Calibration and Model Discrepancy

Tony O'Hagan, MUCM, Sheffield

Outline

- Why model discrepancy
- The meaning of parameters
- Modelling discrepancy
- Conclusions

Why model discrepancy

Is calibration even possible?

The calibration problem

- The problem
  - We have a simulator that predicts a real world phenomenon
  - We have some observations of the real world
  - We want to use those to learn about some unknown parameters
- Formally, the simulator takes two kinds of inputs
  - The calibration parameters θ
  - Control inputs x
  - Simulator is written $y = f(x, \theta)$
  - Observation z_i is obtained at control input values x_i

Traditional formulation

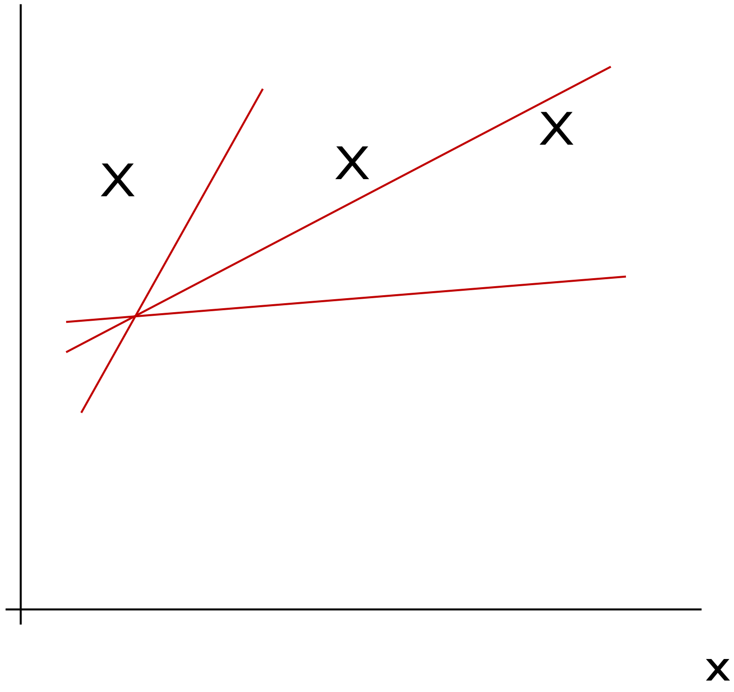
- Write

  $$z_i = f(x_i, \theta) + \varepsilon_i$$

  - where ε_i are independent observation errors
- Estimate θ, e.g. by minimising sum of squared residuals
  - Call estimate t and predict real world process at a new x value by f(x, t)
- Two things wrong with this
  - The formulation is wrong because the simulation model is inevitably wrong
    - So errors are not independent
  - Traditional calibration ignores uncertainty about θ
    - Treats it as now known to equal t

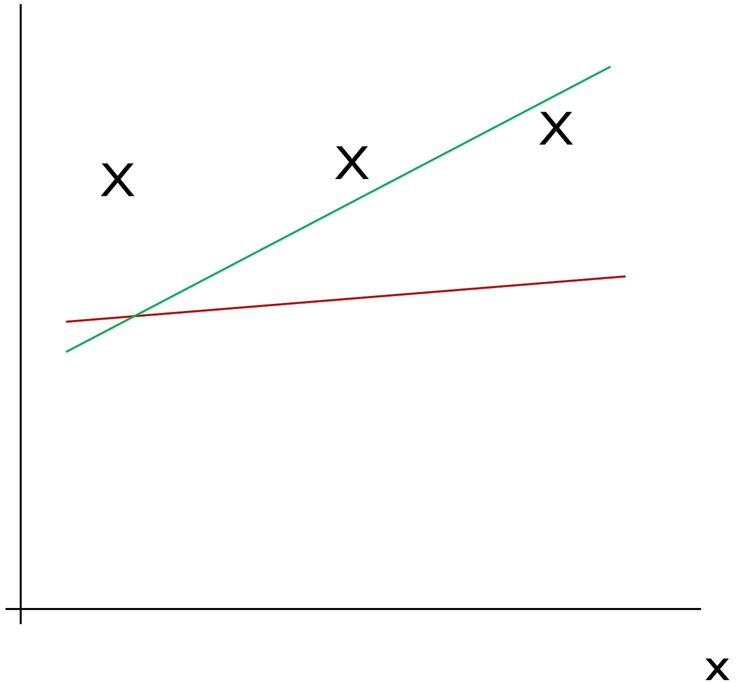
Little example

- One control input, one calibration parameter
  - Three observations marked with crosses
  - Red lines are possible simulator outputs
  - Calibration parameter just changes slope of line
- No value of the calibration parameter gets close to all the observations
  - And yet they seem to lie on a straight line



Is calibration even possible?

- Green line is best fit
  - Minimises sum of squared residuals
  - Red line seems better
    - But with constant bias
  - Green seems to be over-fitting
    - Errors don't look independent
- Can we learn the true value of the calibration parameter?
  - With more data
    - Keeping close to a straight line
  - Over a different range of x



Model discrepancy

- The little example suggests that we need to allow that the model does not correctly represent reality
  - For *any* values of the calibration parameters
- The simulator outputs deviate systematically from reality
  - Call it model bias or model discrepancy
- It is claimed that acknowledging model discrepancy may allow us to achieve more realistic and appropriate estimates of calibration parameters
- But to evaluate that claim, look at a simpler problem

The meaning of parameters

What can we learn from simple models?

What do parameters mean?

- All models are wrong
  - "All models are wrong but some are useful"
    - George E P Box, 1979
- So, what does a parameter (in an admittedly wrong model) mean?
  - How do we specify prior information about a parameter
    - when we know the model is wrong?
  - What have we learnt when we make inference about a parameter
    - in a model we know to be wrong?

Example: Poisson sample

- Suppose we model data as a sample from a Poisson distribution with parameter λ
- We need a prior distribution
  - Do we ask for beliefs about the population mean or variance?
  - Or about the proportion p of zeros in the population?
    - And infer a prior for $\lambda = -\log p$
  - Given that the Poisson assumption is wrong these are all asking about different things
- When we derive a posterior distribution for λ
  - Is it a belief statement about the population mean or variance?
  - Or even about anything real?

Example: Linear regression

- Suppose we assume a simple linear regression model with slope β and intercept α
  - We are interested in the strength of the relationship as represented by β
  - But the model is wrong; the relationship is not truly linear
- We know if we switch to a quadratic model the coefficient of x will change
  - If we assume a linear relationship when the truth is different, e.g. quadratic, the slope will depend on the range of x over which we fit
  - How can we elicit prior beliefs about such a parameter?
  - What do inferences about it mean?

Example: A simple machine

- A machine produces an amount of work y which depends on the amount of effort t put into it
  - Model is $y = \beta t + \varepsilon$
  - Where β is the rate at which effort is converted to work
  - And ε is observation error, assumed iid
  - True value of β is 0.65
- Graph shows observed data
  - All points lie below $y = 0.65t$
  - Because the model is wrong
  - Losses due to friction etc.
  - Fitted slope is 0.568

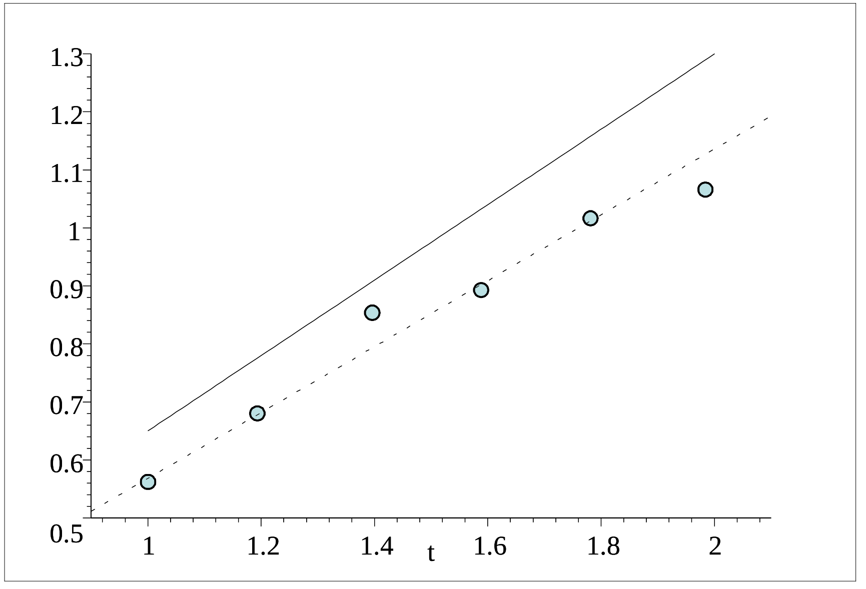


Figure 1

Simple machine – true model

- The true model is shown as the dashed line here
- In this example, the efficiency parameter β is physically meaningful
  - Theoretical value is 0.65
  - This value is of interest to experimenters
    - They have genuine prior information
  - They want the experiment to help them identify this true value
  - But because of model error the estimate is biased
    - And given enough data it will over-rule any prior information

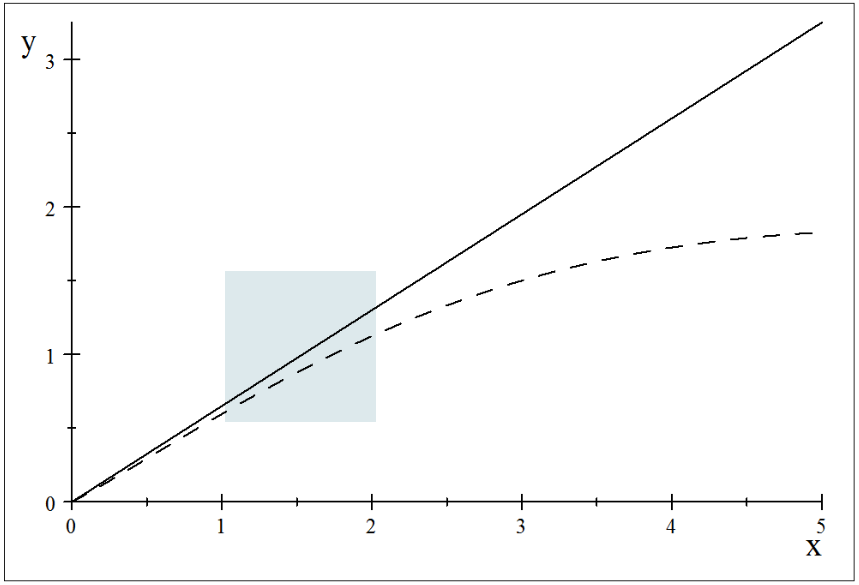


Figure 2

Calibration is just nonlinear regression

- Returning to the context of computer models

  $$y = f(x, \theta) + \varepsilon$$

  - where f is a computer simulator of some phenomenon
- We can view this as just a nonlinear regression model
  - The regression function $f(x, \theta)$ is complex and we can't try alternatives (as we would do in regression modelling)
  - But we have all the same problems as in simple regression
- Given that the model is wrong:
  - What do the calibration parameters θ mean?
  - We can't expect to learn their 'true' values from observations
    - Even with unlimited data

Tuning and physical parameters

- Simulator parameters may be physical or just for tuning
- We adjust tuning parameters so the model fits reality better
  - We are not really interested in their 'true' values
  - We calibrate tuning parameters for prediction
- Physical parameters are different
  - We are often really interested in true physical values
    - And we like to think that calibration can help us learn about them
  - But the model is inevitably wrong, so estimates are distorted
    - And getting more data does not take us closer to their true values
  - Calibration to learn about physical parameters is a delusion
    - Unless … ?

Modelling discrepancy

Is model discrepancy the answer?

Model discrepancy

- In the context of computer models, it is necessary to acknowledge model discrepancy
  - There is a difference between the model with best/true parameter values and reality

$$y = f(x, \theta) + \delta(x) + \varepsilon$$

  - where $\delta(x)$ accounts for this discrepancy
    - Will typically itself have uncertain parameters
- Kennedy and O'Hagan (JRSSB, 2001) introduced this model discrepancy
  - Modelled it as a zero-mean Gaussian process
  - They claimed it acknowledges additional uncertainty
  - And mitigates against over-fitting of θ

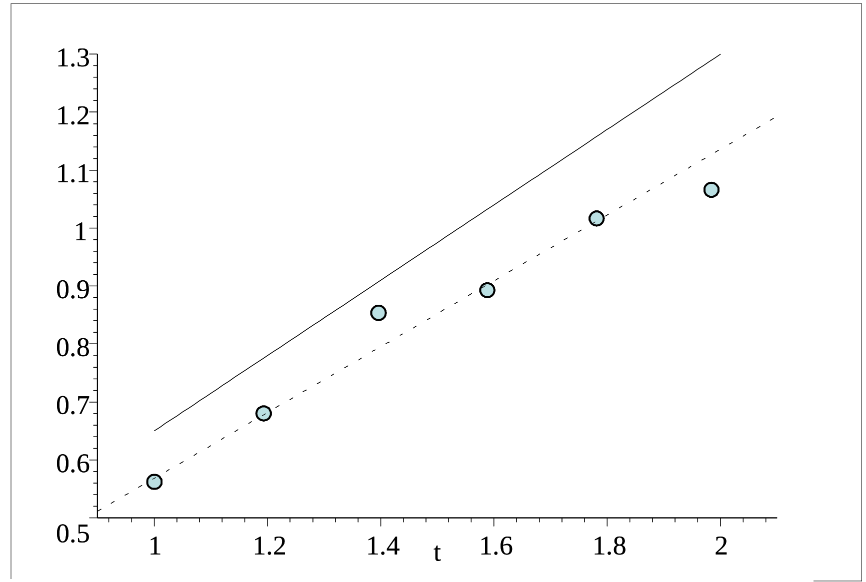
Simple machine revisited

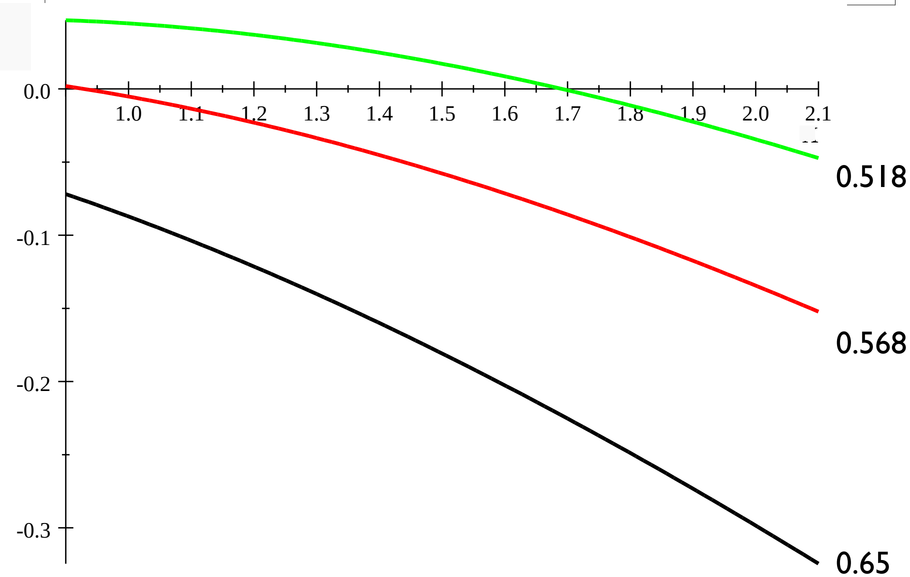
- So add this model discrepancy term to the linear model of the simple machine

$$y = \beta t + \delta(t) + \varepsilon$$

  - With $\delta(t)$ modelled as a zero-mean GP
    - As in Kennedy and O'Hagan
- Now the estimate of β is 0.518
  - It's even further from the true value of 0.65!
  - Without model discrepancy we got 0.568
- What's going on?

- In order to get the right answer we have to infer:
  - The true value of the efficiency parameter is 0.65
    - The solid line
  - Model discrepancy is negative for all t
    - And more so for larger t
    - Like the black curve below
- But the GP says:
  - It's much more likely to be the green curve





Nonidentifiability

- Formulation with model discrepancy is not identifiable
- For any θ, there is a $\delta(x)$
  - Reality is some function $\zeta(x) = f(x, \theta) + \delta(x)$
  - Given θ, model discrepancy is $\delta(x) = \zeta(x) - f(x, \theta)$
  - As in the three curves in the previous example
- Suppose we had an unlimited number of observations
  - We would learn reality's true function $\zeta(x)$ exactly
  - But we would still not learn θ
    - It could in principle be anything
  - In a Bayesian analysis, the prior distribution is used to resolve nonidentifiability

The joint posterior

- Calibration leads to a joint posterior distribution for θ and $\delta(x)$
- But nonidentifiability means there are many equally good fits $(\theta, \delta(x))$ to the data
  - Induces strong correlation between θ and $\delta(x)$
  - This may be compounded by the fact that simulators often have large numbers of parameters
    - (Near-)redundancy means that different θ values produce (almost) identical predictions
    - Sometimes called equifinality
- Within this set, the prior distributions for θ and $\delta(x)$ count

Modelling the discrepancy

- The nonparametric GP term allows the model to fit and predict reality accurately given enough data
- But it **doesn't** mean physical parameters are correctly estimated
  - The separation between original model and discrepancy is unidentified
  - Estimates depend on prior information
  - Unless the real model discrepancy is just the kind expected a priori the physical parameter estimates will still be biased
- It is necessary to **think** about the model discrepancy
  - In the machine example, a prior distribution saying $\delta(t)$ is likely to be negative and decreasing will produce a better answer

What do I mean by 'better'?

- Posterior distribution for θ will typically be quite wide
  - And won't become degenerate with infinite data
- Values of θ with very low posterior probability will typically either
  - Have very low prior probability or
  - Imply a $\delta(x)$ with very low prior probability
- Assuming we don't get prior information about θ wrong, it's important to model $\delta(x)$ carefully
  - Flexibly but informatively
- So 'better' means getting a posterior distribution that covers the true θ

Reification

- In the computer models context, Goldstein and Rougier (JSPI, 2009) introduced a formal mechanism for modelling model discrepancy
  - Called reification
- Based on imagining hypothetical improved model(s)
  - The reified model is such that we have no knowledge of how it might differ from reality,
  - So homogeneous zero-mean discrepancy is appropriate
  - We may also be able to consider the next-generation model in which specific improvements have been made
- Their framework may be over-kill, but
  - if we want parameters to have meaning we have to think seriously about model discrepancy

Extrapolation

- Here's an example of how important discrepancy modelling is
- Consider prediction using a regression model but for x values far from the data
  - We often hear how extrapolation is a bad idea, because of model discrepancy
  - But if $\delta(x)$ has finite variance the impact is bounded, no matter how far we extrapolate
  - For unbounded impact in large extrapolations we could model discrepancy using something like a random walk
    - Or a localised regression model (O'Hagan, JRSSB, 1978)

Conclusions

And challenges!

Key messages

- If you don't include some representation of model discrepancy, you can expect to get nonsense
  - Posterior distribution of θ will converge to wrong value
  - Overfitting and spurious accuracy
- Even if you do include model discrepancy, it's essential to **think** about it and model it carefully
  - Use knowledge about aspects of reality that are not adequately represented in the model
- Even if you do think about it and model it carefully,
  - You will not be able to learn the true physical values of calibration parameters
    - Not even with an unlimited number of physical observations
  - Posterior distribution of θ will not converge
    - But should cover the true value

Challenges

- We need to gain much more practical experience of how calibration works when we incorporate model discrepancy
- A major challenge is how to model the discrepancy
  - 'Flexibly but informatively'

Thanks to …

- Colleagues in the MUCM project
  - Managing Uncertainty in Complex Models
  - http://mucm.ac.uk/