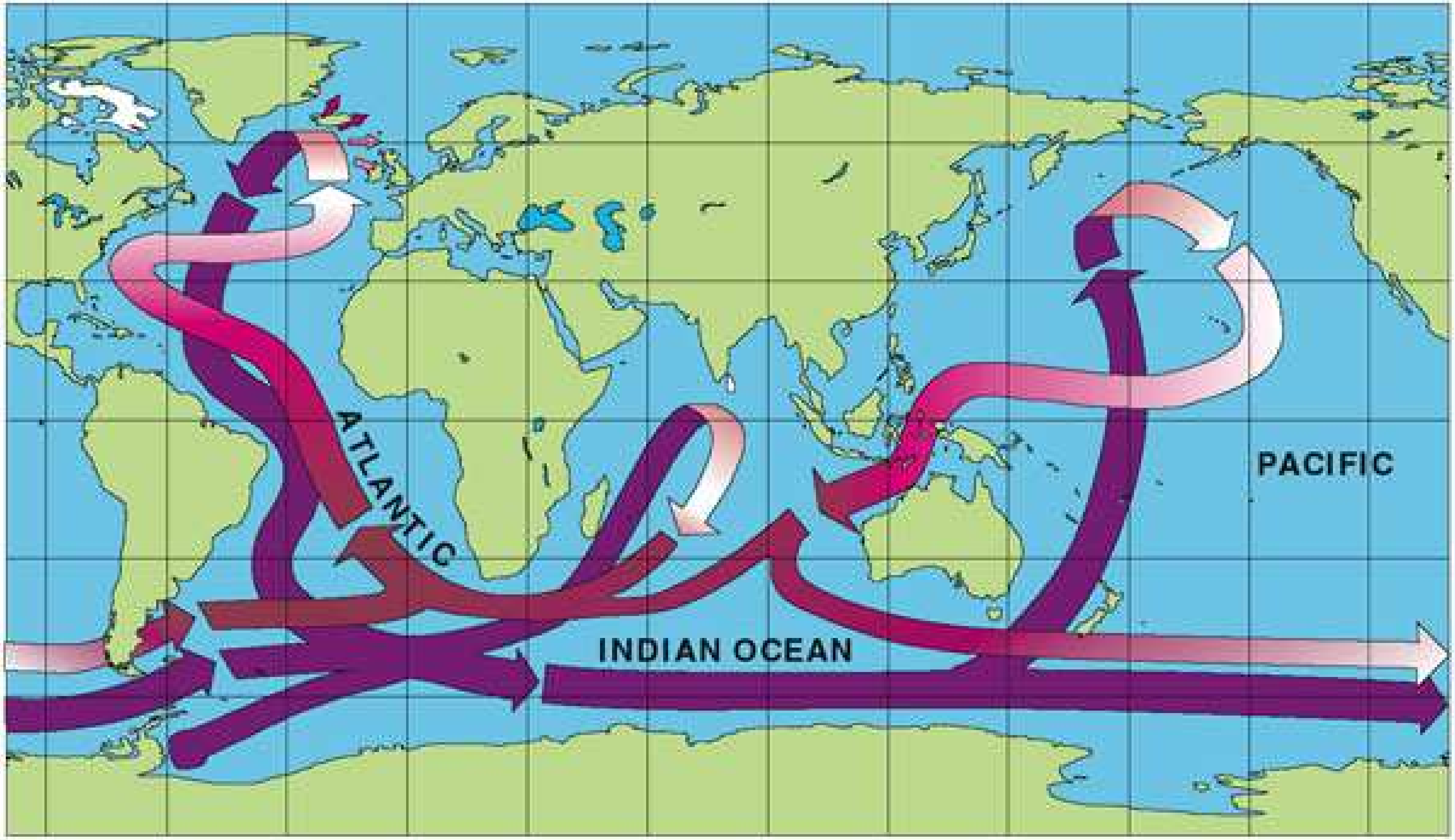
Bayes linear graphical models and computer simulators for complex physical system

Michael Goldstein
Dept. Mathematical Sciences,
Durham University *

* Graphical design: Jonathan Cumming

Global circulation



Thermohaline shutdown

Many atmosphere-ocean models show a slowdown of thermohaline circulation in simulations of the 21st century with the expected rise in greenhouse gases. [This is due to a combination of effects which reduce the density of surface waters, which makes it harder for them to sink.]

Thermohaline shutdown

Many atmosphere-ocean models show a slowdown of thermohaline circulation in simulations of the 21st century with the expected rise in greenhouse gases. [This is due to a combination of effects which reduce the density of surface waters, which makes it harder for them to sink.]

“ The weight of evidence makes it clear that climate change is a real and present danger ... Yet those who think climate change just means Indian summers in Manchester should be told that the chances of the Gulf stream - the Atlantic thermohaline circulation that keeps Britain warm - shutting down are now thought to be greater than 50%.”

[**Burying carbon** Leader Column Thursday February 3, 2005 The Guardian]

Thermohaline shutdown

Many atmosphere-ocean models show a slowdown of thermohaline circulation in simulations of the 21st century with the expected rise in greenhouse gases. [This is due to a combination of effects which reduce the density of surface waters, which makes it harder for them to sink.]

“ The weight of evidence makes it clear that climate change is a real and present danger ... Yet those who think climate change just means Indian summers in Manchester should be told that the chances of the Gulf stream - the Atlantic thermohaline circulation that keeps Britain warm - shutting down are now thought to be greater than 50%.”

[**Burying carbon** Leader Column Thursday February 3, 2005 The Guardian]

QUESTIONS

What does this statement mean?

Thermohaline shutdown

Many atmosphere-ocean models show a slowdown of thermohaline circulation in simulations of the 21st century with the expected rise in greenhouse gases. [This is due to a combination of effects which reduce the density of surface waters, which makes it harder for them to sink.]

“ The weight of evidence makes it clear that climate change is a real and present danger ... Yet those who think climate change just means Indian summers in Manchester should be told that the chances of the Gulf stream - the Atlantic thermohaline circulation that keeps Britain warm - shutting down are now thought to be greater than 50%.”

[**Burying carbon** Leader Column Thursday February 3, 2005 The Guardian]

QUESTIONS

What does this statement mean?

What analysis could possibly be done to justify (or contradict) this conclusion?

The state of the art in climate modelling

Large climate models take months to run on supercomputers. One of the biggest computers in the world is the Earth Simulator in Japan, which is often used for running climate models.

Leading climate models

One leading climate model at the moment is HadCM3, based at the UK Met Office. One component of this model is HadAM3, the atmospheric module. In a simple experiment to study the effect of CO_2-doubling (Murphy et al, 2004, Nature), this is coupled with simple mixed-layer ocean sea-ice models.

Leading climate models

One leading climate model at the moment is HadCM3, based at the UK Met Office. One component of this model is HadAM3, the atmospheric module. In a simple experiment to study the effect of CO2-doubling (Murphy et al, 2004, Nature), this is coupled with simple mixed-layer ocean sea-ice models.
The climate model (HadSM3) has about 100 uncertain parameters, including:

1. *Large scale cloud.* Six parameters
2. *Convection.* Six parameters
3. *Sea ice.* Two parameters
4. *Radiation.* Four parameters
5. *Dynamics.* Four parameters
6. *Land surface.* Four parameters
7. *Boundary layer.* Four parameters

Leading climate models

One leading climate model at the moment is HadCM3, based at the UK Met Office. One component of this model is HadAM3, the atmospheric module. In a simple experiment to study the effect of CO2-doubling (Murphy et al, 2004, Nature), this is coupled with simple mixed-layer ocean sea-ice models.
The climate model (HadSM3) has about 100 uncertain parameters, including:

1. *Large scale cloud.* Six parameters
2. *Convection.* Six parameters
3. *Sea ice.* Two parameters
4. *Radiation.* Four parameters
5. *Dynamics.* Four parameters
6. *Land surface.* Four parameters
7. *Boundary layer.* Four parameters

We have a few hundred evaluations of HadSM3, made over about three years. These evaluations are a central resource for the UK Climate Impacts Programme 2008 (UKCIP08), intended as a fairly definitive view about how climate change will impact the UK, including climate uncertainty statements.

How to relate a model and a physical systems?

- Basic ingredients:

How to relate a model and a physical systems?

- Basic ingredients:
  x^*: system properties (unknown)

How to relate a model and a physical systems?

- Basic ingredients:
  x^*: system properties (unknown)
  y: system behaviour (influenced by x^*)

How to relate a model and a physical systems?

- Basic ingredients:
  x^*: system properties (unknown)
  y: system behaviour (influenced by x^*)
  z: partial observation of y (with error)

How to relate a model and a physical systems?

- Basic ingredients:

  x^*: system properties (unknown)

  y: system behaviour (influenced by x^*)

  z: partial observation of y (with error)
- Ideally, we would like to construct a deterministic model F, embodying the **laws of nature**, which satisfies

$$y = F(x^*)$$

How to relate a model and a physical systems?

- Basic ingredients:
  x^*: system properties (unknown)
  y: system behaviour (influenced by x^*)
  z: partial observation of y (with error)
- Ideally, we would like to construct a deterministic model F, embodying the **laws of nature**, which satisfies

$$y = F(x^*)$$

- In practice, however, the our actual model F is inadequate:
  - F simplifies the physics;
  - F approximates the solution of the physical equations

How to relate a model and a physical systems?

- Basic ingredients:
  x^*: system properties (unknown)
  y: system behaviour (influenced by x^*)
  z: partial observation of y (with error)
- Ideally, we would like to construct a deterministic model F, embodying the **laws of nature**, which satisfies

$$y = F(x^*)$$

- In practice, however, the our actual model F is inadequate:
  - F simplifies the physics;
  - F approximates the solution of the physical equations
- This raises the basic question as to what does the imperfect F tell us about the system values (x^*, y)?

How to relate a model and a physical systems?

- Basic ingredients:
  x^*: system properties (unknown)
  y: system behaviour (influenced by x^*)
  z: partial observation of y (with error)
- Ideally, we would like to construct a deterministic model F, embodying the **laws of nature**, which satisfies

$$y = F(x^*)$$

- In practice, however, the our actual model F is inadequate:
  - F simplifies the physics;
  - F approximates the solution of the physical equations
- This raises the basic question as to what does the imperfect F tell us about the system values (x^*, y)?
- How about several different (imperfect) models for (x^*, y)?

How to relate a model and a physical systems?

- Basic ingredients:
  x^*: system properties (unknown)
  y: system behaviour (influenced by x^*)
  z: partial observation of y (with error)
- Ideally, we would like to construct a deterministic model F, embodying the **laws of nature**, which satisfies

$$y = F(x^*)$$

- In practice, however, the our actual model F is inadequate:
  - F simplifies the physics;
  - F approximates the solution of the physical equations
- This raises the basic question as to what does the imperfect F tell us about the system values (x^*, y)?
- How about several different (imperfect) models for (x^*, y)?
- In particular, input and output very high dimensional and evaluating $F(x)$ for any x may be VERY expensive.

Relating the model and the system

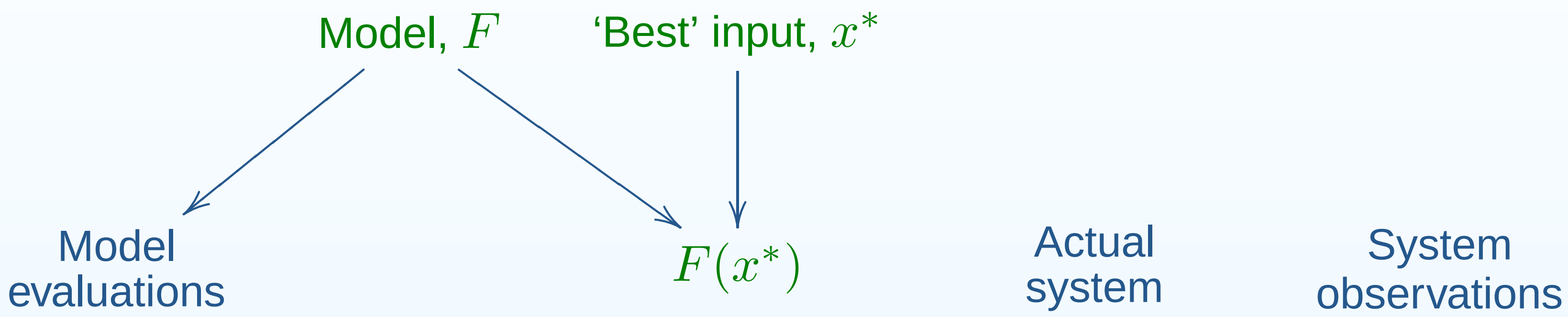
Model evaluations

Actual system

System observations

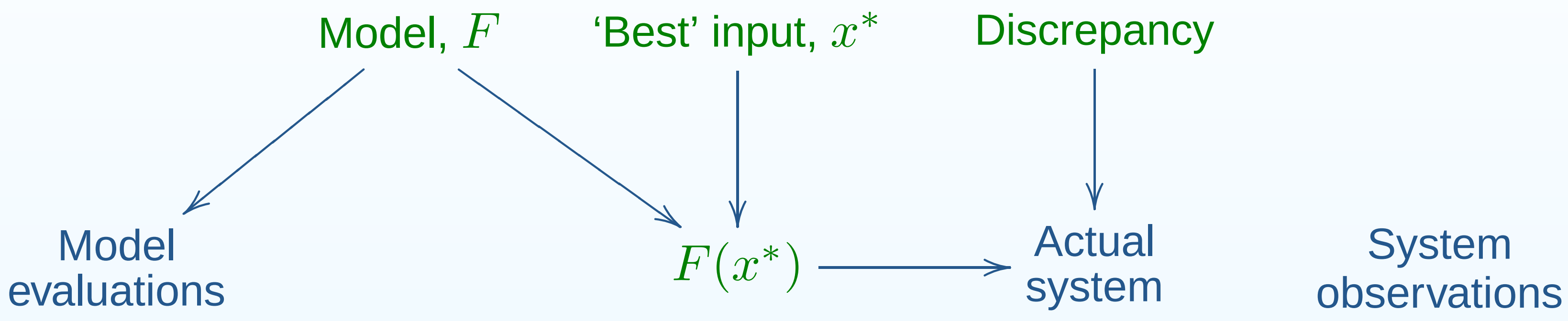
1. We start with a collection of model evaluations, and some system observations
2. We link the evaluations to the notion of a ‘best’ evaluation
3. We link the ‘best’ evaluation to the actual system
4. We incorporate measurement error into the observations
5. Our aim is to develop a unified Bayesian treatment of all these sources of uncertainty, within a natural graphical framework.

Relating the model and the system



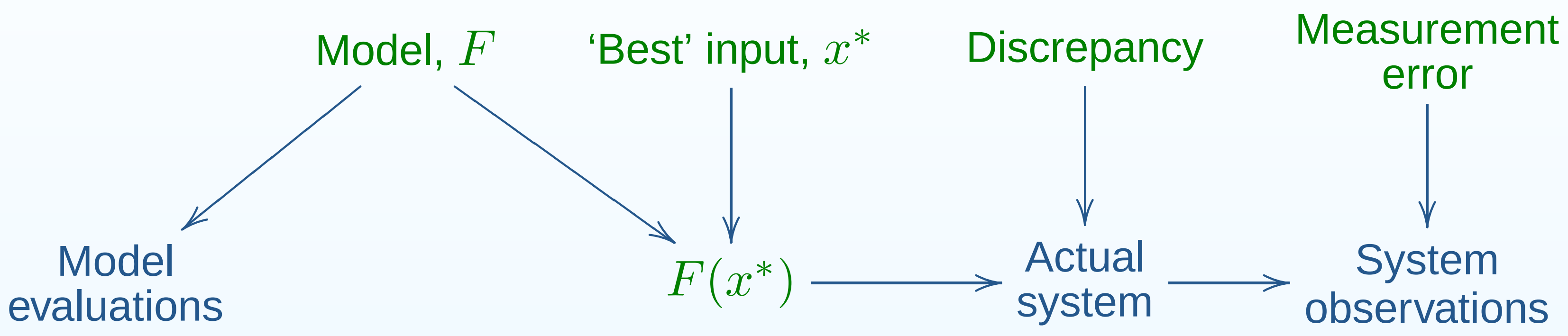
1. We start with a collection of model evaluations, and some system observations
2. We link the evaluations to the notion of a 'best' evaluation
3. We link the 'best' evaluation to the actual system
4. We incorporate measurement error into the observations
5. Our aim is to develop a unified Bayesian treatment of all these sources of uncertainty, within a natural graphical framework.

Relating the model and the system



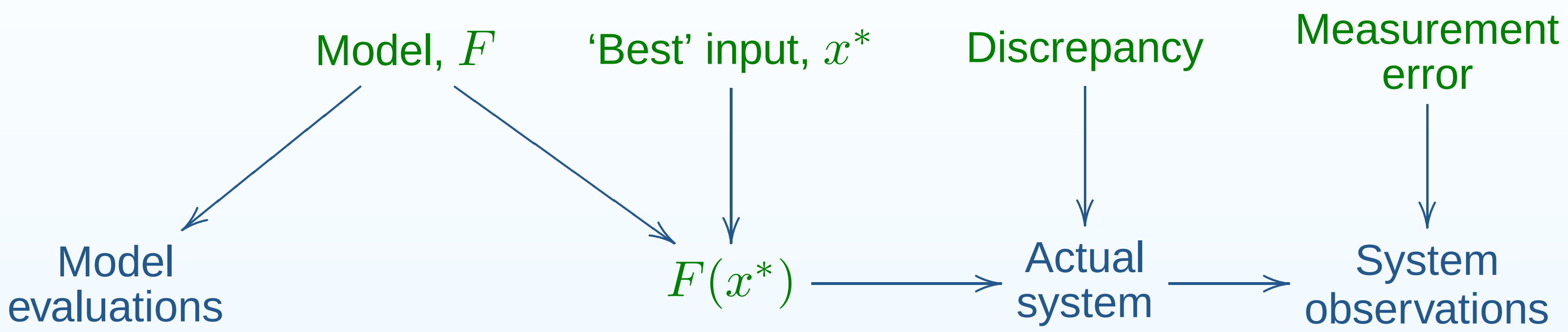
1. We start with a collection of model evaluations, and some system observations
2. We link the evaluations to the notion of a 'best' evaluation
3. We link the 'best' evaluation to the actual system
4. We incorporate measurement error into the observations
5. Our aim is to develop a unified Bayesian treatment of all these sources of uncertainty, within a natural graphical framework.

Relating the model and the system



1. We start with a collection of model evaluations, and some system observations
2. We link the evaluations to the notion of a 'best' evaluation
3. We link the 'best' evaluation to the actual system
4. We incorporate measurement error into the observations
5. Our aim is to develop a unified Bayesian treatment of all these sources of uncertainty, within a natural graphical framework.

Relating the model and the system



1. We start with a collection of model evaluations, and some system observations
2. We link the evaluations to the notion of a 'best' evaluation
3. We link the 'best' evaluation to the actual system
4. We incorporate measurement error into the observations
5. Our aim is to develop a unified Bayesian treatment of all these sources of uncertainty, within a natural graphical framework.

Representing beliefs about F using emulators

An *emulator* is a probabilistic belief specification for a deterministic function. Our emulator for component i of F might be

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

where $B = \{\beta_{ij}\}$ are unknown scalars, g_{ij} are known deterministic functions of x, and $u(x)$ is a weakly stationary stochastic process.

Representing beliefs about F using emulators

An *emulator* is a probabilistic belief specification for a deterministic function. Our emulator for component i of F might be

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

where $B = \{\beta_{ij}\}$ are unknown scalars, g_{ij} are known deterministic functions of x, and $u(x)$ is a weakly stationary stochastic process.
We fit the emulator, $f = Bg(x) + u(x)$, given a collection of model evaluations, using our favourite statistical tools - generalised least squares, maximum likelihood, Bayes - with a generous helping of expert judgement.

Representing beliefs about F using emulators

An *emulator* is a probabilistic belief specification for a deterministic function. Our emulator for component i of F might be

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

where $B = \{\beta_{ij}\}$ are unknown scalars, g_{ij} are known deterministic functions of x, and $u(x)$ is a weakly stationary stochastic process.

We fit the emulator, $f = Bg(x) + u(x)$, given a collection of model evaluations, using our favourite statistical tools - generalised least squares, maximum likelihood, Bayes - with a generous helping of expert judgement.

$Bg(x)$ represents **global** variation and $u(x)$ represents **local** variation in F

Representing beliefs about F using emulators

An *emulator* is a probabilistic belief specification for a deterministic function. Our emulator for component i of F might be

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

where $B = \{\beta_{ij}\}$ are unknown scalars, g_{ij} are known deterministic functions of x, and $u(x)$ is a weakly stationary stochastic process.

We fit the emulator, $f = Bg(x) + u(x)$, given a collection of model evaluations, using our favourite statistical tools - generalised least squares, maximum likelihood, Bayes - with a generous helping of expert judgement.

$Bg(x)$ represents **global** variation and $u(x)$ represents **local** variation in F

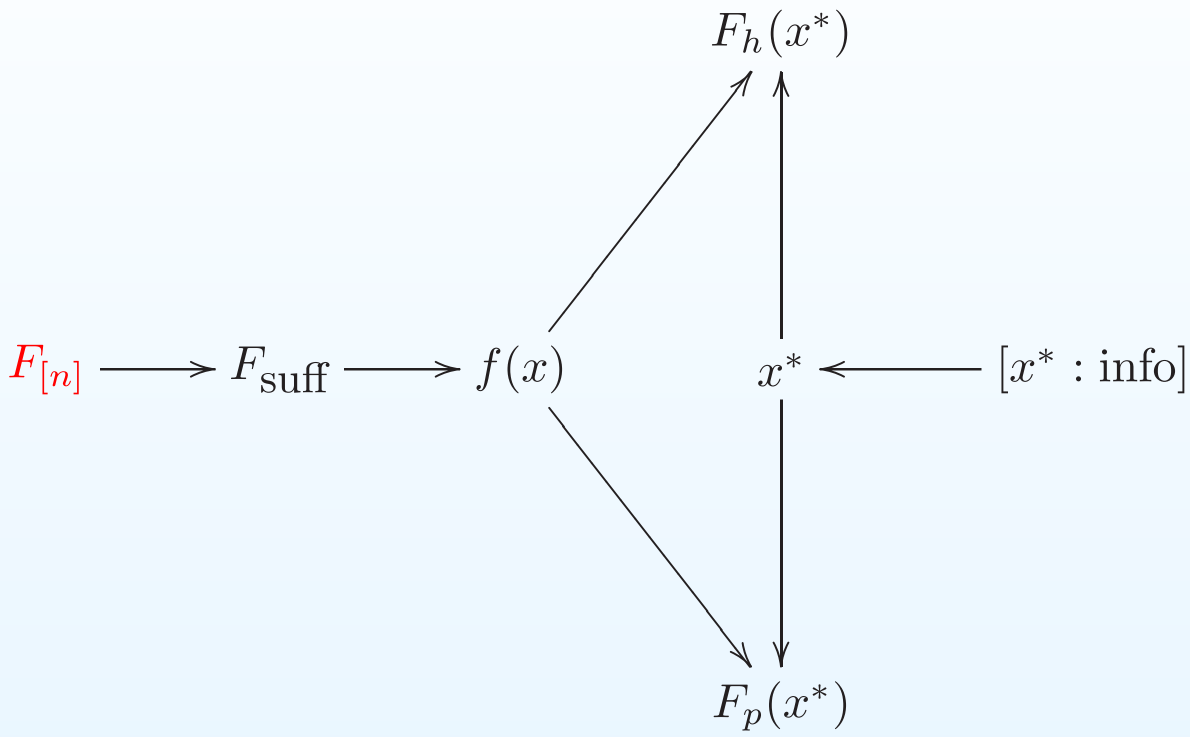
When the input dimension is high, relative to the number of function evaluations we can make, then most of what we may learn about the function comes through the global component. For simplicity, we therefore often suppose that our simulator judgements can be summarised by the global behaviour (as we don't learn much about local behaviour).

Function evaluations and emulator

$$F_{[n]} \longrightarrow F_{\text{suff}} \longrightarrow f(x)$$

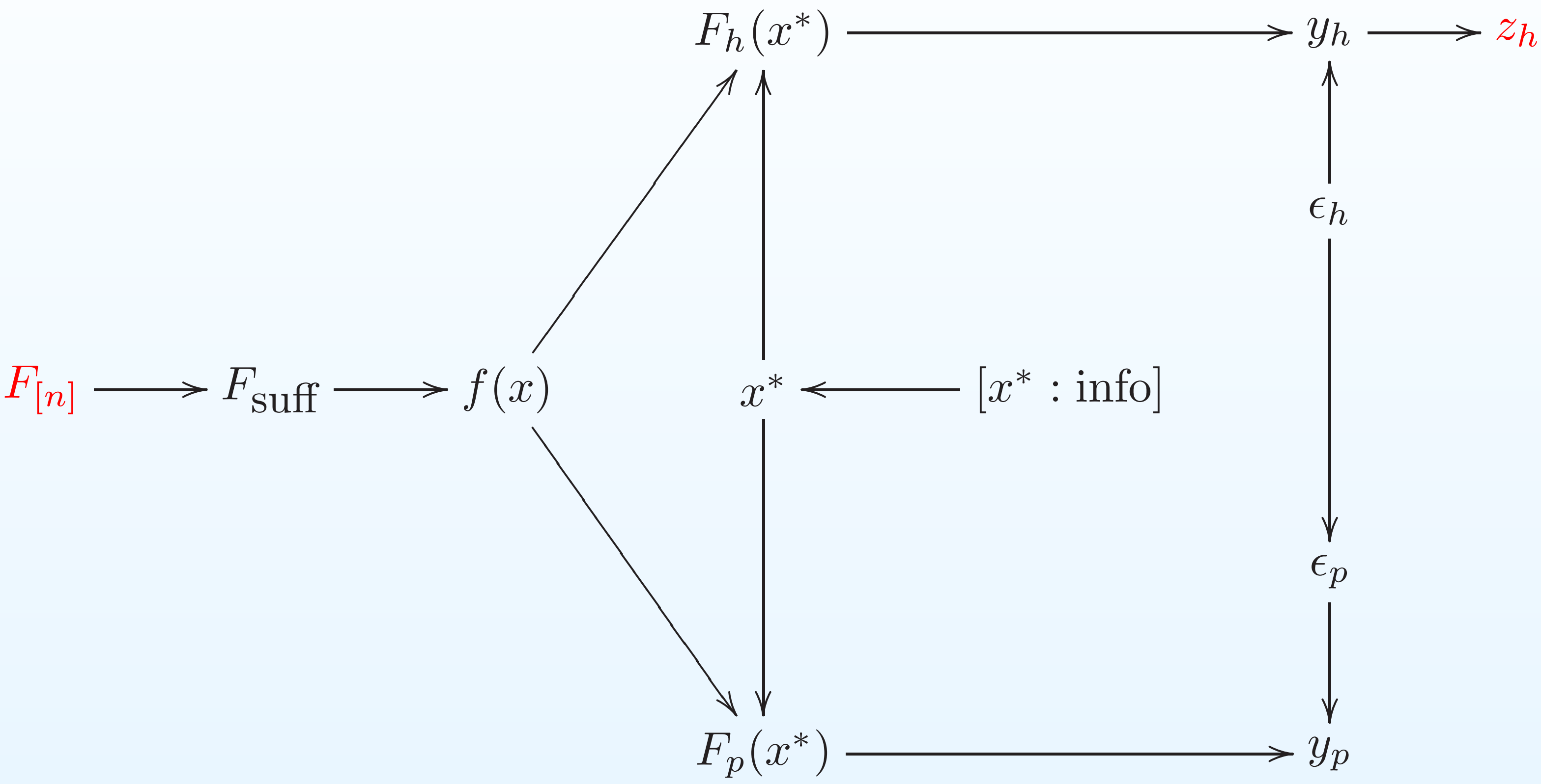
$F_{[n]} = (F(x_1), F(x_2), \ldots)$: evaluations of F at inputs $x_1, x_2, \ldots$
F_{suff}: the global information from $F_{[n]}$ which forms emulator $f(x)$

Emulator and best evaluation



True system properties x^* with emulator $f(x)$ influence beliefs for
$F_h(x^*)$: components of F corresponding to **h**istorical outputs of F
$F_p(x^*)$: components of F corresponding to outputs of F to **p**redict

Best evaluation and system



$F_h(x^*)$ is informative for historical system values y_h observed with error as z_h
$F_p(x^*)$ is informative for system values y_p to predict.
ϵ_h, ϵ_p: the corresponding discrepancy terms between model and system

Bayes linear approach

For large scale problems a full Bayes analysis is very hard because
(i) it is difficult to give a meaningful full prior probability specification over high dimensional spaces;
(ii) the computations, for learning from data (observations and computer runs) and choosing informative runs, may be technically difficult;
(iii) the likelihood surface is extremely complicated, and any full Bayes calculation may be extremely non-robust.

Bayes linear approach

For large scale problems a full Bayes analysis is very hard because
(i) it is difficult to give a meaningful full prior probability specification over high dimensional spaces;
(ii) the computations, for learning from data (observations and computer runs) and choosing informative runs, may be technically difficult;
(iii) the likelihood surface is extremely complicated, and any full Bayes calculation may be extremely non-robust.
However, the idea of the Bayesian approach, namely capturing our expert prior judgements in stochastic form and modifying them by appropriate rules given observations, is conceptually appropriate (and there is no obvious alternative).

Bayes linear approach

For large scale problems a full Bayes analysis is very hard because
(i) it is difficult to give a meaningful full prior probability specification over high dimensional spaces;
(ii) the computations, for learning from data (observations and computer runs) and choosing informative runs, may be technically difficult;
(iii) the likelihood surface is extremely complicated, and any full Bayes calculation may be extremely non-robust.

However, the idea of the Bayesian approach, namely capturing our expert prior judgements in stochastic form and modifying them by appropriate rules given observations, is conceptually appropriate (and there is no obvious alternative).

The Bayes Linear approach is (relatively) simple in terms of belief specification and analysis, as it is based only on the mean, variance and covariance specification which, following de Finetti, we take as primitive.

For a full account, see
Michael Goldstein and David Wooff (2007) Bayes Linear Statistics: Theory and Methods, Wiley.

Geometric description

For any collection $\boldsymbol{C} = (C_1, C_2, ...)$ of random quantities, we denote by $\langle \boldsymbol{C} \rangle$ the collection of (finite) linear combinations $\sum_i r_i C_i$ of the elements of $\boldsymbol{C}$. We view $\langle \boldsymbol{C} \rangle$ as a vector space.

Geometric description

For any collection $\boldsymbol{C} = (C_1, C_2, ...)$ of random quantities, we denote by $\langle \boldsymbol{C} \rangle$ the collection of (finite) linear combinations $\sum_i r_i C_i$ of the elements of $\boldsymbol{C}$. We view $\langle \boldsymbol{C} \rangle$ as a vector space.

Prior covariance is an inner product on $\langle \boldsymbol{C} \rangle$. If $\boldsymbol{C}$ is the union of the elements of the vectors $\boldsymbol{B}$ and $\boldsymbol{D},$ then the adjusted expectation of $Y \in \langle \boldsymbol{B} \rangle$ given $\boldsymbol{D}$, $\mathrm{E}_{\boldsymbol{D}}(X)$, is the orthogonal projection of Y into the linear subspace $\langle \boldsymbol{D} \rangle$, and adjusted variance, $\mathrm{Var}_{\boldsymbol{D}}(X)$, is the squared distance between Y and $\langle \boldsymbol{D} \rangle$.

Geometric description

For any collection $\boldsymbol{C} = (C_1, C_2, ...)$ of random quantities, we denote by $\langle \boldsymbol{C} \rangle$ the collection of (finite) linear combinations $\sum_i r_i C_i$ of the elements of $\boldsymbol{C}$. We view $\langle \boldsymbol{C} \rangle$ as a vector space.

Prior covariance is an inner product on $\langle \boldsymbol{C} \rangle$. If $\boldsymbol{C}$ is the union of the elements of the vectors $\boldsymbol{B}$ and $\boldsymbol{D}$, then the adjusted expectation of $Y \in \langle \boldsymbol{B} \rangle$ given $\boldsymbol{D}$, $\mathrm{E}_{\boldsymbol{D}}(X)$, is the orthogonal projection of Y into the linear subspace $\langle \boldsymbol{D} \rangle$, and adjusted variance, $\mathrm{Var}_{\boldsymbol{D}}(X)$, is the squared distance between Y and $\langle \boldsymbol{D} \rangle$.

In the usual Bayes formalism, $\langle C \rangle$ is the collection of all random variables defined over the outcome space. The inner product space is the Hilbert space of square integrable functions over the outcome space, with respect to the prior measure over the outcomes, with the covariance inner product.

Geometric description

For any collection $\boldsymbol{C} = (C_1, C_2, ...)$ of random quantities, we denote by $\langle \boldsymbol{C} \rangle$ the collection of (finite) linear combinations $\sum_i r_i C_i$ of the elements of $\boldsymbol{C}$. We view $\langle \boldsymbol{C} \rangle$ as a vector space.

Prior covariance is an inner product on $\langle \boldsymbol{C} \rangle$. If $\boldsymbol{C}$ is the union of the elements of the vectors $\boldsymbol{B}$ and $\boldsymbol{D}$, then the adjusted expectation of $Y \in \langle \boldsymbol{B} \rangle$ given $\boldsymbol{D}$, $\mathrm{E}_{\boldsymbol{D}}(X)$, is the orthogonal projection of Y into the linear subspace $\langle \boldsymbol{D} \rangle$, and adjusted variance, $\mathrm{Var}_{\boldsymbol{D}}(X)$, is the squared distance between Y and $\langle \boldsymbol{D} \rangle$.

In the usual Bayes formalism, $\langle C \rangle$ is the collection of all random variables defined over the outcome space. The inner product space is the Hilbert space of square integrable functions over the outcome space, with respect to the prior measure over the outcomes, with the covariance inner product.

Conditional expectation, given a sample, corresponds to orthogonal projection into the subspace of all functions over the sample space.

Geometric description

For any collection $\boldsymbol{C} = (C_1, C_2, ...)$ of random quantities, we denote by $\langle \boldsymbol{C} \rangle$ the collection of (finite) linear combinations $\sum_i r_i C_i$ of the elements of $\boldsymbol{C}$. We view $\langle \boldsymbol{C} \rangle$ as a vector space.

Prior covariance is an inner product on $\langle \boldsymbol{C} \rangle$. If $\boldsymbol{C}$ is the union of the elements of the vectors $\boldsymbol{B}$ and $\boldsymbol{D}$, then the adjusted expectation of $Y \in \langle \boldsymbol{B} \rangle$ given $\boldsymbol{D}$, $\mathrm{E}_{\boldsymbol{D}}(X)$, is the orthogonal projection of Y into the linear subspace $\langle \boldsymbol{D} \rangle$, and adjusted variance, $\mathrm{Var}_{\boldsymbol{D}}(X)$, is the squared distance between Y and $\langle \boldsymbol{D} \rangle$.

In the usual Bayes formalism, $\langle C \rangle$ is the collection of all random variables defined over the outcome space. The inner product space is the Hilbert space of square integrable functions over the outcome space, with respect to the prior measure over the outcomes, with the covariance inner product.

Conditional expectation, given a sample, corresponds to orthogonal projection into the subspace of all functions over the sample space.

Bayes linear analysis allows us to restrict prior specification and subsequent projection into the largest subspace of this full space that we are able to specify prior beliefs over.

Bayes linear adjustment

$\mathsf{E}_z[y]$, $\mathsf{Var}_z[y]$, the expectation and variance for y adjusted by z, are given by

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

Bayes linear adjustment

$\mathsf{E}_z[y]$, $\mathsf{Var}_z[y]$, the expectation and variance for y adjusted by z, are given by

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

Bayes linear analysis may be viewed as the appropriate analysis given a partial specification based on expectation (with methodology for modelling, interpretation and diagnostic analysis).

Bayes linear adjustment

$\mathsf{E}_z[y]$, $\mathsf{Var}_z[y]$, the expectation and variance for y adjusted by z, are given by

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

Bayes linear analysis may be viewed as the appropriate analysis given a partial specification based on expectation (with methodology for modelling, interpretation and diagnostic analysis).
Temporal sure preference principle implies that your actual posterior expectation, $\mathrm{E}_T(\boldsymbol{B})$, at time T when you have observed $\boldsymbol{D}$, satisfies the relation

$$\mathrm{E}_T(\boldsymbol{B}) = \mathrm{E}_{\boldsymbol{D}}(\boldsymbol{B}) + \boldsymbol{R},$$

where $\boldsymbol{R}$ has, a priori, zero expectation and is uncorrelated with $\boldsymbol{D}$.

Bayes linear adjustment

$\mathsf{E}_z[y]$, $\mathsf{Var}_z[y]$, the expectation and variance for y adjusted by z, are given by

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

Bayes linear analysis may be viewed as the appropriate analysis given a partial specification based on expectation (with methodology for modelling, interpretation and diagnostic analysis).
Temporal sure preference principle implies that your actual posterior expectation, $\mathrm{E}_T(\boldsymbol{B})$, at time T when you have observed $\boldsymbol{D}$, satisfies the relation

$$\mathrm{E}_T(\boldsymbol{B}) = \mathrm{E}_{\boldsymbol{D}}(\boldsymbol{B}) + \boldsymbol{R},$$

where $\boldsymbol{R}$ has, a priori, zero expectation and is uncorrelated with $\boldsymbol{D}$.
(If $\boldsymbol{D}$ represents a partition, then $\mathrm{E}_{\boldsymbol{D}}(\boldsymbol{B}) = \mathrm{E}(\boldsymbol{B}|\boldsymbol{D})$, and $\mathrm{E}(\boldsymbol{R}|\boldsymbol{D}) = 0$.)

Bayes linear belief nets

A, B, C are collections of random quantities.

C *separates* A and B, denoted $\lfloor A \perp\!\!\!\perp B \rfloor \, / \, C$, if $\mathrm{E}_{C \cup B}(A) = \mathrm{E}_C(A)$

Bayes linear belief nets

A, B, C are collections of random quantities.
C *separates* A and B, denoted $\lfloor A \perp\!\!\!\perp B \rfloor \,/\, C$, if $\mathrm{E}_{C \cup B}(A) = \mathrm{E}_C(A)$
Geometrically, $\lfloor A \perp\!\!\!\perp B \rfloor \,/\, C$ if the orthogonal complements of A and B in C,
$[A - \mathrm{E}_C(A)], [B - \mathrm{E}_C(B)]$ are orthogonal.

Bayes linear belief nets

A, B, C are collections of random quantities.

C *separates* A and B, denoted $\lfloor A \perp\!\!\!\perp B \rfloor \,/\, C$, if $\mathrm{E}_{C\cup B}(A) = \mathrm{E}_C(A)$

Geometrically, $\lfloor A \perp\!\!\!\perp B \rfloor \,/\, C$ if the orthogonal complements of A and B in C, $[A - \mathrm{E}_C(A)], [B - \mathrm{E}_C(B)]$ are orthogonal.

$\lfloor A \perp\!\!\!\perp B \rfloor \,/\, C$ is a generalised conditional independence property

Therefore, graphical models expressing such belief separations (geometrically - the orthogonalities between subspaces) will have many of the same formal properties as do probabilistic graphical models.

Bayes linear belief nets

A, B, C are collections of random quantities.

C *separates* A and B, denoted $\lfloor A \perp\!\!\!\perp B \rfloor \,/\, C$, if $\mathrm{E}_{C \cup B}(A) = \mathrm{E}_C(A)$

Geometrically, $\lfloor A \perp\!\!\!\perp B \rfloor \,/\, C$ if the orthogonal complements of A and B in C, $[A - \mathrm{E}_C(A)], [B - \mathrm{E}_C(B)]$ are orthogonal.

$\lfloor A \perp\!\!\!\perp B \rfloor \,/\, C$ is a generalised conditional independence property

Therefore, graphical models expressing such belief separations (geometrically - the orthogonalities between subspaces) will have many of the same formal properties as do probabilistic graphical models.

Notes

[1] Close relationship with Gaussian graphical models

Bayes linear belief nets

A, B, C are collections of random quantities.

C *separates* A and B, denoted $\lfloor A \perp\!\!\!\perp B \rfloor / C$, if $\mathrm{E}_{C \cup B}(A) = \mathrm{E}_C(A)$

Geometrically, $\lfloor A \perp\!\!\!\perp B \rfloor / C$ if the orthogonal complements of A and B in C, $[A - \mathrm{E}_C(A)], [B - \mathrm{E}_C(B)]$ are orthogonal.

$\lfloor A \perp\!\!\!\perp B \rfloor / C$ is a generalised conditional independence property

Therefore, graphical models expressing such belief separations (geometrically - the orthogonalities between subspaces) will have many of the same formal properties as do probabilistic graphical models.

Notes

[1] Close relationship with Gaussian graphical models

[2] We may distinguish information about the projections from information about the inner product.

Bayes linear belief nets

A, B, C are collections of random quantities.

C *separates* A and B, denoted $\lfloor A \perp\!\!\!\perp B \rfloor \,/\, C$, if $\mathrm{E}_{C\cup B}(A) = \mathrm{E}_C(A)$

Geometrically, $\lfloor A \perp\!\!\!\perp B \rfloor \,/\, C$ if the orthogonal complements of A and B in C, $[A - \mathrm{E}_C(A)], [B - \mathrm{E}_C(B)]$ are orthogonal.

$\lfloor A \perp\!\!\!\perp B \rfloor \,/\, C$ is a generalised conditional independence property

Therefore, graphical models expressing such belief separations (geometrically - the orthogonalities between subspaces) will have many of the same formal properties as do probabilistic graphical models.

Notes

[1] Close relationship with Gaussian graphical models

[2] We may distinguish information about the projections from information about the inner product.

[3] On a Bayes linear graphical model, we can introduce the actual posterior expectations, $\mathrm{E}_T(\boldsymbol{B})$, as well as the adjusted expectations, $\mathrm{E}_{\boldsymbol{D}}(\boldsymbol{B})$.

Calibration via history matching

History Matching is concerned with learning about best inputs, x^*, using simulator evaluations and data, z. Using the emulator we obtain, for each input choice x, the adjusted values of $\mathrm{E}(f(x))$ and $\mathrm{Var}(f(x))$. We rule out regions of x space for which $F(x)$ is judged to be a very poor match to observed z.

Calibration via history matching

History Matching is concerned with learning about best inputs, x^*, using simulator evaluations and data, z. Using the emulator we obtain, for each input choice x, the adjusted values of $\mathrm{E}(f(x))$ and $\mathrm{Var}(f(x))$. We rule out regions of x space for which $F(x)$ is judged to be a very poor match to observed z.
To achieve this, we calculate, for each output $F_i(x)$, the implausibility:

$$I_{(i)}(x) = |\mathrm{E}(f_i(x)) - z_i|^2 / \mathrm{Var}(f_i(x) - z_i)$$

Calibration via history matching

History Matching is concerned with learning about best inputs, x^*, using simulator evaluations and data, z. Using the emulator we obtain, for each input choice x, the adjusted values of $\mathrm{E}(f(x))$ and $\mathrm{Var}(f(x))$. We rule out regions of x space for which $F(x)$ is judged to be a very poor match to observed z.
To achieve this, we calculate, for each output $F_i(x)$, the implausibility:

$$I_{(i)}(x) = |\mathrm{E}(f_i(x)) - z_i|^2 / \mathrm{Var}(f_i(x) - z_i)$$

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

Calibration via history matching

History Matching is concerned with learning about best inputs, x^*, using simulator evaluations and data, z. Using the emulator we obtain, for each input choice x, the adjusted values of $\mathrm{E}(f(x))$ and $\mathrm{Var}(f(x))$. We rule out regions of x space for which $F(x)$ is judged to be a very poor match to observed z.

To achieve this, we calculate, for each output $F_i(x)$, the implausibility:

$$I_{(i)}(x) = |\mathrm{E}(f_i(x)) - z_i|^2 / \mathrm{Var}(f_i(x) - z_i)$$

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

We iteratively refocus on the 'non-implausible' regions of the input space, by further model runs and refitting our emulator over the sub-region and repeating the analysis. This process is a form of iterative global search.

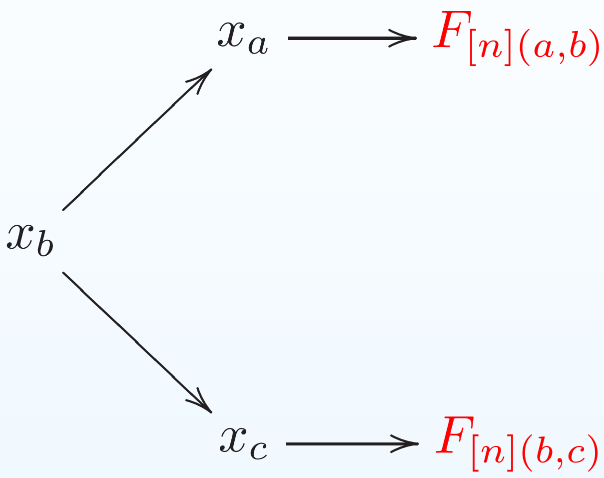
Calibration via history matching

History Matching is concerned with learning about best inputs, x^*, using simulator evaluations and data, z. Using the emulator we obtain, for each input choice x, the adjusted values of $\mathrm{E}(f(x))$ and $\mathrm{Var}(f(x))$. We rule out regions of x space for which $F(x)$ is judged to be a very poor match to observed z.
To achieve this, we calculate, for each output $F_i(x)$, the implausibility:

$$I_{(i)}(x) = |\mathrm{E}(f_i(x)) - z_i|^2 / \mathrm{Var}(f_i(x) - z_i)$$

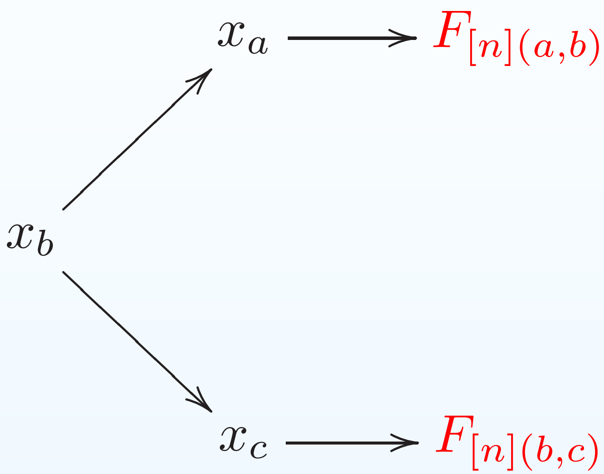
This calculation can be performed univariately, or over sub-vectors. The implausibilities are then combined, such as by using $I_M(x) = \max_i I_{(i)}(x)$, identifying regions of x with large $I_M(x)$ as unlikely to be good choices for x^*.
We iteratively refocus on the ‘non-implausible’ regions of the input space, by further model runs and refitting our emulator over the sub-region and repeating the analysis. This process is a form of iterative global search.
If all values of x are implausible, this is highly informative!

Causal structure and design



Functional graphical models are causal models on the functional inputs.

Causal structure and design

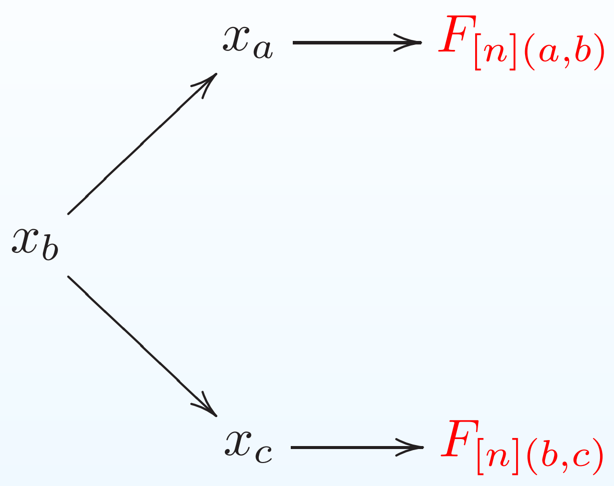


Functional graphical models are causal models on the functional inputs.

Here the outputs divide into three sets x_a, x_b, x_c.

Outputs $F_{(a,b)}$, depend only on x_a, x_b. Outputs $F_{(b,c)}$, depend only on x_b, x_c

Causal structure and design



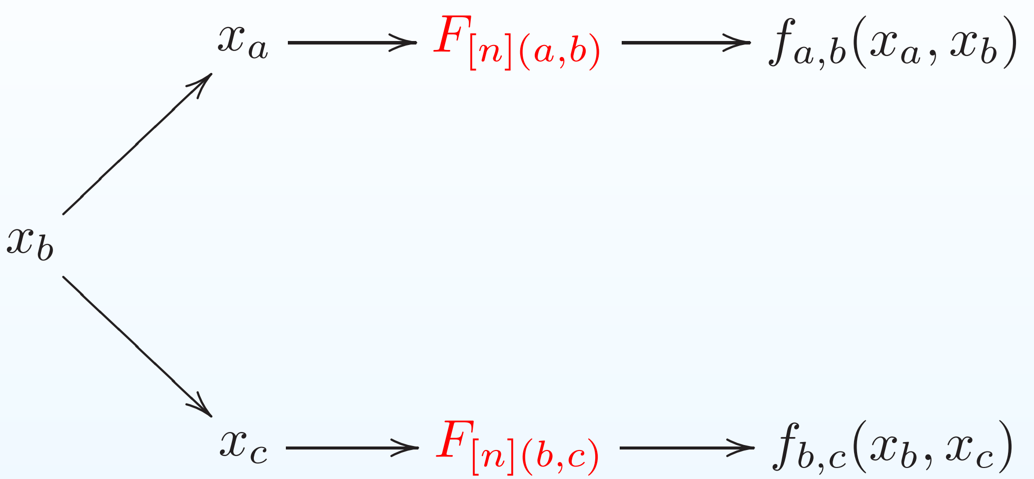
Functional graphical models are causal models on the functional inputs.

Here the outputs divide into three sets x_a, x_b, x_c.

Outputs $F_{(a,b)}$, depend only on x_a, x_b. Outputs $F_{(b,c)}$, depend only on x_b, x_c

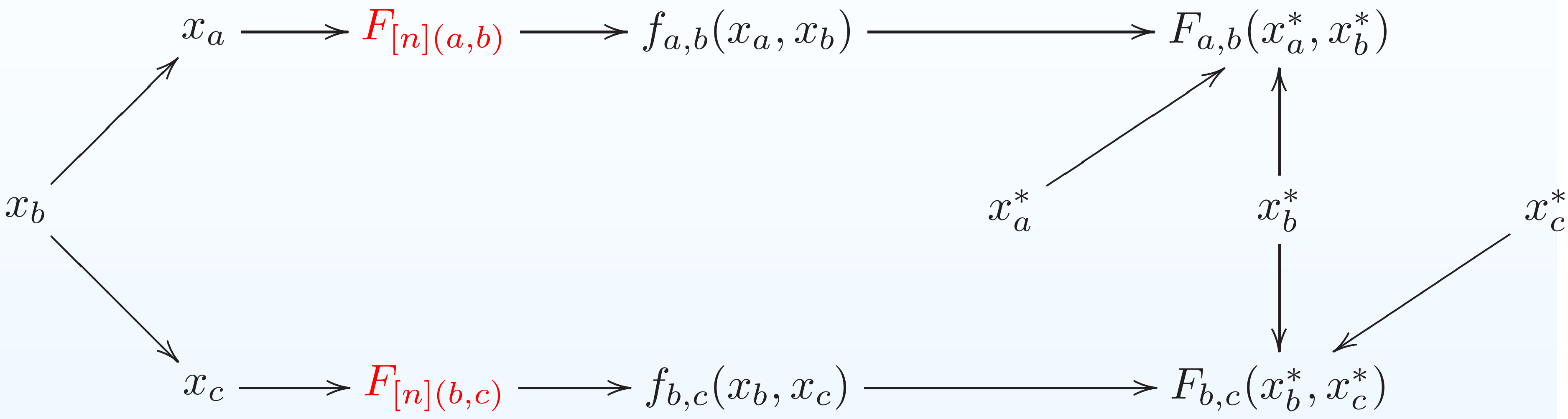
Therefore, we can design a collection of n evaluations, $F_{[n](a,b)}$ and $F_{[n](b,c)}$ independently given our design for x_b (which is enormously helpful in reducing dimensionality)

Design and emulation



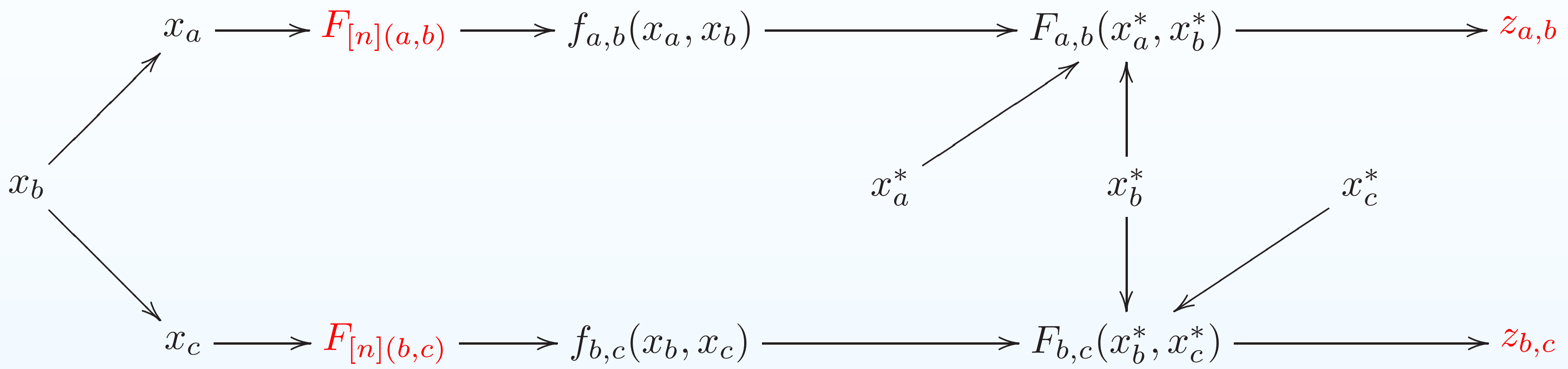
Evaluations, $F_{[n](a,b)}$ and $F_{[n](b,c)}$ are inputs to the corresponding emulators $f_{a,b}(x_a, x_b)$, $f_{b,c}(x_b, x_c)$

Emulation and best evaluations



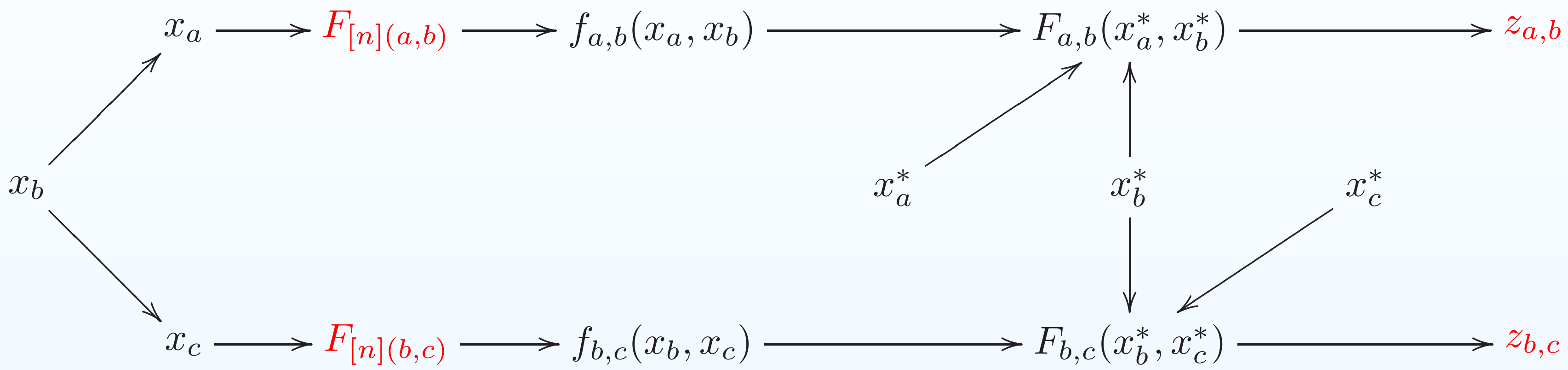
The emulators combine with the true values x^* to generate judgements for model runs at true inputs

Emulation and Calibration



The link to data observations z allows us to assess our implausibility measures over the input space, x, by local computation.

Emulation and Calibration



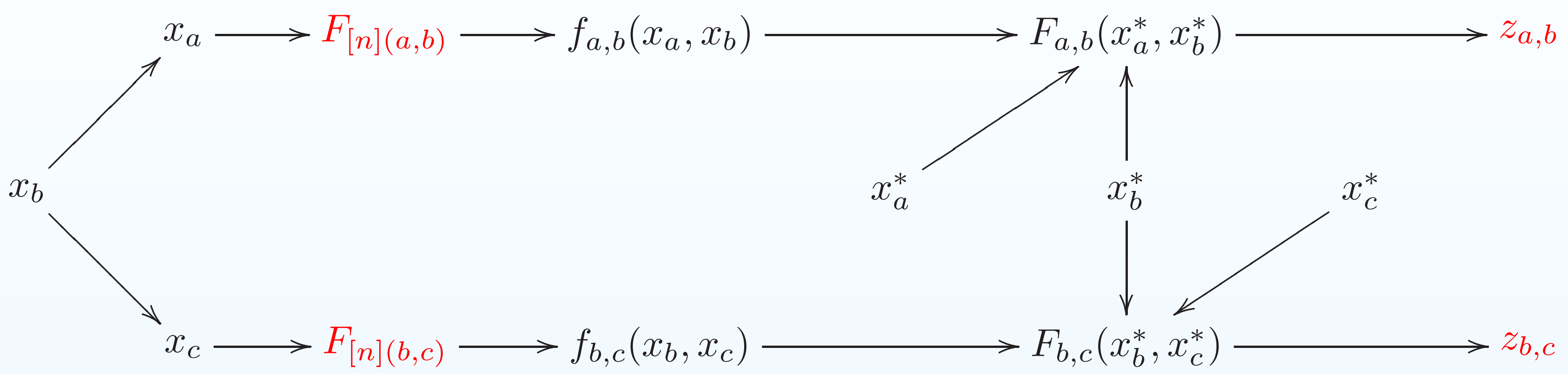
The link to data observations z allows us to assess our implausibility measures over the input space, x, by local computation.

In the above diagram, we collect the implausibility measure to x_b from

(i) the x_a, x_b pair, based on $z_{a,b}$

(ii) the x_b, x_c pair, based on $z_{b,c}$

Emulation and Calibration



The link to data observations z allows us to assess our implausibility measures over the input space, x, by local computation.

In the above diagram, we collect the implausibility measure to x_b from

(i) the x_a, x_b pair, based on $z_{a,b}$

(ii) the x_b, x_c pair, based on $z_{b,c}$

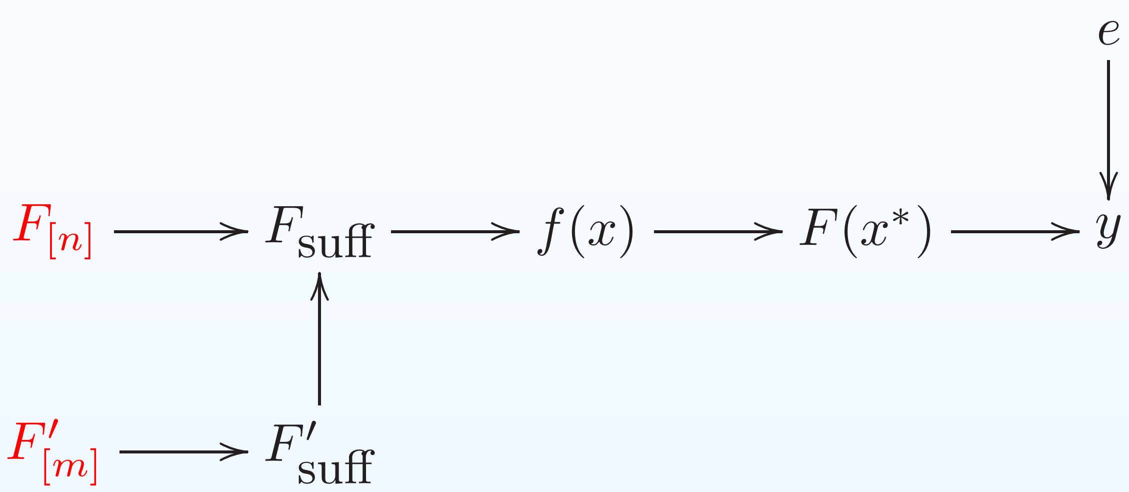
We then distribute the combined implausibility measure back to x_a and x_c.

Small samples

$$
\begin{array}{ccccccccc}
 & & & & & & & & e \\
 & & & & & & & & \downarrow \\
F_{[n]} & \longrightarrow & F_{\text{suff}} & \longrightarrow & f(x) & \longrightarrow & F(x^*) & \longrightarrow & y
\end{array}
$$

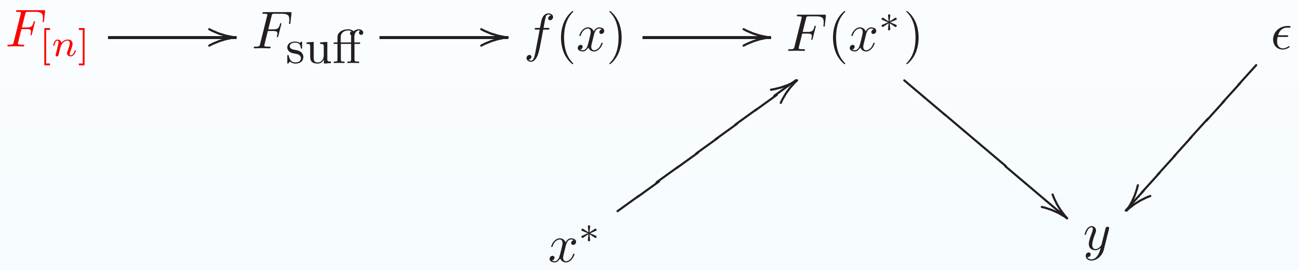
Often, we can only make a few evaluations of our computer simulator, so that our evaluation $F_{[n]}$ is based on small value of n.

Small samples and fast approximations



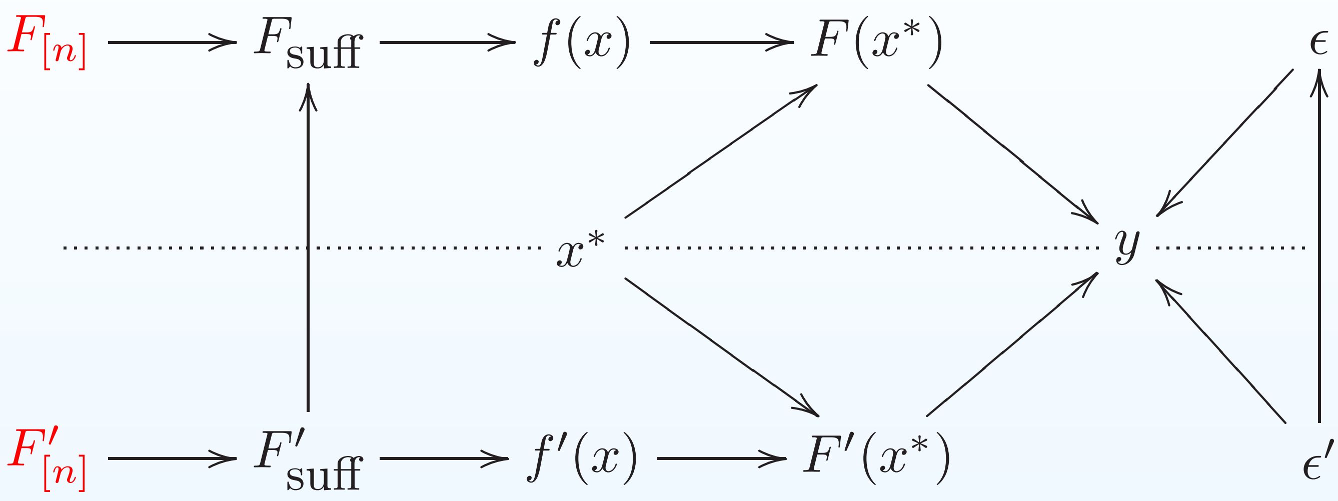
We may be able to make many evaluations, $F'_{[m]}$ of a simpler approximate version of the model as a basis for the inference.

A graphical puzzle



We link evaluations of our simulator F through our emulator to the system values.

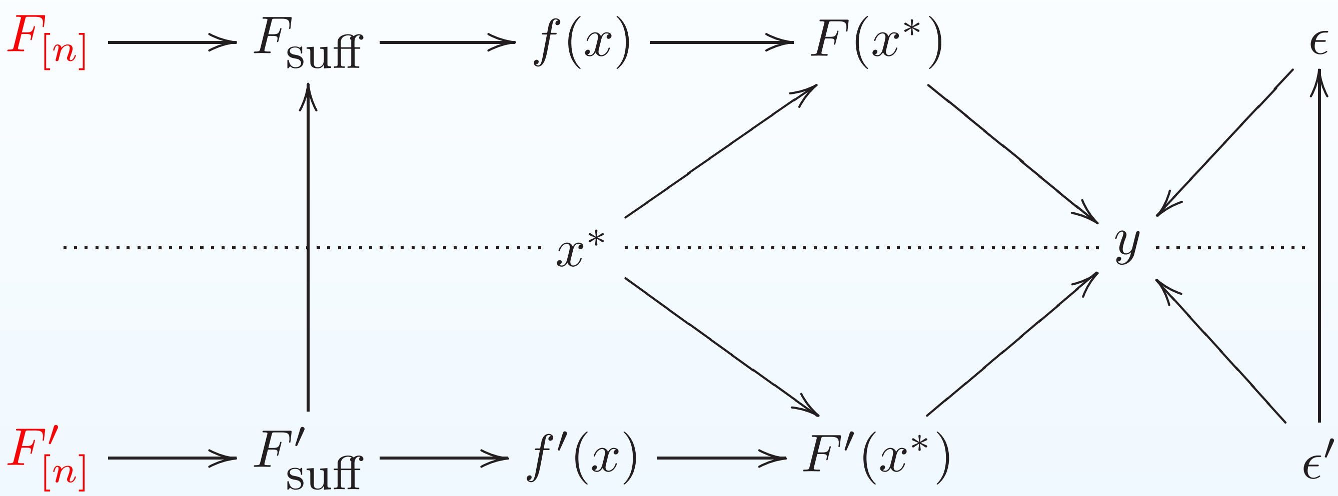
A graphical puzzle



Now add the fast approximation F' to the graph.

But suppose that, last year, the fast approximation was the full model, for which we had already drawn the corresponding version of this graph.

A graphical puzzle



Now add the fast approximation F' to the graph.

But suppose that, last year, the fast approximation was the full model, for which we had already drawn the corresponding version of this graph.

Comment: you can't get all of the conditional orthogonalities in the above diagram without imposing unreasonable constraints on the system.

The Best input

How does learning about F inform us about y?

The Best input

How does learning about F inform us about y?

The simplest (and therefore most popular) way to relate uncertainty about the simulator and the system is the so-called “Best Input Approach”.
We proceed as though there exists a value x^* independent of the function F such that the value of $F(x^*)$ summarises all of the information that the simulator conveys about the system. This means that we consider the model discrepancy, $\epsilon = y - F(x^*)$, to be independent of F, x^*.

The Best input

How does learning about F inform us about y?

The simplest (and therefore most popular) way to relate uncertainty about the simulator and the system is the so-called "Best Input Approach".
We proceed as though there exists a value x^* independent of the function F such that the value of $F(x^*)$ summarises all of the information that the simulator conveys about the system. This means that we consider the model discrepancy, $\epsilon = y - F(x^*)$, to be independent of F, x^*.

This formulation raises serious questions.
In particular, does x^* correspond to "true" system properties?

The Best input

How does learning about F inform us about y?

The simplest (and therefore most popular) way to relate uncertainty about the simulator and the system is the so-called “Best Input Approach”.
We proceed as though there exists a value x^* independent of the function F such that the value of $F(x^*)$ summarises all of the information that the simulator conveys about the system. This means that we consider the model discrepancy, $\epsilon = y - F(x^*)$, to be independent of F, x^*.

This formulation raises serious questions.
In particular, does x^* correspond to “true” system properties?

If so, why should they give best fit to our imperfect model?

The Best input

How does learning about F inform us about y?

The simplest (and therefore most popular) way to relate uncertainty about the simulator and the system is the so-called “Best Input Approach”.
We proceed as though there exists a value x^* independent of the function F such that the value of $F(x^*)$ summarises all of the information that the simulator conveys about the system. This means that we consider the model discrepancy, $\epsilon = y - F(x^*)$, to be independent of F, x^*.

This formulation raises serious questions.
In particular, does x^* correspond to “true” system properties?

If so, why should they give best fit to our imperfect model?

If not, why should there even be any such “best” inputs, and what does it mean to have expert judgements about x^*?

The Best input

How does learning about F inform us about y?

The simplest (and therefore most popular) way to relate uncertainty about the simulator and the system is the so-called “Best Input Approach”.
We proceed as though there exists a value x^* independent of the function F such that the value of $F(x^*)$ summarises all of the information that the simulator conveys about the system. This means that we consider the model discrepancy, $\epsilon = y - F(x^*)$, to be independent of F, x^*.

This formulation raises serious questions.
In particular, does x^* correspond to “true” system properties?

If so, why should they give best fit to our imperfect model?

If not, why should there even be any such “best” inputs, and what does it mean to have expert judgements about x^*?

Further, surprising contradictions arise when we try to construct joint specifications linking collections of models to the physical system in this way.

Reification

Why should we consider that there is a 'best input' x^*?

Reification

Why should we consider that there is a 'best input' x^*?

What does a simulator F really tell us about a physical system y?

Reification

Why should we consider that there is a ‘best input’ x^*?

What does a simulator F really tell us about a physical system y?

How do we combine the information about y from a collection of simulators?

Reification

Why should we consider that there is a ‘best input’ x^*?
What does a simulator F really tell us about a physical system y?
How do we combine the information about y from a collection of simulators?

Consider both our inputs x and the simulator F as abstractions/simplifications of real physical quantities and processes (through approximations in physics, solution methods, level of detail, limitations of current understanding) to a much more realistic simulator F^*, for which real, physical x^* would be the best input, in the sense that $(y - F^*(x^*))$ would be judged independent of (x^*, F^*).

Reification

Why should we consider that there is a 'best input' x^*?
What does a simulator F really tell us about a physical system y?
How do we combine the information about y from a collection of simulators?

Consider both our inputs x and the simulator F as abstractions/simplifications of real physical quantities and processes (through approximations in physics, solution methods, level of detail, limitations of current understanding) to a much more realistic simulator F^*, for which real, physical x^* would be the best input, in the sense that $(y - F^*(x^*))$ would be judged independent of (x^*, F^*).

We call F^* the reified simulator (from reify: to treat an abstract concept as if it was real).

Reification

Why should we consider that there is a 'best input' x^*?
What does a simulator F really tell us about a physical system y?
How do we combine the information about y from a collection of simulators?

Consider both our inputs x and the simulator F as abstractions/simplifications of real physical quantities and processes (through approximations in physics, solution methods, level of detail, limitations of current understanding) to a much more realistic simulator F^*, for which real, physical x^* would be the best input, in the sense that $(y - F^*(x^*))$ would be judged independent of (x^*, F^*).

We call F^* the reified simulator (from reify: to treat an abstract concept as if it was real).

Reifying principle
[1] Simulator F is informative for y, because F is informative for F^* and $F^*(x^*)$ is informative for y.

Reification

Why should we consider that there is a 'best input' x^*?
What does a simulator F really tell us about a physical system y?
How do we combine the information about y from a collection of simulators?

Consider both our inputs x and the simulator F as abstractions/simplifications of real physical quantities and processes (through approximations in physics, solution methods, level of detail, limitations of current understanding) to a much more realistic simulator F^*, for which real, physical x^* would be the best input, in the sense that $(y - F^*(x^*))$ would be judged independent of (x^*, F^*).

We call F^* the reified simulator (from reify: to treat an abstract concept as if it was real).

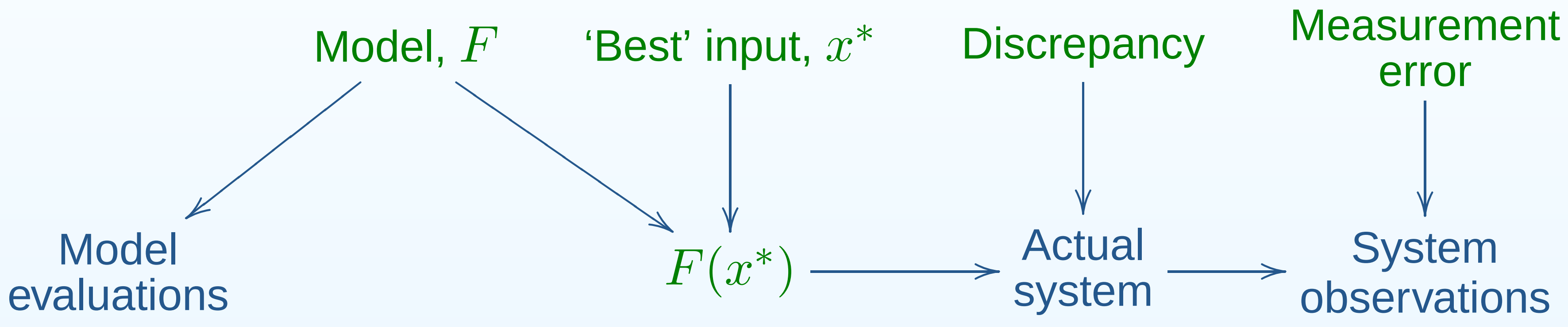
Reifying principle
[1] Simulator F is informative for y, because F is informative for F^* and $F^*(x^*)$ is informative for y.
[2] A collection of simulators $F_1, F_2, \ldots$ is jointly informative for y, as the simulators are jointly informative for F^*.

Relating the model and the system (2)

Our model F is informative for y because F is informative for reified model F^*

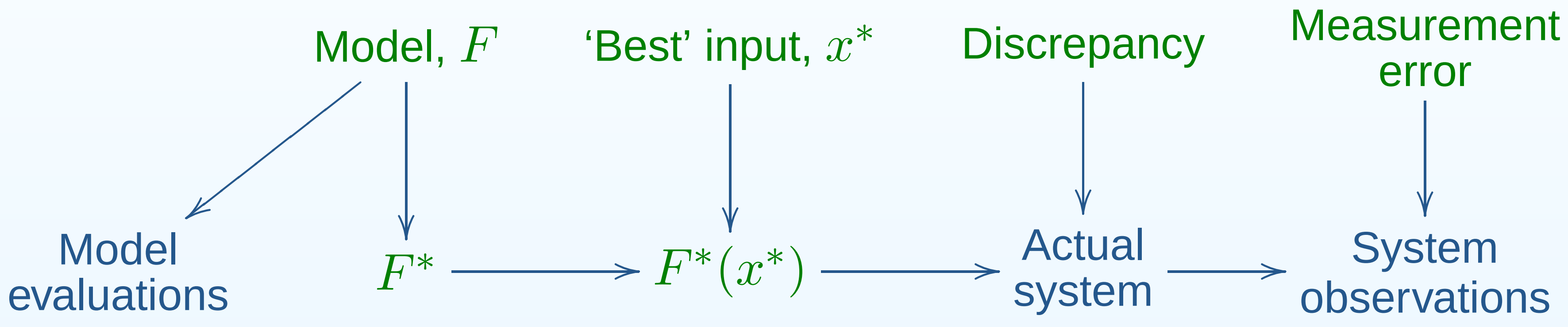
Relating the model and the system (2)

Our model F is informative for y because F is informative for reified model F^*



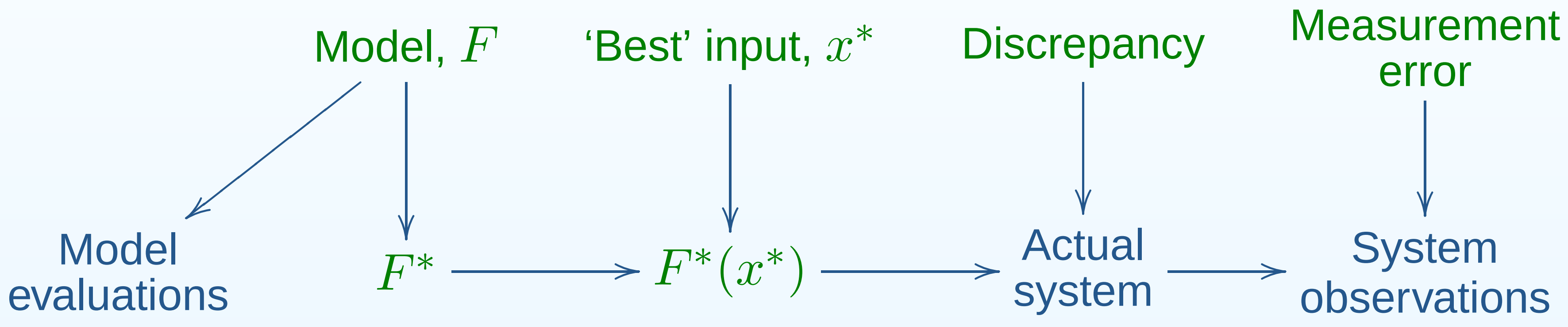
Relating the model and the system (2)

Our model F is informative for y because F is informative for reified model F^*



Relating the model and the system (2)

Our model F is informative for y because F is informative for reified model F^*



Comment: Statistical graphical models need reification too!

Linking F and F^* using emulators

Suppose that our emulator for F is

$$f(x) = Bg(x) + u(x)$$

Linking F and F^* using emulators

Suppose that our emulator for F is

$$f(x) = Bg(x) + u(x)$$

Our simplest emulator for F^* might be

$$f^*(x, w) = B^*g(x) + u^*(x) + u^*(x, w)$$

where we might model our judgements as $B^* = CB + \Gamma$, correlate $u(x)$ and $u^*(x)$, while $u^*(x, w)$, with additional parameters, w, is uncorrelated with remainder.

Linking F and F^* using emulators

Suppose that our emulator for F is

$$f(x) = Bg(x) + u(x)$$

Our simplest emulator for F^* might be

$$f^*(x, w) = B^* g(x) + u^*(x) + u^*(x, w)$$

where we might model our judgements as $B^* = CB + \Gamma$, correlate $u(x)$ and $u^*(x)$, while $u^*(x, w)$, with additional parameters, w, is uncorrelated with remainder.

Structured reification: systematic probabilistic modelling for all those aspects of model deficiency whose effects we are prepared to consider explicitly.

Reified inference structure

$$F_{[n]} \longrightarrow F_{\text{suff}}$$

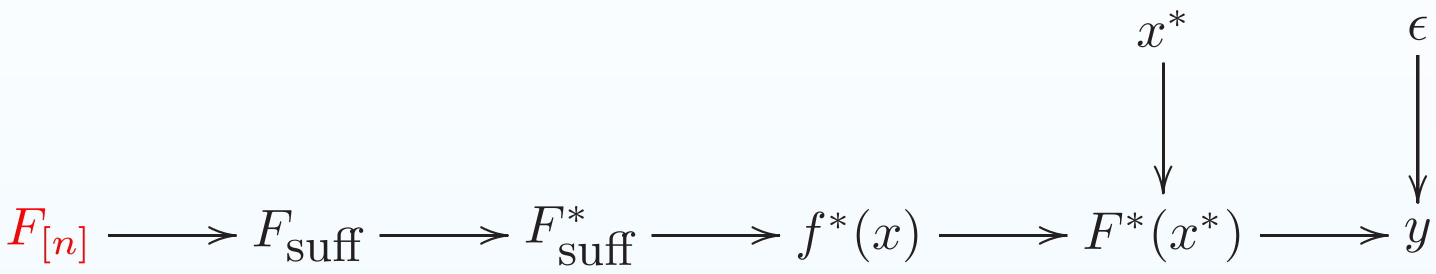
$F_{[n]}$: n evaluations of F at inputs $x_1, x_2, \ldots$
F_{suff}: the global information from $F_{[n]}$.

Reified inference structure

$$F_{[n]} \longrightarrow F_{\text{suff}} \longrightarrow F^*_{\text{suff}}$$

F^*_{suff}: corresponding global information for reified emulator $f^*(x)$

Reified inference structure



True system properties x^* with emulator $f^*(x)$ influence beliefs for $F(x^*)$, which is informative for system values y, with discrepancy ϵ.

Reifified inference structure

$$\begin{array}{ccccccccccc} & & & & & & & & x^* & & \epsilon \\ & & & & & & & & \downarrow & & \downarrow \\ F_{[n]} & \longrightarrow & F_{\text{suff}} & \longrightarrow & F^*_{\text{suff}} & \longrightarrow & f^*(x) & \longrightarrow & F^*(x^*) & \longrightarrow & y \end{array}$$

True system properties x^* with emulator $f^*(x)$ influence beliefs for $F(x^*)$, which is informative for system values y, with discrepancy ϵ.

Comment: All our calibration and forecasting methodology is unchanged - all that has changed is our description of the joint covariance structure.

A Reified influence diagram

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

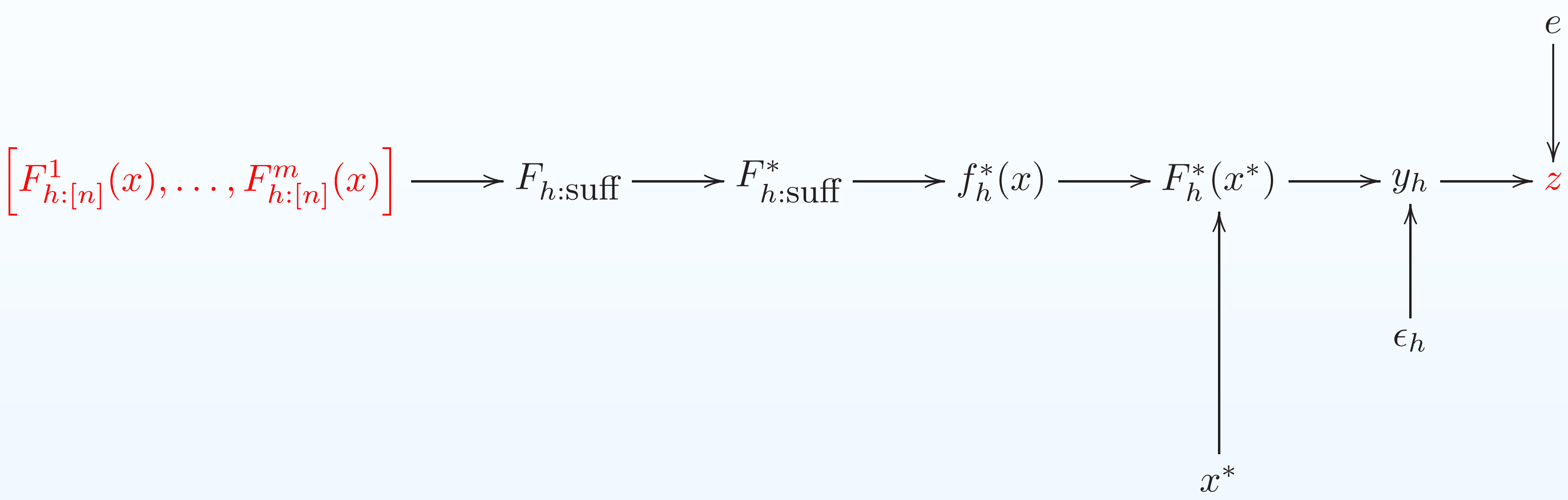
Evaluations of the simulator at each of m initial conditions
for historical components of simulator

A Reified influence diagram

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

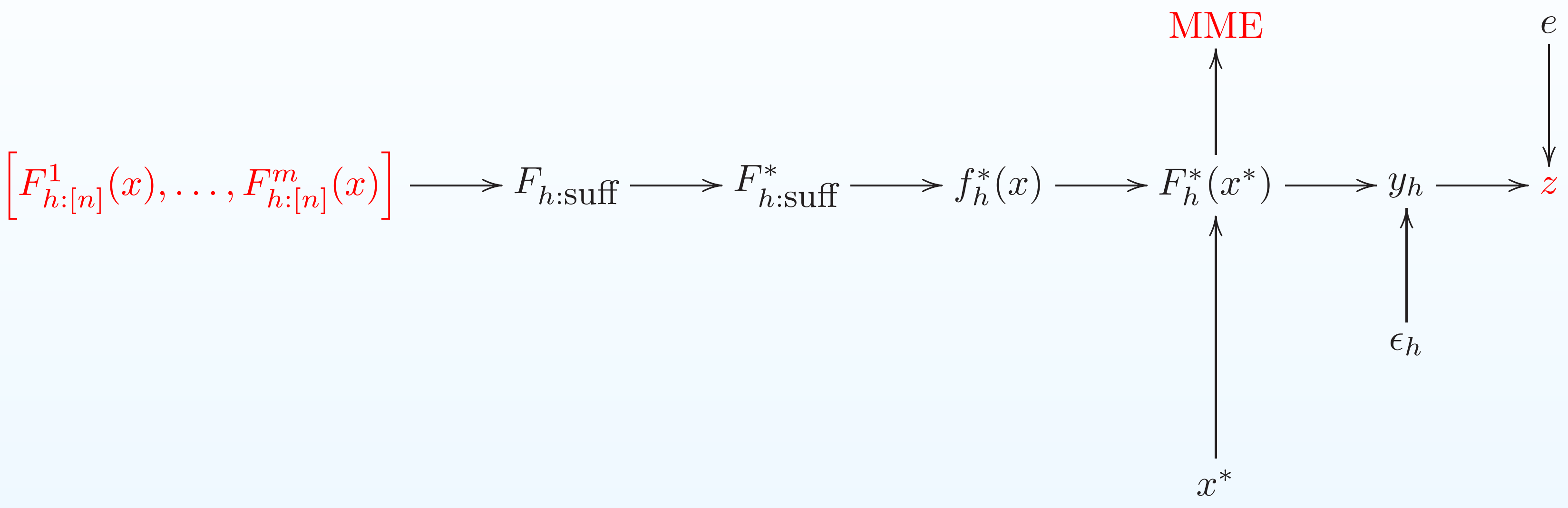
Global information $F_{\text{h:suff}}$ (from second order exchangeability modelling).
passes to Reified global form and to reified emulator.

A Reified influence diagram



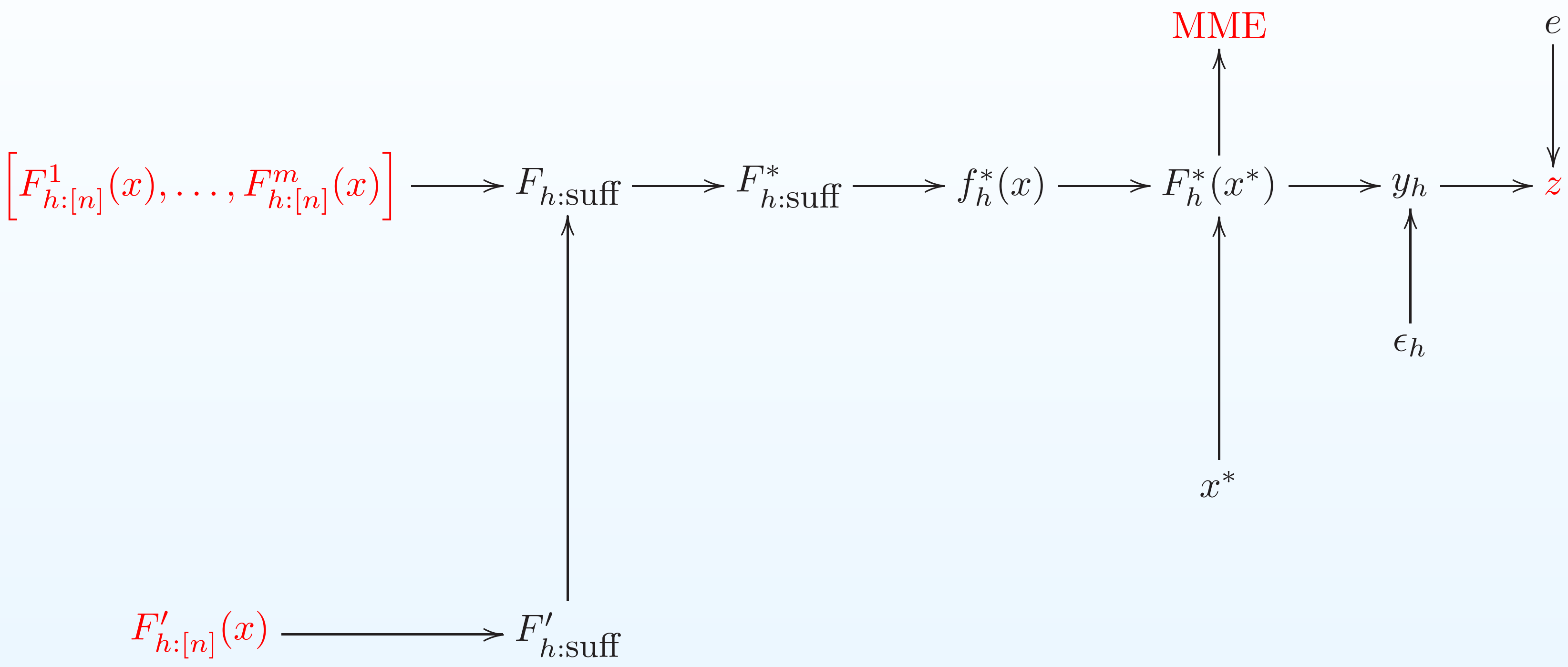
Link with x^* to reified function, at true initial condition, linked to data z

A Reified influence diagram



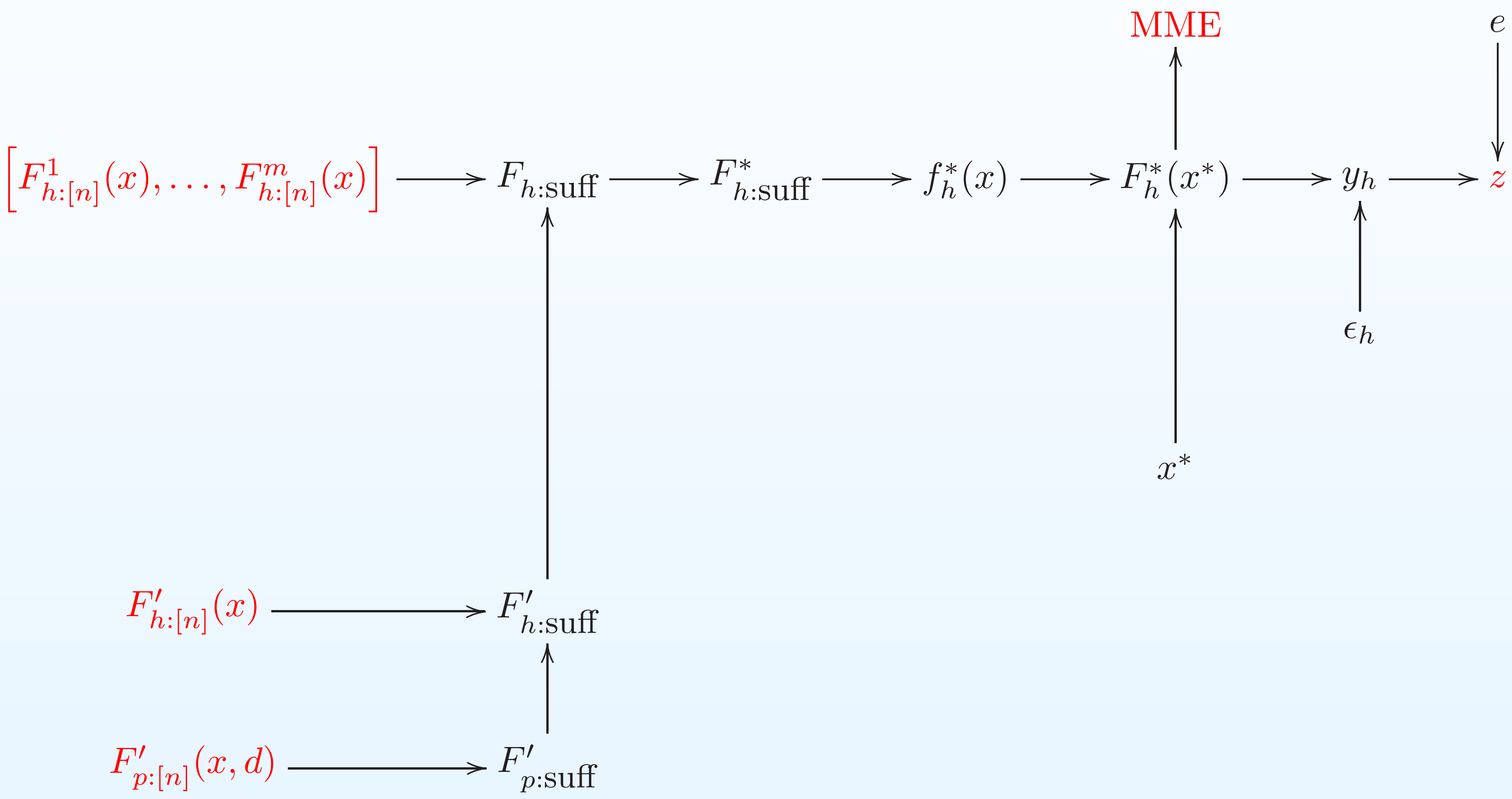
Add observation of a related multi-model ensemble (MME) consisting of tuned runs from related models (more exchangeability modelling).

A Reified influence diagram



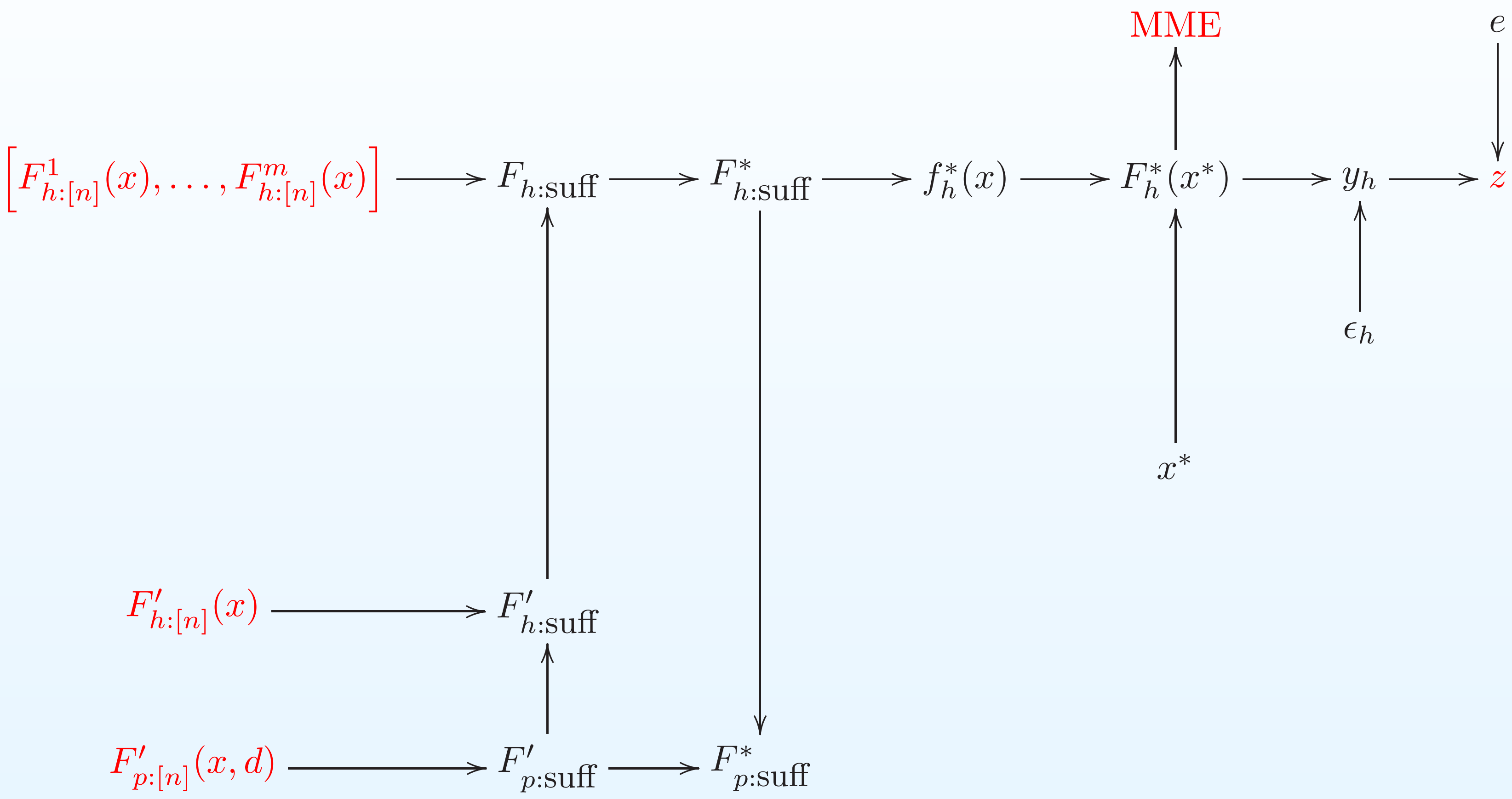
Add a set of evaluations from a fast approximation

A Reified influence diagram



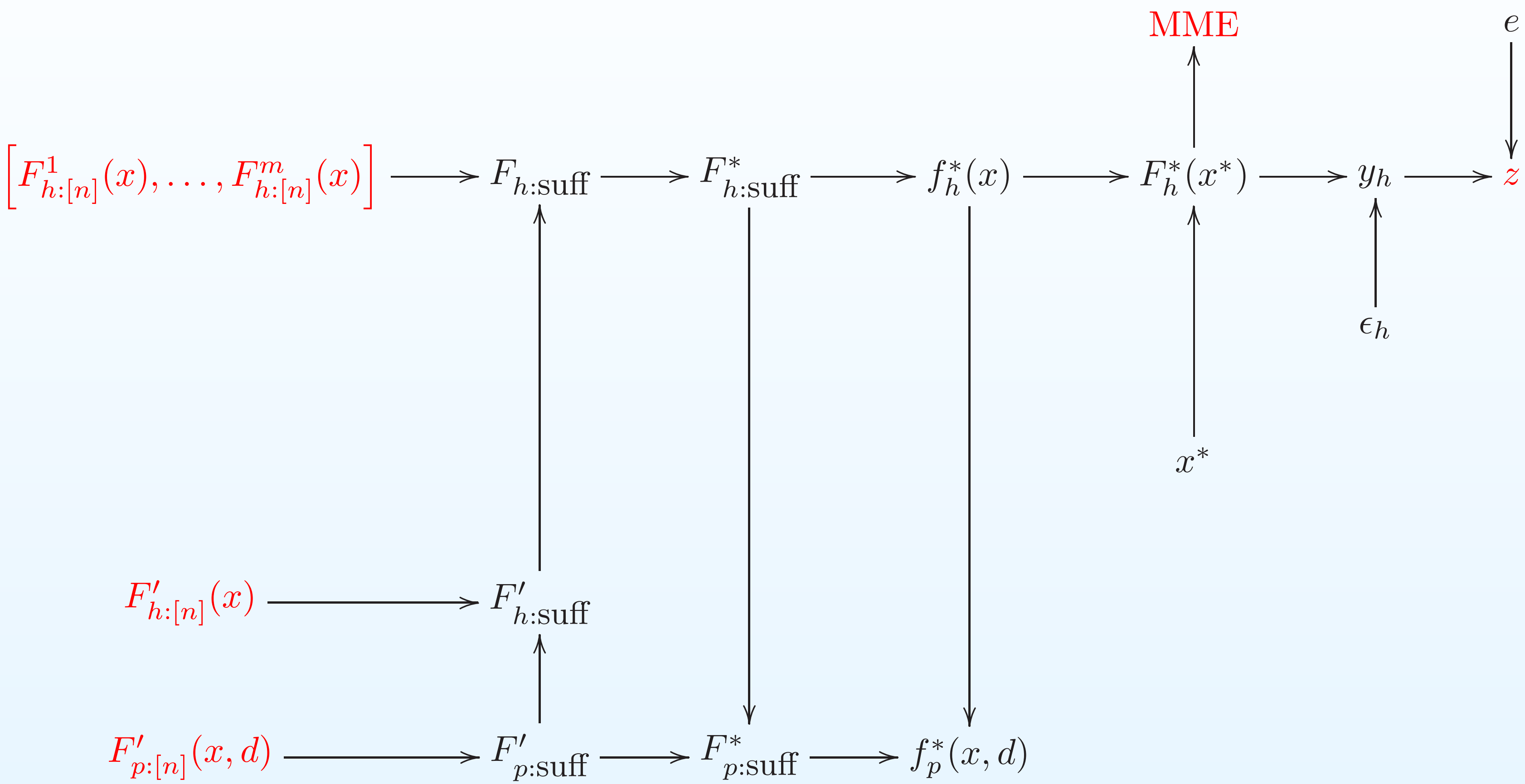
Add evaluations of fast simulator for outcomes to be predicted, with decision choices d

A Reified influence diagram



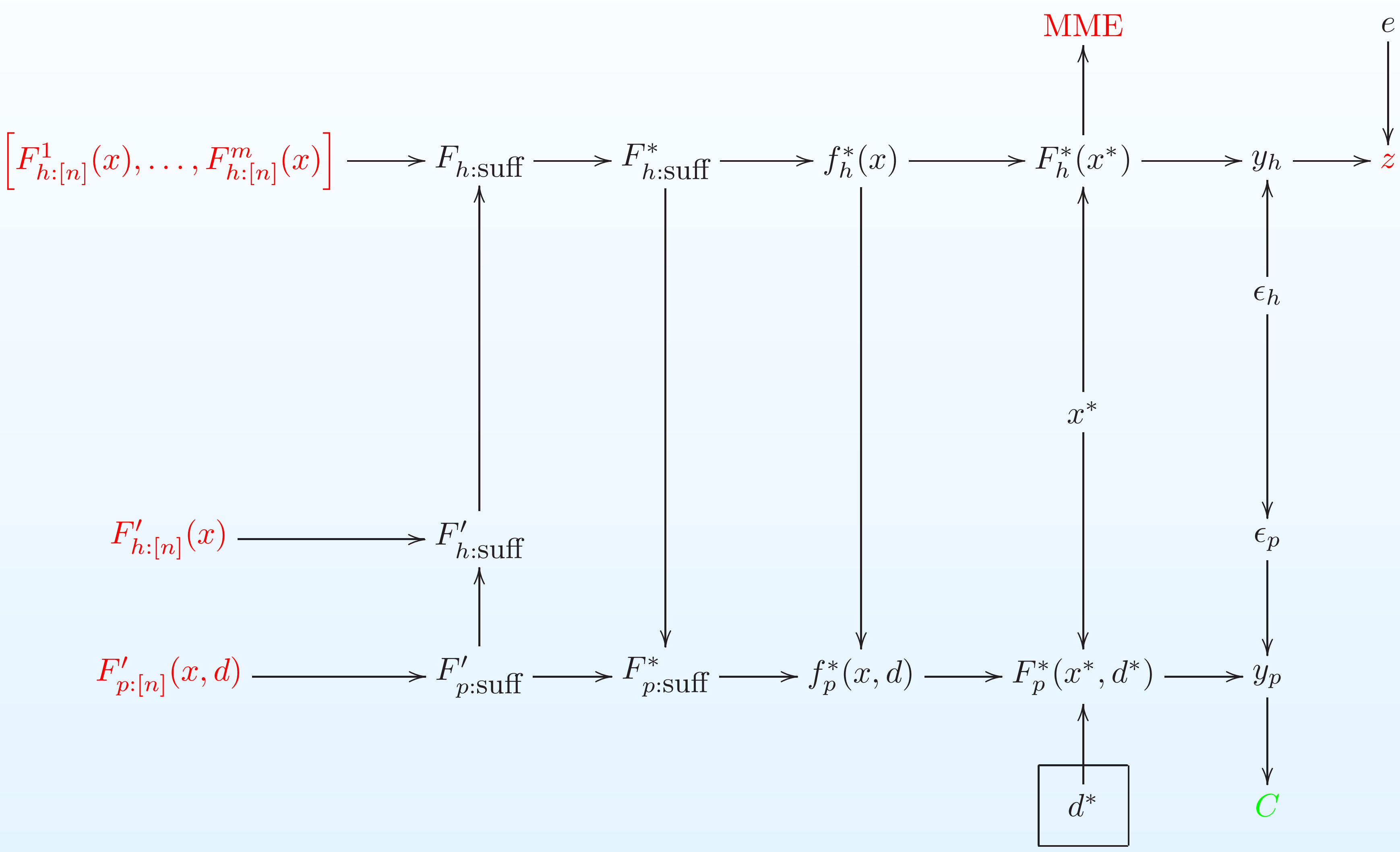
Link to reified global terms for quantities to be predicted

A Reified influence diagram



And to reified global emulator, based on inputs and decisions

A Reified influence diagram



And link, through true future values y_p, to the overall utility cost C of making decision choice d^*.

Best current judgements for complex systems

To assess best current judgements about complex systems, it is enormously helpful to have an overall framework to unify all the uncertainties arising from

Uncertain model parameters, outputs and discrepancies

Uncertain observations/initial conditions/forcing functions

Uncertain relationships between different modelling approaches

Uncertain effects of our attempts to influence the system

Best current judgements for complex systems

To assess best current judgements about complex systems, it is enormously helpful to have an overall framework to unify all the uncertainties arising from

Uncertain model parameters, outputs and discrepancies

Uncertain observations/initial conditions/forcing functions

Uncertain relationships between different modelling approaches

Uncertain effects of our attempts to influence the system

Bayes linear influence diagrams provide a conceptual/graphical framework for unifying our qualitative and quantitative knowledge about all such uncertainties within a structure which is both logical and tractable, so that we can focus on science rather than technical/computational issues.

We need new methodology to construct the general language and tool kit required for making this synthesis in principle, and a close joint effort between statisticians and scientists to achieve the synthesis in practice.

Best current judgements for complex systems

To assess best current judgements about complex systems, it is enormously helpful to have an overall framework to unify all the uncertainties arising from

Uncertain model parameters, outputs and discrepancies

Uncertain observations/initial conditions/forcing functions

Uncertain relationships between different modelling approaches

Uncertain effects of our attempts to influence the system

Bayes linear influence diagrams provide a conceptual/graphical framework for unifying our qualitative and quantitative knowledge about all such uncertainties within a structure which is both logical and tractable, so that we can focus on science rather than technical/computational issues.

We need new methodology to construct the general language and tool kit required for making this synthesis in principle, and a close joint effort between statisticians and scientists to achieve the synthesis in practice.

Such analysis poses serious challenges, but they are no harder than all of the other modelling, computational and observational challenges involved with studying large scale physical systems.

References

P.S. Craig, M. Goldstein, J.C.Rougier, A.H. Seheult, (2001) Bayesian Forecasting Using Large Computer Models, JASA, 96, 717-729

M. Goldstein (1997) Prior inferences for posterior judgements, in Structures and norms in Science, M.C.D. Chiara et. al. eds., 55-71, Kluwer.

M. Goldstein and J.C.Rougier (2008). Reified Bayesian modelling and inference for physical systems (with discussion), JSPI, to appear, .

M. Goldstein and D.A.Wooff (2007) Bayes linear statistics: theory and methods (esp. Chapter 10) Wiley

References

P.S. Craig, M. Goldstein, J.C.Rougier, A.H. Seheult, (2001) Bayesian Forecasting Using Large Computer Models, JASA, 96, 717-729

M. Goldstein (1997) Prior inferences for posterior judgements, in Structures and norms in Science, M.C.D. Chiara et. al. eds., 55-71, Kluwer.

M. Goldstein and J.C.Rougier (2008). Reified Bayesian modelling and inference for physical systems (with discussion), JSPI, to appear, .

M. Goldstein and D.A.Wooff (2007) Bayes linear statistics: theory and methods (esp. Chapter 10) Wiley

And check out the website for the
Managing Uncertainty in Complex Models (MUCM) project
[A consortium of Aston, Durham, LSE, Sheffield and Southampton all hard at work on developing technology for computer model uncertainty problems.]