

Total UQ

Total Tony O'Hagan

Outline

- ▶ **Uncertainty**
 - ▶ Uncertainties in the use of computer models
 - ▶ Uncertainty and probability
- ▶ **Quantification**
 - ▶ Input uncertainties
 - ▶ Code uncertainty
 - ▶ Model discrepancy
 - ▶ Validation
- ▶ **The TUQ manifesto**
- ▶ **Some challenges**



Uncertainty

Is all around

Three principal uncertainties

- ▶ I write reality $\zeta(\mathbf{x})$ as

$$\zeta(\mathbf{x}) = f(\mathbf{x}, \theta) + \delta(\mathbf{x})$$

- ▶ Where as usual $f(\cdot, \cdot)$ represents the simulator, \mathbf{x} are control inputs, θ are calibration parameters and $\delta(\cdot)$ is model discrepancy
- ▶ The principal uncertainties about reality are
 - ▶ Inputs \mathbf{x} and θ
 - ▶ The simulator output $f(\cdot, \cdot)$ at untried inputs
 - ▶ The model discrepancy $\delta(\cdot)$
- ▶ If we want to say anything about reality we must quantify all of these uncertainties

The right way to quantify uncertainty

- ▶ **Probability**
 - ▶ This is not negotiable
 - ▶ No argument
 - ▶ Any other measure does not have the right properties
- ▶ **But quantifying uncertainty is not easy and is not exact**
 - ▶ Any statement of probability in the real world is an imprecise measurement
 - ▶ We have to admit that there is always measurement error
- ▶ **There are various ways of describing imprecise probabilities**
 - ▶ Interval probabilities, Dempster-Shafer etc
 - ▶ They are not alternative ways to quantify uncertainty
 - ▶ Just ways to express imprecision in probability measurement
 - ▶ We can argue about how best to do it
 - ▶ And even whether it's worth doing
 - ▶ But to pretend we can be exact about measurement error is just as absurd as to pretend probabilities are exact

Probability is subjective

- ▶ Frequency probability only has meaning for repeatable events
 - ▶ Almost every uncertainty that is interesting to quantify relates to a one-off, non-repeatable instance
 - ▶ Parameters, the simulator, the model discrepancy
 - ▶ A probability statement about a parameter cannot mean how often it will take some value if repeated indefinitely
- ▶ Only subjective probability works
 - ▶ Probability is an expression of a person's rational degree of belief in something based on the evidence available to them
 - ▶ Subjective does not imply bias, irrationality, superstition, etc
 - ▶ Subjective probability is scientific

Probability is conditional

- ▶ **Uncertainty changes as we get information**
 - ▶ All probabilities are conditional on the available information
- ▶ **Two uncertain quantities are independent if and only if your probability distribution for one does not change when you learn something about the other**
 - ▶ Try always to think of independence this way
- ▶ **Things are rarely totally independent**
 - ▶ When we learn something, our whole uncertainty system is likely to change



Quantification

That means numbers

Measuring probability

- ▶ Uncertainty and beliefs are in your head
- ▶ The way to quantify them is to think
- ▶ Thinking may involve
 - ▶ Directly evaluating probabilities
 - ▶ Elicitation
 - ▶ Using probability theory
 - ▶ To express uncertainties of interest in terms of others that are easier to measure (elaboration)
 - ▶ Using data
 - ▶ Judgements about a model for the data that together with judgements about prior knowledge (elicitation) imply a posterior distribution
 - ▶ A combination of these
 - ▶ Data often do not relate to exactly the uncertain quantity in question
 - Data gaps
 - ▶ Elicit or model the link to adapt the posterior that comes from the data

Input uncertainties

- ▶ We must quantify
 - ▶ Point estimates or ranges are not enough
- ▶ All of the above are relevant for quantifying input uncertainty
 - ▶ Elicitation is hugely important
 - ▶ Also bridging data gaps
- ▶ Distinction between physical and tuning parameters
 - ▶ What is the true value of a tuning parameter?
 - ▶ Need to define it as giving best fit over some range or distribution in x space
- ▶ Changing with new information
 - ▶ Calibration!

Code uncertainty

- ▶ Simulators are often complex and it is unrealistic to run them many thousands of times
 - ▶ So there is uncertainty about $f(.,.)$ at any untried input configuration
- ▶ How to quantify that uncertainty?
 - ▶ A probability distribution for an uncertain function is necessarily very complex, too
 - ▶ A stochastic process
 - ▶ I use a stationary Gaussian process –
 - Both “Gaussian” and “stationary” are quite innocuous
 - ▶ Are there others that I should consider?
 - Note the discussion of validation later
- ▶ Changing with new information
 - ▶ Training data – prior to posterior

Model discrepancy

- ▶ I talked a lot about this last week!
- ▶ It's an essential component of the problem
 - ▶ Linking the simulator to reality
 - ▶ Another function, therefore quantified as a stochastic process
- ▶ Often uncertainty is high
- ▶ The most challenging area of UQ
 - ▶ Needs a lot more research
- ▶ Changing with new information
 - ▶ When we observe reality we learn about all the uncertain components of the problem
 - ▶ Inputs (calibration parameters), outputs, simulator, discrepancy
 - ▶ Calibration (learning about calibration parameters) and data assimilation (learning about the state in a dynamic model) are unrealistic simplifications

Validation

- ▶ Uncertainty quantification should be validated
- ▶ If I express 95% probability intervals for a number of uncertain things, then 95% of these intervals should turn out to contain the true values
 - ▶ Neither more not less
 - ▶ By making predictions for observables we can check this
 - ▶ Confusingly, in the world of expert judgement, this kind of validation is called ‘calibration’
- ▶ Unless the quantification of uncertainty is valid, it’s useless
- ▶ Before we use any UQ system to make important predictions, we must validate it
 - ▶ As far as possible

VALID

What can we validate?

- ▶ Our surrogate/emulator for the simulator
 - ▶ An emulator quantifies uncertainty about the simulator
 - ▶ Any surrogate that doesn't do that has no place in UQ
 - ▶ We can validate by making additional simulator runs and comparing with the emulator's probabilistic predictions
 - ▶ Any surrogate that can't be validated like this has no place in UQ
 - ▶ If it fails the validation, then revise the emulator using the validation data
 - ▶ And any other targeted runs suggested by the validation failure
 - ▶ Get new validation runs and try again
 - ▶ Validating emulators should be completely routine
 - ▶ Otherwise you are not doing UQ

What else can we validate?

- ▶ We can't directly validate uncertainty quantifications of calibration parameters or model discrepancy
 - ▶ Because we cannot directly observe them
- ▶ We can observe reality
 - ▶ But invariably with observation error
 - ▶ We can calibrate the combined UQ of θ , $\delta(\cdot)$ and the observation error
 - ▶ Remember nonidentifiability
- ▶ Note the two-fold use of observations
 - ▶ Validate
 - ▶ If valid – use observations to refine the UQ system
 - ▶ If not valid – use observations to refine the system, but this still needs validation

Fitness for purpose

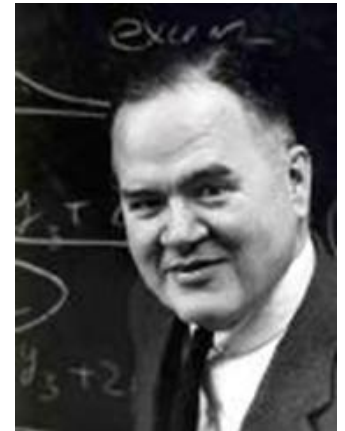
- ▶ We can have many different valid uncertainty quantifications for the same variables
 - ▶ E.g. given a sample, predict new observations using sample mean and variance, or use covariate and regression fit
- ▶ If one valid UQ system makes predictions with greater precision, then this is better
- ▶ If predictions are sufficiently precise for a given task then the UQ is fit for purpose
- ▶ Traditional ‘validation’ of computer models mixes up what I call validation with fitness for purpose
 - ▶ Validation is absolute
 - ▶ Fitness for purpose is relative to the task

The TUQ Manifesto

To the barricades!

What is TUQ?

- ▶ Total Uncertainty Quantification
 - ▶ Or TUQ for short
 - ▶ I pronounce it to rhyme with Duke
 - ▶ But that won't work for Americans, Spanish-speakers etc!
- ▶ It describes a philosophical approach to UQ that is characterised by the preceding slides
 - ▶ A commitment to doing UQ that way
- ▶ A TUQie is someone who has pledged to uphold the principles in the TUQ Manifesto
 - ▶ Apologies to the memory of John Wilder Tukey!



The Manifesto

RESTORING THE SUPREMACY AND SOVEREIGNTY OF JESUS CHRIST

JESUS MANIFESTO



Ω

LEONARD SWEET FRANK VIOLA

40 CENTS


THE Communist Manifesto

382-14

KARL MARX & FREDERICK ENGELS

INTERNATIONAL PUBLISHERS



MANIFESTE DADA 1918.

leur travail les gâteaux des vertébrés, tel que le la source d'un monde caudal à travers les amoncelles. Que ceuvre d'art n'est jamais belle, par dévotion, obéissance, pour tout. La critique est donc inutile, elle est inutile qui subvertissent, pour l'honneur, et sous le monde caudal de généralité. C'est-on avait trouvé la haine psychique condamnée à tout l'honneur ! L'honneur du monde et la haine couvent sous leurs ailes larges et non-ventilées : la mort, les béans, les journaux. Comment venait couronner l'homme ? Le principe : "sans les principes" est une apparence. "C'est-à-dire est une apparence, mais plus acceptable, couronné la méchanceté aussi. Pas de plus, il nous reste après le carnage, l'espèce d'une humanité perdue.

le par-tout de moi puisque je ne veux continuer, je n'ai pas le droit d'oublier d'autres dans mes rêves, je n'obéis personne à moi-même et tout le monde lui son art à se faire, s'il connaît le jeu mouvant en fibres vers les couches actives, ne s'effie qui descendent dans les nues sans flouir de cadavres et de squelette brutes. Malheureusement les choses partent, dans les orbites agrandies par le double, les yeux blancs comme les lèvres des anges.

▲▲▲▲▲ DADA est un bon sens d'indépendance, de refus envers la convention. Ceux qui appartiennent à une gardet leur liberté. Nous avons reconstruit jusque l'édifice. Nous avons aimé des amoncelles niches et futuristes : laboratoires d'idées formelles. Faisons l'art pour gagner l'argent et sauver les gens bourgeois ! Les rêves sont l'annonce des monstres et l'indéfini, glisse le long de la ligne de votre esprit. Tous les groupements d'articles ont abouti à cette œuvre en échevaillant sur de nouveaux courants. Le porte-œuvre une possibilité de se valoir dans les yeux blancs l'œuvre, dans la terre grasse. Il nous avons le droit de présenter, car nous avons connu les Nations et l'art. Strenuous vous d'énergie sont entourent le filon dans la chair sémoussée. Nous sommes rétrospectivement de modifications en abondance typique de végétation vertigineuse, gomme et plus est nous avec, nous saignons et brûlons la nuit, notre sang est vigiler.

Le rythme saupit de la simple façon de regarder l'objet. Chaque regard une base 30 expositions plus bas que ses yeux, les caducres la regardent tout d'un bas, d'êtres compliquant l'apparence en faisant une action perpendiculaire et en l'étranglant saignant à côté. Et l'oubli pourriss les caducres, et les grands rhinocéros et la matière qu'ils courent débilités. Le Manifeste voit la même base et nous voyons l'écriture d'édifice un à côté de l'autre, et ajoute malicieusement quelques lignes-herbes. Cela s'explique que la haine soit son tonnerre et sa rage et sa peinture débilis au placement des lignes vertigineuses. Le polone nouveau crée un monde, sous les éléments sont agités, les rêves, sont saurs et définis sans aggraver. L'œuvre nouvelle présente : il se peut plus, introduction symbolique et dissolutive, mais complètement en pièces, brutes, dans, et sous des organisations nouvelles peuvent être trouvés de tous les côtés par le mot impuissant de la sensation soumise. L'œuvre d'œuvre générale ou plastique en totale.

* en 1918 dans le CABARET VOLTVAIRE à Zurich.

Some challenges

Things worth doing

Moving forward on TUQ

- ▶ The SAMSI working groups will be looking at some things
 - ▶ Surrogates WG to look at different kinds of surrogate
 - ▶ Data assimilation and validation WGs embracing model discrepancy
- ▶ Here are a few suggestions which I think will not be covered in those WGs
 - ▶ Maybe we should have a TUQ working group? 😊
- ▶ I have an ulterior motive in proposing these
 - ▶ I am retired and I don't have postdocs or grad students of my own
 - ▶ So I try to persuade other people's postdocs/grad students to work on things I think are worthwhile!

1. Model discrepancy – simple stats

- ▶ In my talk last week, I asked
 - ▶ What do parameters mean in a model that is wrong?
- ▶ I looked at a couple of simple statistical models
 - ▶ Poisson sample
 - ▶ Simple regression
- ▶ I think there is a nice paper in there, but more work is needed
 - ▶ More exploration of those or other examples

2. Validation diagnostics

- ▶ This is a little-researched area
- ▶ We have some diagnostics built on traditional regression diagnostics (looking at residuals)
 - ▶ But we need more ideas
 - ▶ In particular, what role could scoring rules play?

3. DA with GP emulator

- ▶ **Methods like the ensemble Kalman filter represent uncertainty using samples**
 - ▶ One reason is the difficulty of propagating uncertainty through the simulation model
- ▶ **If we have a GP emulator for the simulator, we can do this propagation analytically**
 - ▶ At least for means and variances
 - ▶ And then make the Gaussian assumptions that EnKF makes

4. Multi-output calibration

- ▶ **Most simulators produce multiple outputs**
 - ▶ We may have observations on several different outputs
 - ▶ We want to combine these to get good calibration
- ▶ **Issues to be explored**
 - ▶ For physical parameters we need to model multivariate discrepancy (carefully!)
 - ▶ For tuning parameters, should we allow different values for different outputs?

References

- ▶ **UQ for model inputs (elicitation, elaboration, Bayes)**
 - ▶ O'Hagan, A. (2010). Probabilistic uncertainty specification: expert judgement of input uncertainties in a mechanistic model of carbon flux. To appear in Environmental Modeling and Software. (<http://www.tonyohagan.co.uk/academic/pub.html>)

- ▶ **Validation diagnostics**
 - ▶ Bastos, L. S. and O'Hagan, A. (2009). Diagnostics for Gaussian process emulators. Technometrics 51, 425-438.