresses local variation in f: $u_i(x)$ is a weakly second order stationary stochastic process, with (for example) correlation function

$$\mathrm{Corr}(u_i(x), u_i(x')) = \exp(-(\tfrac{\|x-x'\|}{\theta_i})^2)$$

We fit the emulators, given a collection of carefully chosen model evaluations, using our favourite statistical tools - generalised least squares, maximum likelihood, Bayes - with a generous helping of expert judgement.

Form of the emulator

We may represent beliefs about component f_i of f, using an emulator:

$$f_i(x) = \sum_j \beta_{ij} g_{ij}(x) \oplus u_i(x)$$

$Bg(x)$ expresses global variation in f: $B = \{\beta_{ij}\}$ are unknown scalars, g_{ij} are known deterministic functions of x (for example, polynomials)
$u(x)$ expresses local variation in f: $u_i(x)$ is a weakly second order stationary stochastic process, with (for example) correlation function

$$\mathrm{Corr}(u_i(x), u_i(x')) = \exp(-(\tfrac{\|x-x'\|}{\theta_i})^2)$$

We fit the emulators, given a collection of carefully chosen model evaluations, using our favourite statistical tools - generalised least squares, maximum likelihood, Bayes - with a generous helping of expert judgement.
We need careful (multi-output) experimental design to choose informative model evaluations, and detailed diagnostics to check emulator validity.

Linked emulators

If the simulator is really slow to evaluate, then we emulate by jointly modelling the simulator with a fast approximate version, f^c, plus older generations of the simulator which we've already emulated and so forth.

Linked emulators

If the simulator is really slow to evaluate, then we emulate by jointly modelling the simulator with a fast approximate version, f^c, plus older generations of the simulator which we've already emulated and so forth.

So, for example, based on many fast simulator evaluations, we build emulator

$$f_i^c(x) = \sum_j \beta_{ij}^c g_{ij}(x) \oplus u_i^c(x)$$

Linked emulators

If the simulator is really slow to evaluate, then we emulate by jointly modelling the simulator with a fast approximate version, f^c, plus older generations of the simulator which we've already emulated and so forth.

So, for example, based on many fast simulator evaluations, we build emulator

$$f_i^c(x) = \sum_j \beta_{ij}^c g_{ij}(x) \oplus u_i^c(x)$$

We use this form as the prior for the emulator for $f_i(x)$.

Then a relatively small number of evaluations of $f_i(x)$, using relations such as

$$\beta_{ij} = \alpha_i \beta_{ij}^c + \gamma_{ij}$$

lets us adjust the prior emulator to an appropriate posterior emulator for $f_i(x)$.

Linked emulators

If the simulator is really slow to evaluate, then we emulate by jointly modelling the simulator with a fast approximate version, f^c, plus older generations of the simulator which we've already emulated and so forth.
So, for example, based on many fast simulator evaluations, we build emulator

$$f_i^c(x) = \sum_j \beta_{ij}^c g_{ij}(x) \oplus u_i^c(x)$$

We use this form as the prior for the emulator for $f_i(x)$.
Then a relatively small number of evaluations of $f_i(x)$, using relations such as

$$\beta_{ij} = \alpha_i \beta_{ij}^c + \gamma_{ij}$$

lets us adjust the prior emulator to an appropriate posterior emulator for $f_i(x)$.
[This approach exploits the heuristic that we need many more function evaluations to identify the qualitative form of the model (i.e. choose appropriate forms $g_{ij}(x)$, etc) than to assess the quantitative form of all of the terms in the model - particularly if we fit meaningful regression components.]

Illustration from RAPID (thanks to Danny Williamson)

One of the main aims of the RAPIT programme is to assess the risk of shutdown of the AMOC (Atlantic Meridional Overturning Circulation)which transports heat from the tropics to Northern Europe and how this risk depends on the future emissions scenario for CO_2.

Illustration from RAPID (thanks to Danny Williamson)

One of the main aims of the RAPIT programme is to assess the risk of shutdown of the AMOC (Atlantic Meridional Overturning Circulation)which transports heat from the tropics to Northern Europe and how this risk depends on the future emissions scenario for CO_2.

RAPIT aims to use large ensembles of the UK Met Office climate model HadCM3, run through climateprediction.net

[An ensemble of many thousands of runs has been collected.]

Illustration from RAPID (thanks to Danny Williamson)

One of the main aims of the RAPIT programme is to assess the risk of shutdown of the AMOC (Atlantic Meridional Overturning Circulation)which transports heat from the tropics to Northern Europe and how this risk depends on the future emissions scenario for CO2.

RAPIT aims to use large ensembles of the UK Met Office climate model HadCM3, run through climateprediction.net

[An ensemble of many thousands of runs has been collected.]

As a preliminary demonstration of concept for the Met Office, we were asked to develop an emulator for HadCM3, based on 24 runs of the simulator, with a variety of parameter choices and future CO2 scenarios.

Illustration from RAPID (thanks to Danny Williamson)

One of the main aims of the RAPIT programme is to assess the risk of shutdown of the AMOC (Atlantic Meridional Overturning Circulation)which transports heat from the tropics to Northern Europe and how this risk depends on the future emissions scenario for CO2.

RAPIT aims to use large ensembles of the UK Met Office climate model HadCM3, run through climateprediction.net

[An ensemble of many thousands of runs has been collected.]

As a preliminary demonstration of concept for the Met Office, we were asked to develop an emulator for HadCM3, based on 24 runs of the simulator, with a variety of parameter choices and future CO2 scenarios.

We had access to some runs of FAMOUS (a lower resolution model), which consisted of 6 scenarios for future CO2 forcing, and between 40 and 80 runs of FAMOUS under each scenario, with different parameter choices.

[And very little time to do the analysis.]

Model outputs

In this experiment only 3 input parameters were varied (an entrainment coefficient in the model atmosphere, a vertical mixing parameter in the ocean, and the solar constant).

Our output of interest was a 170 year time series of AMOC values.

Model outputs

In this experiment only 3 input parameters were varied (an entrainment coefficient in the model atmosphere, a vertical mixing parameter in the ocean, and the solar constant).

Our output of interest was a 170 year time series of AMOC values.

The series is noisy and and the location and direction of spikes in the series was not important.

Model outputs

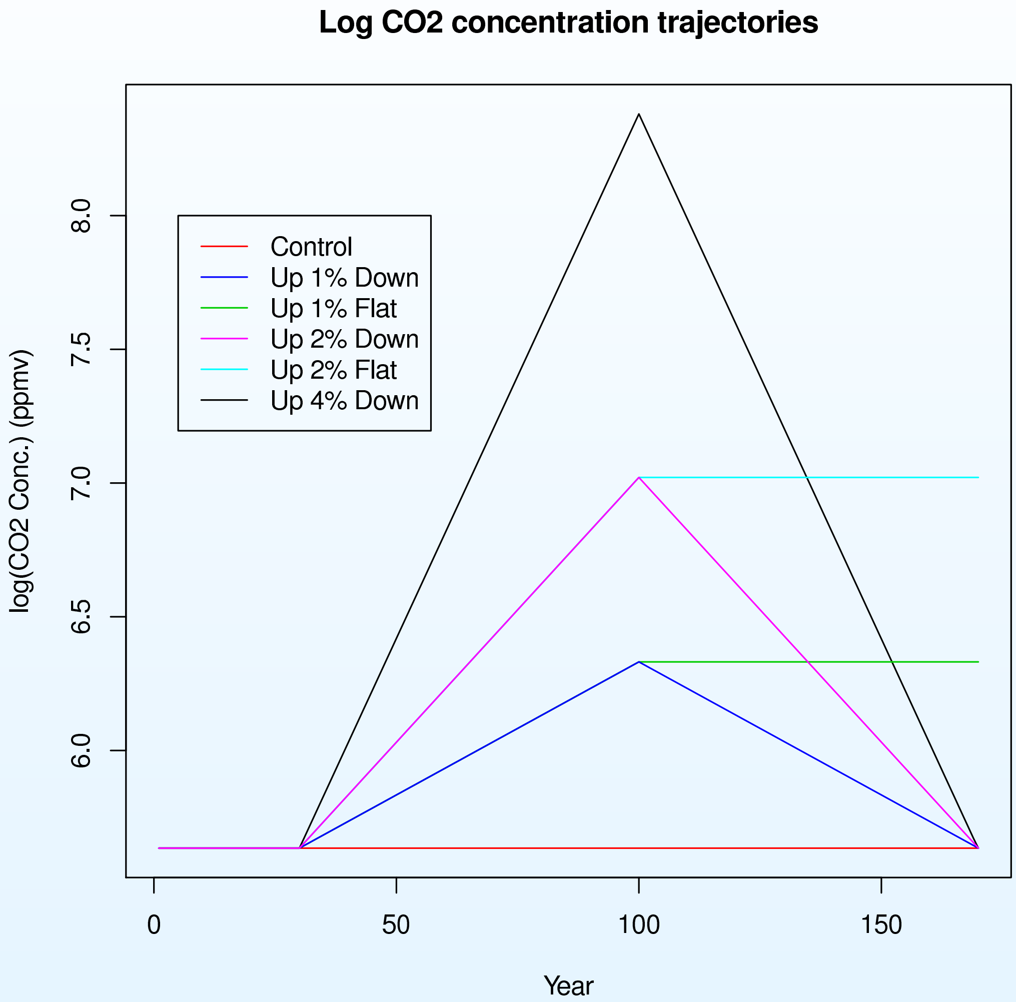
In this experiment only 3 input parameters were varied (an entrainment coefficient in the model atmosphere, a vertical mixing parameter in the ocean, and the solar constant).

Our output of interest was a 170 year time series of AMOC values.

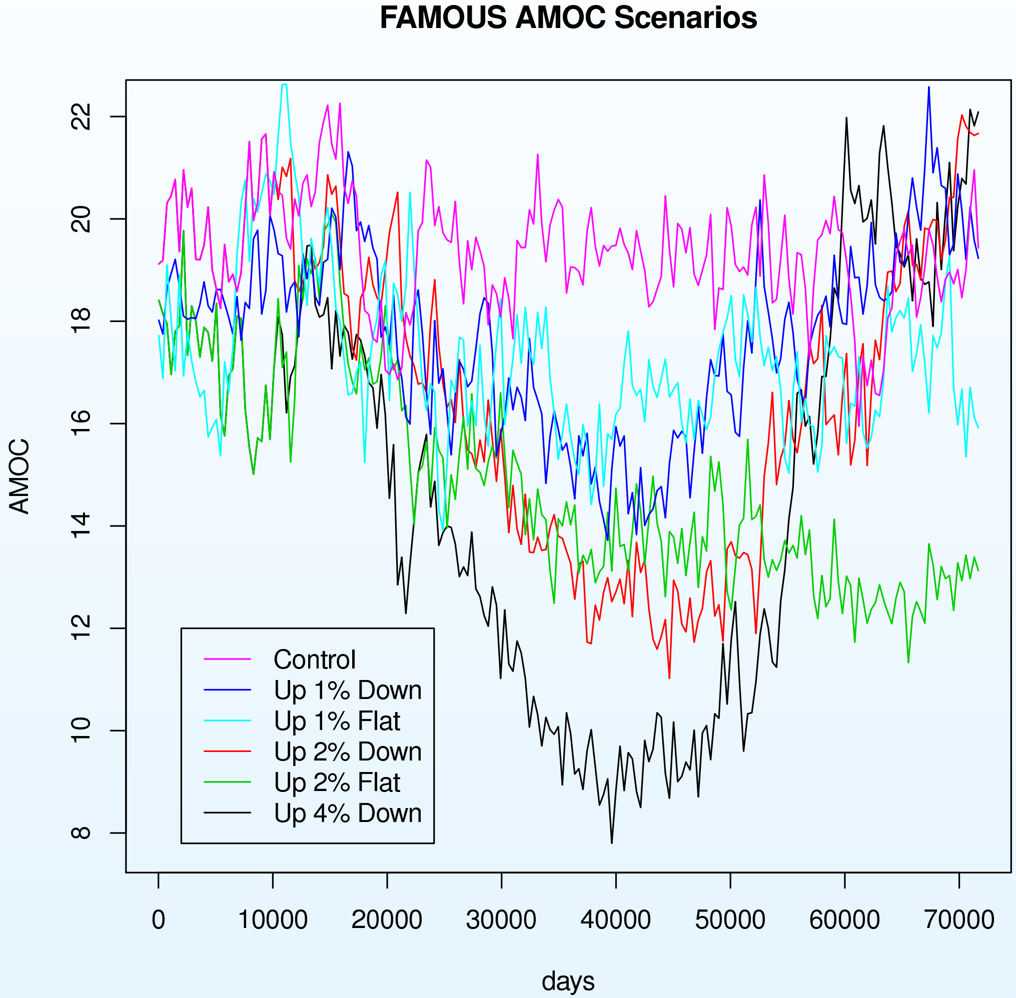
The series is noisy and and the location and direction of spikes in the series was not important.

Interest concerned aspects such as:
the value and location of the smoothed minimum of the series
the amount that AMOC responds to CO_2 forcing and recovers if CO_2 forcing is reduced.

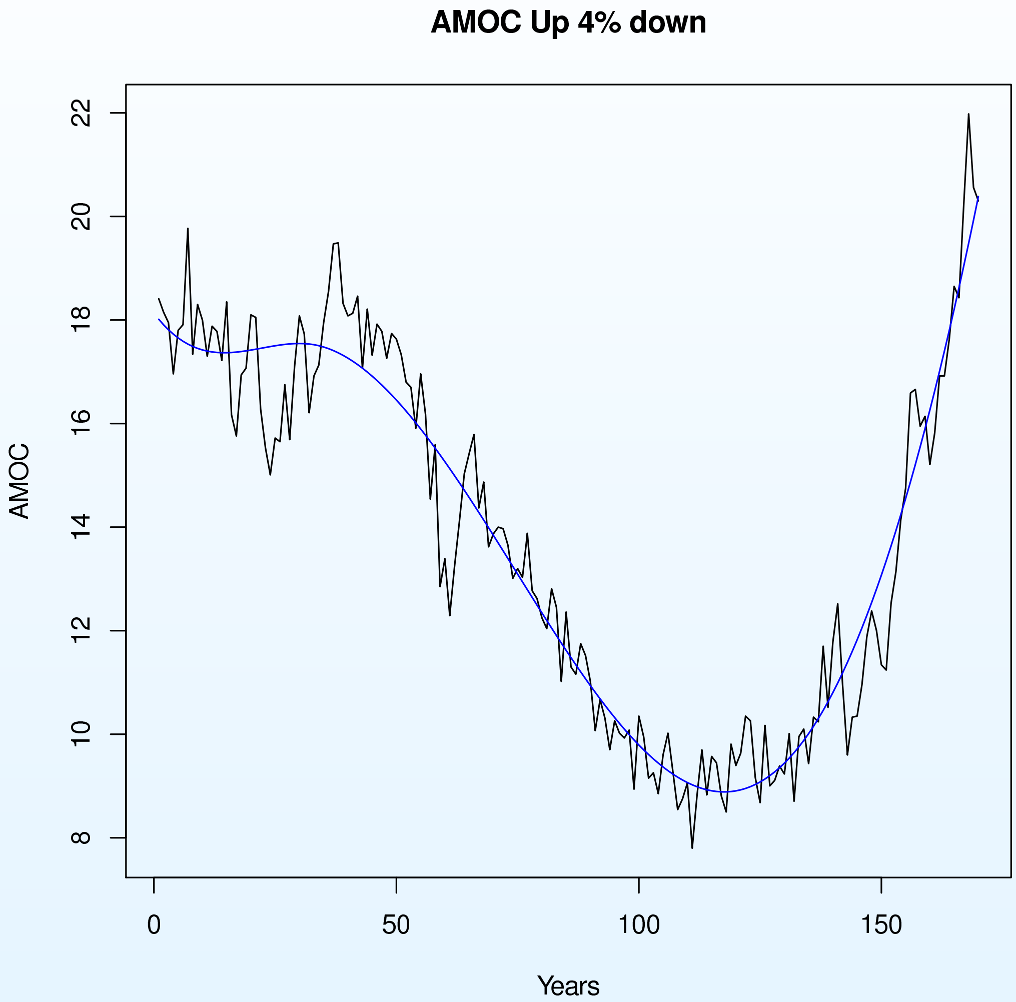
CO2 Scenarios



Famous Scenarios

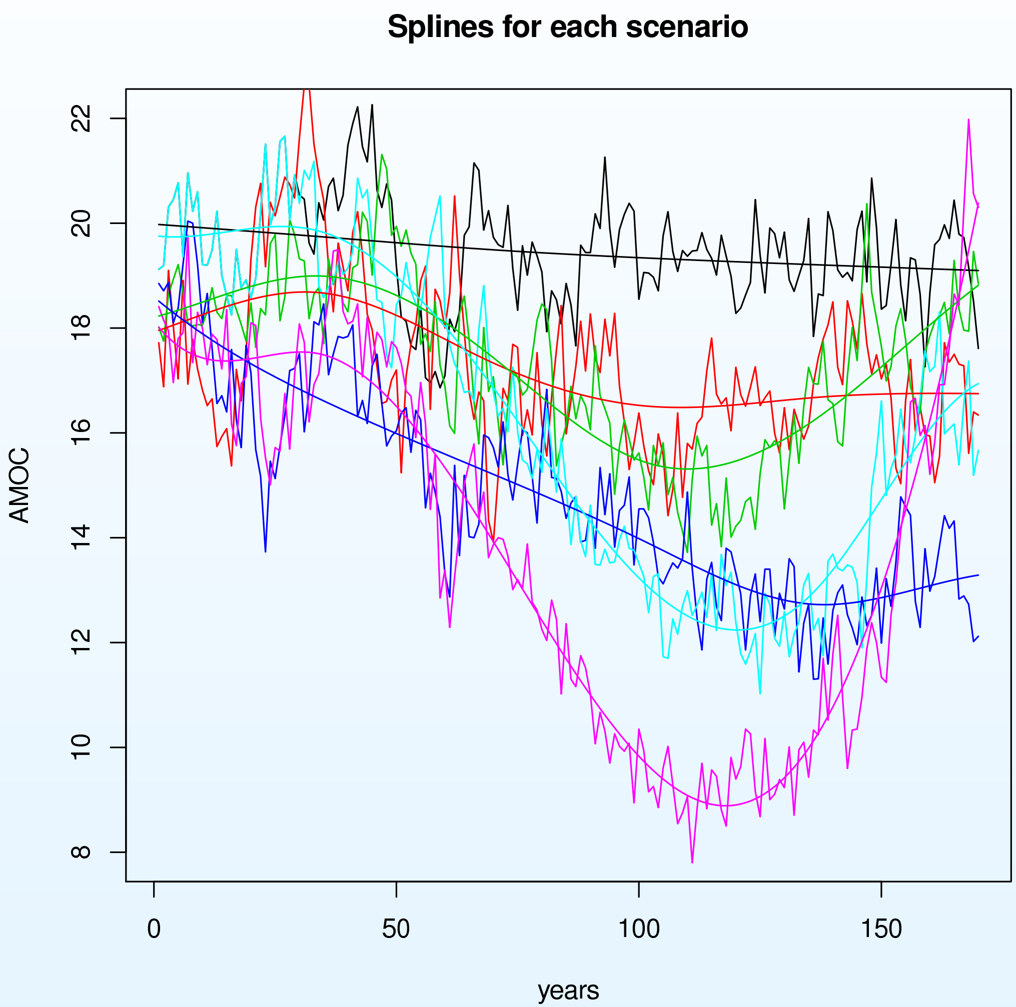


Smoothing



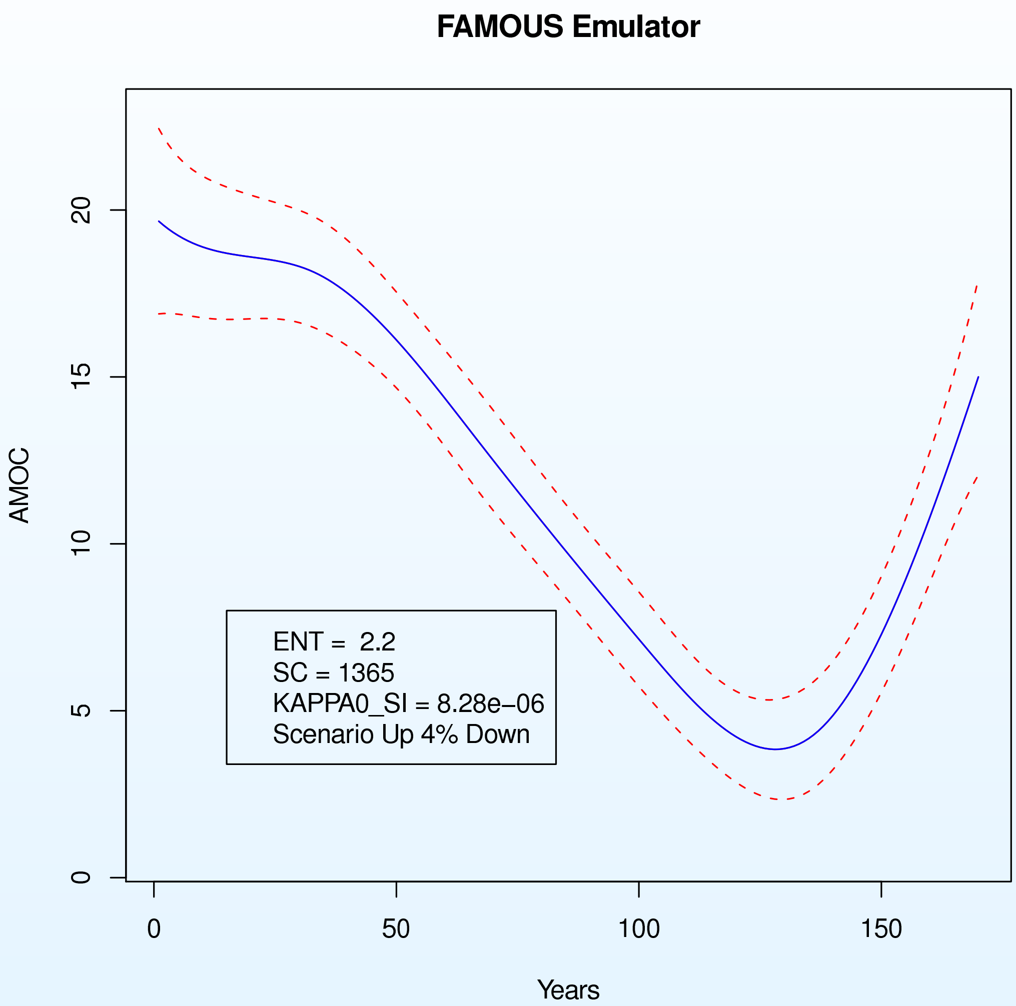
Smooth by fitting splines $f^s(x,t) = \Sigma_j c_j(x) B_j(t)$ where $B_j(t)$ are basis functions over t and $c_j(x)$ are chosen to give the 'best' smooth fit to the time series.

Smoothing



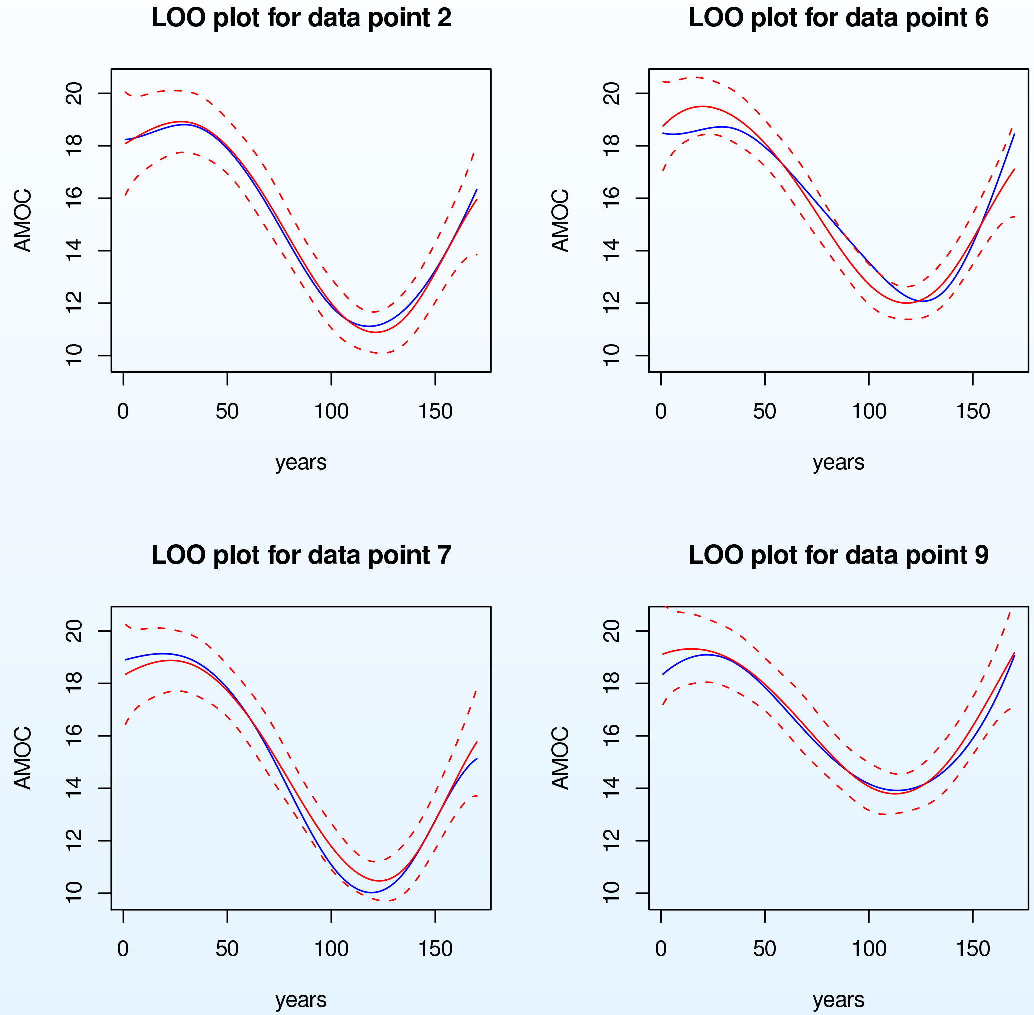
Splines for each scenario
AMOC
years
8
10
12
14
16
18
20
22
0
50
100
150

FAMOUS emulation



Emulate f^s by emulating each coefficient $c_j(x)$ in $f^s(x,t) = \Sigma_j c_j(x) B_j(t)$ (separately for each CO2 scenario)

Diagnostics (leave one out)



Test approach by building emulators leaving out each observed run in turn, and checking whether the run falls within the stated uncertainty limits.

Emulating HadCM3

We now have an emulator for the smoothed version of FAMOUS, for each of the 6 CO_2 scenarios. Next steps

[1] Extend the FAMOUS emulator across all choices of CO_2 scenario.

[We do this using fast geometric arguments, exploiting the speed of working in inner product spaces. For example, we have a different covariance matrix for local variation at each of 6 CO_2 scenarios. We extend this specification to all possible CO_2 scenarios by identifying each covariance matrix as an element of an appropriate inner product space, and adjusting beliefs over covariance matrix space by projection.]

Emulating HadCM3

We now have an emulator for the smoothed version of FAMOUS, for each of the 6 CO_2 scenarios. Next steps

[1] Extend the FAMOUS emulator across all choices of CO_2 scenario.

[We do this using fast geometric arguments, exploiting the speed of working in inner product spaces. For example, we have a different covariance matrix for local variation at each of 6 CO_2 scenarios. We extend this specification to all possible CO_2 scenarios by identifying each covariance matrix as an element of an appropriate inner product space, and adjusting beliefs over covariance matrix space by projection.]

[2] Develop relationships between the elements of the emulator for FAMOUS and the corresponding emulator for HadCM3, using 8 paired runs, and expert judgements. This gives an informed prior for the HadCM3 emulator.

Emulating HadCM3

We now have an emulator for the smoothed version of FAMOUS, for each of the 6 CO_2 scenarios. Next steps

[1] Extend the FAMOUS emulator across all choices of CO_2 scenario.

[We do this using fast geometric arguments, exploiting the speed of working in inner product spaces. For example, we have a different covariance matrix for local variation at each of 6 CO_2 scenarios. We extend this specification to all possible CO_2 scenarios by identifying each covariance matrix as an element of an appropriate inner product space, and adjusting beliefs over covariance matrix space by projection.]

[2] Develop relationships between the elements of the emulator for FAMOUS and the corresponding emulator for HadCM3, using 8 paired runs, and expert judgements. This gives an informed prior for the HadCM3 emulator.

[3] Use the remaining design of 16 runs of HadCM3 (a 16 run Latin hypercube over different parameter choices and CO_2 scenarios) to Bayes linear update the emulator for HadCM3.

Emulating HadCM3

We now have an emulator for the smoothed version of FAMOUS, for each of the 6 CO2 scenarios. Next steps

[1] Extend the FAMOUS emulator across all choices of CO2 scenario.

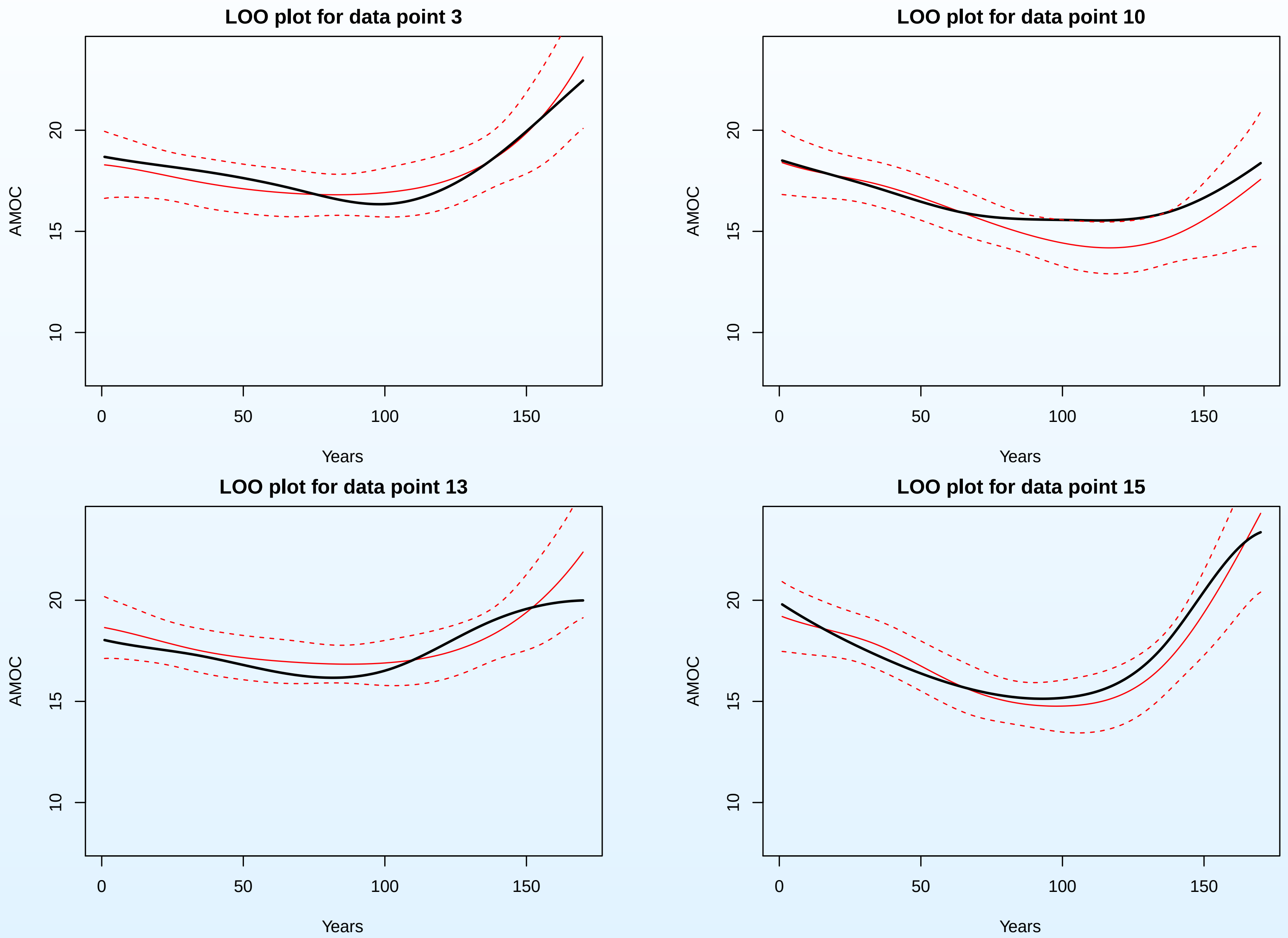
[We do this using fast geometric arguments, exploiting the speed of working in inner product spaces. For example, we have a different covariance matrix for local variation at each of 6 CO2 scenarios. We extend this specification to all possible CO2 scenarios by identifying each covariance matrix as an element of an appropriate inner product space, and adjusting beliefs over covariance matrix space by projection.]

[2] Develop relationships between the elements of the emulator for FAMOUS and the corresponding emulator for HadCM3, using 8 paired runs, and expert judgements. This gives an informed prior for the HadCM3 emulator.

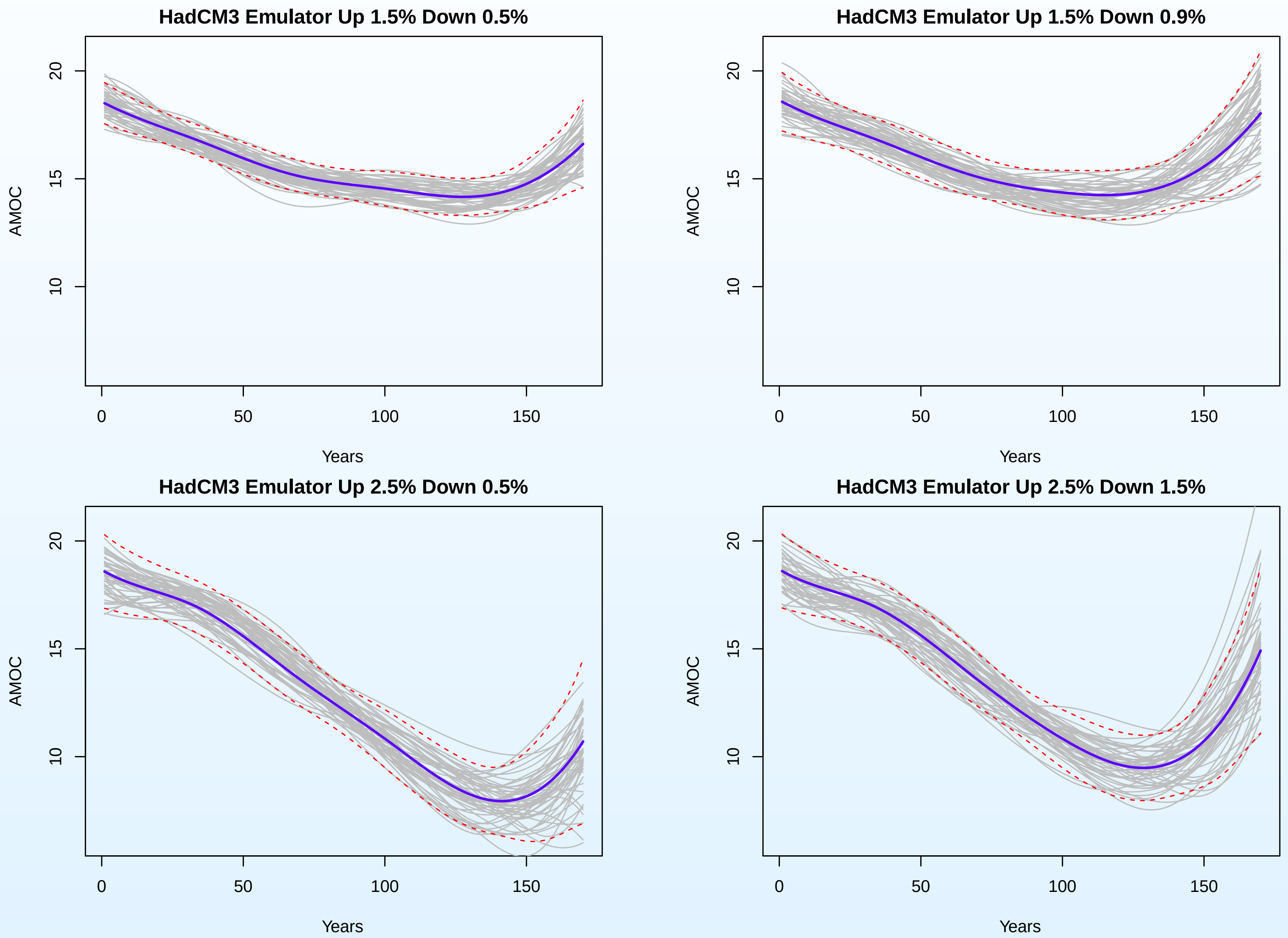
[3] Use the remaining design of 16 runs of HadCM3 (a 16 run Latin hypercube over different parameter choices and CO2 scenarios) to Bayes linear update the emulator for HadCM3.

[4] Diagnostic checking, tuning etc.

Emulating HadCM3:diagnostics



Emulating HadCM3



Simulators for Oil Reservoirs

An oil reservoir is an underground region of porous rock which contains oil and/or gas. The hydrocarbons are trapped above by a layer of impermeable rock and below by a body of water, thus creating the reservoir. The oil and gas are pumped out of the reservoir and fluids are pumped into the reservoir (to boost production).

Simulators for Oil Reservoirs

An oil reservoir is an underground region of porous rock which contains oil and/or gas. The hydrocarbons are trapped above by a layer of impermeable rock and below by a body of water, thus creating the reservoir. The oil and gas are pumped out of the reservoir and fluids are pumped into the reservoir (to boost production).

A reservoir simulator models the flows and distributions of contents of the reservoir over time.

Simulators for Oil Reservoirs

An oil reservoir is an underground region of porous rock which contains oil and/or gas. The hydrocarbons are trapped above by a layer of impermeable rock and below by a body of water, thus creating the reservoir. The oil and gas are pumped out of the reservoir and fluids are pumped into the reservoir (to boost production).

A reservoir simulator models the flows and distributions of contents of the reservoir over time.

The simulator inputs are the features of the geology of the reservoir
such as permeability, porosity, fault transmissibility, aquifer features, saturation properties

Simulators for Oil Reservoirs

An oil reservoir is an underground region of porous rock which contains oil and/or gas. The hydrocarbons are trapped above by a layer of impermeable rock and below by a body of water, thus creating the reservoir. The oil and gas are pumped out of the reservoir and fluids are pumped into the reservoir (to boost production).

A reservoir simulator models the flows and distributions of contents of the reservoir over time.

The simulator inputs are the features of the geology of the reservoir such as permeability, porosity, fault transmissibility, aquifer features, saturation properties

The simulator outputs comprise the behaviour of the various wells and injectors in the reservoir. Output is time series on features such as bottom hole pressures, oil, gas and water rates.

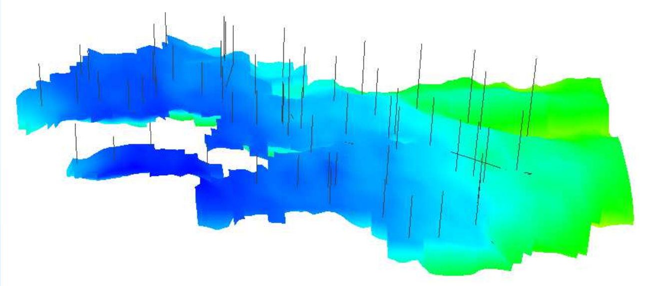
A reservoir example: (thanks to Jonathan Cumming)

The model, based on grid size $38 \times 87 \times 25$, with 43 production and 13 injection wells, simulates 10 years of production, 1.5–3 hours per simulation.

A reservoir example: (thanks to Jonathan Cumming)

The model, based on grid size $38 \times 87 \times 25$, with 43 production and 13 injection wells, simulates 10 years of production, 1.5–3 hours per simulation.

Inputs Field multipliers for porosity (ϕ), permeabilities (k_x, k_z), critical saturation (crw), and aquifer properties (A_p, A_h)
Outputs Oil production rate for a 3-year period, for the 10 production wells active in that period. 4-month averages over the time series



Coarse and Accurate Emulators

The computer model is expensive to evaluate, so we use 'coarse' model, F^c (by coarsening vertical gridding by factor of 10), to capture qualitative features of F. F^c is substantially faster, allowing 1000 model runs (in a Latin Hypercube over the input parameters).

Coarse and Accurate Emulators

The computer model is expensive to evaluate, so we use ‘coarse’ model, F^c (by coarsening vertical gridding by factor of 10), to capture qualitative features of F. F^c is substantially faster, allowing 1000 model runs (in a Latin Hypercube over the input parameters).

For chosen outputs from F^c, we construct emulators f_i^c,from these runs of form

$$f_i^c(\boldsymbol{x}) = \boldsymbol{g}_i(\boldsymbol{x}_{[i]})^T \boldsymbol{\beta}_i^c + w_i^c(\boldsymbol{x})$$

where $\boldsymbol{x}_{[i]}$ – a subset of inputs which account for most of the variation in F, the *active variables*

Coarse and Accurate Emulators

The computer model is expensive to evaluate, so we use ‘coarse’ model, F^c (by coarsening vertical gridding by factor of 10), to capture qualitative features of F. F^c is substantially faster, allowing 1000 model runs (in a Latin Hypercube over the input parameters).

For chosen outputs from F^c, we construct emulators f_i^c,from these runs of form

$$f_i^c(\boldsymbol{x}) = \boldsymbol{g}_i(\boldsymbol{x}_{[i]})^T \boldsymbol{\beta}_i^c + w_i^c(\boldsymbol{x})$$

where $\boldsymbol{x}_{[i]}$ – a subset of inputs which account for most of the variation in F, the *active variables*

We now make 20 runs of F, based on a careful choice of small design. We (Bayes linear) update our emulator f for F, supposing the full model emulator to have the form

$f_i(\boldsymbol{x}) = \boldsymbol{g}_i(\boldsymbol{x}_{[i]})^T \beta_i + w_i^c(\boldsymbol{x})\beta_{w_i} + w_i^a(\boldsymbol{x})$

(linked via equations relating corresponding pairs of coefficients, etc.)

Design considerations

When the simulator is expensive to evaluate, it is important to make good choices of inputs for simulator evaluations.

(For example, varying one input at a time is very wasteful.)

Design considerations

When the simulator is expensive to evaluate, it is important to make good choices of inputs for simulator evaluations.

(For example, varying one input at a time is very wasteful.)

The key feature of simulator design is that we aim to learn about many outputs from one experimental design.

General purpose designs (like Latin hypercubes) are unlikely to be good choices for this.

Design considerations

When the simulator is expensive to evaluate, it is important to make good choices of inputs for simulator evaluations.

(For example, varying one input at a time is very wasteful.)

The key feature of simulator design is that we aim to learn about many outputs from one experimental design.

General purpose designs (like Latin hypercubes) are unlikely to be good choices for this.

If a particular output f_i is largely determined by a subset $x_{[i]}$ of **active** inputs, then we require our design to be balanced and informative for that subset (for example so that we can interpolate in the key dimensions).

Design considerations

When the simulator is expensive to evaluate, it is important to make good choices of inputs for simulator evaluations.
(For example, varying one input at a time is very wasteful.)
The key feature of simulator design is that we aim to learn about many outputs from one experimental design.
General purpose designs (like Latin hypercubes) are unlikely to be good choices for this.
If a particular output f_i is largely determined by a subset $x_{[i]}$ of **active** inputs, then we require our design to be balanced and informative for that subset (for example so that we can interpolate in the key dimensions).
Typically, the active input sets will be different for different outputs. However, these active sets will not be disjoint.

Border block designs

To deal with the interlocking sets of active variables, we must construct **border-block** designs. eg. if
$x_{[1]} = \{x_1, x_2, x_3, x_4\}, x_{[2]} = \{x_1, x_2, x_5, x_6\}$,
Then the border is $\{x_1, x_2\}$ and the blocks are $\{x_3, x_4\}$, and $\{x_5, x_6\}$

Border block designs

To deal with the interlocking sets of active variables, we must construct **border-block** designs. eg. if

$x_{[1]} = \{x_1, x_2, x_3, x_4\}, x_{[2]} = \{x_1, x_2, x_5, x_6\}$,

Then the border is $\{x_1, x_2\}$ and the blocks are $\{x_3, x_4\}$, and $\{x_5, x_6\}$

For fixed border values, we can design independently for each block. Having fixed all the blocks, we can then design for the border. Then we iterate this process.

Border block designs

To deal with the interlocking sets of active variables, we must construct **border-block** designs. eg. if

$x_{[1]} = \{x_1, x_2, x_3, x_4\}, x_{[2]} = \{x_1, x_2, x_5, x_6\}$,

Then the border is $\{x_1, x_2\}$ and the blocks are $\{x_3, x_4\}$, and $\{x_5, x_6\}$

For fixed border values, we can design independently for each block. Having fixed all the blocks, we can then design for the border. Then we iterate this process.

How do we identify the border/block separation?

Many runs of fast simulator

These also suggest the types of emulators that we should design for, in terms of natural scales for inputs, key interactions and so forth.

Emulation Summaries

Well	Time	$\boldsymbol{x}_{[i]}$	No. Model Terms	Coarse Simulator R^2	Accurate Simulator $\tilde{R}^2$
B2	4	ϕ, crw, A_p	9	0.886	0.951
B2	8	ϕ, crw, A_p	7	0.959	0.958
B2	12	ϕ, crw, A_p	10	0.978	0.995
B2	16	ϕ, crw, k_z	7	0.970	0.995
B2	20	ϕ, crw, k_x	11	0.967	0.986
B2	24	ϕ, crw, k_x	10	0.970	0.970
B2	28	ϕ, crw, k_x	10	0.975	0.981
B2	32	ϕ, crw, k_x	11	0.980	0.951
B2	36	ϕ, crw, k_x	11	0.983	0.967

History matching

Model calibration aims to identify “true” input parameters x^*. However

History matching

Model calibration aims to identify "true" input parameters x^*. However

(i) We may not believe in a unique true input value for the model;

History matching

Model calibration aims to identify “true” input parameters x^*. However

(i) We may not believe in a unique true input value for the model;

(ii) We may be unsure whether there are any good choices of input parameters

History matching

Model calibration aims to identify "true" input parameters x^*. However

(i) We may not believe in a unique true input value for the model;

(ii) We may be unsure whether there are any good choices of input parameters

(iii) Full probabilistic calibration analysis may be very difficult/non-robust.

History matching

Model calibration aims to identify "true" input parameters x^*. However
(i) We may not believe in a unique true input value for the model;
(ii) We may be unsure whether there are any good choices of input parameters
(iii) Full probabilistic calibration analysis may be very difficult/non-robust.
A conceptually simple procedure is "history matching", i.e. finding the collection, $C(z)$, of all input choices x for which you judge the match of the model outputs $f_h(x)$ to observed data, z, to be acceptably small, taking into account all of the uncertainties in the problem.

History matching

Model calibration aims to identify “true” input parameters x^*. However
(i) We may not believe in a unique true input value for the model;
(ii) We may be unsure whether there are any good choices of input parameters
(iii) Full probabilistic calibration analysis may be very difficult/non-robust.

A conceptually simple procedure is “history matching”, i.e. finding the collection, $C(z)$, of all input choices x for which you judge the match of the model outputs $f_h(x)$ to observed data, z, to be acceptably small, taking into account all of the uncertainties in the problem.

If $C(z)$ is empty, then that is clear evidence of problems with the model (missing physics, coding errors, etc.) or with the data.

History matching

Model calibration aims to identify "true" input parameters x^*. However

(i) We may not believe in a unique true input value for the model;

(ii) We may be unsure whether there are any good choices of input parameters

(iii) Full probabilistic calibration analysis may be very difficult/non-robust.

A conceptually simple procedure is "history matching", i.e. finding the collection, $C(z)$, of all input choices x for which you judge the match of the model outputs $f_h(x)$ to observed data, z, to be acceptably small, taking into account all of the uncertainties in the problem.

If $C(z)$ is empty, then that is clear evidence of problems with the model (missing physics, coding errors, etc.) or with the data.

If $C(z)$ is non-empty, then an analysis of its elements reveals the constraints on the parameter space imposed by the data.

Further the model projections $f(x) : x \in C(z)$ over future outcomes, reveal the futures consistent with the model physics and the historical data.

History matching

Model calibration aims to identify "true" input parameters x^*. However
(i) We may not believe in a unique true input value for the model;
(ii) We may be unsure whether there are any good choices of input parameters
(iii) Full probabilistic calibration analysis may be very difficult/non-robust.

A conceptually simple procedure is "history matching", i.e. finding the collection, $C(z)$, of all input choices x for which you judge the match of the model outputs $f_h(x)$ to observed data, z, to be acceptably small, taking into account all of the uncertainties in the problem.

If $C(z)$ is empty, then that is clear evidence of problems with the model (missing physics, coding errors, etc.) or with the data.

If $C(z)$ is non-empty, then an analysis of its elements reveals the constraints on the parameter space imposed by the data.

Further the model projections $f(x) : x \in C(z)$ over future outcomes, reveal the futures consistent with the model physics and the historical data.

If the data is informative for the parameter space, then $C(z)$ will typically form a tiny percentage of the original parameter space, so that even if we do wish to calibrate the model, history matching is a useful prior step.

History matching by implausibility

We history match using an ”‘implausibility measure”’ $I(x)$ based on a probabilistic metric (eg. number of sd between z and $f_h(x)$).

History matching by implausibility

We history match using an "‘implausibility measure"’ $I(x)$ based on a probabilistic metric (eg. number of sd between z and $f_h(x)$).
For example, if we are matching a single output, then we might choose

$$I(x) = \frac{(z - \mathrm{E}(f_h(x)))^2}{\mathrm{Var}(z - \mathrm{E}(f_h(x)))}$$

where $\mathrm{Var}(z - \mathrm{E}(f_h(x)))$ is the sum of the measurement variance the structural discrepancy variance and the emulator variance.

History matching by implausibility

We history match using an ”‘implausibility measure”’ $I(x)$ based on a probabilistic metric (eg. number of sd between z and $f_h(x)$).
For example, if we are matching a single output, then we might choose

$$I(x) = \frac{(z - \mathrm{E}(f_h(x)))^2}{\mathrm{Var}(z - \mathrm{E}(f_h(x)))}$$

where $\mathrm{Var}(z - \mathrm{E}(f_h(x)))$ is the sum of the measurement variance the structural discrepancy variance and the emulator variance.
The implausibility calculation can be performed univariately, or by multivariate calculation over sub-vectors. The implausibilities are then combined, such as by using $I_M(x) = \max_i I_{(i)}(x)$, and can then be used to identify regions of x with large $I_M(x)$ as implausible, i.e. unlikely to be good choices for x^*.

History matching by implausibility

We history match using an "'implausibility measure"' $I(x)$ based on a probabilistic metric (eg. number of sd between z and $f_h(x)$).

For example, if we are matching a single output, then we might choose

$$I(x) = \frac{(z - \mathrm{E}(f_h(x)))^2}{\mathrm{Var}(z - \mathrm{E}(f_h(x)))}$$

where $\mathrm{Var}(z - \mathrm{E}(f_h(x)))$ is the sum of the measurement variance the structural discrepancy variance and the emulator variance.

The implausibility calculation can be performed univariately, or by multivariate calculation over sub-vectors. The implausibilities are then combined, such as by using $I_M(x) = \max_i I_{(i)}(x)$, and can then be used to identify regions of x with large $I_M(x)$ as implausible, i.e. unlikely to be good choices for x^*.

With this information, we can then refocus our analysis on the 'non-implausible' regions of the input space, by (i) making more simulator runs & (ii) refitting our emulator over such sub-regions and repeating the analysis.

History matching by implausibility

We history match using an "'implausibility measure"' $I(x)$ based on a probabilistic metric (eg. number of sd between z and $f_h(x)$).
For example, if we are matching a single output, then we might choose

$$I(x) = \frac{(z - \mathrm{E}(f_h(x)))^2}{\mathrm{Var}(z - \mathrm{E}(f_h(x)))}$$

where $\mathrm{Var}(z - \mathrm{E}(f_h(x)))$ is the sum of the measurement variance the structural discrepancy variance and the emulator variance.
The implausibility calculation can be performed univariately, or by multivariate calculation over sub-vectors. The implausibilities are then combined, such as by using $I_M(x) = \max_i I_{(i)}(x)$, and can then be used to identify regions of x with large $I_M(x)$ as implausible, i.e. unlikely to be good choices for x^*.
With this information, we can then refocus our analysis on the 'non-implausible' regions of the input space, by (i) making more simulator runs & (ii) refitting our emulator over such sub-regions and repeating the analysis.
This process is a form of iterative global search aimed at finding all choices of x^* which would give good fits to historical data.

Refocusing

We make the restriction $\mathcal{X}^* = \{\boldsymbol{x} : \mathcal{I}(\boldsymbol{x}) \leq 4\} \simeq \{\boldsymbol{x} : \phi < 0.79\}$ and eliminate 90% of the input space

Refocusing

We make the restriction $\mathcal{X}^* = \{\boldsymbol{x} : \mathcal{I}(\boldsymbol{x}) \leq 4\} \simeq \{\boldsymbol{x} : \phi < 0.79\}$ and eliminate 90% of the input space

Since reducing the space many of the old model runs are no longer valid, we supplement with additional evaluations (262+100 coarse runs, 6+20 accurate runs)

Refocusing

We make the restriction $\mathcal{X}^* = \{\boldsymbol{x} : \mathcal{I}(\boldsymbol{x}) \leq 4\} \simeq \{\boldsymbol{x} : \phi < 0.79\}$ and eliminate 90% of the input space

Since reducing the space many of the old model runs are no longer valid, we supplement with additional evaluations (262+100 coarse runs, 6+20 accurate runs)

We re-fit the coarse and fine emulators, using the old emulator structure as a starting point. We concentrate on emulating the final 4 time points in original data.

Refocusing

We make the restriction $\mathcal{X}^* = \{\boldsymbol{x} : \mathcal{I}(\boldsymbol{x}) \leq 4\} \simeq \{\boldsymbol{x} : \phi < 0.79\}$ and eliminate 90% of the input space

Since reducing the space many of the old model runs are no longer valid, we supplement with additional evaluations (262+100 coarse runs, 6+20 accurate runs)

We re-fit the coarse and fine emulators, using the old emulator structure as a starting point. We concentrate on emulating the final 4 time points in original data.

We introduce an additional point 1 year beyond the end of the previous series as a value to be forecast (actually, we do have real world observations of this value, to reality check our forecasts).

Forecasting

The mean and variance of $F(x)$ are obtained from the mean function and variance function of the emulator f for F.

Using these values, we compute the mean and variance of $F^* = F(x^*)$ (by first conditioning on x^* and then integrating out x^*).

Forecasting

The mean and variance of $F(x)$ are obtained from the mean function and variance function of the emulator f for F.

Using these values, we compute the mean and variance of $F^* = F(x^*)$ (by first conditioning on x^* and then integrating out x^*).

Given $\mathrm{E}(F^*)$, $\mathrm{Var}(F^*)$, and the model discrepancy, ϵ and sampling error e variances, it is now straightforward to compute the joint mean and variance of the collection (y, z) (as $y = F^* \oplus \epsilon, z = y \oplus e$).

Forecasting

The mean and variance of $F(x)$ are obtained from the mean function and variance function of the emulator f for F.

Using these values, we compute the mean and variance of $F^* = F(x^*)$ (by first conditioning on x^* and then integrating out x^*).

Given $\mathrm{E}(F^*)$, $\mathrm{Var}(F^*)$, and the model discrepancy, ϵ and sampling error e variances, it is now straightforward to compute the joint mean and variance of the collection (y, z) (as $y = F^* \oplus \epsilon, z = y \oplus e$).

We now evaluate the adjusted mean and variance for y_p adjusted by z using the Bayes linear adjustment formulae.
(When the forecast variance is large, then we have methods to improve forecast accuracy by assessing local variation near x^*.)

Forecasting

The mean and variance of $F(x)$ are obtained from the mean function and variance function of the emulator f for F.

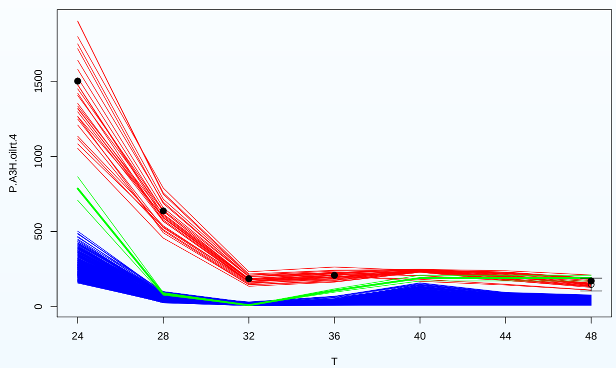
Using these values, we compute the mean and variance of $F^* = F(x^*)$ (by first conditioning on x^* and then integrating out x^*).

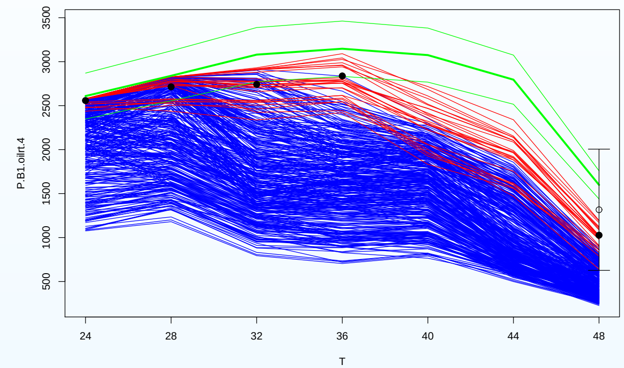
Given $\mathrm{E}(F^*)$, $\mathrm{Var}(F^*)$, and the model discrepancy, ϵ and sampling error e variances, it is now straightforward to compute the joint mean and variance of the collection (y, z) (as $y = F^* \oplus \epsilon, z = y \oplus e$).

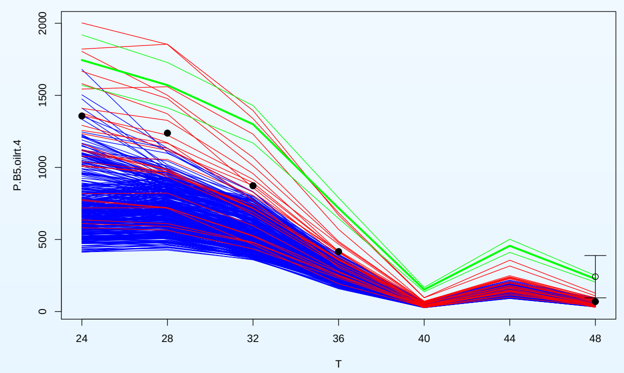
We now evaluate the adjusted mean and variance for y_p adjusted by z using the Bayes linear adjustment formulae.
(When the forecast variance is large, then we have methods to improve forecast accuracy by assessing local variation near x^*.)

This analysis is tractable even for large systems with complex descriptions of structural discrepancy, so allows us to vary decision inputs to achieve “optimal control” for the system.

Forecasting Results







Green lines indicate data z with error bounds of 2sd(e).

Red and blue lines represent the range of the runs of $F(x)$ and $F^c(x)$

Solid black dots correspond to $\mathrm{E}(F^*)$.

The system forecast is indicated by a hollow circle with attached error bars.

(Note how system forecasts correct simulator forecast towards data.)

The Galform Model (thanks to Ian Vernon)

The Cosmologists at the Durham Institute for Computational Cosmology are interested in modelling galaxy formation in the presence of Dark Matter.

First a Dark Matter simulation is performed over a volume of (1.63 billion light years$)^3$. This takes 3 months on a supercomputer.

The Galform Model (thanks to Ian Vernon)

The Cosmologists at the Durham Institute for Computational Cosmology are interested in modelling galaxy formation in the presence of Dark Matter.

First a Dark Matter simulation is performed over a volume of (1.63 billion light years$)^3$. This takes 3 months on a supercomputer.

Galform takes the results of this simulation and models the evolution and attributes of approximately 1 million galaxies.

The Galform Model (thanks to Ian Vernon)

The Cosmologists at the Durham Institute for Computational Cosmology are interested in modelling galaxy formation in the presence of Dark Matter.

First a Dark Matter simulation is performed over a volume of (1.63 billion light years$)^3$. This takes 3 months on a supercomputer.

Galform takes the results of this simulation and models the evolution and attributes of approximately 1 million galaxies.

Galform requires the specification of 17 unknown inputs in order to run.

The Galform Model (thanks to Ian Vernon)

The Cosmologists at the Durham Institute for Computational Cosmology are interested in modelling galaxy formation in the presence of Dark Matter.

First a Dark Matter simulation is performed over a volume of (1.63 billion light years$)^3$. This takes 3 months on a supercomputer.

Galform takes the results of this simulation and models the evolution and attributes of approximately 1 million galaxies.

Galform requires the specification of 17 unknown inputs in order to run.

It takes approximately 1 day to complete 1 run (using a single processor).

The Galform Model (thanks to Ian Vernon)

The Cosmologists at the Durham Institute for Computational Cosmology are interested in modelling galaxy formation in the presence of Dark Matter.

First a Dark Matter simulation is performed over a volume of (1.63 billion light years$)^3$. This takes 3 months on a supercomputer.

Galform takes the results of this simulation and models the evolution and attributes of approximately 1 million galaxies.

Galform requires the specification of 17 unknown inputs in order to run.

It takes approximately 1 day to complete 1 run (using a single processor).

The Galform model produces lots of outputs, some of which can be compared to observed data from the real universe.

The Galform Model (thanks to Ian Vernon)

The Cosmologists at the Durham Institute for Computational Cosmology are interested in modelling galaxy formation in the presence of Dark Matter.

First a Dark Matter simulation is performed over a volume of (1.63 billion light years$)^3$. This takes 3 months on a supercomputer.

Galform takes the results of this simulation and models the evolution and attributes of approximately 1 million galaxies.

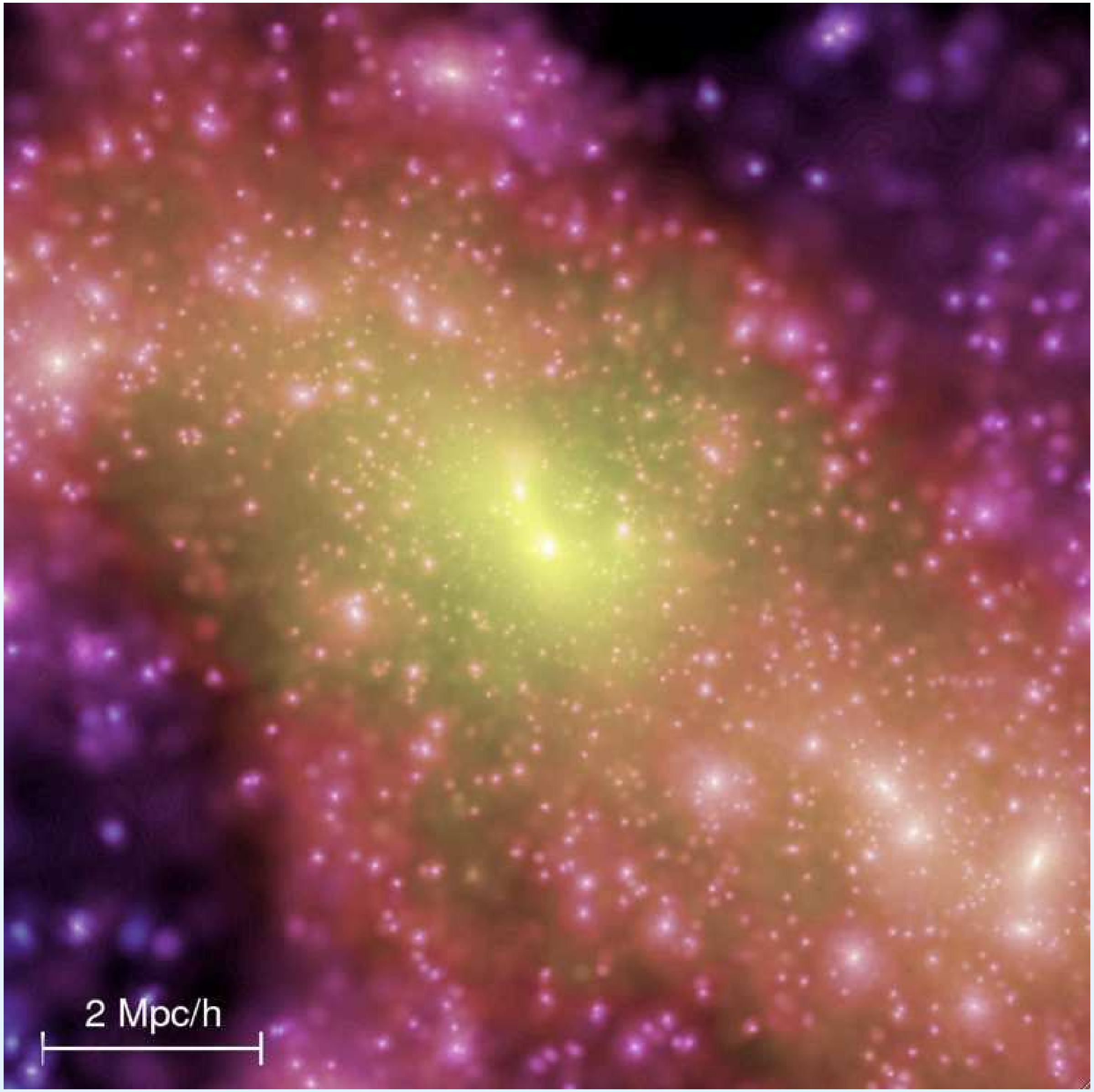
Galform requires the specification of 17 unknown inputs in order to run.

It takes approximately 1 day to complete 1 run (using a single processor).

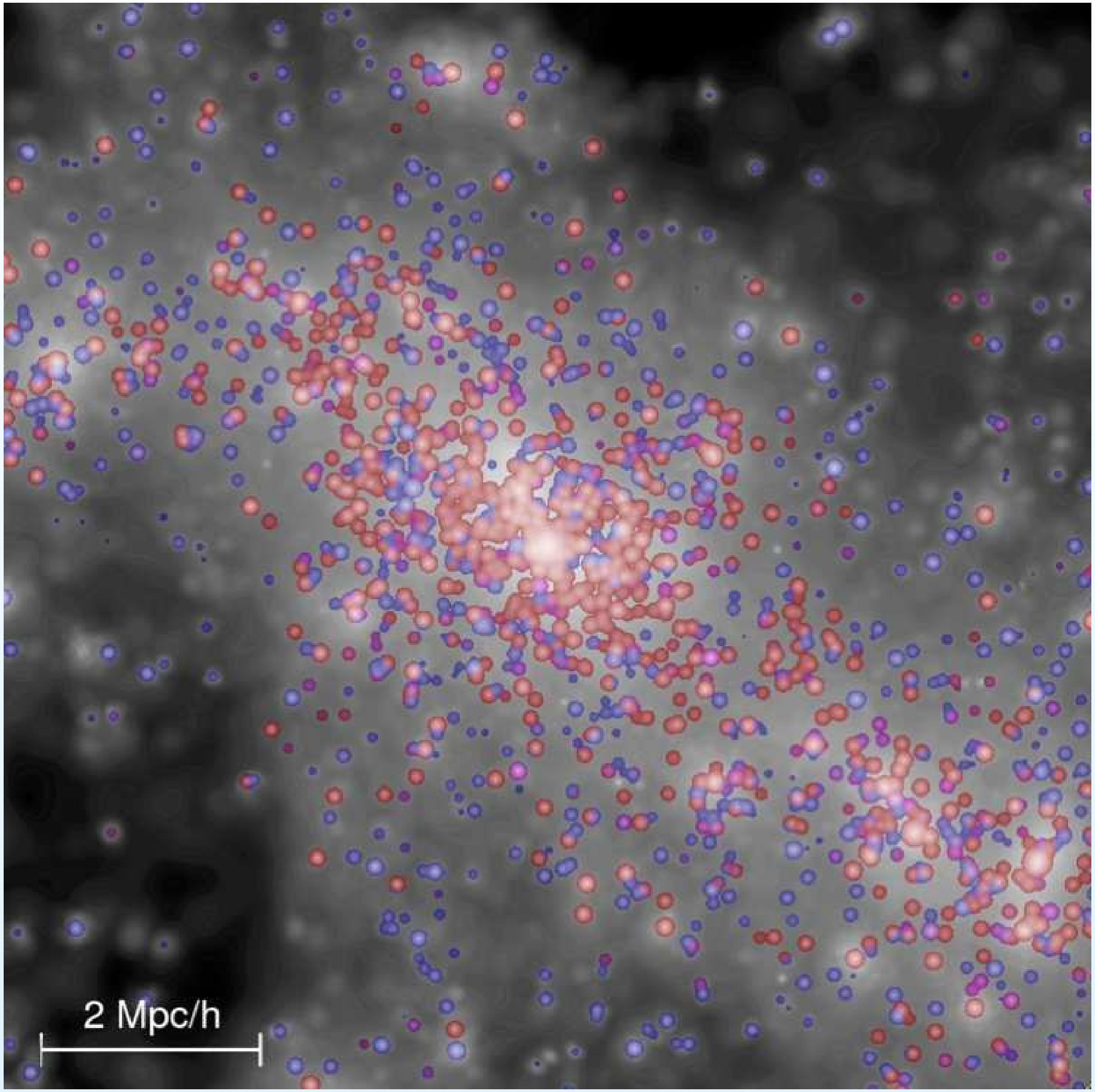
The Galform model produces lots of outputs, some of which can be compared to observed data from the real universe.

Questions of interest. Are there any choices of input parameters that will make Galform output resemble the observed universe? If so, how many different parameter choices are there?

The Dark Matter Simulation



The Galform Model



Inputs

To perform one run, we need to specify the following 17 inputs:

vhotdisk:	100 - 550	**VCUT**:	20 - 50
aReheat:	0.2 - 1.2	**ZCUT**:	6 - 9
alphacool:	0.2 - 1.2	**alphastar**:	-3.2 - -0.3
vhotburst:	100 - 550	**tau0mrg**:	0.8 - 2.7
epsilonStar:	0.001 - 0.1	**fellip**:	0.1 - 0.35
stabledisk:	0.65 - 0.95	**fburst**:	0.01 - 0.15
alphahot:	2 - 3.7	**FSMBH**:	0.001 - 0.01
yield:	0.02 - 0.05	**eSMBH**:	0.004 - 0.05
tdisk:	0 - 1		

Inputs

To perform one run, we need to specify the following 17 inputs:

vhotdisk:	100 - 550	**VCUT**:	20 - 50
aReheat:	0.2 - 1.2	**ZCUT**:	6 - 9
alphacool:	0.2 - 1.2	**alphastar**:	-3.2 - -0.3
vhotburst:	100 - 550	**tau0mrg**:	0.8 - 2.7
epsilonStar:	0.001 - 0.1	**fellip**:	0.1 - 0.35
stabledisk:	0.65 - 0.95	**fburst**:	0.01 - 0.15
alphahot:	2 - 3.7	**FSMBH**:	0.001 - 0.01
yield:	0.02 - 0.05	**eSMBH**:	0.004 - 0.05
tdisk:	0 - 1		

Galform provides multiple output data sets. Initially we analyse luminosity functions giving the number of galaxies per unit volume, for each luminosity.

Bj Luminosity: corresponds to density of young (blue) galaxies

K Luminosity: corresponds to density of old (red) galaxies

Inputs

To perform one run, we need to specify the following 17 inputs:

vhotdisk:	100 - 550	**VCUT**:	20 - 50
aReheat:	0.2 - 1.2	**ZCUT**:	6 - 9
alphacool:	0.2 - 1.2	**alphastar**:	-3.2 - -0.3
vhotburst:	100 - 550	**tau0mrg**:	0.8 - 2.7
epsilonStar:	0.001 - 0.1	**fellip**:	0.1 - 0.35
stabledisk:	0.65 - 0.95	**fburst**:	0.01 - 0.15
alphahot:	2 - 3.7	**FSMBH**:	0.001 - 0.01
yield:	0.02 - 0.05	**eSMBH**:	0.004 - 0.05
tdisk:	0 - 1		

Galform provides multiple output data sets. Initially we analyse luminosity functions giving the number of galaxies per unit volume, for each luminosity.

Bj Luminosity: corresponds to density of young (blue) galaxies

K Luminosity: corresponds to density of old (red) galaxies

We choose 11 outputs that are representative of the Luminosity functions and emulate the functions $f_i(x)$.

Summary of Results

Implausibility calculation requires adding together the emulator, measurement and structural uncertainties, all of which we assess.

Summary of Results

Implausibility calculation requires adding together the emulator, measurement and structural uncertainties, all of which we assess.

We then carry out the iterative history matching procedure, through 4 waves.

Summary of Results

Implausibility calculation requires adding together the emulator, measurement and structural uncertainties, all of which we assess.

We then carry out the iterative history matching procedure, through 4 waves.

In wave 5, we evaluate many good fits to data, and we stop. Some of these choices give simultaneous matches to data sets that the Cosmologists have been unable to match before.

Summary of Results

Implausibility calculation requires adding together the emulator, measurement and structural uncertainties, all of which we assess.

We then carry out the iterative history matching procedure, through 4 waves.

In wave 5, we evaluate many good fits to data, and we stop. Some of these choices give simultaneous matches to data sets that the Cosmologists have been unable to match before.

	No. Model Runs	No. Active Vars	Space Remaining
Wave 1	1000	5	14.9 %
Wave 2	1414	8	5.9 %
Wave 3	1620	8	1.6 %
Wave 4	2011	10	0.12 %

Linking models to reality: exchangeability

Durham University

There are many interesting features involved in relating Galform output to actual observations that we make on the universe.

Linking models to reality: exchangeability

There are many interesting features involved in relating Galform output to actual observations that we make on the universe.

One of the most important is as follows. We need to compensate for our lack of knowledge about the state of dark matter over all of time and space (which the simulator requires). This is typical of initial condition/forcing function uncertainty and is a large factor in the mismatch between model and reality.

Linking models to reality: exchangeability

There are many interesting features involved in relating Galform output to actual observations that we make on the universe.

One of the most important is as follows. We need to compensate for our lack of knowledge about the state of dark matter over all of time and space (which the simulator requires). This is typical of initial condition/forcing function uncertainty and is a large factor in the mismatch between model and reality.

What Galform provides is 520 exchangeable computer models, $f^r(x)$ (one for each description of dark matter).

Linking models to reality: exchangeability

There are many interesting features involved in relating Galform output to actual observations that we make on the universe.

One of the most important is as follows. We need to compensate for our lack of knowledge about the state of dark matter over all of time and space (which the simulator requires). This is typical of initial condition/forcing function uncertainty and is a large factor in the mismatch between model and reality.

What Galform provides is 520 exchangeable computer models, $f^r(x)$ (one for each description of dark matter).

The exchangeability representation for random functions allows us to express each function as

$f^r(x) = M(x) \oplus R^r(x)$

$M(x)$ is the mean function

$R^r(x)$ are the (uncorrelated, exchangeable, mean zero) residual functions.

Linking models to reality: exchangeability 2

$$f^r(x) = M(x) \oplus R^r(x)$$

Linking models to reality: exchangeability 2

$f^r(x) = M(x) \oplus R^r(x)$

If each function is emulated as

$f^r(x) = B^r g(x) + e^r(x)$,

then we have

$B^r = M_B + R_B^r$, $e^r(x) = M_e(x) + R_e^r(x)$

Linking models to reality: exchangeability 2

$f^r(x) = M(x) \oplus R^r(x)$

If each function is emulated as

$f^r(x) = B^r g(x) + e^r(x)$,

then we have

$B^r = M_B + R^r_B$, $e^r(x) = M_e(x) + R^r_e(x)$

If we consider dark matter in our universe to be exchangeable with the 520 individual simulations, then our emulator for Galform evaluated on the correct dark matter configuration is

$f^*(x) = (M_B + R^*_B)g(x) + M_e(x) + R^*_e(x)$

We cannot evaluate this simulator (because we don't know the appropriate dark matter configuration) but we can emulate it, based on a detailed analysis of the Galform experiments.

Linking models to reality: exchangeability and reification

This is an example of the idea of reification (from reify: to treat an abstract concept as if it were real), as follows.

The reason that the evaluations of the simulator are informative for the physical system is that the evaluations are informative about the general relationships between system properties, x, and system behaviour y.

Linking models to reality: exchangeability and reification

This is an example of the idea of reification (from reify: to treat an abstract concept as if it were real), as follows.

The reason that the evaluations of the simulator are informative for the physical system is that the evaluations are informative about the general relationships between system properties, x, and system behaviour y.

More generally, evaluations of a collection of models are jointly informative for the physical system as they are jointly informative for these general relationships.

Linking models to reality: reification

Therefore, our inference from model to reality proceeds in two parts.

[1] Emulate the relationship between system properties and system behaviour (we call this the ”‘reified model”’).

[2] Decompose the difference between model and physical system as

[A] The difference between our simulator and the reified form.

[B] The difference between the reified form at the physically appropriate choice of x and the actual system behaviour y.

Linking models to reality: reification

Therefore, our inference from model to reality proceeds in two parts.

[1] Emulate the relationship between system properties and system behaviour (we call this the ”‘reified model”’).

[2] Decompose the difference between model and physical system as

[A] The difference between our simulator and the reified form.

[B] The difference between the reified form at the physically appropriate choice of x and the actual system behaviour y.

$$\begin{tikzcd} & x^* \arrow[dr] & & & \\ f \arrow[r] & f^* \arrow[r] & f^*(x^*) \arrow[r] & y \arrow[r] & z \end{tikzcd}$$

Linking models to reality: reification

Therefore, our inference from model to reality proceeds in two parts.

[1] Emulate the relationship between system properties and system behaviour (we call this the "'reified model"').

[2] Decompose the difference between model and physical system as

[A] The difference between our simulator and the reified form.

[B] The difference between the reified form at the physically appropriate choice of x and the actual system behaviour y.

$$
\begin{tikzcd}
 & & x^* \arrow[dr] & & & \\
f \arrow[r] & f^* \arrow[rr] & & f^*(x^*) \arrow[r] & y \arrow[r] & z
\end{tikzcd}
$$

Comment Reification retains tractability in linking computer evaluations and system values with system behaviour, while (i) removing logical problems in treatment of discrepancy, (ii) providing guidance for discrepancy modelling, (iii) showing how to combine multi-model information (through f^*).

A Reified influence diagram

$$\left[F^{1}_{h:[n]}(x), \ldots, F^{m}_{h:[n]}(x)\right]$$

Evaluations of the simulator at each of m initial conditions for historical components of simulator

A Reified influence diagram

$$\left[F^1_{h:[n]}(x), \ldots, F^m_{h:[n]}(x)\right] \longrightarrow F_{h:\mathrm{suff}} \longrightarrow F^*_{h:\mathrm{suff}} \longrightarrow f^*_h(x)$$

Global information $F_{\mathrm{h:suff}}$ (from second order exchangeability modelling).
passes to Reified global form and to reified emulator.

A Reified influence diagram

$$
\begin{tikzcd}
 & & & & & & e \arrow[d] \\
\left[F^{1}_{h:[n]}(x), \ldots, F^{m}_{h:[n]}(x)\right] \arrow[r] & F_{h:\mathrm{suff}} \arrow[r] & F^{*}_{h:\mathrm{suff}} \arrow[r] & f^{*}_{h}(x) \arrow[r] & F^{*}_{h}(x^{*}) \arrow[r] & y_h \arrow[r] & z \\
 & & & & & \epsilon_h \arrow[u] & \\
 & & & & x^{*} \arrow[uu] & &
\end{tikzcd}
$$

Link with x^* to reified function, at true initial condition, linked to data z

A Reified influence diagram

$$
\begin{tikzcd}
 & & & & \text{MME} & & e \arrow[d] \\
\left[F^1_{h:[n]}(x), \ldots, F^m_{h:[n]}(x)\right] \arrow[r] & F_{h:\text{suff}} \arrow[r] & F^*_{h:\text{suff}} \arrow[r] & f^*_h(x) \arrow[r] & F^*_h(x^*) \arrow[r] \arrow[u] & y_h \arrow[r] & z \\
 & & & & & \epsilon_h \arrow[u] & \\
 & & & & x^* \arrow[uu] & &
\end{tikzcd}
$$

Add observation of a related multi-model ensemble (MME) consisting of tuned runs from related models (more exchangeability modelling).

A Reified influence diagram

$$
\begin{tikzcd}
 & & & & \mathrm{MME} & & e \arrow[d] \\
\left[F^1_{h:[n]}(x), \dots, F^m_{h:[n]}(x)\right] \arrow[r] & F_{h:\mathrm{suff}} \arrow[r] & F^*_{h:\mathrm{suff}} \arrow[r] & f^*_h(x) \arrow[r] & F^*_h(x^*) \arrow[r] \arrow[u] & y_h \arrow[r] & z \\
 & & & & & \epsilon_h \arrow[u] & \\
 & & & & x^* \arrow[uu] & & \\
F'_{h:[n]}(x) \arrow[r] & F'_{h:\mathrm{suff}} \arrow[uuu] & & & & &
\end{tikzcd}
$$

Add a set of evaluations from a fast approximation

A Reified influence diagram

$$\begin{tikzcd}
 & & & & \mathrm{MME} & & e \arrow[d] \\
\left[F^1_{h:[n]}(x), \ldots, F^m_{h:[n]}(x)\right] \arrow[r] & F_{h:\mathrm{suff}} \arrow[r] & F^*_{h:\mathrm{suff}} \arrow[r] & f^*_h(x) \arrow[r] & F^*_h(x^*) \arrow[u] \arrow[r] & y_h \arrow[r] & z \\
 & & & & & \epsilon_h \arrow[u] & \\
 & & & & x^* \arrow[uu] & & \\
F'_{h:[n]}(x) \arrow[r] & F'_{h:\mathrm{suff}} \arrow[uuu] & & & & & \\
F'_{p:[n]}(x, d) \arrow[r] & F'_{p:\mathrm{suff}} \arrow[u] & & & & &
\end{tikzcd}$$

Add evaluations of fast simulator for outcomes to be predicted, with decision choices d

A Reified influence diagram

$$
\begin{tikzcd}
 & & & & \mathrm{MME} & & e \arrow[d] \\
\left[F^1_{h:[n]}(x), \ldots, F^m_{h:[n]}(x)\right] \arrow[r] & F_{h:\mathrm{suff}} \arrow[r] & F^*_{h:\mathrm{suff}} \arrow[r] \arrow[dddd] & f^*_h(x) \arrow[r] & F^*_h(x^*) \arrow[r] \arrow[u] & y_h \arrow[r] & z \\
 & & & & & \epsilon_h \arrow[u] & \\
 & & & & x^* \arrow[uu] & & \\
F'_{h:[n]}(x) \arrow[r] & F'_{h:\mathrm{suff}} \arrow[uuu] & & & & & \\
F'_{p:[n]}(x, d) \arrow[r] & F'_{p:\mathrm{suff}} \arrow[r] \arrow[u] & F^*_{p:\mathrm{suff}} & & & &
\end{tikzcd}
$$

Link to reified global terms for quantities to be predicted

A Reified influence diagram

$$
\begin{tikzcd}
 & & & & \mathrm{MME} & & e \arrow[d] \\
\left[F^1_{h:[n]}(x), \ldots, F^m_{h:[n]}(x)\right] \arrow[r] & F_{h:\mathrm{suff}} \arrow[r] & F^*_{h:\mathrm{suff}} \arrow[r] \arrow[dddd] & f^*_h(x) \arrow[r] \arrow[dddd] & F^*_h(x^*) \arrow[r] \arrow[u] & y_h \arrow[r] & z \\
 & & & & & \epsilon_h \arrow[u] & \\
 & & & & x^* \arrow[uu] & & \\
F'_{h:[n]}(x) \arrow[r] & F'_{h:\mathrm{suff}} \arrow[uuu] & & & & & \\
F'_{p:[n]}(x,d) \arrow[r] & F'_{p:\mathrm{suff}} \arrow[r] \arrow[u] & F^*_{p:\mathrm{suff}} \arrow[r] & f^*_p(x,d) & & &
\end{tikzcd}
$$

And to reified global emulator, based on inputs and decisions

A Reified influence diagram

$$
\begin{tikzcd}
 & & & & \mathrm{MME} & & e \arrow[d] \\
\left[F^1_{h:[n]}(x), \ldots, F^m_{h:[n]}(x)\right] \arrow[r] & F_{h:\mathrm{suff}} \arrow[r] & F^*_{h:\mathrm{suff}} \arrow[r] \arrow[dddd] & f^*_h(x) \arrow[r] \arrow[dddd] & F^*_h(x^*) \arrow[r] \arrow[u] & y_h \arrow[r] & z \\
 & & & & & \epsilon_h \arrow[u] \arrow[dd] & \\
 & & & & x^* \arrow[uu] \arrow[dd] & & \\
F'_{h:[n]}(x) \arrow[r] & F'_{h:\mathrm{suff}} \arrow[uuu] & & & & \epsilon_p \arrow[d] & \\
F'_{p:[n]}(x,d) \arrow[r] & F'_{p:\mathrm{suff}} \arrow[r] \arrow[u] & F^*_{p:\mathrm{suff}} \arrow[r] & f^*_p(x,d) \arrow[r] & F^*_p(x^*,d^*) \arrow[r] & y_p \arrow[d] & \\
 & & & & \boxed{d^*} \arrow[u] & C &
\end{tikzcd}
$$

And link, through true future values y_p, to the overall utility cost C of making decision choice d^* [Attach more models to diagram at $F^*(x^*)$]

Concluding comments

We have discussed many aspects of uncertainty in climate analyses. In particular:

Concluding comments

We have discussed many aspects of uncertainty in climate analyses. In particular:

Multi-output, multi-level, design and emulation

Concluding comments

We have discussed many aspects of uncertainty in climate analyses. In particular:

Multi-output, multi-level, design and emulation

Reified analysis relating (collections of) models to the world

Concluding comments

We have discussed many aspects of uncertainty in climate analyses. In particular:

Multi-output, multi-level, design and emulation

Reified analysis relating (collections of) models to the world

History matching and system control by Bayes (linear) projection

Concluding comments

We have discussed many aspects of uncertainty in climate analyses. In particular:

Multi-output, multi-level, design and emulation

Reified analysis relating (collections of) models to the world

History matching and system control by Bayes (linear) projection

Beyond this, the fundamental issue that we should be addressing is how does all of the information from all of our climate models (and everything else) become synthesised into reasoned uncertainty judgements about the future of our planet?

Concluding comments

We have discussed many aspects of uncertainty in climate analyses. In particular:

Multi-output, multi-level, design and emulation

Reified analysis relating (collections of) models to the world

History matching and system control by Bayes (linear) projection

Beyond this, the fundamental issue that we should be addressing is how does all of the information from all of our climate models (and everything else) become synthesised into reasoned uncertainty judgements about the future of our planet?

This issue is partly technical

(but we could address that).

Concluding comments

We have discussed many aspects of uncertainty in climate analyses. In particular:

Multi-output, multi-level, design and emulation

Reified analysis relating (collections of) models to the world

History matching and system control by Bayes (linear) projection

Beyond this, the fundamental issue that we should be addressing is

how does all of the information from all of our climate models (and everything else) become synthesised into reasoned uncertainty judgements about the future of our planet?

This issue is partly technical

(but we could address that).

The main problems are cultural - the lack of people, resources, and will to recognise the urgency and importance of the task

(but we could change that).

References for applications

Reservoir Application

P.S. Craig, M. Goldstein, A.H. Seheult, J.A. Smith (1997). Pressure matching for hydocarbon reservoirs: a case study in the use of Bayes linear strategies for large computer experiments (with discussion), in Case Studies in Bayesian Statistics, vol. III, eds. C. Gastonis et al. 37-93. Springer-Verlag.

J. Cumming, M. Goldstein Bayes Linear Uncertainty Analysis for Oil Reservoirs Based on Multiscale Computer Experiments (2009), in the Handbook of Applied Bayesian Analysis,eds A. O'Hagan, M. West, OUP

Galaxy Application

I. Vernon, M. Goldstein, R. Bower (2010), "*Galaxy Formation: a Bayesian Uncertainty Analysis* (with discussion) ", *Bayesian Analysis*, 5(4): 619–670.

Climate Application

D. Williamson, M. Goldstein, A.Blaker (2012) Fast Linked Analyses for Scenario based Hierarchies, JRSSC, 61, 665-691.

General references

M. Goldstein Subjective Bayesian analysis: principles and practice (2006) Bayesian Analysis, 1, 403-420 (and 'Rejoinder to discussion': 465-472)

M. Goldstein and J.C.Rougier (2009). Reified Bayesian modelling and inference for physical systems (with discussion), JSPI, 139, 1221-1239

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

M.C. Kennedy, A. O'Hagan (2001). Bayesian calibration of computer models (with discussion). Journal of the Royal Statistical Society, B,63, 425-464

T. Santner, B. Williams, W. Notz (2003). The Design and Analysis of Computer Experiments. Springer Verlag: New York.

.