Addressing soil contamination in IPPC Site Baseline and Site Closure reports

G. Lethbridge & M.F. Wenham
Shell Global Solutions
Addressing soil contamination in IPPC Site Baseline and Site Closure reports

Summary:

The Environment Agencies in the UK are proposing to take a different approach to new contamination of soil and groundwater that occurs under PPC, compared to that taken to historic contamination (pre-PPC) under Part IIa (EPA, 1990). Ideally they would like to see all the contamination that took place under PPC removed, or failing that to reduce the levels to As Low as Reasonably Practical (ALARP) using Best Available Technology (BAT). This is in contrast to the risk-based approach to historic contamination allowed under Part IIa (EPA, 1990).

The argument against a fully risk-based approach to contamination taking place under PPC is to prevent further degradation of soil and groundwater quality and avoid the danger of PPC becoming a “license to pollute”. There are significant cost implications for remediation of contamination occurring under the proposed PPC regime. The additional costs over and above those that would be incurred for a fully risk-based solution will deliver no additional real environmental benefit.

Given the massive heterogeneity of contaminant distribution in soil and the likely high baseline levels of existing (pre-PPC) contamination, site owners are concerned about the very real dangers of pre-PPC contamination, which can be managed cost-effectively under the risk-based Part IIa (EPA, 1990) legislation, being mistaken for contamination which took place during the lifetime of a PPC permit. The only way to demonstrate with a known degree of confidence whether any new contamination has been added during the lifetime of a PPC permit is to use statistics.

The statistical methods that have been evaluated in this project require sampling densities that are 1-2 orders of magnitude greater than those currently used for site investigations for historic contamination managed under Part IIa (EPA, 1990). This would result in the cost of investigation being equal to or greater than the current cost of remediation in many cases.

In the final analysis, on the grounds of practicality, it is difficult to envisage anything other than a risk-based approach to contamination occurring under PPC.

This report recommends:

- Further dialogue with the Agencies to attempt to align the approach to soil and groundwater contamination under PPC with that under Part IIa (EPA, 1990)
- A greater focus on the site’s Environmental Management System (EMS), quality of product containment, spill/leak prevention measures, spill reporting and corrective action recording as a means of demonstrating whether or not significant new contamination has taken place during the lifetime of a PPC permit.
Environmental forensic measurements of contamination (e.g. fingerprints of complex mixtures, ratios of recalcitrant components in complex mixtures, isotope ratios) may be useful tools where it does become necessary to distinguish between pre-PPC and PPC contamination.

Following the completion of this study, the Environment Agency have recognised the difficulties for site owners of taking a statistical approach to comparing soil contamination data in PPC Baseline and Site Closure reports. They have indicated that they are willing to consider the other criteria recommended in this report as part of a “weight of evidence” approach to judging whether or not significant new contamination has taken place during the lifetime of a PPC permit.
CONTENTS

1. Introduction 4
   1.1 Background 4
   1.2 Concerns over the proposed regulatory approach to soil and groundwater contamination under PPC 5
   1.3 Objectives 6

2. The challenge of the heterogeneous distribution of soil contamination 7

3. Evaluation of a statistical approach to soil contamination decision making under PPC 12
   3.1 Methods 12
      3.1.1 Basic assumptions 12
      3.1.2 Matrices of simulations 14
      3.1.3 How much contamination was added? 15
      3.1.4 Sampling designs 16
      3.1.5 Measures of increased contamination 19
   3.2 Results 20
      3.2.1 Paired t-test with raw values of TPH and null hypothesis of no change in contamination during the PPC period 20
      3.2.2 Paired t-test with logged values of TPH and null hypothesis of no change in contamination during the PPC period 20
      3.2.3 What happens if the locations of the pre-PPC baseline samples are lost? 22
      3.2.4 Effects on detection rate of sampling design, measurement s.d and changes in post/pre-PPC contamination ratio 24
         3.2.4.1 Paired t-tests – existing contamination at site 24
         3.2.4.2 Paired t-tests – greenfield site 24
      3.2.5 Calculations of differences in contamination between pre- and post sampling that could be detected as a function of sampling density and power 25
   3.3 Contamination heterogeneity issues 27
      3.3.1 Point comparisons of differences with LSD on log-scale – sites with pre-existing contamination 27
      3.3.2 Point comparisons of differences with LSD on log scale – greenfield sites 27
   3.4 Conclusions from the statistics 28

4. Alternative approaches to soil contamination decision making under PPC 33
   4.1 Environmental Management Systems 33
   4.2 Environmental forensics 34

5. Overall conclusions 36

6. References 38

APPENDICES 39
Addressing soil contamination in IPPC Site Baseline and Site Closure reports

1. INTRODUCTION

1.1. Background

The European Union (EU) Integrated Pollution Prevention and Control (IPPC) Directive is designed to minimise future environmental impacts arising from industrial and certain other activities which have the potential to contaminate the environment. In the UK, the IPPC Directive is implemented via so-called PPC Regulations such as the Pollution Prevention and Control Act 1999 and the Pollution Prevention and Control (England and Wales) Regulations 2000. The PPC regime uses a permitting system as a means of enforcing a high degree of environmental protection.

An application for a permit to operate a process covered by the PPC regulations requires that a Site Baseline Condition Report be submitted which establishes the baseline conditions of soil and groundwater quality at the site at the start of operations under the new regime. In the future, when the PPC licensed process ceases and the facility is closed, the operator must submit a Site Closure Report describing the soil and groundwater quality to determine whether any soil and/or groundwater contamination has taken place whilst the permitted process has been in operation. If any contamination has taken place during the lifetime of the PPC permit, the site must be returned to a “satisfactory state” before the permit can be surrendered.

The Environment Agency for England and Wales (EA) and the Scottish Environmental Protection Agency (SEPA) are proposing a stringent interpretation of “satisfactory state” to mean removal of all contamination of soil and/or groundwater that took place during the lifetime of the PPC permit, regardless of whether it poses a risk to human health or the environment. The reason for this stringent approach is that the EA and SEPA wish to prevent future “environmental degradation” and feel that a risk-based approach to remediation of contamination occurring during the operation of a PPC permitted operation could become a “license to pollute”.

Under these circumstances, new contamination of soil and/or groundwater arising during the operation of a PPC permitted activity would be treated differently to historic contamination already in the ground which is subject to the contaminated land regulations in Part IIa of the Environmental Protection Act 1990 which are risk-based. In Part IIa (EPA 1990), the need for and extent of any remediation is determined by the level of risk to human health or the environment posed by the contamination.
1.2 Concerns over the proposed regulatory approach to soil and groundwater contamination under IPPC

The proposed regulatory approach to new soil and groundwater contamination under the PPC regime raises a number of fundamental concerns:

- Under most circumstances it is technically impossible to remove all the contamination beneath sites (other than by excavation and landfilling or incineration) due to mass transfer limitations and very slow desorption kinetics. This is one of the reasons (alongside others such as cost:benefit considerations) why historical contamination is dealt with on a risk basis under Part IIa (EPA, 1990). Disposal of contaminated soil to landfill is actively being discouraged by regulatory authorities on the grounds of sustainability. From July 2004, the EU Landfill Directive requires that all wastes classified as hazardous (which includes soils containing chemical contamination above certain thresholds) be treated to reduce chemical concentrations below certain levels prior disposal in a landfill. However, none of the treatment processes are capable of removing all the contamination as required under PPC.

- PPC could result in significant expenditure by site operators remediating contamination that poses no risks to human health and the environment; i.e. there may be no environmental benefit for the expenditure.

- What constitutes a site baseline for a particular chemical or group of chemicals against which decisions will be made as to whether new contamination has taken place during the lifetime of the PPC permit? The EA and SEPA have both produced guidance on Site Baseline Condition Reports to be submitted as part of a PPC permit application which fails to address, other than in very broad terms, the setting of baselines for soil and groundwater quality.
  - Given the massive heterogeneity of contaminant distribution patterns in the ground, it is unlikely to be a single concentration.
  - The large uncertainties surrounding the estimation of contaminant mass in the soil from a number of point measurements (distributed both laterally and vertically) of concentration means that a single mass per unit area measurement would be highly unreliable.
  - A site baseline is more likely to be a range of values representing different locations at the site.
  - Contamination is distributed in three dimensions. How will the vertical distribution of contamination be addressed in the site baseline? Will there be a different baseline range for each depth slice and if so, how thick will each depth slice be?

- How are data on soil and groundwater quality in the Site Closure Report to be compared to those in the Site Baseline Condition Report to decide whether contamination of the ground took place during the lifetime of the PPC permit? It is crucial to know this before embarking on the site investigation for the Site Baseline Condition Report, since it will determine the type, location and number of sampling points.

- The risk of false positives; i.e. the site is deemed to have been contaminated during the lifetime of the PPC permit when in fact it was not. This could arise when the contamination detected during the site investigation for the Site Closure Report had actually been present prior to the granting of the PPC permit, but had not been detected during the site investigation for the Site Baseline Condition Report. The financial consequences of this are likely to be
severe in that contamination which should be dealt with under the risk-based Part IIa (EPA, 1990) regime will now have to be dealt with under PPC. The cost of remediation under PPC is likely to be far greater than under Part IIa (EPA, 1990). Due to the highly heterogeneous nature of soil contamination, the challenge of distinguishing between pre- and post-PPC contamination is not a trivial task (see Section 2). False positives could also arise in other ways such as:

- On-site migration of contamination from a neighbouring property up hydraulic gradient of the site.
- Atmospheric deposition of contamination arising from point sources (e.g. incinerators, smelters etc) down wind of the site or from more widespread diffuse sources (e.g. motor vehicle exhausts). For example, it is well known that heavy metal and polyaromatic hydrocarbon (PAHs) contamination of soil can arise from atmospheric deposition of particulates and that water-soluble contaminants can be washed out of the atmosphere during rainfall.
- Biological or chemical transformation of contamination already in the ground prior to PPC to new species that were not present or detected when the Site Baseline Condition Report was submitted with the PPC permit application.
- Process areas in large manufacturing complexes (e.g. refineries and chemicals plants), which have a high likelihood of soil and groundwater contamination due to the nature of the facilities and their operation, represent a particularly difficult challenge for Site Baseline Condition setting. It will be impossible to carry out an intrusive investigation underneath such plant until it is decommissioned and dismantled and highly unlikely that an intrusive investigation will be possible even close to the plant due to safety considerations (risk of electrocution or fire hazard from underground services and pipelines) whilst it is in operation which is often continuous.

1.3. Objectives

The objectives of this report are to:

- Highlight concerns of the regulated industries over the proposed different approach to future soil and groundwater contamination arising under the PPC regime compared to historical contamination under Part IIa (EPA, 1990).
- Evaluate potential approaches that may be used to determine whether new contamination of the ground has taken place over the lifetime of a PPC permit.
2. THE CHALLENGE OF THE HETEROGENEOUS DISTRIBUTION OF SOIL CONTAMINATION

The key challenge to determining whether new contamination of a site has taken place under the lifetime of a PPC permit is that of the highly heterogeneous distribution of chemicals, of either natural or anthropogenic origin, in the ground. Very few sites are likely to have non-detect baselines for the following reasons:

- Many PPC permitted activities will take place on ground that is already contaminated from decades of industrial activity. In many cases, the chemicals used in the PPC permitted activity will be the same as those that have been handled on site prior to the activities coming under the PPC regime and the same as those already in the ground. This historical contamination may or may not have been dealt with under Part IIa (EPA, 1990). Either way, the baseline for these chemicals will not be non-detect and could be considerable.

- Many industrial sites in the UK are located on made ground (fill) that can vary in thickness from as little as 0.5 metres up to 4 metres. Made ground can be composed of: industrial wastes such as ash, clinker and slag, river, harbour and estuary dredgings, building waste such as hard core and sub-soil and other types of waste materials. Many of these components are high in heavy metals and by the very nature of made ground their distribution is highly heterogeneous. Heavy metal contamination originating from made ground should rarely be an issue under Part IIa (EPA, 1990), because they are rarely mobile and hence normally pose negligible risk to human health and the environment (absence of completed source – pathway – receptor linkages).

- Even green-field sites are unlikely to have non-detect baselines for certain potential contaminants such as heavy metals and hydrocarbons. Most soils contain some level of TPH (total petroleum hydrocarbon) positive material. The higher the organic carbon content, the higher the background value. For example, peat can contain several hundred mg/kg of native TPH positive material of plant origin.

Heterogeneous distribution of contamination occurs at both the small and large scale. Figure 2.1 illustrates the scale of the problem. It shows the lateral and vertical distribution of TPH in a medium-fine sandy soil (dredged beach material) in a 15 metre x 15 metre area of a former tank farm used for storing crude oil. The whole area was hand-augered on 2.5 metre centres and TPH measured at 0-25 cm, 25-50 cm and 50-75 cm below ground level.

The data are summarised in Table 2.1.
Table 2.1 Heterogeneity of TPH in a medium fine sandy soil beneath a former tank farm storing crude oil

<table>
<thead>
<tr>
<th>Data points</th>
<th>0 – 25 cm</th>
<th>25 – 50 cm</th>
<th>50-75 cm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum value</td>
<td>124 mg/kg</td>
<td>58 mg/kg</td>
<td>50 mg/kg</td>
</tr>
<tr>
<td>Maximum value</td>
<td>17,091 mg/kg</td>
<td>16,462 mg/kg</td>
<td>2,353 mg/kg</td>
</tr>
<tr>
<td>Mean</td>
<td>6,221 mg/kg</td>
<td>2,392 mg/kg</td>
<td>479 mg/kg</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>5,960 mg/kg</td>
<td>3,844 mg/kg</td>
<td>626 mg/kg</td>
</tr>
<tr>
<td>Coefficient of variation</td>
<td>96%</td>
<td>161%</td>
<td>131%</td>
</tr>
</tbody>
</table>

The distribution of TPH in soil in three other 15 m x 15 m areas was assessed by the same approach and equally high variability observed.

**Figure 2.1. Distribution of TPH (mg/kg) in a medium-fine sandy soil beneath a former tank farm storing crude oil**

To facilitate visualisation of this heterogeneity, trial pits (1 m X 1 m X 1m) were dug in the same areas (Figure 2.2). The natural (uncontaminated) colour of the sandy soil is off-white, so the crude oil contamination shows up really well against this background. The photographs in Figure 2.2 demonstrate that even in a relatively homogeneous medium-fine sandy soil there are small-scale preferential flow paths, which result in highly contaminated regions being adjacent to visually clean regions.

With this high degree of heterogeneity, it is clear that the only way to definitively determine whether or not new contamination of the ground has taken place during the
lifetime of the PPC permit with a known level of confidence is to use some form of statistical test to compare site investigation data from the Site Baseline Condition Report with that from the Site Closure Report. Statistical methods are rarely used in site investigations for historical contamination under Part IIa (EPA, 1990), on the grounds that the increased costs are difficult to justify in cost:benefit terms when the objective is to manage risk, rather than characterise sources in great detail.

**Figure 2.2.** One metre vertical sections through a medium fine sandy soil beneath a former tank farm storing crude oil
The question is, how can this massive heterogeneity be handled in any statistical comparison of data from the Site Baseline Condition Report with that from the Site Closure Report to avoid false positives (site judged to have been contaminated during the lifetime of the PPC permit when it has not), or false negatives (site judged not to have been contaminated during the lifetime of the PPC permit when it has been)? Theoretically, the closer together two samples are taken, the lower the variability between them would be expected to be.

To test this hypothesis in practice at the former tank farm locations, a series of seven soil borings were hand-augered as close as practically possible (15 cm apart) and TPH measurements made at the same depths as previously (Figure 2.3). In general, the inter sample variability was less between these samples, than those collected over the whole 15 m x 15 m area. The coefficient of variation was reduced from > 100% to < 50% (Table 2.2). The limitations on how close two soil borings could be made to each other, meant that there were locations where the variability was not significantly reduced, but in general, sampling closer together reduced variability.

Figure 2.3. Small-scale variability in TPH (mg/kg) in a medium-fine sandy soil beneath a former tank farm storing crude oil

Table 2.2. Example of small scale variability in TPH (mg/kg) in a medium-fine sandy soil beneath a former tank farm storing crude oil

<table>
<thead>
<tr>
<th>Data points</th>
<th>0 – 25 cm</th>
<th>25 – 50 cm</th>
<th>50-75 cm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data points</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Minimum value</td>
<td>7,060 mg/kg</td>
<td>3,158 mg/kg</td>
<td>313 mg/kg</td>
</tr>
<tr>
<td>Maximum value</td>
<td>12,104 mg/kg</td>
<td>6,215 mg/kg</td>
<td>697 mg/kg</td>
</tr>
<tr>
<td>Mean</td>
<td>9,536 mg/kg</td>
<td>4,622 mg/kg</td>
<td>558 mg/kg</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>1,873 mg/kg</td>
<td>1,281 mg/kg</td>
<td>131 mg/kg</td>
</tr>
<tr>
<td>Coefficient of variation</td>
<td>20%</td>
<td>28%</td>
<td>23%</td>
</tr>
</tbody>
</table>
The inherent variability in the soil becomes a problem when differences in contamination between locations or over time need to be detected. The background variability of the soil can be viewed as the “noise”, whilst the difference in contamination over time (between the Site Baseline Condition investigation and the Site Closure investigation) is the “signal”. The larger the noise, the larger the minimum detectable signal is. The inherent variability in soil contamination levels is made up of a component due to soil and contamination heterogeneity and a component due to measurement/sampling variability. The variability between samples resulting from soil heterogeneity increases as the samples are taken further apart. Two neighbouring analysis samples taken from the same core will differ by an amount proportional to only the measurement/sampling variability. If taking samples in as near as possible the same location could reduce the variability, our ability to detect contamination during the IPPC period would improve.

Due to the absence of detailed measurements of hydrocarbon contamination over periods similar to those over which the PPC would operate, a decision was taken to simulate the accumulation of hydrocarbon contamination. These simulated hydrocarbon levels were used to test different sampling and statistical analysis strategies. The results of this exercise are described in the following sections.
3. EVALUATION OF A STATISTICAL APPROACH TO SOIL CONTAMINATION DECISION MAKING UNDER PPC

3.1. Methods

After reviewing various site investigation reports it was decided that the best approach was to simulate a site using MATLAB software in order to create a model site that had all the necessary features and provided for the most flexible approach.

MATLAB software and the statistics toolbox provide a high-performance numerical computing environment where statistical simulations can be quickly and easily carried out. MATLAB itself provides a core programming language so that commands can be automated, while the statistics toolbox provides standard statistical functions.

Since different areas of an industrial site vary in the chances of them being contaminated and the extent to which they are contaminated according to the activities carried out, separate simulations were carried out for areas of low (e.g. offices and car/lorry park), medium (e.g. pumping stations, oil-water separator) and high (e.g. process areas, tank farm, product loading and unloading areas) risk of contamination.

Although the contamination used in the model is petroleum hydrocarbon (TPH), the principles established apply to any type of soil contamination.

3.1.1 Basic assumptions

The following basic assumptions were used in the simulations.

- For the period before PPC, the expected number of spills per year over a 100m x 100m area as a function of risk (low, medium and high) and spill size (small, medium and large) were fixed at the following levels:

<table>
<thead>
<tr>
<th>Spill frequencies</th>
<th>Risk Category</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
</tr>
<tr>
<td>Small – 10m²</td>
<td>2</td>
</tr>
<tr>
<td>Medium – 100m²</td>
<td>0</td>
</tr>
<tr>
<td>Large – 1000m²</td>
<td>0</td>
</tr>
</tbody>
</table>

- Spills were constructed by superimposing circles of contamination in layers. The smallest spills consisted of one single layer whereas the largest spills had three layers. Each layer added TPH contamination to the area over the zone covered by the circle. This gave the spills an approximate concentration gradient. The dimensions of the layers are shown in the next table.

<table>
<thead>
<tr>
<th>Radii of spill layers in metres</th>
<th>1st layer</th>
<th>2nd layer</th>
<th>3rd layer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small – 10m²</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Medium – 100m²</td>
<td>5</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Large – 1000m²</td>
<td>17</td>
<td>11</td>
<td>5</td>
</tr>
</tbody>
</table>
A diagram showing the TPH concentration in small, medium and large spills is shown below. The spills are not perfectly circular because of the resolution of the grid.

Illustration of three single spills showing concentration contours:
When many spills accumulate over years of simulated time the resulting contamination appears as below, the scale units for TPH concentration are mg/kg and spills were allowed to accumulate for 40 years of operation prior to PPC.

The simulated contamination produced spill maps which were reasonably representative of contamination at real sites.

- The scaling of the spill maps was fixed for all the current simulations. However, the ability to change the maximum concentration was provided in case simulations of contamination by other chemicals are required in the future.
- Measured values below 100mg/kg were below the detection limits of the analysis method and these were therefore set at 100mg/kg.

### 3.1.2 Matrices of simulation runs

It was assumed that the PPC permit was in operation for 20 years. (NB. The post-IPPC time will affect the detection rates – the longer contamination is allowed to accumulate the easier it will be to detect.) It is reasonable to assume that the spill frequency during the lifetime of the PPC permit will be considerably less than that prior to PPC as the latter covered a period of time when contamination prevention measures were not as good as they are now. Three scenarios were considered where the spill frequency during the lifetime of the PPC permit was reduced by 75%, 90% and 100% (i.e. no spills) compared to the pre-PPC frequency.

Several sets of simulations were carried out, in order to explore various effects. All the simulations were repeated for low, medium and high contamination risk areas. The sets of simulations can be described as follows:
Ten runs over a wide range of sampling densities at constant Post/Pre-PPC contamination ratio (0.25) and constant measurement s.d. (0.25).
- Sampling density levels: 0.0008, 0.0012, 0.003, 0.005, 0.02, 0.04, 0.06, 0.08, 0.1 and 0.125 per m².

24 runs over four sampling densities, three levels of Post/Pre-PPC contamination ratio and two levels of measurement s.d.
- Sampling density levels used: 0.0008, 0.0012, 0.003 and 0.005 per m².
- Post/Pre-PPC contamination ratios used: 0.25, 0.1 and 0, which represent reductions in spill frequency of 75%, 90% and 100% respectively.
- Measurement error s.d.’s of 0.25 and 0.5 were used. These were derived from real site investigation data.

Three runs at a constant sampling density of 0.02 per m² and zero measurement error at each of the three levels of Post/Pre-PPC contamination ratio.

The lowest density was chosen to match approximately the density range of 0.0006-0.0009 m² in typical use for site investigations for historic contamination under Part IIa. The Post/Pre-PPC contamination ratio applies to the annual spill rates before and after the start of the PPC period. The figure of 0.25 for Post/Pre-PPC contamination means that the post-PPC contamination rate has been reduced by 75% compared to that before the PPC period. Likewise, a figure of 0.1 is equivalent to a reduction of 90% and a zero means there has been no contamination during the lifetime of the PPC permit.

TPH levels in soil are typically lognormally distributed (Wenham, 2002). The standard deviation is therefore calculated on a log scale based on data from a site where sets of six measurements of soil TPH content were made from a 15cm radius around one central measurement. The s.d. value of 0.25 is the standard deviation of the natural logarithms of these seven measurements. The higher level of s.d. was chosen arbitrarily to be twice the experimental level. The three runs at 0.02 samples per m² form a subsidiary matrix with a high sampling density and zero measurement error to simulate “perfect” sampling. A density of 0.02 should, in theory, give a 100% chance of detecting a 5m-diameter spill. The very high sampling densities (up to 0.123 per m²) were included to enable 2m diameter spills to be detected. (See Appendix A for a table of theoretical spill detection probabilities as a function of spill size, sampling density and sampling design).

For the operation of a PPC permitted process on a previously greenfield site, a reduced matrix of twelve runs was used, i.e. using just two sampling densities (0.0012 and 0.005 per m²), three levels of added contamination and the same two levels of measurement s.d. as when pre-existing contamination was present. The absolute levels of added contamination were kept the same as when pre-existing contamination was present.

### 3.1.3 How much contamination was added?
It is possible to quantify how much contamination was added during the simulations of the PPC period by examining the added contamination at each of the 10,000 locations in the experimental 100x100 grid. These quantities apply both to sites where there was pre-existing contamination and to greenfield sites.


<table>
<thead>
<tr>
<th>Post/Pre spill ratio</th>
<th>Low risk</th>
<th>Medium Risk</th>
<th>High Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average contamination increase over whole area, mg/kg TPH</td>
<td>29.6</td>
<td>368.9</td>
</tr>
<tr>
<td>0.25</td>
<td>Number of locations contaminated out of 10,000</td>
<td>126</td>
<td>1330</td>
</tr>
<tr>
<td></td>
<td>Average contamination increase over contaminated locations only, mg/kg TPH</td>
<td>2349</td>
<td>2774</td>
</tr>
<tr>
<td>0.1</td>
<td>Average contamination increase over whole area, mg/kg TPH</td>
<td>12</td>
<td>151</td>
</tr>
<tr>
<td></td>
<td>Number of locations contaminated out of 10,000</td>
<td>52</td>
<td>562</td>
</tr>
<tr>
<td></td>
<td>Average contamination increase over contaminated locations only, mg/kg TPH</td>
<td>2307</td>
<td>2686</td>
</tr>
</tbody>
</table>

The change in average contamination over the whole area changes in proportion to the change in Post/Pre-PPC spill ratio. The change in the number of contaminated locations and the average contamination increase per contaminated location do not follow a simple pattern.

### 3.1.4 Sampling designs

Five designs of non-targeted sampling schemes were chosen: simple random sampling, stratified random sampling, square grid sampling, herringbone sampling and triangular grid sampling. They are illustrated in the following diagrams for three of the densities.
Simple random sampling

Stratified random sampling
Square grid sampling

Herringbone grid sampling – shift = ¼ grid square
The different sampling designs were chosen to explore any differences in the ability to detect contamination that might result. Herringbone sampling (ref CLR4) is believed to offer advantages in the detection of specific spill shapes, e.g. elliptical spills. In theory, triangular grid sampling offers improved efficiency in detecting circular spills, since to detect a spill of a particular radius, the sampling density using a triangular grid is only half that required if a square grid is used (see Appendix A).

3.1.5 Measures of Increased Contamination

Various methods of detecting increased contamination over the range of conditions were evaluated:

1) The paired t-test using raw values of TPH and a null hypothesis\(^1\) of no change in contamination during the IPPC period (i.e. the IPPC draft document approach\(^2\))

2) The paired t-test using logs of values of TPH and a null hypothesis of no change in contamination during the IPPC period (i.e. a variant of the IPPC draft document approach to take account of the lognormal distribution of soil contamination values)

---

\(^1\) See Appendix B for an explanation and illustration of hypothesis testing
\(^2\) Page 31 of the IPPC document gives an example where the paired t-test is used, with a null hypothesis of no increase in contamination during the IPPC period. A two-tailed test is used which implies that the alternative hypothesis is that an increase or decrease in contamination has occurred during the IPPC period.
3) The two-sample t-test on logged and raw values of TPH. This simulates the situation where the locations of the original samples have been lost.

4) The number of point comparisons whose post-IPPC minus pre-IPPC contamination on a log scale is greater than the Least Significant Difference on the log scale. [The Least Significant Difference of the smallest difference between two single measurements that could not have arisen by chance, given the reproducibility of the measurement and sampling procedure. A 5% significance level was used in the calculations.]

3.2. Results

We’ll look at the results for each of the measures of increased contamination in turn.

3.2.1 Paired t-test with raw values of TPH and null hypothesis of no change in contamination during the PPC period.

Post/pre-PPC contamination ratio of 0.25
Increased contamination is detected 80% of the time when the sampling density is more than 0.005 per m² for high risk, just below 0.02 per m² for medium risk and about 0.125 per m² for low risk areas. The ability to detect added contamination when a null hypothesis of “added contamination equals zero” is used is related to the amount of extra contamination added during the simulation. The very small amount of contamination added to the low risk area requires a high sampling density before it can be detected, whereas the very large amount added to the high risk area is detected more easily. However, the sampling density required to detect the contamination added to the high risk area is very much higher (order of magnitude) than the sampling densities typically employed at the moment for site investigations for historic contamination. At the current typical sampling density of 0.0008 per m² the detection rate for contamination added at a level of over 4000mg/kg is only around 11%.

The method advocated in the draft PPC document applied to data obtained at sampling densities typical of current site investigations is likely to miss contamination added during the IPPC period almost 90% of the time.

To increase the detection rate using the draft PPC method to around 80%, the sampling density needs to be increased by a factor of about 12, to approximately 0.01 per m².

Note: The paired t-test using raw values was not simulated on areas that had been greenfield sites before PPC.

3.2.2 Paired t-test with logged values of TPH and null hypothesis of no change in contamination during the IPPC period.

Post/pre-PPC contamination ratio of 0.25
Contamination is detected 80% of the time when the sampling density is more than 0.005 per m² for high risk, just above 0.01 per m² for medium risk and about 0.04 per m² for low risk areas. These sampling densities are generally lower than when the method is applied using raw TPH values. The very small amount of contamination added to the low risk area still requires a relatively high sampling density before it can
be detected. At the current typical sampling density of 0.0008 per m² the detection rate for contamination added at a level of over 4000mg/kg is only around 12% even when log-TPH is used.

The method advocated in the draft PPC document applied to logs of TPH data obtained at sampling densities typical of current site investigations is likely to miss contamination added during the PPC period almost 90% of the time.

To increase the detection rate using log-TPH values and the draft PPC method to around 80%, the sampling density needs to be increased by about a factor of 9 to approximately 0.007 per m².

Note: The paired t-test using log-TPH values was not simulated on areas that had been greenfield sites before PPC.

Taking logs of TPH values improves contamination detection rates, so a lower sampling density could be used.

The following three graphs show the different contamination detection rates for the three risk areas over the full sampling density range. These data were used to make the preceding comments. The null hypothesis is that, on average, there has been no contamination added. This is not the case for each of the three risk levels, so the null hypothesis should, ideally, be rejected. We therefore accept the alternative hypothesis that the soil is contaminated and theoretically requires remediation under PPC. The vertical axis is the number of times this decision is made per 1000 realisations.

[H₁: Δ > 0 in the diagram on the left means that the alternative hypothesis used was that the added contamination during PPC was greater than zero.]
3.2.3 What happens if the locations of pre-PPC baseline samples are lost?

The paired t-test method relies on being able to take a sample in the same location (or as close as practically possible) after the PPC period as before it. If this were not possible, a two-sample t-test would have to be used instead. This test compares estimates of the average contamination over the whole area before and after the PPC period. (The paired t-test compares the average of the changes in contamination at the sampling points with a specific value.)

When the average at a single point is calculated, the spatial variability between points is removed and the variability in the measurements is composed only of measurement error and local sampling variability since it is unlikely for the second sample to be taken in exactly the same location years later.

When the average contamination over the whole area is estimated the standard error of this estimate will be large, since spatial variability is included. The two-sample t-test will have less power than the paired t-test since the standard error of the
difference between the whole area estimates is inflated compared to the standard error of the average of the paired differences.

Additional simulations were carried out to quantify the expected drop in power. They all used a Post/Pre-PPC contamination ratio of 0.25 and a measurement s.d. of 0.25. The following graphs show that the results followed the expected trend since the detection rate using the two-sample t-test is always much lower than for the paired t-test.
Every effort should be made to preserve the sampling locations, since the power of any statistical test will be much higher if a paired t-test can be used.

### 3.2.4 Effects on detection rate of sampling design, measurement s.d. and changes in Post/Pre IPPC contamination ratio

These effects were explored at the low end of the sampling density range (0.0008 to 0.005 per m$^2$) and at 0.02 per m$^2$, using twenty-four runs at three levels of Post/Pre-PPC contamination ratio (0.25, 0.1 and 0) and two levels of measurement s.d. (0.25 and 0.5). The t-tests were done exclusively on the logarithms of the TPH values, using a null hypothesis of no added contamination.

Graphs of these results are shown in Appendix C.

#### 3.2.4.1 Paired t-tests – existing contamination at site

**No new contamination post IPPC**

We would expect to detect a significant difference 5% of the time even when there has been no new contamination. This is the same as saying that the chance of a false positive is 5%. This is an automatic feature of using a significance level of 5% for the t-test. This was the case for the high and medium risks areas, irrespective of measurement s.d and sampling pattern.

This did not happen for low risk areas – the detection rate was much lower, <2%. This is most likely due to there being very little contamination there to detect and what is there has a small radius (see later comment re. paired t-test results for the greenfield site). Many “errors” were generated with the t-test in the low-risk case because the values before and after the start of PPC were both less than the detection limit. These were counted as “no contamination detected”. Graphs of these results are on the top of the first three pages of Appendix C.

**Contamination added post IPPC**

Detection is always better where the sampling and measurement s.d. is lower.
Detection rates are less than 20% for sampling densities around 0.001/m², irrespective of the risk level and even where the annual contamination rate is 25% of the pre-PPC rate.

Higher sampling densities give better detection rates – up to 65% for high contamination risk areas and the higher rate of post-PPC contamination. There is no effect of sampling pattern. Graphs of these results are on the first three pages of Appendix C.

3.2.4.2 Paired t-tests – greenfield site
Each of the three contamination risk (low, medium and high) areas all started with the same small amount of background TPH contamination. Most locations had greenfield contamination levels less than 100mg/kg and fewer than 5% had maximum levels of 200mg/kg. The absolute amounts of contamination added during the PPC period were made the same as when contamination had already been present.

For low risk areas there was no appreciable difference in detection rate between sites with existing contamination and greenfield sites. Both gave detection rates of less than about 3%. This is probably because the pre-existing contamination on a low risk area and a greenfield site were similar.

The “perfect” sampling density gave a detection rate of only 25%, which was lower than expected, until you consider that the “perfect” sampling density should have given 100% detection for spills of 5m in diameter. The spills in the low risk site were actually only 2m in diameter, so we should have expected this low detection rate.

For medium and high risk areas, the detection rate was much higher for greenfield sites compared to sites with pre-existing contamination. In fact, the density of 0.0012m² with the lower s.d. gave nearly 100% detection for high risk areas on a greenfield site, whereas the corresponding figure for a site with pre-existing contamination was 8%. The chance of a false positive for medium and high risk areas was less than 5% for all sampling densities investigated.

The detection rate improved with smaller s.d and higher sampling density and again there was no effect of sampling pattern. Graphs of the greenfield site results are on the last three pages of Appendix C.

3.2.5 Calculations of Differences in Contamination between Pre and Post-IPPC Sampling that could be detected as a function of sampling density and power

So far this report has presented probabilities of detecting various levels of added contamination under a variety of sampling schemes. It would be informative to examine the levels of contamination that could reasonably be detected for particular sampling densities at various levels of statistical power.

We will assume a null hypothesis of no added contamination during the IPPC period, i.e. “innocent until proved guilty”. Type I error is kept constant at 5% and the calculations are based on the levels of contamination on an area that carried a high
risk of contamination both before and during the PPC period. The calculations were done approximately using the automatic sample size calculator in the MINITAB statistical package. The one-sample t-test (i.e. the paired t-test) was selected. These calculations were checked against the actual results of the simulations shown in the graphs on the first three pages of Appendix C.

The results of these calculations are shown graphically in the following two figures; there is one graph per level of s.d. The coloured points are values taken from the Appendix C graphs, which were used to check the calculations.
Approximate Contamination Increases Detectable at Various Sampling Densities - s.d. = 0.5

The graphs above show the detectable contamination increase as a function of sample density for four power levels. For example, on the first graph (for the lower of the two s.d.’s) if contamination during the PPC period has increased by 5000mg/kg on average, sampling at a density of 0.0008m$^{-2}$ will detect it around 20% of the time. This means that if you sampled five locations that had been contaminated with 5000mg/kg of contamination during the PPC period, you will fail to spot that extra contamination on four of those five locations.

The second graph shows the same results, but for the higher variability level. The situation here is even worse; for the same degree of contamination (5000mg/kg added during PPC), sampling at 0.0008m$^{-2}$ will miss the contamination on nine out of ten occasions.

From a statistical point of view, it is desirable that the sampling detects added contamination at least 80% of the time. Clearly detecting additional contamination 80% of the time is going to be impossible unless very high density (and therefore costly) sampling schemes are carried out.

3.3. Contamination heterogeneity issues

So far the statistical methods have explored ways of identifying whether the average contamination in an area of a site has increased during the IPPC period. We know from experience that the extent of contamination can be very variable. If a significant increase in contamination is detected in an area, the recommendation would be to remediate the entire area. This may not be the best course of action for two reasons.

- It may not be cost effective if only one part of the area is actually contaminated
• The contamination over part of the area will need to be severe in order for the difference in average contamination to be detectable, so some zones of contamination are likely to be missed.

We attempted to find methods of examining contamination heterogeneity and these are outlined in the next two sections.

3.3.1  Point Comparisons of differences with LSD on log scale – sites with pre-existing contamination

The Least Significant Difference (LSD) is the largest difference between two individual values that is not due to chance.

\[ LSD = z_{1-\alpha} \sqrt{2} s \]

where \( z \) is a standard normal deviate\(^3\) and \( s \) is the measurement s.d.

The percentage of each set of samples where the post- to pre-PPC difference is greater than the LSD averages out at between 0.25% and 10% depending on the s.d., the risk and the post-PPC contamination rate. High contamination-risk areas, low s.d. and high post-PPC contamination rate give the highest percentage.

A way of turning these results into a test might be to count the number of simulations where at least \( m \) pairs of post- and pre-PPC results registered a difference greater than the LSD.

3.3.2  Point Comparisons of differences with LSD on log scale – greenfield sites

The comparison of differences method is very much better when you are dealing with a greenfield site. The percentage of pairs greater than the LSD matches the true value when the Post/Pre-PPC ratio is 0.01 or greater for all types of site. When there has been no new contamination added, the maximum percentage of pairs where a difference was detected erroneously is less than 0.2%. There are no effects of sampling layout. This method looks very promising as a way of detecting contamination on a greenfield site.

3.3.3  Devising a measure of heterogeneity

Supposing we have an area of soil that we have sampled before and after the PPC period. Let’s suppose that the values in each square in the diagrams below are the logarithms of the added contamination within each grid square. If the area had been contaminated homogeneously we might get the distribution of contamination on the

\(^3\) \( z_{1-\alpha} \)-values are the inverse of the normal cumulative distribution function. They define the limits between which normally distributed variables lie \((1-\alpha)\times100\% \) of the time.
left. If there had been a larger spill in the four shaded grid squares in the top right hand corner, we might get the picture on the right.

<table>
<thead>
<tr>
<th>7.46</th>
<th>7.76</th>
<th>7.57</th>
<th>7.22</th>
<th>7.65</th>
<th>7.46</th>
<th>7.76</th>
<th>7.57</th>
<th>8.18</th>
<th>8.08</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.00</td>
<td>7.61</td>
<td>7.48</td>
<td>7.53</td>
<td>7.58</td>
<td>7.00</td>
<td>7.61</td>
<td>7.48</td>
<td>8.62</td>
<td>8.70</td>
</tr>
<tr>
<td>7.36</td>
<td>8.12</td>
<td>7.64</td>
<td>7.41</td>
<td>7.09</td>
<td>7.36</td>
<td>8.12</td>
<td>7.64</td>
<td>7.41</td>
<td>7.09</td>
</tr>
<tr>
<td>7.25</td>
<td>7.32</td>
<td>7.12</td>
<td>7.72</td>
<td>7.71</td>
<td>7.25</td>
<td>7.32</td>
<td>7.12</td>
<td>7.72</td>
<td>7.71</td>
</tr>
<tr>
<td>7.69</td>
<td>7.29</td>
<td>8.26</td>
<td>8.10</td>
<td>7.68</td>
<td>7.69</td>
<td>7.29</td>
<td>8.26</td>
<td>8.10</td>
<td>7.68</td>
</tr>
</tbody>
</table>

The following statistic can be used to quantify the inhomogeneity of the site:

\[ I_{\text{actual}} = \sum_{j} \sum_{i} (\Delta y_j - \Delta y_i)^2 \|x_i - x_j\| \]

where \( \Delta y_j \) is the difference in contamination at the \( i \)th sampling point and \( \|x_i - x_j\| \) is the absolute Euclidean distance between sampling locations \( i \) and \( j \) in metres. The value of \( I_{\text{actual}} \) for the situation on the left is 3104 metres, whereas for the situation on the right it is 6713 metres. The small amount of inhomogeneity in the top right hand corner of the diagram on the right is enough to more than double the value of \( I_{\text{actual}} \).

In practice, \( I \) could take a variety of values, depending on the distribution of the contamination. We used simulations to calculate the normal range of \( I \) values when no inhomogeneity is present, if an experimental value is higher than the 95th percentile of the theoretical \( I \) values, we have an indication that gross inhomogeneity is present.

If we detect gross inhomogeneity, we could then partition the area until the resulting sub-areas showed no further inhomogeneity. Tests for contamination could be applied separately to each homogeneous sub-area and only those sub-areas showing significant contamination increases during the IPPC period would then have to be remediated. This approach would, in theory, overcome the two issues of inhomogeneity listed previously.

Another set of simulations was carried out as in the following flow chart.

---

4 The \( \Delta y \) values are dimensionless, since they are the difference of the logarithms of concentrations. This is because \( \log(y_1) - \log(y_2) = \log\left(\frac{y_1}{y_2}\right) \) so the units for \( y \) of mg/kg cancel.
Start

Take 25 samples from area of mixed contamination risk

Calculate $I_{actual}$ for samples

Take 25 samples from low, medium and high single risk areas

Calculate $I_{p50}$ for low medium and high risk areas

End
The process was repeated for several combinations of simulation parameters. The results are tabulated below.

<table>
<thead>
<tr>
<th>Simulation parameters</th>
<th>$I_{95%}$ low risk</th>
<th>$I_{95%}$ medium risk</th>
<th>$I_{95%}$ high risk</th>
<th>$I_{actual}$ mixed risk</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post/Pre ratio=0.25</td>
<td>43</td>
<td>13774</td>
<td>1442</td>
<td>32202</td>
</tr>
<tr>
<td># samples=25</td>
<td>41</td>
<td>13506</td>
<td>560</td>
<td>65</td>
</tr>
<tr>
<td>Post/Pre ratio=1</td>
<td>0</td>
<td>43413</td>
<td>1151</td>
<td>29809</td>
</tr>
<tr>
<td># samples=25</td>
<td>700</td>
<td>32661</td>
<td>1329</td>
<td>40907</td>
</tr>
<tr>
<td>Post/Pre ratio=0.25</td>
<td>0</td>
<td>15580</td>
<td>259</td>
<td>2327</td>
</tr>
<tr>
<td># samples=16</td>
<td>12243</td>
<td>6720</td>
<td>198</td>
<td>326</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>406</td>
<td>453</td>
<td>6551</td>
</tr>
<tr>
<td></td>
<td>6471</td>
<td>11037</td>
<td>200</td>
<td>199</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>250</td>
<td>369</td>
<td>7951</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>447</td>
<td>316</td>
<td>6364</td>
</tr>
<tr>
<td></td>
<td>3.2</td>
<td>6298</td>
<td>204</td>
<td>93</td>
</tr>
</tbody>
</table>

The table above shows that especially for low and medium risk areas the value of $I_{95\%}$ is extremely variable. The main reason for this is that the calculation of these values has been done on a single realisation of contamination, which was itself quite inhomogeneous because of the method used to simulate spills. If many realisations had been made, the values would probably have been more stable. However, this would have been very time-consuming. Also, no account was taken of the lognormal distribution of TPH values and taking logs may have stabilised the single risk area $I_{95\%}$ values.

### 3.4 Conclusions from the statistics

1. The sampling density of 0.0008 per m$^2$ in use typically for site investigations is too low to detect even high levels of contamination added during the IPPC period when the paired t-test is used with a null hypothesis of no added contamination during PPC.
2. The detection rate for a paired t-test with a null hypothesis of no added contamination can be improved if logarithms of the TPH values are taken first.
3. If the original sampling locations are lost it becomes very difficult to detect contamination added during the IPPC period since the standard deviation of the data includes spatial variation.
4. Detection of contamination is always better where the sampling and measurement s.d. is lower.
5. There is no effect of sampling pattern on contamination detection rate.
6. The sampling density of 0.02 per m$^2$ gave perfect detection on greenfield sites for contamination on medium and high risk areas where spills were at least 5m in diameter. This finding is in line with theoretical calculations.
7. Very high sampling densities are required to achieve optimal detection rates for contamination added during the IPPC period. Sampling at such high densities would cost more than the remediation of the site.

8. Comparing estimates of the difference in contamination at a single sampling point with the Least Significant Difference may prove to be a useful way of detecting hot spots of contamination. If this method could be perfected it would save remediating the whole area when just a small part of it had been contaminated.

9. An initial attempt to formulate a statistic to detect heterogeneous contamination was not particularly successful. Changing the way the statistic is calculated may help, but this will be more computationally intensive.
4. ALTERNATIVE APPROACHES TO SOIL CONTAMINATION Decision Making Under PPC

Given the prohibitively large sampling costs required for the statistical methods advocated in the draft IPPC document to detect any new soil contamination that occurred during the lifetime of a PPC permit are, what are the alternatives?

4.1 Environmental Management Systems

Given the potentially high costs of remediating contaminated soil and groundwater it is obviously preferable to prevent contamination in the first place by appropriate use of:

- Primary containment
- Secondary containment
- Measures to prevent tank overfills
- Tank and pipeline inspection and integrity testing regimes
- Environmentally aware operating procedures
- Environmental awareness staff training

It makes sense to match the type of containment and spill/leak prevention measures to the level of environmental risk posed by a particular tank, pipeline or facility, rather than a “one size fits all” approach. Thus the highest specification containment and most intensive spill/leak detection regimes are reserved for the highest risk sites. This approach ensures the greatest environmental protection for a given budget. The Institute of Petroleum is currently developing guidance on this subject for above ground storage tanks.

A potential consequence of the Agencies not allowing a risk-based approach to soil contamination occurring during the lifetime of a PPC permit is that risk-based containment and spill/leak prevention is not allowed. The risk arising from this is that companies spread their investment in containment and spill/leak prevention measures evenly across their facilities with the result that the high risk sites do not get the level of investment they require to adequately manage the risks. Thus a blanket approach to soil contamination under PPC could be counter-productive and result in greater risk to the environment.

It is unrealistic to expect that even with the best containment, spill/leak prevention measures, operating procedures and staff training in place, that spills and leaks will not occur during the lifetime of a PPC permit, although they should occur far less frequently than they have in the past. It is best to detect leaks as early as possible using wet stock management tools, chemical sensors located in the ground or traditional groundwater monitoring. When leaks are detected or spills occur they are best cleaned-up immediately. In general, the cost of remediation for mobile contamination increases exponentially with the length of time it is left in the ground to migrate.

The EA and SEPA have indicated that they will take into account the following to determine the extent of, or even the need for, a site investigation when the PPC permit is surrendered:

- The site’s Environmental Management System (EMS)
• Quality of primary containment in place
• Quality of secondary containment in place
• Tank and pipeline inspection and integrity regimes and associated records
• Spill prevention measures
• Leak detection measures
• Spill reporting and subsequent clean-up
• Incident investigation reporting and corrective actions
• Groundwater monitoring reports
• Remediation reports

Thus a good EMS coupled with good housekeeping and record keeping may be sufficient to convince the agencies that no site investigation is required for surrender of the PPC permit or it may at least eliminate large portions of the site from such a need.

It is beyond the scope of this report to go into these measures in any detail.

4.2. Environmental forensics

One way of dealing with contaminant heterogeneity when monitoring the biodegradation of crude oils or refined petroleum products in the environment is to select recalcitrant components referred to as biomarkers (e.g. hopanes) in these complex mixtures and monitor the change in the ratio of the other components to these recalcitrant markers. Another approach for individual contaminants is to measure changes in stable isotope ratios (e.g. $^{12}$C:$^{13}$C, $^{14}$N:$^{15}$N), which take place when a compound is subject to biodegradation. Enzymes preferentially act on lighter isotopes so that during biodegradation contaminants become enriched in heavier isotopes.

Both these methods enable rates of contaminant degradation to be calculated without the need for large numbers of samples to demonstrate significant reductions in contaminant concentrations.

These and other types of environmental forensics analysis may be useful in helping to determine whether the contamination identified during a site closure investigation prior to the surrender of a PPC permit is the same as that detected in the site baseline investigation, without the need for large numbers of samples. They may not work in every case, but have considerable potential to be of value.

Environmental forensics is a rapidly growing field of which two examples have been given. Methods vary from the relatively simple to the complex. They include:
• Knowledge of site history
• Good record keeping (written and photographic)
• Chemical fingerprinting of complex mixtures
• Ratio analysis of individual components in complex mixtures
• Stable and radioactive isotope ratio analysis
• Principal components analysis
• Fate and transport modelling
It is beyond the scope of this report to discuss these methods in any detail. For more information consult Sullivan et al., (2001) and Murphy and Morrison (2002) and the Journal of Environmental Forensics published by AEHS.
5. OVERALL CONCLUSIONS

The statistical methods that have been evaluated in this project require sampling densities that are 1-2 orders of magnitude greater than those currently used for site investigations for historic contamination managed under Part IIa (EPA, 1990). This would result in the cost of investigation being equal to or greater than the current cost of remediation in many cases (Table 5.1). This cannot be justified on cost:benefit grounds.

Table 5.1. Comparison of site investigation costs at different sampling densities with current remediation costs

<table>
<thead>
<tr>
<th>Sampling density (m⁻²)</th>
<th>Cost (£m⁻²)*</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0008</td>
<td>0.6</td>
<td>Typical of site investigations for historic contamination under Part IIa</td>
</tr>
<tr>
<td>0.008</td>
<td>6.0</td>
<td></td>
</tr>
<tr>
<td>0.02</td>
<td>15.0</td>
<td>Theoretically 100% detection of a 5 m diameter spill</td>
</tr>
<tr>
<td>0.2</td>
<td>150</td>
<td>Taken from example given in EA draft Guidance (EA, 2002)</td>
</tr>
<tr>
<td>Cost of remediation</td>
<td>5 - 25</td>
<td>Current remediation costs</td>
</tr>
</tbody>
</table>

* Assumes £750 per location made up as follows:
  - cost of trial pit = £300
  - cost of borehole = £1,000
  - ratio of trial pits:boreholes = 1:1
  - analysis costs = £100 per location

In the final analysis, on the grounds of practicality, it is difficult to envisage anything other than a risk-based approach to contamination occurring under PPC.

This report recommends:
- Further dialogue with the Agencies to attempt to align the approach to soil and groundwater contamination under PPC with that under Part IIa (EPA, 1990)
- A greater focus on the site’s Environmental Management System (EMS), quality of product containment, spill/leak prevention measures, spill reporting and corrective action recording as a means of demonstrating whether or not significant new contamination has taken place during the lifetime of a PPC permit.

Environmental forensic measurements of contamination (e.g. fingerprints of complex mixtures, ratios of recalcitrant components in complex mixtures, isotope ratios) may be useful tools where it does become necessary to distinguish between pre-PPC and PPC contamination.

Following the completion of this study, the Environment Agency have recognised the difficulties for site owners of taking a statistical approach to comparing soil contamination data in PPC Baseline and Site Closure reports. They have indicated that they are willing to consider the other criteria recommended in this report as part
of a “weight of evidence” approach to judging whether or not significant new contamination has taken place during the lifetime of a PPC permit.
6. REFERENCES


Appendix A: Table of theoretical spill detection probabilities as a function of spill size, sampling density and sampling design

**Square grid**

<table>
<thead>
<tr>
<th>Chance of detection</th>
<th>Spill radius r (m)</th>
<th>Grid spacing, R (m)</th>
<th>Sampling density (per m²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>5</td>
<td>39.63</td>
<td>0.00064</td>
</tr>
<tr>
<td>0.1</td>
<td>5</td>
<td>28.02</td>
<td>0.00127</td>
</tr>
<tr>
<td>0.2</td>
<td>5</td>
<td>19.82</td>
<td>0.00255</td>
</tr>
<tr>
<td>0.3</td>
<td>5</td>
<td>16.18</td>
<td>0.00382</td>
</tr>
<tr>
<td>0.4</td>
<td>5</td>
<td>14.01</td>
<td>0.00509</td>
</tr>
<tr>
<td>0.5</td>
<td>5</td>
<td>12.53</td>
<td>0.00637</td>
</tr>
<tr>
<td>0.6</td>
<td>5</td>
<td>11.44</td>
<td>0.00764</td>
</tr>
<tr>
<td>0.785</td>
<td>5</td>
<td>10.00</td>
<td>0.01000</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>7.07</td>
<td>0.02000</td>
</tr>
</tbody>
</table>

**Triangular grid**

<table>
<thead>
<tr>
<th>Chance of detection</th>
<th>Spill radius r (m)</th>
<th>Grid spacing, R (m)</th>
<th>Sampling density (per m²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>5</td>
<td>60.23</td>
<td>0.00032</td>
</tr>
<tr>
<td>0.1</td>
<td>5</td>
<td>42.59</td>
<td>0.00064</td>
</tr>
<tr>
<td>0.2</td>
<td>5</td>
<td>30.11</td>
<td>0.00127</td>
</tr>
<tr>
<td>0.3</td>
<td>5</td>
<td>24.59</td>
<td>0.00191</td>
</tr>
<tr>
<td>0.4</td>
<td>5</td>
<td>21.29</td>
<td>0.00255</td>
</tr>
<tr>
<td>0.5</td>
<td>5</td>
<td>19.05</td>
<td>0.00318</td>
</tr>
<tr>
<td>0.6046</td>
<td>5</td>
<td>17.32</td>
<td>0.00385</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>8.66</td>
<td>0.01540</td>
</tr>
</tbody>
</table>

**Herringbone grid**

<table>
<thead>
<tr>
<th>Chance of detection</th>
<th>Spill radius r (m)</th>
<th>Grid spacing, R (m)</th>
<th>Shift, s (m)</th>
<th>Sampling density (per m²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>5</td>
<td>39.63</td>
<td>9.91</td>
<td>0.00064</td>
</tr>
<tr>
<td>0.1</td>
<td>5</td>
<td>28.02</td>
<td>7.01</td>
<td>0.00127</td>
</tr>
<tr>
<td>0.2</td>
<td>5</td>
<td>19.82</td>
<td>4.95</td>
<td>0.00255</td>
</tr>
<tr>
<td>0.3</td>
<td>5</td>
<td>16.18</td>
<td>4.05</td>
<td>0.00382</td>
</tr>
<tr>
<td>0.4</td>
<td>5</td>
<td>14.01</td>
<td>3.50</td>
<td>0.00509</td>
</tr>
<tr>
<td>0.5</td>
<td>5</td>
<td>12.53</td>
<td>3.13</td>
<td>0.00637</td>
</tr>
<tr>
<td>0.6</td>
<td>5</td>
<td>11.44</td>
<td>2.86</td>
<td>0.00764</td>
</tr>
<tr>
<td>0.7</td>
<td>5</td>
<td>10.59</td>
<td>2.65</td>
<td>0.00891</td>
</tr>
</tbody>
</table>
Appendix B: Explanation and illustration of hypothesis testing

Many statistical tests rely on the idea of hypothesis tests. The basic process is as follows:

1. Construct a pair of hypotheses – a null hypothesis, \( H_0 \) and an alternative hypothesis, \( H_1 \).
2. Collect data.
3. Calculate a test statistic.
4. Compare the test statistic with a critical value from tables.
5. Accept or reject the null hypothesis.

The null hypothesis is a statement of what you initially believe. The alternative hypothesis is a statement of what you will be persuaded to believe if there is sufficient evidence.

Null hypotheses are typically used in an experimental context where one might be looking for the effect of a drug in relieving symptoms or the effect of an additive in improving process yield. Statisticians frequently take the conservative approach and choose to take the pessimistic view that the drug or additive does nothing and then design the study such that, if the effect is genuine, the experiment will stand a reasonable chance of detecting the effect.

The effect is either genuine or not and the test can either identify an effect or not. There are therefore \( 2 \times 2 = 4 \) possible outcomes which are usually tabulated like this. I’ll use the letter \( B \) to represent the size of the effect.

<table>
<thead>
<tr>
<th>True situation</th>
<th>Accept ( H_0 )</th>
<th>Reject ( H_0 ) – Accept ( H_1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( H_0 : B=0 )</td>
<td>Correct decision. Probability = 1-( \alpha )</td>
<td>Wrong decision. Probability = ( \beta ) or Type II error</td>
</tr>
<tr>
<td>( H_1 : B&lt;&gt;0 )</td>
<td>Wrong decision. Probability = ( \alpha ) or Type I error</td>
<td>Correct decision. Probability = 1-( \beta ) or Power</td>
</tr>
</tbody>
</table>

The Type I error is also called “false positive” in everyday terms.

Tests are usually set up so that the Type I error is stringently controlled. Values are limited to 5% or below. We talk about effects being “highly significant” if the Type I error is below 1%. The lower the Type I error, the more unlikely it is for the result to have occurred by chance. If it is important for you to be really sure an effect is genuine, you need to start with a null hypothesis that represents the “safe” option.

People are usually less concerned about Type II errors and power – in some cases they are rarely considered. This does not mean they are not important. A power level of 80% is considered good.

The last issue we need to look at is “one-sided” and “two-sided” tests. The “sidedness” refers to the alternative hypothesis. If you are only looking for a positive effect
or perhaps it is physically impossible for the effect to be negative, you would do a “one-sided” test and set the alternative hypothesis at $B>0$. The converse would be the case for effects where you were only interested in a decrease in the response. If the effect could be positive or negative and effects of both signs were of interest, you would use a two-sided test and select $B<>0$ as your alternative hypothesis.

By chance, in the absence of any real effect, the test statistic has a certain distribution, which we’ll assume is normal for the purposes of illustrating it.

The limits on the diagram above are in fact the Critical Values, to which the test statistic is compared. The area under the curve beyond the limits is equal to the Type I error, i.e. typically 5%. If the absolute value of the test statistic is greater then the critical value the null hypothesis is rejected.

It is slightly easier to reject the null hypothesis if you select a one-sided test, since the critical value is lower, so you have to select the alternative hypothesis objectively.

In the case of our soil remediation hypothesis, we should ideally assume that we remediate only if an increase in contamination is detected over the IPPC period. We are not interested in whether the contamination has stayed the same or gone down – it makes no difference to our actions.
Appendix C: Graphs showing effects on detection rate of sampling design, measurement s.d. and changes in Post/Pre PPC contamination ratio

Results for low risk areas with existing contamination

![Graphs showing effects on detection rate of sampling design, measurement s.d. and changes in Post/Pre PPC contamination ratio](image-url)
Results for medium risk areas with existing contamination
Results for high risk areas with existing contamination
Results for low risk areas on former greenfield site
Results for medium risk areas on former greenfield site
Results for high risk areas on former greenfield site