Parameter estimation for system submodels with limited or missing data using a data-free inference procedure

SAND2018-5444C

Tiernan Casey, Habib N. Najm

Combustion Research Facility
Sandia National Laboratories
Livermore, CA, USA
tcasey@sandia.gov

SIAM Conference on Uncertainty Quantification

Garden Grove, CA, USA
April 16-19, 2018

Introduction

Models for physical systems of practical interest often are often characterized by hierarchies of models, e.g. continuum approximations

Such approaches inevitably require closure models for sub-continuum processes, e.g. constitutive relations

This may involve the evaluation of detailed models at the sub-continuum level, from which statistics may be extracted, or more commonly carefully designed experiments are performed to constrain these models with (typically noisy) measurement data

A robust treatment of uncertainty in such systems should involve analysis these experimental data sources, perhaps computing a joint probability density function (JPDF) to describe the model parameters

Introduction

Models for physical systems of practical interest often are often characterized by hierarchies of models, e.g. continuum approximations

Such approaches inevitably require closure models for sub-continuum processes, e.g. constitutive relations

This may involve the evaluation of detailed models at the sub-continuum level, from which statistics may be extracted, or more commonly carefully designed experiments are performed to constrain these models with (typically noisy) measurement data

A robust treatment of uncertainty in such systems should involve analysis these experimental data sources, perhaps computing a joint probability density function (JPDF) to describe the model parameters

Introduction

Models for physical systems of practical interest often are often characterized by hierarchies of models, e.g. continuum approximations

Such approaches inevitably require closure models for sub-continuum processes, e.g. constitutive relations

This may involve the evaluation of detailed models at the sub-continuum level, from which statistics may be extracted, or more commonly carefully designed experiments are performed to constrain these models with (typically noisy) measurement data

A robust treatment of uncertainty in such systems should involve analysis these experimental data sources, perhaps computing a joint probability density function (JPDF) to describe the model parameters

Introduction

Models for physical systems of practical interest often are often characterized by hierarchies of models, e.g. continuum approximations

Such approaches inevitably require closure models for sub-continuum processes, e.g. constitutive relations

This may involve the evaluation of detailed models at the sub-continuum level, from which statistics may be extracted, or more commonly carefully designed experiments are performed to constrain these models with (typically noisy) measurement data

A robust treatment of uncertainty in such systems should involve analysis these experimental data sources, perhaps computing a joint probability density function (JPDF) to describe the model parameters

Model parameter estimation in the absence of data

Given noisy evidential data, $\mathbf{D}_0$, (e.g. from an experiment) statistical methodologies (e.g. Bayesian inference) can be employed to estimate joint posterior densities on model parameters, λ:

$$p(\lambda|\mathbf{D}_0) \propto p(\mathbf{D}_0|\lambda)p(\lambda) \quad (1)$$

However, raw experimental data are usually unreported, and for many legacy experiments the underlying data are missing or lost forever

What is reported are limited (low-dimensional) summary statistics, $S(\mathbf{D}_0)$, of the data, typically in the form of nominal fitting model parameters, interval uncertainty

Also perhaps nominal information about the experiment and fitting context (fitting model form, experiment initial and boundary conditions, etc.)

This limited form of reporting often omits important information about the parametric uncertainty e.g. correlations, hindering uncertainty quantification efforts

Model parameter estimation in the absence of data

Given noisy evidential data, $\mathbf{D}_0$, (e.g. from an experiment) statistical methodologies (e.g. Bayesian inference) can be employed to estimate joint posterior densities on model parameters, λ:

$$p(\lambda|\mathbf{D}_0) \propto p(\mathbf{D}_0|\lambda)p(\lambda) \tag{1}$$

However, raw experimental data are usually unreported, and for many legacy experiments the underlying data are missing or lost forever

What is reported are limited (low-dimensional) summary statistics, $S(\mathbf{D}_0)$, of the data, typically in the form of nominal fitting model parameters, interval uncertainty

Also perhaps nominal information about the experiment and fitting context (fitting model form, experiment initial and boundary conditions, etc.)

This limited form of reporting often omits important information about the parametric uncertainty e.g. correlations, hindering uncertainty quantification efforts

Model parameter estimation in the absence of data

Given noisy evidential data, $\mathbf{D}_0$, (e.g. from an experiment) statistical methodologies (e.g. Bayesian inference) can be employed to estimate joint posterior densities on model parameters, λ:

$$p(\lambda|\mathbf{D}_0) \propto p(\mathbf{D}_0|\lambda)p(\lambda) \qquad (1)$$

However, raw experimental data are usually unreported, and for many legacy experiments the underlying data are missing or lost forever

What is reported are limited (low-dimensional) summary statistics, $S(\mathbf{D}_0)$, of the data, typically in the form of nominal fitting model parameters, interval uncertainty

Also perhaps nominal information about the experiment and fitting context (fitting model form, experiment initial and boundary conditions, etc.)

This limited form of reporting often omits important information about the parametric uncertainty e.g. correlations, hindering uncertainty quantification efforts

Model parameter estimation in the absence of data

Given noisy evidential data, $\mathbf{D}_0$, (e.g. from an experiment) statistical methodologies (e.g. Bayesian inference) can be employed to estimate joint posterior densities on model parameters, λ:

$$p(\lambda|\mathbf{D}_0) \propto p(\mathbf{D}_0|\lambda)p(\lambda) \quad (1)$$

However, raw experimental data are usually unreported, and for many legacy experiments the underlying data are missing or lost forever

What is reported are limited (low-dimensional) summary statistics, $S(\mathbf{D}_0)$, of the data, typically in the form of nominal fitting model parameters, interval uncertainty

Also perhaps nominal information about the experiment and fitting context (fitting model form, experiment initial and boundary conditions, etc.)

This limited form of reporting often omits important information about the parametric uncertainty e.g. correlations, hindering uncertainty quantification efforts

Model parameter estimation in the absence of data

Given noisy evidential data, $\mathbf{D}_0$, (e.g. from an experiment) statistical methodologies (e.g. Bayesian inference) can be employed to estimate joint posterior densities on model parameters, λ:

$$p(\lambda|\mathbf{D}_0) \propto p(\mathbf{D}_0|\lambda)p(\lambda) \tag{1}$$

However, raw experimental data are usually unreported, and for many legacy experiments the underlying data are missing or lost forever

What is reported are limited (low-dimensional) summary statistics, $S(\mathbf{D}_0)$, of the data, typically in the form of nominal fitting model parameters, interval uncertainty

Also perhaps nominal information about the experiment and fitting context (fitting model form, experiment initial and boundary conditions, etc.)

This limited form of reporting often omits important information about the parametric uncertainty e.g. correlations, hindering uncertainty quantification efforts

Data-free inference (DFI)

Can we estimate the desired JPDF given the available information?

If the reported statistics of the data, $S(\mathbf{D}_0)$, are **sufficient** for the parameters λ, then computing $p(\lambda | S(\mathbf{D}_0))$ is equivalent to computing the desired $p(\lambda | \mathbf{D}_0)$.

The reported statistics $S(\mathbf{D}_0)$, are a unique mapping of the missing data. Estimating the inverse mapping would give access to plausible data sets, $\mathbf{D}$, consistent with these statistics, of which $\mathbf{D}_0$ is a member.

In a Bayesian framework, we can construct an inference to explore a space of data and associated parameters that satisfy the statistics, interpreted as constraints, with distribution $\Rightarrow p(\lambda, \mathbf{D} | S(\mathbf{D}_0))$

Applying Bayes' Law directly:

$$p(\lambda, \mathbf{D} | S(\mathbf{D}_0)) \propto p(S(\mathbf{D}_0) | \lambda, \mathbf{D})\, p(\lambda | \mathbf{D})\, p(\mathbf{D}) \quad (2)$$

$p(\lambda | \mathbf{D})$: can be computed if the fitting model is known

$p(S(\mathbf{D}_0) | \lambda, \mathbf{D})$: since statistics of the data are often reported as statistics of the uncertain fitting parameters λ, computing or estimating this likelihood for arbitrary $\mathbf{D}$ and λ is not straightforward

Data-free inference (DFI)

Can we estimate the desired JPDF given the available information?

If the reported statistics of the data, $S(\mathbf{D}_0)$, are **sufficient** for the parameters λ, then computing $p(\lambda|S(\mathbf{D}_0))$ is equivalent to computing the desired $p(\lambda|\mathbf{D}_0)$.

The reported statistics $S(\mathbf{D}_0)$, are a unique mapping of the missing data. Estimating the inverse mapping would give access to plausible data sets, **D**, consistent with these statistics, of which $\mathbf{D}_0$ is a member.

In a Bayesian framework, we can construct an inference to explore a space of data and associated parameters that satisfy the statistics, interpreted as constraints, with distribution $\Rightarrow p(\lambda, \mathbf{D}|S(\mathbf{D}_0))$

Applying Bayes' Law directly:

$$p(\lambda, \mathbf{D}|S(\mathbf{D}_0)) \propto p(S(\mathbf{D}_0)|\lambda, \mathbf{D})\, p(\lambda|\mathbf{D})\, p(\mathbf{D}) \quad (2)$$

$p(\lambda|\mathbf{D})$: can be computed if the fitting model is known

$p(S(\mathbf{D}_0)|\lambda, \mathbf{D})$: since statistics of the data are often reported as statistics of the uncertain fitting parameters λ, computing or estimating this likelihood for arbitrary **D** and λ is not straightforward

Data-free inference (DFI)

Can we estimate the desired JPDF given the available information?

If the reported statistics of the data, $S(\mathbf{D}_0)$, are **sufficient** for the parameters λ, then computing $p(\lambda|S(\mathbf{D}_0))$ is equivalent to computing the desired $p(\lambda|\mathbf{D}_0)$.

The reported statistics $S(\mathbf{D}_0)$, are a unique mapping of the missing data. Estimating the inverse mapping would give access to plausible data sets, $\mathbf{D}$, consistent with these statistics, of which $\mathbf{D}_0$ is a member.

In a Bayesian framework, we can construct an inference to explore a space of data and associated parameters that satisfy the statistics, interpreted as constraints, with distribution $\Rightarrow p(\lambda, \mathbf{D}|S(\mathbf{D}_0))$

Applying Bayes' Law directly:

$$p(\lambda, \mathbf{D}|S(\mathbf{D}_0)) \propto p(S(\mathbf{D}_0)|\lambda, \mathbf{D})\, p(\lambda|\mathbf{D})\, p(\mathbf{D}) \tag{2}$$

$p(\lambda|\mathbf{D})$: can be computed if the fitting model is known

$p(S(\mathbf{D}_0)|\lambda, \mathbf{D})$: since statistics of the data are often reported as statistics of the uncertain fitting parameters λ, computing or estimating this likelihood for arbitrary $\mathbf{D}$ and λ is not straightforward

Data-free inference (DFI)

Can we estimate the desired JPDF given the available information?

If the reported statistics of the data, $S(\mathbf{D}_0)$, are **sufficient** for the parameters λ, then computing $p(\lambda|S(\mathbf{D}_0))$ is equivalent to computing the desired $p(\lambda|\mathbf{D}_0)$.

The reported statistics $S(\mathbf{D}_0)$, are a unique mapping of the missing data. Estimating the inverse mapping would give access to plausible data sets, $\mathbf{D}$, consistent with these statistics, of which $\mathbf{D}_0$ is a member.

In a Bayesian framework, we can construct an inference to explore a space of data and associated parameters that satisfy the statistics, interpreted as constraints, with distribution $\Rightarrow p(\lambda, \mathbf{D}|S(\mathbf{D}_0))$

Applying Bayes' Law directly:

$$p(\lambda, \mathbf{D}|S(\mathbf{D}_0)) \propto p(S(\mathbf{D}_0)|\lambda, \mathbf{D})\, p(\lambda|\mathbf{D})\, p(\mathbf{D}) \quad (2)$$

$p(\lambda|\mathbf{D})$: can be computed if the fitting model is known

$p(S(\mathbf{D}_0)|\lambda, \mathbf{D})$: since statistics of the data are often reported as statistics of the uncertain fitting parameters λ, computing or estimating this likelihood for arbitrary $\mathbf{D}$ and λ is not straightforward

Data-free inference (DFI)

Can we estimate the desired JPDF given the available information?

If the reported statistics of the data, $S(\mathbf{D}_0)$, are **sufficient** for the parameters λ, then computing $p(\lambda|S(\mathbf{D}_0))$ is equivalent to computing the desired $p(\lambda|\mathbf{D}_0)$.

The reported statistics $S(\mathbf{D}_0)$, are a unique mapping of the missing data. Estimating the inverse mapping would give access to plausible data sets, $\mathbf{D}$, consistent with these statistics, of which $\mathbf{D}_0$ is a member.

In a Bayesian framework, we can construct an inference to explore a space of data and associated parameters that satisfy the statistics, interpreted as constraints, with distribution $\Rightarrow p(\lambda, \mathbf{D}|S(\mathbf{D}_0))$

Applying Bayes' Law directly:

$$p(\lambda, \mathbf{D}|S(\mathbf{D}_0)) \propto p(S(\mathbf{D}_0)|\lambda, \mathbf{D})\, p(\lambda|\mathbf{D})\, p(\mathbf{D}) \tag{2}$$

$p(\lambda|\mathbf{D})$: can be computed if the fitting model is known

$p(S(\mathbf{D}_0)|\lambda, \mathbf{D})$: since statistics of the data are often reported as statistics of the uncertain fitting parameters λ, computing or estimating this likelihood for arbitrary $\mathbf{D}$ and λ is not straightforward

Sandia National Laboratories

Data-free inference (DFI)

Can we estimate the desired JPDF given the available information?

If the reported statistics of the data, $S\left(\mathbf{D}_0\right)$, are **sufficient** for the parameters λ, then computing $p(\lambda|S\left(\mathbf{D}_0\right))$ is equivalent to computing the desired $p(\lambda|\mathbf{D}_0)$.

The reported statistics $S\left(\mathbf{D}_0\right)$, are a unique mapping of the missing data. Estimating the inverse mapping would give access to plausible data sets, $\mathbf{D}$, consistent with these statistics, of which $\mathbf{D}_0$ is a member.

In a Bayesian framework, we can construct an inference to explore a space of data and associated parameters that satisfy the statistics, interpreted as constraints, with distribution $\Rightarrow p(\lambda, \mathbf{D}|S\left(\mathbf{D}_0\right))$

Applying Bayes' Law directly:

$$p\left(\lambda, \mathbf{D}|S\left(\mathbf{D}_0\right)\right) \propto p\left(S\left(\mathbf{D}_0\right)|\lambda, \mathbf{D}\right) p\left(\lambda|\mathbf{D}\right) p(\mathbf{D}) \tag{2}$$

$p\left(\lambda|\mathbf{D}\right)$: can be computed if the fitting model is known

$p\left(S\left(\mathbf{D}_0\right)|\lambda, \mathbf{D}\right)$: since statistics of the data are often reported as statistics of the uncertain fitting parameters λ, computing or estimating this likelihood for arbitrary $\mathbf{D}$ and λ is not straightforward

Data-free inference: Two-step algorithm and Approximate Bayesian Computation (ABC)

Using a two-step approach, first constructing an inference on data:

$$p\left(\mathbf{D}|S\left(\mathbf{D}_0\right)\right) \propto p\left(S\left(\mathbf{D}_0\right)|\mathbf{D}\right)p(\mathbf{D}) \tag{3}$$

$p\left(S\left(\mathbf{D}_0\right)|\mathbf{D}\right)$: Approximate Bayesian Computation (ABC) approaches can be invoked to construct an approximate likelihood that enforces consistency of proposed data, $\mathbf{D}$, by comparing statistics of λ through the associated parameter posterior $p(\lambda|\mathbf{D})$ (which we can compute) with the reported statistics.

This requires constructing an inference on parameters:

$$p\left(\lambda|\mathbf{D}\right) \propto p(\mathbf{D}|\lambda)p(\lambda) \tag{4}$$

$p\left(S\left(\mathbf{D}_0\right)|\mathbf{D}\right)$ can then be designed to enforce consistency (approximately) e.g.:

$$p\left(S\left(\mathbf{D}_0\right)|\mathbf{D}\right) \approx \exp\left(-\delta\left(\frac{S\left(\mathbf{D}_0\right)-S\left(\mathbf{D}\right)}{S\left(\mathbf{D}_0\right)}\right)^2\right) \tag{5}$$

a Gaussian penalty kernel.

Data-free inference: Two-step algorithm and Approximate Bayesian Computation (ABC)

Using a two-step approach, first constructing an inference on data:

$$p\left(\mathbf{D} | S\left(\mathbf{D}_0\right)\right) \propto p\left(S\left(\mathbf{D}_0\right) | \mathbf{D}\right) p(\mathbf{D}) \tag{3}$$

$p\left(S\left(\mathbf{D}_0\right) | \mathbf{D}\right)$: Approximate Bayesian Computation (ABC) approaches can be invoked to construct an approximate likelihood that enforces consistency of proposed data, **D**, by comparing statistics of λ through the associated parameter posterior $p(\lambda | \mathbf{D})$ (which we can compute) with the reported statistics.

This requires constructing an inference on parameters:

$$p\left(\lambda | \mathbf{D}\right) \propto p(\mathbf{D} | \lambda) p(\lambda) \tag{4}$$

$p\left(S\left(\mathbf{D}_0\right) | \mathbf{D}\right)$ can then be designed to enforce consistency (approximately) e.g.:

$$p\left(S\left(\mathbf{D}_0\right) | \mathbf{D}\right) \approx \exp\left(-\delta\left(\frac{S\left(\mathbf{D}_0\right) - S\left(\mathbf{D}\right)}{S\left(\mathbf{D}_0\right)}\right)^2\right) \tag{5}$$

a Gaussian penalty kernel.

Data-free inference: Two-step algorithm and Approximate Bayesian Computation (ABC)

Using a two-step approach, first constructing an inference on data:

$$p\left(\mathbf{D}|S\left(\mathbf{D}_0\right)\right) \propto p\left(S\left(\mathbf{D}_0\right)|\mathbf{D}\right)p(\mathbf{D}) \quad (3)$$

$p\left(S\left(\mathbf{D}_0\right)|\mathbf{D}\right)$: Approximate Bayesian Computation (ABC) approaches can be invoked to construct an approximate likelihood that enforces consistency of proposed data, **D**, by comparing statistics of λ through the associated parameter posterior $p(\lambda|\mathbf{D})$ (which we can compute) with the reported statistics.

This requires constructing an inference on parameters:

$$p\left(\lambda|\mathbf{D}\right) \propto p(\mathbf{D}|\lambda)p(\lambda) \quad (4)$$

$p\left(S\left(\mathbf{D}_0\right)|\mathbf{D}\right)$ can then be designed to enforce consistency (approximately) e.g.:

$$p\left(S\left(\mathbf{D}_0\right)|\mathbf{D}\right) \approx \exp\left(-\delta\left(\frac{S\left(\mathbf{D}_0\right)-S\left(\mathbf{D}\right)}{S\left(\mathbf{D}_0\right)}\right)^2\right) \quad (5)$$

a Gaussian penalty kernel.

Data-free inference: Two-step algorithm and Approximate Bayesian Computation (ABC)

Using a two-step approach, first constructing an inference on data:

$$p\left(\mathbf{D}|S\left(\mathbf{D}_0\right)\right) \propto p\left(S\left(\mathbf{D}_0\right)|\mathbf{D}\right) p(\mathbf{D}) \quad (3)$$

$p\left(S\left(\mathbf{D}_0\right)|\mathbf{D}\right)$: Approximate Bayesian Computation (ABC) approaches can be invoked to construct an approximate likelihood that enforces consistency of proposed data, **D**, by comparing statistics of λ through the associated parameter posterior $p(\lambda|\mathbf{D})$ (which we can compute) with the reported statistics.

This requires constructing an inference on parameters:

$$p\left(\lambda|\mathbf{D}\right) \propto p(\mathbf{D}|\lambda)p(\lambda) \quad (4)$$

$p\left(S\left(\mathbf{D}_0\right)|\mathbf{D}\right)$ can then be designed to enforce consistency (approximately) e.g.:

$$p\left(S\left(\mathbf{D}_0\right)|\mathbf{D}\right) \approx \exp\left(-\delta\left(\frac{S\left(\mathbf{D}_0\right) - S\left(\mathbf{D}\right)}{S\left(\mathbf{D}_0\right)}\right)^2\right) \quad (5)$$

a Gaussian penalty kernel.

Data-free inference: Maximum Entropy (MaxEnt)

With this nested inference construction we can generate samples from $p\left(\lambda, \mathbf{D} | S\left(\mathbf{D}_0\right)\right)$, with $\mathbf{D}$ from the outer inference on data and λ from the inner inference on parameters

Computing $p(\lambda | S\left(\mathbf{D}_0\right))$ is then achieved by marginalizing over data-space:

$$p(\lambda | S\left(\mathbf{D}_0\right)) = \int_{\mathbf{D}} p\left(\lambda, \mathbf{D}' | S\left(\mathbf{D}_0\right)\right) d\mathbf{D}' = \int_{\mathbf{D}} p\left(\lambda | \mathbf{D}'\right) p(\mathbf{D}' | S\left(\mathbf{D}_0\right)) d\mathbf{D}' \quad (6)$$

corresponding to linear pooling of consistent parameter posteriors.

From an entropy perspective of Bayesian updating, the posterior $p(\lambda | \mathbf{D}_0)$ maximizes entropy relative to the prior.

For situations where the posterior conditioning variable is described by a distribution, the relative-entropy maximizing distribution corresponds to linear pooling of posteriors according to that distribution. This is precisely the case with the marginalization in Eq. 6, making $p(\lambda | S\left(\mathbf{D}_0\right))$ the maximum entropy (MaxEnt) density given the available information

Data-free inference: Maximum Entropy (MaxEnt)

With this nested inference construction we can generate samples from $p\left(\lambda, \mathbf{D} | S\left(\mathbf{D}_0\right)\right)$, with $\mathbf{D}$ from the outer inference on data and λ from the inner inference on parameters

Computing $p(\lambda | S\left(\mathbf{D}_0\right))$ is then achieved by marginalizing over data-space:

$$p(\lambda | S\left(\mathbf{D}_0\right)) = \int_{\mathbf{D}} p\left(\lambda, \mathbf{D}' | S\left(\mathbf{D}_0\right)\right) d\mathbf{D}' = \int_{\mathbf{D}} p\left(\lambda | \mathbf{D}'\right) p(\mathbf{D}' | S\left(\mathbf{D}_0\right)) d\mathbf{D}' \quad (6)$$

corresponding to linear pooling of consistent parameter posteriors.

From an entropy perspective of Bayesian updating, the posterior $p(\lambda | \mathbf{D}_0)$ maximizes entropy relative to the prior.

For situations where the posterior conditioning variable is described by a distribution, the relative-entropy maximizing distribution corresponds to linear pooling of posteriors according to that distribution. This is precisely the case with the marginalization in Eq. 6, making $p(\lambda | S\left(\mathbf{D}_0\right))$ the maximum entropy (MaxEnt) density given the available information

Data-free inference: Maximum Entropy (MaxEnt)

With this nested inference construction we can generate samples from $p\left(\lambda, \mathbf{D}|S\left(\mathbf{D}_0\right)\right)$, with $\mathbf{D}$ from the outer inference on data and λ from the inner inference on parameters

Computing $p(\lambda|S\left(\mathbf{D}_0\right))$ is then achieved by marginalizing over data-space:

$$p(\lambda|S\left(\mathbf{D}_0\right)) = \int_{\mathbf{D}} p\left(\lambda, \mathbf{D}'|S\left(\mathbf{D}_0\right)\right) d\mathbf{D}' = \int_{\mathbf{D}} p\left(\lambda|\mathbf{D}'\right) p(\mathbf{D}'|S\left(\mathbf{D}_0\right))d\mathbf{D}' \quad (6)$$

corresponding to linear pooling of consistent parameter posteriors.

From an entropy perspective of Bayesian updating, the posterior $p(\lambda|\mathbf{D}_0)$ maximizes entropy relative to the prior.

For situations where the posterior conditioning variable is described by a distribution, the relative-entropy maximizing distribution corresponds to linear pooling of posteriors according to that distribution. This is precisely the case with the marginalization in Eq. 6, making $p(\lambda|S\left(\mathbf{D}_0\right))$ the maximum entropy (MaxEnt) density given the available information

Data-free inference: Maximum Entropy (MaxEnt)

With this nested inference construction we can generate samples from $p\left(\lambda, \mathbf{D}|S\left(\mathbf{D}_0\right)\right)$, with $\mathbf{D}$ from the outer inference on data and λ from the inner inference on parameters

Computing $p(\lambda|S\left(\mathbf{D}_0\right))$ is then achieved by marginalizing over data-space:

$$p(\lambda|S\left(\mathbf{D}_0\right)) = \int_{\mathbf{D}} p\left(\lambda, \mathbf{D}'|S\left(\mathbf{D}_0\right)\right) d\mathbf{D}' = \int_{\mathbf{D}} p\left(\lambda|\mathbf{D}'\right) p(\mathbf{D}'|S\left(\mathbf{D}_0\right))d\mathbf{D}' \quad (6)$$

corresponding to linear pooling of consistent parameter posteriors.

From an entropy perspective of Bayesian updating, the posterior $p(\lambda|\mathbf{D}_0)$ maximizes entropy relative to the prior.

For situations where the posterior conditioning variable is described by a distribution, the relative-entropy maximizing distribution corresponds to linear pooling of posteriors according to that distribution. This is precisely the case with the marginalization in Eq. 6, making $p(\lambda|S\left(\mathbf{D}_0\right))$ the maximum entropy (MaxEnt) density given the available information

Example - line fit to noisy data

Consider the case of noisy data, fit with a simple line model assuming zero-mean additive *iid* Gaussian noise. The noisy data model is:

$$y(x) = mx + c + \sigma\varepsilon \tag{7}$$

where ε is a standard normally distributed random variable, σ is the standard deviation of the noise, and the model parameters are $\lambda = \{m, c\}$

This stochastic data model implies a data likelihood:

$$p(\mathbf{D}|\lambda) = \frac{1}{\sqrt{2\pi\sigma}} \prod_i \exp\left(-\frac{(y(x_i) - [mx_i + c])^2}{2\sigma^2}\right) \tag{8}$$

which enables construction of an inference on λ given data

Since the true data $\mathbf{D}_0$, or the posterior $p(\lambda|\mathbf{D}_0)$, are typically unreported, the DFI construction employs the available information to approximate them

Example - line fit to noisy data

Consider the case of noisy data, fit with a simple line model assuming zero-mean additive *iid* Gaussian noise. The noisy data model is:

$$y(x) = mx + c + \sigma\varepsilon \tag{7}$$

where ε is a standard normally distributed random variable, σ is the standard deviation of the noise, and the model parameters are $\lambda = \{m, c\}$

This stochastic data model implies a data likelihood:

$$p(\mathbf{D}|\lambda) = \frac{1}{\sqrt{2\pi\sigma}} \prod_i \exp\left(-\frac{(y(x_i) - [mx_i + c])^2}{2\sigma^2}\right) \tag{8}$$

which enables construction of an inference on λ given data

Since the true data $\mathbf{D}_0$, or the posterior $p(\lambda|\mathbf{D}_0)$, are typically unreported, the DFI construction employs the available information to approximate them

Example - line fit to noisy data

Consider the case of noisy data, fit with a simple line model assuming zero-mean additive *iid* Gaussian noise. The noisy data model is:

$$y(x) = mx + c + \sigma\varepsilon \tag{7}$$

where ε is a standard normally distributed random variable, σ is the standard deviation of the noise, and the model parameters are $\lambda = \{m, c\}$

This stochastic data model implies a data likelihood:

$$p(\mathbf{D}|\lambda) = \frac{1}{\sqrt{2\pi\sigma}} \prod_i \exp\left(-\frac{(y(x_i) - [mx_i + c])^2}{2\sigma^2}\right) \tag{8}$$

which enables construction of an inference on λ given data

Since the true data $\mathbf{D}_0$, or the posterior $p(\lambda|\mathbf{D}_0)$, are typically unreported, the DFI construction employs the available information to approximate them

Example - line fit to noisy data

i.e:

$$S(\mathbf{D}_0) = \{\mu_m, \mu_c, \sigma_m, \sigma_c\} \tag{9}$$

The two-step DFI construction, outer (data) inference:

$$p(\mathbf{D}|S(\mathbf{D}_0)) = \frac{p(S(\mathbf{D}_0)|\mathbf{D})p(\mathbf{D})}{p(S(\mathbf{D}_0))} \tag{10}$$

where $p(S(\mathbf{D}_0)|\mathbf{D})$ is constructed using ABC:

$$p\left(S(\mathbf{D}_0)|\mathbf{D}\right) = \prod_k \exp\left(-\delta_k\left(\frac{S_k\left(\mathbf{D}_0\right) - S_k\left(\mathbf{D}\right)}{S_k\left(\mathbf{D}_0\right)}\right)^2\right) \tag{11}$$

and the statistics of the proposed data, $S_k\left(\mathbf{D}\right)$ are determined by solving an inner (parameter) inference:

$$p(\lambda|\mathbf{D}) = \frac{p(\mathbf{D}|\lambda)p(\lambda)}{p(\mathbf{D})} \tag{12}$$

Both outer and inner inferences performed numerically using Markov Chain Monte Carlo methods to compute samples $\{\mathbf{D}, \lambda\}$

Example - line fit to noisy data

i.e:

$$S(\mathbf{D}_0) = \{\mu_m, \mu_c, \sigma_m, \sigma_c\} \tag{9}$$

The two-step DFI construction, outer (data) inference:

$$p(\mathbf{D}|S(\mathbf{D}_0)) = \frac{p(S(\mathbf{D}_0)|\mathbf{D})p(\mathbf{D})}{p(S(\mathbf{D}_0))} \tag{10}$$

where $p(S(\mathbf{D}_0)|\mathbf{D})$ is constructed using ABC:

$$p\left(S(\mathbf{D}_0)|\mathbf{D}\right) = \prod_k \exp\left(-\delta_k\left(\frac{S_k\left(\mathbf{D}_0\right) - S_k\left(\mathbf{D}\right)}{S_k\left(\mathbf{D}_0\right)}\right)^2\right) \tag{11}$$

and the statistics of the proposed data, $S_k\left(\mathbf{D}\right)$ are determined by solving an inner (parameter) inference:

$$p(\lambda|\mathbf{D}) = \frac{p(\mathbf{D}|\lambda)p(\lambda)}{p(\mathbf{D})} \tag{12}$$

Both outer and inner inferences performed numerically using Markov Chain Monte Carlo methods to compute samples $\{\mathbf{D}, \lambda\}$

Example - line fit to noisy data

i.e:

$$S(\mathbf{D}_0) = \{\mu_m, \mu_c, \sigma_m, \sigma_c\} \tag{9}$$

The two-step DFI construction, outer (data) inference:

$$p(\mathbf{D}|S(\mathbf{D}_0)) = \frac{p(S(\mathbf{D}_0)|\mathbf{D})p(\mathbf{D})}{p(S(\mathbf{D}_0))} \tag{10}$$

where $p(S(\mathbf{D}_0)|\mathbf{D})$ is constructed using ABC:

$$p\left(S(\mathbf{D}_0)|\mathbf{D}\right) = \prod_k \exp\left(-\delta_k \left(\frac{S_k\left(\mathbf{D}_0\right) - S_k\left(\mathbf{D}\right)}{S_k\left(\mathbf{D}_0\right)}\right)^2\right) \tag{11}$$

and the statistics of the proposed data, $S_k\left(\mathbf{D}\right)$ are determined by solving an inner (parameter) inference:

$$p(\lambda|\mathbf{D}) = \frac{p(\mathbf{D}|\lambda)p(\lambda)}{p(\mathbf{D})} \tag{12}$$

Both outer and inner inferences performed numerically using Markov Chain Monte Carlo methods to compute samples $\{\mathbf{D}, \lambda\}$

Example - line fit to noisy data

i.e:

$$S(\mathbf{D}_0) = \{\mu_m, \mu_c, \sigma_m, \sigma_c\} \tag{9}$$

The two-step DFI construction, outer (data) inference:

$$p(\mathbf{D}|S(\mathbf{D}_0)) = \frac{p(S(\mathbf{D}_0)|\mathbf{D})p(\mathbf{D})}{p(S(\mathbf{D}_0))} \tag{10}$$

where $p(S(\mathbf{D}_0)|\mathbf{D})$ is constructed using ABC:

$$p\left(S(\mathbf{D}_0)|\mathbf{D}\right) = \prod_k \exp\left(-\delta_k\left(\frac{S_k\left(\mathbf{D}_0\right) - S_k\left(\mathbf{D}\right)}{S_k\left(\mathbf{D}_0\right)}\right)^2\right) \tag{11}$$

and the statistics of the proposed data, $S_k\left(\mathbf{D}\right)$ are determined by solving an inner (parameter) inference:

$$p(\lambda|\mathbf{D}) = \frac{p(\mathbf{D}|\lambda)p(\lambda)}{p(\mathbf{D})} \tag{12}$$

Both outer and inner inferences performed numerically using Markov Chain Monte Carlo methods to compute samples $\{\mathbf{D}, \lambda\}$

Example - line fit to noisy data

Define the data space, number of data points (e.g. defined in experiment)

Initialize the data in a region of high likelihood

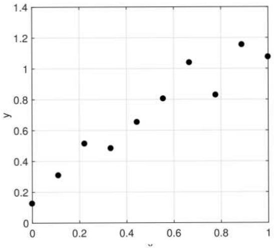
Typically run the outer data inference to discover 1000s of data sets consistent with the statistics

Linear pooling achieved by averaging over all consistent parameter posteriors

- Is the statistics are sufficient, each consistent posterior should be identical, but some discrepancy from numerical approximations
- Each individual consistent posterior constructed from consistent parameter samples using Kernel Density Estimation (KDE)

Result is the full approximation of the missing posterior, $p(\lambda|S(\mathbf{D}_0))$

sample noisy data set (**D**) from outer inference



Example - line fit to noisy data

Define the data space, number of data points (e.g. defined in experiment)

Initialize the data in a region of high likelihood

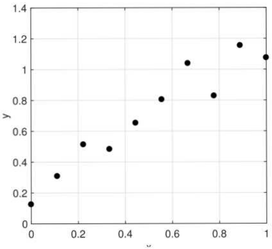
Typically run the outer data inference to discover 1000s of data sets consistent with the statistics

Linear pooling achieved by averaging over all consistent parameter posteriors

- Is the statistics are sufficient, each consistent posterior should be identical, but some discrepancy from numerical approximations
- Each individual consistent posterior constructed from consistent parameter samples using Kernel Density Estimation (KDE)

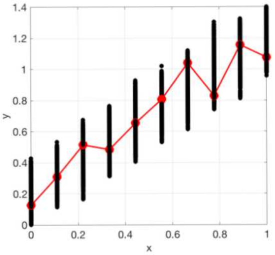
Result is the full approximation of the missing posterior, $p(\lambda | S(\mathbf{D}_0))$

sample noisy data set (**D**) from outer inference



Example - line fit to noisy data

10,000 data sets (**D**) from outer DFI inference



Define the data space, number of data points (e.g. defined in experiment)

Initialize the data in a region of high likelihood

Typically run the outer data inference to discover 1000s of data sets consistent with the statistics

Linear pooling achieved by averaging over all consistent parameter posteriors

- Is the statistics are sufficient, each consistent posterior should be identical, but some discrepancy from numerical approximations
- Each individual consistent posterior constructed from consistent parameter samples using Kernel Density Estimation (KDE)

Result is the full approximation of the missing posterior, $p(\lambda|S(\mathbf{D}_0))$

Example - line fit to noisy data

Define the data space, number of data points (e.g. defined in experiment)

Initialize the data in a region of high likelihood

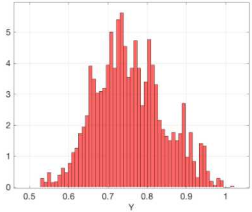
Typically run the outer data inference to discover 1000s of data sets consistent with the statistics

Linear pooling achieved by averaging over all consistent parameter posteriors

- Is the statistics are sufficient, each consistent posterior should be identical, but some discrepancy from numerical approximations
- Each individual consistent posterior constructed from consistent parameter samples using Kernel Density Estimation (KDE)

Result is the full approximation of the missing posterior, $p(\lambda|S(\mathbf{D}_0))$

histogram of data samples for $\mathbf{D}(x = 5\Delta x)$



Example - line fit to noisy data

Define the data space, number of data points (e.g. defined in experiment)

Initialize the data in a region of high likelihood

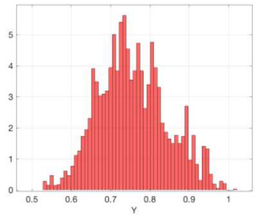
Typically run the outer data inference to discover 1000s of data sets consistent with the statistics

Linear pooling achieved by averaging over all consistent parameter posteriors

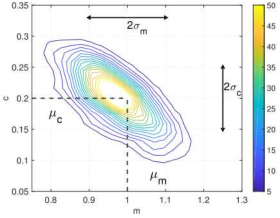
- Is the statistics are sufficient, each consistent posterior should be identical, but some discrepancy from numerical approximations
- Each individual consistent posterior constructed from consistent parameter samples using Kernel Density Estimation (KDE)

Result is the full approximation of the missing posterior, $p(\lambda|S(\mathbf{D}_0))$

histogram of data samples for $\mathbf{D}(x = \Delta x)$



pooled posterior $p(\lambda|S)$



Example - line fit to noisy data

Define the data space, number of data points (e.g. defined in experiment)

Initialize the data in a region of high likelihood

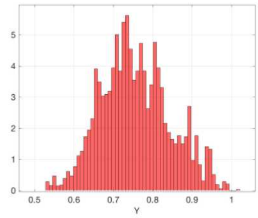
Typically run the outer data inference to discover 1000s of data sets consistent with the statistics

Linear pooling achieved by averaging over all consistent parameter posteriors

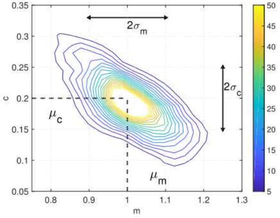
- Is the statistics are sufficient, each consistent posterior should be identical, but some discrepancy from numerical approximations
- Each individual consistent posterior constructed from consistent parameter samples using Kernel Density Estimation (KDE)

Result is the full approximation of the missing posterior, $p(\lambda|S(\mathbf{D}_0))$

histogram of data samples for $\mathbf{D}(x = \Delta x)$



pooled posterior $p(\lambda|S)$



Example - line fit to noisy data

Define the data space, number of data points (e.g. defined in experiment)

Initialize the data in a region of high likelihood

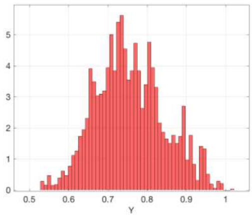
Typically run the outer data inference to discover 1000s of data sets consistent with the statistics

Linear pooling achieved by averaging over all consistent parameter posteriors

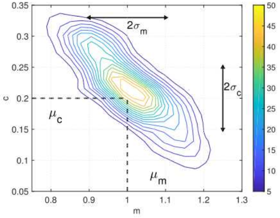
- Is the statistics are sufficient, each consistent posterior should be identical, but some discrepancy from numerical approximations
- Each individual consistent posterior constructed from consistent parameter samples using Kernel Density Estimation (KDE)

Result is the full approximation of the missing posterior, $p(\lambda|S(\mathbf{D}_0))$

histogram of data samples for $\mathbf{D}(x = \Delta x)$



pooled posterior $p(\lambda|S)$



Example - line fit to noisy data

Define the data space, number of data points (e.g. defined in experiment)

Initialize the data in a region of high likelihood

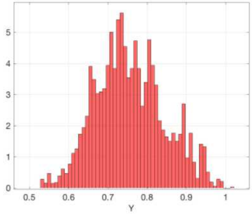
Typically run the outer data inference to discover 1000s of data sets consistent with the statistics

Linear pooling achieved by averaging over all consistent parameter posteriors

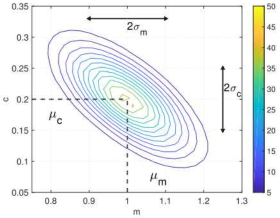
- Is the statistics are sufficient, each consistent posterior should be identical, but some discrepancy from numerical approximations
- Each individual consistent posterior constructed from consistent parameter samples using Kernel Density Estimation (KDE)

Result is the full approximation of the missing posterior, $p(\lambda|S(\mathbf{D}_0))$

histogram of data samples for $\mathbf{D}(x = \Delta x)$



pooled posterior $p(\lambda|S)$



Data-free inference: Model Replacement, Data Fusion

$$p(\lambda|S) = \int_{\mathbf{D}} p\left(\lambda, \mathbf{D}'|S\right) d\mathbf{D}' = \int_{\mathbf{D}} p\left(\lambda|\mathbf{D}'\right) p(\mathbf{D}'|S) d\mathbf{D}' \tag{13}$$

MaxEnt parameter posterior corresponding to linear pooling of consistent parameter posteriors.

Model replacement: linearly pooling over a different posterior

$$p(\theta|S) = \int_{\mathbf{D}} p\left(\theta, \mathbf{D}'|S\right) d\mathbf{D}' = \int_{\mathbf{D}} p\left(\theta|\mathbf{D}'\right) p(\mathbf{D}'|S) d\mathbf{D}' \tag{14}$$

Why do this? Desire a different fitting context than that used by the experimentalists

Data fusion: pooling over a posterior computed using multiple consistent data sources, e.g. from different reported experiments

$$\begin{aligned} p(\theta|S_1, S_2) &= \int_{\mathbf{D}_1, \mathbf{D}_2} p\left(\theta, \mathbf{D}'_1, \mathbf{D}'_2|S_1, S_2\right) d\mathbf{D}'_1 d\mathbf{D}'_2 \\ &= \int_{\mathbf{D}_1, \mathbf{D}_2} p\left(\theta|\mathbf{D}'_1, \mathbf{D}'_2\right) p(\mathbf{D}'_1|S_1) p(\mathbf{D}'_2|S_2) d\mathbf{D}'_1 d\mathbf{D}'_2 \end{aligned} \tag{15}$$

Data-free inference: Model Replacement, Data Fusion

$$p(\lambda|S) = \int_{\mathbf{D}} p\left(\lambda, \mathbf{D}'|S\right) d\mathbf{D}' = \int_{\mathbf{D}} p\left(\lambda|\mathbf{D}'\right) p(\mathbf{D}'|S) d\mathbf{D}' \quad (13)$$

MaxEnt parameter posterior corresponding to linear pooling of consistent parameter posteriors.

Model replacement: linearly pooling over a different posterior

$$p(\theta|S) = \int_{\mathbf{D}} p\left(\theta, \mathbf{D}'|S\right) d\mathbf{D}' = \int_{\mathbf{D}} p\left(\theta|\mathbf{D}'\right) p(\mathbf{D}'|S) d\mathbf{D}' \quad (14)$$

Why do this? Desire a different fitting context than that used by the experimentalists

Data fusion: pooling over a posterior computed using multiple consistent data sources, e.g. from different reported experiments

$$\begin{aligned} p(\theta|S_1, S_2) &= \int_{\mathbf{D}_1, \mathbf{D}_2} p\left(\theta, \mathbf{D}_1', \mathbf{D}_2'|S_1, S_2\right) d\mathbf{D}_1' d\mathbf{D}_2' \\ &= \int_{\mathbf{D}_1, \mathbf{D}_2} p\left(\theta|\mathbf{D}_1', \mathbf{D}_2'\right) p(\mathbf{D}_1'|S_1) p(\mathbf{D}_2'|S_2) d\mathbf{D}_1' d\mathbf{D}_2' \end{aligned} \quad (15)$$

Data-free inference: Model Replacement, Data Fusion

$$p(\lambda|S) = \int_{\mathbf{D}} p\left(\lambda, \mathbf{D}'|S\right) d\mathbf{D}' = \int_{\mathbf{D}} p\left(\lambda|\mathbf{D}'\right) p(\mathbf{D}'|S) d\mathbf{D}' \tag{13}$$

MaxEnt parameter posterior corresponding to linear pooling of consistent parameter posteriors.

Model replacement: linearly pooling over a different posterior

$$p(\theta|S) = \int_{\mathbf{D}} p\left(\theta, \mathbf{D}'|S\right) d\mathbf{D}' = \int_{\mathbf{D}} p\left(\theta|\mathbf{D}'\right) p(\mathbf{D}'|S) d\mathbf{D}' \tag{14}$$

Why do this? Desire a different fitting context than that used by the experimentalists

Data fusion: pooling over a posterior computed using multiple consistent data sources, e.g. from different reported experiments

$$\begin{aligned} p(\theta|S_1, S_2) &= \int_{\mathbf{D}_1, \mathbf{D}_2} p\left(\theta, \mathbf{D}_1', \mathbf{D}_2'|S_1, S_2\right) d\mathbf{D}_1' d\mathbf{D}_2' \\ &= \int_{\mathbf{D}_1, \mathbf{D}_2} p\left(\theta|\mathbf{D}_1', \mathbf{D}_2'\right) p(\mathbf{D}_1'|S_1) p(\mathbf{D}_2'|S_2) d\mathbf{D}_1' d\mathbf{D}_2' \end{aligned} \tag{15}$$

Hierarchical models

Often we are interested in analyzing data at different conditions (e.g. temperature) to construct hierarchical models that span these conditions.

- Chemical kinetics experiments: noisy exponential decay data model at different temperatures

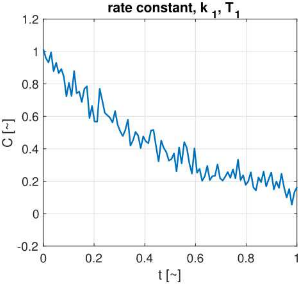
$$c(t) = \exp\left(-k(T)t\right) + \sigma\varepsilon \tag{16}$$

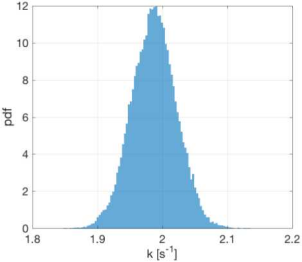
- Hierarchical (Arrhenius) model couples the exponential decay constants across temperatures

$$k(T) = A\exp\left(-\frac{E}{T}\right) \tag{17}$$

DFI approach:

- Infer consistent data for the reported statistics of $k(T)$





Hierarchical models

Often we are interested in analyzing data at different conditions (e.g. temperature) to construct hierarchical models that span these conditions.

Chemical kinetics experiments: noisy exponential decay data model at different temperatures

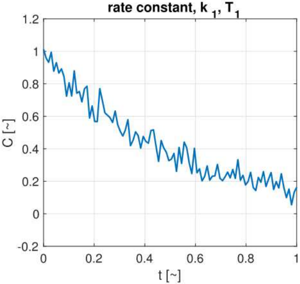
$$c(t) = \exp\left(-k(T)t\right) + \sigma\varepsilon \qquad (16)$$

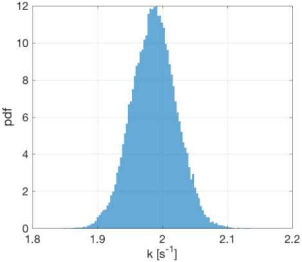
Hierarchical (Arrhenius) model couples the exponential decay constants across temperatures

$$k(T) = A\exp\left(-\frac{E}{T}\right) \qquad (17)$$

DFI approach:

Infer consistent data for the reported statistics of $k(T)$





Hierarchical models

Often we are interested in analyzing data at different conditions (e.g. temperature) to construct hierarchical models that span these conditions.

Chemical kinetics experiments: noisy exponential decay data model at different temperatures

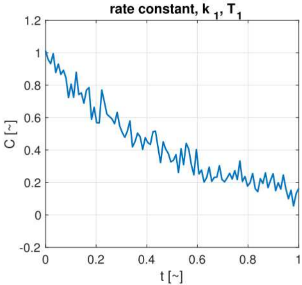
$$c(t) = \exp\left(-k(T)t\right) + \sigma\varepsilon \quad (16)$$

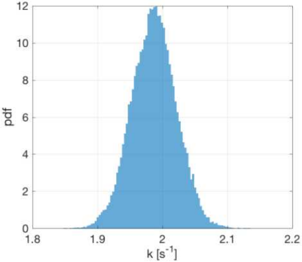
Hierarchical (Arrhenius) model couples the exponential decay constants across temperatures

$$k(T) = A\exp\left(-\frac{E}{T}\right) \quad (17)$$

DFI approach:

Infer consistent data for the reported statistics of $k(T)$





Hierarchical models

Often we are interested in analyzing data at different conditions (e.g. temperature) to construct hierarchical models that span these conditions.

Chemical kinetics experiments: noisy exponential decay data model at different temperatures

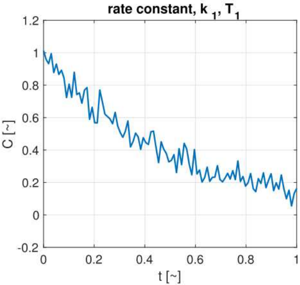
$$c(t) = \exp\left(-k(T)t\right) + \sigma\varepsilon \qquad (16)$$

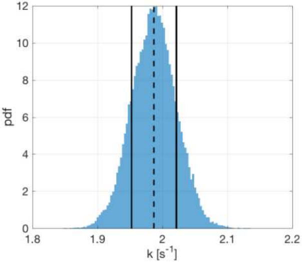
Hierarchical (Arrhenius) model couples the exponential decay constants across temperatures

$$k(T) = A\exp\left(-\frac{E}{T}\right) \qquad (17)$$

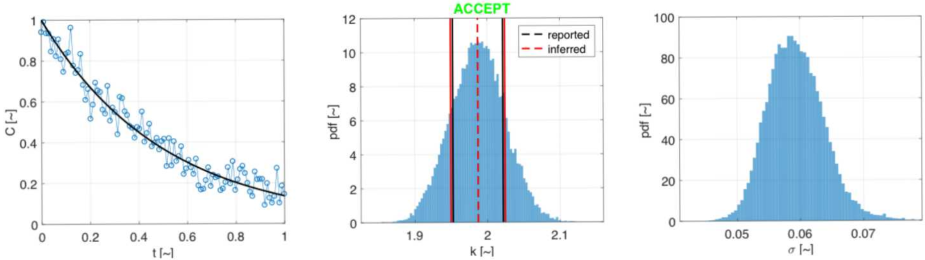
DFI approach:

Infer consistent data for the reported statistics of $k(T)$

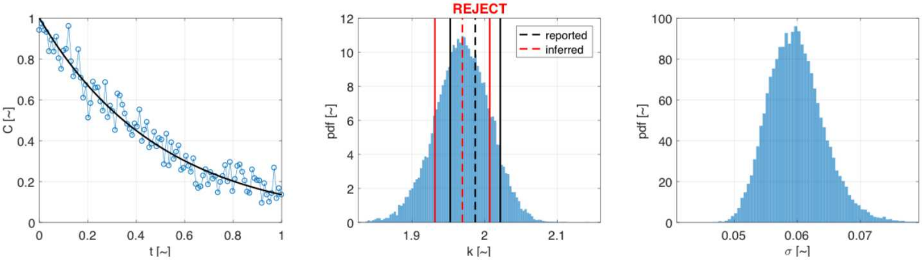




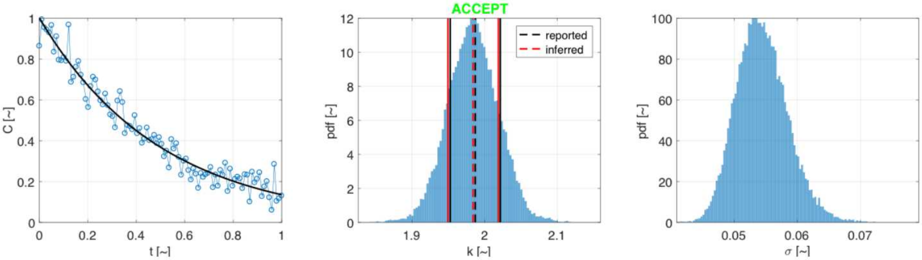
Hierarchical models



Hierarchical models



Hierarchical models



Hierarchical models

Infer an ensemble of data consistent with the available information

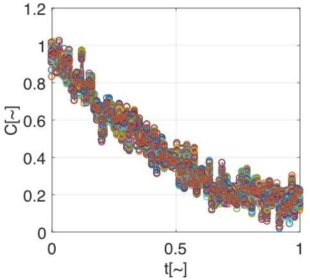
- Could pool over the $k(T)$ posteriors
- But really want to calibrate the hierarchical model across temperatures

$$c_2(t) = \exp\left(-k(T_2)t\right) + \sigma\varepsilon \tag{18}$$

- Infer data for this second experiment, given statistics on $k(T_2)$

$$k(T) = A\exp\left(-\frac{E}{T}\right) \tag{19}$$

Use data fusion and model replacement to learn the posterior on the parameters of the Arrhenius (hierarchical) model, $p(\log A, E|S_1, S_2)$



Hierarchical models

Infer an ensemble of data consistent with the available information

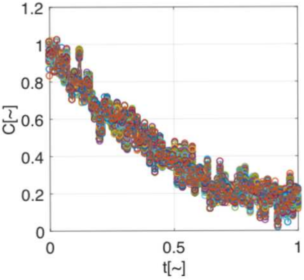
- Could pool over the $k(T)$ posteriors
- But really want to calibrate the hierarchical model across temperatures

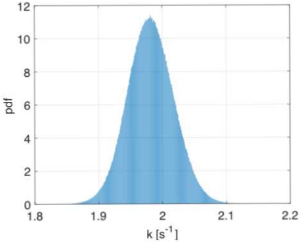
$$c_2(t) = \exp\left(-k(T_2)t\right) + \sigma\varepsilon \qquad (18)$$

- Infer data for this second experiment, given statistics on $k(T_2)$

$$k(T) = A\exp\left(-\frac{E}{T}\right) \qquad (19)$$

Use data fusion and model replacement to learn the posterior on the parameters of the Arrhenius (hierarchical) model, $p(\log A, E|S_1, S_2)$



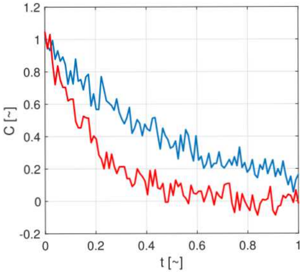


Hierarchical models

Infer an ensemble of data consistent with the available information

- Could pool over the $k(T)$ posteriors
- But really want to calibrate the hierarchical model across temperatures

$$c_2(t) = \exp\left(-k(T_2)t\right) + \sigma\varepsilon \qquad (18)$$



- Infer data for this second experiment, given statistics on $k(T_2)$

$$k(T) = A\exp\left(-\frac{E}{T}\right) \qquad (19)$$

Use data fusion and model replacement to learn the posterior on the parameters of the Arrhenius (hierarchical) model, $p(\log A, E|S_1, S_2)$

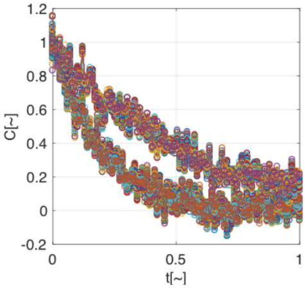
Hierarchical models

Infer an ensemble of data consistent with the available information

- Could pool over the $k(T)$ posteriors
- But really want to calibrate the hierarchical model across temperatures

$$c_2(t) = \exp\left(-k(T_2)t\right) + \sigma\varepsilon \tag{18}$$

- Infer data for this second experiment, given statistics on $k(T_2)$



$$k(T) = A\exp\left(-\frac{E}{T}\right) \tag{19}$$

Use data fusion and model replacement to learn the posterior on the parameters of the Arrhenius (hierarchical) model, $p(\log A, E|S_1, S_2)$

Hierarchical models

Infer an ensemble of data consistent with the available information

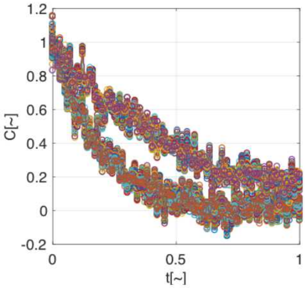
- Could pool over the $k(T)$ posteriors
- But really want to calibrate the hierarchical model across temperatures

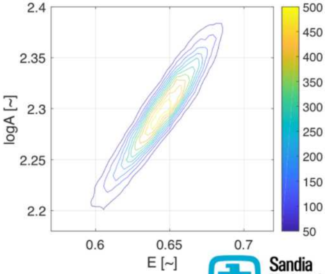
$$c_2(t) = \exp\left(-k(T_2)t\right) + \sigma\varepsilon \qquad (18)$$

- Infer data for this second experiment, given statistics on $k(T_2)$

$$k(T) = A\exp\left(-\frac{E}{T}\right) \qquad (19)$$

Use data fusion and model replacement to learn the posterior on the parameters of the Arrhenius (hierarchical) model, $p(\log A, E|S_1, S_2)$

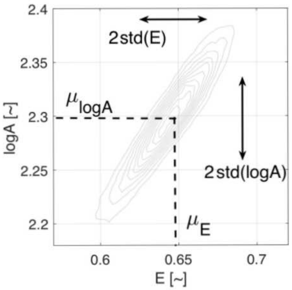




Hierarchical models

What if the statistics are reported on the parameters of the hierarchical model, instead of on each $k(T)$? We can infer data at all experimental conditions using these statistics as constraints on the data at all experiment temperatures.
e.g.

$$k(T) = A \pm \Delta A \exp\left(-\frac{E \pm \Delta E}{T}\right) \tag{20}$$
$$S = \{\mu_{\log A}, \mu_E, \sigma_{\log A}, \sigma_E\}$$



Posterior recovery fails (w.r.t. correlation) fails when we lose information about uncertain parameters that couple data across experiments, i.e. error in E

Hierarchical models

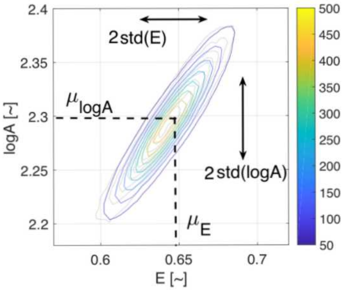
What if the statistics are reported on the parameters of the hierarchical model, instead of on each $k(T)$? We can infer data at all experimental conditions using these statistics as constraints on the data at all experiment temperatures.

e.g.

$$k(T) = A \pm \Delta A \exp\left(-\frac{E \pm \Delta E}{T}\right) \tag{20}$$

$$S = \{\mu_{\log A}, \mu_E, \sigma_{\log A}, \sigma_E\}$$

Posterior recovery fails (w.r.t. correlation) fails when we lose information about uncertain parameters that couple data across experiments, i.e. error in E

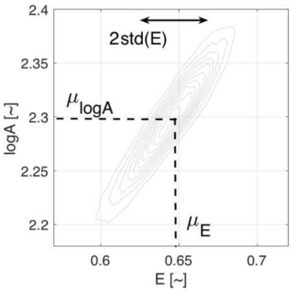


Hierarchical models

What if the statistics are reported on the parameters of the hierarchical model, instead of on each $k(T)$? We can infer data at all experimental conditions using these statistics as constraints on the data at all experiment temperatures.

e.g.

$$k(T) = A\text{exp}\left(-\frac{E \pm \Delta E}{T}\right) \tag{20}$$
$$S = \{\mu_{\log A}, \mu_E, \sigma_E\}$$



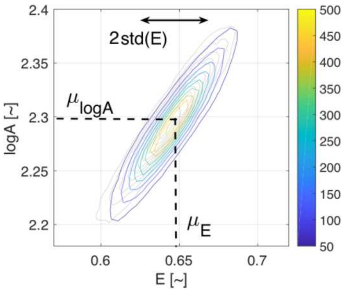
Posterior recovery fails (w.r.t. correlation) fails when we lose information about uncertain parameters that couple data across experiments, i.e. error in E

Hierarchical models

What if the statistics are reported on the parameters of the hierarchical model, instead of on each $k(T)$? We can infer data at all experimental conditions using these statistics as constraints on the data at all experiment temperatures.

e.g.

$$k(T) = A\text{exp}\left(-\frac{E \pm \Delta E}{T}\right) \tag{20}$$
$$S = \{\mu_{\log A}, \mu_E, \sigma_E\}$$



Posterior recovery fails (w.r.t. correlation) fails when we lose information about uncertain parameters that couple data across experiments, i.e. error in E

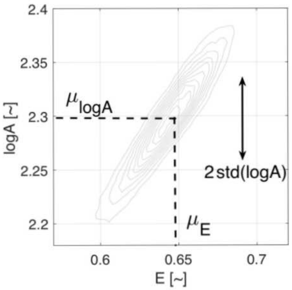
Hierarchical models

What if the statistics are reported on the parameters of the hierarchical model, instead of on each $k(T)$? We can infer data at all experimental conditions using these statistics as constraints on the data at all experiment temperatures.

e.g.

$$k(T) = A \pm \Delta A \text{exp}\left(-\frac{E}{T}\right) \qquad (20)$$

$$S = \{\mu_{\log A}, \mu_E, \sigma_{\log A}\}$$

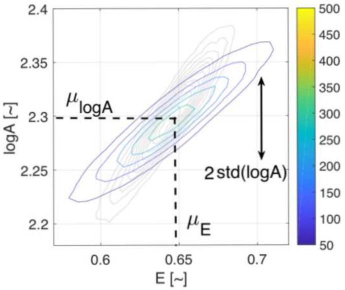


Posterior recovery fails (w.r.t. correlation) fails when we lose information about uncertain parameters that couple data across experiments, i.e. error in E

Hierarchical models

What if the statistics are reported on the parameters of the hierarchical model, instead of on each $k(T)$? We can infer data at all experimental conditions using these statistics as constraints on the data at all experiment temperatures.
e.g.

$$k(T) = A \pm \Delta A \text{exp}\left(-\frac{E}{T}\right) \tag{20}$$
$$S = \{\mu_{\log A}, \mu_E, \sigma_{\log A}\}$$



Posterior recovery fails (w.r.t. correlation) fails when we lose information about uncertain parameters that couple data across experiments, i.e. error in E

Real experiments - combustion chemical kinetics

Can extend this hierarchical learning approach to real experiments, e.g. H_2-O_2 chemical kinetics

chemical ODE model: assembly of parameterized reactions, $k_i(T) = A_i \exp\left(-\frac{E_i}{T}\right)$

1: $H+O_2 \longrightarrow OH+O$
Pirraglia et al., J. Phys. Chem. (1989)
2: $H+H_2O \longrightarrow OH+H_2$
Michael et al. J. Phys. Chem. (1988)

Missing data: H atom concentration decay vs. time, temperatures

Exp. 2 data fitting: ODE model governed only by reaction 2

Exp. 1 data fitting: coupled ODE model governed by reaction 1 with a nuisance reaction with specified rate parameters, 10% of the H decay (reaction 2)

```
ELEMENTS
O  H  AR  N
END
SPECIES
H2      H       O       O2      OH      H2O     HO2     H2O2
AR      N2
END
!THERMO
! Insert GRI-Mech thermodynamics here or use in default file
!END
REACTIONS
H+O2<=>O+OH                                 3.550E+15   -0.410   16600.00
H2+OH<=>H2O+H                               2.160E+08    1.510    3430.00
H2O2(+M)<=>2OH(+M)                          2.950E+14    0.000   48400.00
     LOW  /  1.200E+17    0.000   45500.0/
O+H2<=>H+OH                                 5.080E+04    2.670    6290.00
O+H2O<=>2OH                                 2.970E+06    2.020   13400.00
H2+M<=>2H+M                                 4.580E+19   -1.400  104380.00
H2/2.5/ H2O/12.0/
H2+AR<=>2H+AR                               5.840E+18   -1.100  104380.00
2O+M<=>O2+M                                 6.160E+15   -0.500       0.00
H2/2.5/ H2O/12.0/
2O+AR<=>O2+AR                               1.890E+13    0.000   -1790.00
O+H+M<=>OH+M                                4.710E+18   -1.000       0.00
H2/2.5/ H2O/12.0/
H+OH+M<=>H2O+M                              3.800E+22   -2.000       0.00
H2/2.5/ H2O/12.0/ AR/0.38/
H+O2(+M)<=>HO2(+M)                          1.480E+12    0.600       0.00
     LOW  /  9.040E+19   -1.500  490.00/
     TROE/   0.50  1.0E+10  1.0E-10 /
HO2+H<=>H2+O2                               1.660E+13    0.000     820.00
HO2+H<=>2OH                                 7.080E+13    0.000     300.00
HO2+O<=>OH+O2                               3.250E+13    0.000       0.00
HO2+OH<=>H2O+O2                             2.890E+13    0.000    -500.00
2HO2<=>O2+H2O2                              4.200E+14    0.000   11980.00
 DUPLICATE
2HO2<=>O2+H2O2                              1.300E+11    0.000   -1603.00
 DUPLICATE
H2O2+H<=>H2O+OH                             2.410E+13    0.000    3970.00
H2O2+H<=>H2+HO2                             4.820E+13    0.000    7950.00
H2O2+O<=>OH+HO2                             9.550E+06    2.000    3970.00
H2O2+OH<=>H2O+HO2                           1.000E+12    0.000       0.00
 DUPLICATE
H2O2+OH<=>H2O+HO2                           5.800E+14    0.000    9560.00
 DUPLICATE
END
```

Real experiments - combustion chemical kinetics

Can extend this hierarchical learning approach to real experiments, e.g. H_2-O_2 chemical kinetics

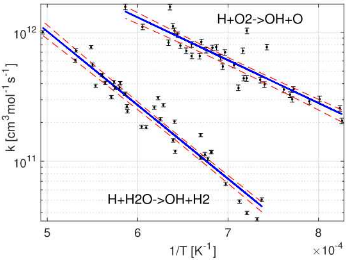
chemical ODE model: assembly of parameterized reactions, $k_i(T) = A_i \exp\left(-\frac{E_i}{T}\right)$

1: $H + O_2 \longrightarrow OH + O$
Pirraglia et al., J. Phys. Chem. (1989)
2: $H + H_2O \longrightarrow OH + H_2$
Michael et al. J. Phys. Chem. (1988)

Missing data: H atom concentration decay vs. time, temperatures

Exp. 2 data fitting: ODE model governed only by reaction 2

Exp. 1 data fitting: coupled ODE model governed by reaction 1 with a nuisance reaction with specified rate parameters, 10% of the H decay (reaction 2)



Real experiments - combustion chemical kinetics

Can extend this hierarchical learning approach to real experiments, e.g. H_2-O_2 chemical kinetics

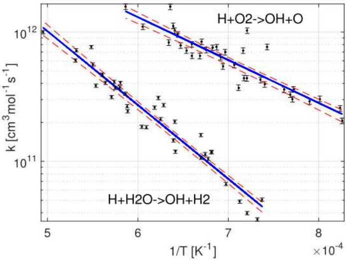
chemical ODE model: assembly of parameterized reactions, $k_i(T) = A_i \exp\left(-\frac{E_i}{T}\right)$

1: $H + O_2 \longrightarrow OH + O$
Pirraglia et al., J. Phys. Chem. (1989)
2: $H + H_2O \longrightarrow OH + H_2$
Michael et al. J. Phys. Chem. (1988)

Missing data: H atom concentration decay vs. time, temperatures

Exp. 2 data fitting: ODE model governed only by reaction 2

Exp. 1 data fitting: coupled ODE model governed by reaction 1 with a nuisance reaction with specified rate parameters, 10% of the H decay (reaction 2)

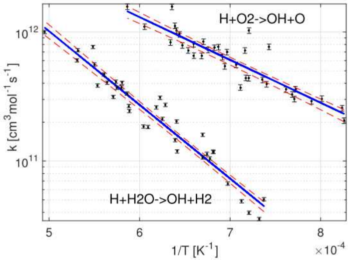


Real experiments - combustion chemical kinetics

Can extend this hierarchical learning approach to real experiments, e.g. H_2-O_2 chemical kinetics

chemical ODE model: assembly of parameterized reactions, $k_i(T) = A_i \exp\left(-\frac{E_i}{T}\right)$

1: $H+O_2 \longrightarrow OH+O$
Pirraglia et al., J. Phys. Chem. (1989)
2: $H+H_2O \longrightarrow OH+H_2$
Michael et al. J. Phys. Chem. (1988)



Missing data: H atom concentration decay vs. time, temperatures

Exp. 2 data fitting: ODE model governed only by reaction 2

Exp. 1 data fitting: coupled ODE model governed by reaction 1 with a nuisance reaction with specified rate parameters, 10% of the H decay (reaction 2)

Real experiments - combustion chemical kinetics

Can extend this hierarchical learning approach to real experiments, e.g. H_2-O_2 chemical kinetics

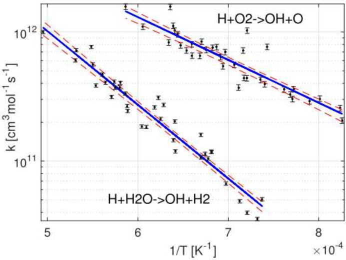
chemical ODE model: assembly of parameterized reactions, $k_i(T) = A_i \exp\left(-\frac{E_i}{T}\right)$

1: $H + O_2 \longrightarrow OH + O$
Pirraglia et al., J. Phys. Chem. (1989)
2: $H + H_2O \longrightarrow OH + H_2$
Michael et al. J. Phys. Chem. (1988)

Missing data: H atom concentration decay vs. time, temperatures

Exp. 2 data fitting: ODE model governed only by reaction 2

Exp. 1 data fitting: coupled ODE model governed by reaction 1 with a nuisance reaction with specified rate parameters, 10% of the H decay (reaction 2)



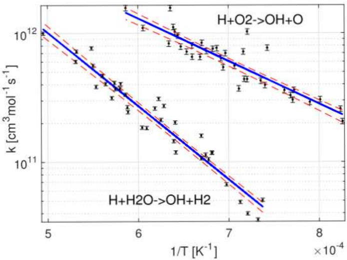
Real experiments - combustion chemical kinetics

Reaction 1:

Perform DFI procedure using reported $k(T)$ information at each temperature (35 in total)
Use data fusion and model replacement to learn the Arrhenius parameters of reaction 1 across all the data

Reaction 2:

repeat the procedure to learn the Arrhenius parameters of reaction 2



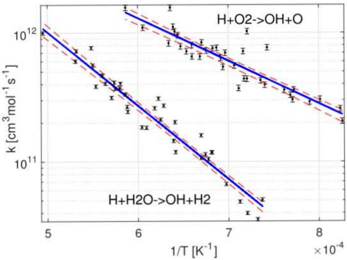
Real experiments - combustion chemical kinetics

Reaction 1:

- Perform DFI procedure using reported $k(T)$ information at each temperature (35 in total)
- Use data fusion and model replacement to learn the Arrhenius parameters of reaction 1 across all the data

Reaction 2:

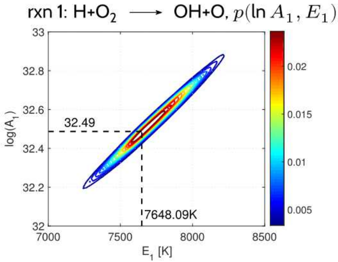
- repeat the procedure to learn the Arrhenius parameters of reaction 2



Real experiments - combustion chemical kinetics

Reaction 1:

- Perform DFI procedure using reported $k(T)$ information at each temperature (35 in total)
- Use data fusion and model replacement to learn the Arrhenius parameters of reaction 1 across all the data



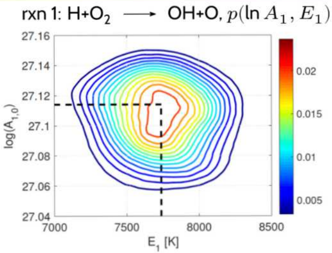
Reaction 2:

- repeat the procedure to learn the Arrhenius parameters of reaction 2

Real experiments - combustion chemical kinetics

Reaction 1:

> Perform DFI procedure using reported $k(T)$ information at each temperature (35 in total)
> Use data fusion and model replacement to learn the Arrhenius parameters of reaction 1 across all the data



Reaction 2:

> repeat the procedure to learn the Arrhenius parameters of reaction 2

Real experiments - combustion chemical kinetics

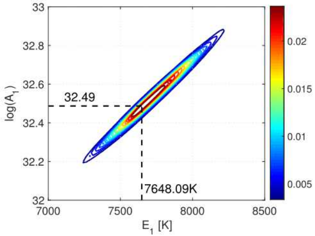
Reaction 1:

Perform DFI procedure using reported $k(T)$ information at each temperature (35 in total)
Use data fusion and model replacement to learn the Arrhenius parameters of reaction 1 across all the data

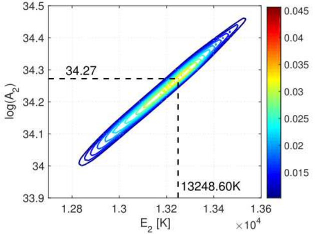
Reaction 2:

repeat the procedure to learn the Arrhenius parameters of reaction 2

rxn 1: $H+O_2 \longrightarrow OH+O$, $p(\ln A_1, E_1)$



rxn 2: $H+H_2O \longrightarrow OH+H_2$, $p(\ln A_2, E_2)$



Multiple reactions - combined learning

Learning all parameters across both reactions simultaneously

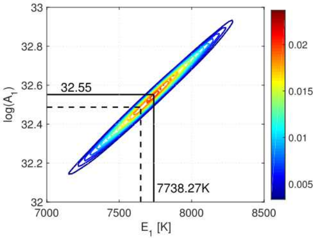
- Data from rxn 1 experiment is informative on both sets of parameters
- Marginal PDFs augmented compared to separate learning

Reveal subtle correlations between parameters across reactions

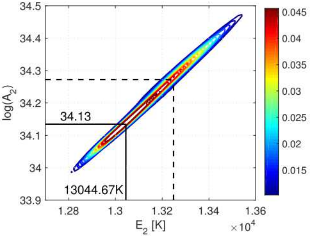
- Cross-reaction correlation arises from experiment design, activity of 'nuisance' reactions
- Analysis reveals temperature range where reactions are most coupled
- Opportunity to learn parameters of both reactions from a single experiment

Can be extended to experiments for all reactions in the chemical model, to construct the full JPDF

rxn 1: $H+O_2 \longrightarrow OH+O$, $p(\ln A_1, E_1)$



rxn 2: $H+H_2O \longrightarrow OH+H_2$, $p(\ln A_2, E_2)$



Multiple reactions - combined learning

Learning all parameters across both reactions simultaneously

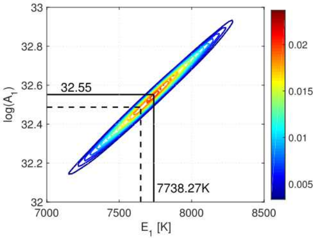
- Data from rxn 1 experiment is informative on both sets of parameters
- Marginal PDFs augmented compared to separate learning

Reveal subtle correlations between parameters across reactions

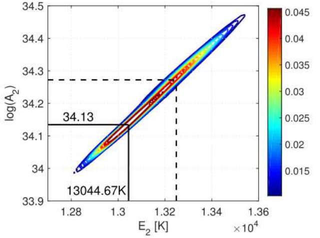
- Cross-reaction correlation arises from experiment design, activity of 'nuisance' reactions
- Analysis reveals temperature range where reactions are most coupled
- Opportunity to learn parameters of both reactions from a single experiment

Can be extended to experiments for all reactions in the chemical model, to construct the full JPDF

rxn 1: $H+O_2 \longrightarrow OH+O$, $p(\ln A_1, E_1)$



rxn 2: $H+H_2O \longrightarrow OH+H_2$, $p(\ln A_2, E_2)$



Multiple reactions - combined learning

Learning all parameters across both reactions simultaneously

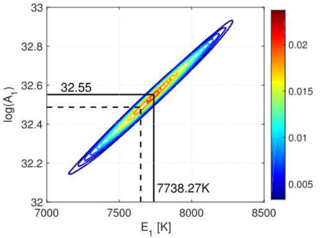
- Data from rxn 1 experiment is informative on both sets of parameters
- Marginal PDFs augmented compared to separate learning

Reveal subtle correlations between parameters across reactions

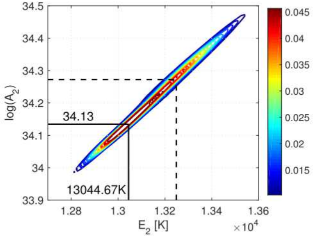
- Cross-reaction correlation arises from experiment design, activity of 'nuisance' reactions
- Analysis reveals temperature range where reactions are most coupled
- Opportunity to learn parameters of both reactions from a single experiment

Can be extended to experiments for all reactions in the chemical model, to construct the full JPDF

rxn 1: $H+O_2 \longrightarrow OH+O$, $p(\ln A_1, E_1)$



rxn 2: $H+H_2O \longrightarrow OH+H_2$, $p(\ln A_2, E_2)$



Multiple reactions - combined learning

Learning all parameters across both reactions simultaneously

- Data from rxn 1 experiment is informative on both sets of parameters
- Marginal PDFs augmented compared to separate learning

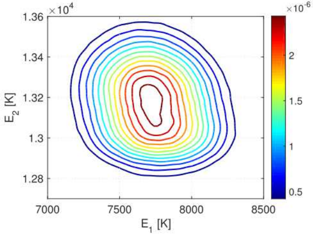
Reveal subtle correlations between parameters across reactions

- Cross-reaction correlation arises from experiment design, activity of 'nuisance' reactions
- Analysis reveals temperature range where reactions are most coupled
- Opportunity to learn parameters of both reactions from a single experiment

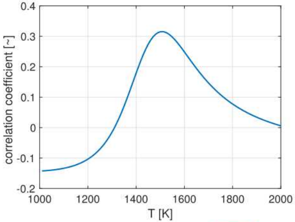
Can be extended to experiments for all reactions in the chemical model, to construct the full JPDF

correlation between activation energies,

$p(E_1, E_2)$



$\rho_{k_1(T),k_2(T)}$



Multiple reactions - combined learning

Learning all parameters across both reactions simultaneously

- Data from rxn 1 experiment is informative on both sets of parameters
- Marginal PDFs augmented compared to separate learning

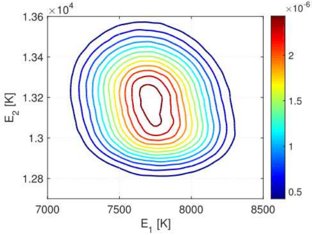
Reveal subtle correlations between parameters across reactions

- Cross-reaction correlation arises from experiment design, activity of 'nuisance' reactions
- Analysis reveals temperature range where reactions are most coupled
- Opportunity to learn parameters of both reactions from a single experiment

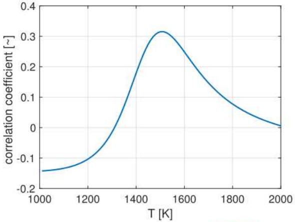
Can be extended to experiments for all reactions in the chemical model, to construct the full JPDF

correlation between activation energies,

$p(E_1, E_2)$



$\rho_{k_1(T), k_2(T)}$



Multiple reactions - combined learning

Learning all parameters across both reactions simultaneously

- Data from rxn 1 experiment is informative on both sets of parameters
- Marginal PDFs augmented compared to separate learning

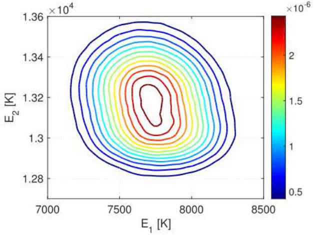
Reveal subtle correlations between parameters across reactions

- Cross-reaction correlation arises from experiment design, activity of 'nuisance' reactions
- Analysis reveals temperature range where reactions are most coupled
- Opportunity to learn parameters of both reactions from a single experiment

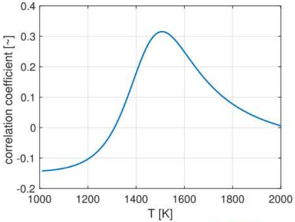
Can be extended to experiments for all reactions in the chemical model, to construct the full JPDF

correlation between activation energies,

$p(E_1, E_2)$



$\rho_{k_1(T),k_2(T)}$



Multiple reactions - combined learning

Learning all parameters across both reactions simultaneously

- Data from rxn 1 experiment is informative on both sets of parameters
- Marginal PDFs augmented compared to separate learning

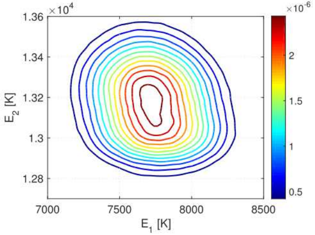
Reveal subtle correlations between parameters across reactions

- Cross-reaction correlation arises from experiment design, activity of 'nuisance' reactions
- Analysis reveals temperature range where reactions are most coupled
- Opportunity to learn parameters of both reactions from a single experiment

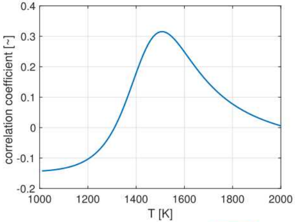
Can be extended to experiments for all reactions in the chemical model, to construct the full JPDF

correlation between activation energies,

$p(E_1, E_2)$



$\rho_{k_1(T),k_2(T)}$



Multiple reactions - combined learning

Learning all parameters across both reactions simultaneously

- Data from rxn 1 experiment is informative on both sets of parameters
- Marginal PDFs augmented compared to separate learning

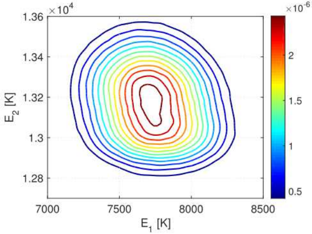
Reveal subtle correlations between parameters across reactions

- Cross-reaction correlation arises from experiment design, activity of 'nuisance' reactions
- Analysis reveals temperature range where reactions are most coupled
- Opportunity to learn parameters of both reactions from a single experiment

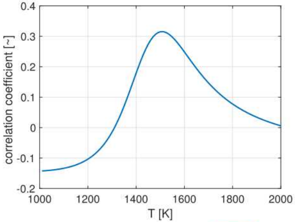
Can be extended to experiments for all reactions in the chemical model, to construct the full JPDF

correlation between activation energies,

$p(E_1, E_2)$



$\rho_{k_1(T), k_2(T)}$



Uncertainty propagation - predictive modeling

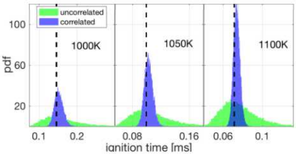
Drawing samples from the JPDF of all parameters, can propagate uncertainty through combustion models

- Ignition delay time
- Laminar flame speed
- Pressure-temperature explosion limits

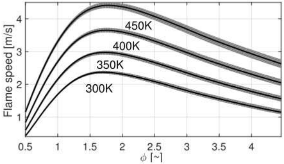
Parameter correlation information is crucial

- Drastic over-prediction of uncertainty in quantities of interest (QoIs) when parameters are assumed independently distributed according to their marginal PDFs

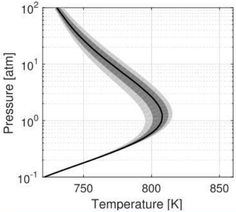
ignition time: impact of correlation



laminar flame speed



explosion limit curve



Uncertainty propagation - predictive modeling

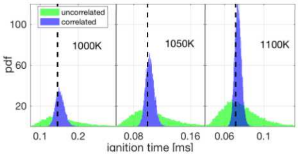
Drawing samples from the JPDF of all parameters, can propagate uncertainty through combustion models

- Ignition delay time
- Laminar flame speed
- Pressure-temperature explosion limits

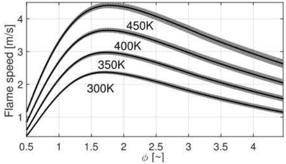
Parameter correlation information is crucial

- Drastic over-prediction of uncertainty in quantities of interest (QoIs) when parameters are assumed independently distributed according to their marginal PDFs

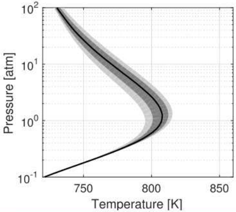
ignition time: impact of correlation



laminar flame speed



explosion limit curve



Sandia National Laboratories

Summary

For estimating parameter uncertainty when supporting data is unreported or missing, we have recourse to a data-free inference (DFI) procedure for constructing a MaxEnt posterior density consistent with available information

Demonstrated for simple data models, and hierarchical models with variety of reported summary statistics at different modeling levels

The MaxEnt linear pooling procedure enables combination of hypothetical data sets, model replacement for a desired fitting context

Applied to real missing experimental data sets in combustion chemical kinetics. Discover parameter uncertainty structure implied by the reported information

When missing data sets are informative on multiple parameters in the model, recover correlation of model parameters across reactions in the combined learning setting, avoiding bias in the predictive model specification

Identify missed opportunities for learning, enable model selection

Summary

For estimating parameter uncertainty when supporting data is unreported or missing, we have recourse to a data-free inference (DFI) procedure for constructing a MaxEnt posterior density consistent with available information

Demonstrated for simple data models, and hierarchical models with variety of reported summary statistics at different modeling levels

The MaxEnt linear pooling procedure enables combination of hypothetical data sets, model replacement for a desired fitting context

Applied to real missing experimental data sets in combustion chemical kinetics. Discover parameter uncertainty structure implied by the reported information

When missing data sets are informative on multiple parameters in the model, recover correlation of model parameters across reactions in the combined learning setting, avoiding bias in the predictive model specification

Identify missed opportunities for learning, enable model selection

Summary

For estimating parameter uncertainty when supporting data is unreported or missing, we have recourse to a data-free inference (DFI) procedure for constructing a MaxEnt posterior density consistent with available information

Demonstrated for simple data models, and hierarchical models with variety of reported summary statistics at different modeling levels

The MaxEnt linear pooling procedure enables combination of hypothetical data sets, model replacement for a desired fitting context

Applied to real missing experimental data sets in combustion chemical kinetics. Discover parameter uncertainty structure implied by the reported information

When missing data sets are informative on multiple parameters in the model, recover correlation of model parameters across reactions in the combined learning setting, avoiding bias in the predictive model specification

Identify missed opportunities for learning, enable model selection

Summary

For estimating parameter uncertainty when supporting data is unreported or missing, we have recourse to a data-free inference (DFI) procedure for constructing a MaxEnt posterior density consistent with available information

Demonstrated for simple data models, and hierarchical models with variety of reported summary statistics at different modeling levels

The MaxEnt linear pooling procedure enables combination of hypothetical data sets, model replacement for a desired fitting context

Applied to real missing experimental data sets in combustion chemical kinetics. Discover parameter uncertainty structure implied by the reported information

When missing data sets are informative on multiple parameters in the model, recover correlation of model parameters across reactions in the combined learning setting, avoiding bias in the predictive model specification

Identify missed opportunities for learning, enable model selection

Summary

For estimating parameter uncertainty when supporting data is unreported or missing, we have recourse to a data-free inference (DFI) procedure for constructing a MaxEnt posterior density consistent with available information

Demonstrated for simple data models, and hierarchical models with variety of reported summary statistics at different modeling levels

The MaxEnt linear pooling procedure enables combination of hypothetical data sets, model replacement for a desired fitting context

Applied to real missing experimental data sets in combustion chemical kinetics. Discover parameter uncertainty structure implied by the reported information

When missing data sets are informative on multiple parameters in the model, recover correlation of model parameters across reactions in the combined learning setting, avoiding bias in the predictive model specification

Identify missed opportunities for learning, enable model selection

Summary

For estimating parameter uncertainty when supporting data is unreported or missing, we have recourse to a data-free inference (DFI) procedure for constructing a MaxEnt posterior density consistent with available information

Demonstrated for simple data models, and hierarchical models with variety of reported summary statistics at different modeling levels

The MaxEnt linear pooling procedure enables combination of hypothetical data sets, model replacement for a desired fitting context

Applied to real missing experimental data sets in combustion chemical kinetics. Discover parameter uncertainty structure implied by the reported information

When missing data sets are informative on multiple parameters in the model, recover correlation of model parameters across reactions in the combined learning setting, avoiding bias in the predictive model specification

Identify missed opportunities for learning, enable model selection

Acknowledgement

M. Khalil, X. Huan, B.J. Debusschere, K. Sargsyan, C. Safta, K. Chowdhary, R.D. Berry

This work was supported by:

DOE Office of Basic Energy Sciences, Div. of Chem. Sci., Geosci., & Biosci.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. The views expressed in the article do not necessarily represent the views of the U.S. Department of Energy or the United States Government.