

Seminar Series in Uncertainty and Estimation: Understanding, Communication, and Decision-Based Data Specification

PJ Chamberlain Ph.D., Enviro-Sci Consulting

About These Seminars: Virtually all scientific and engineering projects possess some degree of uncertainty in their results. Project managers and team members who carefully participate in this seminar will become equipped with a top-down framework for identifying and controlling sources of uncertainty, in both data and the estimates derived from them, from an end use and decision uncertainty perspective. This framework will serve as your orienting "grid" analogous to a free body diagram in engineering - regardless of the complexity you will identify the essentials and be able to collaborate with the appropriate project support to define and achieve the project's quantitative goals. You will become conversant in the essentials of uncertainty and estimation, improving your technical interaction and concurrence with project members, and you'll more effectively assist your clients in determining their actual needs and deliverables relating to data, analysis, and the client's decisionmaking process.

The bottom line value of this seminar is that it will help you to become more competitive in your existing areas of business, and you'll also expand the capabilities you can offer your clients.

Topics & Methods Discussed: *Concepts in Uncertainty and Probability, Stochastic Decision Analysis, Decision Performance, Uncertainty Analysis, Assurance Probabilities, Why Minimize Uncertainty?, End-Use and Decision-Driven Sampling Plans and Data Quality Objectives, Bayes Methods & Bayes Risk, Posterior Risk, Public Health/Ecological Risks, Worker Hazard Costs, Sampling & Remediation Costs, Multi-Criteria Decision Methods, Spatial Sampling Methods: Grid, Judgmental, Random/Stratified, Ranked Set, Adaptive Cluster, Incremental & Temporal, and Composite Sampling, Value of Perfect Information Method, Variance Components, Laboratory Uncertainty, Total Spatial Sampling Uncertainty, Robust Outlier Rejection, Gaussian Plume Guided Sampling, Optimal and Robust Nonparametric Regression, Monte Carlo Methods for Physics models and Integrated Decision/Sampling models, Data Cleaning, Hypothesis Testing, Prediction Intervals, Spatial Correlation & Kriging, Bootstrap Resampling and Estimation, Latin Hypercube Sampling, High Level Nuclear Waste Sampling, Nuclear Wasteform Simulant, Classification/Misclassification Probabilities, Neural Network Based Classification Decisions, Forensic Sampling for International Nuclear Treaty Verification, Sensitivity Analysis of Coupled Mechanistic Models, Inverse Based Calibration of Finite Element/Finite Difference Models.*

Who should participate? Program and Project Managers, Risk Managers, Scientists, Engineers, and Technical Editors will become equipped in this seminar.

What methods will be taught?

No specific methods will be rigorously taught in this series. Rather, you will gain an invaluable conceptual understanding so that you may effectively staff, co-scope, manage, and produce high quality deliverables in projects involving data, uncertainty, estimation, or decisionmaking. Also with this understanding and some experience, you will develop an acumen for discerning where these aspects arise in a new project, increasing your effectiveness with clients.

To this end, in this seminar you'll be systematically exposed to both the simple essentials and several advanced methods, including some from decision analysis. From a conceptual standpoint, your overall understanding will exceed that acquired by merely taking several statistics courses.

Is a background in calculus and statistics required?

No prior background is required. In fact having no previous exposure could be beneficial because it will help you maintain your focus on the concepts.

Seminar Series in Uncertainty and Estimation: Understanding, Communication, and Decision-Based Data Specification

PJ Chamberlain Ph.D., Enviro-Sci Consulting

Background: Since 1995, Dr. Pete Chamberlain has been Principal Scientist and Owner of Enviro-Sci Consulting in Big Lake, Alaska. Pete is also a practicing Structural Engineer. Prior to founding Enviro-Sci, Pete was a Senior Research Scientist with Battelle-Pacific Northwest National Laboratory (PNNL) in Richland, Washington, near the Hanford nuclear site. While at PNNL, initially he provided statistical consulting support to a numerous nuclear, geophysics, engineering, and chemistry projects at the Hanford site. Later, having expanded his skills through education and mentoring, he worked at PNNL as an Environmental Scientist and focused his work on multimedia contaminant fate/migration modeling, human health risk assessments at several prominent Superfund sites, development of integrated migration/exposure models for multiply-contaminated media, and finite element groundwater/contaminant model development.

Pete has also served as Visiting Assistant Professor at Texas A&M University in collaboration with the Office of Texas State Chemist.

Dr. Chamberlain holds a joint Ph.D. from Texas A&M University in Statistics (linear models, theoretical probability, parametric inference and decision theory) and in Agricultural Economics (industrial engineering methods applied to agriculture, stochastic/optimal control). Pete also has a B.S. level background in Structural Engineering and has done advanced study in Hydrology, Groundwater Modeling, Chemical Engineering, and Finite Element Modeling.

Owing to his diverse educational and experience background, Pete is a Multidisciplinary Consulting Scientist, and his perspectives enable him to bring other technical expertise into his statistical analyses, from crafting trend analysis tests realistically consistent with groundwater plume movement, to developing decision methodologies, to assessing uncertainty of multiattribute based decisions.

This series is dedicated to a mechanical engineer who possessed much of the understanding that I'll impart to you. Although a multiproject manager, this man actually developed the framework and secured the funding for my seminal statistical project "Decision Simulation." As part of the Hanford Single-Shell nuclear waste tanks characterization program, this project bidirectionally integrated the performance of a proposed stochastic risk-criteria based decision rule (for the disposition of each tank - i.e. in situ treatment vs. remove and process) and the tank sampling and analysis design. The framework attained some prominence and was credited to me, but I and my team only implemented his concept with refinements. Upon its completion, we presented it to the National Academy of Sciences.

Pricing and Other Terms

Pricing: 3-10 attendees.....\$1,200/person
11-20 attendees.....\$1,100/person
>20 attendees.....\$1,000/person

Payment: 50% payment to reserve a seminar date, (nonrefundable). balance due 7 days prior to presentation date, nonrefundable on/after date of travel

Materials: Registration includes copies of R-Cran based software routines and spreadsheet example. Unless otherwise specified, copyright of presentation materials is retained by Enviro-Sci Consulting.

Seminar Schedule:

Day 1 (pm, 3 hrs) **Session 1** ***A Comprehensive Example: From Dice to Decision Analysis***
Day 2 (am, 3 hrs) **Session 2** ***Essentials of Uncertainty, Probability, Estimation, & Decisionmaking***
(pm, 3 hrs) **Session 3** ***Project Applications and Examples; Self-Graded Examination***

Seminar Series in Uncertainty and Estimation: Understanding, Communication, and Decision-Based Data Specification

PJ Chamberlain Ph.D., Enviro-Sci Consulting

Dear Prospective Hosting Organization:

This series is equally targeted toward technical contributors, risk managers, project and program managers, and technical staff managers involved in high level program promotion. The stochastic decisionmaking topics covered will be of particular interest, not only as a basis for defining data quality goals, but also in their own right. The application of decisionmaking methodologies to complex problems naturally finds it's way to the top of organizations, then becomes propagated to other leaders in the private and government sectors who seek a rational and practical framework for solving notoriously intractable problems.

These seminars cover a huge amount of information, but an overall conceptual theme will become evident. During the presentation I'm always on the lookout for moments of wit and humor and I welcome this from the audience, which adds an element of fun and helps cognition and retention. After the presentations, the attendees like to stay and either discuss the subject matter and explore applications or solutions in their own projects, or just listen and eventually ask some questions. I'm always glad for their interest and participation, and time permitting, I'm happy to address all questions and observations.

Thank you, and I look forward to imparting some valuable skills to your staff and thereby improving the competitive position of your organization.

Pete

*Peter Chamberlain Ph.D.,
Owner, Principal Scientist, Structural Engineer
Enviro-Sci Consulting
Big Lake, Alaska 99652
(907) 232-8974 envirosci@enviro-sci.com
www.enviro-sci.com*



Seminar Series in Uncertainty and Estimation: Understanding, Communication, and Decision-Based Data Specification

PJ Chamberlain Ph.D., Enviro-Sci Consulting

Session 1: A Comprehensive Example - From Dice to Decision Analysis (3 hrs)

From simple *dice throws* to generating *error components and final measurements*....

then on to *estimating variance components* of these data....

then on to determining required *numbers of subsamples* at analytical, field, and spatial levels....

and also determining *maximum data tolerances*, first from simple *cost-oriented goals*....

then using *decision performance goals*, for alternative *decision rules* and alternative uncertainty-based *decision performance measures* that consider public health risks, ecological risks, worker hazards, sampling costs, and remediation losses.

This session is foundational to the entire series, serving as an instructional laboratory. Students will each participate in the data generation and estimation processes. They will utilize the instructor-provided applications to obtain the example calculations and graphics and they may assess, on their own, "what if" results in decisions/design/precision due to arbitrary changes in sample design, data variation, cost, and risk factors. From this session, participants will begin to develop an intuitive understanding of the interrelation between decision objectives, sample design, and data precision. They will also be well positioned to comprehend and absorb the material in Session 2, which both reinforces and expands upon the concepts and principles in this first session

Session 2: Essentials of Uncertainty, Probability, Estimation, and Decisionmaking (3 hrs)

Uncertainty

Randomness - Definition, Types and Causes

Some Contributions to total Uncertainty in Data - Additive components of uncertainty

Uncertainty in Final Data

Randomness in data acquisition - sampling, probe placement, etc.

Randomness in any preparation of sample material

Randomness in instrumentation's reading

Data Uncertainty - inherent variation in characteristic of interest

Coding Uncertainty

Covariance of Data – what is it? Definition and Variance-Covariance Matrix

Data Verification Concepts - plotting methods, outlier analysis

Designing an experiment to enable estimating each component of

variation, (re)designs that reduce the larger components, once identified

Controlling Uncertainty in Data, Consistent with Data Quality Objectives

Personnel Training, Quality Standards, Incentive Methods

Commensurate Instrumentation

Impact of spatial and temporal data acquisition on data uncertainty and uncertainty in estimates

Seminar Series in Uncertainty and Estimation: Understanding, Communication, and Decision-Based Data Specification

PJ Chamberlain Ph.D., Enviro-Sci Consulting

Session 2: Essentials (Cont'd)

Spatial Sampling

Physical distribution of a trait of interest in an area (2D) or volume (3D) often considered to be random, i.e. a stochastic process whose values at different locations are mutually correlated

Probability distribution of data is determined both by how sampling is done and the physical distribution of the property

Examples of Spatial Distributions – uniform, hotspot, Gaussian, geologically influenced, etc.

Sampling Regimes - Random, Judgmental, Stratified, Grid, Ranked Set, Adaptive Cluster, etc.

Composite Sampling - possible advantages

Spatial Covariance – unavoidable, affects uncertainty of estimated mean, but can optimize by selecting appropriate sampling plan.

Its impact on our spatial component of variation

Functional Estimation of Spatial Covariance – Semi-Variogram

What our overall variance covariance matrix looks like considering variance components and spatial covariance

Incorporating spatial sampling considerations into our algorithms to determine optimal number and configuration of field samples or to determine optimal levels of total spatial uncertainty (precision)

Assessing incremental or temporal sampling plans, mid-sampling – can reassess Expected

Risks or use Value of Sample Information methods.

Probability – What is it and how is it interpreted?

Definition and Interpretations of Probability

Various Probability Distributions for continuous, discrete, and mixed data

Joint Probability – Concept and Distributions

Estimation – Estimating Underlying Features of the Data

Estimates as functions of the data, which estimate the true underlying values of interest

Examples of estimates for various parameters

Bias in estimation – undesirable but sometimes acceptable if estimate has lower variance and bias can be corrected

Outlier-resistant methods - robust estimates

Predictions based on parameter estimates: regression response surfaces, temporal predictions, spatial predictions (kriging also utilizes data directly)

Uncertainty of Estimates and Predictions

Why uncertainty is undesirable

Uncertainty in parameter estimates, regression predictions, and temporal/spatial predictions

Some graphical representations of point estimate and prediction uncertainties

Reduce uncertainty in estimates by redesigning sampling/data acquisition and analysis regime for lower variance estimation

Estimation uncertainty also improved by reduction in variation of the data

Which does the End User/Decisionmaker actually need - Estimates or Assurance Probabilities?

Estimates and conservative estimates (upper bound, lower bound) vs. assurance/exceedance probabilities for alternative values

Seminar Series in Uncertainty and Estimation: Understanding, Communication, and Decision-Based Data Specification

PJ Chamberlain Ph.D., Enviro-Sci Consulting

Session 2: Essentials (Cont'd)

Rolling It All Up to the Decision Level

Goal of Decision – To make the *correct decision with high probability*

Decision Uncertainty and Reliability -

How data variation causes variation in the decision

Computing Decision Uncertainty - parametric, Monte Carlo, aparametric

Decision Rule Types, Correct, Incorrect Decisions, Continuous Decision Sets

Evaluating Decision Performance - Determine Probabilities of Correct, Incorrect, and In-Between Decisions

Underlying parameter for data unknown - requires assuming specific parameter value

Hypothesis Testing - Special case of decisionmaking using 0-1 decision rule

The Whole 9 Yards: Treating underlying parameters as random variables - e.g. due to spatial uncertainty or temporal uncertainty - *The Bayes Approach*

Computing Data Uncertainty Requirements Based on Required Decision Success Rates
Measurement & analysis uncertainty

Best spatial or temporal sampling plan - via mathematical inversion or simulating entire data and decision process - National Academy of Sciences Presentation "Decision Simulation" in Session 3

Choosing the Best Decision Rule - Comparing the performances of various decision rules - Assumed Parameter, Bayes

A Look at Monte Carlo Simulation – Let the Computer do the Work!

Assess Uncertainty of Estimates or Actions which depend up the data, without doing calculus solutions.

Rapid for simple problems, the only way for complex problems or results of finite element or other numerical algorithms

How it works

Random Number Generation. Based on the uniform RNG for most Monte Carlo methods, but requires estimates of the probability distributions of the data and all variables to be randomized

Latin Hypercube Sampling – insures coverage within each distribution, can induce correlations between random variables

Evaluating Results – tabulate means, variances, empirical distributions, other features of either data or functions of data (e.g. risks, costs, losses)

Can evaluate Expected Losses, Posterior Risk, Bayes Risk, for alternative Decision Rules, under alternative assumptions and sampling plans – i.e. sample data points are the randomization unit

Data Resampling – Sometimes an alternative to parametric Monte Carlo methods

Summary of Session 2: What You Should Remember

Seminar Series in Uncertainty and Estimation: Understanding, Communication, and Decision-Based Data Specification

PJ Chamberlain Ph.D., Enviro-Sci Consulting

Session 3: Project Applications and Examples (3 hrs)

A Statistician's Role - Project Support Spectrum

Approaches to a problem

Spectrum of statistician's deliverables for alternative project types

Design/Sampling and Analysis Plan

What is it?

Experimental design

Spatial/physical sampling

Optimal plan - minimizes uncertainty in estimates, decision uncertainty, costs, etc.

Designs which facilitate estimating variance components

Spatial and temporal considerations

Analysis of Data: Quality Assurance and Estimation

Data validation using multiway, temporal, spatial plots

Outlier Analysis to identify "false" data points

Variance component analysis – why?

Why Uncertainty is Undesirable - Illustrations

Swamps-out the detection of important differences

Compromises decision rule integrity

Causes mal-selection of important technologies

Bringing it All Together (so far) - Examples

Estimates of Underlying Parameters or Features of Data

Estimating underlying mean, variance, quantiles - parametric, nonparametric, small sample asymptotics

Parameters as functions of time, spatial parameters, finite element inverse method

Predictions using parameter estimates - regression, temporal, spatial - kriging, spatial autoregression

Testing (Hypotheses)

Ordinary Hypotheses

Complicated Hypotheses - trend behavior, spatial/physical extent of characteristic of interest

Monte Carlo based testing

Assurance Probabilities

Exceedance Probabilities

Prediction Envelopes

Probability Contours

Misclassification Probabilities

Probabilistic Performance Assessment, Probabilistic Risk Assessment

Some Project Applications

Contaminated Nuclear Wasteform Simulant

Single Shell Tank Decision Simulation Project/NAS Presentation

Treaty Verification: Integrated Sampling, and Classification/Decision Framework

Other Projects and Applications

**Seminar Series in Uncertainty and Estimation: Understanding, Communication, and
Decision-Based Data Specification**

PJ Chamberlain Ph.D., Enviro-Sci Consulting

Summary of Session 3: What You Should Remember

**Recap: Session 1 example, Session 2 principles, Session 3 project
applications**

Examination (multiple choice, self-graded)

Closing