U.S. Department of Energy
Office of Civilian Radioactive Waste Management

Uncertainty Importance Analysis for TSPA-SR

Presented to:
DOE/NRC Technical Exchange on TSPA-SR

Presented by:
Srikanta Mishra
M&O Performance Assessment

June 7, 2000

YUCCA MOUNTAIN PROJECT

Overview of Presentation

- Introduction
- Uncertainty Importance Analysis
- Overview of Techniques for Uncertainty Importance Analysis
  - Regression Analysis
  - Scatter Plot Analysis
  - Classification Tree Analysis
- Interpretation of Results
- Documentation of Results
- Summary

Introduction

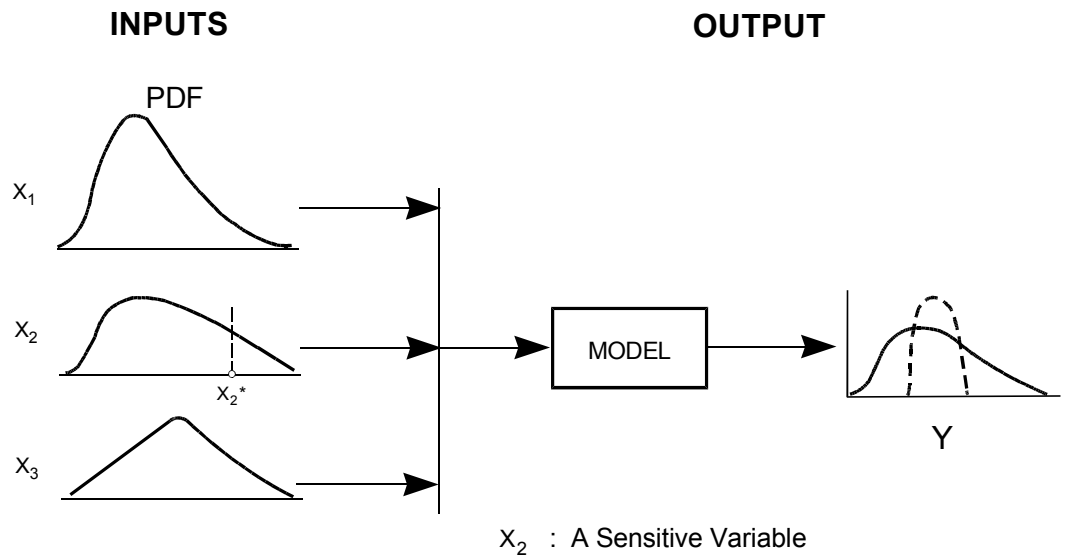
- **Inputs of TSPA model (i.e., scenarios, mathematical/conceptual models and parameters) are uncertain and/or spatially variable**
- **Uncertain inputs yield uncertain outputs (predicted distributions of dose)**
- **Output distributions generated by Monte Carlo simulations**

Uncertainty Importance Analysis

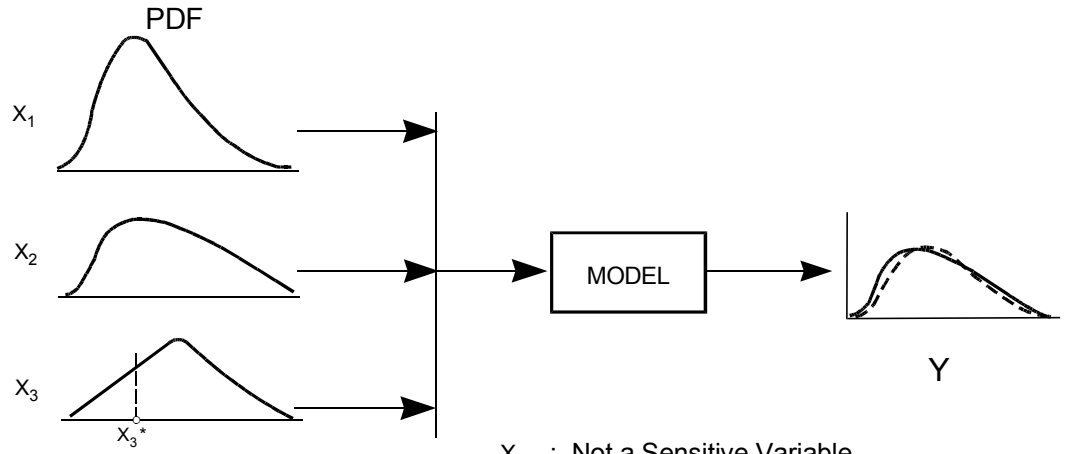
- **Uncertainty Analysis - quantification of spread and variance of dose**
- **Sensitivity Analysis - Identification of the input variables whose 'spread' is a dominant cause of output uncertainty**
- **"Uncertainty Importance Analysis" describes Sensitivity Analysis in the context of Monte Carlo simulations**
- **Application**

  If the spread in the most sensitive variables is reduced, the spread in the dose would be reduced (a desirable objective)

Uncertainty Importance
A Graphical Interpretation



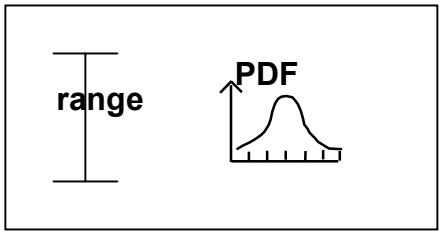
X_2 : A Sensitive Variable



X_3 : Not a Sensitive Variable

Uncertainty Importance

- **Component Factors**
  - **Uncertainty in the input variable**
  - **Sensitivity of the input variable in the process model**



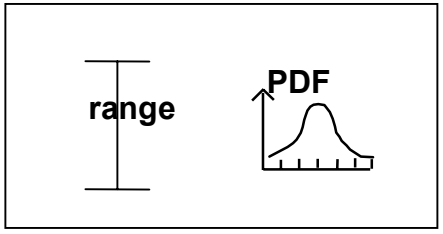
Uncertainty in input

X

$$S = \frac{\text{Change in output}}{\text{Change in input}}$$

Sensitivity Coefficient

=



Uncertainty in Output

Overview of Techniques

- **Regression analysis**
  - **Quantitative input-output model built via rank regression to determine most important contributors to output variance (spread)**
- **Scatter plot analysis**
  - **Visual measure of association between model output and uncertain inputs**
- **Classification tree analysis**
  - **Binary decision tree built from input-output data to identify which variables control extreme realizations**

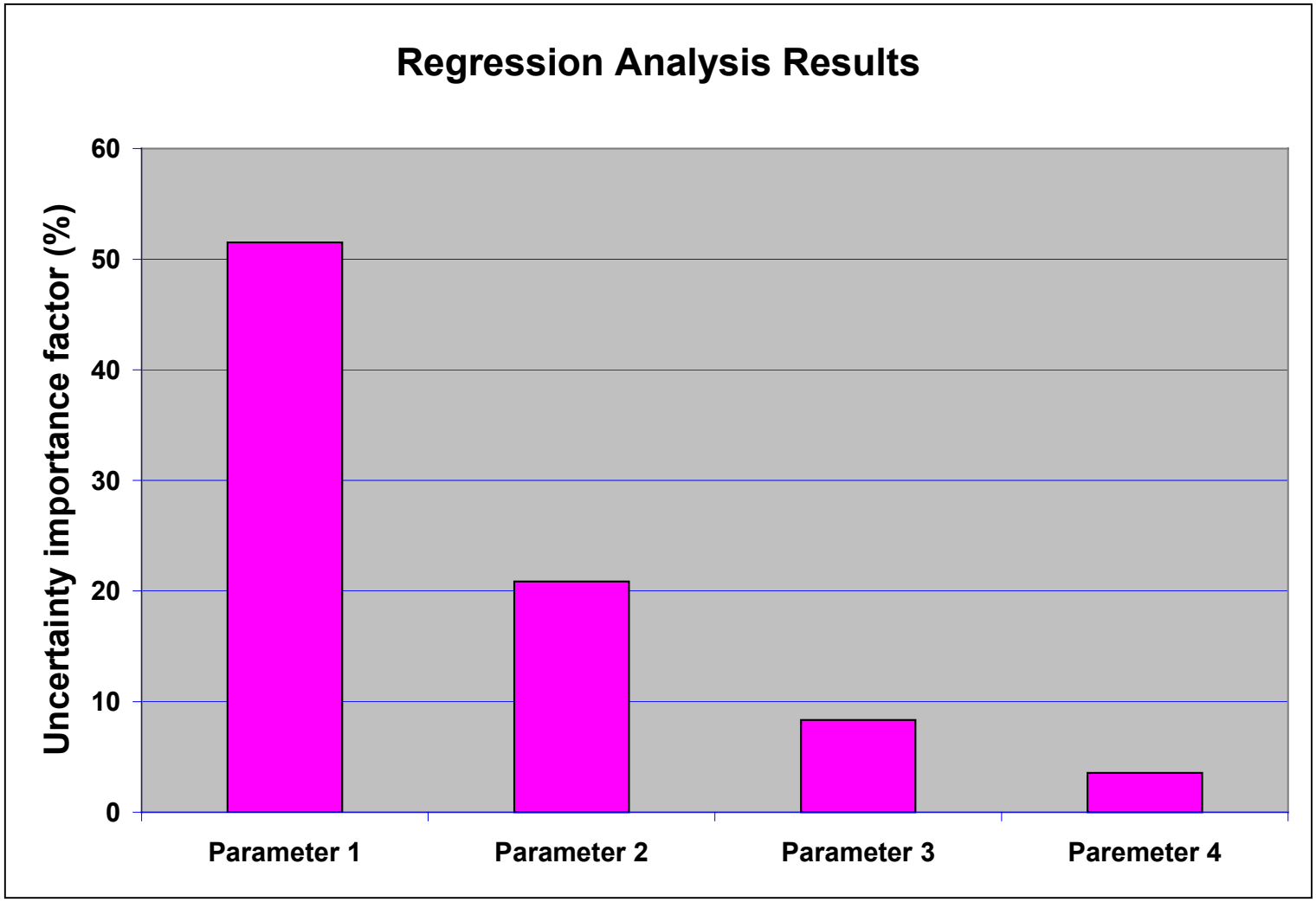
Regression Analysis - Basics

- Quantification of strength of input-output relationship in TSPA model
- Based on step-wise linear rank regression model fitted between GOLDSIM output and all randomly sampled input variables
- Parameters ranked on the basis of how their exclusion would degrade explanatory power of regression model (uncertainty importance factor)
- Importance ranking reflection of (a) input uncertainty and (b) sensitivity of output to input

Regression Analysis - Steps

- **Build a sequence of multivariate linear rank regression models between output and inputs**
- **At each step, admit the variable which accounts for the largest amount of unexplained variance until no more regression coefficients pass statistical significance tests**
- **Importance ranking metrics**
  - **Partial correlation => correlation between output and input after removing linear influence of all other inputs**
  - **R^2-loss => loss in explanatory power of current model if a variable is excluded from regression model**

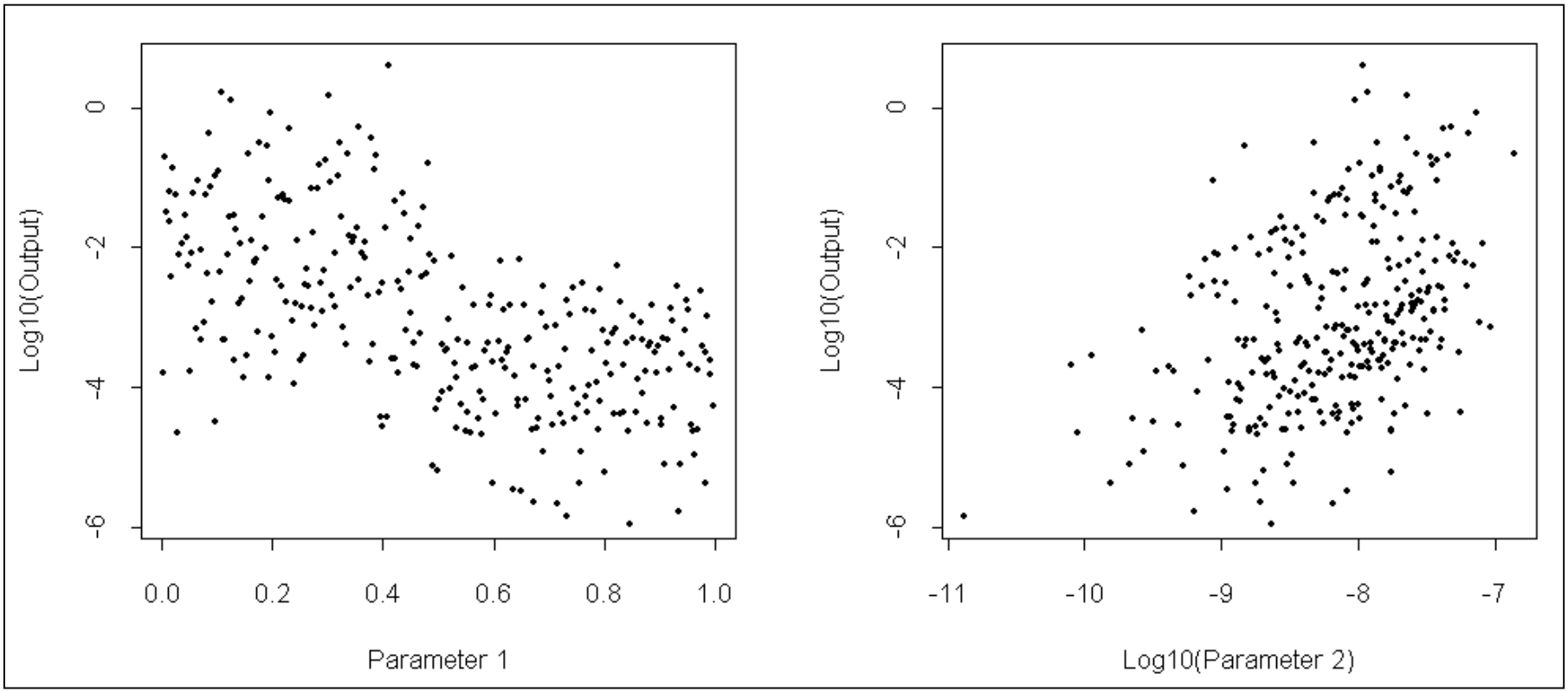
Regression Analysis - Example



Scatter Plot - Basics

- Simple graphical tool for determining nature of relationship between output and inputs
- Trends in plot indicate whether relationship is random, positive or negative
- Useful supplement for visual examination of regression analysis results
- Can also help reveal threshold phenomena and non-linearities in input-output relationship

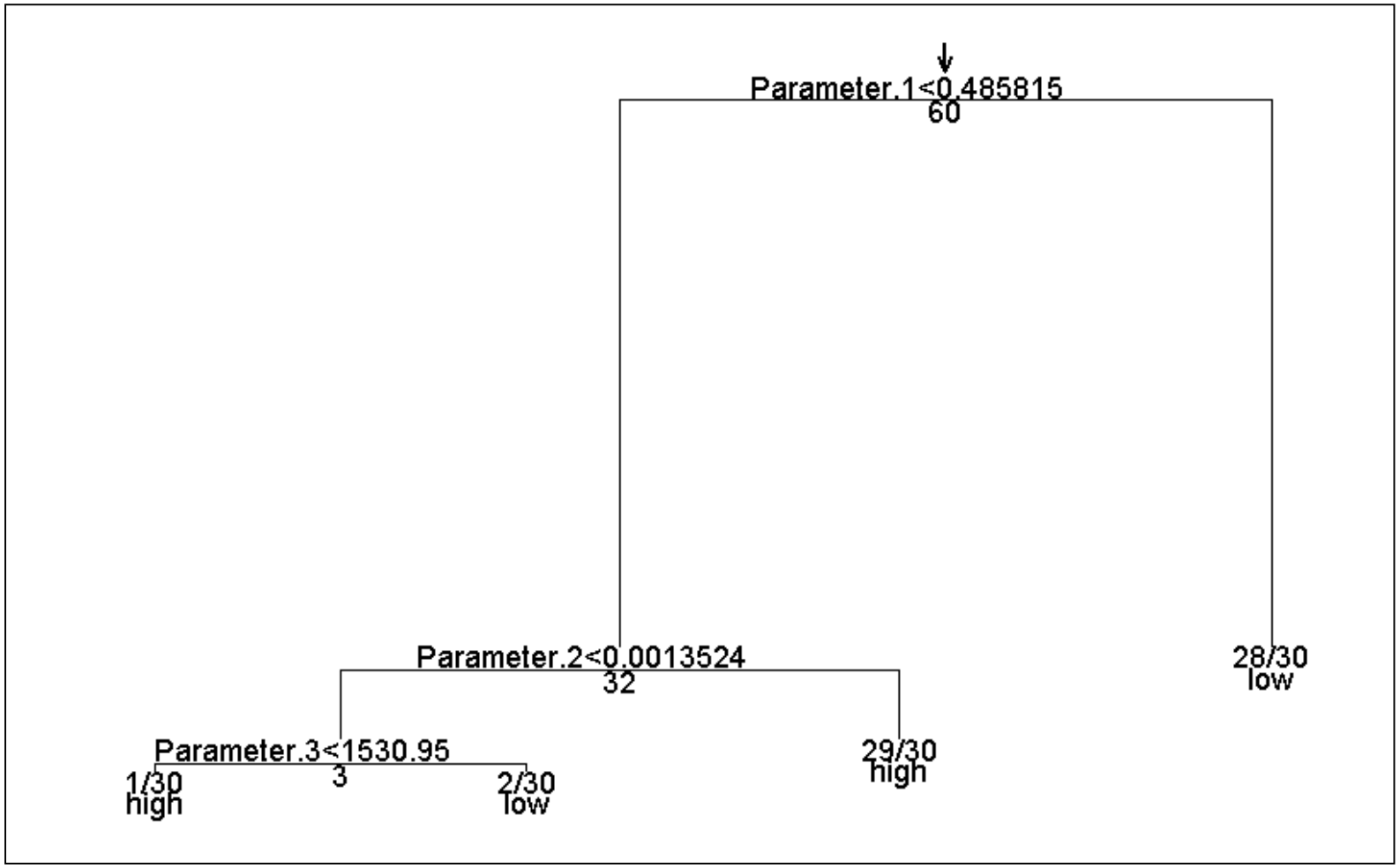
Scatter Plot Analysis - Example



Classification Tree - Basics

- Binary decision tree built to identify variables controlling extreme realizations
- CART algorithm recursively finds best split from among all variables based on purity (measure of similarity of outcomes within subgroup) until all subgroups contain only one outcome
- Application to TSPA data involves converting numerical results to categorical values
  - Doses below 10th percentile => “low”
  - Doses above 90th percentile => “high”

Classification Tree Analysis - Example



Interpretation of Results

- Results can be used to screen candidate variables for "one-off" sensitivity analyses (using 5^{th}-95^{th} percentile values)
- System-level importance ranking may differ from subsystem-level ranking
- Importance ranking conditional to range of uncertainty used in existing model
- Results should be used in conjunction with sensitivity and barrier importance analyses to guide future data collection / modeling

Documentation of Uncertainty Importance Analysis

- **Overview of Uncertainty Importance Analysis in TSPA-SR**
  - **Section 2.2.4 Uncertainty Analysis**
- **Results of Uncertainty Importance Analysis**
  - **Section 5 Sensitivity Analysis**
  - **Section 5.1 Results of Uncertainty Importance Analysis**

Summary

- **Multiple path approach to evaluating uncertainty importance for TSPA-SR**
- **Results provide valuable support to interpretation of TSPA-SR model**