UNIVERSITY OF LEEDS

CYPHER SUMMER SCHOOL

GUI-HDMR demonstration

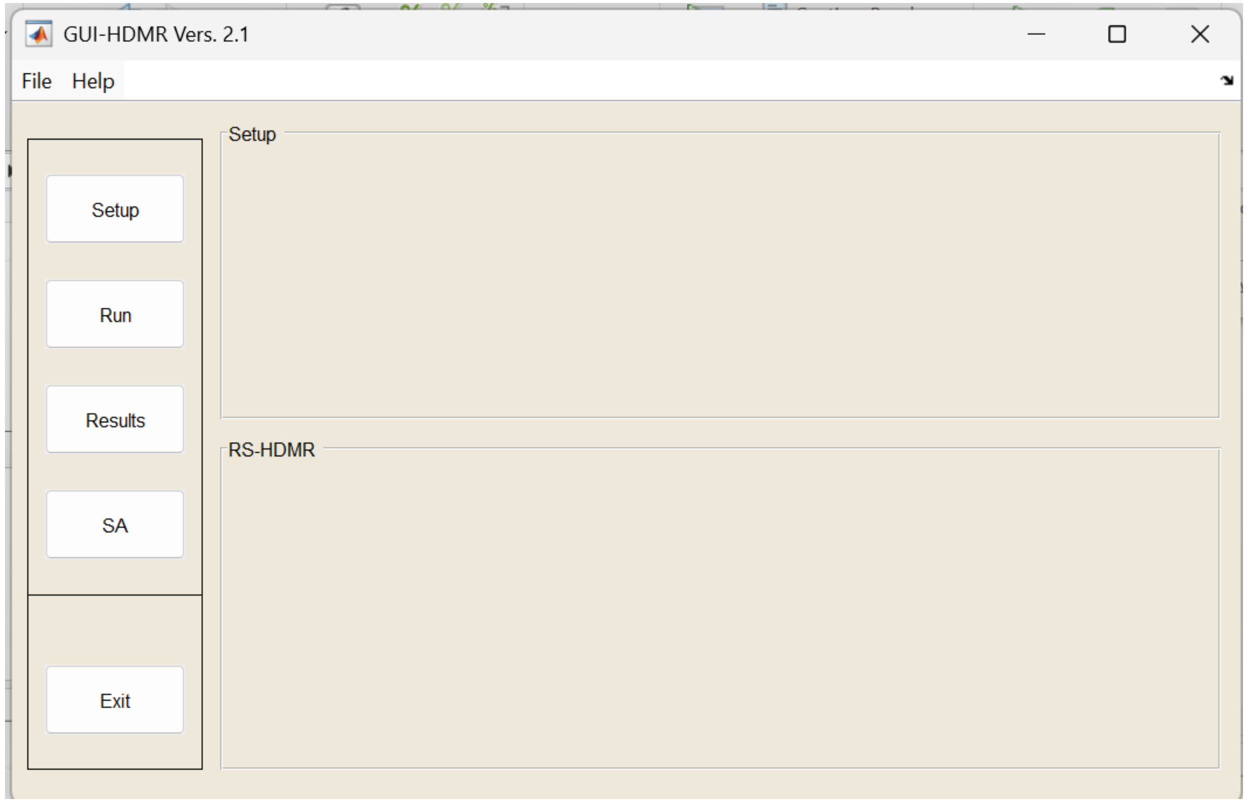
Professor Alison Tomlin
University of Leeds
School of Chemical
and Process Engineering

GUI-HDMR

- GUI-HDMR is a freely available Matlab toolbox with a graphical user interface.
- The software provides a straightforward and efficient approach to explore the input-output mapping of a complex model with a large number of input parameters.
- The mapping is generated across a sample by running the detailed model.
- The HDMR code then fits a response surface model to the input-ouput mapping.
- The larger the sample that can be afforded, the better the fit will be.
- Variance based sensitivity indices are determined in an automatic way in order to rank the importance of model input parameters and to explore the influence of parameter interactions.
- The set of input values can be any Monte Carlo sample. However, a quasi-random sampling method is preferable.
- This guarantees that the input space is covered more uniformly than by using random values and it provides a better convergence rate.

Opening the code

- If using Matlab directly open Matlab and move to the directory where you saved **Working code 2025. Open this directory.**
- Type gui_hdmr in the Matlab command line.
- If using the pre-compiled executable using Runtime you should just be able to double click on the downloaded executable file (assuming Runtime has installed correctly).



The GUI window should appear.

Click on **Setup** and we will begin to upload files.

Ishigami test function

- We will start with one of the test functions used to validate the code since it has only 3 input parameters and has analytical solutions for the variance based sensitivity indices.
- See: Ziehn, T., Tomlin, A. S., 2009. GUI-HDMR - A Software Tool for Global Sensitivity Analysis of Complex Models. Environmental Modelling & Software, doi:10.1016/j.envsoft.2008.12.002.
- The Ishigami function has 3 input parameters and a non-linear response:

  $f(\mathbf{x}) = \sin(x_1) + a\sin^2(x_2) + bx_3^4 \sin(x_1)$,

  where x_i is uniformly distributed within $(-\pi, \pi)$ and the constants $a = 7$, $b = 0.1$.
- We can see the dependence on x_3, which has no additive effect (first order) on the model output but contributes via interaction with x_1.

Testing different HDMR set-ups

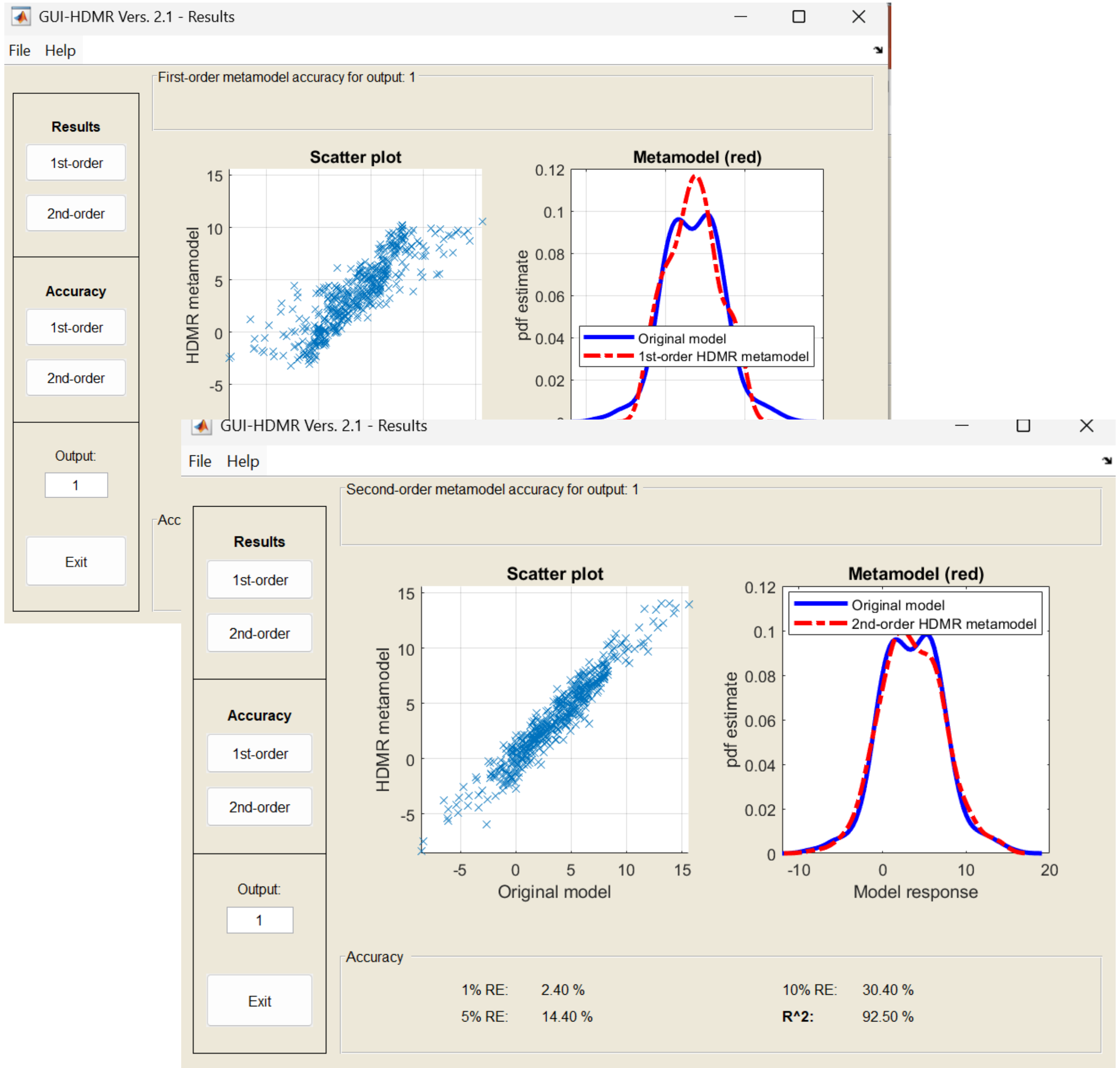
Sensitivity Index	Analytical result
S_1	0.3139
S_2	0.4424
S_3	0
$S_{1,3}$	0.2437
$S_{2,3}$	0
$S_{1,2}$	0

From the setup window load up the input, output and ranges files.

There are 10,000 available samples. Start with a small sample and build up.

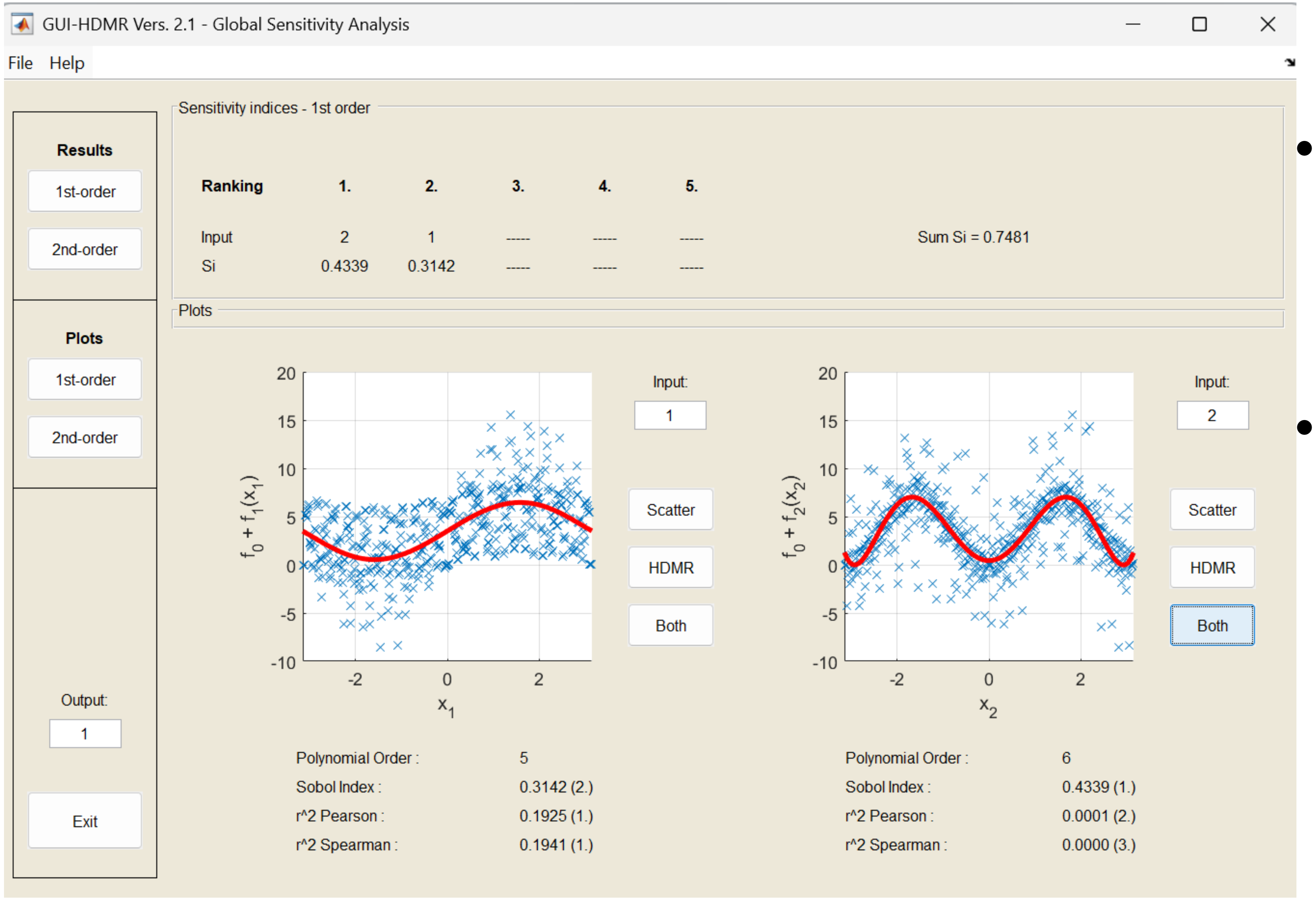
See how close you can get to the analytical results by testing different sample sizes and polynomial orders.

Results window



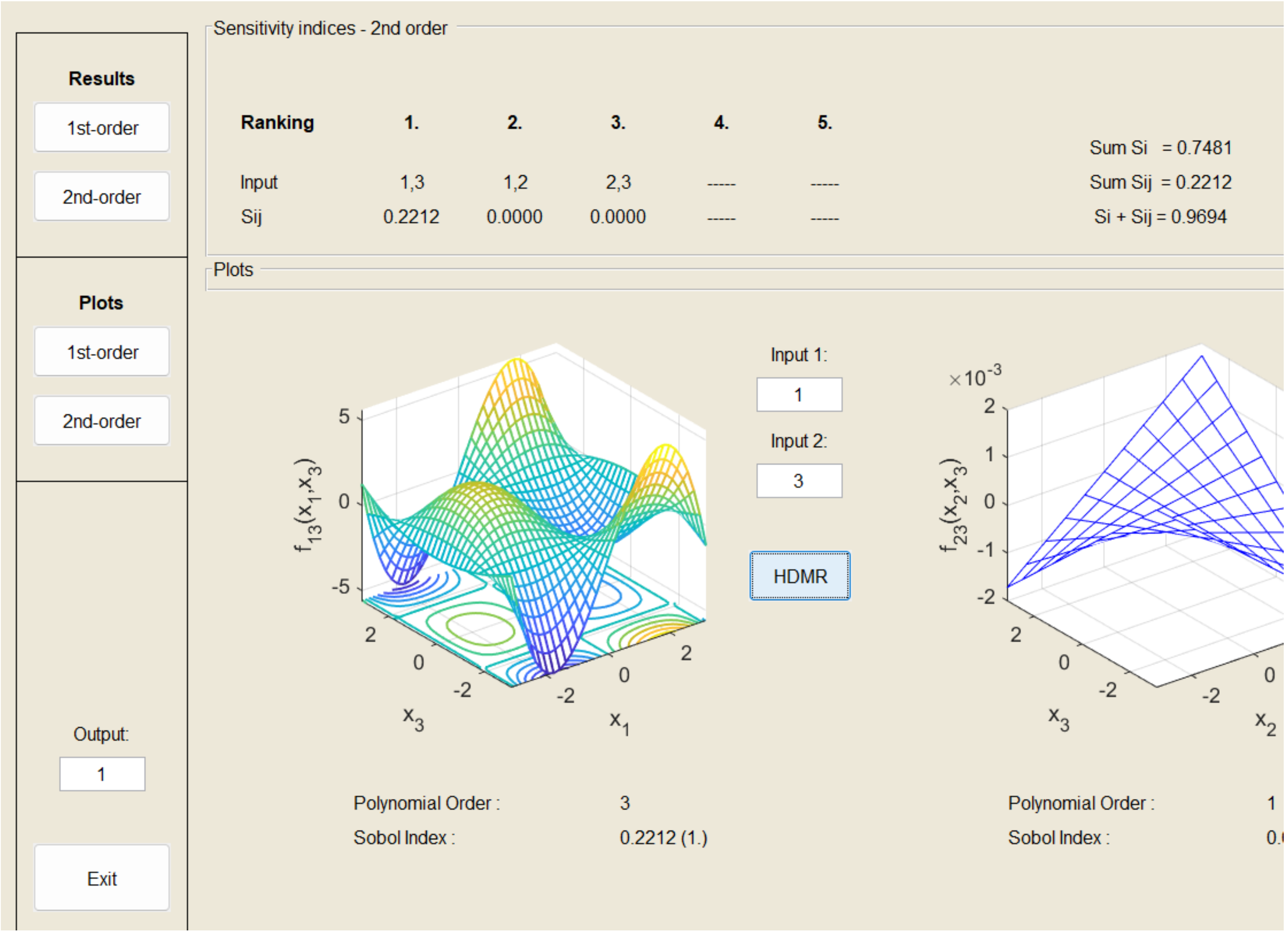
- Results window shows you how close your fitted HDMR model is to the original data set.
- Unsurprisingly, the first order model using a small sample size not great.
- However, even with 256 points the second order model is reasonable: R^2 = 92.5 %.
- See if you can do better.

SA window – 1st-order



- You can plot the scatter plots from original data, the HDMR component function or both.
- The component function shows the response to the selected parameter and the scatter the variability due to the other parameters.

2^{nd} – order



- Correctly, only $S_{1,3}$ has been fitted as being non-zero.
- The sum of the indices is 0.9694 for this 2048 sample.
- Should be 1 so a small error..

Street Canyon Modelling Test Case

- The model was developed to help understand the causes of high air pollution (particularly NO_2) within narrow street canyons in the city of York, UK.

- High traffic volumes lead to congested streets and vehicle high emissions.
- Constricted air flows and flow recirculation zones are caused by the “canyon” shape of the streets.
- There are a lot of factors affecting NO_2 concentrations.
  - Traffic demand and emissions factors.
  - Wind speed and direction.
  - Levels of turbulence and boundary layer descriptions.
  - Fast time-scale chemistry.

Model description

- A coupled modelling systems was developed including:
  - Micro-scale traffic modelling, including behaviour at signalled junctions.
  - Vehicle emissions modelling based on individual traffic parameters.
  - A turbulent flow model (k-epsilon based), coupled with a Lagrangian model for pollutant dispersion.
  - NOx/Ozone chemistry.

Arrhenius parameters for reactions:

R1 $O + O_2 + N_2 = O_3 + N_2$

R2 $O + O_2 + O_2 = O_3 + O_2$

R3 $O + O_3 = 2O_2$

R4 $O + NO + M = NO_2 + M$

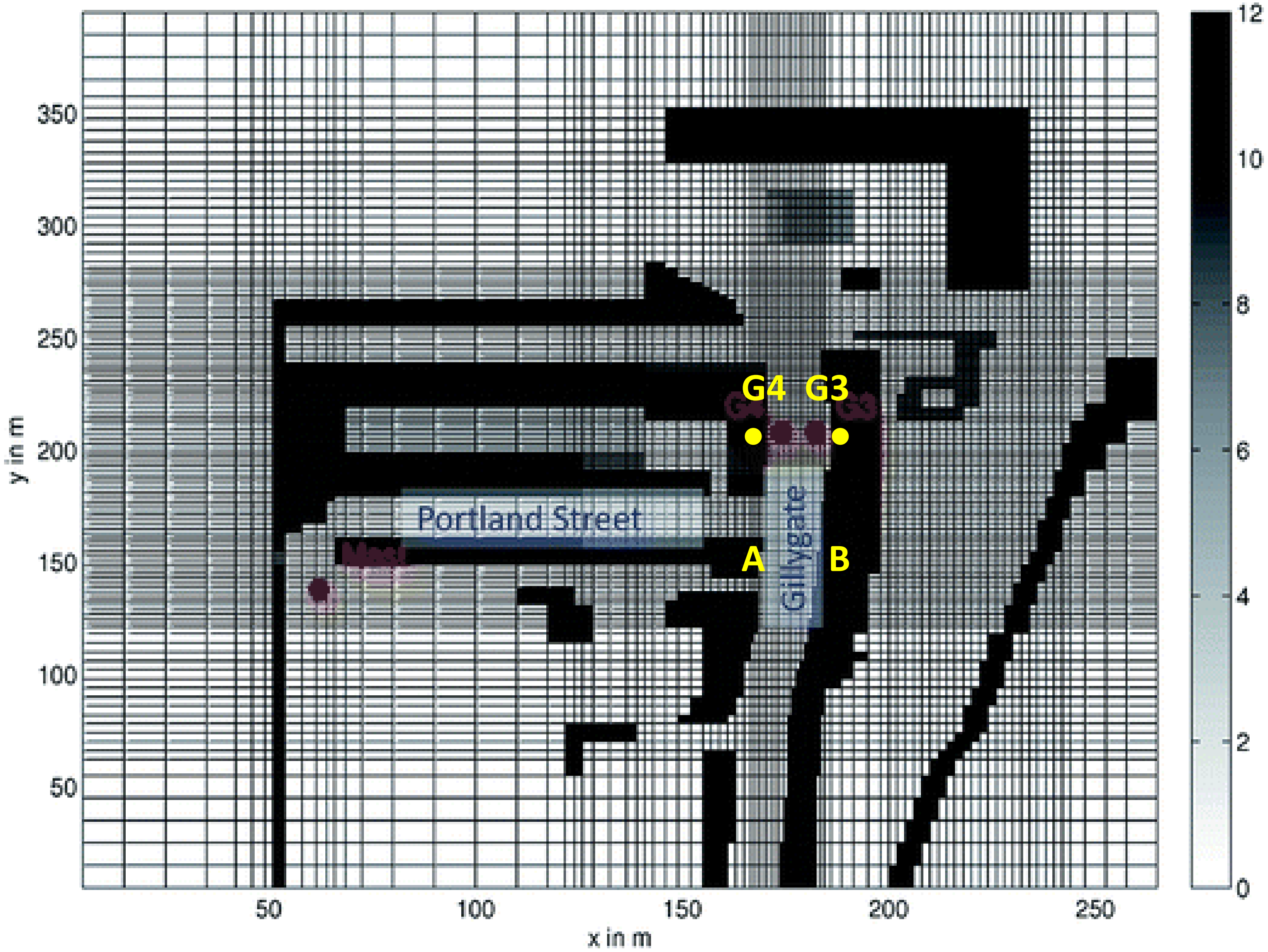
R5 $O + NO_2 = NO + O_2$

R6 $NO + O_3 = NO_2 + O_2$

Photolysis rate J, R7 $O_3 = O + O_{2,}$ R8 $NO_2 = NO + O$

Paper: The treatment of uncertainties in reactive pollution dispersion models at urban scales, 2016, AS Tomlin, T Ziehn, P Goodman, JE Tate, NS Dixon, Faraday Discussions 189, 567-587

Model grid



- G3 and G4 are locations within Gillgate where measurements were taken over 1 month.
- Other locations in the streets were modelled.
- 26 parameters were varied in total as described in:

parameters_combined.txt

Running HDMR for the canyon model

- The whole model is computationally expensive and so only ~500 model runs could be afforded.
- The importance ranking of the different parameters may vary with location in the street. You have been given modelled NO_2 concentrations for 4 locations (G3, G4, A, B) within an output file.
- Files needed are:
  - sobol_550_26.txt
  - NO2 concentrations.txt
  - ranges.txt

Load up these files and try to get the best fits with a 512 sample size.

This will be much slower to run than the Ishigami function with 26 input parameters.

How to avoid overfitting

- Try first without using variance reduction methods or setting a threshold for the sensitivity indices.
- Do you get large numbers of possibly spurious 2^{nd} order component functions?
    - Does this affect the 2^{nd} order accuracy?
- What happens when you use say the ratio control method and set thresholds for the component functions?
- With a large number of parameters and small sample size, the variance reduction methods do make quite a difference to the accuracy.
- How much of the variance is described by your best model fit?

Final Remarks

- This is not a **black box** method (in common with other fitting methods like NN for example).
- The success of the HDMR fitting (and thus the calculation of sensitivity indices) depends on the sample size and the nature of the response surface.
- For RSs with 2nd order interactions, you will need a much larger sample size to achieve a good quality fit.
- For models with only 1st order effects, 256 runs using a low discrepancy sample can often be enough.
- Some trial and error is involved, since you do not know a priori whether you have significant 2nd order effects.
    - These are likely to be present when < 80% of variance is represented by 1st order indices.
    - The scatter plot (HDMR vs. full model runs) will have unevenly spread residuals about the straight line for the 1st order fit if higher order terms are important.
- If you discover 2nd order effects are present you can add to the sample size and re-run the HDMR to achieve better representation of the 2nd order terms.
- The variance reduction methods can really help to exclude spurious 2nd order component functions that affect the final accuracy.