Calibrate your model!

Jonathan Rougier

Department of Mathematics
University of Bristol

mailto:j.c.rougier@bristol.ac.uk

BRISK/Cabot/CREDIBLE Summer School:
Uncertainty and risk in natural hazards

7-11 July 2014

Setting the scene

Although quantitative models vary very widely, the following scheme seems to cover the vast majority of cases:

$$\text{sim}(s_i; x, \theta) \quad i = 1, \dots, n$$

where:

- **'sim'** Short for 'simulator', to be preferred to 'model'.
- s_i An index of the simulator output, perhaps a location, but in general a tuple representing location, a time, and a type.
- x Control and concomitant variables. Control variables are set by the experimenter, while concomitant variables are uncontrolled but measurable (such as environmental variables).
- θ Simulator parameters. These tend to be imperfectly known; sometimes they have no system analogue at all.

Various approaches

In all cases, the basic assertion is that there exists a parameter value θ^* such that

$$\mathrm{system}(s_i; x) \approx \mathsf{sim}(s_i; x, \theta^*) \quad \text{for all } s_i \text{ and } x.$$

The 'universality' of θ^* is crucial.

Given some system observations (which may include measurement errors) three popular 'levels' of calibration are:

1. History matching. Ruling out bad candidates for θ^*.
2. Tuning. Finding an acceptable candidate for θ^*.
3. Calibration. Computing a probability distribution for θ^*.

I use 'calibrate' to cover all three, but in fact we'll only be looking at (1).

Three golden rules

1. Focus on a small number of parameters.

   It's best to focus on the parameters that you think will be important, or to determine these in an initial screening study.

2. Calibrate against a small number of outputs.

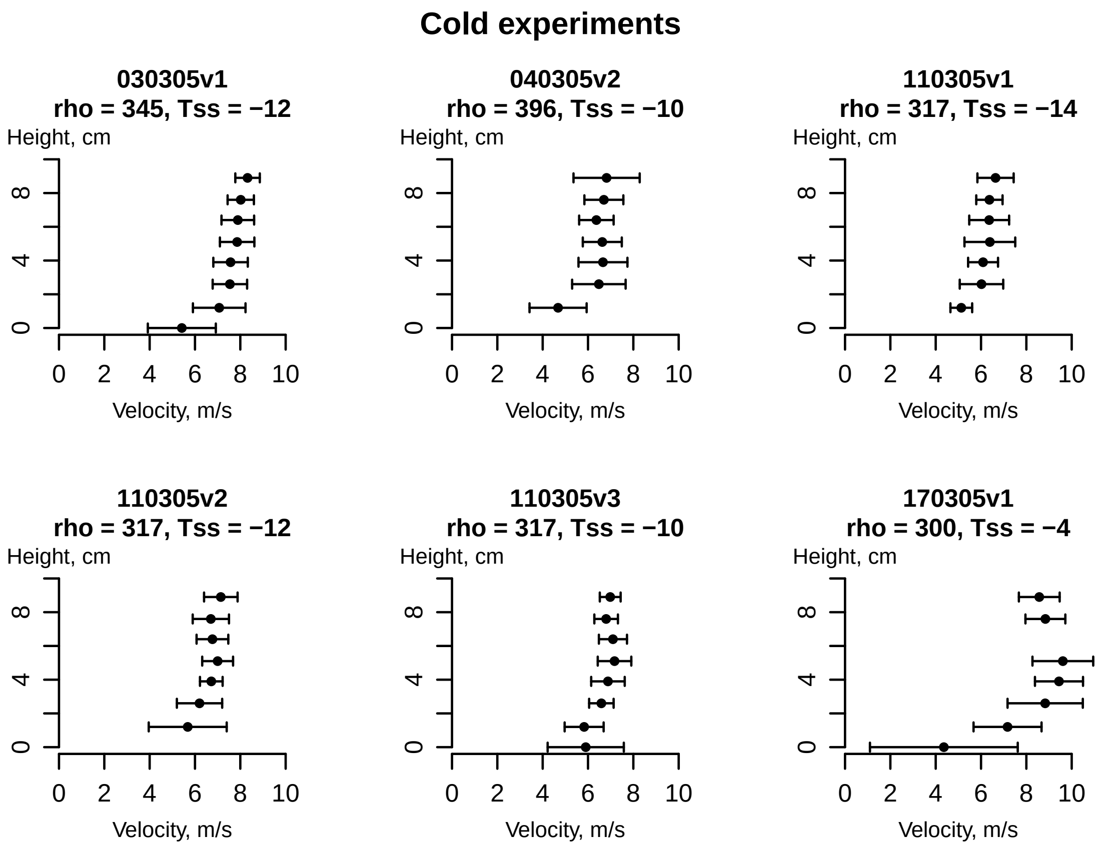
   Identify the 'crucial' outputs that are good for discriminating good from bad candidate values for θ^*, perhaps consider clumping or thinning outputs. (Re-parameterise spatial fields.)

3. Proceed sequentially.

   Perhaps an initial screening study to identify important parameters and crucial outputs. If possible, always use sequences of batches of simulator runs to successively refine the ranges of the candidate values for θ^*.

Enough with the scene-setting!

The experimental data



The Herschel-Bulkley Rheology Model

Gives us a semi-empirical model for the velocity profile:

$$v(z) = \begin{cases} v_h + (v_0 - v_h)\left(1 - \frac{z}{h}\right)^{\frac{1+\alpha}{\alpha}} & 0 \le z < h \\ v_h & z \ge h. \end{cases}$$

where

$$v_h = v_0 + \frac{h}{t_c}\frac{\alpha}{1+\alpha}\left(\frac{h}{H-h}\right)^{1/\alpha}$$

and h solves $\tau_c = (H-h)g\rho\sin\theta$.

$v(\cdot)$	Velocity (m/s)	θ	Inclination (°)
z	Height ordinate (m)	ρ	Snow density (kg/m^3)
h, H	Height to plug layer, top of flow (m)	α	Stress coefficient
v_0, v_h	Basal, plug-layer velocity (m/s)	τ_c	Stress (Pa)
T_{ss}	Snow-surface temperature (°C)	t_c	Time constant (s)
g	Acceleration (m/s^2)		

Write a wrapper for the observations

- On the one hand, we have a collection of observations, often arising from several different experiments, y^{obs} say, where

$$y^{\text{obs}} = \{y_1^{\text{obs}}, \ldots, y_n^{\text{obs}}\}.$$

- On the other, we have our numerical simulator for what happens in a single experiment, sim.
- Our wrapper combines these two together: it is a script that uses the numerical simulator to generate pseudo-observations that match the experimental results, as a function of the simulator parameters

$$\text{wrap}(\theta^{(1)}) \to y^{(1)}$$
$$\text{wrap}(\theta^{(2)}) \to y^{(2)}$$
$$\vdots$$
$$\text{wrap}(\theta^{*}) \to y^{\text{obs}} \quad (?)$$

First stage of the experiment

Although it is ill-advised to give a firm recommendation, think about spending about 30–50% of your computational budget on this first stage.

1. Choice of 'active' parameters (transformations?); set ranges for each active parameter.
2. Initial exploratory design varies all parameters at once: a space-filling maximin latin hypercube is very popular.
3. Identify a subset/function of the outputs and plausible ranges for these, possibly Principal Variables Analysis.
4. Parallel Coordinates Plot for screening, colouring by implausibility.

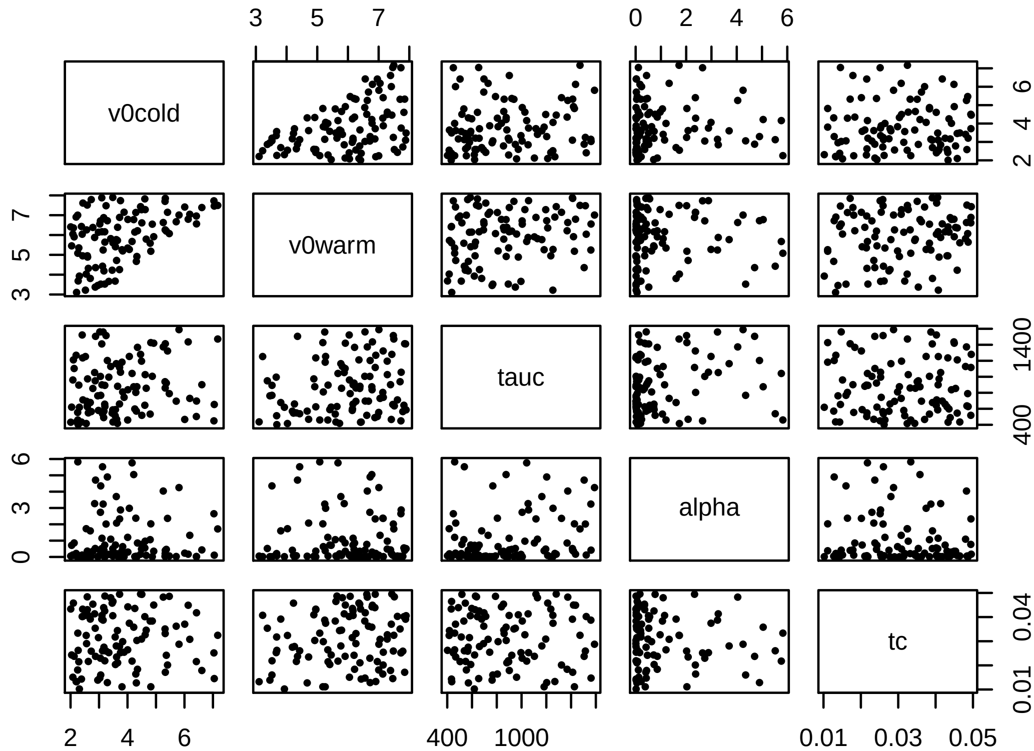
And then iterate ...

Space-filling designs

Sometimes you just want to put n points into a rectangular subset of p-dimensional space.

- Putting them in at random is simple, but clumpy. A better alternative is a maximin latin hypercube.
- Maximin designs attempt to maximise the minimum inter-point Euclidean distance. A latin hypercube has uniform 1D margins.
- Sooner or later you will also want to extend your design with additional points. Again, we can do that in a space-filling way using the maximin criterion.
- And the same goes for shrinking your design. One reason to shrink a design is to impose additional constraints by generating too many rows, and then rejecting some of them.

2D margins of the first design (100 runs)



Implausibility

Implausibility is a measure of the consistency between a specified value of θ and the observations.

- For a single output j, define the implausibility of parameters θ as

$$\mathcal{I}_j(\theta) := \frac{|y_j^{\text{obs}} - \text{sim}(s_j; x_j, \theta)|}{\sqrt{\sigma_j^2 + \tau_j^2}}$$

  where σ_j^2 is the observation error variance, and τ_j^2 is the structural error variance (which you must specify).

- Structural Error: the error in your simulator that cannot be tuned away through careful adjustment of the parameters.

- For a set of outputs $\mathcal{J}$ (this is one popular definition)

$$\mathcal{I}_{\mathcal{J}}(\theta) := \max_{j \in \mathcal{J}} \{ \mathcal{I}_j(\theta) \}.$$

Principal variables

Often, it is a lot more convenient to work with a subset of the outputs than it is to work with all of the outputs, or with linear combinations of the outputs. Principal variables is a method for choosing a good subset.

J.A. Cumming and D.A. Wooff, 2007, Dimensional Reduction Via Principal Variables, *Computational Statistics and Data Analysis*, **52**, pp. 550–565.

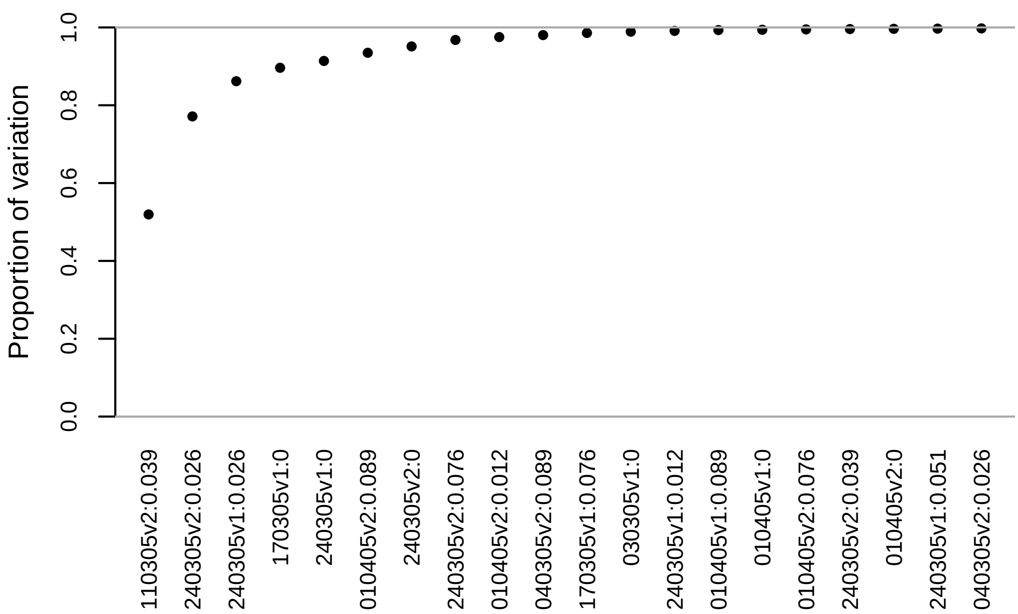
- There is one wrinkle for calibration. We would like to choose a subset $\mathcal{J}$ such that if $j \in \mathcal{J}$, then j being implausible makes a lot of other outputs not in $\mathcal{J}$ implausible as well. So we apply principal variables on the matrix of implausibilities,

$$I_{ij} = \mathcal{I}_j(\theta^{(i)})$$

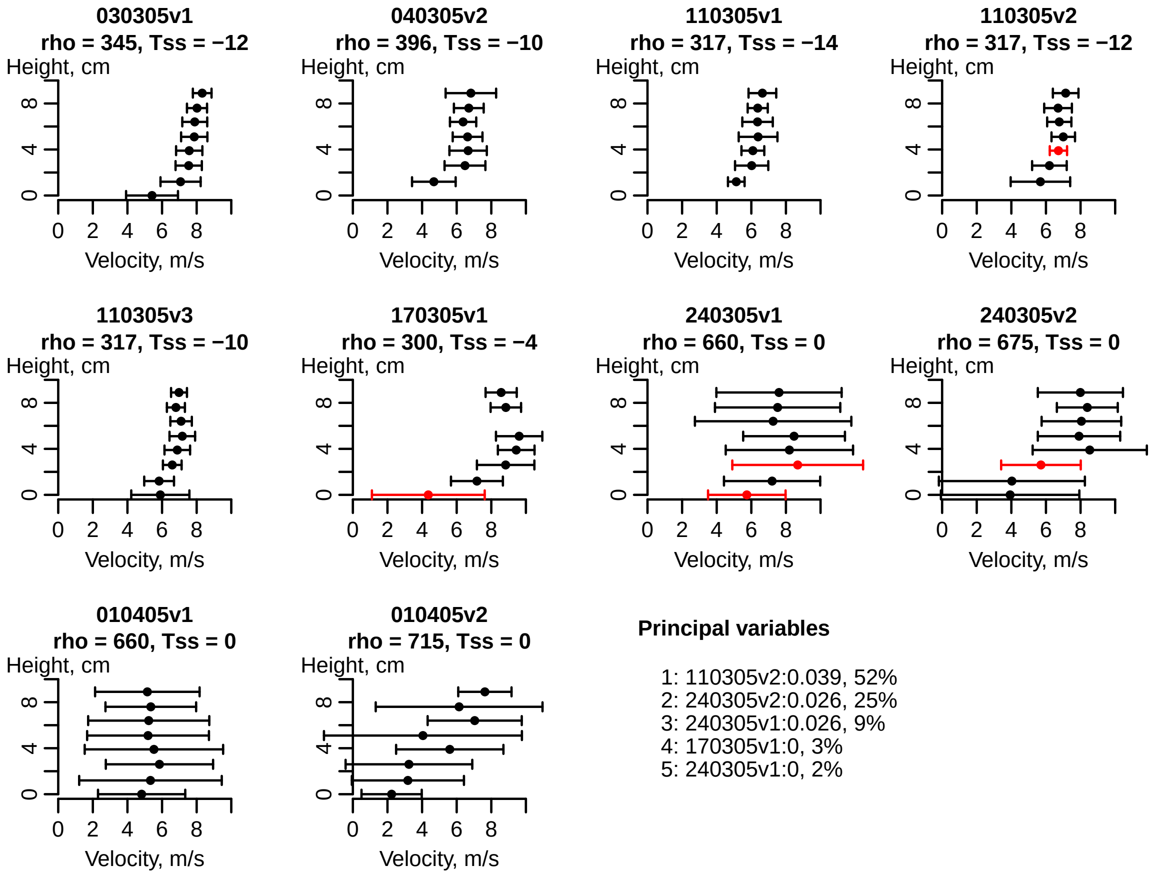
where $\theta^{(i)}$ are the parameter values for the ith run.

Scree plot for principal variables

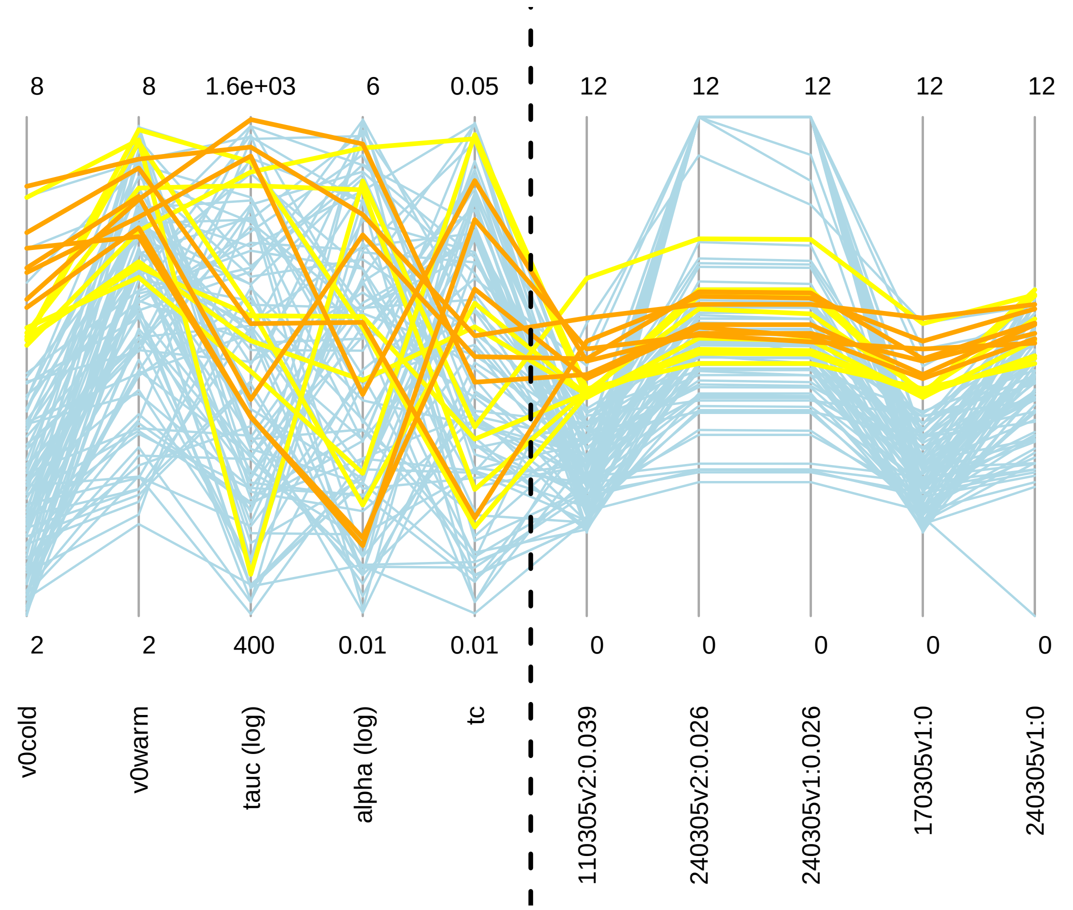
This is computed on the matrix of implausibilities, and shows the proportion of explained variation.



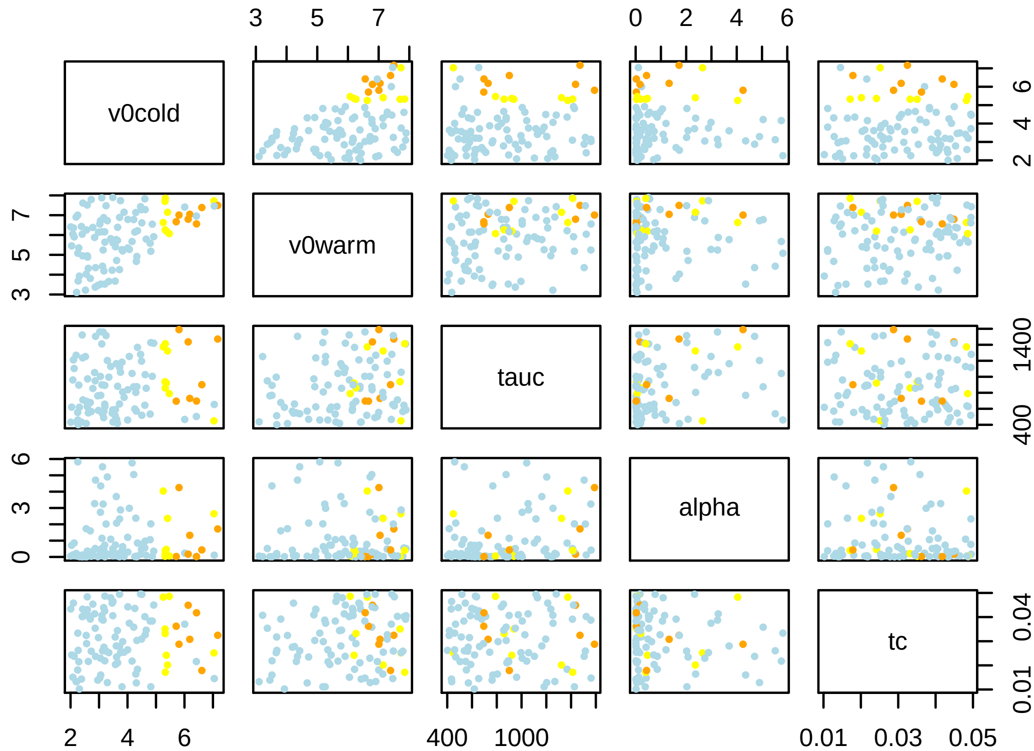
Location of first few principal variables



Parallel coordinate plot coloured by implausibility



2D margins coloured by implausibility

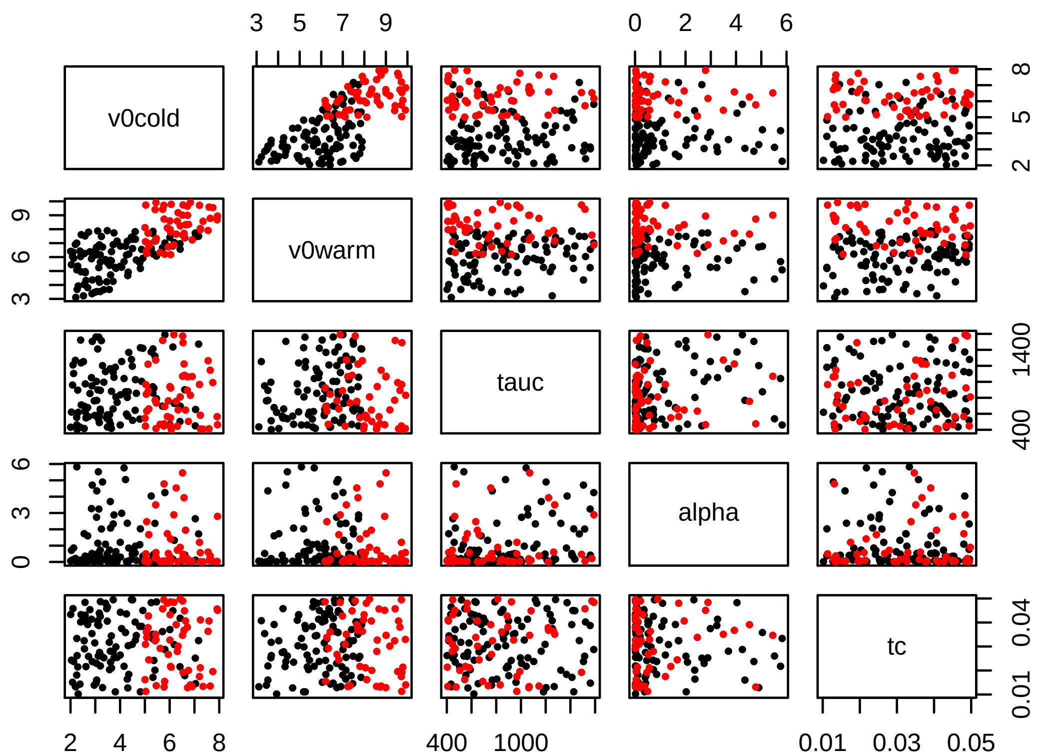


Time to take stock

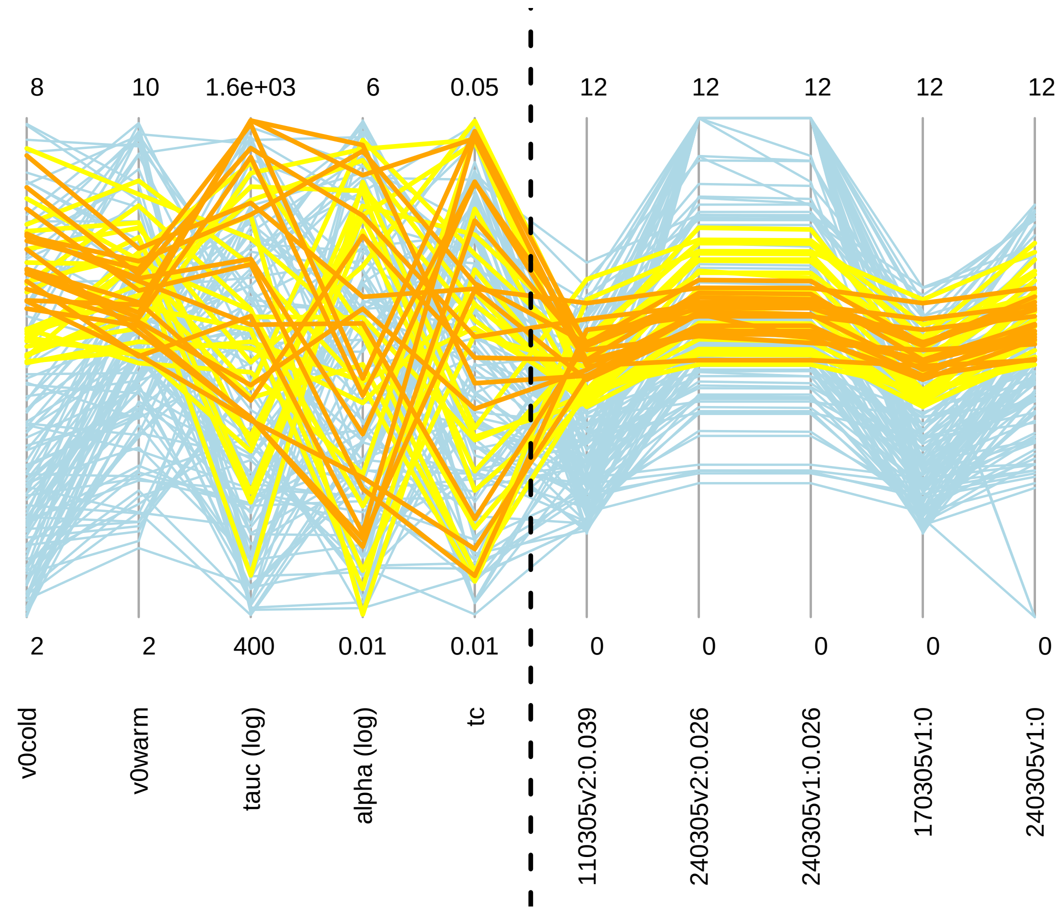
- It looks as though I can raise the lower bound on `v0cold` to 5 m/s, and I also want to push upper bound of `v0warm` to 10 m/s. Of course I still want to maintain `v0cold` $\leq$ `v0warm`.
- These types of adjustments are fiddly but extremely common. The `boxdesign` functions in the `calibrate` package are your friend.
  1. Generate too many new points (with the updated ranges). These are selected to be space-filling wrt the current design.
  2. Impose restrictions such as `v0cold` $\leq$ `v0warm` by deleting rows.
  3. Still too many points?—choose a space-filling subset of the rows that remain.
  4. Apply the wrapper to the resulting design, and then combine the results (both design and outputs).

  Note: I tend to think of all of the runs together, and seldom have a need (apart from sanity checking) for keeping them separate.

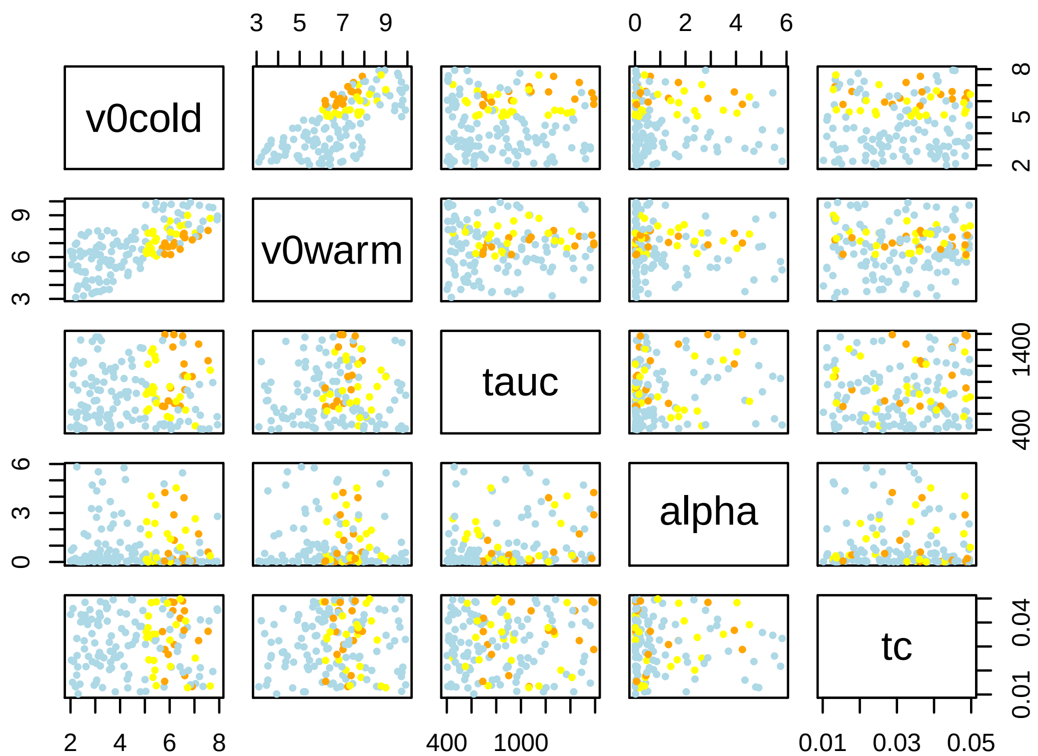
2D margins of the first & second designs (100 + 50 runs)



Parallel coordinate plot (both designs) coloured by implausibility



2D margins (both designs) coloured by implausibility

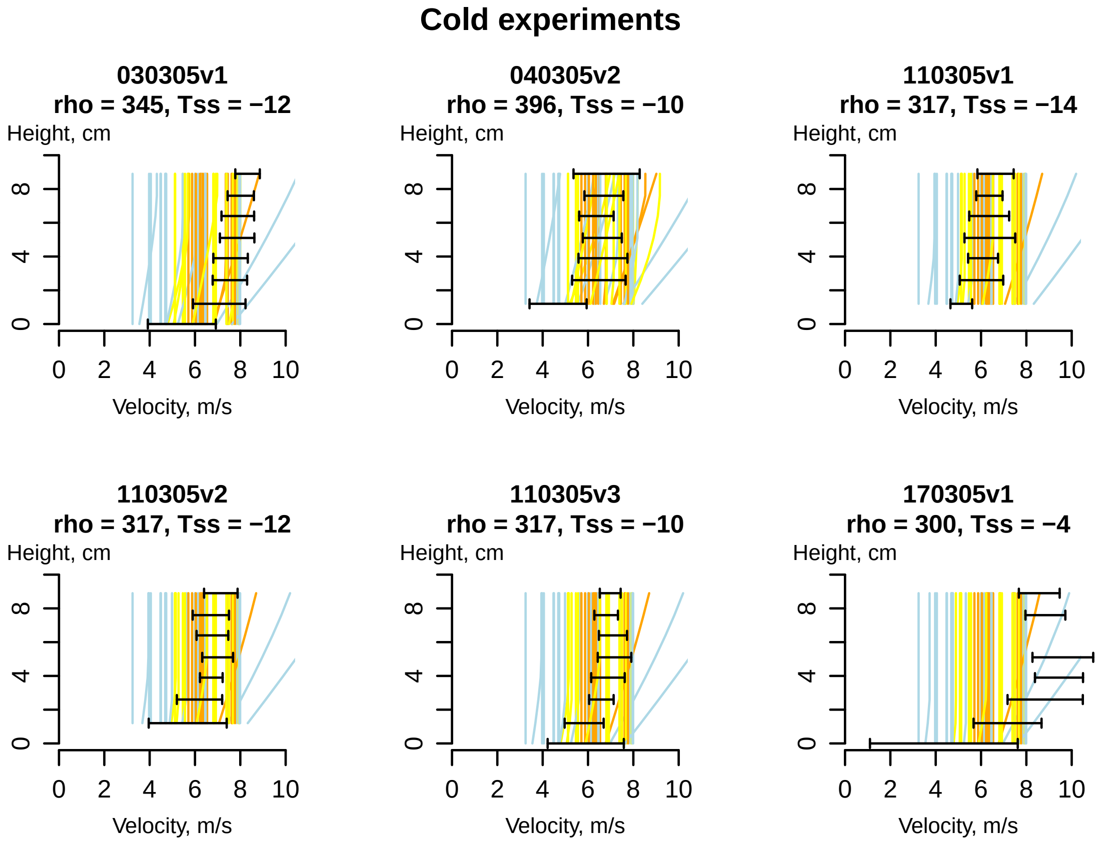


The next stage

It looks as though I could push the upper bounds of `tauc` and `tc` higher, but let's call it a day!

- Sooner or later you will get tired of tweaking the ranges of the individual parameters. Then you will want to blow some/the rest of your computing budget on runs with parameter values which are not ruled out by the observations.
- For a threshold on not-ruled-out, $\mathcal{I}_{\mathcal{J}}(\theta) \leq 3$ is a sensible choice, that arises originally from the 3-sigma rule. Points with implausibilities below this threshold can be thought of as elements of a 95% confidence set for θ^*, the 'true but unknown' parameter.
- It is a common task to generate new points that are like a subset of existing points, implemented in the `morepoints` function.

How did the 'similar to not-ruled-out' points do?



Resources

- Download the `calibrate` package for the R computing environment from `http://www.maths.bris.ac.uk/~mazjcr/#software`. Send me an email if you are using it and I will keep you updated.
- Download the script file for creating all the pictures from the same place: `avalanche2014.R` is the code and `chute.txt` the observations. These slides are available at `calibrateYourModel2014.pdf`.
- This application was more complicated than presented here. More details are available in our paper, at `http://dx.doi.org/10.1111/j.1467-9876.2010.00717.x` although note that this was not an exercise in history matching.