Guidance on Assigning Values to Uncertain Parameters in Subsurface Contaminant Fate and Transport Modelling

National Groundwater & Contaminated Land Centre report NC/99/38/3

Guidance on Assigning Values to Uncertain Parameters in Subsurface Contaminant Fate and Transport Modelling

A McMahon, J Heathcote, M Carey, A Erskine & J Barker

June 2001

Research contractor:
Entec UK Ltd

National Groundwater & Contaminated Land Centre
Environment Agency
Olton Court
10 Warwick Road
Olton
Solihull
B92 7HX

National Groundwater & Contaminated Land Centre report
NC/99/38/3

Publishing Organisation
Environment Agency
Rio House
Waterside Drive
Aztec West
Almondsbury
Bristol BS32 4UD

Tel: 01454 624400 Fax: 01454 624409

ISBN: 1 857 05605 1

Dissemination status
Internal: Released to Regions
External: Released to Public Domain

Statement of Use
This document provides guidance to Environment Agency staff on assigning values to uncertain parameters in subsurface fate and transport models. This document forms one of a set of three reports produced under this project.

Research Contractor

This document was produced under Project NC/99/38 by:

Entec UK Ltd
160-162 Abbey Foregate
Shrewsbury
Shropshire
SY2 6BZ

The Environment Agency's Project Manager was Jonathan Smith, National Groundwater & Contaminated Land Centre.

The Project Board comprised of Dr Bridget Butler, Dr John Davys, Sarah Evers, Paul Hulme, Dr Janet MacMillan, Tony Marsland, Dr Martin Shepley and Jenny Thomas.

The Agency is grateful for peer review comments from David Hall of Golder Associates.

CONTENTS

List of Tables

List of Figures

List of Appendices

EXECUTIVE SUMMARY

The Environment Agency has duties and powers to ensure the protection of groundwater and the remediation of contaminated land and groundwater. These responsibilities are covered by the Groundwater Regulations 1998, Water Resources Act 1991 (WRA 1991) and the Environmental Protection Act 1990 (EPA 1990). To assist in this regulatory role, environmental risk assessment methodologies are employed by the Agency as part of the decision-making process. With regard to the protection and remediation of groundwater, contaminant fate and transport models are often used. They may take the form of deterministic models or probabilistic tools, such as LandSim and ConSim.

In using contaminant fate and transport models there are three main sources of uncertainty:

- Conceptual - is the system sufficiently well understood and defined? Unless the conceptual model is right, any further analysis will be spurious.
- Model - does the mathematical model adequately describe the conceptual system and its behaviour?
- Parameter - are the parameters values used in the model known and adequately described?

The purpose of this document is to provide guidance, both internally for the Agency and for external bodies, on parameter inputs to quantitative probabilistic approaches to contaminant fate and transport modelling. These probabilistic approaches address the uncertainty and variability of input parameters by using probability distributions to describe each parameter. This document includes guidance on the theory of probability distributions, the practical application of these distributions in contaminant transport models and the interpretation of results. This document also emphasises the importance of a robust conceptual model.

A key step in contaminant transport modelling is the selection of values for input parameters. In a probabilistic model, the range of values for an input parameter is defined by a probability density function (PDF). These take account of uncertainty and variability which may arise from a number of sources especially:

- Intrinsic Uncertainty;
- Experimental Uncertainty;
- Heterogeneity;
- Time Variation.

This approach does not account for an incorrect conceptual model. The statistical definitions and characteristics of the commonly used probability distributions are explained in the text, and provide a theoretical background to the application of probabilistic models.

In order to produce defensible results, contaminant transport models require the input of site-specific data wherever possible, but often data are limited or not available. When data are available they must be critically reviewed to establish their validity and their appropriate use in defining PDFs. Key questions include:

- Are the data from the same (or a relevant) population?

- Are the data time dependent?
- Are extreme values representative of the system?
- Are measurements representative of the system as a whole?
- Are data reasonable and physically possible?
- Can data be scaled up from the scale of measurements to the scale of the problem?

This data review should form part of the conceptual model.

The report describes methods for fitting PDFs to field data (even if these data are limited), together with methods for upscaling field measurements to provide a probabilistic description of the average system behaviour.

Where data are not available, parameter values and ranges may need to be obtained from other sources. Sources may include published literature or reliance on expert judgement with appropriate explanation. Other, less widely used methods are available, such as the Maximum Entropy and Bayesian Methods.

The interpretation of model results is a key step in probabilistic modelling. It is important to understand the sensitivity of the model to the parameter distributions that have been used and to check model results are valid when compared with field data. The report provides guidance on understanding the sensitivity of models to parameter distributions.

Keywords

Contaminant transport, uncertainty, risk-assessment, probabilistic, probability density functions, PDF.

GLOSSARY

Analytical model	Exact mathematical solutions of the flow and/or transport equation for all points in time and space. In order to produce these exact solutions, the flow/transport equations have to be simplified (e.g. very limited, if any, representation of the spatial and temporal variation of the real system).
Bayesian method	A technique of re-calculating probable outcomes by incorporating new data, combining what is known with what is expected.
Binomial distribution	Describes the number of times an event occurs in a fixed number of trials.
Compliance point	Location where a target concentration must be achieved.
Conceptual model	A simplified representation of how the real system is believed to behave based on a qualitative analysis of field data. A quantitative conceptual model includes preliminary calculations for key processes.
Continuous variable	Parameter that varies continuously in space or time.
Convergence	When an infinite series or sequence of numbers tends to a limit.
Confidence interval	Interval constructed to have a known probability of containing the true value of an unknown parameter.
Covariance	A measure of the correlation between two parameters.
Cumulative distribution function (CDF)	Mathematical function of a parameter representing the probability distribution, usually conceptualised as a graph showing the percentage of values less than or equal to a given value.
Dependent parameter	Two parameters are dependent if knowledge of the value of one of them alters the probability distribution of the other.
Deterministic model	A model where all elements and parameters of the model are assigned unique values.
Discrete variable	Variable that is defined by a series of distinct values, but which cannot take intermediate values.
Dispersivity	A property that quantifies the physical dispersion of a solute being transported in a porous medium.
Distribution	Description of the frequency of observations.
Event	Single occurrence.
Exponential distribution	Used to describe events occurring randomly with time.

Gamma distribution	This distribution is always positive but is skewed (not symmetrical about its mean). It is often used to represent the time between occurrence of events.
Geometric mean	Exponential of mean of logarithms, the nth root of the product of n values.
Harmonic mean	Reciprocal of the arithmetic mean of the reciprocals of the observations.
Heterogeneity	Variability in space.
Histogram	A column graph showing numbers of measured values occurring within equally-spaced intervals.
Hydraulic conductivity	A coefficient of proportionality describing the rate at which water can move through a permeable medium.
Hydraulic gradient	The rate change in total hydraulic head with change in distance in a given direction. (dimensionless).
Latin Hypercube method	A sampling technique used by Monte-Carlo analysis where the parameter space is sub-divided and parameter values are then picked from each interval as a way of ensuring that values are taken from the entire probability distribution.
Log-normal distribution	A probability distribution whose logarithms are distributed normally.
Log-uniform distribution	A probability distribution whose logarithms are distributed evenly between an upper and lower bound.
Log-triangular distribution	A probability distribution whose logarithms form a distribution function that looks like a triangle.
Maximum entropy method	A theoretical approach that optimises distributions given the information available.
Mean (arithmetic)	Arithmetic average of a set of values, 1/*n*th of the sum of n values.
Median	Value for which there is a 50% probability of the actual value from a distribution being greater.
Mode	Most likely value of a set of observations.
Mathematical model	Mathematical expression(s) or governing equations which approximate the observed relationships between the input parameters (recharge, abstractions, transmissivity etc) and the outputs (groundwater head, river flows, etc). These governing equations may be solved using *analytical or numerical* techniques.

Model	A simplification of reality in order to aid in the understanding of and/or predict the outcomes of the real system. In this report the term 'model' is used to describe the code or equations plus the data.
Monte-Carlo analysis	A method of carrying out a calculation using probability distributions rather than numbers. It is simple and powerful and involves repeated sampling from the input distributions.
Normal distribution	A probability distribution characterised by a mean and standard deviation.
Numerical model	Solution of the flow and/or transport equation using numerical approximations, i.e. inputs are specified at certain points in time and space which allows for a more realistic variation of parameters than in *analytical models.* However, outputs are also produced only at these same specified points in time and space.
Parameter (hydrogeological)	Physical property of the system under investigation (e.g. hydraulic conductivity).
Parameter (distribution)	Characteristic of a theoretical probability distribution (e.g. mean, standard deviation).
Parameter space	A representation of all the values that could be taken by all the parameters (usually imagined as an *n*-dimensional volume, where *n* is the number of parameters).
Percentile	The value below which occur a specified proportion of observations (in an ordered set of observations).
Population	All the possible outcomes of an event, e.g. hydraulic conductivity of all possible $1cm^3$ of an aquifer.
Porosity	The ratio of the volume of void spaces in a rock or sediment to the total volume of the rock or sediment. (dimensionless).
Probability	Any outcome of an event can be allocated a probability p (with $0 \leq p \leq 1$) such that, if the event happened an infinite number of times, the proportion of times that this outcome occurred would be p. If p=0 then the event never occurs, and if p=1 the event always occurs.
Probability density function (PDF)	Mathematical function representing the probability distribution of a parameter, i.e. the likelihood that a given value will occur.
Probability distribution	The probabilities associated with the possible outcomes of an event. The 'event' in the context of this document is usually the value of a parameter.

Parameter of distribution	Characteristic of a theoretical distribution (e.g. mean, standard deviation).
Probabilistic model	An aggregation of model realisations, where the input parameters to each realisation are characterised by probability distributions.
Realisation	Single calculation with a single set of parameter values (in the context of repeated parameter sampling).
Receptor	An entity (e.g. human, animal, controlled water, vegetable, building, air) which is vulnerable to the adverse effects of a hazardous substance or agent.
Remedial target	The goal of remedial activity set for the site; may take the form of a maximum or minimum permitted concentration in the soil or groundwater.
Recharge	The quantity of water that reaches a water resource such as an aquifer, calculated as rainfall less runoff, evapotranspiration and soil storage.
Retardation	A measure of the reduction in solute velocity relative to the velocity of the advecting groundwater caused by processes such as adsorption.
Sample	A sub-set of the population.
Scale dependency	The tendency of a parameter to take different values depending on the scale over which it is being measured.
Standard deviation	Measurement of the variability of a distribution. Square root of the variance.
Skewed distribution	A distribution which has a degree of asymmetry about the centre value of the distribution.
Sensitivity analysis	A process of identifying the model parameters that have most effect on the model output.
Stochastic field	Used to describe the uncertainty of a parameter which varies in space.
Triangular distribution	A simple probability distribution with a PDF graph that looks like a triangle and is defined by a minimum, most likely and maximum value.
Uncertainty	The degree to which a well-defined and located parameter (e.g. the horizontal hydraulic conductivity of a 1 cm cube of rock at a defined location) is unknown.
Uniform distribution	A simple probability distribution giving equal chance for a range of values given a minimum and maximum value.
Upscaling	The process of deriving an effective value for a parameter applicable to the scale of interest, using information about the

	value of the parameter at a smaller scale, its variability, and the process of interest.
Variability	The degree to which a well defined parameter varies in space and/or time, e.g. the hydraulic conductivity of all possible 1 cm cubes of rock from a particular aquifer horizon.
Variance	Measurement of the variability of a distribution. Square of the standard deviation.

1. Introduction

1.1 Background

The Environment Agency (the Agency) has duties (i.e. obligations) and powers (i.e. ability and discretion) to ensure the protection of groundwater and the remediation of contaminated land and groundwater. These responsibilities are covered by the Groundwater Regulations 1998, the Water Resources Act 1991 (WRA 1991) and the Environmental Protection Act 1990 (EPA 1990). Fundamental to the Agency's regulatory role is the assessment of risk to the environment and determination of the need for protection or remediation. The Agency employs the principle of risk assessment (the risk of a contaminant source causing harm or pollution via a given pathway at an identified receptor) to assist with decision making for problems involving contaminant transport and also encourages external bodies to adopt the risk assessment philosophy.

Risks may be assessed qualitatively (e.g. a high, medium or low risk of pollution) or quantitatively (e.g. by predicting the concentration and consequences of a contaminant at a specified location at a certain time). This report deals specifically with quantitative approaches for contaminant transport modelling in groundwater, as part of quantitative risk assessment, which can be divided into two categories: deterministic and probabilistic. Deterministic assessments involve the assignment of a single value to each parameter, and the calculation results in a single number. This approach implies a high degree of certainty in the input data, e.g. the input parameter can be defined by a single value or its variability is known everywhere. A high proportion of environmental risk assessments involve studies of the subsurface where such a level of certainty is not present.

Probabilistic approaches provide methods of addressing uncertainty or variability in a known and structured way using probability distributions of values; as knowledge increases the corresponding reduction in uncertainty can be incorporated. A range of possible outcomes, which can be described by a probability distribution, will be generated by a probabilistic model as a result of the combination of different input parameter values (e.g. by calculating the percentage chance that a specified concentration will be exceeded at a specified location at a certain time).

Uncertainty in defining parameter values results both from our lack of knowledge or understanding of a system and from potential errors in measurements or test results. However, most parameters also have natural variability. Hydraulic conductivity, for example, will vary spatially within an aquifer and this variability will be reflected in test results. Both natural variability and uncertainty need to be considered when assigning values to input parameters (Hertwich *et al.*, 1999).

The probability density function (PDF) is a commonly used method to describe the likelihood of a parameter having any particular value and its use is the topic of this document.

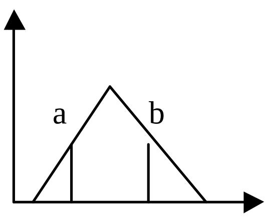
1.2 Purpose of this document

This document provides practical guidance on assigning values, in the form of probability distributions, to uncertain or variable parameters used in contaminant fate and transport models of the subsurface, and highlights the potential problems of their inappropriate application. The document is not intended to give a rigorous scientific explanation of probability distributions or probability density functions, but to guide the reader on understanding the main issues. The emphasis is on the practical application of the theory to environmental risk assessment and contaminant transport modelling.

This document is not intended as a 'recipe book', but as a guide to using data that have been collected as part of a properly structured investigation and to assigning values to parameters which cannot be measured. The document does not specifically deal with the development of a conceptual model, but does emphasise its importance to contaminant fate and transport modelling (see also Environment Agency, 2001a).

1.3 Definition of terms

Many of the terms in popular use in the context of contaminant fate and transport modelling and environmental risk assessment are often used loosely and this can lead to misunderstanding. The glossary in this report aims to provide a clear definition of the technical terms used in the report. However, there are a number of key terms that are critical to understanding the concepts in this document and these are reproduced below:

Key Definitions:	
Uncertainty	The degree to which a well-defined and located parameter (e.g. the horizontal hydraulic conductivity of a 1 cm^3 of rock at a defined location) is unknown.
Variability	The degree to which a well defined parameter varies in space and/or time, e.g. the hydraulic conductivity of all possible 1 cm cubes of rock from a particular aquifer horizon.
Probability	Any outcome of an event can be allocated a probability *p* (with $0 \le p \le 1$) such that, if the event happened an infinite number of times, the proportion of times that this outcome occurred would be *p*. If p=0 then the event never occurs, and if p=1 the event always occurs.
Probability distribution	The probabilities associated with the possible outcomes of an event. The 'event' in the context of this document is usually the parameter having a particular value (or narrow range of values).
Probability density function (PDF) 	Mathematical function of a parameter representing the probability distribution, usually conceptualised as a graph of PDF against the parameter value. In simplistic terms, the magnitude of the function represents the relative likelihood of the parameter taking that value. In the case of a discrete set of possible values, the function represents the probability itself but for continuously variable parameters this definition breaks down and we have to redefine the PDF more rigorously. In this case, the area under the PDF curve between *a* and *b* represents the probability of the parameter being between *a* and *b*.
Model	A simplification of reality in order to aid in the understanding of and/or predict the outcomes of the real system. In this report the term 'model' is used to describe the code or equations plus the data.
Probabilistic model	An aggregation of model realisations, where the input parameters to each realisation are characterised by probability distributions.
Deterministic model	A model where all elements of the model are assigned a single value.

1.4 Target audience

This document is aimed at environmental professionals undertaking or reviewing environmental risk assessments and contaminant fate and transport modelling. Users need to be numerate and have an understanding of the principles of hydrogeology, the movement and behaviour of contaminants in the ground, the principles of risk assessment and the constraints of the legislation/regulatory policy within which decisions are made. This document assumes that the reader has the necessary understanding of these principles. To add to this understanding, this document aims to show how a knowledge of uncertainty in hydrogeology can be incorporated within the risk assessment to provide an indication of the possible environmental impacts.

1.5 Relationship to other procedures

This guidance note is one of a number of technical guidance documents produced by the Agency's National Groundwater and Contaminated Land Centre and is aimed at improving understanding and capability, both inside and outside the Agency, in the risk based approach to environmental protection. This document is one of a series of three technical guidance notes produced on the subject of contaminant fate and transport modelling in the subsurface. The other two documents in this series are:

- *Guide to Good Practice for the Development of Conceptual Models and the Selection and Application of Mathematical Models of Contaminant Transport Processes in the Subsurface* (Environment Agency, 2001a).
- *Guidance on the Assessment and Interrogation of Subsurface Analytical Contaminant Fate and Transport Models* (Environment Agency, 2001b).

The first document provides guidance on:

- Whether a mathematical modelling approach is justified;
- Development of a conceptual model and how this should then translated through to a mathematical model;
- Selection of a modelling approach;
- Construction and testing of the mathematical model;
- Assessment of model results;
- Presentation and reporting.

The second document provides guidance to Agency staff on how to critically examine a submitted model.

This document is intended as a supporting document, providing specific guidance on dealing with uncertainty in fate and transport modelling and assigning values to uncertain parameters.

These documents are intended to be used in conjunction with the Agency R&D report *Methodology for the Derivation of Remedial Targets for Soil and Groundwater to Protect Water Resources* (Environment Agency, 1999a) which presents a framework for deriving remedial targets for soil and groundwater to protect water resources and includes a brief discussion of contaminant fate and transport models.

Two risk assessment modelling software packages have also been developed for use by the Agency in assessing the potential impacts on water quality:

- *LandSim* allows probabilistic quantitative assessments of landfill site performance and its likely impact on water quality (Environment Agency, 1996, 2001c).
- *ConSim* allows probabilistic quantitative assessment of the likely impact on water quality from contaminated sites (Environment Agency, 1999b).

These packages allow uncertainties in parameter definition to be considered, and describe data uncertainty using probability density functions (PDFs). This document is directly relevant to users of the above packages and is designed to provide guidance that complements the information in the user manuals.

1.6 Report layout

This document should be used in conjunction with the *Guide to Good Practice for the Development of Conceptual Models and the Selection and Application of Mathematical Models of Contaminant Transport Processes in the Subsurface* (Environment Agency, 2001a) which contains more detailed and comprehensive discussion of the key topics.

Chapter 2 gives an overview of the general approach to contaminant fate and transport modelling and the parameters that need to be considered. This chapter also discusses parameter uncertainty and outlines the procedure for defining and applying PDFs.

Chapter 3 explains PDFs in detail and how they can be defined. The most commonly used types of PDF are described and the significance of independent and dependent parameters is discussed.

Chapter 4 looks at data requirements and the practicalities of selecting appropriate PDFs based on the data available.

Chapter 5 describes modelling methods which use PDFs and discusses the number of simulations required to obtain statistically valid results.

Chapter 6 sets out an example problem which is used to illustrate the influence of PDFs and combinations of PDFs for different parameters on modelling results.

Chapter 7 discusses the interpretation of modelling results, how they relate to observed behaviour and the importance of checking that modelling results are credible. The potential pitfalls of interpreting results of modelling with PDFs are also outlined.

2. Overall approach

2.1 The application of a probabilistic approach to fate and transport modelling

The main steps in the application of a fate and transport model are summarised in Figure 2.1. Essential to this process is the development of a conceptual model to describe system behaviour, determining the objectives for the model (including whether the use of a mathematical model is appropriate), and selecting a computer model that adequately represents the system behaviour. The selection of hydrogeological parameter values forms a key component of this process and is the focus of this document. Further detailed guidance on the application of fate and transport models is given in Environment Agency (2001a and 2001b).

Recognition and understanding of uncertainty is key to fate and transport modelling. Uncertainty will be associated with:

- Our understanding of the system (conceptual uncertainty). For example, is the decrease in contaminant concentration away from the source due to natural degradation and what processes are giving rise to and controlling degradation?
- Whether the mathematical model adequately describes the system behaviour (modelling uncertainty). Application of a mathematical model will require a number of assumptions to be made about the system behaviour (for example, can sorption be described by a linear isotherm?). Uncertainty may be attached to whether these assumptions are valid.
- Definition of parameter values (parameter uncertainty). For example uncertainty may be associated with our measurement of a parameter value or our knowledge of the natural variation in a parameter value. A parameter may, however, not be knowable (intrinsic uncertainty) i.e. future rainfall events. This is discussed in greater detail in Sections 2.2 and 3.1.

The development of the conceptual and mathematical model should be an iterative exercise, with the model being continually challenged and refined through reference to field observations. Nevertheless uncertainty will still be associated with these models. The key is understanding the effect of uncertainty and deciding whether we can accept this level of uncertainty in using the results from a fate and transport model.

This document sets out approaches to managing parameter uncertainty, but these approaches do not provide a solution to conceptual uncertainty. The latter often is the main problem in assessing contaminant fate and transport and therefore it is important that investigations are of a sufficient standard to define the system with confidence. In many cases further investigations may be required if development of the conceptual model and mathematical model have shown inconsistencies and/or shortfalls in our definition of the system behaviour.

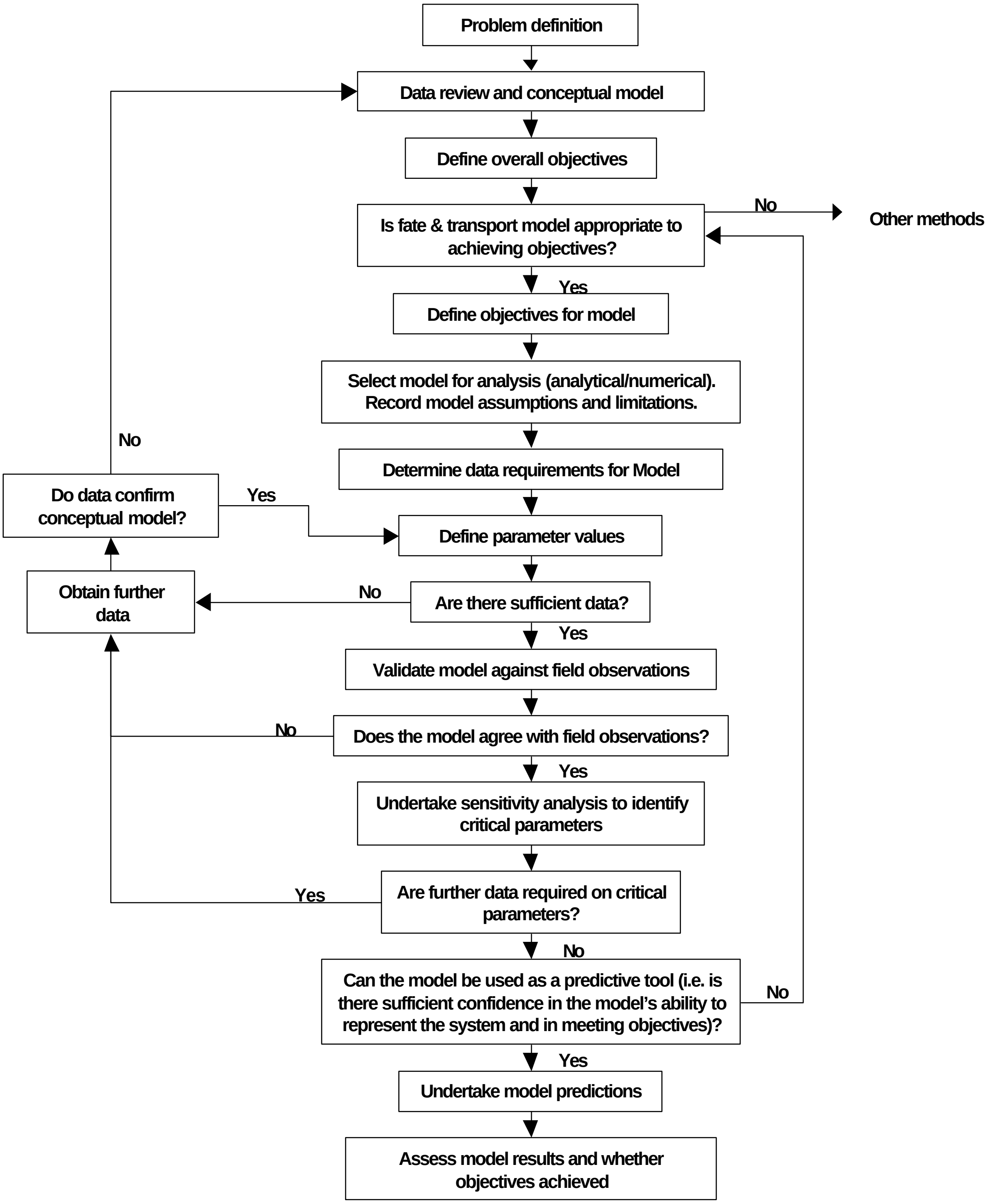


Figure 2.1 Basic steps in the application of a fate and transport model

It is important to recognise that an incorrect conceptual model will invalidate the results of the study. Even though a model may reproduce field observations, this does not necessarily mean that it is correct (although it will obviously improve our confidence in the model) as (for example) different combinations of assumptions and/or parameter values may produce the same model results. A more realistic test will be whether the model continues to match field conditions under different conditions.

The main parameters that are relevant to fate and transport modelling in the unsaturated and saturated zones are listed in Table 2.1. This table also identifies some of the sources of uncertainty in defining these parameters and some of the common assumptions that are made in mathematical models. The actual parameters used will depend on the type (analytical/numerical) and complexity of the model.

Model parameters may be defined as single values (as in a deterministic model) or as a range of values (as in a probabilistic model). A deterministic model will give the result as a single value, whereas a probabilistic model will give a range of results.

In practice, a parameter value will either be:

- A single value (such as volume of contaminant released), but for which there may be uncertainty about the actual volume;
- A parameter which varies naturally (such as spatial variation in clay content within the aquifer) and there is uncertainty about this variation and how our finite set of field measurements describes this variation.

This is discussed in further detail in Section 2.2.

In a probabilistic simulation, the possible combinations of parameter values are used as input to the model, resulting in a range of possible outcomes as illustrated by Figure 2.2. In relation to the use of fate and transport models in risk assessments, a probabilistic approach will provide information on the likelihood of particular impacts on a receptor, taking account of uncertainty in defining the parameters that control contaminant transport.

Sensitivity analyses can be used in combination with either type of model to examine the influence of changing a parameter value on the calculated result and thereby identify which parameters have the greatest influence on the result.

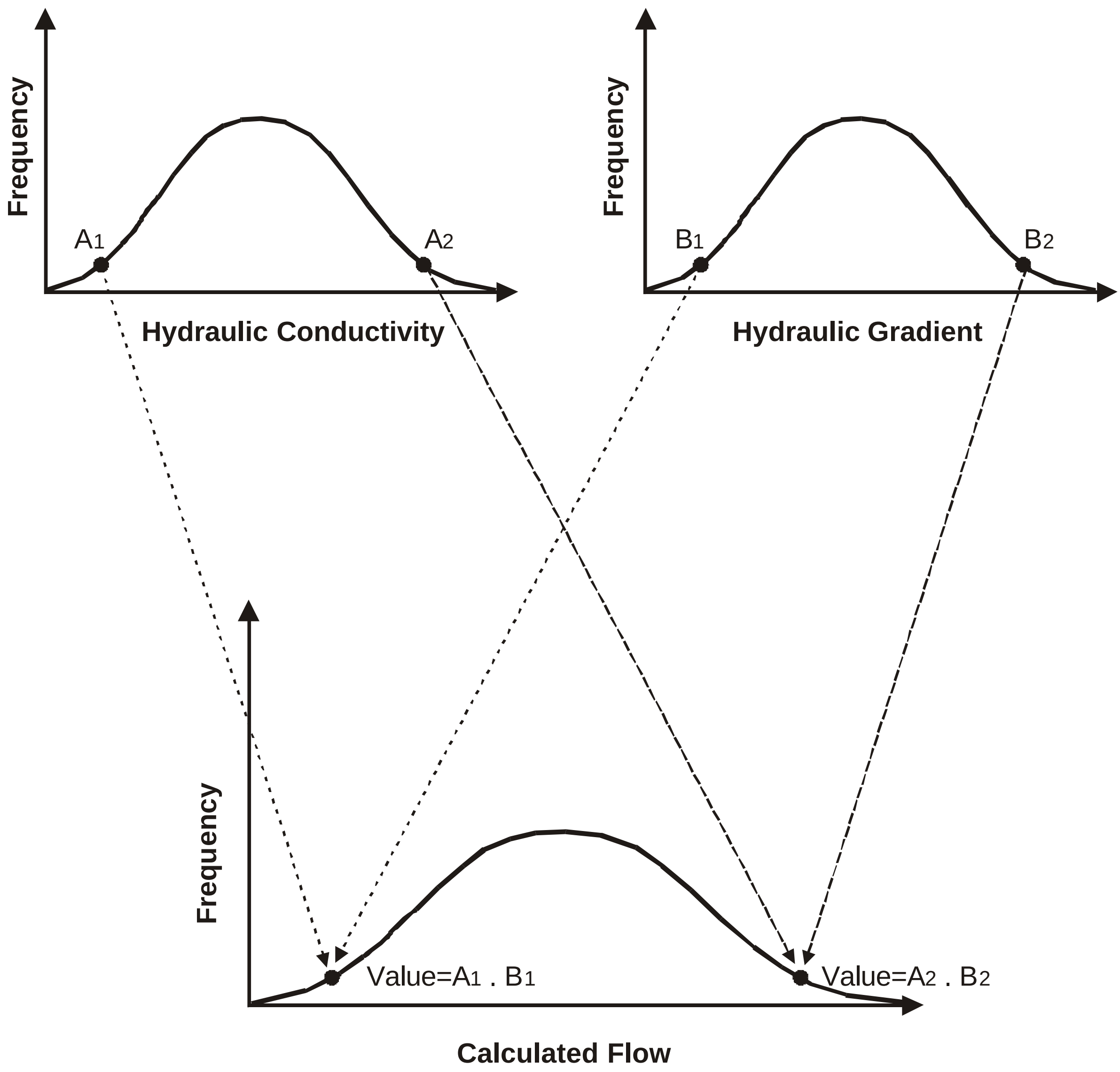


Figure 2.2 Combination of Parameter Distributions

Table 2.1 Influence of model parameters on contaminant transport

Parameter	Influence on contaminant transport	Example PDF	Main uncertainties
Source term	• Mass of contaminant entering the system. • Contaminant concentrations in groundwater.	Log-triangular (mass) Uniform (timing)	➢ Mass and timing of contaminant of release. ➢ Source concentration and geometry (borehole investigation may not allow plume geometry to have been defined such that maximum concentration is underestimated).
Recharge	• Dilution • Contaminant loading (leaching)	Uniform, Normal	➢ Seasonal variation in effective rainfall and leaching of contaminants. ➢ Indirect recharge (leaking drains, leakage from rivers, surface water run-off from areas of hard standing). ➢ Recharge is typically derived indirectly from measurement of rainfall and evaporation. Measurement of these values may be uncertain.
Hydraulic conductivity (k)	• Rate of contaminant transport (advection) and arrival time at receptor. • Calculated groundwater dilution.	Log-normal, Log Triangular	➢ Contaminant transport is sensitive to this parameter. Field measurements can often vary by more than an order of magnitude (due to the natural heterogeneity of most aquifers). ➢ Important parameter to determine by field measurement - literature values unlikely to be sufficiently precise, although Aquifer Properties Manual (BGS/Environment Agency 1997 and 2000) data may be adequate for some problems. ➢ Values can be measured using range of techniques including laboratory testing, falling head tests, pumping tests. Each of these methods results in measurement of hydraulic conductivity over different volumes which may not be directly relevant to the correct 'average' for the whole volume modelled.
Hydraulic gradient (*i*)	• Rate and direction of groundwater flow. • Calculated groundwater dilution.	Uniform	➢ Important to determine by field measurements (minimum of three boreholes required). ➢ Anomalous gradients may be obtained where monitoring borehole intersects more than one aquifer horizon. ➢ Hydraulic gradient and direction of flow can vary with time. ➢ Hydraulic gradient is dependent on hydraulic conductivity. Steep gradients unlikely to occur in zones of high permeability.

Table 2.1 (continued) Influence of model parameters on contaminant transport

Parameter	Influence on contaminant transport	Example PDF	Main uncertainties
Porosity (*n*)	• Rate of contaminant movement and arrival time at receptor.	Uniform, Normal	➢ Measurement of the transport porosity is often difficult and expensive and therefore the value of this parameter is often assumed based on measurement of total porosity, literature values or expert opinion. ➢ Relative importance of fissure flow and intergranular flow. ➢ Importance of fissure-pore water diffusion in dual porosity aquifers in contaminant transport.
Groundwater velocity	• Rate of contaminant movement and arrival time at receptor.		➢ Typically calculated as a function of hydraulic conductivity, gradient, porosity (see above).
Dispersivity	• Spreading of contaminant. Arrival time at receptor. • Reduction in contaminant concentrations.	Triangular, Uniform	➢ Scale dependent. ➢ Rarely measured in the field (due to time and cost of measurement). Values typically assumed to be a function of pathway length (for example longitudinal dispersivity often assumed to be one tenth of pathway length). Important to determine how dispersion is being used to represent contaminant movement.
Diffusion	• Spreading of contaminant	Triangular, Uniform	➢ Usually only significant where rates of groundwater flow are low, e.g. strata characterised by values of hydraulic conductivity of less than 1×10^{-9} m/s. Rarely measured in the field due to relatively slow rates of contaminant movement. Values for diffusion coefficient usually based on literature.
Mixing depth/aquifer thickness	• Dilution by groundwater flow • Significance of vertical dispersion (for thin aquifers vertical dispersion should be negligible)	Uniform	➢ Typically estimated based on experience, theoretical calculation, hydrographs (groundwater level variation), borehole logs (high k zones) and/or vertical dispersivity. Mixing depth will typically be less than the aquifer thickness.
Bulk density	• Used in calculation of contaminant retardation (see below)	Uniform, Normal	➢ Measurement is straight forward and relatively cheap once samples have been obtained. Literature values typically fall in narrow range and can reasonably be used; consequently calculations of retardation rates are relatively insensitive to this parameter.
Sorption/retardation	• Rate of contaminant migration.	Triangular, Uniform	➢ Typically represented as a linear reversible reaction. Sorption may be more accurately represented by a non-linear isotherm.
Partition coefficient (K_d)	• Used in calculation of retardation of contaminant or in soil water partitioning • Rate of contaminant migration	Triangular, Uniform (use Log distributions if large range in values)	➢ Partitioning can be sensitive to soil or groundwater pH, pK_a, H, f_{oc} and contaminant K_{oc}, and values can range by more than an order of magnitude. Typically K_{oc} and H based on literature values, and combined with field derived f_{oc} to derive site-specific K_d.

Table 2.1 (continued) Influence of model parameters on contaminant transport

Parameter	Influence on contaminant transport	Example PDF	Main uncertainties
Organic partition coefficient (K_{OC})	• Used in calculation of retardation of contaminant and soil water partitioning. • Rate of contaminant migration	Triangular, Uniform	➢ Often based on literature values, although a range of different values may be given in literature sources.
Fraction of organic carbon (f_{OC})	• Calculation of partition coefficient	Triangular, Uniform	➢ For low f_{OC} values (less than 0.001), organic transport may be dependent on mineral surface area.
Cation exchange capacity (CEC)	• Delay for breakthrough of cations (e.g. potassium, ammonium)	Triangular, Uniform	➢ Measurements' sensitivity to pH, Eh, solute concentration. ➢ Aquifers have a finite capacity for cation exchange. Cations will compete for available exchange sites and this is typically handled by specifying a reaction efficiency (usually based on literature values or expert opinion) as a measure of available sites. Cation exchange is normally a reversible process.
Biodegradation	• Reduction of contaminant mass and concentration.	Triangular, Uniform	➢ Determination of degradation rates often based on literature values which: ▪ may not be appropriate to UK conditions (e.g. aquifer type, water temperature); ▪ may be for different conditions from those observed at a site (e.g. anaerobic conditions may occur at site, whereas the literature value may be for aerobic conditions); ▪ laboratory values which may not be applicable to field conditions. ➢ Determination of degradation rates from field observations usually reliant on having time-series of monitoring data covering a number of years (See EA R&D Publication 95 (EA, 2000)).

Probabilistic models allow parameter uncertainty to be taken into account in the analysis. However, they should not be viewed as an alternative to obtaining site-specific data, and they cannot compensate for the deficiencies of an incorrect conceptual model or an incorrect or inappropriate mathematical model.

The results of a probabilistic model can, however, be used to guide site investigations. For example, a fate and transport model may have been used to predict the possible extent of a contaminant plume. Boreholes could then be sited within this predicted plume to provide data that could help constrain the uncertainty in parameters describing transport of contaminants.

2.2 Lack of knowledge: uncertainty and variability

In theory, if the values of all the relevant hydrogeological parameters are known, and the contaminant release history is known, it is possible to calculate the expected concentration of a contaminant at any subsequent time and location of interest. In practice, such precise calculations are rarely possible because of lack of knowledge in key parameters. Two basic kinds of lack of knowledge can be distinguished: uncertainty in a parameter that clearly has a single value (e.g. the catastrophic failure of a storage tank is a discrete event, even though the precise date of failure may not be known), and variability in a parameter that is a function of location or time (e.g. the hydraulic conductivity of the material comprising an aquifer).

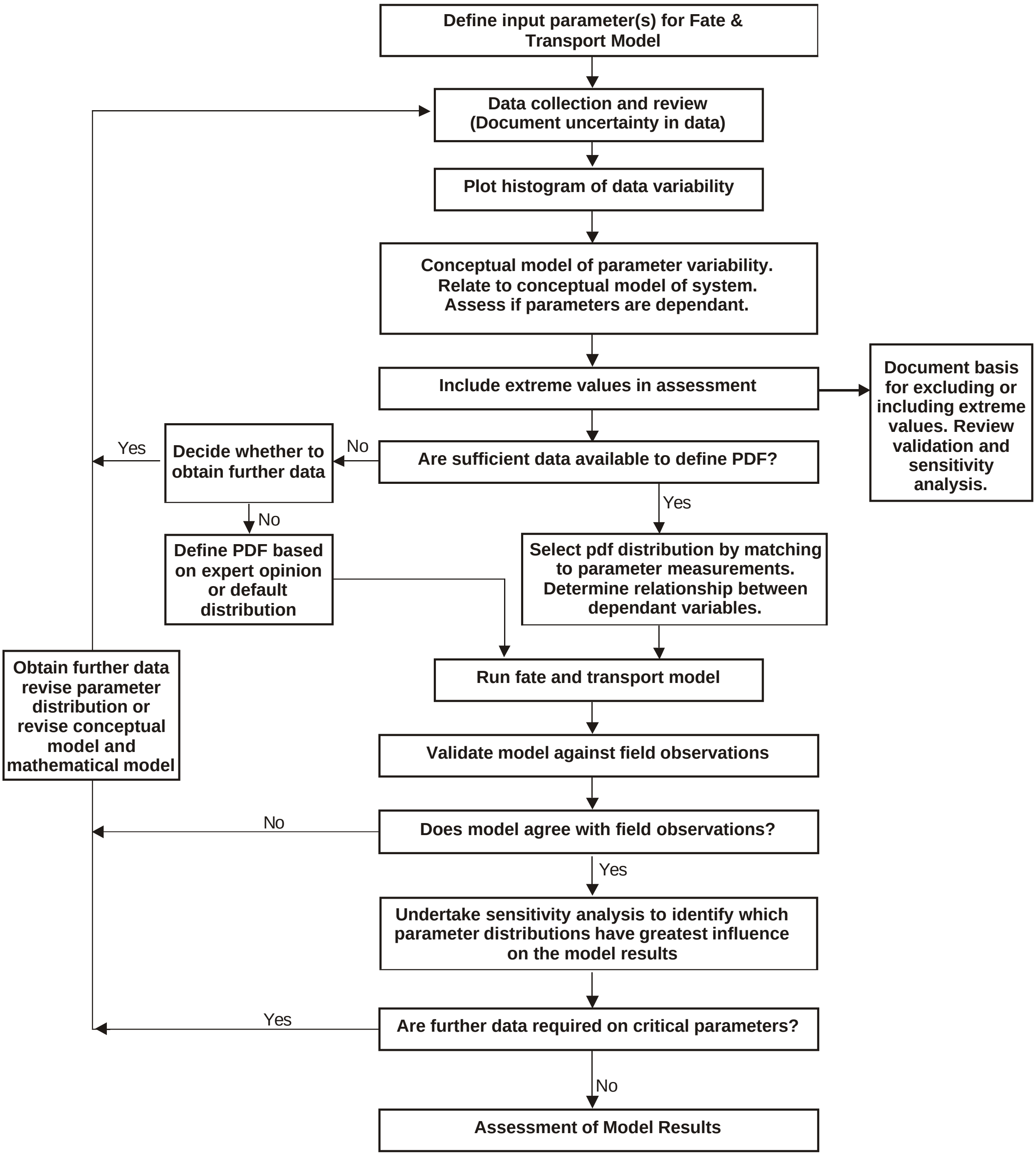
A parameter having a single, albeit uncertain, value can be described by a probability density function or PDF. The PDF describes how likely it is that the parameter has any particular value.

A parameter that varies in space, for example hydraulic conductivity, could be defined if sufficient data were available. However, this will seldom be the case. As a result our knowledge of the hydraulic conductivity at every point, other than at the point of measurement, is uncertain. This is what is known as a stochastic field. In practice, there will usually be some relationship between the hydraulic conductivities at adjacent points. It may be possible to derive a statistical description of the spatial variability of a parameter, for example using geostatistics. From this description, it is possible to generate realisations of the spatial distribution of a parameter value. The generation of these realisations is a very large subject, beyond the scope of this guidance (refer to Cressie, 1991 and Deutsh et al., 1992).

In practice for many models, stochastic fields will be simplified into single parameters. For example, rather than modelling the spatial variation of hydraulic conductivity between contaminant source and receptor, a single appropriate average value of hydraulic conductivity may be used. Because the detailed distribution of hydraulic conductivity is unknown, this overall average value is uncertain. An uncertain single value can be described by a PDF. There are statistical techniques that can be used to derive the PDF of the appropriate average, given measurements of point values. A mechanism for converting spatial variability into uncertainty in a single 'average' value is covered in Appendix B.

2.3 Procedure for defining and applying PDFs

The key stages in defining and using PDFs are shown in Figure 2.3 and summarised below. The chapters noted in brackets indicate the chapters in this document which expand on these stages.



Note: This flow diagram assumes that data are available to define parameter values

Figure 2.3 Main Stages in the Application and Use of Probability Density Functions

Data collection and review *(Chapter 4).* Assessment of the available data to determine parameter ranges. Extreme values should be included unless there is reason to suspect analytical problems or anomalous results. Where limited data are available, expert judgement should be used to determine minimum and maximum parameter values and the likely parameter distribution.

i) ***Understanding of patterns, trends and variation*** *(Chapter 4).* Before considering the PDF, the distribution of the parameters must be understood and any important features incorporated into the conceptual model

ii) ***Selection of appropriate PDFs*** *(Chapter 4).* Selection of an appropriate PDF (Uniform, Log-normal etc) to represent the observed parameter variation and the uncertainty in its definition.

iii) ***Run the fate and transport model*** *(Chapter 5).* A Monte Carlo simulation is the most commonly used technique. The model is repeatedly run using data sets chosen from the PDFs, until repeatability (refer to Section 5.1) in the model results has been achieved.

iv) ***Model refinement and validation (or testing the model against observed data)*** *(Chapter 7).* Model results should be compared with field observations to test the validity of the model. The model and the field observations should then be examined to determine if they are both still credible. In the case where the results are not credible, this may point to the inappropriate use of mathematical model (it does not represent the conceptual model), or the inadequacy of the conceptual model, In some cases it may be appropriate to revise the PDFs where, for example, extreme values had originally been included in the data set but which resulted in model predictions that were inconsistent with the field observations.

v) ***Sensitivity analyses*** *(Chapter 7).* Identification of the parameters whose uncertainty has the greatest influence on the model results. For sensitive parameters consideration should then be given as to whether further information should be obtained.

vi) ***Review of results*** *(Chapter 7).* Critical review of the model results to ensure that assumptions in constructing the parameter distributions and the model are valid. Dependent on this comparison then the following will be appropriate: use of the model as a predictive tool; revision of the conceptual model; reassessment of the data and the PDF used to define the model parameters; and/or the collection of further data.

The overall process in the development of a fate and transport model is detailed in Environment Agency (2001a). The development of the conceptual and mathematical model (including parameter values) is an iterative process, with each model being updated or revised as more data become available.

3. Definition and purpose of probability density functions

3.1 Introduction

This section describes and illustrates some of the principles of probability that underpin uncertainty theory. Readers who wish to become more familiar with probability theory should refer to a probability textbook which introduces the concepts in detail, such as Till (1974), Davis (1973) or Gilbert (1987).

3.2 Why use probability distributions?

Probability distributions are used in order to describe uncertainty and variability. When we carry out a calculation we select values for parameters even though these values may not be well-known. Some parameters can be measured fairly directly without the need for interpretation (e.g. the area of a piece of land), but even then there is uncertainty associated with how accurately it was measured (*experimental uncertainty*). More often, there is uncertainty associated with the fact that a parameter is continuously varying in space or varying in time.

The major sources of uncertainty in a calculation are:

- **Incorrect conceptual model**. This form of uncertainty occurs when our calculation is not correct because we have oversimplified or misrepresented the physics of the situation and arises from a fundamental misunderstanding or oversimplification of the system – an incorrect conceptual model. This form of uncertainty is not addressed by the approach presented in this document and needs to be evaluated if valid decisions are to be made following the review of the model results (see Environment Agency, 2001b for further guidance on conceptual models).

- **Intrinsic uncertainty.** A parameter may not be 'knowable' (for example, because it relates to the future - such as the number of holes that will occur in a landfill liner throughout the life of the landfill). A parameter that has been estimated or relies on expert judgement has intrinsic uncertainty.

- **Time variation**. This derives from the fact that some parameters vary in time (e.g. thickness of unsaturated zone due to water level variations) and, therefore, we will not know the value at any time where it has not been measured. This is a form of intrinsic uncertainty.

- **Experimental uncertainty**. This derives from the limitations in accuracy and precision of the measurement technique. Precision refers to the fact that the measurement can only be resolved to a certain extent (say to the nearest centimetre). A lack of accuracy means that the measured value differs from the true value (e.g. due to a wrongly calibrated tape measure). A measured parameter always has experimental uncertainty.

- **Heterogeneity (spatial variation)**. This derives from the fact that if a parameter (e.g. clay content) varies spatially, we will not know the precise value at any point where it has not been measured. What we may ultimately need to know is the 'net' or effective value of the parameter over the scale of interest– which depends on the *upscaling* method used (see

Section 4.4 and Appendix B). The uncertainty in the effective parameter is related to the variability in the sampled parameter.

3.3 Samples and populations

It is necessary to make a distinction between the *population*, that is all possible measurements (which may be an infinite number), and the *sample* of the population, which consists of the measurements that are available. The population is completely described by its *distribution* which can in turn be defined by parameters (e.g. population mean (μ) and variance (σ^2)). These parameters will generally be unknown as it will rarely be possible to make all the necessary measurements to define the population.

Statistics can be derived for the sample of measurements (e.g. sample mean and standard deviation). These statistics are not the same as the parameters (of distribution) of the population. For example, the sample mean is an estimator of the population mean (but it is not a unique value - if a different set of measurements or sample of the population were taken, the estimated sample mean would be different, albeit often only slightly).

Unfortunately, the terminology does not differentiate between the two, with words such as 'mean' used to signify both the average of the sample of measurements and the mean of the population (sometimes called the *expected value* of the distribution). The former is in fact a 'statistic' and is an 'estimator' of the population mean. To assist with clarity, Table 3.1 defines the properties of a probability distribution and also includes the statistic generally used to estimate them from a sample

Appendix A provides an example of the calculation of statistics for a data set.

Table 3.1 Parameters of probability distributions and their statistics

Parameter	Description	Parameters of population distribution	Statistic of sample measurements
Arithmetic mean	Commonly understood as ‘average’	$m = \int_{-\infty}^{\infty} xf(x)dx$	$\overline{X} = \frac{1}{n}\sum X_i$
Geometric mean	Inverse logarithm of the mean of the logarithms	$\exp\left[\int_{-\infty}^{\infty} \ln(x) f(x)dx\right]$	$\overline{X}$ = $^{n}Ö(X_1.X_2.X_3.X_4\ldots. X_n)$ or $\overline{X}$ = $\exp[\Sigma(\ln X_i)/n]$
Harmonic mean	Inverse of the mean of the inverses	$\left[\int_{-\infty}^{\infty} \frac{f(x)}{x} dx\right]^{-1}$	$\overline{X}$ = $n/\Sigma(1/X_i)$
Median	Value for which there is a 50% probability of it being exceeded	*M* such that $\int_{-\infty}^{M} f(x)dx = 0.5$	After ranking the data, the ½(n+1)th value. If there are an even number of measurements, then take the average of the ½n^{th} value and the ½(n+1)th value
Mode	Most likely value	*M* such that *f(M)>f(x)* for all *x* not equal to *M*	Group the data in intervals and the mode is the most populated interval. (Note: A population and sample may have more than one mode)
Variance	Indicates how spread out the distribution is.	$s^2 = \int_{-\infty}^{\infty} (x-m)^2 f(x)dx$	$V = \frac{1}{n-1}\sum (X_i - \overline{X})^2$
Standard Deviation	Square root of variance	s	SD = $\sqrt{V}$

Notes: In the variance, the (n-1) term is sometimes approximated to n, if n is very large. The probability density function is f(x). In the statistics column, it is assumed that the sample consists of n readings X_1, X_2, X_3,…, X_n. *The population consists of all possible measurements (this may be an infinite sample). It is assumed that all the distributions are continuous rather than discrete.*

3.4 Common probability distributions

This section explains and defines the most common probability distributions. Table 3.2 gives the mathematical expression of each of these distributions and Figure 3.1 shows a graphical representation. Figure 3.1 has been drawn so that each of the distributions presented have been derived using the same mean (10), standard deviation (3) and ensuring that the area under the curve is always 1. The exception was the exponential distribution for which the parameters of the distribution were set to give an area under the curve of 1.

Probability density functions require different numbers of parameters to be defined. For example, the Uniform distribution requires a minimum and maximum value to be defined, whereas the Triangular distribution requires a minimum, most likely and a maximum value. Thus the definition of a PDF is not necessarily the answer for all data shortage problems as an uncertain hydrogeological parameter value is described in terms of two or more uncertain statistical parameters. This may not represent an increase in knowledge!

3.4.1 NORMAL

The most common observed distribution is the Normal distribution (also called the Gaussian distribution). The Normal distribution implies a symmetrical grouping of the sample around a specific value (mean) with less chance of a sample further away from this value. The graph of the PDF is often referred to as the “bell-curve”. An example of a normally distributed parameter is the clay content of samples from a certain geological unit.

Probabilities from Normal distributions can be described as either a one-tailed or two-tailed distributions (Figure 3.2). A two-tailed probability is used when both high and low extreme values need to be taken account of (i.e. the range in values across the mean). A one-tailed probability is used when outcomes above or below a certain value are of concern.

It is useful to note that 68.3% of a Normal distribution occurs within one standard deviation on either side of the mean (two-tailed probability), 95.4% within two standard deviations and 99.7% within three standard deviations. For a one tailed distribution, the 95-percentile, that is the value which is greater than 95% of the population, will be at 1.64 standard deviations from the mean (closer than 2 standard deviations because we are looking at the tail on one side only). This is illustrated in Figure 3.2.

The Normal distribution function can theoretically take any value from minus infinity to plus infinity (i.e. it is *unbounded*). In principle, therefore, any parameter known always to be positive cannot be perfectly normal because there is zero probability that a negative number will occur. In practice, the Normal distribution is the best distribution for very many observations in nature and as long as the mean is more than three standard deviations away from zero (in which case the distribution produces a probability of a negative number of only 0.15%), it can be safely used.

In some software (e.g. ConSim, LandSim), a MODIFIED NORMAL distribution is defined which is a Normal distribution with the negative tail cut off and the PDF adjusted upwards by a factor everywhere else to recover the lost probability (i.e. so that the area under the curve is still 1). This distribution is useful because it is similar to the Normal distribution and cannot be negative. For example, the ConSim package constrains the Normal distribution between zero and infinity to prevent negative values being used in the analysis (otherwise defining a Normal distribution for porosity of 0.1 with a standard deviation of 0.2 would introduce negative values into the calculation).

The Normal distribution is symmetrical about its mean. If there is evidence that a distribution is skewed, then the Normal distribution should not be used.

The Normal distribution can usually be used to describe the variation in porosity measurements and clay content.

3.4.2 UNIFORM

The Uniform distribution distributes probability equally between two extreme values. Example: If monitoring occurs on the first of each month and there is no contamination on 1 January but there is contamination on the 1 February a Uniform distribution assumes that there is an equal probability that the contamination event happened on any of the intervening days.

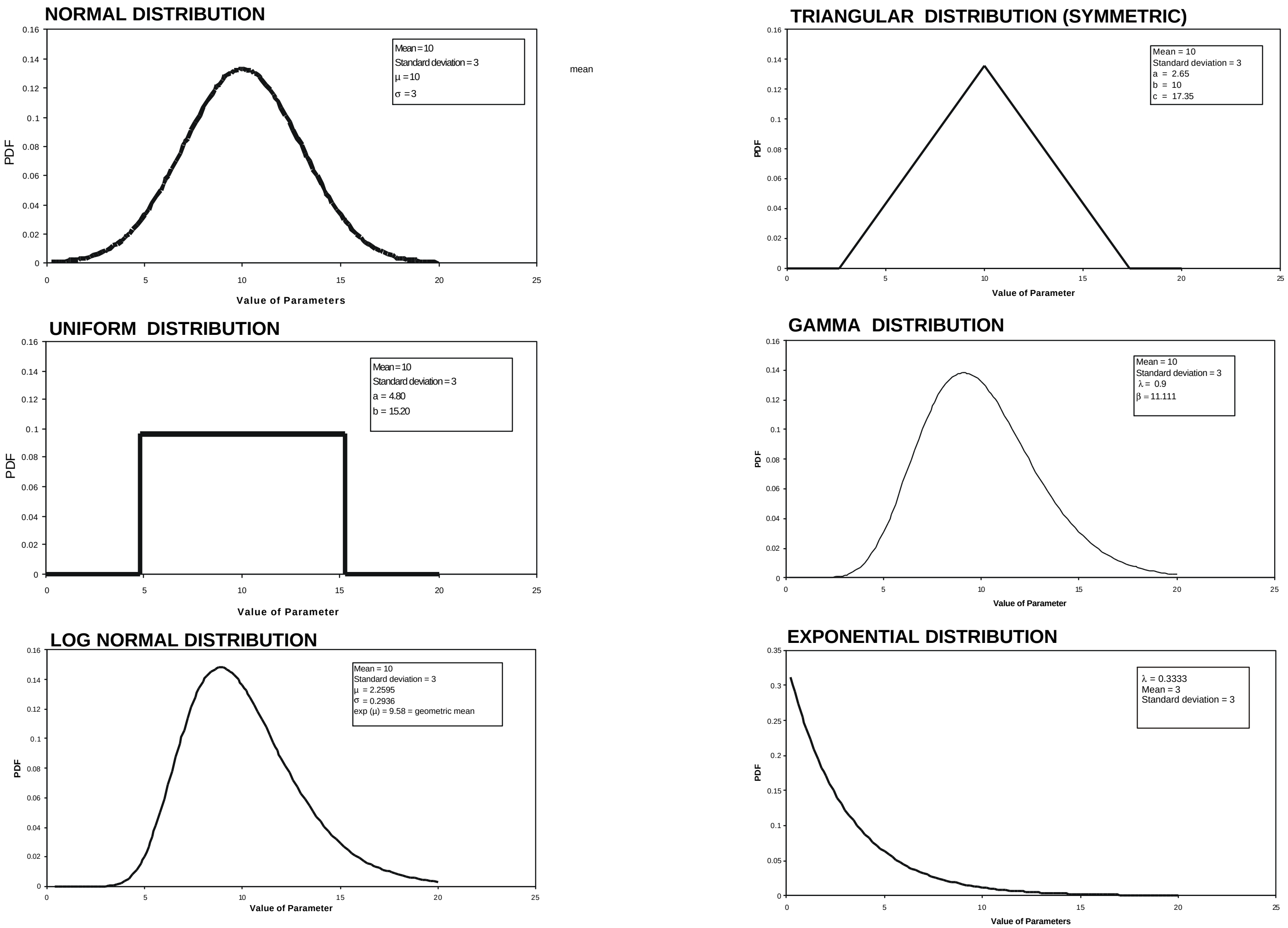


Figure 3.1 Graphical Representation of Common Probability Distributions

In selecting a minimum and maximum value for a Uniform or a Triangular PDF (see below) based on observed measurements, it is important to recognise that these are unlikely to describe the actual population range. For this reason, it may be appropriate to use expert opinion to define higher and lower values than those determined from observed measurements (although this would need to be justified). However, care needs to be exercised as the maximum and minimum measurements may be anomalous (due to measurement error) or inappropriate to the problem (i.e. a very low value of hydraulic conductivity may relate to a clay lens within a sand and gravel deposit, and where contaminant movement is via the coarser sand and gravel fraction).

Uniform distributions typically are used to describe variation in porosity and measured lengths (e.g. distances, water table fluctuations etc).

3.4.3 LOG-NORMAL

A distribution is Log-normal if the logarithms of the values are distributed normally. A common example is hydraulic conductivity, which has been observed to vary in the field by several orders of magnitude. The true average of a Log-normal distribution is the geometric mean (which is the exponential of the arithmetic mean of the natural logarithms). There are two common ways of mathematically expressing this distribution as a PDF (Palisade Corporation @Risk Model). In Table 3.2, the version in which μ is the arithmetic mean of the logarithms and σ is the standard deviation of the logarithms is presented (Till, 1967) and is used in this report. The arithmetic mean of the logarithms is also the geometric mean. A Log-normal distribution contains no negative values and is skewed.

It is important to check, when using software packages, which parameters of the distribution need to be entered. For example Excel™ requires the mean and standard deviation of the logarithms, whereas ConSim asks for the arithmetic mean and standard deviation of the raw data set. The Crystal Ball Package gives a number of options, the default is the arithmetic mean and standard deviation of the raw data, but the geometric mean and standard deviation can also be input.

The arithmetic mean and standard deviation of a log normally distributed data set are related to the arithmetic mean and standard deviation of the logarithms of the data by the following equations (Till, 1967).

$$m = \exp\left(m_n + \tfrac{1}{2}\sigma_n^2\right)$$

$$\sigma = m^2\left(\exp\left(\sigma_n^2\right) - 1\right)$$

where;

m_n = mean of natural log or geometric mean of the data set

σ_n = standard deviation of natural logs

m = arithmetic mean of the data set

σ = standard deviation of the data set.

If the data set does not follow a Log-normal distribution then these two equations do not hold (see Appendix A).

A Log distribution is recommended for distributions that span more than one order of magnitude and which appear to be skewed. It is not appropriate where distributions appear to be linear, i.e. no bias to one end or the other, in which case a Uniform distribution should be used.

The Log-normal distribution is typically used to describe the variation in hydraulic conductivity values.

Note on logarithms:

Logarithms may be taken using any number as a base, but are commonly made using either e or 10

$$e = 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} ... \cong 2.7183$$

logs to base e are written $\log_e$ or ln

logs to base 10 are written $\log_{10}$ or log

They are related by ln x = ln10 log x

i.e. $\ln_x$ = 2.3026 log x

Log-normal distributions can be calculated using either base, but they must be used consistently, bearing in mind the multiplier between them. There is a mathematical advantage in some circumstances in using ln.

3.4.4 TRIANGULAR

The Triangular distribution can be regarded as a simple approximation of the Normal distribution. It is called triangular because of the shape of the PDF. It has three parameters: a minimum value, a maximum value and a most likely value (mode). As for the Uniform distribution (Section 3.4.2), maximum and minimum values should be defined based on expert judgement rather than the observed range of measurements. By definition, values cannot be lower than the minimum or higher than the maximum. Most experts have a feeling for the range of a physical property in terms of a maximum credible value and minimum credible value rather than as a standard deviation or some other parameter of distribution. Understanding of these three parameters (maximum, minimum and mode) is fairly intuitive. Consequently for Uniform, Triangular and Log-triangular distributions the range of values are constrained compared to Normal and Log-normal distribution. For this reason, the Triangular distribution is useful for estimated and elicited distributions, where the variation (between lowest and highest values) is less than an order of magnitude.

3.4.5 LOG TRIANGULAR

A distribution is Log-triangular if the logarithms of all the measurements are distributed in a Triangular distribution. As with the Triangular distribution, this distribution is useful because the parameters are intuitive. The distribution is useful for estimated and elicited distributions, where the variation (between lowest and highest values) is greater than an order of magnitude.

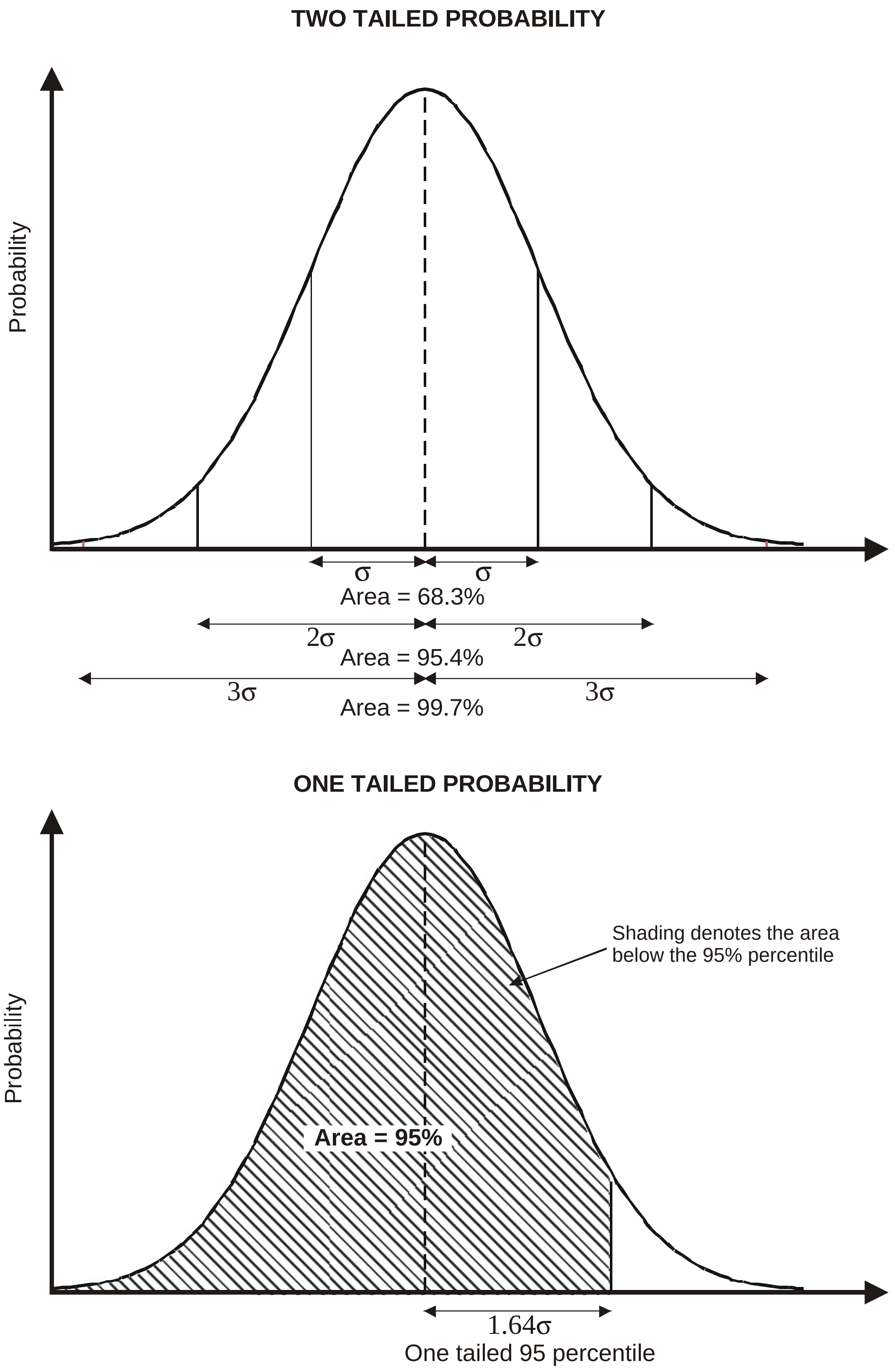


Figure 3.2 Some Properties of the Normal Distribution

02449-01.S007b

3.4.6 Other Distributions

There are a number of other less commonly used distributions including Exponential, Gamma, Binomial, Poisson, and Log-uniform. These distributions are generally less appropriate for describing aquifer or contaminant properties (there is typically insufficient data to justify a close fit of data to those distributions), although the Exponential, Binomial and Poisson distribution can be used to describe events, such as the failure of a pipeline or a liner which may result in the release of a contaminant. These other distributions are described below.

EXPONENTIAL

This distribution is always positive but has zero as its mode. It is commonly used to describe the time between events that occur randomly, but at a steady long-term expected number of events per time period. The mean and standard deviation of this distribution are identical.

GAMMA

This distribution is always positive but is skewed (not symmetrical about its mean). It is often used to represent contaminant concentrations or the time between occurrence of events. It represents the expected number of arrivals in a given time period. It is in fact the sum of *n* Exponential distributions.

BINOMIAL

The binomial distribution describes the likelihood of a given event occurring in a fixed number of trials, usually in the context that it occurs or does not occur. This distribution can only take discrete values (such as 0, 1, 2 etc), and not fractions. An example of its use is in LandSim, where it is used to define probability of the failure, or not, of leachate drainage pipes.

POISSON

The poisson distribution describes the number of times an event occurs in a given interval (e.g. number of fractures per m^3 of rock). This distribution can only take discrete values (such as 0, 1, 2 etc), and cannot be a fraction.

LOG-UNIFORM

The distribution is described by the log of the minimum and maximum values. It is used where these values range by more than an order of magnitude, i.e. the range is very poorly defined. In the vast majority of cases the use of such a distribution points to the need for more information rather than further analysis.

Table 3.2 Properties of the common probability distributions

Distribution	PDF	Mean	Variance	Mode	Median	Parameters
Uniform	0 $x<a$ $1/(b-a)$ $a<x<b$ 0 $b<x$	$(a+b)/2$	$(b-a)^2/12$	None	$(a+b)/2$	*a* (minimum), *b* (maximum)
Normal	$\frac{1}{s\sqrt{2p}}\exp\left(\frac{-(x-m)^2}{2s^2}\right)$	μ	s^2	m	m	m, s
Log-normal	$\frac{1}{xs\sqrt{2p}}\exp\left(\frac{-(\ln x-m)^2}{2s^2}\right)$	$\exp(m+½s^2)$	$\exp(2m+s^2).(\exp(s^2)-1)$	$\exp(m-s^2)$	$\exp(m)$	m, s
Triangular	0 $x<a$ $2(x-a)/(b-a)(c-a)$ $a<x<b$ $2(c-x)/(c-b)(c-a)$ $b<x<c$ 0 $x>c$	$(a+b+c)/3$	$(a^2+b^2+c^2-ab-ac-bc)/18$	b	$a+\sqrt{[½(c-a)(b-a)]}$ $b>(a+c)/2$ $c-\sqrt{[½(c-a)(c-b)]}$ $b<(a+c)/2$	*a* (minimum), *b* (most likely), *c* (maximum)
Exponential	$x<0$ 0 $x>0$ $\exp(-x/\lambda)/\lambda$	l	l^2	0	$l \ln 2$	λ
Gamma	$x<0$ 0 $x>0$ $\frac{x^{b-1}}{l^b\Gamma(b)}\exp(-x/l)$	lb	l^2b	$l(b-1)$ $b>1$ 0 $b<1$		l, b

Note: $\Gamma(x)$ is the gamma function. If x is a positive integer it takes the value x! or x-factorial.

β is shape factor

λ is rate

where a, b and c are constants

The above distributions are illustrated in Figure 3.1.

3.5 Uncertainty analysis

When several different distributions are combined in a calculation, the resulting distribution has its own characteristics and may not even be of the same type as any of the constituent distributions. To find the probability of the output parameter exceeding a certain value, it is necessary to search throughout the parameter space (all the possible parameter values) for the sub-space where the combination takes this value or larger and then find the probabilities of each of the parameters falling in this sub-space. Unfortunately the mathematics of determining such distributions are extremely complex and even the multiplication of two Normal distributions becomes very complicated.

Routinely when we attempt to calculate an impact we use scoping values in our equations. Sometimes they may be best-case estimates and sometimes they may be worst-case estimates, both of which are examples of combinations of extreme values of parameters. One problem with combinations of extreme estimates is that these may be far more unlikely than we appreciate. This is because the chance of all the parameters taking their extreme values simultaneously is far more remote than the chance of any one of them taking its extreme value individually. This is considered further in Section 5.2.

The realisation of this fact is one of the reasons for using uncertainty analysis and trying to establish exactly how much of a worst case we are talking about. Is it 1 in a million chance, 1 in a billion, 1 in 10? These are all unlikely but there is significant difference between them. The example in Box 3.1 shows the result of combining two Uniform distributions. We use this example to illustrate that "worst-case" analysis is sometimes far more conservative than it seems. Whilst a worst-case analysis may be justified under the precautionary approach, decisions based solely on such results may have financial implications disproportionate to the actual level of risk.

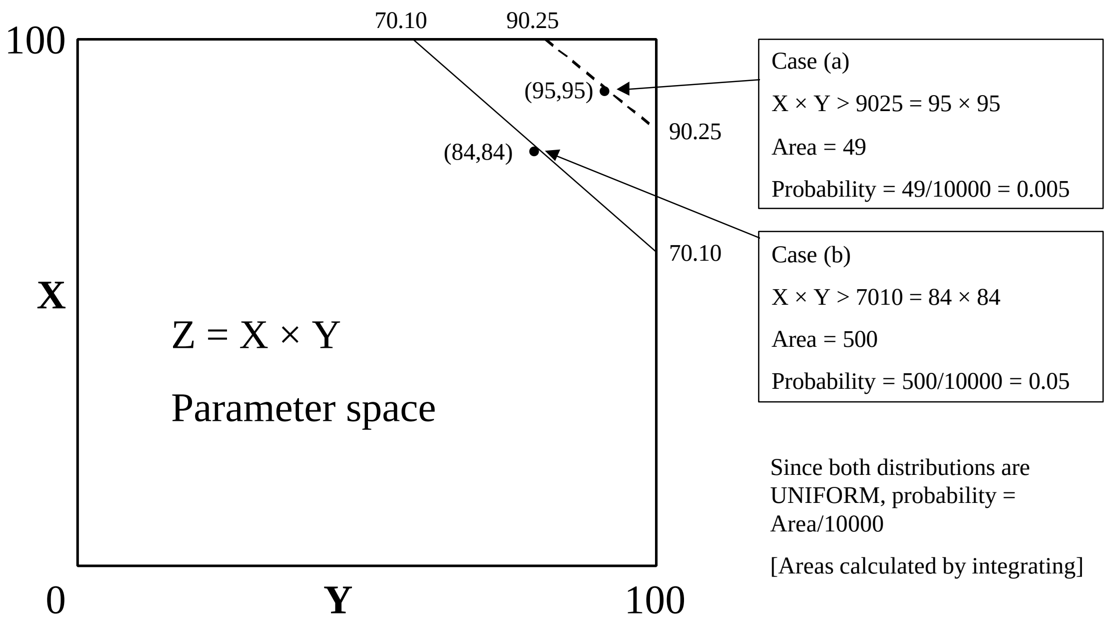
Box 3.1. How conservative is worst case?

Suppose we have two Uniform distributions independently distributed between 0 and 100 (imagine picking blindly a number from 1 to 100 twice and multiplying them). The 95-percentile at the upper range of each distribution is therefore 95. In other words, there is only a 5% chance of choosing a number larger than this. What is the 95-percentile of the product of the two?

Using the 'worst-case' approach, we would simply multiply the two 95-percentiles and get 9025. The chances of actually getting a product as large as this are surprisingly small (0.49% in fact). This can be demonstrated by determining the number of pairs of parameters that give a product ≥9025, as illustrated by case (a) Figure 3.3. The true probability of this being exceeded is 0.0049 (0.49%), i.e. the value of 9025 is the 99.51%ile. Figure 3.3 illustrates how this was calculated, by calculating the area of the parameter-space where the value is exceeded.

By comparison the 95-percentile of the product of two Uniform distributions ranging from 1 to 100 is 7010 as illustrated by case (b), Figure 3.3, i.e. the product of two 95%iles is more conservative than the 95%ile of the product.

This example serves to illustrate the importance of understanding the effect of combining "worst case" parameters i.e. combining two 95% values gives a result equivalent to the 99.51%.



Conclusion:

Case (a) $X_{95\%} \times Y_{95\%} = Z_{99.5\%}$

Case (b) $X_{84\%} \times Y_{84\%} = Z_{95\%}$

Figure 3.3 The product of two uniform distributions

Monte-Carlo analysis

There are methods of calculating approximately what the resultant means and variances would be from combining PDFs (they involve Taylor series expansions about the mean). However the power of modern computers has given us a much simpler approach. Monte-Carlo analysis does not require difficult mathematics and can deal with complex combinations of PDFs including calculations with conditional statements (i.e. if-statements) such as those involved in fate and transport calculations.

Monte-Carlo analysis involves generating sets of values for the input parameters, by picking them from their respective distributions. Each set of input parameters is then used to calculate a realisation of the output result. If this sampling is carried out thousands of times, the distribution of results tends to approach the theoretical distribution that was difficult to derive mathematically. Section 5.1 describes Monte-Carlo analysis in more detail.

The histogram of the results is the main output of this technique and represents an experimentally-derived estimate of the true PDF of the calculated output. The Monte-Carlo result will get closer and closer to the mathematical answer, as more and more realisations are carried out. It is a sort of experiment and, as such, produces an approximation to the answer. Each time a Monte-Carlo analysis is carried out the derived histogram will be different, but with a large enough number of realisations, an acceptable level of accuracy and repeatability can be obtained.

3.6 Dependent parameters

One limitation of Monte-Carlo analysis is that the sampling generally assumes all the parameters are independent of one another. Unfortunately this is not always true.

For example, the hydraulic gradient is not independent of the hydraulic conductivity of an aquifer since there is a negative correlation between the two. For example, shallow hydraulic gradients are typically associated with higher values of hydraulic conductivity and steeper gradients by lower values of hydraulic conductivity.

Some probabilistic models (e.g. LandSim and ConSim) do not allow dependency to be taken into account; the expectation is that expert users will input reasonable values given their knowledge of the hydrogeology. However, it should be recognised that the Monte Carlo method *can and will* pick extreme values from dependent ranges that would not be expected to occur in nature e.g. high hydraulic conductivity with low hydraulic gradient.

Some modelling packages allow adjustments to the Monte-Carlo method to be made to take this sort of linear dependence into account.

Linear dependence means that the two parameters are related by the equation:

$$Y = aX + b + E \qquad (1)$$

where X and Y are the parameters, *a* and *b* are constants and E is a normally distributed error with mean equal to 0. The smaller E is, the more tightly the two parameters are related.

To take dependence into account in a calculation, both *a* and *b* need to be estimated (refer to Box 3.2) and an indicator of how closely the two parameters are related (which is determined by the error E). The statistic required by these packages is the *correlation coefficient* (see Box 3.2) between the two parameters. The correlation ranges from -1 (negatively correlated) to +1 (positively correlated) with 0 indicating no correlation. The constants, *a* and *b*, are determined from the PDFs already entered for the parameters. An alternative is to define the second variable in terms of the first using an expression comparable with Equation 1 above.

The correlation coefficient can be estimated from a large number of data-pairs of the two parameters using the statistics given in Box 3.2. In practice, the data-pairs can be entered to the two software packages above and they will calculate all the parameters. However, the definitions of the common parameters and their relationship to the much-quoted *R-squared* statistic, are provided below.

In packages that do not specifically address parameter dependency, a straight forward solution is to define a PDF for the parameter with the greatest uncertainty, and to use a single value for the other dependent parameter. So, for example, hydraulic conductivity (which varies greatly) may be described using a Log-normal PDF, while hydraulic gradient (which is generally less variable) is described by a single (most likely) value.

Box 3.2 Definition of dependency statistics

Correlation coefficient (r) between two parameters x and y is defined as:

$$r = C/(SD_x SD_y)$$

where C is the *covariance* defined as $C = \Sigma(x_i - \bar{x})(y_i - \bar{y})/(n-1)$,

$\bar{x} = \Sigma x_i/n$, and

$\bar{y} = \Sigma y_i/n$.

n = number of samples

SD = standard deviation of sample measurements

R-squared is a related parameter which statistical packages usually report after regressing the data. Data regression is the calculation of the constants *a* and *b* by minimising the squares of the residual values. *R*-squared represents the proportion of variation in X explained by the model.

$$R^2 = 1 - \Sigma(y_i - a'x_i - b')^2/((n-1)SD_y^2)$$

where a' is the estimator for *a*, defined by $a' = \Sigma(x_i - \bar{x})(y_i - \bar{y})/ \Sigma(x_i - \bar{x})(x_i - \bar{x})$

and b' is the estimator for *b*, defined by $b' = \bar{y} - a'\bar{x}$

4. Choosing a PDF and data requirements

4.1 Procedure for defining the uncertainty in a parameter value

In most modelling and risk assessment scenarios the possible range of values for some parameters will not be known. There may be very limited or, in some cases, no site-specific data with which to define some parameter ranges and a method of selecting appropriate values and PDFs must be found. If there is no site-specific information on any of the parameters then modelling is inappropriate.

Certain parameters can readily be measured in the field (e.g. hydraulic conductivity, hydraulic gradient), but others are generally not measured (e.g. dispersivity) and values must be obtained from other sources. Where site-specific data are available they may be limited and consideration must be given to the way in which they are used to generate PDFs.

This chapter provides guidance on the review and screening of data and on the various approaches to choosing an appropriate PDF.

4.2 Data requirements

It is not the purpose of this document to describe methods and techniques of acquiring field data, although some reference is made to the reliability or data quality of some methods. Published good practice guidelines or standards should be used where applicable and parameter testing should, where possible, be carried out by laboratories accredited for the relevant test(s). The process of collecting site-specific data, which will then be used to characterise a site and input into a quantitative risk assessment, is critical to the quality of the data and hence the quality of the model and results.

The first stage in the assessment is to determine which parameters are required for input to the fate and transport model. This will be dependent on the type of model selected. For each parameter either a single value (deterministic) or range of values (probability distribution) will need to be defined. The possible cases are:

i) **Sufficient** site-specific data are available;

ii) **Limited** site-specific data are available;

iii) **Insufficient** site data are available, but literature values are available from similar geological/hydrogeological regimes or experts are available who can estimate parameters for the site. It is stressed that care needs to be taken in using literature values and deciding whether they are appropriate. Literature values and expert opinion values should, in general, only be used for model input parameters that:

- have values quoted in the literature covering only a small range, e.g. bulk density;
- are relatively insensitive parameters in the analysis (e.g. variation of this parameter within its credible range does not significantly affect the model outcome);
- can be justified based on comparison or correlation with similar site(s).

iv) **Absent.** Data are very restricted or absent. In this case more site investigation is necessary before modelling, although literature values or expert opinion values help as part of an initial scoping exercise.

Before choosing a PDF the data should be reviewed critically in order to establish the following:

- **Are the data from the same (or a relevant) population?** For example, a large number of permeability tests may have been undertaken at a site underlain by the Millstone Grit – this is a layered aquifer, and each layer may be characterised by a different hydraulic conductivity and therefore it may not be appropriate to group all of the test results. Alternatively, it may be reasonable to combine results from a number of different locations to obtain a larger data set if it can be demonstrated that they are part of the same population (e.g. test results from different sites may be combined if it is shown they relate to the same geological horizon).

- **Are data time dependent?** Some parameters are variable in time and if any trends exist, they must be identified and incorporated into the conceptual model. For example, it does not make sense to calculate the mean of a series of steadily decreasing leachate concentration measurements.

- **Are extreme values representative of the system?** Do they describe an important part of the system or are they truly anomalous? In the *initial* definition of a PDF, extreme values should be retained unless the measurement can be justified to be anomalous. Comparison of model results with field observations may then provide a basis for including or excluding extreme values.

- **Are measurements representative of the system as a whole?** For example, values of hydraulic conductivity may be available from public water supply abstraction boreholes but these are likely to have been located in zones of high hydraulic conductivity and will not necessarily be representative of the aquifer as a whole. They will be more representative than values derived from laboratory testing of cored samples because of the sample size (i.e. aquifer scale compared to core scale). This is called *bias* in the sampling procedure and illustrates the importance of thinking about what is being measured and whether it is representative of the part of the system behaviour we are interested in.

- **Are data reasonable?** In some cases it may be appropriate to check the measurements with similar sites or established databases (the BGS/Agency Aquifer Properties Manuals provides a useful data source) to check if data are anomalous. Are the numbers physically possible? For example, values of 10^{-6} for storage coefficient are not feasible.

- **Should the data be scaled up?** Upscaling refers to deriving a value for a parameter at the scale of interest (i.e. regional value) based on measurements made at a smaller scale (i.e. laboratory measurements). The decision to upscale should be based on the conceptual understanding of the system behaviour, i.e. is it described by individual measurements or by the 'average' of the measurements. If the latter, then the 'average' should be calculated. The uncertainty in the 'average' will generally be less than the uncertainty in the individual measurements (see Section 4.4). This is illustrated by Figure 4.1 where it would be appropriate to upscale the measurements of hydraulic conductivity for model 1, but inappropriate for model 2. This figure illustrates the importance of thinking about the geology of the deposit as this may give some idea about the geometry of different lithologies, i.e. whether beds may be continuous or discontinuous.

Cautionary note on upscaling: laboratory scale measurements may not be representative at field scale. Even pumping test results may not be representative of regional aquifer values.

Where sufficient data are available, graphical techniques (including histograms, cross-plots, cumulative frequency plots) should be used to develop a qualitative understanding of the data (i.e. presence of outliers) and the type of distribution that may be appropriate (see Section 4.3.2).

When presenting parameter distributions it is important that there is an audit trail and that the process of defining PDFs is documented. There must be precise definition of the parameter whose distribution is determined, and the data used to support the definition. It is recommended that the following information should be provided:

- raw data – the source and nature of the measurements should be described and the data should be presented in a table and, preferably, in graphical format including the use of histograms or frequency plots;
- results of any statistical analysis of the data (mean, standard deviation etc);
- the basis for excluding/including any extreme values;
- information on how measurements were undertaken, including any uncertainty in the measurements.

4.3 How to select an appropriate PDF

The best approach to selecting parameter values and PDFs will vary depending on the amount of site-specific data available for each parameter. Three scenarios are considered in this section. The numbers of data points given are for guidelines only.

i) **Sufficient data** are available to warrant statistical analysis to determine a PDF (more than 30 points), Section 4.3.2;

ii) **Limited data** are available (roughly 5-30 data points) and you may have to assume the type of PDF, but the parameters can be estimated from the data, Section 4.3.3;

iii) **Insufficient data.** There are no or very limited data (4 or less) and you will have to incorporate elicitation to determine the type of PDF and its parameters, Section 4.3.4.

The approach to defining a PDF must ensure that it is consistent with the assumptions incorporated into the conceptual model of the system. This is discussed in the first section below. It will also depend on whether the hydrogeological parameter described by the data is upscaled. Is it a single uncertain value or is it a stochastic field, for which an appropriate uncertain average is needed? This point is discussed in detail in Section 4.4.

Finally, it should be stressed that the definition of a probability distribution and its parameters must incorporate clear thinking. For example, when did a petrol tank leak if it was full in March and empty in June when next inspected? A Uniform distribution between March and June reflects the fact that the leak could have occurred at any time between these dates with equal probability.

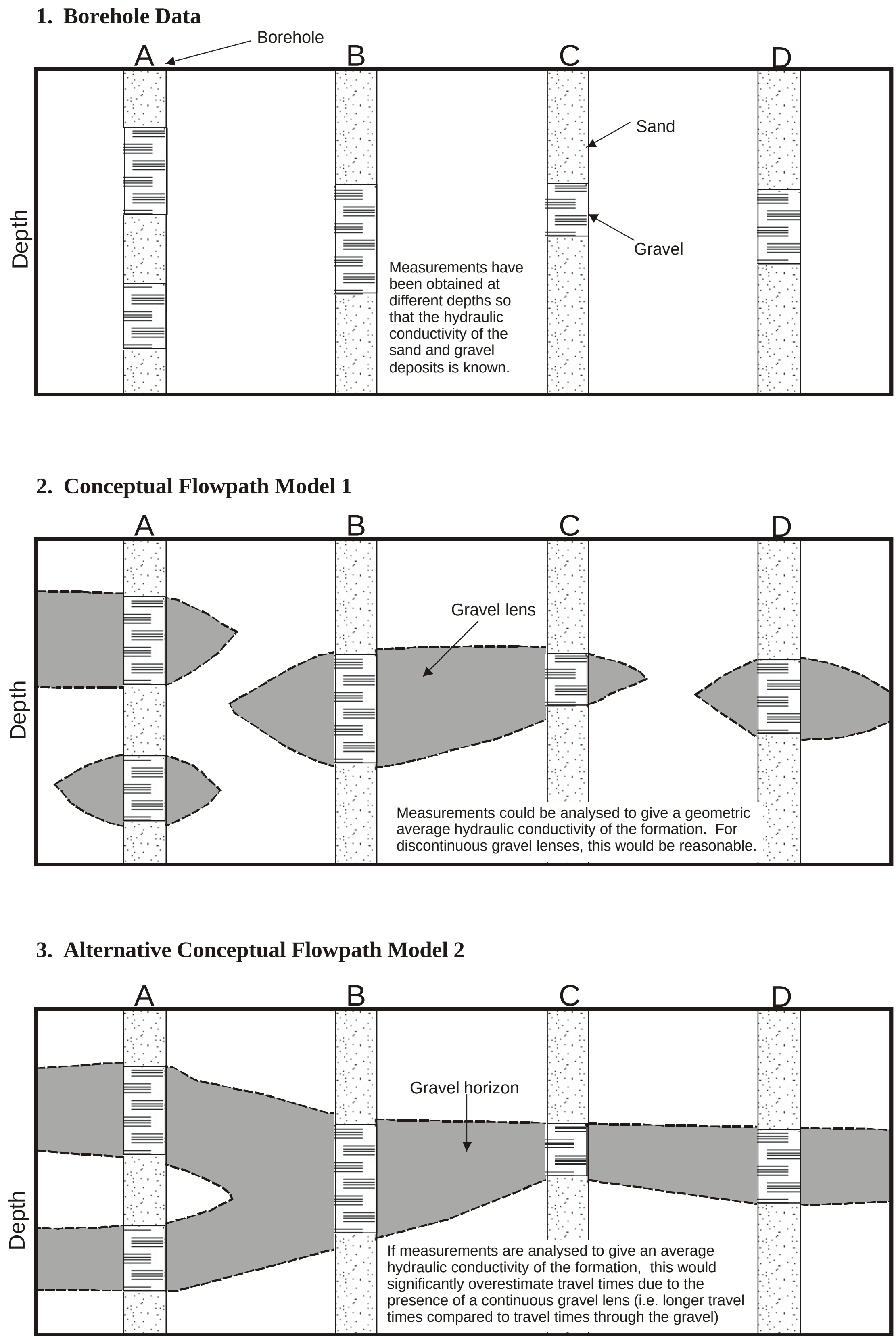


Figure 4.1 Scaling up and Uncertainty

4.3.1 Patterns and trends

Before any statistical analysis is carried out, the data need to be carefully analysed for patterns and trends. The term 'pattern' is used here to refer to spatial distribution and 'trend' to time-dependent data. If a trend or pattern is identified then this should be incorporated into the conceptual model. Only after such patterns and trends have been identified, and the decision made as to whether the data refer to a single population, should we proceed with the derivation of a PDF.

An example of each of these is presented in the boxes below.

Box 4.1 Example of data trends

Time-dependent measurements of chloride concentration in leachate have been taken for a period of five years from a single sump. A histogram of the data could be plotted and might show a roughly Uniform distribution with a mean of 1000 mg/l ranging from 1500 mg/l to 500 mg/l. However, when the data are plotted as a graph against time, it is clear that the strength was 1500 mg/l five years ago and declines at about 200 mg/l per year.

It would be wrong to categorise this parameter as uniform. The observed decline in leachate strength with time must be incorporated into the conceptual model. This could be done as a linear or exponential decline in strength. Uncertainty is included by inspection of the scatter on the plotted graph and including the uncertainty in defining the two new parameters (starting strength and decline rate) as PDFs.

Box 4.2 Example of pattern in data (see Figure 4.1)

Four boreholes have been drilled (Figure 4.1). Slug tests conducted on each of the boreholes have given three different values of hydraulic conductivity. The geometric mean is 10^{-8} m/s. However, the geologist has noticed that the geology consists of alternating sand and clay layers and that the conductivities measured seem to depend on the proportion of the two materials. He also knows from laboratory tests that the clay generally has a hydraulic conductivity of approximately 10^{-9} m/s and the sand an hydraulic conductivity of approximately 10^{-7} m/s. He recognises that transport depends on whether the sand lenses are continuous (Model 2) or not (Model 1). If they are, then the effective conductivity of the entire horizontal flow path would be approximately 10^{-7} m/s and if they are not, it would be approximately 10^{-8} m/s (i.e. midway between the values of hydraulic conductivity for the sand and clay lithologies). After further research into the sedimentology of the area he decides that they are probably, but not certainly, continuous (i.e. Model 2).

Possible approaches include:

1. Assume the hydraulic conductivity of the formation is described by a Log-normal distribution with a mean of 10^{-7} m/s (sand) to determine if there is likely to be a breakthrough at the receptor, as this is likely to represent a worst case.
2. Check the assessment using a Log-normal distribution with a mean of 10^{-9} m/s (clay), as this may, for example, explain the absence of any observed breakthrough of a contaminant at a monitoring borehole.
3. An alternative approach would be to assume that a special two-valued PDF is appropriate which has a 80% probability of being Log-normal around 10^{-7} m/s (sand) and a 20% probability of being Log-normal around 10^{-9} m/s (clay). The modeller incorporates this

into the spreadsheet by defining the hydraulic conductivity as the product of a Log-normal PDF around 10^{-7} m/s and a conditional statement based on a Uniform distribution that takes the value 1 (sand) or 0.01 (clay) e.g. for Excel™, IF(UNIFORM(0,1)<0.2, 0.01,1)]. This approach introduces another uncertain parameter which needs to be estimated; the relative likelihood of the two distributions.

This example shows a number of different approaches that could be used. The important conclusion is to determine what the results of the analysis show, and is our uncertainty important to any decision making process, or is the degree of uncertainty unacceptable and therefore further data are needed?

This example also introduces the concept of upscaling. The important conductivity is not the distribution at any one point but the distribution of the average of the entire flowpath.

In general, if a pattern is discerned that shows that the data come from two or more separate populations, then the data should be divided up and a PDF then derived for each population according to the procedures described in Section 4.3.2.

4.3.2 PDF selection - sufficient data

The process of selecting a PDF is:

i) Check the characteristics of the data for patterns, trends, bias, time-dependence etc as described previously. This will usually involve constructing a graph against time, or plotting maps and sections. Incorporate the understanding gained into the conceptual model. Only proceed when the data-points come from a single population.

ii) Plot the data as a histogram, frequency plot, or cumulative frequency plot (e.g. use the Histogram function in Excel™). If the curve has more than one mode, this is an indicator that the samples may be from more than one population and (i) above should be revisited. However if there is no other indicator that identifies the two populations, then it may be appropriate to devise a bimodal PDF.

The preparation of probability plots (involving plotting the data in order according to its percentile) is a useful graphical method of identifying the PDF. If the data form a straight line on normal graph paper, then this is a reasonable justification for assuming normality without formal proof.

iii) If there are fewer than 30 data points, it will probably be necessary to assume the type of PDF as it may not be possible to prove the distribution with a high degree of statistical certainty. This will be the typical case for most site investigations such that the choice of a PDF will involve some expert judgement, however, guidance on some of the factors to be considered is given in Sections 4.3.3 and 4.3.4.

iv) Choose the distribution type most likely to fit the data from the shape of the histogram and knowledge of the property. Probability plots are also useful if the distribution is uncertain. The PDF chosen should not be more complex than necessary. With a small number of data points, an uneven histogram does not necessarily indicate a complicated PDF - the variability of the system and the small sample both have an influence. For example, if there are two peaks on the histogram for 5 measurements, this does not necessarily mean the PDF has two peaks. If there are still two peaks after 5000 measurements, then it probably has and you may wish to see if the data come from two separate populations. The key is always to think about what the data are telling us about the system, what we know and what we are uncertain about. Other information about the system should be incorporated in thinking about system behaviour, for example inspection of geological exposures may tell us something about the layering of different lithologies.

If the curve is *skewed* then Log or Gamma distributions may be most appropriate.

If the data range over more than one order of magnitude then a Log distribution function (e.g. Log-normal, Log-triangular) should be considered.

v) Estimate the parameters of the PDF from the data (Table 3.2, Table 4.1). Some commercial packages have the option to fit PDFs to data; alternatively, most spreadsheets include statistical analysis packages. The estimation of the PDF parameters should be made using the data. For example, the parameter μ, the population mean, in a Normal distribution is estimated by the sample mean (arithmetic) and the parameter σ^2 is estimated by the sample variance (see Table 4.1).

However, very occasionally, there may be a good reason to use a different *estimator* to determine parameters. For example, if there are outliers where the measurement accuracy is compromised, using the median to estimate the parameter μ eliminates the influence of such outliers (Gilbert, 1987).

In chemical data, dealing with NDs (non-detects) or LTs (less thans) is an important issue. The most common approaches are to set the LTs at half the detection limit, or to use the median as an estimator of the sample mean. The NDs or LTs should not be set as zero. In general these methods should be adequate but, if not, consideration should be given to an alternative testing method with a lower detection limit.

Table 4.1 Estimator statistics

Distribution	Population Parameter of Distribution	Recommended Estimator	Notes
Normal	*m*	Arithmetic mean of sample	Use median if suspect outliers or exclude outliers from data set
	s	Standard deviation of sample	Use difference between the 84% and 16% percentile values if suspect outliers, *s* = (84%ile-16%ile)/2
Log-normal	*m*	Natural logarithm of geometric mean or arithmetic mean of the logs of the data of sample	A number of software packages require the arithmetic mean and standard deviation of the data. See Section 3.4.3.
	s	Standard deviation of natural logarithms of the data of sample	

Note: "Log" in the table above means natural log, or log to the base e, or ln. Refer also to Section 3.4.1 and 3.4.3.

vi) Use a statistical test procedure to determine if the selected distribution fits the data, i.e. is it statistically valid? For example, if a Normal distribution is proposed then a statistical test for normality (Box 4.3) should be undertaken. Such procedures include Chi-squared, Kolmogorov-Smirnov, Shapiro-Wilks and D'Agostino (Gilbert, 1987). The details to follow for the Chi-squared test are given in Box 4.3 and a worked example is included in Appendix A. The Chi-squared test can be used to test the fit of any selected distribution. The Shapiro-Wilks and D'Agostine tests are better tests but are specific for Normal distributions (less effort and more reliable).

Box 4.3 Use of Chi-squared to investigate normality

One statistical test for whether a sample fits a certain distribution is the Chi-squared procedure/test (p70-73, Till, 1974). This method can be used for any distribution but the example here checks the data against the Normal distribution (an example is given in Appendix A):

1) Calculate mean ($\bar{x}$) and variance $(SD)^2$ of sample measurements (x) (these will be estimators of the population).

2) Form a histogram of $z = (x - \bar{x})/SD$ with equal intervals. The size of interval and the end intervals should be selected to ensure that there are at least 5 values in each interval. This histogram should be normal with mean equal to zero and variance of 1. z is a standardised value which expresses the deviation of each x value about the mean in units of standard deviation.

3) Calculate the expected frequency (EF) for each interval if the distribution were normal with mean zero and variance 1, where:

   Where EF = probability of the z-value occurring in that class interval × total number of readings.

   The probability can be derived using look up tables that define the area under a normal curve mean zero and variance 1 (alternatively the Excel™ function NORMSDIST can be used).

4) Calculate χ^2 (the Chi-squared statistic) using

$$\chi^2 = \Sigma\,(OF-EF)^2/EF$$

   where the sum is over all the intervals, OF is the observed frequency and EF is the expected frequency.

5) Look up the Chi-squared function (from a table, or alternatively use the Excel™ function CHIINV(level of confidence, m-3) with m-3 degrees of freedom at the desired level of confidence, e.g. 0.05 (5% confidence); where m is the number of intervals.

6) If the χ^2 statistic is larger than the Chi-squared function then reject the hypothesis of normality at 5% confidence. There is a less than 5% chance of this sample representing a Normal distribution.

This method can also be applied to other distributions in a similar way.

If the test (e.g. Chi-squared) fails, then try other possible distributions until it succeeds. If none of the distributions fits well, then the alternative is to define your own distribution based on the data (which is possible in most software packages).

An alternative method for testing whether the measurements fit a distribution is the Kolmogorov-Smirnov test. There is a good account of this in Vose (1996).

Most modern software packages favour the Shapiro-Wilks test to test for normality or log-normality. This test is described briefly in Box 4.4; and more detail is provided by Gilbert (1987). A worked example is included in Appendix A. For data sets that comprise more than 50 measurements then the D'Agostino test is the recommended method (Gilbert, 1987).

vii) Run the model.

viii) Reality Check. The results of the modelling must be examined critically against any field data that exist (refer to Chapter 7).

Box 4.4 Use of Shapiro-Wilks to prove normality

This method tests specifically for normality and is described in Gilbert, 1987 (p159). An example of this method is illustrated in Appendix A.

1) Calculate $k = n/2$ if n is even or $(n-1)/2$ if n is odd (where n = number of samples);

2) Look up in a table (Table A6, Gilbert, 1987) the coefficients $a_1, a_2, \ldots a_k$ for the relevant n;

3) Order the data so that $x_1 < x_2 < \ldots < x_n$

4) Calculate W (the Shapiro-Wilks statistic) according to:

$$W = \frac{\left[\sum_{i=1}^{k} a_i (x_{n-i+1} - x_i)\right]^2}{\sum_{i=1}^{n} (x_i - \overline{x})^2}$$

5) Look up the Shapiro-Wilks function from a table (Table A7, Gilbert, 1987) at the desired level of confidence, e.g. 0.05 (5% confidence);

6) If the W statistic is less than the Shapiro-Wilks function then reject the hypothesis of normality at 5% confidence.

4.3.3 PDF Selection - Limited Data

When there are only few site-specific data some assumptions must be made about the likely distribution of the parameter values whilst still taking into account the measured values. Experience suggests that certain parameters commonly exhibit certain types of distributions and that there are other preferred distributions based on the number of data points available.

Some recommendations are listed below:

i) The recommended default distribution for hydraulic conductivity is Log-normal (BGS/ Environment Agency, 1997). In all cases where data spans an order of magnitude a log-based distribution should be used. If sufficient data are available to define that distribution, a PDF other than Log-normal may be more applicable.

ii) Distances are often best estimated as uniform ranging between the maximum and minimum plausible values. For example, in the case of distance from a landfill to a compliance borehole, the PDF for the distance should be uniform between the distance from the front and back of the landfill site to the borehole.

iii) Normal distributions are recommended for other parameters (e.g. porosity, infiltration), if there are more than 10 readings where the histogram suggests normality. The distribution should be checked to ensure that negative values are not assigned to a parameter which cannot be negative e.g. porosity. Also the mean value should be at least three standard deviations from zero. Some software packages (e.g. ConSim and LandSim) will automatically set the minimum value as zero, if the specified distribution results in negative values; it should be noted that this will skew the distribution slightly.

iv) Triangular or Log-triangular distributions are generally recommended if there are fewer than 10 readings since they are intuitive to estimate. Table 2.1 gives suggested distributions for different parameters. However, they can be over constrained - it is unlikely that a set of 10 measurements will include either the lowest or highest real measurement. There are a number of alternative methods that can be used where limited information are available, such as Maximum Entropy (Kapur, 1989).

v) For parameters where there is greater than one order of magnitude variation, then a logarithmic distribution should be considered.

Table 2.1 provides examples of PDFs used for different parameters. However, good practice should be to test that the selected PDF fits the data (Section 4.3.4 point VI). Figure 3.1 shows the shape of different PDF's and a useful check is to see if the selected PDF looks right when compared to the data.

The estimation of the PDF parameters should be made using site-specific data wherever possible. For example, the parameter μ in a Normal distribution is estimated by the sample mean and the parameter σ^2 is estimated by the sample variance. Similarly, for the Log-normal distribution, the parameter μ is estimated by the geometric sample mean (i.e. $\log_e(\mu)$ is estimated by the mean of the logarithms) and the parameter $\log_e(\sigma^2)$ is estimated by the variance of the logarithms of the sample data (Table 4.1).

4.3.4 PDF selection - insufficient data

Where there are few or no site-specific data, some method of defining PDFs is required if modelling is to be undertaken. This approach should be limited to initial assessments or where no data exist for <u>some</u> of the parameters. If analysis shows that the missing data are critical to the decision-making process, then more data must be obtained, i.e. the process should be iterative. **It is emphasised that the best response to this situation is to obtain more data**.

The most common methods of dealing with a lack of data (where it is not possible to obtain more) are to use expert opinion (elicitation) or values derived from literature sources.

i) Elicitation of Expert Opinion

A potential error is to define a range of parameter values that is too narrow. The likelihood of doing this can be minimised by starting with the definition of maximum and minimum values, rather than starting with the most likely. When estimating a parameter distribution, it is also more sensible to define minimum and maximum values, owing to the difficulty in estimating the standard deviation of a population, which is not intuitive to most people. Maxima and minima should always be justified with reasons why they cannot be exceeded.

'Expert' opinion can be used to define PDFs, but a common problem is that 'experts' can give very different distributions for the same parameter. It is also essential that expert opinion should be used to determine what is not known about a parameter. Care should be taken for PDFs derived by expert opinion, as this will only represent an educated 'guess', and should not be viewed with certainty. A number of techniques have been developed to overcome this including the 'Delphi' method, originally developed by the Rand Corporation (Dalkey & Helmer, 1963). The procedure follows the lines of:

1. Judgements along with their rationales are obtained independently from each member of an expert panel.
2. The summarised results are fed back to those experts in a manner that is carefully controlled so as to eliminate the pressures associated with group meetings.
3. The process is then iterated and a convergence of opinion occurs under most circumstances.

It is normally found that initial (individual) parameter ranges are unrealistically narrow and only partly overlapping. During the iteration procedure, the experts tend to broaden their uncertainty range and revise their best estimate towards the panel mean to an extent dependent on their confidence in their knowledge.

The Delphi method forms part of a broader probability assessment process developed by the Decision Analysis Department of SRI International (Speltzler et al, 1975, Stael von Holstein & Matheson 1979, Merkhover 1987).

Alternatively, a consensus approach may be used, taking into account the views and experience of the Environment Agency as well as others.

The tendency to make overly precise estimates can be reduced by training the expert panel in concepts of degree of belief.

These procedures have been extensively used in relation to the disposal of radioactive waste (e.g. Davis *et al*,. 1983). **It is important that the thinking behind elicited parameter distribution is documented.**

ii) Literature sources

Parameter values can be obtained from the literature if site-specific data are not available or need to be supplemented. Text books and journals may provide case studies with examples of parameter values for specific locations but these are unlikely to be directly relevant to the site for which data are required. Other literature sources include compilations of data from a number of references and may provide parameter ranges that are more statistically valid. The BGS/Agency Major and Minor Aquifer Properties Manuals, for example, provides a significant data base of measurements of the aquifer properties (hydraulic conductivity, porosity) for aquifers in England and Wales. **If literature values are used they should be relevant to site conditions and justified in the relevant reports/documentation.** If subsequent analysis indicates that these values are critical to the assessment then site -specific data should be obtained.

Caution is required when using literature values and the relevance and validity of the data should be checked. The data must be appropriate to the scenario being modelled.

Two approaches that can be applied where there are limited data are the Bayesian method and the principle of Maximum Entropy. The Bayesian method (Appendix C) permits re-assessing the PDF as a result of the acquisition of new data. For example, the expected range of hydraulic conductivity of a material (derived from the literature) can be combined with a very small number of site-specific measurements. (If all the original data are available, they can be combined in the usual way, but this is often not possible). The principle of Maximum Entropy (Kapur, 1989) is a theoretical method of optimising the form of the PDF assuming only the facts known (such as the mean and the fact that the parameter is positive).

4.3.5 Reality check

The results from a probabilistic model should be compared with field observations, wherever possible, to determine if the model results are plausible (see Chapter 7). For example, a model may have predicted no impact at the 95-percentile, whereas field data may show clear evidence of an impact. In this case, the model needs to be re-examined. This comparison will be subjective, as the model results will be described by a range of values or outcomes and these are being compared with single measurements. For sites which are well defined by monitoring data, then consideration should be given to a deterministic approach, as by modifying the input parameters to obtain a best fit to the observed data, the range of parameter values can be constrained (refer to Environment Agency, 2001b).

4.4 Relationship between heterogeneity and uncertainty (upscaling)

For some contaminant problems, the ‘average’ parameter value may best describe the system behaviour. In this case, measurements of a parameter at one scale (e.g. laboratory measurements) can be used to define the parameter at a larger scale. This approach of using sample measurements to define the ‘average’ system behaviour is described as upscaling. Where the system is believed to be heterogeneous, then upscaling should be used with care.

For example, a number of measurements may have been made on clay content, and these measurements could be described by the mean and standard deviation of the data set. The conceptual model of the system may have concluded that the migration of a contaminant is controlled by the total clay content along the pathway. In this case, the key parameter would be the average or mean clay content, rather than the measured range in values.

In calculating the mean of the sample data, there will be some uncertainty as to how far away this will be from the mean of the population (e.g. if every possible measurement of clay content could be made). This uncertainty may decrease as more measurements are obtained. In Appendix B, a methodology is presented to determine a PDF to describe the uncertainty of the mean of a set of measurements, i.e. how likely is it that this value corresponds to the population mean.

It is essential that there should be a clear understanding of whether upscaling is appropriate to a problem and that this approach can be justified. Parameter values should only be upscaled with extreme care where heterogeneity in the system behaviour is considered important in controlling the migration of contaminants. This is illustrated by Models 1 and 2 in Figure 4.1. For Model 1, the use of an 'average' value of hydraulic property to describe the movement of contaminants would be appropriate. Whilst for Model 2, where continuous layers of sand and gravel are present, then upscaling would be inappropriate unless the structure of the heterogeneity and associated parameter variability is well understood. Upscaling will result in a distribution with a narrower range. In most cases it is unlikely that understanding will be sufficient to perform upscaling in a meaningful manner.

In most analytical expressions for describing contaminant transport, the parameter in the equation represents a sort of 'average' value representative of the entire flow path. This is not necessarily an arithmetic mean although it is intuitive that it would be some value between the largest and the smallest encountered.

4.5 Summary of Procedure for Selecting a PDF

The overall procedure for choosing a PDF is as follows (Section 4.3.2):

i) Review the data, check for problems etc;

ii) Plot data (e.g. histograms, frequency plots);

iii) Select PDF which best describes the data (based on statistical analysis of the data set, expert judgement based on knowledge of the parameter);

iv) Estimate parameters of the distribution (Table 3.2, Table 4.1);

v) Test if selected distribution fits the data through the use of statistical methods (e.g. Box 4.3, Box 4.4) and/or comparison of the distribution with the data set.

This procedure assumes sufficient data are available, Sections 4.3.3 and 4.3.4 give guidance on the approach to be adopted for PDF selection where data are limited or insufficient.

5. Methods of modelling

5.1 Probability space

When using probabilistic models, it is also essential to consider parameter space. This concept is illustrated by Figure 5.1 which shows the boundary of possible values for two hydrogeological variables based on their PDFs. On the assumption that the PDFs have been selected conservatively (i.e. the uncertainty in these variables has been taken into account), then the actual system should fall within this bounded area. In practice, the system will be defined by more than two variables, and the definition of parameter space will be more complex.

All possible outcomes of a probabilistic analysis will fall within this parameter space. An initial check that can be used, before undertaking an extensive series of probabilistic realisations, is to calculate the model results based on hydrogeological parameter values taken at the boundaries of the parameter space. This analysis may show that all of the results are outside the area of concern (e.g. below a remedial target defined for a receptor) and, therefore, that further work is unnecessary. The latter approach is also useful in providing an independent check of the probabilistic analysis.

The benefit from considering parameter space is that the plausible parameter space may be smaller than the space defined by the extreme values, and consequently the range of possible outcomes is smaller. An example of this is a space defined by porosity and hydraulic conductivity. For most rocks, high hydraulic conductivities are associated with high rather than low porosities, and *vice versa*. Thus it may be possible to eliminate areas of the parameter space based on knowledge of the dependency of different parameters, as illustrated by Figure 5.1. A second example is where values of hydraulic conductivity and gradient could be constrained by comparing the flow for a groundwater catchment calculated from Darcy's Law with the groundwater water recharge over the catchment.

This concept of parameter space may, therefore, prove useful in interpreting and refining the results of probabilistic runs, by eliminating results from highly improbable areas of the parameter space where evidence shows the answer cannot lie. This may reduce the range of estimated risk.

5.2 Monte Carlo method

Monte Carlo simulation is the most widely used method of uncertainty analysis (refer also to Section 3.5). The basic method involves:

i) Definition of the probability distribution for each model parameter;

ii) Repeated running of the model with parameter sets chosen from these distributions. The result from each run (referred to in this report as a realisation) is recorded to build up a picture of the possible outcomes from combining different parameter values;

iii) Analysis of the results from each model calculation or realisation to obtain a probability distribution (as illustrated by Figure 2.2) of the possible outcomes from combining the parameter values. The results are typically expressed in terms of the likelihood that a given value will be exceeded (Figure 5.2).

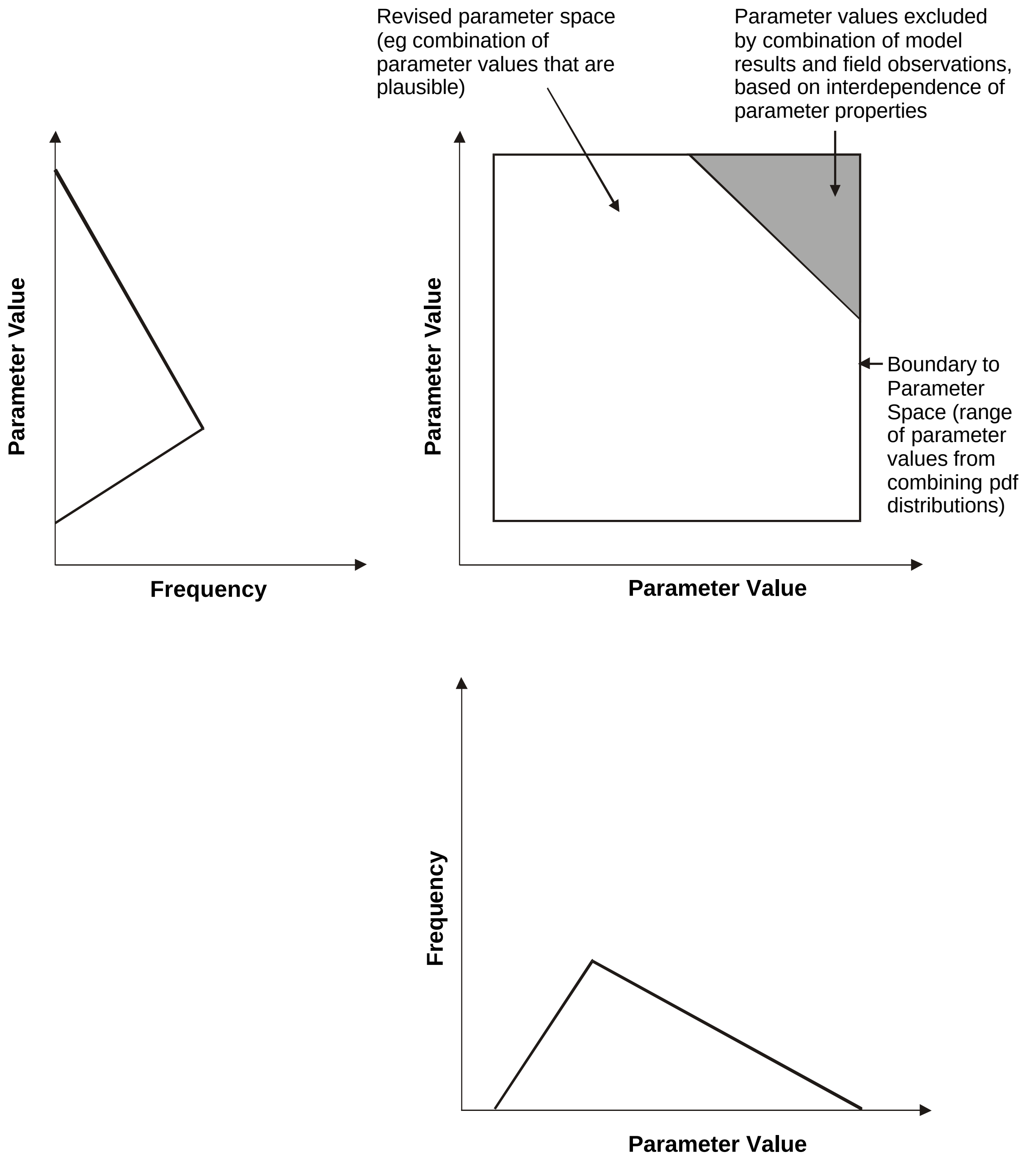


Figure 5.1 Illustration of Parameter Space

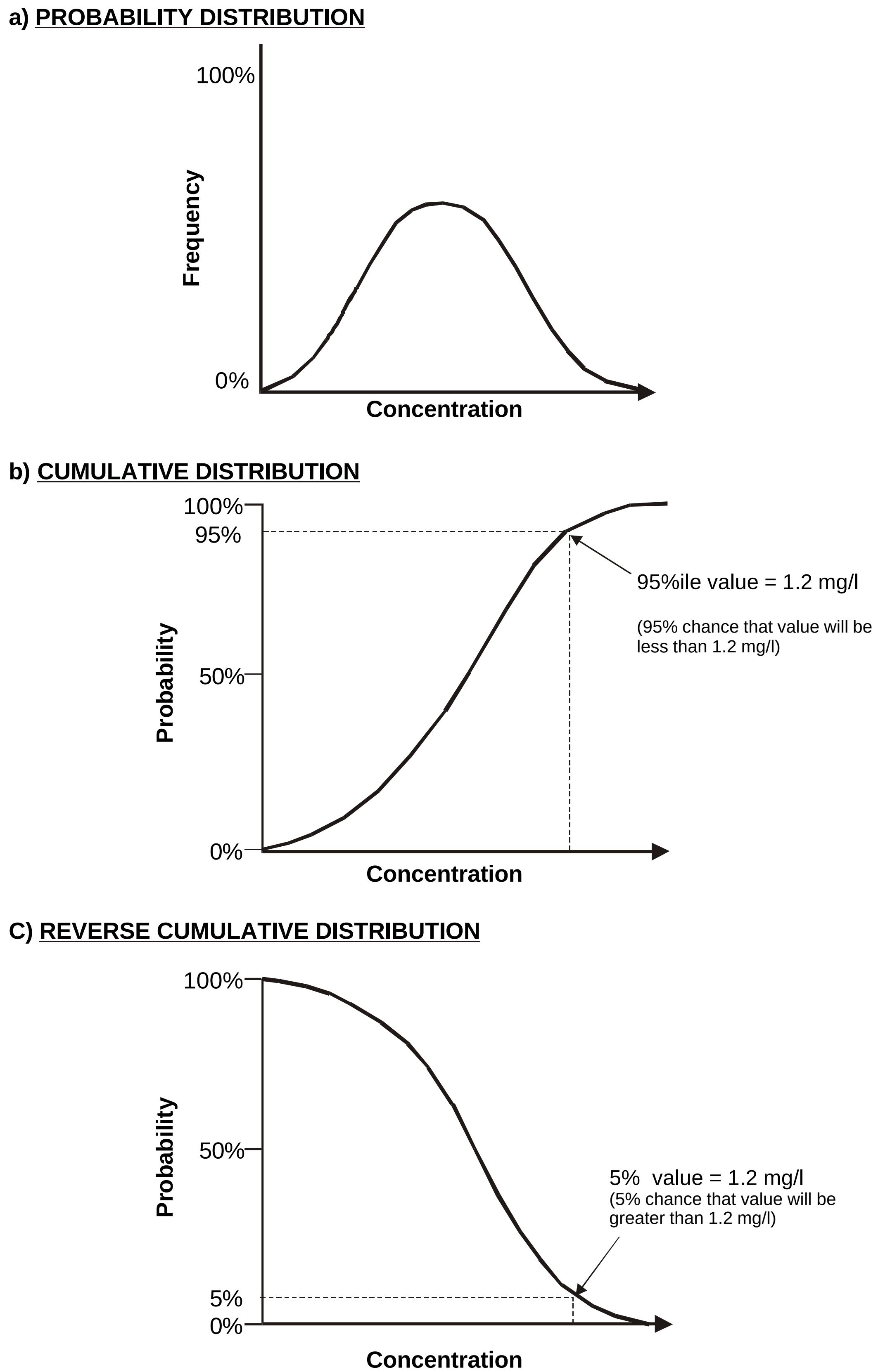


Figure 5.2 Distribution Curves

Data input values to the model will either be randomly selected from their defined parameter distribution, or an algorithm used to select the parameter value. An example of the latter is the Latin Hypercube method. In this technique, the probability distribution is divided into intervals of equal probability and data are then picked from each interval as a way of ensuring that values are taken from the entire probability distribution. This decreases the number of realisations required to achieve convergence.

The accuracy or repeatability of a probabilistic simulation will be dependent on the number of realisations. This is illustrated by Figure 5.3, which shows the cumulative frequency distribution for the output from a Monte Carlo run at different points during the simulation. At the start of the simulation, the distribution curve describing the model results is very variable, but with an increasing number of simulations the smoothness of the distribution increases.

Care needs to be taken in selecting an appropriate number of realisations. For some fate and transport models, running a Monte Carlo simulation can be extremely time consuming. Conversely if too few realisations are undertaken then the results will not be repeatable, particularly at the high or low range. This is illustrated in Table 5.1, where the output from three different LandSim model runs are compared for different numbers of realisations. This table shows that the repeatability of the model results improves with the number of realisations.

Table 5.1 Influence of Realisations on Predicted Leakage (using LandSim v.1.08 Model)

Number of Simulations	**Predicted Leakage (m^3/d) at the 90%ile value**			
	1st Run	**2nd Run**	**3rd Run**	**Range**
200	36.6	32.4	32.9	4.2
500	33.8	37.9	37.6	4.1
3000	31.9	32.2	32.9	1.0

For some models, an element of trial and error will be required to determine the minimum number of realisations required to obtain model convergence i.e. repeatable results. In general, 500 realisations are recommended for initial model runs (to minimise run times) to check the model, and 2000 to 3000 for final model runs. This may be determined by comparison of the model results from one run to the next, or examination of the distribution curve of the results (irregularities in the distribution curve suggest that further model realisations are required).

The greater the number of realisations the more likely that combinations of parameters from the higher and lower ends of their probability distributions will be trialed and consequently the confidence in the result of the Monte-Carlo simulation will increase, i.e. the resulting probability distribution reflects the range of possible results. For example, LandSim recommends that a minimum of 1001 realisations should be run to give 99% confidence (i.e. how certain we are of the result) for the 95%ile result; but that good practice is to run the maximum number of realisations (3000) for the final model run. Further discussion of confidence is given in Gilbert, 1987).

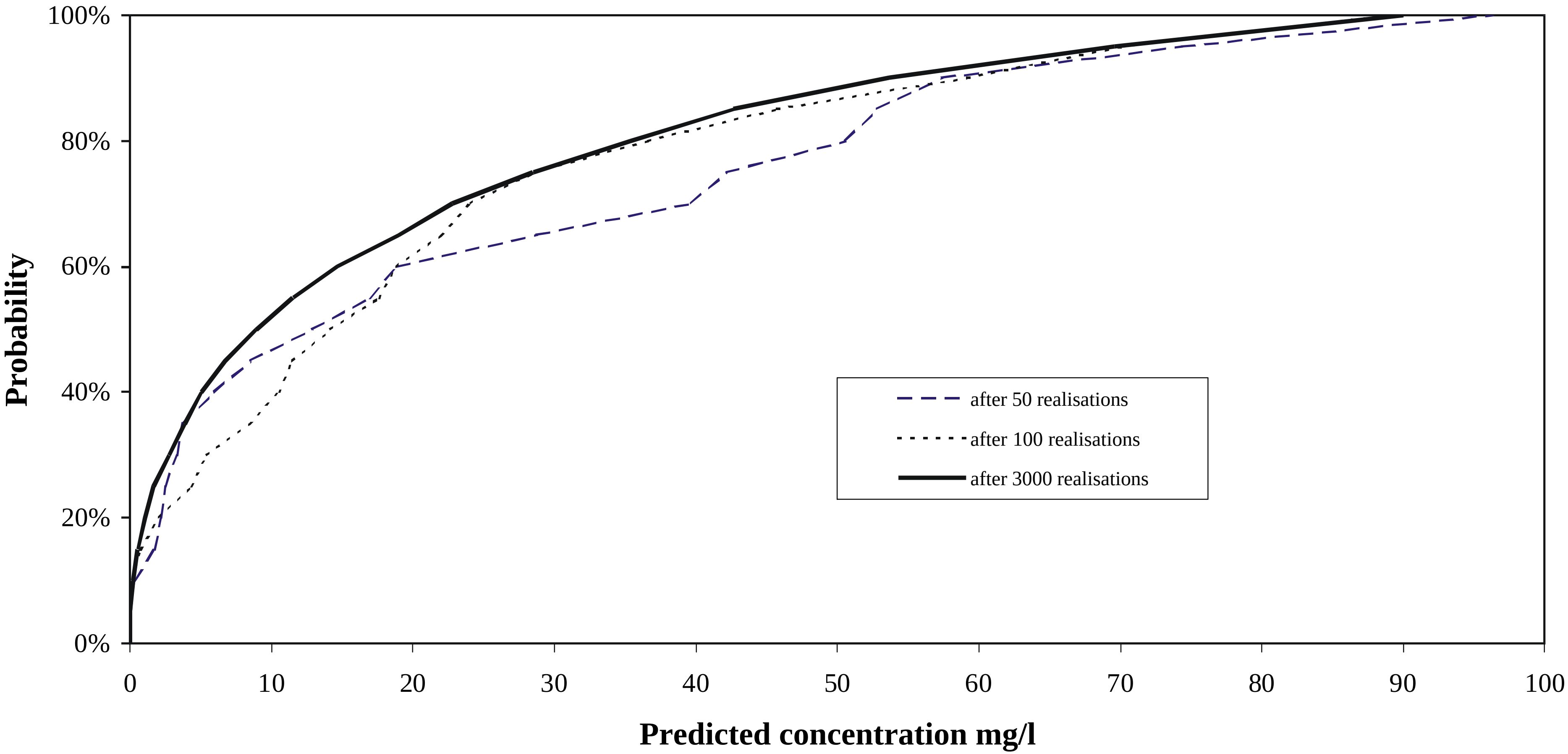


Figure 5.3 Influence of the Number of Simulations on Model Results

5.3 Sensitivity analysis

Sensitivity analyses are used primarily to understand which parameters have the largest effect on the outcome, i.e. it is done to understand how the model works and identify which are the critical parameters. A sensitivity analysis comprises modifying parameter values or parameter distributions and examining the effect of this change on the model results. The results of this analysis should then be used to determine whether further data are needed to define sensitive parameters, i.e. those parameters which have the greatest influence on the model result. Further discussion of sensitivity analyses is given in Environment Agency (2001b).

It is important to note that further data collection may not necessarily result in a narrower or tighter distribution of parameter values (i.e. justification for excluding extreme values from the data set).

The results of the sensitivity analysis can be compared against field observations to check that the results fall within a credible range. They may allow some constraints to be put on the possible values for uncertain parameters. However, because most models will have ten or more parameters, the 'acceptable' parameter space will be very complex indeed, and is unlikely to be fully explorable by a sensitivity analysis.

The assessment should also consider the influence of the type of PDF used to describe the variation in a parameter value, e.g. does the distribution give greater significance to maximum and minimum values.

5.4 Models

A number of probabilistic models have been developed for subsurface contaminant fate and transport modelling, including two Agency codes, LandSim and ConSim.

LandSim is a modelling tool developed for the Agency to assist in the assessment of risk to water quality from landfill sites. The model uses the Monte Carlo approach. Different values can be entered for each input parameter, but the model has pre-assigned distribution[1] (either Triangular, Log-triangular, Normal or Uniform) for each parameter.

ConSim is a modelling tool developed for the Agency to assist in the assessment of the impact of contaminated soils on groundwater quality. The model uses the Monte Carlo approach, and allows the user to choose one of the following PDF distributions for each model parameter (including Uniform, Log-uniform, Triangular, Log-triangular, Normal, Log-normal, Poisson, Binomial and Exponential). For the Normal and Exponential distributions, the package automatically constrains the minimum value as zero and sets the maximum value as infinity.

Other probabilistic modelling tools that are widely available include CrystalBall and @Risk. These packages are linked to a spreadsheet model (Microsoft Excel™) and provide the option of undertaking Monte Carlo analysis. The user is required to input the necessary calculation in the spreadsheet. A range of distribution functions are provided, and the user can constrain some distributions (such as Log-normal), by entering a minimum and maximum value as well as the mean and standard deviation. This has the advantage of removing extreme values that may be generated by the distribution function. This should be done only after a thorough

[1] LandSim release 2 (Environment Agency, 2001c) requires PDFs to be specified for a number of parameter distributions.

consideration of the values and the evidence against them. Extremeness is in itself no justification for removal.

It should be emphasised that the modelling approach and any software adopted should be appropriate within the UK policy/legislative framework, and described in accompanying reports such that the modelling is auditable. The use of in-house spreadsheets and/or code will need to be fully described in submissions to the Agency in order to allow a review to be performed, and should be supported by evidence of quality assurance procedures and independent (non-Agency) model verfication.

6. Worked example of probabilistic transport model

This section examines the use of probabilistic analysis in fate and transport modelling. It also examines the influence of different PDF distributions (Normal, Triangular, Uniform etc) on the output from the Ogata & Banks (Bear, 1979) analytical transport equation on the calculated contaminant concentration at a compliance point. For the purposes of this example the compliance point is taken at a point 600 m down hydraulic gradient of the source.

The parameter values used in this evaluation are given in Table 6.1. To simplify this exercise, PDFs are applied to hydraulic conductivity and dispersivity only and the remaining parameters are defined by a single value, i.e. deterministically.

To provide an initial indication of the likely range in predicted concentrations at the receptor, minimum and maximum values (Table 6.1) for hydraulic conductivity and dispersivity have been combined. The calculated values range from 0 to 98 mg/l (these provide an initial indication of parameter space), and indicate the potential for a breakthrough (defined as concentration greater than 0 mg/l) at the receptor (Table 6.2).

Table 6.1 Parameter values

Parameter	**Value**		
	Median	**Minimum**	**Maximum**
Hydraulic conductivity (m/d)†	25 ‡	0.4	124‡
Hydraulic gradient	0.004		
Effective porosity	0.1		
Longitudinal dispersivity (m)	30	20	40
Retardation factor, R_f	1.5		
Degradation rate (d^{-1})	No decay Entered as zero		
Distance to receptor (m)	600		
Time since contaminant release (d)	200		
Initial concentration (mg/l)	100		

Note (†) The measured values of hydraulic conductivity are 0.4, 4, 13, 14, 16, 21, 25, 34, 35, 41, 48, 71 and 124 m/d and relate to a sand and gravel aquifer where borehole logs indicate a variable sequence of sands, gravels and silty clays. (These values were taken from a larger synthetic set that had a Log-normal distribution).
Note (‡) Based on measured values.

Table 6.2 Initial calculation of contaminant concentrations at the receptor based on minimum, median and maximum assumed values

Hydraulic conductivity m/d	Dispersivity m	Concentration mg/l At receptor after 200 days
0.4	20	0.0
0.4	30	0.0
0.4	40	0.0
25	20	0.0
25	30	0.02
25	40	0.1
124	20	97.9
124	30	95.7
124	40	93.7

In this example, two different conceptual models have been developed to describe how the measurements of hydraulic conductivity could be used to describe the aquifer system, and the effect of these models on the predicted contaminant concentrations. The cases considered are:

Conceptual Model 1

In the first model, it is assumed that any of the measurements of hydraulic conductivity could describe the flowpath to the receptor. That is, it is considered that the values of hydraulic conductivity are an uncertain measurement of a uniform parameter. A probabilistic analysis has been undertaken to consider this uncertainty.

In addition, the influence of assuming different probability distributions to describe the hydraulic conductivity data has also been considered as part of the analysis. The distributions used are:

- Uniform: where minimum and maximum values are based on the observed data set. These maximum and minimum values are likely to underestimate the actual range in values of hydraulic conductivity;
- Triangular: where the minimum, most likely and maximum values are based on the observed data set, and are thus also likely to underestimate the actual range in values of hydraulic conductivity;
- Normal: The arithmetic mean and standard deviation of the data set have been calculated (Table 4.1);
- Log-normal: The arithmetic and standard deviation of the logarithms of the data values have been calculated (Table 4.1).

As noted in Section 4.3.3, a Log-normal distribution would usually be used to describe the variation in hydraulic conductivity.

The predicted concentrations are given in Table 6.3, in terms of the 50 and 95%ile values, and illustrated by Figure 6.1.

Figure 6.2 compares these distributions to the original data set and the following observations are made:

- The Uniform distribution provides the poorest match to the data set.
- The Normal distribution needs to be curtailed (to prevent negative values) and provides poor fit in the middle of the observed range, but reasonable match at the upper end of the range.
- The Log-normal and Triangular distributions provides the best match at the lower end of the data range, but overestimate values at the higher end.

It is emphasised that this comparison is only in relation to a limited set of data.

These matches are reflected in the predicted calculations as follows:

- The Uniform distribution gives highest 50%ile concentrations;
- The Log-normal distribution gives the highest concentration at the 95%ile.

The spread of these results is an indication of conceptual uncertainty in the model.

It is important to think about selected distribution (does it over or underestimate values within the data range), and how it affects the predicted results. The modeller may decide to select the Log-normal distribution, but needs to be confident that the one high value of hydraulic conductivity does not reflect part of the system characterised by a high permeability.

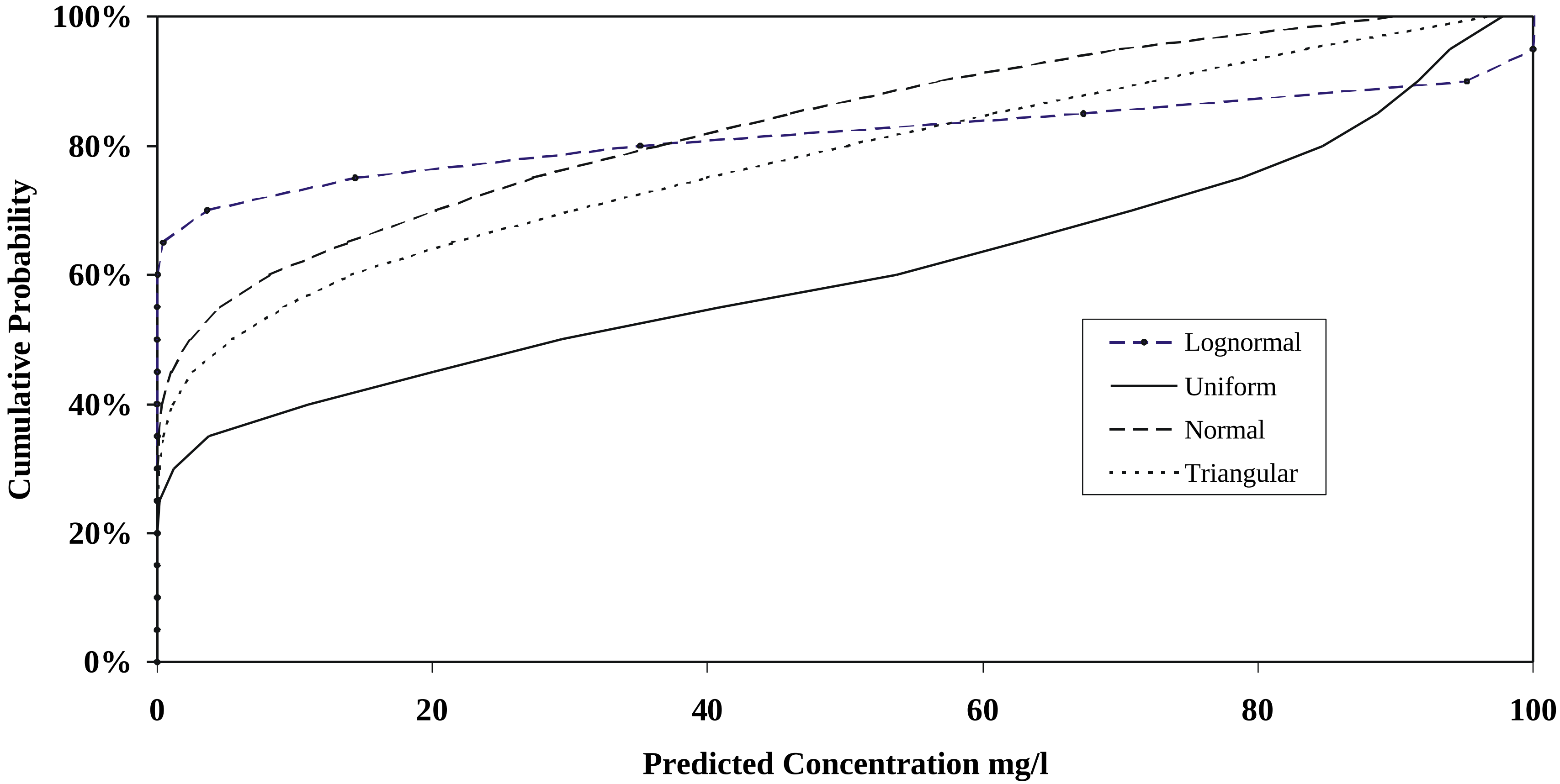


Figure 6.1 Influence of Different PDFs on the Model Results (Conceptual Model 1)

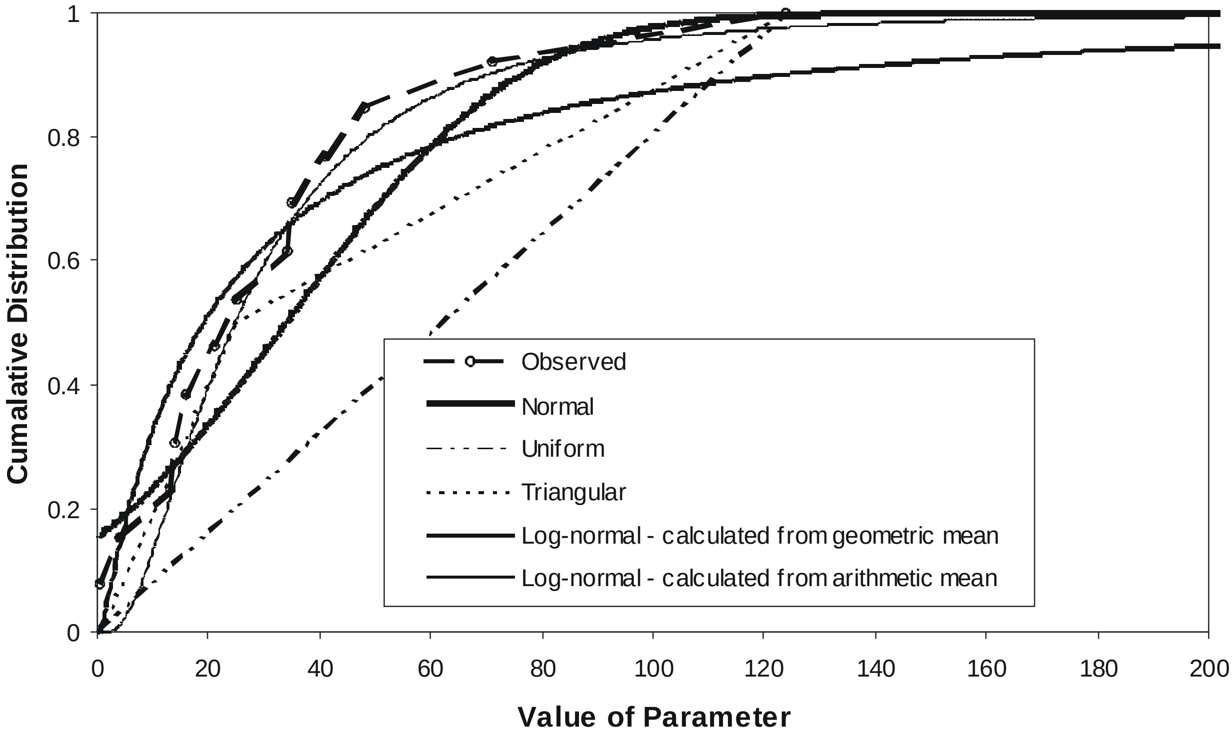


Figure 6.2 Comparison of Probability Distribution Functions for Hydraulic Conductivity (see Table 6.1)

Table 6.3 Influence of different assumed PDF's on predicted concentrations

Distribution	Input Parameters Hydraulic conductivity (m/d)	Predicted Concentration mg/l 50%ile	95%ile
Uniform	Min = 0.4 Max = 124	29	84
Triangular	Min = 0.4 Most likely = 25 Max = 124	5.5	70
Normal[1]	Mean = 34 Standard dev. = 33	2.4	94
Log-normal[2]	Mean of logs= 2.95 Stdev of logs = 1.45	0	100

Note (1) For this distribution, the equivalent minimum and maximum values (i.e. within three standard deviations) are 0.1 m/d and 133 m/d. The minimum value has been truncated as the Normal distribution gives negative values at the lower end of the distribution.

Note (2) For this distribution, the equivalent minimum and maximum values (i.e. within three standard deviations) are 0.25 m/d and 1480 m/d.

Nevertheless the predicted results show that, irrespective of the distribution used in the analysis, contaminant breakthrough would be expected at the 95%ile level (i.e. ≥ 1 in 20 chance of contamination at the receptor). The results show a wider range at the 50%ile, although it is more likely that any decision to implement remedial measures would be based on the 95%ile results and in this case the 95%ile results point to a risk to the receptor.

Conceptual Model 2

In the second conceptual model, it is considered that the values of hydraulic conductivity represent a more random system, i.e. the measurements relate to discontinuous lenses of sand, gravel and silty clay, such that the mean value of the measurements of hydraulic conductivity is more likely to describe the flowpath. The uncertainty lies in whether this mean hydraulic conductivity of the data set is the same as the actual population mean. A probability distribution was calculated to describe the variation in the mean (taken as the geometric mean of the data plus a calculated factor to describe its variance) as described in Appendix B. The resultant distribution is described by a Log-normal distribution (mean = 2.95, standard deviation = 1.45).

The predicted contaminant concentrations are depicted in Figure 6.3 and compared with the Log-normal distribution for conceptual model 1. This predicted contaminant profile is markedly different from those calculated under the first conceptual model, but still shows a contaminant breakthrough at the 95 %ile level, albeit at a lower concentration (about 12 mg/l). This observation may provide added support to the decision to undertake remedial action. In other cases, the difference in results obtained from different conceptual models may demonstrate that more information is required about the system (e.g. whether the gravel horizons are laterally continuous), if both conceptual models have equal merit.

It should be noted that other conceptual models of the data may also be valid, for example the deposit may comprise an alternating sequence of sands, gravels and silty clays and that the measurements of hydraulic conductivity relate to the different horizons. In this case, the important parameter may be the higher value of hydraulic conductivity, particularly if this is the only measurement of the hydraulic conductivity of the gravel layer. In this case additional information may need to be incorporated in defining the probability distribution. For example, information from comparable sites on the distribution of hydraulic conductivity values for gravel deposits.

Discussion

This example illustrates that useful results can be obtained by considering different conceptual models to describe the system behaviour. By examining the data in different ways, the possible sensitivity of conclusions to these assumptions can be seen, often leading to greater confidence in decisions, even if this is to obtain further information about the system behaviour. It also serves to illustrate the importance of the conceptual model, and of considering different possible interpretations of the data.

In this particular case, where there is a possibility of no significant risk, it may be worthwhile to collect more data to refine the conceptual model, before embarking on a protracted remediation exercise. However, using a number of conceptual models indicated the possibility at the 5% confidence level of a contaminant breakthrough, and therefore if a 95% confidence in no pollution is needed, the decision to take remedial action is robust and unaffected by conceptual uncertainty.

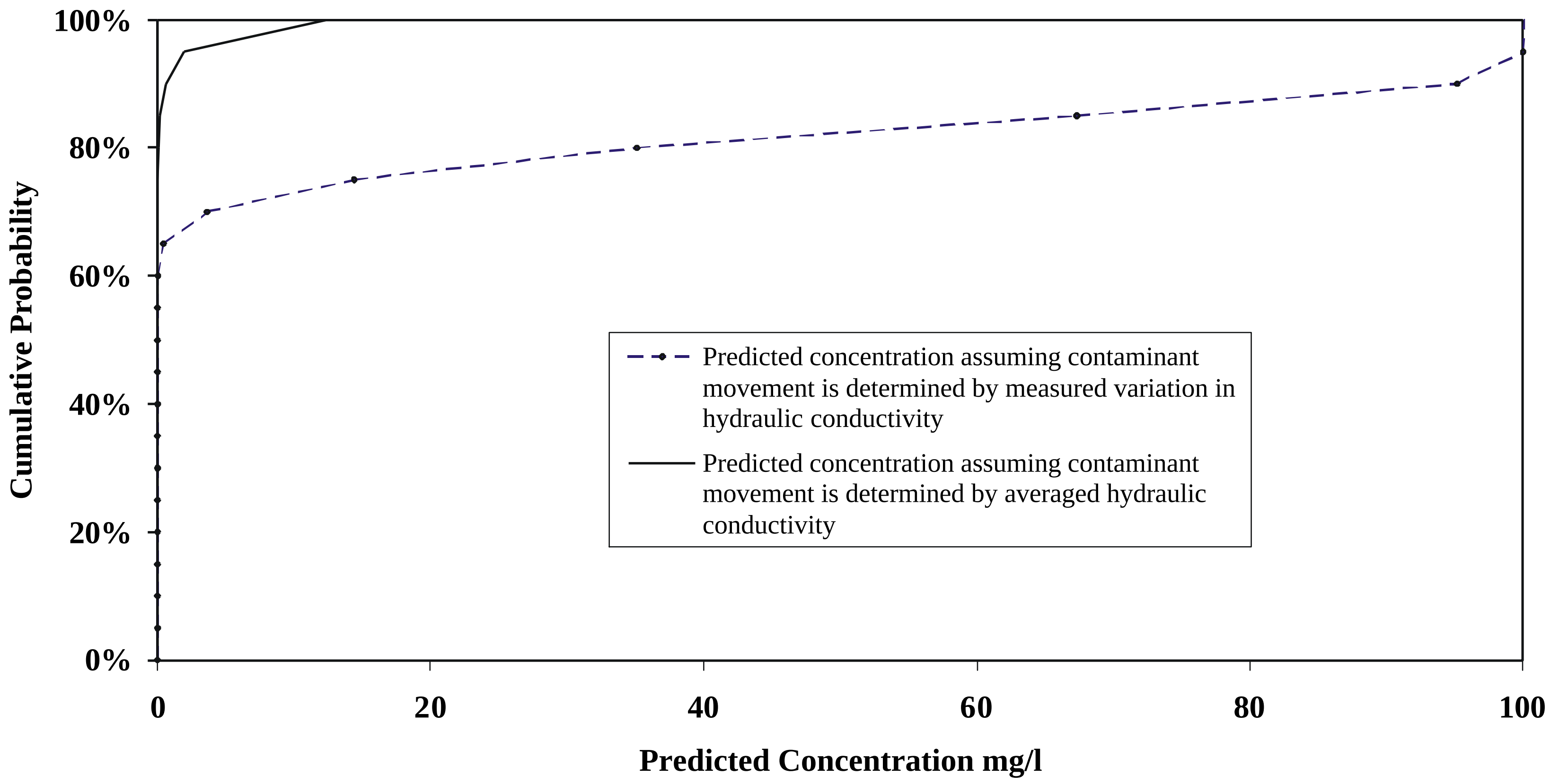


Figure 6.3 Influence of Different Conceptual Models of the Variation of Hydraulic Conductivity on the Model Results (Conceptual Model 2)

7. Interpretation of results and reality check

7.1 Introduction

This section provides guidance on how the results from a probabilistic model simulation should be interpreted, including:

- checking that the model results are valid when compared with field observations;
- understanding what factors or distributions have had the greatest influence on the model output.

7.2 What do the results mean?

In a fate and transport probabilistic model simulation, a distribution of possible outcomes is calculated, based on the PDF distributions that have been defined for each model input. The results are typically presented (Figure 5.2) as a:

- probability distribution (PDF);
- cumulative distribution (CDF);
- reverse or complementary cumulative distribution (CCDF).

These plots show the percentage chance or likelihood that the result will be greater or less than a given value. For example, using Figure 5.2b (cumulative frequency distribution graph), the 95%ile value of 1.2 mg/l (for the calculated contaminant concentration at an identified receptor), means that for the possible combinations of input parameters, there is a 95% chance (19 in 20) that the possible outcomes will be less than this value, or a 5% chance (1 in 20) of this value being exceeded. In effect we are saying that we can be 95% certain that the contaminant concentration at the receptor will not exceed 1.2 mg/l. This is still dependent on our defined parameter distributions, and whether these truly describe the site conditions. It is important to recognise that incorporating uncertainty into the assessment does not mean that we can be certain about the predicted result. It will also be dependent on whether the conceptual model is correct and whether an appropriate mathematical model has been used (refer to Environment Agency, 2001a).

The model results at the higher and lower end of the probability range will be most sensitive to the extreme values in the data set and to the number of simulations.

Care needs to be taken when recording the results to ensure that the plot is read in the correct way. An easy error is to misread cumulative and reverse cumulative distribution plots. For example, the 95%ile value on a cumulative graph may mistakenly be recorded as having a 95% chance of being exceeded, when in fact the opposite is true.

The slope of a cumulative frequency graph will reflect the variability in the parameter definition: the steeper the graph, the smaller the range in values that have been used to define each parameter input distribution (this does not necessarily imply that more confidence can be attached to the results).

It is important to recognise that by combining the possible range of values for a number of different parameters (particularly if extreme values have been used in defining the PDF), that there is a potential for a wide range of possible outcomes. In some cases these may not be credible when compared with site characteristics (refer to Section 5.3).

A temptation in undertaking risk assessments is to take, say, the 95%ile value from a number of different parameter distributions and to combine them in a deterministic model run (refer to Box 3.1). Although this may be seen as a method of determining a worst case scenario, the resulting answer, if say the 95%ile value had been taken from three distributions, would be approximately equivalent to the 99.99%ile value and as a consequence may not represent a reasonable solution.

The results from a fate and transport model are usually compared with a standard, for example, a remedial target to protect a groundwater or surface water receptor (Environment Agency, 1999a). Where a probabilistic analysis has been undertaken, the criterion for deciding whether the model results comply with or exceed the standard needs to be determined, i.e. what percentage of the model results exceed the standard. For example, if the 90%ile is chosen, then we are 90% certain that the identified standard will not be exceeded. Typically the 95%ile is used as the criterion for assessing acceptable risks of water pollution against an environmental standard, although this should be determined based on the nature and significance of the hazard and the sensitivity of the receptor. It will be also dependent on the input data being realistic, i.e. the parameter distribution is not unduly conservative or optimistic.

It is important to understand whether the combinations of parameter values that contribute to the tail of the distribution are realistic - especially if dependence was not incorporated into the uncertainty analysis. For example, certain combinations of high hydraulic conductivity, low porosity and a steep hydraulic gradient may not be realistic. Data collection should be targeted to parameters that have been found to have the greatest influence on the modelled impact.

7.3 Constraining the analysis

In interpreting the results, it is essential that they should be reviewed in the context of what is credible and consistent with field observations. This recognises that:

- a probabilistic analysis will result in the combination of maximum and minimum values from the parameter distributions, giving a wide range of possible outcomes. Consequently extreme values can have a significant influence on the model results. The real system is unlikely (by definition) to contain a large number of extreme values;
- parameters may be dependent, such that combining the extreme ends of different parameter distributions may be unrealistic. However, any parameter combinations that are excluded must be excluded for defensible reasons and be documented.

It may be appropriate to back-substitute the values producing these realisations, to check the hydraulics are reasonable. It is important to consider dependent values, as well as the one of immediate interest. For example, the immediate interest may be the concentration of a contaminant at a particular point and time. Associated with this calculation will be a calculation of groundwater flow. For example, the model may have predicted that as a consequence of combining the range of parameter values for hydraulic conductivity, hydraulic gradient, aquifer thickness and aquifer width that the range of possible groundwater flows is 10 to 3000 m^3/d. However, consideration of the groundwater catchment area, and aquifer recharge rate, may indicate that the groundwater flow could only realistically fall in the range

400 to 1200 m^3/d. In this case the larger range of calculated model flows may be a result of including extreme values for hydraulic conductivity and/or that no account was given of the dependency of hydraulic gradient on hydraulic conductivity. In the latter case, more detailed examination of the field data may have shown that the steeper hydraulic gradients may have been a local feature of the site and were associated with zones of lower hydraulic conductivity.

Some commercial modelling packages provide the option of constraining components of the analysis; for example, a maximum groundwater flow could be specified. This illustrates the need to revise the model input parameter distributions, based on field observations, and that typically, the overall procedure will be iterative. The development of a model should be an iterative exercise, with the model being continually refined through reference to field observations.

Conversely, it is also important to recognise that the model results may identify limitations in the definition of the site. For example, field monitoring may have shown no evidence for contamination downgradient of a site, whereas model predictions (using parameter values derived from field investigations) may have indicated that groundwater contamination would be expected. This could indicate an inadequate monitoring network, such as inappropriately constructed or placed boreholes, rather than errors in the selected parameter distributions, and the need to undertake further investigation.

7.4 Review of model results

In assessing the results, the following factors should be considered:

i) How confident are we in the conceptual model?

ii) Are the model results consistent with field observations and the conceptual model, i.e. is the observed system behaviour within the range of behaviours predicted by the model? (remember that reality is a single realisation, probably different from any of those used in the probabilistic simulation). The criteria for deciding whether a mathematical model provides an acceptable representation should be determined as part of the model conceptualisation (refer to Environment Agency, 2000a). Where the model fails to represent the system, then consideration should be given to whether:

- the conceptual model is valid or needs reassessing (for example, the validity of excluding high values of hydraulic conductivity from the data set, when field observations suggests that the system behaviour is determined by very permeable pathways). It is important to recognise that an incorrect conceptual model will invalidate the results of the study;
- the system is sufficiently well defined by the site investigation;
- data values used to define the PDF are valid, particularly where extreme values have been excluded or included in this analysis. For example, hydraulic conductivity testing of a sand and gravel aquifer may have given the following set of measurements for hydraulic conductivity:

  1, 2.1, 2.7, 4.3, 7.8, 8.2, 9.1, 11.2, 150 m/d

  In defining the parameter distribution, a decision will need to have been made as to whether to include the value of 150 m/d in the dataset. Is it an incorrect result or does it indicate the presence of a permeable gravel lens within the deposit, and which may represent a highly permeable pathway? Comparison of the model results with field observations may provide a basis for including or excluding this high value. Gilbert (1987) provides further discussion on the handling of outliers in the data set;

- the PDF distribution is appropriate (e.g. does it adequately describe the data set);
- the applicability of the mathematical solution and the possible influence of its assumptions and limitations.

iii) To what degree are the results affected by the variability of the system and uncertainty in defining the system?

iv) To what degree have the results been influenced by the assumptions and limitations of the conceptual model and the model/method used to represent the problem?

v) Is the analysis sensitive to certain parameter distributions? If so, are these adequately constrained by measurement or observations of system behaviour?

vi) Have extreme values in the data set biased the results?

vii) Was it valid to exclude some of the data values or combinations of data values from the analysis (e.g. are apparently extreme data values a true representation of the system)?

viii) Have sufficient realisations been made such that the model results are repeatable?

ix) Is there a conceptual understanding of the data and do these adequately describe the system behaviour. For example, values of hydraulic conductivity may be combined from tests undertaken in different layers of a multi-layered aquifer, whereas contaminant migration was only occurring through one of the layers and this is characterised by a smaller variation in hydraulic conductivity when compared with the entire data set.

x) Are the tails in a predicted parameter distribution realistic, in terms of what are credible combinations of extreme parameter values?

xi) If analysis shows that the missing data are critical to the decision-making process, then more data must be obtained.

A potential danger in using PDFs is to be over-certain that the results are right, when in reality limited information was available to define the parameter distribution.

Depending on the review of the model results, one of the following will be appropriate:

- the results from the analysis can be used in the decision making process;
- further data are required to define the system;
- the conceptual model and method of analysis should be reviewed.

It is essential that the conceptual model and modelling approach should be challenged continually and reassessed, i.e. how confident are we in our understanding and representation of the system behaviour.

7.5 Summary

In certain circumstances, probabilistic models provide a tool to allow uncertainty in the definition of parameter values and/or the heterogeneity of the system to be taken into account in contaminant fate and transport modelling. However, they should not be used as an alternative to obtaining site-specific data or to the development of a defensible conceptual model of the system behaviour (Environment Agency 2001a), particularly as this is often the main uncertainty in any contaminant fate and transport modelling exercise.

8. REFERENCES

BGS/ Environment Agency, 1997.	*The Physical Properties of Major Aquifers in England and Wales.* BGS Technical Report WD/97/34. Environment Agency R&D Publication 8.
BGS/ Environment Agency, 2000.	*The Physical Properties of Minor Aquifers in England and Wales.* BGS Technical Report WD/00/04. Environment Agency R&D Publication 68.
Bear J, 1979.	Hydraulics of Groundwater. McGraw-Hill Book Company.
Cressie, N A C, 1991	Statistics and spatial data. John Wiley and Sons Inc.
Dalkey N and Helmer O, 1963	An experimental application of the delphi method to the user of experts. *Management Science* **9**(3): 458-467, 1963.
Davis J D, Runchal A K and Bauman N A, 1983	Delphi analysis of radionuclide release scenarios for a nuclear waste repository at the Hanford Site, Washington state: An initial Iteration, RHO-BW-ST-42P, Rockwell Hanford Operations, Richland, WA.
Davis, J C, 1973	Statistics and data analysis in geology, 2nd Edition. John Wiley and Sons Inc.
De Marsily, G, 1986	Quantitative Hydrogeology, Academic Press
Decisioneering Inc, 1996	Crystal Ball Version 4.0 User Manual. Decisioneering, Inc.
DeGroot M H, 1970	Optimal Statistical Decisions, McGraw-Hill Book Company, New York, New York.
Deutsh, C V & Journel, A G, 1992	Geostatistical Software Library and Users Guide. Oxford University Press.
Environment Agency, 1996	LandSim: Landfill Performance Simulation by Monte Carlo Method. Prepared by Golder Associates (UK) Ltd.
Environment Agency, 1999a	Methodology for the Derivation of Remedial Targets for Soil and Groundwater to Protect Water Resources. R&D Publication 20. Prepared by Aspinwall and Company.
Environment Agency, 1999b	ConSim: Contamination Impact on Groundwater: Simulation by Monte Carlo Method. Prepared by Golder Associates Ltd.
Environment Agency, 2000	Guidance on the assessment and monitoring of natural attenuation of contaminants in groundwater. R&D Publication 95. Prepared by Enviros Aspinwall & LGC.
Environment Agency, 2001a	Guidance on the Assessment and Interrogation of Subsurface Analytical Contaminant Fate and Transport Models. NGWCLC report NC/99/38/1. Prepared by Entec UK Ltd.

Environment Agency, 2001b	Guide to Good Practice for the Development of Conceptual Models and the Selection and Application of Mathematical Models of Contaminant Transport Processes in the Subsurface. NGWCLC report NC/99/38/2. Prepared by Entec.
Environment Agency, 2001c	LandSim Release 2: Landfill Performance Simulation by Monte Carlo Method. Environment Agency Publication 120. Prepared by Golder Associates (UK) Ltd, Nottingham.
Gelhar L W, 1993	Stochastic Subsurface Hydrology, Prentice Hall, New Jersey.
Gilbert R O, 1987	Statistical methods for environmental pollution monitoring, 2nd Edition. John Wiley and Sons Inc.
Hertwich, E.G., McKone, T.E & Pease, W.S., 1999	Parameter uncertainty and variability in evaluative fate and exposure models. *Risk Analysis* **19** (6)
Journel A G & Huijbregts C, 1978	Mining Geostatistics. Academic Press, New York
Kapur J J, 1989	Maximum Entropy Models in Science and Engineering. Wiley Eastern Ltd
Matheron G, 1967	Eléménts pour une théorie des milieux. Mason et Cie.
Merkhover M W, 1987	IEEE Transactions on Systems, *Man and Cybernetics* **17**(5).
Palisade Corporation, 1997	Guide to Using @Risk. Palisade Corporation.
Speltzer C S and Stael von Holstein C A S, 1975	Probability Encoding in Decision Analysis. Management Science 22, 340-358.
Stael von Holstein C A S and Matheson J E, 1979	A manual for encoding probability distributions. SRI Project 770778, SRI International, Menlo Park, CA.
Till R, 1974	Statistical Methods for the Earth Scientist. Macmillan.
Vanmarcke, E 1983	Random fields: Analysis and synthesis, MIT Press
Vose D, 1996	Quantitative Risk Analyses: A Guide to Monte-Carlo Simulation Modelling.
Wen X & Gómez-Hernández J J, 1996	Upscaling hydraulic conductivities in heterogeneous media: an overview. *Journal of Hydrology* **183**.

Further reading

Domenico P A & Schwartz F W, 1990	Physical and Chemical Hydrogeology. Wiley
Ganoulis J G, 1994	Engineering Risk Analysis of Water Pollution: Probabilities and Fuzzy Sets.
Jaynes E T, 1957	Information theory and statistical mechanics. Phys. Rev. 106:620-630.

Jaynes E T, 1983 Papers on Probability, Statistics and Statistical Physics (ed. R.D. Roenkrnantz), Dordrecht, D. Reidel.

Shaw E M, 1994 Hydrology in Practice, Third Edition, Chapman & Hall.

Appendix A

Example derivation of a probability distribution function

Field measurements of hydraulic conductivity were obtained as part of an investigation of a sand and gravel aquifer. The values of hydraulic conductivity are 0.1, 0.4, 0.4, 1, 23, 26, 45, 60 m/d. This example describes an approach to deriving a PDF.

As only 8 measurements are available, it is inappropriate to undertake a detailed statistical analysis of the data set to derive the form of the PDF, instead a distribution needs to be assumed. Since the parameter is hydraulic conductivity, and the spread of values is more than an order of magnitude, a Log-normal distribution is assumed (refer to Sections 4.3 and 4.3.3).

The estimator for μ for a Log-normal distribution is the arithmetic mean of the logarithms of the sample and for σ the estimator is the standard deviation of the logarithms of the sample (from Table 4.1, Section 4). These estimators are calculated in Table A1.

Table A1 – Data analysis

Hydraulic Conductivity m/d	Natural Log of Measurement (Ln)
0.1	-2.30
0.4	-0.91
0.4	-0.91
1	0
23	3.13
26	3.25
45	3.80
60	4.09
Mean of samples	Arithmetic mean of logarithms μ = [ln(0.1)+ln(0.4)+ln(0.4)+ln(1)+ln(23)+ln(26)+ln(45)+ln(60)]/8 = **1.27**
Variance	Variance of logarithms σ^2 = $[(\ln(0.1)-\ln(1.27))^2+(\ln(0.4)-\ln(1.27))^2+(\ln(0.4)-\ln(1.27))^2+(\ln(1)-\ln(1.27))^2+(\ln(23)-\ln(1.27))^2+(\ln(26)-\ln(1.27))^2+(\ln(45)-\ln(1.27))^2+(\ln(60)-\ln(1.27))^2]/7$ = **6.54**
Standard deviation	Standard deviation (σ) = $\sqrt{6.54}$ = **2.56**

The mean (μ) of the logarithms of the data values is 1.27 and the standard deviation (σ) of the logarithms of the data values is 2.56. The geometric mean of the data is 3.56 m/d (=exp(1.27)).

This analysis indicates that if the data follow a Log-normal distribution, then the distribution can be described by the following estimators:

$$\mu = \text{mean of logarithms} = 1.27$$

$$\sigma = \text{standard deviation of logarithms} = 2.56$$

This implies that 68% of samples are between 0.28 m/d and 46 m/d (i.e. these values occur within one standard deviation). Checking back to the original data shows that 6 of the eight measurements fall within this range, which is about right. As a final check the PDF should be compared with the original data set to check that it adequately describes the data (Figure A1). Good practice should be to check whether other distributions (e.g. Log-triangular) provide a better fit, i.e. we have assumed in this example that the data are log-normally distributed.

In using Log-normal distributions in software packages it is important to check in what form do the values of mean and standard deviation need to be entered, for example the default option in Crystal Ball is that the arithmetic mean and the standard deviation of the data set should be entered. For the data set in Table A1, these values are 19 m/d and 23 m/d respectively. For a Log-normal distribution the arithmetic and standard deviation are related to the mean of the logarithms and standard deviation of the logarithms by the following expressions (i.e. one can be calculated from the other):

$m = \exp\ (m_n + ½\ \sigma_n^2)$

$\sigma = m^2\ (\exp\ (\sigma_n^2) - 1)$

or

$\sigma_n = \ln(m) - ½\ \ln(\sigma^2/m^2 + 1)$

$\sigma_n = \ln(\sigma^2/m^2 + 1)$

where;

m_n = mean of natural log or geometric mean of the data set

σ_n = standard deviation of natural logs

m = arithmetic mean of the data set

σ = standard deviation of the data set.

However, if the data set are not Log-normally distributed, then these equations will not hold and will provide different answers and different distributions (as shown on Figure A1).

For many problems there will be insufficient data to determine whether a data set follow a given distribution, and consequently may be more appropriate to determine which distribution provides the closest fit to the data. In this case, it is worthwhile calculating a Log-normal distribution using both the arithmetic mean and standard deviation as well as the geometric mean and standard deviation (if the data are from a Log-normal distribution then the calculated PDF curves will be the same).

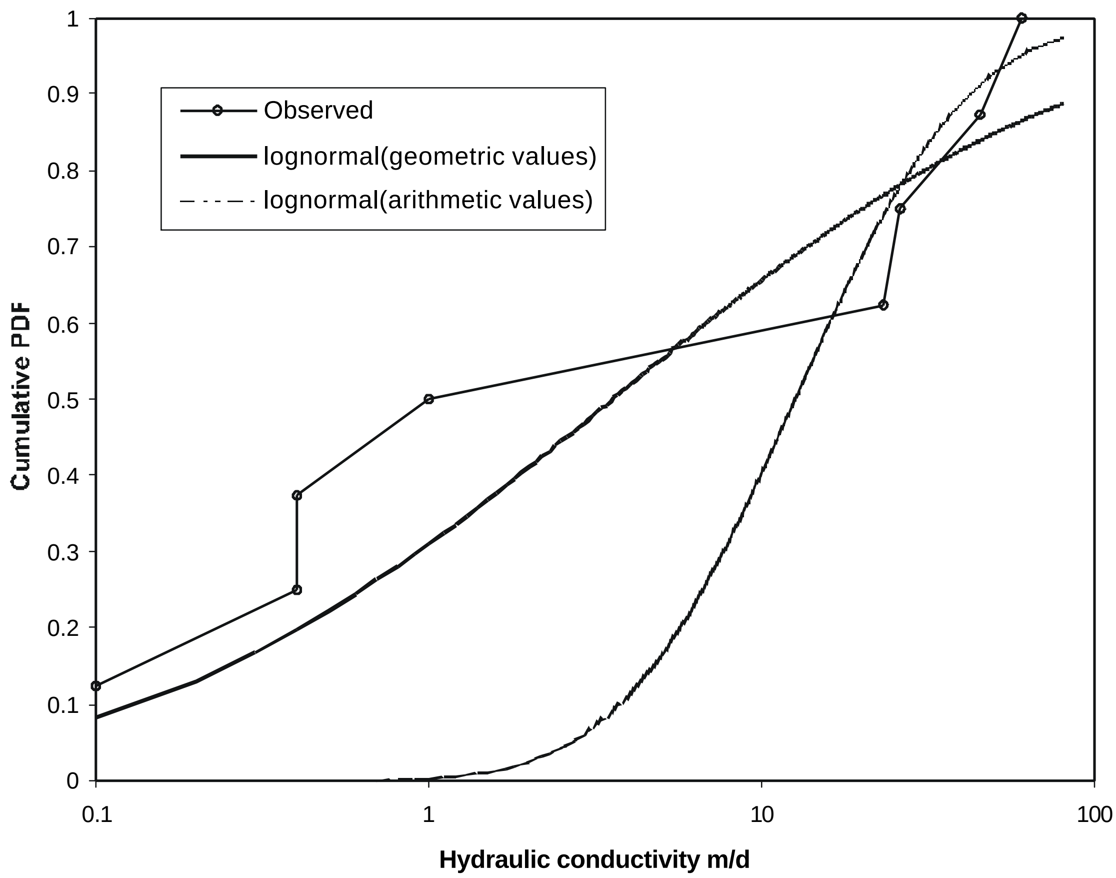


Figure A1 Comparison of Distributions

Example of Chi-squared test to investigate normality

A total of 30 measurements of porosity have been obtained as given below:

Porosity Measurements	*z* value	Porosity Measurements	*z* value
0.06	-2.01	0.21	0.13
0.07	-1.86	0.22	0.27
0.1	-1.44	0.23	0.41
0.1	-1.44	0.23	0.41
0.12	-1.15	0.24	0.56
0.13	-1.01	0.24	0.56
0.15	-0.72	0.24	0.56
0.15	-0.72	0.25	0.70
0.16	-0.58	0.25	0.70
0.17	-0.44	0.27	0.98
0.17	-0.44	0.28	1.12
0.18	-0.30	0.28	1.12
0.19	-0.16	0.3	1.41
0.19	-0.16	0.31	1.55
0.2	-0.01	0.34	1.98
0.2	-0.01		

1. Calculate arithmetic mean ($\bar{x}$) and variance (σ^2) of sample measurements (x)

   Mean = 0.2

   Variance = 0.0049

   Standard deviation = 0.07

2. Form a histogram of $z = (x - \bar{x})/\sigma$ with equal intervals. The calculated values of z are given in the above table and these have been divided into four classes as shown below. *The size of interval and the end intervals should be selected to ensure that there are at least 5 values in each interval (if necessary class intervals can be combined).*

z- class	Frequency
<-1	6
-1 to 0	10
0 to 1	11
> 1	5

3. Calculate the expected frequency (EF) for each interval if the distribution were normal with mean zero and variance 1 as given in the table below.

EF = probability of the z-value occurring in that class interval × total number of readings.

***z*- class**	**Observed Frequency (OF)**	**Probability of a Z-value falling in this range[1]**	**Expected Frequency (EF)[2]**
<-1	6	0.1587	0.237
-1 to 0	10	0.3413	0.032
0 to 1	11	0.3413	0.017
> 1	5	0.1587	0.001

(1) The probability can be derived using look up tables that define the area under a normal curve mean zero and variance 1 (alternatively the Excel™ function NORMSDIST can be used, e.g. NORMSDIST(-1) gives a value of 0.1587.

(2) EF = probability of the z-value occurring in that class interval × total number of readings

4. Calculate χ^2 (the Chi-squared statistic) using:

$$\chi^2 = \Sigma\ (OF\text{-}EF)^2 / EF$$

$$=[((6 - 0.237)^2/0.237) + ((10 - 0.032)^2/0.032) + ((11 – 0.017)^2/0.017) + ((5 - 0.001)^2/0.001))]$$

$$= 0.287$$

where the sum is over intervals, OF is the observed frequency and EF is the expected frequency.

5. The Chi-squared statistic (0.287) should then be compared with the Chi-squared function. This can be derived from look up tables or alternatively calculated using the Excel™ function CHIINV for m-3 degrees of freedom at the desired level of confidence, e.g. 0.05 (5% confidence); where m is the number of intervals. If the Excel™ function is used then:

Chi-squared function = CHIINV(0.05,1) = 3.84

(4 intervals ⇒ m-3 degrees of freedom = 1)

In this case the Chi-squared statistic (0.287) is smaller than the Chi-squared function (3.84) and, therefore, the hypothesis of normality can not be rejected at 5% confidence (i.e. there is a greater than 5% chance of this sample representing a Normal distribution). If the calculated Chi-squared statistic is greater than the Chi-squared function then less than 5% of the sample comes from a Normal distribution, and consequently it can be concluded this data set is not normally distributed. A final useful check is see to whether the histogram looks normal.

Use of Shapiro-Wilks to investigate normality

This statistical test is described in Gilbert, 1987 (p159). This method can only be used for the Normal distribution (or the Log-normal by taking logarithms)

Values of hydraulic conductivity have been obtained from field testing, and the assumption that these can be represented by a Log-normal distribution needs to be checked.

The data values are:

Hydraulic conductivity m/d	Natural logarithm
0.5	-0.69
1	0.00
5	1.61
10	2.30
18	2.89
27	3.30
38	3.64
40	3.69
55	4.01
65	4.17

1. Calculate $k = n/2$ if n is even or $(n\text{-}1)/2$ if n is odd (where n = number of samples = 10).

   Therefore $k = 5$ (i.e. n = 10 values / 2 = 5)

2. Look up in a table (Table A6, Gilbert, 1987) the coefficients a_1, a_2, …a_k. These values are:

   $a_1 = 0.5739$

   $a_2 = 0.3291$

   $a_3 = 0.2141$

   $a_4 = 0.1224$

   $a_5 = 0.0399$

3. Order the data (as above) so that $x_1 < x_2 < \ldots < x_n$

4. Calculate W (the Shapiro-Wilks statistic) according to

$$W = \frac{\left[\sum_{i=1}^{k} a_i (x_{n-i+1} - x_i) \right]^2}{\sum_{i=1}^{n} (x_i - \overline{x})^2}$$

$$W = (0.5739(4.17{-}{-}0.69+0.3291(4.01-0)+0.2141(3.69-1.61)+0.1224(3.64-2.30)+0.0399(3.3-2.89))^2/5.78$$

$$W = 0.87$$

The Shapiro-Wilks function at a 5% confidence level is 0.84 (Table A7, Gilbert, 1987). Since the calculated Shapiro-Wilks statistic of 0.87 is greater than this value, the hypothesis that the data are from a Log-normal distribution can not be rejected at the 5% confidence level i.e. there is less than 5% confidence the data is normally distributed.

Appendix B

Converting heterogeneous variability to uncertainty (upscaling)

For some cases, the key information may be the mean parameter value (or some other 'average') rather than the distribution in parameter values (that describe the heterogeneity of the system). Assessment of which is the relevant parameter describing the system behaviour should form part of the conceptual model.

For example, a number of measurements may have been made on clay content, and these measurements could be described by the mean and standard deviation of the data set. The conceptual model of the system may have indicated that the migration of a contaminant is a function of the total clay content along the pathway. In this case, the key parameter would be the total clay content in the entire volume of material. This is equivalent to the mean clay content per unit volume multiplied by the volume of the block of ground. The calculation needs the average clay content in this volume of rock, but the available measurements are from small core plugs. A process of upscaling is needed to derive the needed large scale value from the available small scale value. In this particular case, the arithmetic mean is a good estimator of the large scale value needed, so the upscaling process becomes that of determining this arithmetic mean. However, the mean estimated from a small number of samples is only an estimate of the actual mean.

In the clay content example, each clay analysis carried out by a laboratory assesses about 0.001 m^3 of material, so, for a 1 m unsaturated zone under a hectare of landfill, the total amount of clay that we want to estimate involves the equivalent of 10^7 analyses. This large number of potential samples may be treated as the population itself in this case

The source of the uncertainty is, therefore, the difference between our estimator of the population mean (the mean of the 20 or so analyses we have carried out) and the actual population mean of the whole block of ground. This difference would decrease as more measurements were obtained. It is important to note that more samples do not materially change the variance on the sample distribution. The uncertainty in the estimate of the population mean, estimated by the sample mean can be calculated from the sample variance. More samples do reduce the uncertainty of the population mean.

The mean total clay content of the volume of ground will be determined by a distribution very close to normal. This is because of the Central Limit Theorem (see Box B1) and the fact that *n* is very large.

Box B1. Central Limit Theorem.

Consider the arithmetic mean of a large number of independent measurements from a single distribution with population mean μ and variance σ^2. It can be proven mathematically that this mean is normally distributed with mean μ and variance σ^2/n as n tends to infinity – whatever the original distribution of measurements is. This is a powerful result and illustrates why the Normal distribution has been given the name 'Normal'.

The statistical theory for the uncertainty between our estimated mean and the true population mean is well established. The distribution of the mean of a set of *n* readings from a Normal distribution is the Students-*t* distribution with *n*-1 degrees of freedom. If *n* is greater than about 30, this distribution is close to normal. This allows us to specify uncertainty bounds on where the true population mean might be as described in Box B2.

Box B2 Example of estimation of the uncertainty on population mean

The uncertainty in the population mean can be estimated using the following equation:

$M \pm f_{(1-a/2)}\, \sigma/\sqrt{n}$

Where

M = mean

σ = standard deviation

n = number of samples

$f_{(1-a/2)}$ = value that cuts off a/2 of the upper tail of a Normal distribution

a = confidence limit (e.g. 95%)

Suppose we have 16 clay samples, assumed to be normally distributed, with mean (M) and standard deviation (σ). To obtain the 95-percentile confidence level on the population mean, we look up the 2.5 percent value from a Students-*t* table with 15 (n-1) degrees of freedom (Appendix D), which is 2.131 (the alternative is to use the Excel™ function TINV(0.05,15)) We select 2.5 percent because we require a two-tailed confidence. The 95-percent confidence range on the population mean is therefore:

$M \pm 2.131\, SD/\sqrt{n} = M \pm 0.533\, SD$

This calculation illustrates that there is considerably less uncertainty on the population mean than the equivalent 95-percentile spread on the actual sample data where $M \pm 1.96\, SD$. (for a Normal distribution 95% of values will fall within 1.96 *SD*).

The Students-*t* distribution is commonly not included in software packages and an alternative approach using the standard deviation of the sample data set to determine the uncertainty in the population mean is set out below, and is based on a conversion to a Normal distribution using the 95-percentile. The Students-*t* distribution is the recommended approach as given in Box B2.

The approach is:

1. Calculate the arithmetic mean of the samples (alternative approach is given for hydraulic conductivity below)

2. Determine the uncertainty in the estimator of the population mean using Table B1, as follows:

- For fewer than 3 samples use Triangular distribution with minimum and maximum values based on observed measurements (for the majority of problems this number of readings is insufficient and more data should be obtained);

- For more than 3 readings multiply the standard deviation by the factors given in Table B1 (i.e. uncertainty in the mean = $M \pm$ factor σ).

Table B1 Conversion of sample standard deviation to standard deviation of population mean

No of readings	Action	No of Readings	Action
1	Elicit max & min and use Triangular	9	Multiply SD by 0.39
2	Elicit max & min and use Triangular	10	Multiply SD by 0.36
3	Elicit max & min and use Triangular	11	Multiply SD by 0.34
4	Multiply SD by 0.81	12	Multiply SD by 0.32
5	Multiply SD by 0.63	13	Multiply SD by 0.31
6	Multiply SD by 0.54	14	Multiply SD by 0.29
7	Multiply SD by 0.47	15	Multiply SD by 0.28
8	Multiply SD by 0.43	16+	Multiply SD by $1/\sqrt{n}$

Note: SD is the standard deviation of the measurements

It is emphasised that upscaling by averaging should be used only where it can be shown that the average value of a parameter is the controlling factor in contaminant transport. If the system behaviour is controlled by a small heterogeneous part of the system, it should not be used as the upscaling is only describing the uncertainty in the mean, not the heterogeneity of the system.

Most parameters may be upscaled using the arithmetic mean of the data. The main exception is hydraulic conductivity. This is a much more complicated subject than it may appear and there is a large volume of literature on the subject. A good overview of upscaling hydraulic conductivity is given by Wen & Gómez-Hernández (1996) and in the text-book by Gelhar (1993). Some authors refer to the reduction in uncertainty that occurs with upscaling as *variance reduction*. It is mentioned in books by de Marsily (1986) and Vanmarcke (1983). Vanmarcke (1983) and Journel and Huijbregts (1978) also provide formulae and type curves for upscaling.

Upscaling the *transport porosity* is even more difficult. Transport porosity determines the relationship of Darcy velocity (as defined by the hydraulic conductivity) and the actual contaminant velocity. It is complicated by the need to decide whether it is the first arrival that is important or the mode of the arrival concentration (i.e. maximum concentration at the receptor).

A pragmatic approach is clearly appropriate. The statements below are based on simple analytic considerations of combining two blocks of different properties together and take no account of dispersion.

1. For variation parallel to the direction of flow, hydraulic conductivity is upscaled using the harmonic mean of the data and porosity with the arithmetic mean (to get travel time).

2. For variation perpendicular to the direction of flow, then hydraulic conductivity is upscaled using the arithmetic mean. Porosity is upscaled differently according to whether

it is the first arrival that counts or the mode of arrival time. For first arrival, the minimum porosity counts and for the mode of arrival time it is the arithmetic mean.

3. For variation in both directions in two dimensions (a quasi-random variability - the usual situation) it is generally considered that the hydraulic conductivity can be upscaled using the geometric mean of the data set (first shown for 2-D situations by Matheron, 1967) and the porosity using the arithmetic mean.

4. For variation in three dimensions, Gelhar (1993), has shown that the effective hydraulic conductivity is the geometric mean multiplied by the term $\exp(\sigma^2/6)$, where σ^2 is the variance of the distribution of the logarithm of the values of hydraulic conductivity.

If hydraulic conductivity is assumed to have a Log-normal distribution, then the factors given in Table B2 apply to the standard deviation of the logarithms. If Gelhar's formula is used to increase the geometric mean, then the original SD of the logarithms should be used

The assessment of whether and how data values need to be upscaled should form part of the conceptual model.

Appendix C

Bayesian Method

The Bayes Theorem provides a method of updating an initial probability distribution (referred to as a 'prior distribution'), using additional data (obtained from further field testing) to produce a revised distribution (referred to as a 'posterior distribution'). For example, a PDF for effective porosity might have been obtained from expert opinion and then a tracer test performed giving a single value of that effective porosity. We can update the 'prior distribution' to obtain a 'posterior distribution' by using Bayes' probability theorem:

$$P(y|x)=P(x|y)\,P(y)\,/\,P(x).$$

where P(y|x) is the revised probability of the parameter y after the data x have been taken into account,

P(x|y) is the conditional probability of the data x given the prior distribution, y

P(y) is the (prior) probability of the parameter before any data are collected

P(x) is the probability of the data (this normalises the function).

An example of the Bayes Method is given below:

EXAMPLE: Bayes' theorem applied to establishing a transmissivity distribution

Suppose we wish to establish a probability distribution for transmissivity in an area. Based on information on aquifer characteristics from an adjacent catchment, we choose to assume that the logarithm of transmissivity is normally distributed and that the prior mean (M) and variance (S^2) of this parameter can be described as follows:

- Prior Mean (M) = log (1000) = 3
- Prior Variance (S^2) = 1

Three pumping tests (n =3) are performed in the study area and, the mean (m) and variance (s^2) of the logarithms of this sample have been estimated assuming a Normal distribution for the logarithms as follows:

- Sample mean (m) = log (10000) = 4
- Sample variance (s^2) = 3

These two sets of information can be combined to give a revised or posterior probability estimate using the following formulae:

Posterior mean = $m = \dfrac{M/S^2 + nm/s^2}{1/S^2 + n/s^2} = \dfrac{3/1 + 3\times 4/3}{1/1 + 3/3} = 3.5$

and posterior variance (s^2) determined from:

$\dfrac{1}{s^2} = \dfrac{1}{S^2} + \dfrac{n}{s^2} = \dfrac{1}{1} + \dfrac{3}{3} = 2$ so $s^2 = 0.5$

In summary the transmissivity distribution can be described by Log-normal(m=3.5, s^2=0.5) which has the geometric mean 3162 m^2/d and 68% of data between 621 and 16 106 m^2/d.

The change in the calculated mean and variance as a result of applying the Bayes methodology is summarised below:

	Mean Transmissivity (m^2/d)	**Variance**
Initial or prior parameters	Log(1000) =3	1
Parameters of test data	Log(10000) = 4	3
Revised or posterior parameters	Log(3162) = 3.5	0.5

Note that, as expected, the revised distribution has a mean that is intermediate between the prior mean and the mean of the measurements and a variance that is smaller than either the prior or measured values.

The decision as to how to incorporate additional data will be dependent on our conceptual model, and the importance that should be attached to the original distribution which may take more account of our conceptual understanding of the system behaviour.

Further examples of the Bayes Method are given in Gilbert, 1987.