

Uncertainty in energy system models

Amy Wilson

Durham University

May 2015

Table of Contents

- 1 Model uncertainty
- 2 Statistical methodology
- 3 Example - generation investment
- 4 Conclusion

Contents

- 1 Model uncertainty
- 2 Statistical methodology
- 3 Example - generation investment
- 4 Conclusion

Models

A large part of the study of energy systems involves the building of models (deterministic or stochastic) to represent particular aspects of an energy system. A model:

- takes a set of inputs (known, uncertain, control parameters)
- produces a set of outputs
- is an approximation of a system.

Models

- Let \mathbf{x} be the model inputs,
- \mathbf{y} be the model outputs,

Then the model can be thought of as a function f , with

$$\mathbf{y} = f(\mathbf{x}).$$

George Box: 'Essentially, all models are wrong, but some are useful'
But how can we tell how useful a model is?

Types of uncertainty

- **Parametric uncertainty:** uncertainty about the appropriate input parameters to use.
 - E.g. demand forecasts, investor risk level, weather data.
- **Structural uncertainty:** how the model itself relates to the real-world process it is modelling
 - E.g. are the data really Normally distributed? what effect does this simplifying assumption have on the results?
- Both types of uncertainty can be found in deterministic and stochastic models.

Why should we account for uncertainty

- Avoid overconfidence in results - giving a single answer is misleading if there is uncertainty.
- Reporting a confidence/probability interval when full account hasn't been taken of all uncertainties can be equally misleading.
- Normally want to reduce our uncertainty about the real-world, not about the model - can't do this unless we model this uncertainty.
- Huge costs can be at stake when policy decisions are based on model output - the cost of doing an uncertainty analysis can be much smaller than the cost of making the wrong decision.

Contents

- 1 Model uncertainty
- 2 Statistical methodology**
- 3 Example - generation investment
- 4 Conclusion

Statistical methods

For models that are quick to run:

- **Parametric uncertainty:**

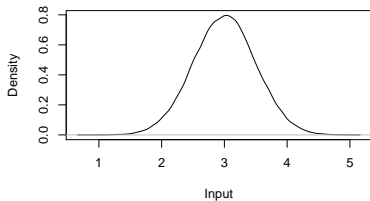
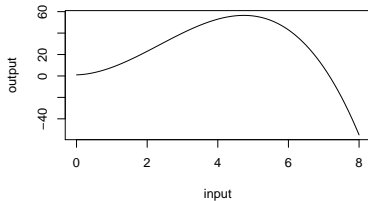
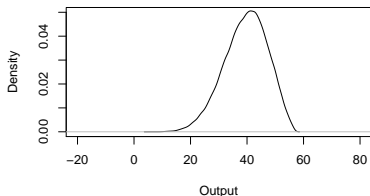
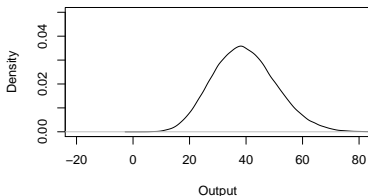
- Monte Carlo simulation (with carefully chosen distributions for your input parameters).
- Scenario-based analyses are unlikely to be sufficient - what about situations that fall in between the scenarios?
- Doing a sensitivity analysis by changing one input at a time and testing how the output changes won't be enough either - what about interactions?

- **Structural discrepancy:** can model this using a combination of:

- historical data
- expert opinion on the shortfalls of the model
- experimentation.

Can be a detailed analysis, dependent on the inputs, or can use a broad-brush approach if data/expert opinion isn't available.

Example

Input**Model****Output – no discrepancy****Output – 20% discrepancy**

Example capacity adequacy

- Estimation of the probability

$$P(\text{demand}_t - \text{wind}_t > \text{available generation}_t)$$

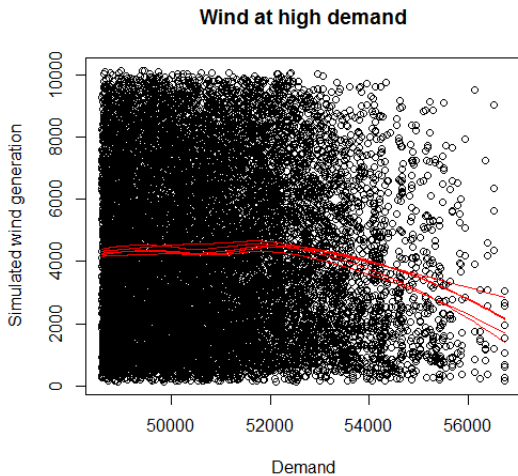
over all times in the UK peak season (winter).

- Inputs are demand forecast, temperature forecast, size of conventional generating units and availability probabilities.
- Wind forecast simulated dependent on the temperature.
- Parametric uncertainty accounted for using bootstrapping (only considers the uncertainties we have seen historically in the data).
- Structural uncertainty considered by relaxing assumption of independence of wind and demand (but still some work to do).

Example capacity adequacy

Year	LOLE (hours per year)	L. bound	U. bound
2005–06	0.45	0.27	0.62
2010–11	3.12	2.24	4.01
2013–14	0.13	0.065	0.22
Overall	1.15	0.52	1.95

Example capacity adequacy



Emulation

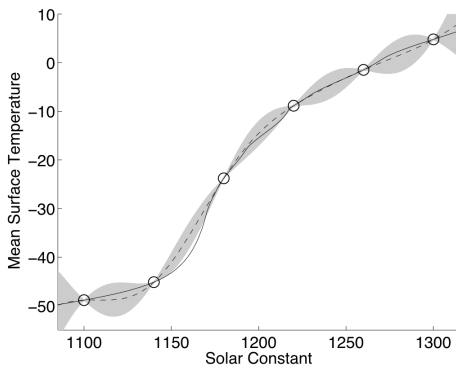
Models that take a long time to evaluate are more difficult, but methods are well established. See www.mucm.ac.uk. Have been used in:

- Climate change models,
- Study of oil wells,
- Study of spread of aids,
- Models of the galaxy!

Emulation general idea

- Idea is to build a statistical model of your model, based on a small number of model evaluations.
- At inputs for which you haven't evaluated the model, you don't know what the model output is.
- However, your statistical model can give you a probability distribution for the output at each input.
- You can then use Monte Carlo simulation with the statistical model instead of the original model.

Emulation general idea



From www.MUCM.ac.uk

Gaussian processes

One particular statistical model that is often used is a Gaussian Process (GP). We again represent the simulator as

$$\mathbf{y} = f(\mathbf{x}).$$

A GP assumes that for m different sets of inputs, $\mathbf{x}_1, \dots, \mathbf{x}_m$, the joint distribution of $f(\mathbf{x}_1), \dots, f(\mathbf{x}_m)$ follows a multivariate Normal distribution, i.e.

$$(f(\mathbf{x}_1), \dots, f(\mathbf{x}_m)) \mid \theta \sim \text{MVN}(\mathbf{m}, V),$$

for some mean \mathbf{m} and some covariance matrix V (both functions of $\mathbf{x}_1, \dots, \mathbf{x}_m$ and some parameters θ to be estimated) .

Emulation Gaussian Process

Standard method is to:

- Set mean to $m(\mathbf{x}, \beta) = \mathbf{h}(\mathbf{x})^T \beta$ for some vector of functions $\mathbf{h}(\mathbf{x})$.
- E.g. for two-dimensional input space, could have

$$\mathbf{h}(\mathbf{x})^T = (1, x_1, x_2),$$

- Set covariance between $f(\mathbf{x})$ and $f(\mathbf{x}')$ to be

$$\text{Cov}(f(\mathbf{x}), f(\mathbf{x}') \mid \beta, \sigma^2) = \sigma^2 c(\mathbf{x}, \mathbf{x}'),$$

for some correlation function $c(\mathbf{x}, \mathbf{x}')$.

- E.g. often see correlation function set to

$$c(\mathbf{x}, \mathbf{x}') = \exp[-(\mathbf{x} - \mathbf{x}')^T D (\mathbf{x} - \mathbf{x}')],$$

with D a diagonal matrix.

Emulation parameter estimation

- To estimate the GP for your energy simulator, need to estimate β, σ^2 and D .
- Have training data $(\mathbf{x}_i, f(\mathbf{x}_i))$ for $i \in \{1, \dots, n\}$ (the inputs you have tested your model at).
- Can use Bayesian methods, specifying a prior distribution for β, σ^2 and D and updating this with the training data to get estimates. To simplify the maths, can instead find D using optimisation methods and treat this as fixed.
- For more details see www.MUCM.ac.uk.

Emulation new input

- Once you have estimates for β , σ^2 and D , it is possible to estimate a distribution for any new set of inputs (those at which the output is unknown).
- This distribution can be used in combination with Monte Carlo simulation to estimate the overall uncertainty in the model output, as well as the sensitivity of the output to groups of inputs.
- Monte Carlo simulation is possible now because the emulator is much faster to evaluate than the simulator.
- To use Monte Carlo simulation, need to specify a prior distribution over the input space.
- Note that as you are using a statistical model of the simulator, you have added in a new uncertainty - the uncertainty due to emulation!

Structural discrepancy

- There are many ways to model structural discrepancy.
- E.g. if \mathbf{y}_r are the 'real-world' observations, rather than the model output, we could set

$$\mathbf{y}_r = f(\mathbf{x}) + \epsilon(\mathbf{x})$$

- And model $\epsilon(\mathbf{x}) \sim N(0, \sigma^2)$
- Or $\epsilon(\mathbf{x}) \sim N(0, \sigma^2(x))$, using a difference variance for different areas of the input space,
- Or can model $\epsilon(\mathbf{x})$ using another Gaussian Process.

Contents

- 1 Model uncertainty
- 2 Statistical methodology
- 3 Example - generation investment**
- 4 Conclusion

Models

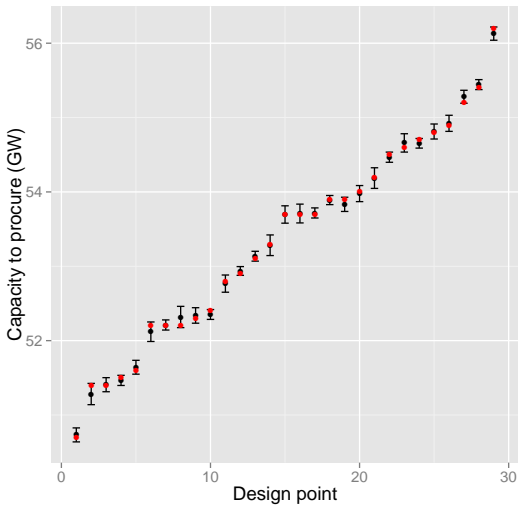
Two long-term generation investment models:

- Example 1: the DDM (dynamic dispatch model). Used by DECC to predict behaviour of energy market so that good policy decisions can be made.
- Example 2: long-term generation investment model developed by Eager, Bialek and Hobbs in 'Dynamic modeling of thermal generation capacity investment: Application to markets with high wind penetration', Power Systems, IEEE Transactions on 27(4) (2012) 2127-2137. (this analysis was done by Meng Xu, Durham University).

Example 1

- Gaussian Process emulator fitted using 6 inputs and 1 output:
 - inputs: annual demand, peak demand, interconnector flow, CCGT availability, wind level, weather conditions.
 - output: capacity to procure.
- No structural discrepancy modelled (yet!).
- 29 design points (so based on 29 model evaluations). One design point was excluded.

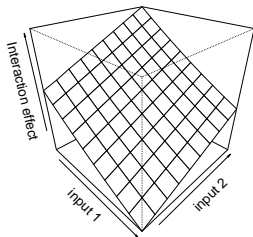
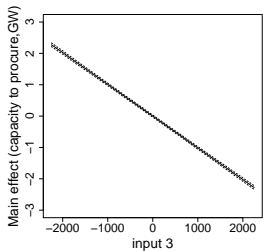
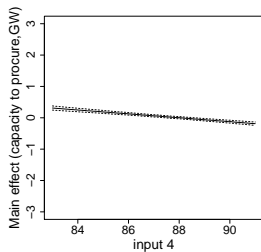
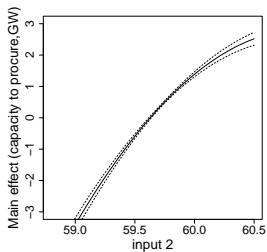
Model fit



Uncertainty in results

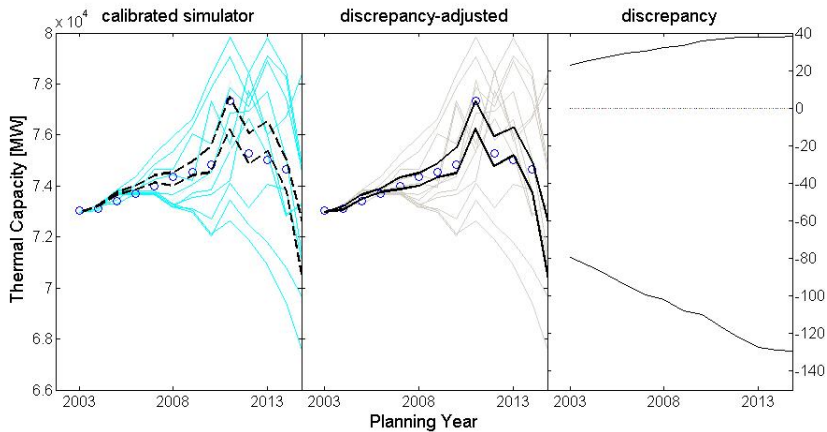
Uncertainty measure	Capacity to procure (GW)		
	Correlation= 0	0.25	0.85
Expected value:	53.59	53.59	53.58
Variance (emulation):	0.00030	0.00031	0.00031
Variance (parameter):	8.64	6.95	2.77
Variance of variance (emulation):	0.30	0.17	0.02

Sensitivity analysis



Example 2

- Gaussian Process emulator fitted using 2 inputs and 1 output:
 - inputs: attitude to risk, price mark-up
 - output: annual installed thermal generation capacity.
- Structural discrepancy modelled using a sum of Normal distributions (discrepancy therefore varies with time, but not with input parameters)
- 12 design points.
- One aim was to calibrate the distribution of the input parameters against historical observations.



Contents

- 1 Model uncertainty
- 2 Statistical methodology
- 3 Example - generation investment
- 4 Conclusion**

Conclusion

- Important to consider uncertainties when modelling. Both parametric and structural.
- Without assessing uncertainties, it is not normally possible to use a model to say anything about the 'real-world'.
- Variety of statistical methods can be used, depending on characteristics of model.
- For models which take a long time to run, emulation is a useful tool.

Acknowledgements

With thanks to:

- Chris Dent
- Michael Goldstein
- Meng Xu