Techniques for Sensitivity Analyses on Non-Monotonic Functions

Osvaldo Pensado and Budhi Sagar

Center for Nuclear Waste Regulatory Analyses (CNWRA)
Southwest Research Institute®, San Antonio, Texas, USA

8 International Conference on Probabilistic Safety Assessment and Management (PSAM8)

May 14–19, 2006, New Orleans, Louisiana

Outline

- Background
- Techniques
- Results
- Conclusions

Background

- The performance of sensitivity techniques for models with non-monotonic dependence on parameters needs to be investigated
- In complex environmental stochastic models, competing effects frequently arise, leading to non-linear and non-monotonic dependencies
  - Temperature, chemistry, hydrodynamics, physical thresholds
- Purpose: evaluate sensitivity techniques previously developed at CNWRA
  - Partitioning method
  - Parameter tree method
  - Mean-based and standard deviation-based index methods

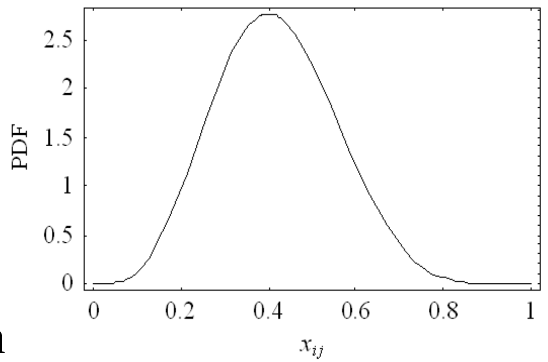
Nomenclature

$y_i = f(x_{i1}, x_{i2}, x_{i3}, \cdots, x_{iN})$ Performance metric, *i*th realization

N = number of parameters

$\mathbf{x}_j = \{x_{ij},\ i = 1, \cdots, R\}$ Sampled vector of *j*th parameter

R = number of samples or realizations



$\mathbf{r}_i = \{x_{i1}, x_{i2}, x_{i3}, \cdots, x_{iN}; y_i\}$ Realization vector, *i*th realization

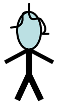
$\mathbf{y} = \{y_i,\ i = 1, \cdots, R\}$ Performance metric vector

CNWRA
A center of excellence in earth sciences and engineering™

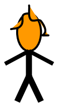
Nomenclature

Analogy

$\mathbf{r}_1 = \{x_{11}, x_{12}, x_{13}, \cdots, x_{1N}; y_1\}$


Peter = (blue skin, short, …; manager)

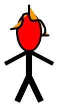
$\mathbf{r}_2$


Jane = (yellow skin, tall, …; engineer)

$\vdots$

$\vdots$

$\mathbf{r}_R$


Lauren

Outstanding set (e.g., highest y_i values) = {3, 105, 132, …}

Outstanding set (e.g., managers) = {John, Beth, Vijay, …}

What makes a realization r_i outstanding?

What makes an individual outstanding?

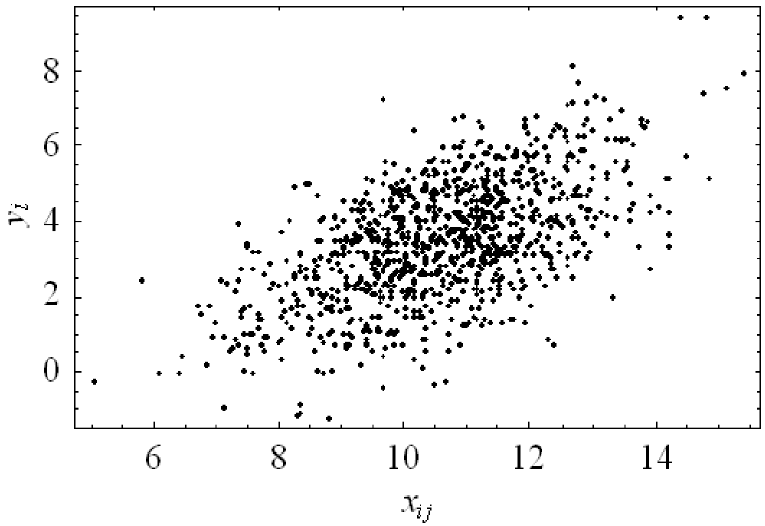
What are the most important parameters?

What are the most important features?

n = size of the outstanding set, $n < N$

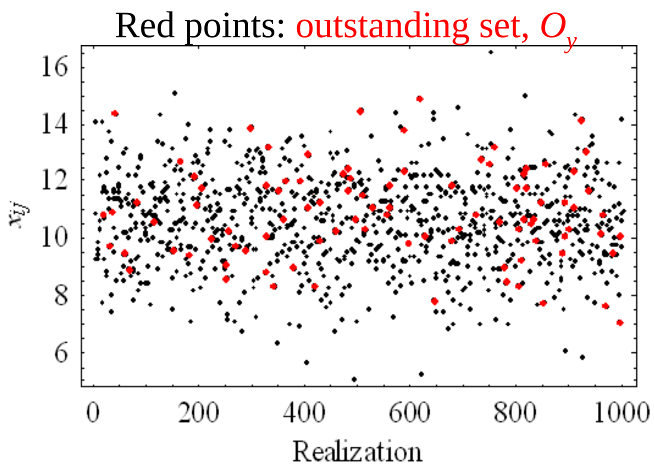
Sensitivity Indices

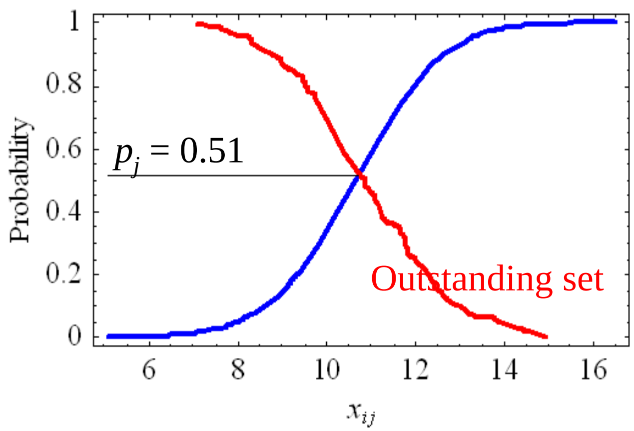
- A sensitivity index is used as a measure of the change in the performance metric, y_i, due to changes in the input parameter x_{ij}



- The most important parameters are those with the highest magnitude of the sensitivity index

Partitioning Method





p: probability intercept between the cumulative distribution function (CDF) of the full set and complementary CDF of the outstanding set.

Sensitivity index:

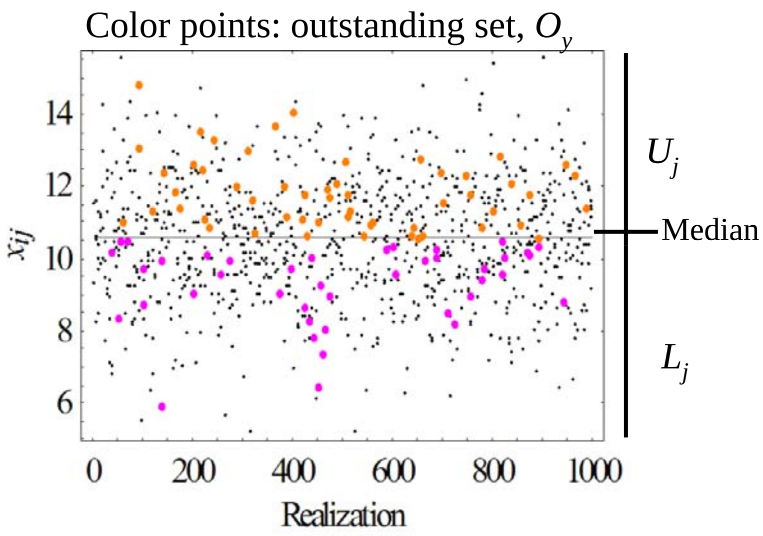
$$I_{pm}^{j} = p_j - 0.5$$

Distribution function for *p* to test for lack of correlation: beta distribution in [0, 1] with shape parameters

$$\alpha = \beta = \frac{0.125 - 0.5 s^2}{s^2} \qquad s = \frac{0.246}{\sqrt{n}}$$

If $|p_j - 0.5| > p_{0.975} - 0.5$, the hypothesis of lack of correlation is rejected.

Parameter Tree Method



Sensitivity index:

$$I_{pt}^{j} = \frac{2}{R}\left(\text{Elements in } U_j \cap O_y - \text{Elements in } L_j \cap O_y\right)$$

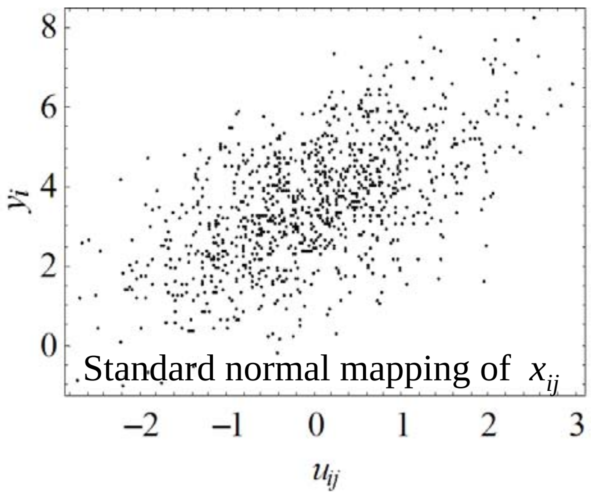
Distribution function for I_{pt}^{j} to test for lack of correlation: beta distribution in [-1, 1] with shape parameters

$$\alpha = \beta = \frac{0.125 - 0.5 s^2}{s^2}$$

$$s = \frac{0.5}{\sqrt{R}}\left[1 - 0.675\left(\frac{n}{R} - 0.5\right)^2 - 13.3\left(\frac{n}{R} - 0.5\right)^4\right]$$

If $|I_{pt}^{j}| > I_{0.975}$, the hypothesis of lack of correlation is rejected.

Mean and Standard Deviation (SD) Based Index Methods



Mean-based index: $I_{mb}^{j} = \dfrac{\mathbf{u}_j \cdot \mathbf{y}}{\|\mathbf{y}\|}$

Standard deviation-based index: $I_{sb}^{j} = \dfrac{(\mathbf{u}_j^{2} - \mathbf{1}) \cdot \mathbf{y}}{\sqrt{2}\|\mathbf{y}\|}$

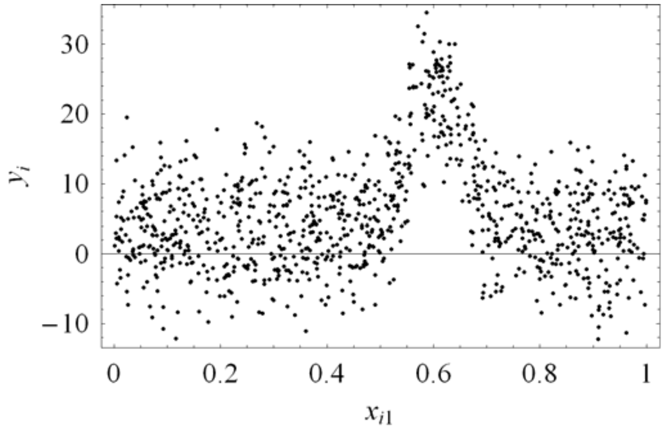
Distribution function for I to test for lack of correlation: standard normal distribution

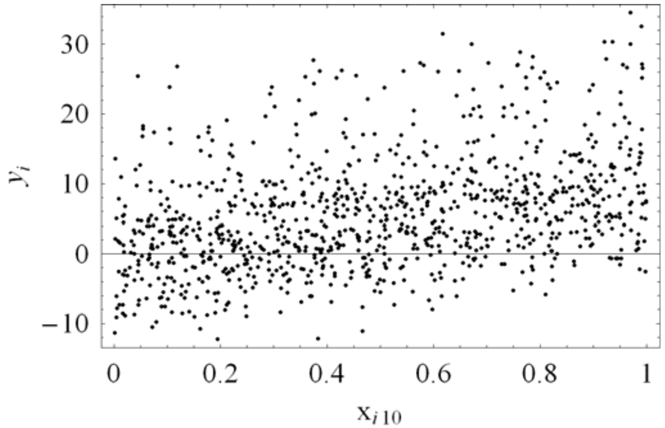
$$I_{f\ \text{quantile}} = \sqrt{2}\,\mathrm{erf}^{-1}(2f-1)$$

If $|I^{j}| > I_{0.975}$, the hypothesis of lack of correlation is rejected.

Example 1

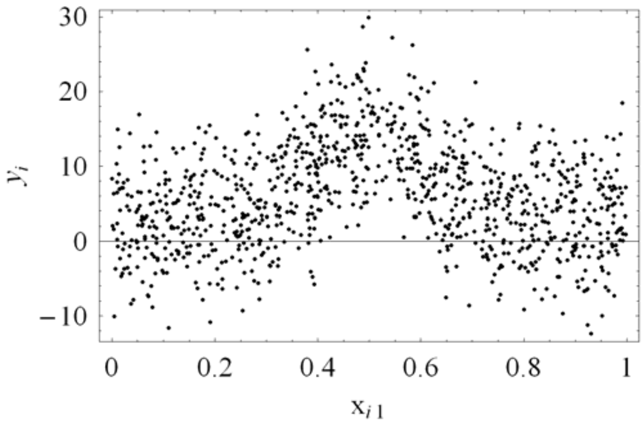
$$y_i = 20e^{-200(x_{i1}-0.6)^2} + \sum_{j=2}^{10} j(-1)^j x_{ij}$$





Method	**Parameter Ranking**	**Index/$I_{97.5}$ (First 5 Values)**
Linear regression	10, 9, 8, 7, 6, 5, 4, 3, 2, **1**	5.83, -5.54, 4.85, 4.34, -3.96
Parameter tree	**1**, 6, 8, 9, 7, 10, 4, 5, 3, 2	5.85, 4.68, 2.11, 1.52, -1.52
Partitioning	**1**, 6, 9, 10, 8, 5, 7, 3, 4, 2	2.10, 2.02, -1.78, 0.96, 0.93
Mean-based	10, 9, 8, 6, 7, 5, 4, 3, 2, 1	4.63, -4.50, 3.62, 3.56, -3.14
SD-based	**1**, 3, 9, 4, 2, 8, 5, 6, 7, 10	-5.80, 1.61, -1.46, 1.23, 0.94

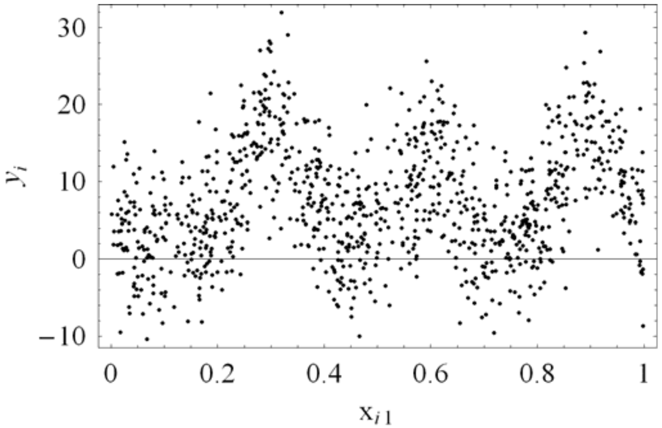
Example 2



$$y_i = 12e^{-50(x_{i1}-0.5)^2} + \sum_{j=2}^{10} j(-1)^j x_{ij}$$

Method	Parameter Ranking	Index/$I_{97.5}$ (First 5 Values)
Linear regression	10, 9, 8, 7, 6, 5, 4, 3, 2, **1**	6.15, -5.87, 5.85, -4.84, 4.03
Parameter tree	10, 9, 8, 7, 6, 5, 4, 2, 3, **1**	3.86, -2.57, 2.11, -1.87, 1.64
Partitioning	10, 9, 8, 7, 6, 5, 4, 2, 3, **1**	3.47, -2.42, 2.06, -1.75, 1.60
Mean-based	10, 9, 8, 7, 6, 5, 4, 2, 3, **1**	4.87, 4.61, -4.38, -3.26, 2.67
SD-based	**1**, 2, 6, 9, 5, 3, 4, 10, 8, 7	-3.55, 0.67, -0.40, 0.25, -0.18

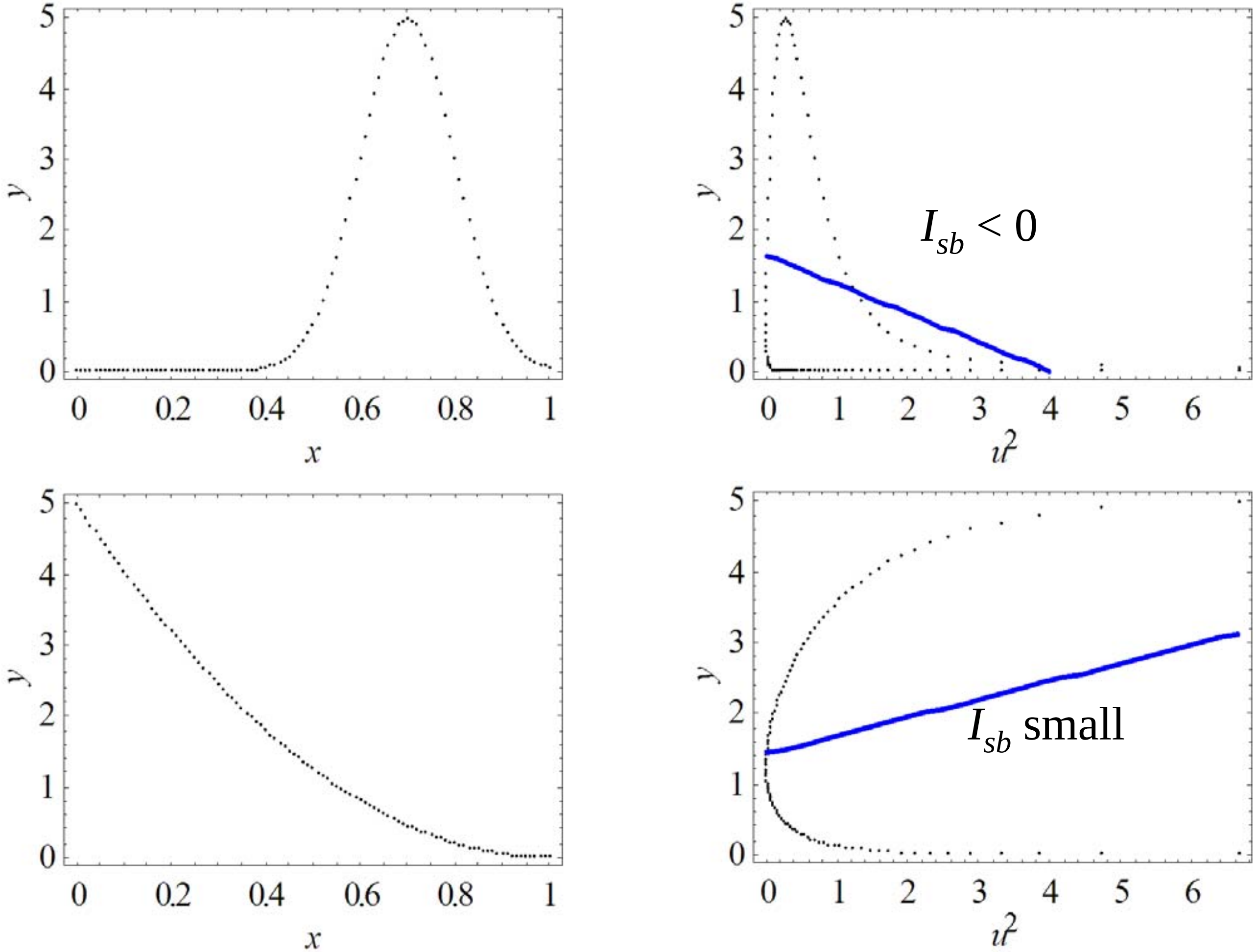
Example 3



$$y_i = 15e^{-200(x_{i1}-0.3)^2} + 10e^{-200(x_{i1}-0.6)^2} + 15e^{-200(x_{i1}-0.9)^2} + \sum_{j=2}^{10} j(-1)^j x_{ij}$$

Method	**Parameter ranking**	**Index/$I_{97.5}$ (first 5 values)**
Linear regression	9, 10, 8, 7, 6, 5, **1**, 4, 3, 2	-6.09, 6.00, 4.82, -4.33, 3.26
Parameter tree	10, 9, 5, 8, 4, 7, 6, 2, 3, **1**	4.33, -2.69, -2.22, 1.76, 1.64
Partitioning	10, 9, 5, 8, 7, 6, 4, 2, 3, **1**	3.27, -2.48, -1.65, 1.60, -1.52
Mean-based	10, 9, 8, 7, 1, 5, 6, 4, 3, 2	4.24, -4.22, 3.18, -2.92, 2.43
SD-based	**1**, 10, 5, 2, 9, 3, 6, 7, 4, 8	-0.90, 0.56, 0.56, -0.50, 0.43

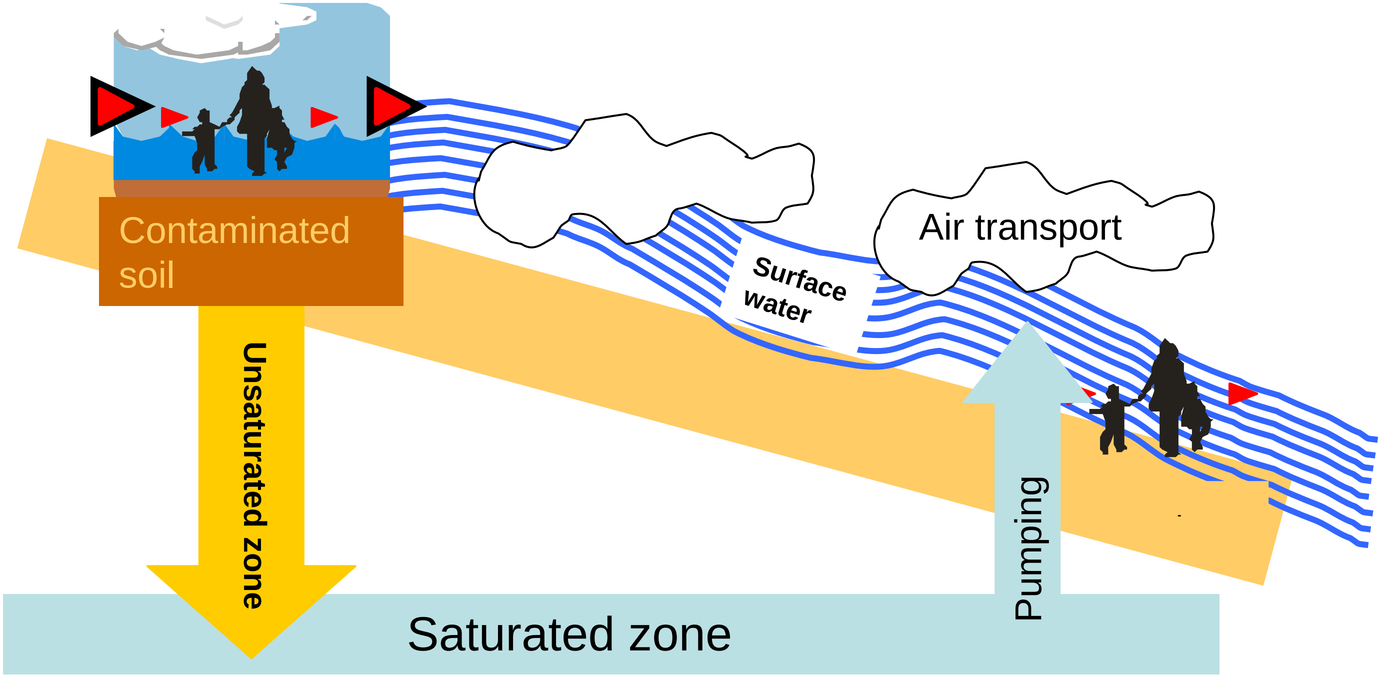
Why Did the SD-Based Index Succeed?



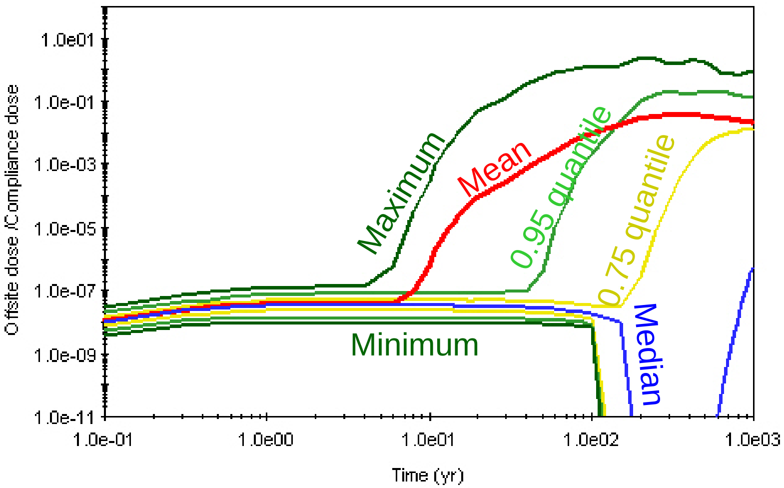
If y is maximized at mid-range in x, I_{sb} is negative.

Example 4

Contaminant transport model of a hypothetical site to estimate far-field doses



Example 4



Example performance metric: maximum dose in 1,000 years.

The models included a relevant non-monotonic dependence between parameter 42 and the maximum dose.

Method	Ranking (First 5 Parameters)	Index/$I_{97.5}$ (First 5 Values)
Linear regression	7, 41, 12, 9, 18	-1.99, -1.97, 1.80, -1.26, -1.13
Parameter tree	41, 7, 12, 32, 28	-4.28, -2.85, 1.21, -1.10, -0.99
Partitioning	41, 7, 32, 12, 14	-3.16, -2.16, -1.05, 1.04, 0.95
Mean-based	41, 7, 12, 18, 9	-3.40, -2.50, 2.05, -1.32, -1.13
SD-based	42, 27, 5, 41, 10	-2.16, -1.10, 0.98, 0.88, 0.84
Linear regression ($\mathbf{u}_j^2$, $\mathbf{y}$)	42, 10, 5, 27, 39	-2.18, 1.21, 1.00, -0.85, -0.74

Only the SD-based index and the linear regression variant identified parameter 42 as significant.

CNWRA
A center of excellence
in earth sciences
and engineering™

Conclusions

- The partitioning and parameter tree methods provide limited capability to identify non-monotonic trends.

- Mapping input parameter vectors into the standard normal distribution to the square can isolate non-monotonic dependencies.

- As a first step, standard techniques (e.g., linear regression) can identify main monotonic dependencies. As a second step, linear regression between the square of the standard normal distribution parameter mapping and the performance metric can identify non-monotonic dependencies.

Disclaimer

- This work was performed by the Center for Nuclear Waste Regulatory Analyses (CNWRA) for the U.S. Nuclear Regulatory Commission (NRC) under Contract No. NRC–02–02–012 on behalf of the NRC Office of Nuclear Material Safety and Safeguards, Division of High–Level Waste Repository Safety.
- This work is an independent product of CNWRA and does not necessarily reflect the view or the regulatory position of the NRC.