Sensitivity Analysis with Functional Inputs

Max D. Morris
Department of Statistics
Iowa State University

MASCOT-NUM, St-Etienne

- **Sensitivity Analysis**: Understanding the "overall impact" of individual inputs or groups of inputs on the output of a computer model.
- **Computer Model**: Focus on deterministic models – numerical implementations of explicitly or implicitly defined functions.
- **Today**: Review and propose a few approaches for extending popular sensitivity/uncertainty ideas developed for scalar-valued inputs to:
  - models for which some inputs are themselves functions, and the output of interest is a scalar. (In fact, the output may be a scalar-valued summary of a function.)
  - further, focus on input functions of one variable, e.g. time.

Examples involving time-varying inputs:

- Regional environment models. Boundary conditions may be time-varying functions.
- Chemical reactor models. "Forcing functions" including temperature, concentration, physical mixing rates.
- Groundwater hydrology models. Rainfall rates, pumping rates.
- Injection molding process models. Heat and pressure schedules.

Notation and Restrictions:

- Model inputs: $(x_1...x_m, z_1(t)...z_n(t)) \in \Delta$
- Model output: $y = f(x_1...x_m, z_1(t)...z_n(t))$
- Attention here is focused on scalar $t \in [0,1]$, where $z_i(t)$ is continuous and "well-behaved"
- Will sometimes substitute a long vector of values over a t-grid for the function:

$$z_i(t) \rightarrow \mathbf{z}_i = \begin{pmatrix} z_i(0.00) \\ z_i(0.01) \\ z_i(0.02) \\ ... \\ z_i(1.00) \end{pmatrix}$$

Three Basic Approaches popular with scalar-input problems, in decreasing order of the number of function evaluations generally required:

- *Variance-based sensitivity analysis* – A multivariate probability distribution is specified for $\mathbf{x}$ over its domain Δ, representing (ideally) situational uncertainty about $\mathbf{x}$. The goal is to understand how variability propogates to y. (e.g. Saltelli et al., 2000)
- *Statistical surrogate-based sensitivity analysis* – y is assumed to be a relatively "well behaved" function of $\mathbf{x}$ that can be formally predicted or estimated via statistical modeling. Sensitivity of y to each x_i is assessed through model parameters (Welch et al., 1992), by computing variance-based indices on the estimate of f, or via a more formal Bayesian approach (Oakley & O'Hagan, 2004).
- *Simple approximation-based sensitivity analysis* – The sensitivity of output to each input is assessed by numerical approximation to $\partial y/\partial x_i$, $i = 1, 2, 3, ..., m$, or to an average of these quantities over Δ or some appropriate subregion (e.g. $\pm 1\%$ about nominal values).

A Toy Function for Examples:

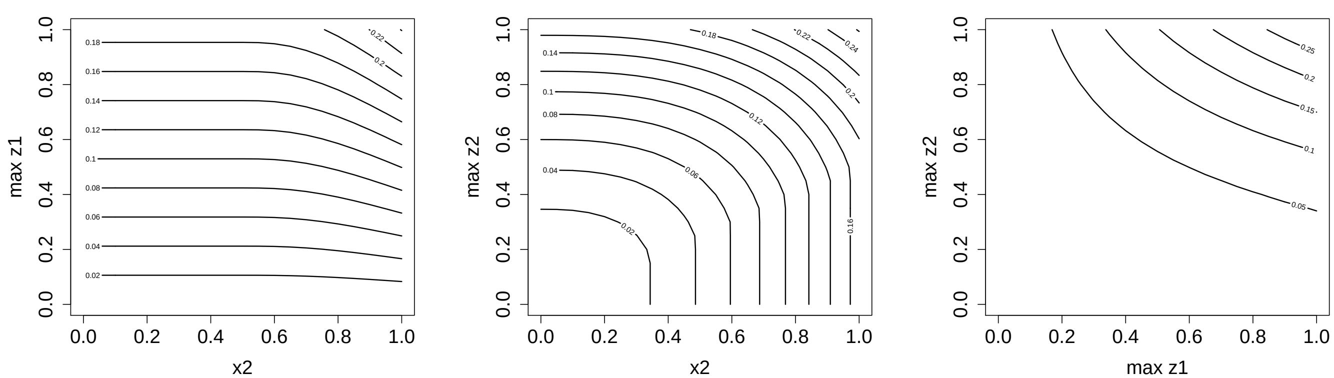
$$y = f(x_1, x_2, z_1, z_2) = \int_{t=0}^{1} \max_{s\in(0,t]} z_1(s) \times \max[(1-t)x_2, z_2(t)]^2 dt$$

- Note that x_1 does nothing

Some pictures:

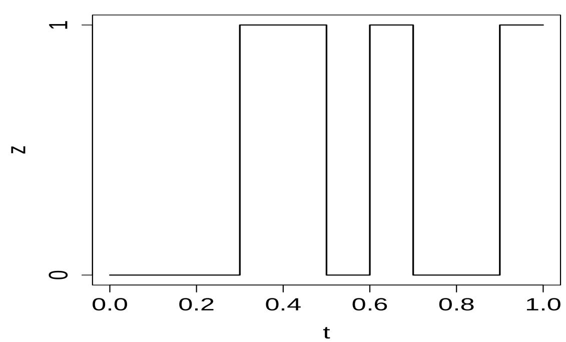
- $z_i(t) = \begin{cases} 2t \max z_i & t < \frac{1}{2} \\ 2(1-t) \max z_i & t \geq \frac{1}{2} \end{cases} \qquad i = 1, 2, \quad \max z_i \in [0, 1]$

- Unreferenced x or z in each panel $= \frac{1}{2}$



1. Simple approximation-based sensitivity analysis

- Fruth, Roustant, and Kuhnt (2014)
- Restrict attention to input functions $z(t)$ that are:
  - piece-wise constant on intervals defined by a grid on t,
    $$G = \{0 = t_0 < t_1 < t_2 < ... < t_g = 1\}$$
  - take one of only two values within each interval



- Use a form of sequential bifurcation (Bettonvil, 1995) to progressively refine G. (Important, but I won't consider this aspect here.)

- For a given G, let $\mathbf{z}_i = (z_{i1}, z_{i2}, ...z_{ig})'$.
- Then $y = f(z_1(t), z_2(t), ..., z_n(t)) = f^*(\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_n)$, i.e. reduction to $g \times n$ two-level scalar-valued inputs ... there is much experimental design literature for this case.
- Define "centered" input values z as $\bar{z} = z - \frac{1}{2}$, so that 0 is the "nominal" value for each input, and $\bar{z} = \pm\frac{1}{2}$.
- The authors use least-squares to fit data from N model runs:

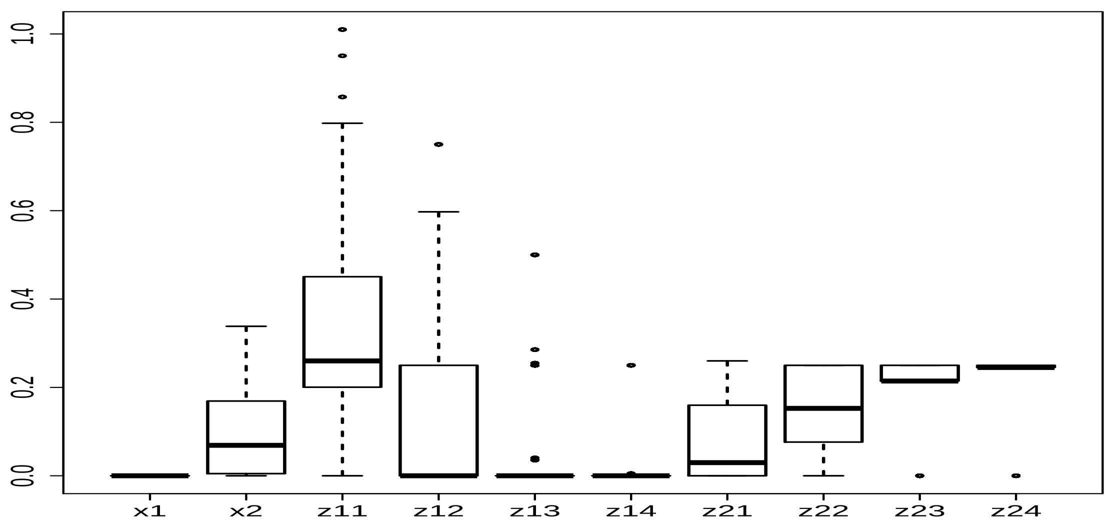
$$(\hat{\alpha}, \hat{\beta}_{ik}, i = 1...n, k = 1...g) = \text{argmin} \sum_{j=1}^{N} [y^j - (\alpha + \sum_{i=1}^{n} \sum_{k=1}^{g} \bar{z}_{ik}^j \beta_{ik})]^2$$

- Then use

$$\hat{H}_{ik} = \hat{\beta}_{ik}/(t_k - t_{k-1})$$

as an index of the sensitivity of y to the value of z_i within the kth interval of the t-grid, normalized to be expressed on a per-unit basis of t.

- What should we hope to be estimating here?
- Suppose $G = \{0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1\}$
- Test function inputs are represented by 2 x's and 8 scalar-valued z's.
- Each input is then associated with $2^{10-1} = 512$ "slopes" associated with the edges of a 10-dimensional hypercube ... here they are:



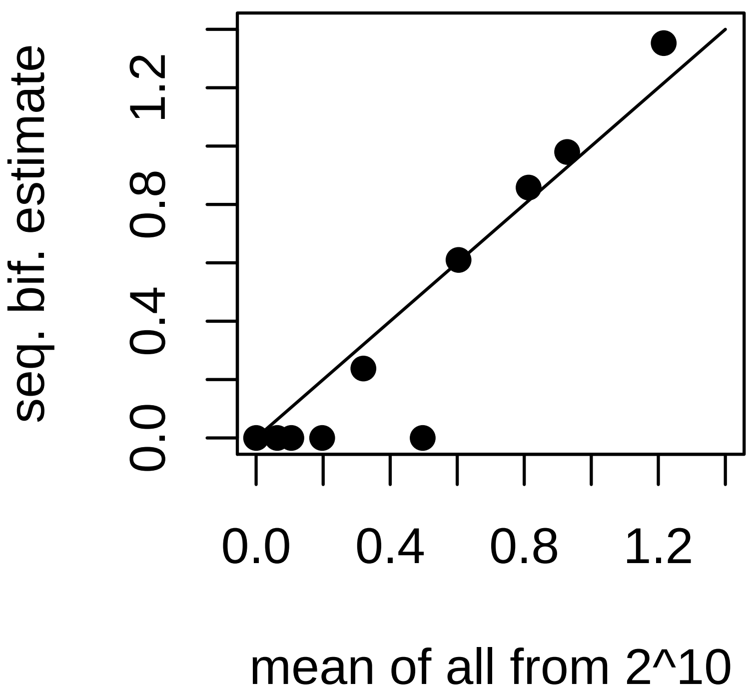
- Basic sequential bifurcation might lead to an accumulated experimental design as follows:

x_1	x_2	z_{11}	z_{12}	z_{13}	z_{14}	z_{21}	z_{22}	z_{23}	z_{24}	y
0	0	0	0	0	0	0	0	0	0	0.0000
1	1	1	1	1	1	1	1	1	1	1.0100
1	1	1	1	1	0	0	0	0	0	0.3384
1	1	1	1	1	1	1	0	0	0	0.3978
1	1	1	1	1	1	1	1	1	0	0.7649
1	1	1	1	1	1	1	1	0	0	0.5504
1	1	0	0	0	0	0	0	0	0	0.0000
1	1	1	1	0	0	0	0	0	0	0.3384
1	1	1	0	0	0	0	0	0	0	0.3384
1	1	1	1	1	1	0	0	0	0	0.3384
1	0	0	0	0	0	0	0	0	0	0.0000

(Note that a different experimental design would have been developed if the inputs had been listed in a different order ...)

- Data collected from this design lead to the following values of $\hat{H}$ (compared to the the "truth" from a full 2^{10} design):

x_1	x_2	z_{12}	z_{12}	z_{13}	z_{14}	z_{21}	z_{22}	z_{23}	z_{24}
0.000	0.000	1.353	<u>0.000</u>	<u>0.000</u>	0.000	0.238	0.610	0.858	0.980
0.000	0.105	1.216	0.497	0.197	0.063	0.320	0.604	0.813	0.928



- These errors are not realizations of random noise in the data (since there is none), but can be thought of as *bias* in estimators that have no variance.
- If $\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1$ is used as the basis of analysis, but the data are actually generated by a "true" model: $\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2$ then the least-squares estimate $\boldsymbol{\beta}_1$ is

$$\hat{\boldsymbol{\beta}}_1 = \boldsymbol{\beta}_1 + (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2\boldsymbol{\beta}_2 = \boldsymbol{\beta}_1 + \mathbf{A}\boldsymbol{\beta}_2$$

- The experimental design determines $\mathbf{X}_1$ and $\mathbf{X}_2$, and hence the alias matrix $\mathbf{A}$.
- Mitchell (1974) proposed using $||\mathbf{A}||$ as an index of design quality for estimating main-effects models when second-order terms are present in the data-generating process.
- We modify this idea slightly here to omit the first row of $\mathbf{A}$ since this corresponds to bias in the model intercept, which is of no real interest to us.

$\hat{H}$ and alias indices for designs of different sizes:

- SB = Sequential Bifurcation (as shown)
- FO SB = Foldover of Sequential Bifurcation design
- PB = minimal Plackett-Burman design
- FO PB = Foldover of Plackett-Burman design
- 2^{10-5}_{III} = Minimum Aberation Regular Fraction of Resolution III
- 2^{10-4}_{IV} = Minimum Aberation Regular Fraction of Resolution IV
- 2^{10-3}_{IV} = (larger) Minimum Aberation Regular Fraction of Resolution IV
- 2^{10} = Full Two-Level Factorial design

design	$x1$	x_2	z_{11}	z_{12}	z_{13}	z_{14}	z_{21}	z_{22}	z_{23}	z_{24}	N	$\|\|\mathbf{A}_2\|\|$	$\|\|\mathbf{A}_3\|\|$
SB	0.000	0.000	1.353	0.000	0.000	0.000	0.238	0.610	0.858	0.980	11	22.50	22.50
FO SB	0.000	0.000	1.197	0.500	0.500	0.500	0.119	0.305	0.429	0.490	20	0	22.50
PB	-0.106	0.104	1.239	0.416	-0.049	-0,035	0.598	0.245	0.394	1.206	12	10.00	5.83
FO PB	-0.047	0.066	1.098	0.445	0.079	-0.213	0.340	0.559	0.842	0.879	24	0	5.83
2^{10-6}_{III}	0.017	0.138	1.394	0.674	0.174	0.076	0.325	0.778	0.555	0.817	16	6.00	4.50
2^{10-5}_{IV}	0.014	0.120	1.217	0.496	0.198	0.055	0.173	0.659	0.825	0.938	32	0	2.50
2^{10-4}_{IV}	0.002	0.105	1.216	0.496	0.197	0.063	0.320	0.603	0.751	0.812	64	0	0.50
2^{10}	0.000	0.105	1.216	0.497	0.197	0.063	0.320	0.604	0.813	0.928	1024	0	0.00

(Underlines are errors of more than 0.10)

2. Variance-based sensitivity analysis

- Iooss and Ribatet (2009), Jacques et al. (2006) advocate a direct extension of the standard approach for scalar inputs, called the *microparameter method*.
- Quick reminder of the popular scalar-input approach

$\mathbf{A}$

x_1	x_2	x_3
.23	.46	.81
.71	.52	.33
...	...	...
.48	.21	.50

$\mathbf{B}$

x_1	x_2	x_3
.53	.27	.26
.21	.04	.37
...	...	...
.88	.49	.94

$\mathbf{A}_1$

x_1	x_2	x_3
.53	.46	.81
.21	.52	.33
...	...	...
.88	.21	.50

$\mathbf{A}_2$

x_1	x_2	x_3
.23	.27	.81
.71	.04	.33
...	...	...
.48	.49	.50

$\mathbf{A}_3$

x_1	x_2	x_3
.23	.46	.26
.71	.52	.37
...	...	...
.48	.21	.94

- Then averages of squared differences of outputs corresponding to paired rows form the basis of sensitivity index estimates:

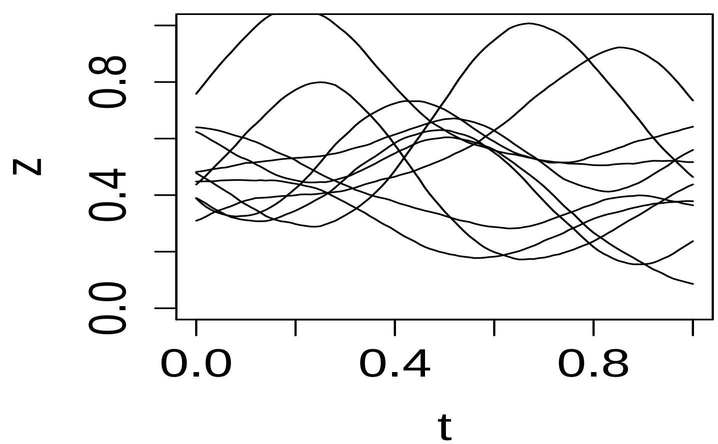
- $\mathbf{A}\&\mathbf{B} \rightarrow \widehat{\mathsf{Var}}(y)$, the unconditional variance
- $\mathbf{B}\&\mathbf{A}_1 \rightarrow E_{x_1}\widehat{\mathsf{Var}}_{x_2,x_3}[y|x_1]$
  - First-Order Sensitivity: $\hat{S}(x_1) = 1 - E_{x_1}\widehat{\mathsf{Var}}_{x_2,x_3}[y|x_1]/\widehat{\mathsf{Var}}(y)$
- $\mathbf{A}\&\mathbf{A}_1 \rightarrow \widehat{E_{x_2,x_2}\mathsf{V}}\mathsf{ar}_{x_1}[y|x_2,x_3]$
  - Total Sensitivity: $\hat{T}(x_1) = \widehat{E_{x_2,x_2}\mathsf{V}}\mathsf{ar}_{x_1}[y|x_2,x_3]/\mathsf{Var}(y)$
- and similarly for other inputs, using a different $\mathbf{A}_i$ but the same $\mathbf{A}$ and $\mathbf{B}$ in each case.

The same approach can be taken when any or all inputs are functional

- Functional inputs (or their vector approximations) are regarded as realizations of stochastic processes (or multivariate distributions)
- For example, a Gaussian process with

$$E(z(t)) = \frac{1}{2},\ \text{Var}(z(t)) = (\tfrac{1}{6})^2$$
$$\text{Corr}(z(t_1), z(t_2)) = e^{-\theta|t_1 - t_2|^{1.99}} \text{ with } \theta = 10:$$

- Realizations:



- In the examples that follow, I use this process model for both z_1 and z_2, and represent them as 101-element vectors $\mathbf{z}_1$ and $\mathbf{z}_2$.

- With:
  - x_1 and $x_2 \sim U[0,1]$, and each of $\mathbf{z}_1$ and $\mathbf{z}_2$ as described above
  - 6 input arrays, 100,000 rows per array (600,000 function evaluations)

  results for the example model are:

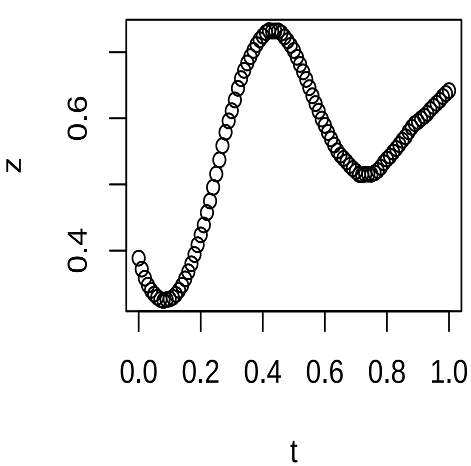
	x_1	x_2	$\mathbf{z}_1$	$\mathbf{z}_2$
$\hat{S}$	0.0092	0.1065	0.2565	0.5937
$\hat{T}$	0.0000	0.1277	0.2896	0.6382

- This is useful, but it offers little insight into *how* z_1 and z_2 influence y.
- Proposal: "Factor" the functional input into one or a few scalar-valued summaries and an *independent* functional residual (of hopefully little importance).

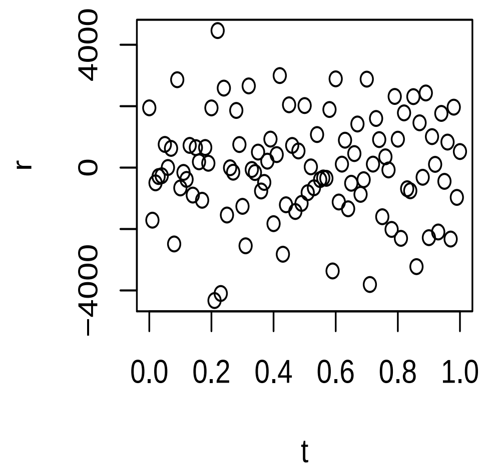
Special case: Gaussian processes: $\mathbf{z} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

- Scalar-valued summaries: $\mathbf{s} = \mathbf{C}'\mathbf{z}$
  - e.g. coefficients of a low-order least-squares polynomial approximation to $\mathbf{z}$
- A "residual": $\mathbf{r} = (\mathbf{I} - \mathbf{C}(\mathbf{C}'\mathbf{C})^{-1}\mathbf{C}')\boldsymbol{\Sigma}^{-1}\mathbf{z}$
- Both $\mathbf{s}$ and $\mathbf{r}$ are multivariate normal, and independent, and $\mathbf{z}$ can be recovered from $\mathbf{s}$ and $\mathbf{r}$

- Example:
  - Univariate $s = \bar{z}$
  - $\mathbf{z}$ generated as before



$\rightarrow s = 0.5522$ and



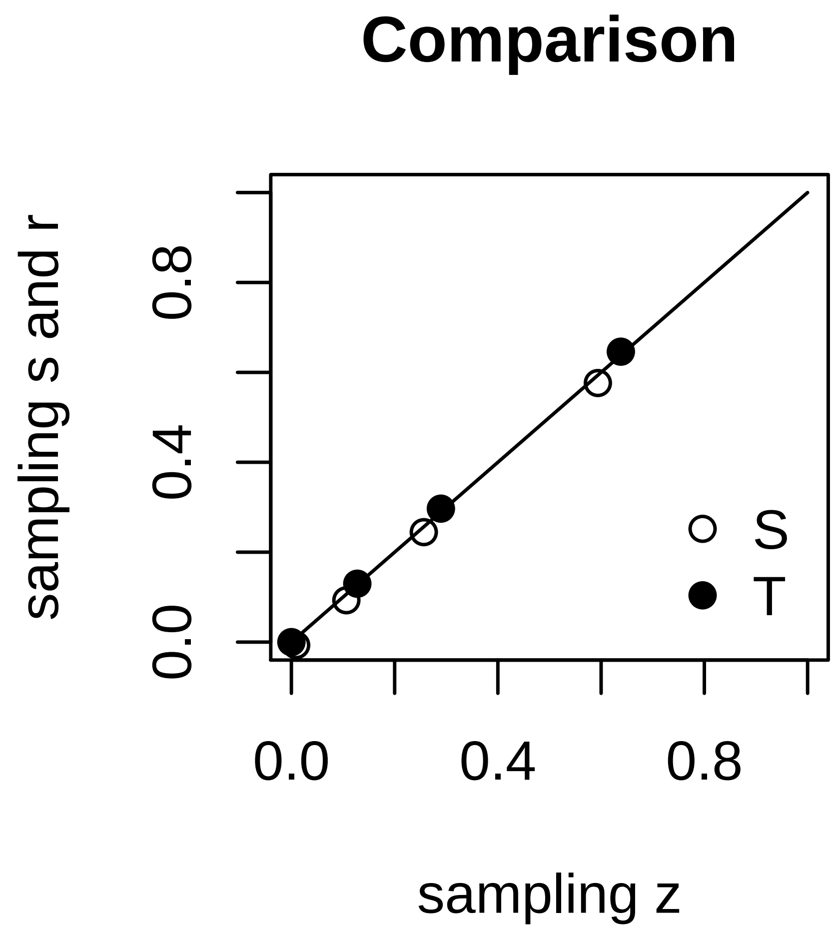
- Hence, our example can be viewed as:

$$y = f(x_1, x_2, s_1, \mathbf{r}_1, s_2, \mathbf{r}_2)$$

- Use $s_1 = \text{ave}(\mathbf{z}_1)$ and $s_2 = \text{ave}(\mathbf{z}_2)$
- Now 8 input arrays, 100,000 rows per array (800,000 function evaluations)

	x_1	x_2	s_1	$\mathbf{r}_1$	s_2	$\mathbf{r}_2$
$\hat{S}$	-0.0068	0.0931	0.1350	0.1096	0.5534	0.0230
$\hat{T}$	0	0.1300	0.1656	0.1313	0.5989	0.0472

- s_2 is important, while $\mathbf{r}_2$ has little impact
- s_1 is more important than $\mathbf{r}_1$, which is comparable to x_2



- Comparing $S(\mathbf{z}_1)$ to $S(s_1) + S(\mathbf{r}_1)$, et cetera

3. Statistical surrogate-based sensitivity analysis

- Ioose and Ribatet (2009) also discussed using a *joint modeling approach* to sensitivity analysis with functional inputs, based on fitting two models to output data.
- The (conditional) mean and variance of the output are modeled as functions of scalar-valued inputs only, i.e.
  - for inputs $= (x_1, ..., x_m, z_1(t), ..., z_n(t))$,
  - estimate models for $E(y|x_1, ..., x_m)$ and $\text{Var}(y|x_1, ..., x_m)$.
- So, for example, $E_{x's}\text{Var}_{z's}(y|x's)$ can be estimated by integrating the dispersion model w.r.t. the distribution of $x's$, et cetera.
- Authors used GLM and GAM in their examples.
- In this form, the approach does not separate the variability associated with different functional inputs.

Here I'll try something related, and refer to it as "semi-modeling":

- Draw F realizations of each input function,

$$z_1^{i_1}(t)...z_n^{i_n}(t), i_1...i_n = 1...F.$$

- Model y only for the selected function values, i.e.

$$y = f(x_1...x_m, i_1...i_n)$$

  where $i_1...i_n$ are categorical variables, each with values $1...F$, indexing associated input function values.

- Given training data, fit a single predictive model of the output:

$$\hat{y} = \hat{f}(x_1...x_m, i_1...i_n)$$

- Then, for example, using a random sample of size R (much larger than F) of each of $x_1 ... x_m, i_1 ... i_n$ and $x'_1 ... x'_m, i'_1 ... i'_n$,
  - $\widehat{\text{Var}}(y) = \frac{1}{2R} \sum_{r=1}^{R} (\hat{y}(x_1^r ... x_m^r, i_1^r ... i_n^r) - \hat{y}(x_1'^r ... x_m'^r, i_1'^r ... i_n'^r))^2$
  - $\hat{T}(x_1) = \frac{1}{2R} \sum_{r=1}^{R} (\hat{y}(x_1^r ... x_m^r, i_1^r ... i_n^r) - \hat{y}(x_1'^r ... x_m^r, i_1^r ... i_n^r))^2 / \widehat{\text{Var}}(y)$
  - $\hat{S}(x_1) =$

  $[\widehat{\text{Var}}(y) - \frac{1}{2R} \sum_{r=1}^{R} (\hat{y}(x_1^r ... x_m^r, i_1^r ... i_n^r) - \hat{y}(x_1^r ... x_m'^r, i_1'^r ... i_n'^r))^2] / \widehat{\text{Var}}(y)$

  and similarly for other inputs, both scalar and functional.

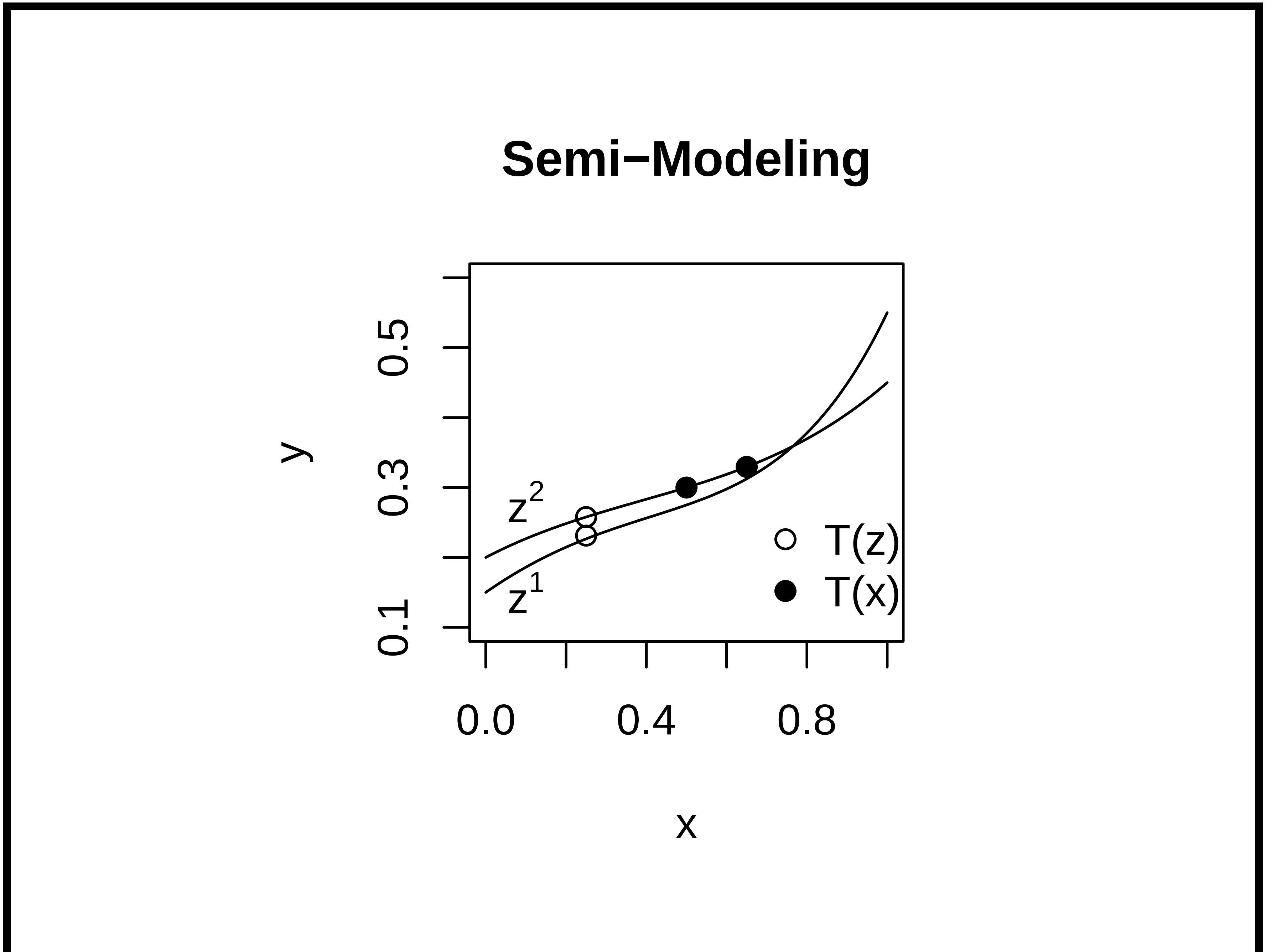
- Here I model y with a stationary Gaussian stochastic process model, where for

$$y = f(x_1 ... x_m, i_1, ... i_n), y' = f(x'_1 ... x'_m, i'_1, ... i'_n),$$

$$E(y) = E(y') = \mu, \text{Var}(y) = \text{Var}(y') = \sigma^2, \text{Cov}(y, y') = \sigma^2 e^{-\text{dist}},$$

$$\text{dist} = \sum_{j=1}^{m} \theta_j (x_j - x'_j)^2 + \sum_{j=1}^{n} \phi_j I(i_j \neq i'_j),$$

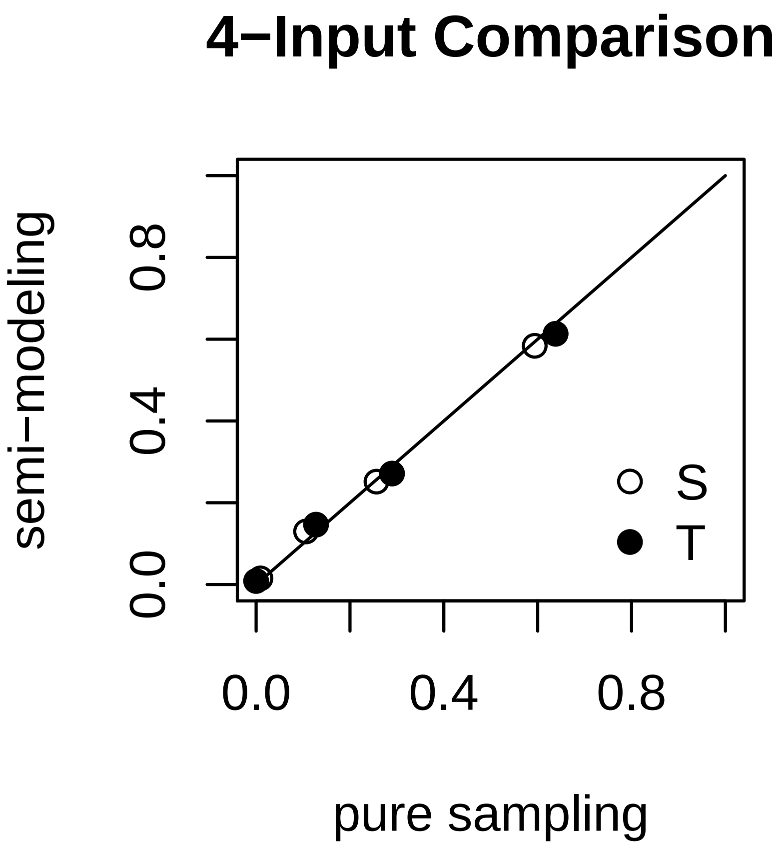
  fitting parameters via maximum likelihood.


Semi−Modeling
y
0.1
0.3
0.5
x
0.0
0.4
0.8
z2
z1
T(z)
T(x)

Results for the example model:

- $F = 50$ realizations of each of x_1, x_2, $\mathbf{z}_1$ and $\mathbf{z}_2$, distributed as before.
- Design constructed by repeating each input value 5 times, and forming the $N =$ 250-run experimental design via the maximin distance criterion.
- Result provides a predictor of y for any combination of x_1, x_2 and any of the 50 drawn realizations for each of $\mathbf{z}_1$ and $\mathbf{z}_2$.
- Results ($R = 10,000$):

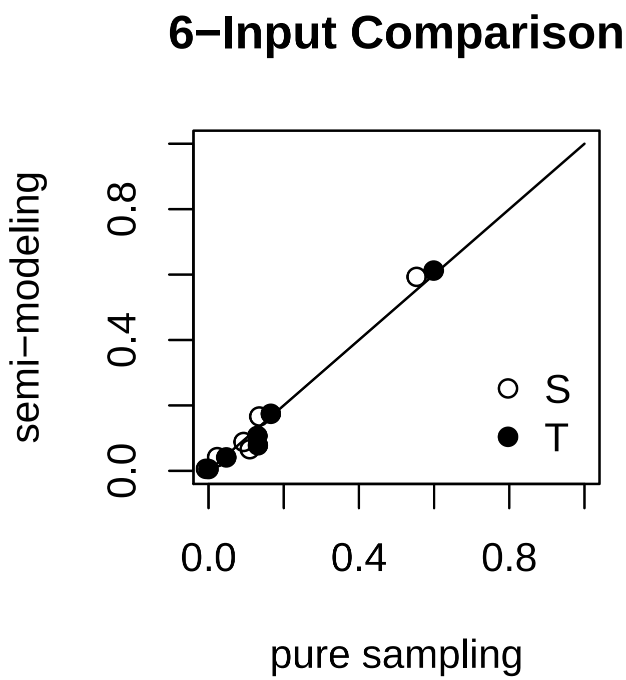
	x_1	x_2	$\mathbf{z}_1$	$\mathbf{z}_2$
$\hat{S}$	0.0151	0.1299	0.2517	0.5839
$\hat{T}$	0.0088	0.1465	0.2715	0.6133

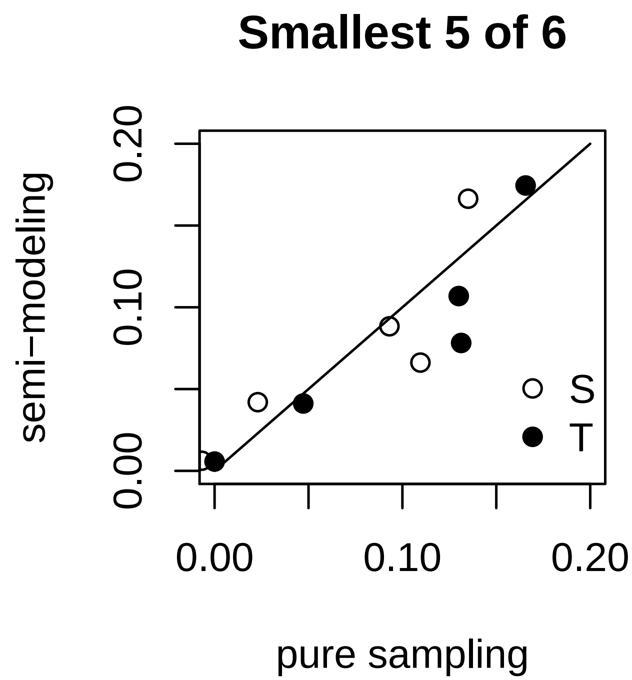


- Results are consistent with those from the pure sampling-based approach, but requiring far fewer function evaluations.

- Replacing $\mathbf{z}_1$ with s_1 and $\mathbf{r}_1$, and $\mathbf{z}_2$ with s_2 and $\mathbf{r}_2$:

	x_1	x_2	s_1	$\mathbf{r}_1$	s_2	$\mathbf{r}_2$
$\hat{S}$	0.0062	0.0884	0.1664	0.0662	0.5931	0.0420
$\hat{T}$	0.0057	0.1069	0.1745	0.0782	0.6124	0.0412





Concluding thoughts:

- Can more bias-resistant alternatives to Sequential Bifurcation be developed for the piecewise constant inputs case (that doesn't require too many runs)?
- Traditional variance-based sensitivity analysis may be most effective if functional inputs can be decomposed into independent (1.) important low-dimensional, and (2.) less important higher-dimensional components.
- Meta-models that are accurate approximations for a moderate sample of functional inputs may improve the efficiency of variance-based sensitivity analysis.

References

- Bettonvil, B. (1995). "Factor Screening by Sequential Bifurcation," *Comm. Statist.: Simulation Comput.* **24**(1) 165-185.
- Fruth, J., O. Roustant, and S. Kuhnt (2014). "Sequential Designs for Sensitivity Analysis of Functional Inputs in Computer Experiments," *Reliab. Eng. Syst. Safety*, to appear.
- Ioss and Ribatet (2009). "Global Sensitivity Analysis of Computer Models with Functional Inputs," *Reliab. Eng. Syst. Safety* **94** 1194-1204.
- Jacques, J., C. Lavergne, and N. Devictor (2006). "Sensitivity Analysis in Presence of Model Uncertainty and Correlated Inputs," *Reliab. Eng. Syst. Safety* **91** 1126-1126.
- Mitchell, T.J. (1974). "Computer Construction of "D-Optimal" First-Order Designs," *Technometrics* **16**(2) 211-220.
- Oakley, J.E., and A. O'Hagan (2004). "Probabilistic Sensitivity Analysis of Complex Models: A Bayesian Approach," *J.R. Statist. Soc. B* **66**(3) 751-769.
- Saltelli, A., K. Chan, and E.M. Scott, eds. (2000). *Sensitivity Analysis*, Wiley, West Sussex, England.
- Welch, W.J., R.J. Buck, J. Sacks, H.P. Wynn, T.J. Mitchell, and M.D. Morris (1992). "Screening, Predicting, and Computer Experiments," *Technometrics* **34**(1) 15-25.