Onshore oil and gas monitoring: assessing the statistical significance of change

Chief Scientist's Group report

October 2019

Version: SC160020

We are the Environment Agency. We protect and improve the environment.

We help people and wildlife adapt to climate change and reduce its impacts, including flooding, drought, sea level rise and coastal erosion.

We improve the quality of our water, land and air by tackling pollution. We work with businesses to help them comply with environmental regulations. A healthy and diverse environment enhances people's lives and contributes to economic growth.

We can't do this alone. We work as part of the Defra group (Department for Environment, Food & Rural Affairs), with the rest of government, local councils, businesses, civil society groups and local communities to create a better place for people and wildlife.

This report is the result of research commissioned and funded by the Environment Agency.

Research at the Environment Agency

Scientific research and analysis underpins everything the Environment Agency does. It helps us to understand and manage the environment effectively. Our own experts work with leading scientific organisations, universities and other parts of the Defra group to bring the best knowledge to bear on the environmental problems that we face now and in the future. Our scientific work is published as summaries and reports, freely available to all.

This report is the result of research commissioned by the Environment Agency's Chief Scientist's Group.

You can find out more about our current science programmes at https://www.gov.uk/government/organisations/environment-agency/about/research

If you have any comments or questions about this report or the Environment Agency's other scientific work, please contact research@environment-agency.gov.uk.

Professor Doug Wilson
Chief Scientist

Foreword

This report has been produced to inform the Environment Agency on the use and implications of statistical techniques for evaluating "baseline" environmental data around possible onshore oil and gas sites. It considers the basic decisions and data needed to establish a baseline, and draws on existing practice and published statistical techniques to derive principles that may help in the design and operation of environmental baseline surveys.

The report considers potential pollutant substances in both air and water (groundwater) with an emphasis on substances released at onshore oil and gas sites, including sites that aim to extract hydrocarbons from shale rock. Specifically, it considers the role of a suitable conceptual model to describe source-pathway-receptor relationships at a site, and to drive the design of surveys and the interpretation of results. It includes likely pollutants from onshore oil and gas sites and aspects of their fate and transport. For example in relation to air, the pollutants considered include methane, nitrogen oxides, particulates and volatile organic compounds (non-methane hydrocarbons).

The shale gas sector is an emerging extractive industry in the UK rather than an established one. By contrast, over the past 20 years a similar sector has emerged and become established in the United States of America (USA). A frequent criticism of the USA experience is that suitable baseline studies were not carried out before work started. This has made it difficult or impossible to identify local environmental changes or to apportion changes between industry contributions and other factors. The need for robust baseline procedures and for data to identify and apportion changes is recognised in the UK. The present work is therefore a preliminary study to address this need, in the absence of established baseline procedures from developments like those in the USA.

The study shows how monitoring strategies and approaches can be developed to better detect whether or not there have been changes in environmental quality around sites, and to characterise any changes and apportion them to oil, gas or other sources. It reviews monitoring and statistical principles that can underpin the development of baselines. It also reviews some published case studies that show how the principles may be applied in practice, although the extent of available case studies was limited.

The report describes how information can be developed and presented to assess sites before, during and after operations. It is not designed to prescribe how information from monitoring, modelling and assessment should be used to make regulatory decisions, because such prescriptive advice would be outside the scope of a research study. The report is not a statement of the Environment Agency's position, and it does not represent Environment Agency guidance on the matter.

Executive summary

Development of the UK's online oil and gas (OOG) resources is at an important stage. After several years with little activity, permission has been given for some exploratory drilling and hydraulic fracturing. This project has examined the requirements for air-quality and groundwater-quality monitoring at OOG facilities for Environment Agency and external use, with particular attention to establishing an environmental baseline. It also considers how to detect any statistically significant changes from the baseline during a facility's operational stage, and how to attribute changes to potential cause(s).

The first phase of the project involved a literature review to identify principles for the statistical assessment of air quality and groundwater monitoring data. It considered: processes and pathways for pollutants, the regulatory context, monitoring techniques, survey design, and statistical techniques. Suggestions for survey design and data analysis were developed at this stage. The second phase involved developing 5 case studies to show how the identified principles for statistical assessment could be applied to existing air and groundwater datasets from the UK. The third phase combined the principles and case study findings to provide options for survey design and statistical data analysis. It describes how datasets may be analysed using appropriate statistical tests and identifies tests that may be used to determine significance; it then indicates how test results may be interpreted. It also considers how signal strengthening and source attribution may be done using statistical techniques.

The findings are presented as a series of notes linked to a five-stage process (see below) that was devised for establishing a baseline and determining change. The first 3 stages address how to design a monitoring survey that covers the establishment of a baseline and different operational phases in the lifecycle of an OOG facility. The final 2 stages address how to perform statistical analysis on the monitoring data, and how to detect statistically significant changes from baseline conditions.

1 Set out monitoring design principles & derive the conceptual model

Guidelines to give
- Key considerations
- An overview of how to approach conceptual model design
- Introduction to how this informs the monitoring design

2 Identify key substances and their "risk" to air and groundwater quality at an operational site

Guidelines to cover
- Key substances, and where to find any associated standards and QA/QC procedures
- Issues related to societal perception and public information

3 Set out site and operation specific monitoring design

- Guidelines to consider how conceptual model and key substance information can be used to inform location and frequency of measurement, the duration of baseline monitoring phase and other explanatory data to be collected
- Guidelines will also be introduced for "Minimum Statistically Based Standards for monitoring" for integration in finalised design

4 Begin stage specific monitoring and statistical data analysis

Guidelines to cover
- Statistical analysis for the establishment of baseline
- Statistical analysis for the determination of change / cause to trigger further data investigation
- How the statistical analysis should inform monitoring design

5 Statistically investigate the data where a "potential risk" has been identified

Guidelines to cover
- How to investigate potential causes of change

Overview of staged process of survey design and statistical analysis

Supporting information, including the case studies, is provided in a separate Annex.

Contents

List of tables and figures

1   Introduction

Development of the UK's onshore oil and gas (OOG) resources is at an important stage. After a number of years with little activity on the ground, a number of companies are now progressing through the regulatory and technical issues towards exploratory drilling and hydraulic fracturing. The issue by the government of the 14th round of Petroleum Exploration and Development Licences potentially opened up a wider range of sites for exploration. This report aims to provide advice on how to evaluate the requirements for monitoring by both the Environment Agency and others. Monitoring needs to:

- satisfy the need for routine assessment of compliance with permits

- be sufficiently comprehensive to meet the information needs of local communities and their representatives

Statistical procedures for baseline setting, reviewing operational monitoring and assessing compliance are not well established for OOG sites. This study reviewed what has been done in other air quality and groundwater contexts, including existing OOG sites, and recommends procedures for use with OOG activities. The aim of this document is to:

- identify appropriate techniques for statistical analysis to determine the statistical significance of any change or demonstrate that little or no change has occurred;

- inform the design of monitoring programmes and assessment of data;

- provide advice on how to further investigate the data to identify potential causes of change (including OOG activities and other cycles such as natural seasonal changes).

1.1   Regulatory framework

The regulatory framework for managing unconventional oil and gas development is set out in guidance produced by the then Department for Energy and Climate Change[1] (DECC 2013). The Environment Agency's role is linked to the environmental permitting process, on which it has published sector guidance for the OOG industry (Environment Agency 2016). It identifies activities that may be regulated by the Environment Agency. These include:

- well pad construction

- drilling exploratory wells

- flow testing and well stimulation, including hydraulic fracturing

- storing and handling crude oil

- treatment of waste gases (including flaring)

---

[1] Now part of the Department for Business, Energy and Industrial Strategy

- handling, storage and disposal of produced waters and flowback fluid

- managing extractive wastes

- extraction of coal mine methane

The Environment Agency is also a statutory consultee for the planning and environmental impact assessment process.

Alongside these statutory obligations, the Infrastructure Act 2015 specifies a requirement for baseline monitoring of methane in groundwater prior to the commencement of OOG extraction. During the passing of the act, undertakings were given that baseline monitoring would be sufficient to ensure that any significant subsequent impacts of OOG could be detected. These commitments are underlined by the European Commission's Recommendation on minimum principles for the exploration and production of hydrocarbons (such as shale gas) using high-volume hydraulic fracturing (2014/70/EU) (European Commission 2014). This places over-reaching obligations on Member States in relation to baseline studies, requiring that a baseline to be determined for:

- quality and flow characteristics of surface and ground water

- water quality at drinking water abstraction points

- air quality

- soil condition

- presence of methane and other volatile organic compounds in water

- seismicity

- land use

- biodiversity

- status of infrastructure and buildings

- existing wells and abandoned structures

It is likely to fall to the Environment Agency to ensure that these obligations are met as the OOG industry develops. This project therefore has an important role in enabling the Environment Agency and other partners to conduct air quality and groundwater monitoring surveys that:

- are carried out to a high standard of monitoring design

- will enable robust statistical analysis, detection and attribution of change

This will help to ensure that investments in monitoring produce optimum results and help to detect any impacts of OOG activities in practice.

1.2   Approach

The project had 3 phases. Supporting information for all the phases is provided in the Annex to this report.

Phase 1: Establish principles

The first phase involved a literature review to identify the principles that form the basis of statistical assessments of air quality and groundwater monitoring data. It considered:

- processes and pathways for pollutants
- the regulatory context
- monitoring techniques
- survey design
- statistical techniques

Outline approaches for survey design and data analysis were developed at this stage.

Phase 2: Review and refine

The second phase involved the preparation of case studies illustrating how principles for statistical assessment identified in Phase 1 could be applied to existing air quality and groundwater datasets from the UK. Existing examples of environmental baseline and operational monitoring studies were used to evaluate the outline in practice; and it was then refined throughout the case study assessments.

The case studies are presented in Section A.3 of the Annex. They cover:

- establishing an air quality baseline at a proposed OOG site
- analysing monitored air quality data at the energy-from-waste facility at Great Blakenham in Suffolk
- investigating the source of nickel concentrations detected at air quality monitoring stations in Sheffield
- establishing a methane baseline in groundwater
- analysing operational groundwater quality data at an existing OOG site

Phase 3: Develop recommendations

The third phase consisted of finalising the approaches for survey design and statistical data analysis. The survey design guide aims to enable the design of air quality and groundwater monitoring surveys so as to optimise the benefit of the investment in time and resources, while gathering sufficient data to enable robust statistical analysis. It sets out key principles for survey design, including reference to existing guidance where relevant.

The approach for statistical analysis describes how datasets should be analysed using appropriate statistical tests. It identifies tests that should be applied to determine significance and indicates how test findings should be interpreted. It also considers statistical methods for analysing data in order to

clarify signals from particular sources (signal strengthening), and to attribute changes to whatever sources or factors have caused them (source attribution).

1.3   Overview

Following the literature review and development of the case studies, a five-stage approach for establishing a baseline and determining change was proposed (Figure 1.1).
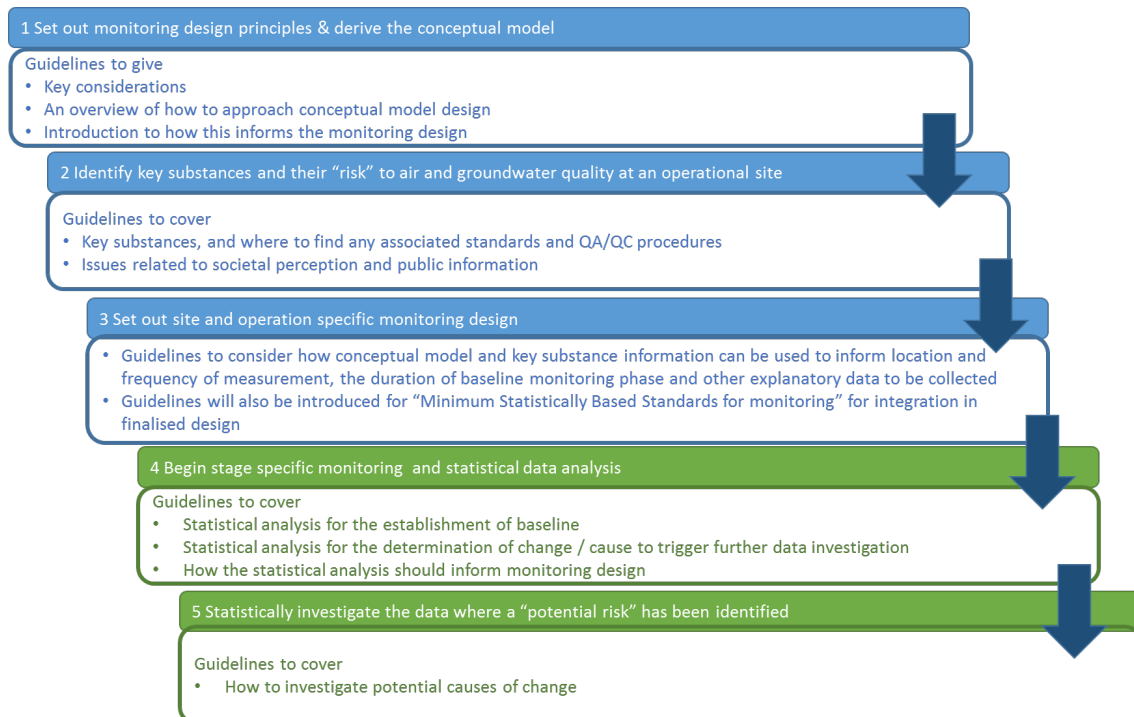
Figure 1.1   Overview of staged process of survey design and statistical analysis

Stage 1 involves the setting out of the overarching principles governing the design of the monitoring programme applicable to both air and groundwater quality, and the development of the conceptual model (Section 2). It is not always practical to consider the monitoring and analysis of air and groundwater quality together in a single approach and so later sections consider them separately. This allows sufficient detail to be provided for both while enabling a divergence in approach where appropriate.

Stage 3 describes how the proposed monitoring design can be refined to make it specific to the particular site and operations (Section 4) for air quality and groundwater quality.

Stage 4 explains how to analyse the baseline data pertaining to air and groundwater quality, and then how to detect any change from baseline conditions (Section 5).

4

Stage 5 looks at how to investigate and attribute potential causes of a statistically significant change (Section 6).

The output for Stages 3 to 5 is presented in a series of "Notes" (Table 1.1) which provide explanations and comments on appropriate methods for each topic in a Stage. A series of Diagrams provide flowcharts and decision trees which summarise the approach and indicate when to consult a particular Note.

Table 1.1 List of Notes for Stages 3 to 5

Number	Title
Stage 3: Site and operation specific monitoring design	
1.1	Design of air quality monitoring programme
1.2	Statistical considerations in air quality monitoring design
1.3	Design of groundwater monitoring programmes
1.4	Statistical considerations in groundwater monitoring design
Stage 4: Lifecycle stage specific monitoring and statistical data analysis	
2.1	QA/QC checks
2.2	Data visualisation
2.3	Outlier detection
2.4	Adequacy
3.1	Test for normality
3.2	Transforming the data
3.3	Trend analysis
3.4	Change detection
3.5	Change difference
3.6	Groundwater assessment criteria
Stage 5: Signal strengthening/change attribution	
4.1-AQ	Collate explanatory variable data
4.2-AQ	Conduct wind sector analysis (WSA)
4.3-AQ	Perform conditional analysis to increase signal strength
4.4-AQ	Filter out distant emissions
4.5-AQ	Other activities
4.6-AQ	Application of parametric techniques for change attribution (linear regression)
4.7-AQ	Application of non-parametric techniques for change attribution (PCA / kernel estimator methods)
4.8-AQ	Interpret attribution test results
4.1-GW	Collate explanatory variable data
4.2-GW	Collate explanatory variable data
4.3-GW	Normality checks
4.4-GW	Application of parametric techniques for change attribution (linear regression)
4.5-GW	Application of non-parametric techniques for change attribution (PCA / kernel estimator methods)
4.6-GW	Interpret attribution test results

1.4 Project limitations and recommendations for further research

This project represents an initial assessment of the state of existing knowledge on the statistical assessment of changes from baseline conditions in air quality and groundwater associated with OOG activities. Coupled with subject knowledge and a wider search of the literature of practice in other areas, this information was used to outline an appropriate methodology for use in future for the OOG sector.

The study has helped to clarify gaps, methods and priorities. However, baseline establishment and change assessment in the industry is still an area of developing science. The case studies have shown that rigorous baseline setting, change detection and source attribution are rarely undertaken in practice. Hence there are only a limited number of examples available to explore and verify in practice.

This document does not cover radioactive substances, such as naturally occurring radioactive materials (NORM).

Statistical assessment of changes from baseline conditions is a knowledge area that requires further evidence and further identification, refinement and evaluation of techniques. Pioneering OOG operators in the UK should expect to have to invest in baseline monitoring and data analysis, in order to provide comprehensive datasets that will allow potential changes to be detected in air and groundwater quality due to OOG activities. Subsequent operators are likely to benefit from that investment e.g. it may enable subsequent monitoring effort to be refined using the experience gained.

Recommendations

Because baseline-setting, change detection and source attribution are rarely conducted to the level anticipated for new OOG sites, they are important topics where knowledge and experience are limited. Recommendations for further research are therefore made:

- Ongoing research is necessary to demonstrate and consolidate systematic procedures, terminology and worked examples for analysis of environmental data for the OOG industry.

- Well-designed modelling studies could help to provide useful worked examples in the absence of suitable existing monitoring studies. For example, air dispersion modelling predictions/assessments could be conducted for realistic site scenarios. These could include sites with nearby roads and agricultural emissions, which could be used to illustrate some of the signal/noise issues and attribution difficulties that might be encountered.

- Monitoring requirements could be kept proportionate by having a hierarchy of monitoring and assessment methods that can be escalated/de-escalated according to risk. This project has not created a complete hierarchy, but it has clarified some of the principles and practices for later inclusion in a complete hierarchy.

2   Stage 1: Monitoring design principles and conceptual model development

This section establishes the most important considerations for monitoring design, provides an overview of how to approach conceptual model design, and outlines how the conceptual model informs monitoring design.

Monitoring will be necessary at the 4 stages of the lifecycle of the OOG facility:

- baseline establishment
- operational
- post-operational and decommissioning
- post-decommissioning

For the purposes of this report, operational stage monitoring is defined as including site preparation, drilling, hydraulic fracturing (if carried out), well completion and production. The post-operational and decommissioning stage is defined as the period of time where activities on site have stopped and the well abandoned (decommissioned) in accordance with the requirements of the Health and Safety Executive. A period of monitoring during this stage would need to show that there is no longer any significant ongoing environmental risk such that the environmental permit could be handed back.

This document focuses on the first 2 stages, that is, baseline and operational monitoring. Each stage of the lifecycle of the OOG facility should be carried out in anticipation of the likely requirements of the next stage of its lifecycle. This is particularly important for the baseline survey stage, as the pollutants that might be emitted during the operational cycle should be anticipated so that the baseline can be focused on these 'target' pollutants. In turn, post-operational and post-decommissioning monitoring will be guided by data collected during the operational stage, which will determine appropriate indicator parameters and monitoring frequencies.

In addition, there could be a time delay between cycles. For example, baseline monitoring might not be followed immediately by operational activities, so there may be a need to update the baseline ahead of operations commencing. However, continuous monitoring from baseline through to operations and decommissioning is recommended.

Crucially, the monitoring design should address the question of what level of change is of interest. The results from the monitoring will need to be capable of detecting that level of change, should it occur; or just as importantly, be capable of demonstrating that no substantive change has occurred.

2.1   Key considerations

Application of statistical methods to monitoring will:

- enable collection of a valid baseline dataset

- help choose appropriate measurement frequencies

- specify the reliability of measurement methods

The most important considerations when establishing an environmental monitoring programme are listed below and addressed in the following sections.

- What is the need for a monitoring strategy?

- Who should perform the monitoring?

- What substances should be measured?

- Where measurements should be carried out?

- How frequently should measurements be carried out?

- For how long should measurements be carried out?

- What accuracy/precision and what level of confidence is required?

- What standards should be applied?

- What defines a 'risk' (for example, breaching of threshold, deterioration through time or statistical change, or all of these)?

- What ancillary data should be considered e.g. data on meteorology; site activities; neighbouring sources?

The need for a monitoring strategy is underpinned by legislative and societal requirements to provide evidence and either detect problems so that they can be mitigated or provide reassurance that OOG activities are not having a detrimental impact on the environment. Operators will be required to carry out the necessary monitoring under the compliance requirements specified by the Environment Agency or another regulator.

2.2　Conceptual model design

A conceptual model represents the characteristics of the site and our understanding of the possible relationships between contaminant sources, transport pathways and sensitive receptors (Environment Agency 2003). The development of the conceptual model forms the main part of a preliminary risk assessment for a site or activity, and the model should be refined or revised as more information and understanding is obtained through the risk assessment process.

The term 'pollutant linkage' is used to describe a particular source, pathway and receptor combination. Information collected during development of the conceptual model may include a desk study and site reconnaissance.

There are various potential sources of contamination for air and water associated with OOG facilities. These sources vary depending on the nature of the operation and the stage of the development. A conceptual model allows these sources to be characterised and sensitive receptors to be identified, and the pathways linking the two to be defined. Detailed understanding of the

source and type of potential pollutants is required, and these may vary with the stage of operation.

This section provides an overview of the approach for developing conceptual models for both air and groundwater quality.

Air quality

Important emission sources to air during the early stage of development (that is, well construction and drilling) include the drilling rigs and pumps used for hydraulic fracturing. During well completion, emissions can often result from the venting and/or flaring of natural gas. Methane and other volatile organic compounds (VOCs) dissolved in flowback water and produced water need to be carefully managed to minimise releases to the environment, but some residual fugitive emissions may occur.

As the development moves into the production stage, important pollutant sources include:

- pumps – which bring the gas to the surface

- compressors

- amine units

- dehydration units

- fugitive emissions (that is, due to leaks in pipes and associated equipment, from well casings, fissures in the ground, wind ablation of proppant stockpiles)

- heavy goods vehicles (HGVs) transporting water and proppant to and from the site

Compressor stations located downstream of the wellhead are also sources of combustion and fugitive emissions (Groundwater Protection Council and All Consulting 2009).

The conceptual model is an essential part of baseline and operational air quality survey design as it enables operators to identify:

- key substances that need to be measured

- locations of air sampling points; first, to obtain an appropriate and representative baseline dataset and thereafter to maximise the likelihood of observing any detectable increase over baseline concentrations

- timing of measurements during the operational stage.

Substances released at ground level and at ambient temperature are likely to be at their highest concentration closest to the site. In contrast, substances released at elevated level and/or at high temperature may disperse further from the site, resulting in the highest concentrations some distance from it. A dispersion model can be a useful to establish the most appropriate monitoring locations, such as places where concentrations due to OOG sources are

elevated, and/or are prominent compared to those due to other sources (see, for example, Case Study 2 in Annex Section A.3.2.

Operators need to ensure that all emissions to air that may affect ambient air quality around the facility are adequately represented by the conceptual model for the site. Figure 2.1 shows an indicative conceptual model of emissions to air arising from an OOG facility. [The dispersion pathway between sources and receptors is not shown explicitly in this diagram but is indicated by the arrows leading to the receptor stage.]

Ozone is not included in this model as it is not a direct emission with local impacts from any site component, but forms over regional scales in the atmosphere from interactions between sunlight and emissions – including emissions from other sources. With the exception of the flares, sulphur dioxide has been excluded as it is assumed all fuels will be low sulphur to comply with the Sulphur Content of Liquid Fuels Regulations and the Gas Safety (Management) Regulations. Vehicle emissions are also omitted from Figure 2.1 as they can be expected to occur throughout the operation (although at higher intensity during drilling and hydraulic fracturing operations) and do not lie within the control of the Environment Agency.

Emissions are listed as being either 'conduited' (that is, a discrete, measurable source, such as an emissions stack), fugitive, or both of these. The model reflects 'normal operation'; however, excessive fugitive emissions and/or flaring should not be considered acceptable under normal operating conditions.

Presenting the conceptual model in this format allows important questions such as the following to be considered.

- Should the primary substances used to detect change be those that appear at all stages in the process (that is, those with the highest occurrences in Figure 2.1, or VOCs and methane)?

- Do substances that are emitted sporadically such as hydrogen sulphide require a different statistical approach to assessing change to those that are contained (for example, using a threshold or limit value)?

For a specific proposed development, the indicative conceptual model shown in Figure 2.1 should be elaborated to reflect the circumstances of the facility. The approach shown in Figure 2.1 could be supplemented with a site plan showing the locations of key sources and potentially sensitive receptors, and also information on local meteorology e.g. wind rose.

Operators would also be expected to provide further detail on the emission characteristics of the proposed facility as part of a permit and/or planning consent application, including:

- gridded locations (easting and northing), heights and diameters of all conduited emissions

- emission characteristics for all conduited sources, including the volume flux (mg/Nm^3), velocity (g/s), temperature, water/oxygen content and pollutant types, including potential VOC species

- details of proposed site layout and locations of nearby sensitive receptor sites

- timescales for each relevant stage of the development

The operator would also be required to provide variable emission profiles reflecting the different stages of the OOG development. The timescales of the development will have a major influence on these profiles and can vary considerably depending on the nature of the site. The time required to construct the well-pad and to carry out drilling, well completion and operation of the site will depend on a number of factors including the site's topography, the number of wells and the experience of the developer. The length of the operation will also depend on:

- the nature of the formation being fractured

- the level of extraction

- the rate at which the fracturing fluid is injected

- the intervals between stages

All these variations will affect the nature and scale of emissions to air. In addition, it will be important to understand the nature of release points in terms of the release point height above ground level and the discharge temperature. These characteristics will affect the dispersion of emissions in the atmosphere and thus the optimum locations for baseline monitoring sites.

OOG facilities are also subject to significant diurnal operational variations. These will depend on the nature of the site, any planning/regulatory controls and the operator's working methods. This will result in a degree of uncertainty when developing conceptual models. The potential for diurnal variations in background concentrations also need to be taken into consideration and must be factored into the monitoring approach adopted by the operator. It is recommended that baseline measurements are recorded on all substances of interest at regular intervals (that is, equally spaced hourly, daily, weekly, fortnightly or monthly) and should continue for at least one seasonal cycle. The use of adaptive monitoring approaches is considered in more detail in Section A.1.4.1 of the Annex.
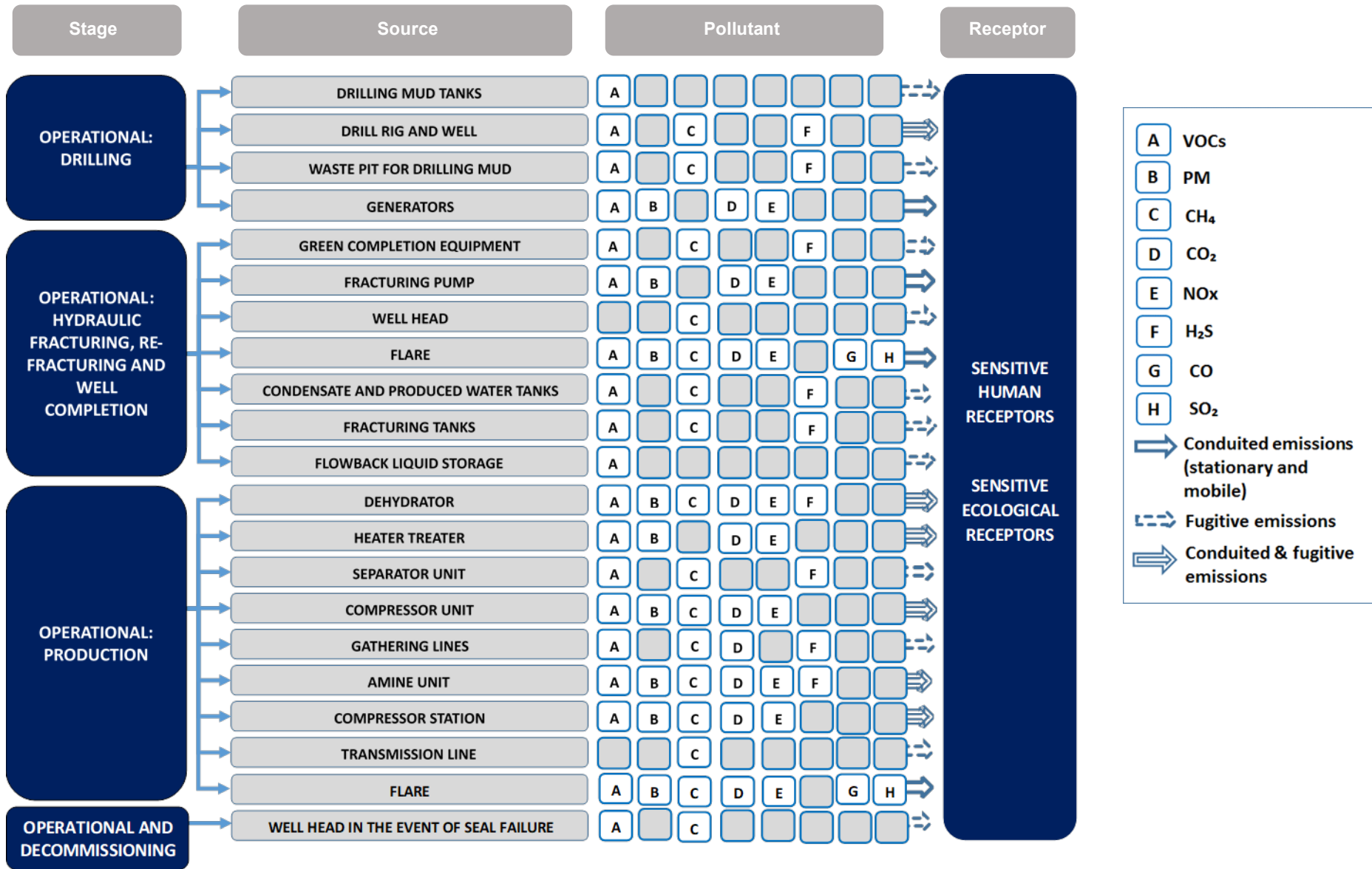
Figure 2.1 Indicative conceptual model of emissions to air from an OOG facility

13

Groundwater quality

The site conceptual model should inform the risk assessment that OOG operators are required to carry out to determine what activities could directly or indirectly pollute groundwater and its receptors (Defra and Environment Agency 2016). Development of a site conceptual model is an iterative process whereby it becomes better defined at each stage of the risk assessment process. The main purpose of the site conceptual model is to 'describe important hydraulic, hydro-chemical and biological processes that are at work in the soil, the unsaturated zone and the groundwater itself' and should 'describe potential environmental impacts associated with the site, and any uncertainties in how the activity will interact with the hydrogeological setting' (Defra and Environment Agency 2016). Linkages between sources, pathways and receptors need to be adequately understood and described.

Where there are uncertainties in the site conceptual model, additional site investigations may be necessary to further characterise the groundwater and subsurface environment (Defra and Environment Agency 2016). This characterisation stage provides the context needed to design the monitoring required for the protection of groundwater. Research by the Irish Environmental Protection Agency (EPA) highlighted that the characterisation stage is not as straightforward for OOG operations as it is for the other industrial operations, as there is less known about deeper groundwater conditions and potential pathways to receptors near the surface (EPA 2016). For example, it is not known what VOC species and their proportions are present in shale gas from different UK reserves. Understanding fracturing in the geological structures and the potential for fractures to be preferential pathways to receptors becomes critical (Council of Canadian Academies 2014, EPA 2016).

Section 3.2 discusses the key substances in groundwater to be monitored. These should not be considered in isolation as groundwater level monitoring is also important for the conceptual model, as it helps to understand the direction of groundwater flow, and therefore the direction and speed of potential contaminant transport. Longer term water level monitoring data, which can be obtained from nearby appropriate groundwater level monitoring stations, are also valuable for understanding seasonal changes in groundwater recharge and potential fluctuations in contaminant concentrations; existing knowledge such as aquifer properties and the findings of baseline reports (Shand et al, 2007) are important information sources for the conceptual model. Groundwater recharge also plays a role in the amount of dilution or contaminant loading from surface sources.

It is standard practice to use the site conceptual model to inform groundwater monitoring (see, for example, Council of Canadian Academies 2014, UKOOG 2015a, Defra and Environment Agency 2016, EPA 2016). Environment Agency guidance on monitoring of groundwater for landfills emphasises the importance of a risk-based approach to the proper design of groundwater monitoring programmes to focus effort on actual risks (Environment Agency 2003). Guidance on developing conceptual models includes European Commission (2010) and Environment Agency (2014).

Figure 2.2 illustrates the sources, pollutants and receptors for potential impacts on groundwater from an OOG facility. These sources could potentially affect groundwater and its receptors if not appropriately mitigated or managed. Potential impacts include:

- contamination from surface spills or leaks

- shallow aquifer contamination from leaking operational or abandoned wells

- leaks of saline water from deep formation waters to shallow aquifers
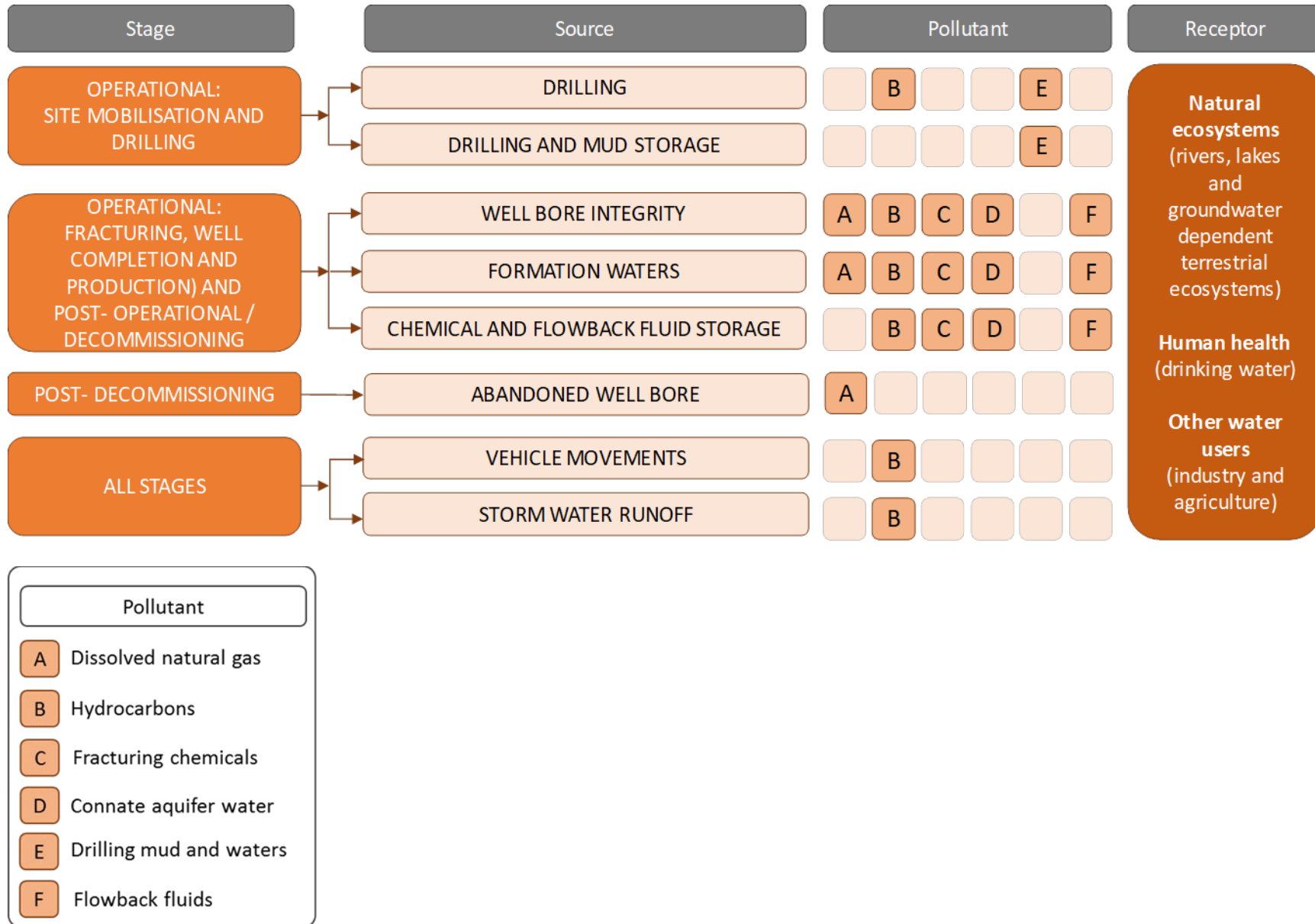
Figure 2.2   Indicative conceptual model of potential contamination of groundwater from an OOG facility

This understanding of potential pathways provides background information on practical elements that need to be considered in the statistical design of monitoring programmes such as the location of monitoring wells and depth of samples.

Groundwater monitoring wells are generally located so that they allow us to provide protection to groundwater receptors. Wells that are located to represent the main pathways are known as sentry wells or warning wells, and are used to detect contamination prior to it reaching a receptor (Council of Canadian Academies 2014). Where there is a compliance target, these wells are known as a compliance point in the UK (Defra and Environment Agency 2016). A properly developed conceptual model and a good understanding of pathways and groundwater flow rates are therefore necessary to determine suitable locations for monitoring wells.

2.3　Using the conceptual model in monitoring design

Several important considerations need to be addressed to achieve good practice in the monitoring design. The development of a conceptual model is useful for a number of reasons, not least in encouraging the planner/operator to think through the site and the operation-specific issues that surround OOG development. For operations that may affect groundwater quality, it is a statutory requirement that a conceptual model is developed during risk assessment.

It is recommended that the conceptual model should be the primary basis for deciding what, where, how frequently and over how long measurements are taken, supplemented by additional information known about the site (for example, dispersion), its operations, and known uncertainties and sensitivities. Legal and permitting requirements will also drive monitoring design. All this means that the requirements for monitoring will vary from site to site.

It is important that consistent questions are applied to the conceptual model and that these can be used to address the main objective:

> 'to ensure that there is robust evidence from a site on which OOG operations will occur, in order to establish the "baseline" condition and to ensure that information captured pre-development is suitable for characterising change and for attributing the potential causes of change should the need arise'.

To meet this objective, it is proposed that:

1. The conceptual model should be used in the first instance to define the monitoring approach.

2. The proposed design should be reviewed by the relevant regulator to consider the level of confidence and degree of precision required..

The level of accuracy/precision to be required must be considered. The level of confidence with which the operator wishes to establish a baseline condition, identify and attribute a change or demonstrate that no significant change has

occurred should also be borne in mind when deciding the monitoring campaign. For example, if contaminants that demonstrate seasonal variability are selected, data to define this seasonal variation must be available for an operator to have high confidence in conclusions reached during the statistical assessment. In particular, it is important that seasonal variations are not incorrectly attributed to changes arising from OOG operations and vice versa.

Table 2.1, which covers both air and groundwater quality, summarises the key questions that need to be addressed during the development of the conceptual model. These questions are driven by an underlying requirement to consider how much precision, accuracy and representativeness are required from a monitoring campaign in order to be able to detect a change, or to confirm an absence of change, to a given level of statistical confidence.

Table 2.1   Key questions to be addressed during conceptual model development for OOG activities to protect air and groundwater

Key question	Considerations
What should be measured to establish the baseline condition?	• Site-specific and operation-specific information relating to the contaminants likely to be produced
	• Contaminants that are not likely to be produced but which may characterise other local sources
	• Conservative markers (for example, salinity) that do not change due to biogeochemical processes or reactions such as sorption, biodegradation or dissolution
	• A selection of the OOG-specific contaminants through all stages (as identified by the conceptual model) should be considered for measurement at the baseline stage for reference during site operation/closedown activities.[1]
	• Where a large number of contaminants are identified that could arise from operations, the cost of monitoring all of them may not be proportionate. The sensitivity of receptors to the effects should therefore be considered alongside the quality of data that might be obtained. Contaminants that generally demonstrate high variability, with the potential to demonstrate only a fractional change as a result of OOG activities, would not represent high quality data compared with measurements that do not show much variation in the baseline condition. This would generally need to be considered on a case-by-case basis after the baseline has been established. To maximise confidence that the right suite of parameters is selected, adopting a similar approach to that of the guidance for the monitoring of landfill leachate, groundwater and surface water (Environment Agency 2003) is recommended whereby a broad range of parameters is monitored in the baseline establishment stage. Following baseline

Key question	Considerations
	assessment and an evaluation of their variability, the range of substances could be reduced in scope in the operational stage.[2,3]
	• Supplementary data should be collected at the site or in the locality to support interpretation of monitoring results. This should include data on groundwater levels and local meteorological data. The need for other data such as on traffic flows on nearby major roads should also be considered.
What should be measured during the operational stage?	• Similar arguments apply to the selection of contaminants for operational monitoring as in the baseline establishment stage. Here, however, the contaminants specific to the operational stage (for example, drilling, production) should be considered as appropriate. An adaptive approach using the type of conceptual model shown in Figures 2.1 and 2.1 could be considered, with preference given to selecting contaminants that occur at multiple stages as appropriate.[4]
Where should measurements occur during the baseline establishment and operational stages?	• The conceptual model can be used to identify the most appropriate sites based on climatic and geological factors (where to best detect and attribute a response), as well as identifying what receptor sites need to be protected.
	• The operator should consider whether it is necessary/preferable to establish a baseline across all or only a part of the site, or whether single or reduced monitoring locations could/should be used in the baseline stage, with further monitoring locations added once the site becomes operational.[5]
How frequently and for how long should measurements be made during the baseline stage?	• There may be some subjectivity in defining the frequency of measurement/duration of monitoring period from a conceptual model. Comparing the contaminants for monitoring with current standards can be used to help define these parameters, while estimates of the rates of flow/dispersion/residence times may help to define a reasonable frequency of measurement. For example, there is little need to monitor a high residence time aquifer on a weekly basis.[6]
	• To establish a baseline, it is necessary to have measurements at a high enough frequency over a long enough duration to establish seasonal cycles in contaminants. Predictions on seasonal variations of

Key question	Considerations
	different contaminants could be established during development of the conceptual model.
How frequently should measurements be made during the operational stage?	• The same measurement frequency as in the baseline establishment stage can be adopted. It may be possible to apply an adaptive approach to the monitoring frequency based on the conceptual model during different stages of the development on a contaminant by contaminant basis. As an example, monitoring might be scheduled around particular activities like flaring or hydraulic fracturing. • Use of continuous air quality monitoring systems can give a high frequency of measurements, which in turn would give flexibility in the choice of an appropriate resolution for statistical analysis.

Notes: [1]This report draws a distinction between 'baseline' and 'background'. Baseline refers to the set of measurements made before there is any activity onsite. Background is used to define the measurements made on alternative or control sites (that is, monitoring locations that are not affected by OOG activities).

[2] Although the list of parameters listed for landfills is not directly relevant to OOG, the approach of assessing a range of parameters and then focusing future monitoring on those contaminants most likely to provide meaningful results is relevant.

[3] More details on the suite of contaminants for consideration are given in Sections A.1.3.5 and A.1.4.2 of the Annex.

[4] More details of an adaptive approach to monitoring are given in Section A.1.4.1 of the Annex.

[5] It is worth acknowledging the precedent set in the guidance for landfill monitoring for groundwater protection, which indicates that a minimum of 3 boreholes should be used for baseline establishment, but also that where a practicality conflict exists between sampling frequency and monitoring extent, preference should be given to increasing sampling frequency (Environment Agency 2003).

[6] Adaptations in accordance with the travel times used to support the selection of an appropriate monitoring frequency for landfill monitoring in groundwater are provided in Section A.1.4.1 of the Annex.

3 Stage 2: Key substances and risks to air and groundwater quality

3.1 Air quality

The primary pollutants of concern for emission to air are associated with fugitive and combustion sources. They include:

- oxides of nitrogen (NOx) – arising from the combustion of fossil fuels (for example, vehicles, compressor engines and flares)

- VOCs – resulting from the dehydration of natural gas

- particulate matter (PM_{10}, that is, <10μm, and $PM_{2.5}$, that is, <2.5μm) – arising from site preparation, construction, vehicle movements and combustion sources

- carbon monoxide (CO) – arising from incomplete combustion of carbon-based fuels in engines and during flaring

- sulphur dioxide (SO_2) – arising from the combustion of sulphur-based fuels

- ozone (O_3) – forms regionally from sunlight and emissions of NOx + VOCs

- methane (CH_4) – fugitive emission from gas processing equipment, particularly when it is being operated under high pressure

The likely source–pathway–receptor characteristics of these emission sources are summarised in Table A.1 in the Annex.

A screening approach is recommended to enable the range of substances measured to be minimised without any significant loss in survey robustness. For example, it may be possible to correlate concentrations of NOx, sulphur dioxide and/or PM_{10} and $PM_{2.5}$. Similarly, it may be possible to correlate levels of methane and individual VOCs. This enables a reduced set of measurements to be made, with the potential for extending the range of measurements if a potential issue is identified.

Where locations are remote from other sources and there are no local receptors, a proportionate and adaptive campaign should also be considered. This could include monitoring different contaminants in different phases and at different frequencies, and scaling back the intensity of monitoring as applicable, subject to agreement with the regulator. However, a cautious approach is required to ensure any correlations identified during the baseline establishment stage hold during the operational stage.

Recorded ambient concentrations of air pollutants associated with OOG developments should be assessed against the air quality standards and guidelines for each pollutant. These values provide an indication of the limits of detection. The measurement campaign needs to be adequate to measure concentrations reliably for assessment against these standards and guidelines.

Due to the range of other factors that influence ambient air quality, however, there is no guarantee that assessment against these standards and guidelines will enable the detection of a change in air quality due to OOG activities. In addition, the occurrence of a detectable change does not necessarily mean there is a significant health or harm impact on sensitive receptors.

The monitoring guidelines for conventional compliance monitoring are matched to legal requirements, and are designed to provide a balance between practicality and the requirements of statistical analysis methods. However, the guidelines may not reflect the time period necessary for a robust application of statistical methods for formal change detection and could potentially lead to the misinterpretation of results obtained from non-formal methods such as conditional analysis.

In addition to the use of appropriate methods for the collection of recorded air quality data, operators must have in place a suitable protocol for the reporting of recorded concentrations to regulators, local residents and environmental groups.

3.2   Groundwater quality

Potential pollutants from OOG activities that may be contaminating groundwater and its receptors are identified in Figure 2.2. The key substances requiring monitoring will be informed by the risk assessment and the proposed operations at a particular site, as well as by the conceptual site model.

The potential sources of pollutants from OOG facilities and associated groundwater quality considerations can be summarised as follows (Vengosh et al. 2014, EPA 2016):

- Stray gas – dissolved natural gas components, including methane and stable isotopes for fingerprinting naturally occurring methane. These will change through the cycle of exploration, pre-production, production and decommissioning.

- Flowback fluid and produced waters from well leaks or storage on the surface with chemicals such as chloride, sodium, bromide, heavy metals and NORM. The concentration and potential range of chemicals is site-specific.

- Chemicals within hydraulic fracturing fluids – additives make up 0.1– 0.5% of hydraulic fracturing fluids (API 2010, AMEC, 2014). The number of different additives registered for use in the USA for hydraulic fracturing fluids is quite high, with the US Environmental Protection Agency (USEPA) reporting 692 unique ingredients for base fluids, proppant and additives in hydraulic fracturing fluids (USEPA 2015).

- Drilling muds or fluids – UK onshore shale gas well guidelines for the exploration and appraisal phase (UKOOG 2015b) recommend that OOG operators use water or water-based fluids. Water-based fluids are primarily composed of water or brine with barite and clay, but sometimes include chemical additives.

- Hydrocarbon contamination from surface spill and leaks

- In addition to potential well leaks, shallow aquifers could be contaminated by the migration of deep saline water or hydraulic fracturing fluids through fractures.

This list is of potential indicators and is not intended to be a monitoring specification. Monitoring programmes should be developed on a site-specific basis and be proportionate to the proposed development.

The likely source–pathway–receptor characteristics of these emission sources are summarised in Figure 2.2 (see also Table A.1 in the Annex). Existing guidance (for example, Environment Agency 2003, UKTAG 2012) explains how to determine appropriate assessment criteria for groundwater quality parameters and how these should be used to determine risk to groundwater receptors. They are not therefore considered as part of this study, whose focus is on how to detect change from the baseline using statistical techniques.

4 Stage 3: Site and operation specific monitoring design

Sections 1 and 2 focus respectively on Stages 1 and 2 of the process of establishing a baseline and determining change, that is, the development of a conceptual model and how it can be used to define appropriate monitoring locations, contaminants for monitoring, and the frequency and duration of the monitoring campaign for the baseline and operational stages.

This section deals with Stage 3, which looks at how the proposed monitoring design can be refined in the context of its objectives to make it specific to the site and operation. The full process is shown in Diagram 1 (Figure 4.1). This starts with the development of the conceptual model, incorporating some of the components discussed in the previous sections.

For **air quality**, the conceptual model, standards to be adopted and principles of measurement are fairly well developed and can be defined in a fairly generic way. This document expands on some of the concepts and provides recommendations specific to OOG (see Note 1.1), It is recommended that these are considered alongside a review of whether the monitoring design is likely to deliver robust data for baseline establishment via reference to both the objective and the key questions of 'what', 'where', 'how often' and 'for how long' (see Note 1.2). This was deemed necessary as some of the air quality standards that might be used to define the monitoring duration require only short durations of measurements and may not capture the full effects of seasonality.

For **groundwater quality**, the conceptual model and standards to be applied for compliance can be regarded as case specific. Guidance already exists for determining hazardous and non-hazardous substances, appropriate standards and thresholds, and how these should be used to determine risk to groundwater receptors (UKTAG 2013, Environment Agency 2017, JAGDAG 2017). As with air quality, it is recommended that the standards are considered alongside the statistical considerations to deliver robust data for baseline establishment (see Note 1.3). In the case of groundwater quality, the primary concern over whether the data collected will be adequate for purpose is about the proposed frequency of observation and the total number of observations collected.
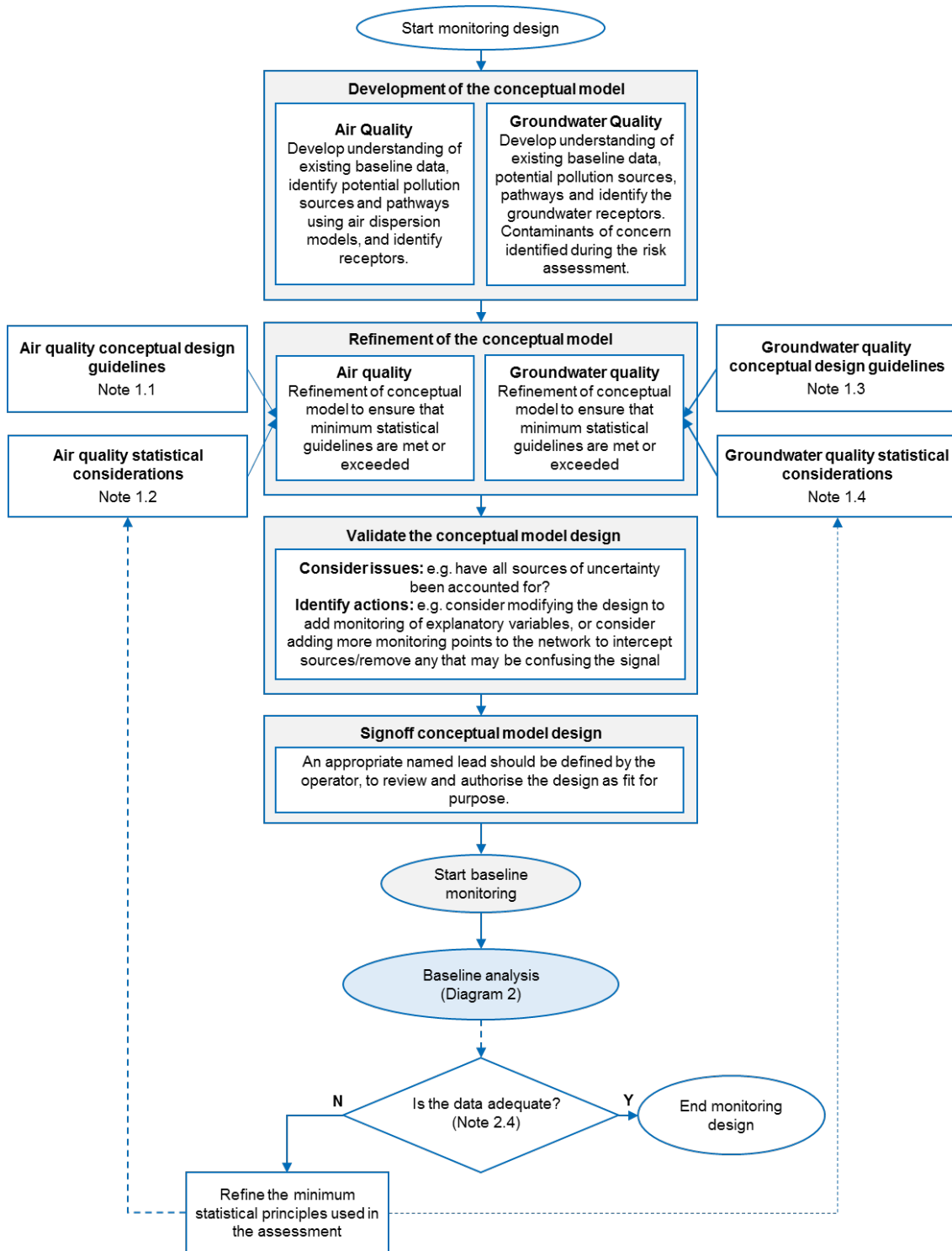
Figure 4.1   Diagram 1: Flowchart for the monitoring design for establishing a pre-development baseline for OOG activities

Following refinement of the conceptual model and signoff of the monitoring design, the next step is to begin monitoring and data assessment. However, the monitoring design stage is not complete until the data collected are determined to be adequate for purpose (assessed in Diagram 2; see Section 5.2). In some circumstances, the operator or regulator may wish to consider extending the duration of baseline monitoring for some substances (e.g. beyond 3 months for PM_{10} beyond) and/or increasing the frequency of observations (e.g. above 6

monthly borehole observations for methane). The circumstance could arise where the data are not deemed adequate because the seasonal cycle is not fully captured or where there are few data points, and there is large expected variation between years. Practical considerations may, however, mean that further refinements to the monitoring design are not feasible. In this case, the statistical consideration would be updated to reflect the recognised constraints but the monitoring design would not be altered. Continuing the sequence of the flowchart, the data would be deemed adequate with reduced confidence and the flowchart of monitoring design for the baseline would be deemed completed.

The minimum duration for baseline establishment is suggested in Note 1.1. It is assumed that there would be a seamless transition from the baseline establishment stage to the operational stage once this minimum duration is reached. If there is a scheduled break between the baseline establishment stage and the commencement of OOG activities, it is recommended that – where practicable – monitoring should continue and the baseline establishment period extended (perhaps using a reduced monitoring frequency as appropriate). A pause in the monitoring could allow time for the baseline to change for reasons unconnected to OOG activity (for example, evolving impacts from road traffic); where monitoring has not been continuous, a correction will need to be applied for the underlying trend. This correction can be confirmed if data are available from 'background' monitoring locations (that is, alternative 'control' monitoring points that are similar to the site but unaffected by the OOG activity). For groundwater this could be an up-gradient well, and for air quality an upwind monitoring station.

4.1   Air quality

Note 1.1: Design of air quality monitoring programme

When designing an air quality survey, the principal objective is to answer the following questions:

- Is there a detectable change in levels of air pollutants at locations affected by OOG activity compared with the levels measured at locations not affected by the development?

- Can this change be attributed to the OOG activity?

- What is the magnitude of any attributable change?

The reference point for evaluating a change in measured concentrations is measurements taken before operations began i.e. the baseline. Detecting change from simultaneous measurements at other locations ('background sites') may also be conducted if there is no baseline. If measurement and analysis tools are combined with meteorological observations, it is possible to attribute changes to sources e.g. by characterising the contributions of individual sources to air-pollutant levels measured at a single location that is impacted by multiple sources.

The design of the monitoring campaign for the assessment of ambient air pollution around OOG facilities will be influenced by several factors including:

- site-specific risks

- the presence of nearby sensitive receptor sites

- the nature of local meteorological conditions

A successful and representative monitoring campaign will consist of an appropriate number of air quality monitors and, if required, weather stations, located in positions that reflect the potential risks posed by the facility.

Where a site has no existing pollution sources nearby, a single appropriately situated monitor may be suitable. More than one monitoring location, however, may be required in other circumstances. For example, in areas with more than one OOG site, there is no guarantee that the background concentrations recorded at one site will be representative of another site. It is therefore recommended that a monitoring station is set up at each well pad, or in a location representative of a cluster of well pads. If the location of future well pads is known in advance, a monitoring station could be strategically placed to represent all future wells. In order to monitor the effects of increased HGV movements on nearby roads, it may also be necessary to establish additional monitoring locations away from the site (for example, at the side of the main access road).

The positioning of a monitoring station is determined by the characteristics of the development site and the local area. Where monitors or samplers are positioned offsite, they should be placed at one or more of the following:

- Site boundary or 'fence line' – where net emission fluxes from the site can be estimated from concentration transects measured at the permit boundary of the site, in support of regulation and national reporting;

- Residential properties, residential areas and other sensitive locations – providing localised measurements at high sensitivity receptors;

- Location of maximum offsite impact, and/or of maximum prominence as shown by the strength of the impact "signal" from on-site sources compared to the "noise" from off-site sources. The location will depend on the height and discharge conditions of the sources e.g. on-site methane emissions at ground level may travel further if they are under high pressure or undergo plume rise. The location of maximum impact and/or prominence is likely to be determined by atmospheric dispersion modelling.

- Background locations – in order to determine ambient concentrations due to other, off-site, sources

- Local air cavities – where pollutants associated with onsite processes may accumulate

The position of onsite monitors will be affected by the management and location of site activities. Both onsite and offsite monitoring will need to consider the availability of power and appropriate access.

Air quality surveys for OOG developments in the UK are required to achieve the following objectives:

- To detect airborne concentrations at levels below 20% of the air quality criteria in line with the maximum amount of percentage reduction requirements set under EU Directive 2008/50/EC

- To detect variations from baseline at around 20% of the levels typically recorded in the vicinity of OOG activities to enable robust conclusions to be drawn regarding the contribution of such activities to measured concentrations during the operational stage, if possible in the context of other factors affecting measured baseline air quality levels.

The following aspects should also be taken into account when designing air quality surveys for OOG developments in the UK.

- Variations in the nature and quantity of substances emitted during the operational lifetime of the site should be allowed for.

- A comprehensive environmental monitoring analysis should enable the contribution to airborne pollutant levels of any unplanned releases to be identified.

- Measurement at different locations around an OOG facility and/or at a range of heights may enable the contribution of different sources to be distinguished.

- A default minimum for baseline monitoring of one calendar year is suggested. Existing guidance suggests that 6 months of data may be sufficient to determine compliance with air quality standards and guidelines. However, analysis of case study data indicates that a full year of data may be needed to characterise any change from baseline due to OOG activity, or to be confident that no change due to OOG activity has occurred in the context of other factors affecting measured pollution levels.

- A screening approach will enable the range of substances measured to be minimised and thus avoid excessive survey costs. For example, it may be possible to correlate levels of nitrogen dioxide, sulphur dioxide and/or PM_{10}/$PM_{2.5}$. Similarly, it may be possible to correlate levels of methane and individual VOCs. This would enable a reduced set of measurements to be made, with the potential to extend the range of measurements if a potential issue is identified. In addition, a proportionate and adaptive campaign should be considered where locations are remote and there are no local receptors. This could include monitoring different contaminants in different stages and at different frequencies – scaling back the intensity of monitoring as applicable.

- As well as baseline and change detection requirements, the monitoring design should also consider what information is required, should a change be detected, to enable attribution to its cause. Examples of these data include non-target source information and local meteorological data (wind direction, speed and temperature). Satellite images and photo-panorama around the monitoring site could also be collected in advance of the baseline monitoring

campaign, and prior to the operation of the site, so that any changes in land use (for example, new agricultural buildings, roads or housing) can be identified.

In collating and reviewing ambient air quality data, operators should adhere to the quality assurance (QA)/quality control (QC) requirements set out in Defra's Local Air Quality Management technical guidance TG16 (Defra 2016).

Outline recommendations for monitoring of air pollutants relating to the duration and frequency of measurements for pollutants of concern are given in Table 4.1. The recommendations are based on EPA guidance for baseline monitoring systems for OOG facilities (EPA 2016), with amendments to reflect the requirements for air quality monitoring and local air quality management in the UK.

Table 4.1 Recommended monitoring criteria for emissions from OOG facilities

Pollutant	Proposed monitoring
NOx	**Site monitoring** • Monitoring should enable short-term peaks and long-term concentrations to be characterised. A cost-effective approach may be to use a small number (say, one or two) continuous monitoring stations to provide robust data on short-term and long-term mean concentrations, supplemented by diffusion tube measurements to provide indicative long-term mean concentrations. • To enable comparison with air quality objectives, and detection of change from baseline where possible, monitoring should be conducted for a minimum of one year to allow the calculation of an annual average and the determination of hourly averages. **Roadside concentrations** • Modelled background concentrations of NOx and nitrogen dioxide can be used to determine traffic-related impacts as set out in the 'Design Manual for Roads and Bridges' (Highways Agency et al. 2007). • To provide a more accurate representative of roadside concentrations, it is recommended that additional monitoring is carried out at a roadside location on one of the main access routes to the site. Ideally this would be carried out over a period of 6 months, covering both winter and summer months; if this is not possible for practical reasons, a shorter period – no less than 3 months – would be acceptable.
Sulphur dioxide	• This should be carried out for no less than one month. However, monitoring may not be necessary if produced gas will not be routinely combusted at the installation. The sulphur content of the gas will also need to be considered when assessing monitoring requirements.
Carbon monoxide	• No supplementary baseline monitoring is required due to a lack of evidence to suggest significant carbon monoxide emissions associated with OOG activities and consistently low background carbon monoxide levels across the UK. • Data from the Automatic Urban and Rural Network (AURN) should be sufficient to determine a baseline.
Ozone	• No supplementary baseline monitoring is required as ozone is not a primary pollutant released directly by OOG processes. It is a secondary photochemical pollutant that forms regionally from NOx and hydrocarbons under

Pollutant	Proposed monitoring
	sunlight, so it may be appropriate to monitor NOx and hydrocarbons as ozone precursor pollutants. • Research for the Environment Agency indicated that OOG activity might be expected to give rise to no more than minor changes in ozone levels, including potentially a small net reduction in ozone concentrations at some locations (Ricardo Energy & Environment in press). • Data from the AURN should be sufficient to determine an ozone baseline if needed.
Particulate matter (PM$_{10}$, PM$_{2.5}$ and total suspended particulates)	**Site monitoring** • To enable comparison with air quality objectives and detect any change from baseline, onsite monitoring should be carried out for a minimum of one year to enable daily and annual averages to be determined, and any change that can be attributed to the operation of the OOG facility to be identified. **Roadside concentrations** • Concentrations of particulates should be determined at roadside locations along the main access routes to the site. This can be done either through establishing an additional monitoring location or by using modelled background concentrations. **Dust** • If the site will be stockpiling large quantities of coarse material (i.e. sand), it may be necessary to carry out additional dust deposition sampling at an appropriate offsite location. However, it may be possible to address this through the application of appropriate dust management practices, detailed in a site-specific dust management plan.
Methane, benzene and non-methane volatile organic compounds (NMVOCs)	• In order to enable the detection of change from baseline, monitoring should be carried out at the well pad to provide hourly concentrations of total NMVOCs and methane over a period of at least one year. • A monitoring programme should be put in place for periodic sampling and assessment of concentrations of NMVOCs (for example, benzene, toluene, ethylbenzene and xylene, BTEX) through the use of gas chromatography–mass spectrometry (GC-MS). • Formaldehyde concentrations are not easily determined by GC-MS and so alternative means of estimating its background concentrations should be considered.

Pollutant	Proposed monitoring
Polycyclic aromatic hydrocarbons (PAHs), assessed as benzo[a]pyrene	• Measurements of benzo[a]pyrene should be carried out within the proposed well pad for at least one year (no temporal resolution is specified).

During the operational stage, it is recommended the following elements of the baseline monitoring campaign be applied:

- Well pad development – PM monitoring only

- Drilling – monitoring for NOx, PM_{10}, $PM_{2.5}$, total NMVOCs, methane, PAHs and sulphur dioxide (minimum one month)

- Hydraulic fracturing – monitoring for NOx, PM_{10}, $PM_{2.5}$, total NMVOCs, methane, PAHs and sulphur dioxide (minimum one month)

- Completion – monitoring for NOx, PM_{10}, $PM_{2.5}$, total NMVOCs, methane, PAHs and sulphur dioxide (minimum one month)

- Production – continuation of methane and total NMVOC monitoring

During production, monitoring for combustion gases should also be undertaken if combustion processes are present and in use (flare, compressor and so on). The operator is recommended to periodically review these data during the production stage. If there is no evidence of any detectable or significant increases in ambient pollutant concentrations due to process operations, monitoring may be discontinued.

As discussed in Section 2.2, the timeframes over which emissions occur will have a significant bearing on their potential air quality impact. It is therefore essential for operators to develop a clear outline of the timescales for OOG developments and toplan their monitoring campaigns accordingly.

Note 1.2: Statistical considerations in air quality monitoring design

The design and recommended monitoring discussed in the preceding section are focussed on demonstrating compliance with air quality standards and good practice, with some modifications for OOG situations. However, they do not necessarily reflect what is needed to collect robust data for the purpose of establishing baselines, detecting changes, or determining that no change has occurred. This report does not give explicit quantitative recommendations for minimum durations or frequencies of measurement. The conceptual model and reference values in the recommendations should be used in the first instance. To increase confidence in detecting change that could arise from OOG activities, however, the design should be subjected to a review stage where the key questions are re-assessed (Table 4.2).

Table 4.2 Key questions to be addressed during air quality monitoring design for OOG activities

Key question	Considerations
What should be measured to establish the baseline condition/ assess change?	• What is the variability and detectability of the contaminants selected? • Are there other unspecified contaminants that have general low background environmental variability which should be considered?
Where should measurements occur during the baseline establishment and operational stages?	• Are multiple monitoring stations used? Can monitors be deployed to triangulate sources, or in pairs upwind and downwind of sources to estimate net pollution fluxes? Such configurations should be considered on a site-by-site basis. • Local conditions, including topography and meteorology, should be considered when selecting monitoring locations. • The confidence with which assessments can be made will be increased with increased spatial coverage. Also, having background sites can clarify background signals.
How frequently and over how long should measurements be made during the baseline establishment period?	• For greater certainty in capturing variability in background environments, it is recommended that a full seasonal cycle is captured i.e. one year. The frequency of measurements is likely to be adequate for the purpose of comparison with conventional air-quality standards (Annex Section A.1.3.5). However, the duration of a period of changed concentrations can be short. As shown in Case Study 1 (Annex Section A.3), several factors may result in short-term increases in ambient pollutant concentrations. For the purpose of detecting or discounting the occurrence of change, these factors need to be characterised as part of an annual cycle. Practical considerations related to sample duration are set out in the Annex.

4.2 Groundwater quality

Note 1.3: Design of groundwater monitoring programmes

The Petroleum Act 1998 (as amended by the Infrastructure Act 2015) requires that, for high volume hydraulic fracturing, baseline monitoring is required for a period of at least 12 months for methane in groundwater (Environment Agency 2016). Environment Agency (2016) further states that the baseline data should consist of 3 sets of data at a minimum in order to determine the natural variation. However, it is emphasised that the hydrological conditions at a

particular site should determine the duration and frequency of baseline monitoring (Environment Agency 2016). In reality, the frequency of groundwater monitoring to establish a baseline will depend on:

- the stability and seasonality of the water quality parameter

- the potential for contamination from existing sources

- pollutant travel times

- the hydrogeology of the site

Examples of groundwater data and analysis are given in the case studies in Section A.3 of the Annex.

The frequency of groundwater monitoring during the operational stage is not specified here as it depends on the pollutant travel times, hydrogeology of the site and the overall risk to receptors (Defra and Environment Agency 2016). Environment Agency guidance for the OOG sector emphasises that monitoring should reflect the different activities at the site, with higher frequencies possibly being required for higher risk activities such as well stimulation (Environment Agency 2016).

When establishing a monitoring programme, background groundwater quality both upgradient and downgradient of the OOG site should be considered. The number of sample locations required will depend on:

- the complexity of the hydrogeology at the site

- receptor location

- the risk from potential sources of pollution from the OOG site

In the context of OOG operations, the site is often considered to be the well pad boundary. However, particularly in the case where there are horizontal wells, adequate spatial coverage is required because the sources of pollution may extend some distance from the well pad towards sensitive receptors (EPA 2016).

Baseline monitoring will help to assess existing conditions against which changes can be identified and tracked. Baseline monitoring is required at the local site-specific level as well as at the regional scale (Environment Agency 2016, EPA 2016). The presence of pre-existing groundwater contamination and the lack of comprehensive baseline monitoring data have made it difficult in some cases to determine the impacts from OOG activities (Brantley et al. 2014).

Regional monitoring undertaken by the British Geological Survey (including the "Baseline chemistry of groundwater in UK aquifers" series see Shand et al, 2007) and the Environment Agency could be used to inform the conceptual model and baseline monitoring design by giving an indication of existing groundwater quality and variability in monitoring parameters. Regional data will prove useful as a longer duration of monitoring might be available and can be used to inform the baseline. The appropriateness of using nearby regional monitoring boreholes to inform the baseline for a particular site would need to be considered on a case-by-case basis.

Note 1.4: Statistical considerations in groundwater monitoring design

The approaches applied to the interpretation of groundwater monitoring data in the UK relate to existing groundwater quality standards, but do not necessarily reflect what is needed to collect robust data for the purpose of establishing baselines, detecting changes, or determining that no change has occurred. This document does not make explicit quantitative recommendations for minimum durations or frequencies of measurement; the conceptual model should be used in the first instance. To increase the confidence in detecting change that could arise from OOG activities, however, the monitoring design should be subjected to a review stage where the key questions are re-assessed (Table 4.3).

Table 4.3 Key questions to be addressed during groundwater monitoring design for OOG activities

Key question	Considerations
What should be measured to establish the baseline condition/ assess change?	• What is the general variability and detectability of the contaminants selected? • Are there other unspecified contaminants that have general low background environmental variability which should be considered?
Where should measurements occur during the baseline establishment and operational stages?	• Are multiple downgradient boreholes to be used? From a statistical perspective, there are advantages and disadvantages to having more frequent observations versus more monitoring sites. Existing guidance from the Environment Agency (2003) on landfill monitoring indicates that, where practicality conflicts arise, preference should be given to more frequent observation. However, evidence for a generic statement on this is under review following research which may indicate that increasing the spatial coverage will give better representation of transient impacts. There is therefore no explicit guidance at present. • Are upgradient boreholes to be used? There is no requirement to have a specific configuration of upgradient and downgradient boreholes, and arrangements should be considered on a site-by-site basis. However, having upgradient and downgradient boreholes could improve confidence in attribution.
How frequently and over how long should measurements be made during the baseline establishment period?	• For greater certainty in capturing background environmental variability, it is recommended that a full seasonal cycle (that is, one year) is captured. Some work has been done to assess the minimum number of data points required for baseline in landfill (16–20 with noted reservations) and some assessment has been made in this project for

Key question	Considerations
	methane (see Note 2.4). For Water Framework Directive trend assessment, a minimum of 8 observations is required.
	• Monitoring frequency and duration will be site and contaminant specific, and frequency should be determined by the need to produce a dataset that the operator and regulator can have sufficient (statistically significant) confidence in. As a rule of thumb, a minimum of monthly measurements over 12 months should be considered for rapid travel time aquifers (<2 years), and quarterly over 2 years for all other aquifers. The requirements of the Petroleum Act 1998 (as amended by the Infrastructure Act 2015) for 12 months of baseline groundwater monitoring are an important consideration.
How frequently should measurements be made during the operational period?	• Is the measurement frequency consistent for baseline and operational stages? To avoid introducing bias into the analysis, it is generally advisable to have consistent, regular observations. An example would be a move from analysis at a quarterly level or to a biannual level; this move might change the variability in the observations, which would need to be accounted for in the analytical method chosen. Having said that, there may be practical reasons for the changing the intensity of the monitoring.

5 Stage 4: Lifecycle stage specific monitoring and statistical data analysis

The preceding sections outline how a baseline and operational stage monitoring survey[2] for OOG activity should be approached in terms of:

- the substances measured

- the location of measuring points

- the frequency of measurement

- the duration of monitoring for the lifecycle stage

This section and the next one are intended to guide the user through the statistical analysis of the monitoring data. This section explains how to analyse the baseline data and then how to detect any change from baseline conditions. Section 6 looks at how to attribute this change to what might have caused it.

The analysis is structured around the recommended approach for groundwater. For air quality, these stages are reversed to reflect the different nature of influences on air quality and monitoring strategies compared with groundwater survey and data analysis. The full approaches for groundwater quality and air quality are shown in Figures 5.1 and 5.2 respectively.

5.1 Overview of the statistical analysis

This section describes the application of statistical analysis tools to establishing a baseline level and detecting a change from this baseline during the generic 'operational stage'. The guidelines used during decommissioning may be assumed to share some common approaches, but are not currently defined.

The guidelines developed in this section cover:

- statistical analysis to establish the baseline (Section 5.2)

- statistical analysis to determine change/cause and thus trigger further data analysis (Section 5.3)

- how the statistical analysis should inform ongoing monitoring design (Section 5.4)

This project reviewed the statistical methods applied to setting the baseline for air quality and groundwater quality at OOG sites, as well as approaches used for baseline and change detection for other types of operations and in other sectors. This review found that:

---

[2] Not covered explicitly as a separate flowchart, but discussed within the notes. If created, this flowchart would be as the baseline establishment phase with the exception of having an already developed conceptual model as opposed to there being a requirement to generate one.

- the methods could be grouped under various broad categories (for example, signal strengthening, change difference[3] and change detection[4])

- there was no bias towards one technique over all the others

Due to the restrictions on available data before operations began, the OOG industry has not implemented formal change detection techniques. In instances where the implications of a change have been examined, this has usually been through inference using a change difference technique.

As a standard approach in starting any statistical analysis and in common with a number of the papers identified by the literature review (irrespective of the final statistical procedure applied), decision-making principles should start with visualisation. Looking at the data both as a time series and as a distribution (using histograms or boxplots), and using these to sense check observations and later analysis, is recommended. Conducting any standard[5] QA/QC tests on the data[6] and testing the data for outliers (if not included in the standard QA/QC checks) is also suggested.

These stages are included in both the analysis used to establish the baseline and in the assessment of change. During baseline establishment, there is a step that examines the adequacy of the data. If this check indicates the data are not satisfactory, adapting the monitoring design (for example, extending the monitoring period) to establish an adequate baseline should be considered.

If this check is acceptable and operations begin, then for groundwater, it will be necessary to attribute a source to the observed change in **groundwater quality** (Figure 5.1) if there is:

- a deterioration in observed water quality (that is, a negative trend is detected)

- a change in water quality is detected through either a formal change detection method (sometimes known as a systematic change) or a difference method (standard hypothesis testing of 2 fixed points in time, which may be most suited to, for example, fugitive releases)

The ordering of this approach is subtly different for **air quality** (Figure 5.2). In view of the wide-ranging environmental influences on air quality survey data, it is considered helpful to filter out any background environmental signal before

---

[3] Change difference: Indicates a method that assesses the difference between 2 or more points in time or space (for example, the average baseline concentration versus the average operational concentration up to a point in time, or operational site concentration versus background concentration).
[4] Change detection: Formal methods for assessing if and when a change has occurred over a continuous time series.
[5] These will be contaminant-specific and referenced in the appropriate guidelines (see Sections A.1.3 and A.1.4 in the Annex).
[6] For the purposes of this project, it is assumed that these procedures are standard for the sector/parameter of interest and are adequate. A review of these methods is considered to be outside the scope of the project.

any change detection analysis is performed. In line with some of the literature studies, the following approach is therefore recommended:

1. Attempt to attribute the sources of pollution at a site.

2. Where necessary, determine if this represents a change from the baseline condition.

Change attribution is considered in Section 6. Although this simplifies detection of changes in air quality due to OOG activity, it presents major challenges in identifying (attributing) other source signals and removing them. Placement of monitors and conditional/directional analysis will be vital for this strategy to work.
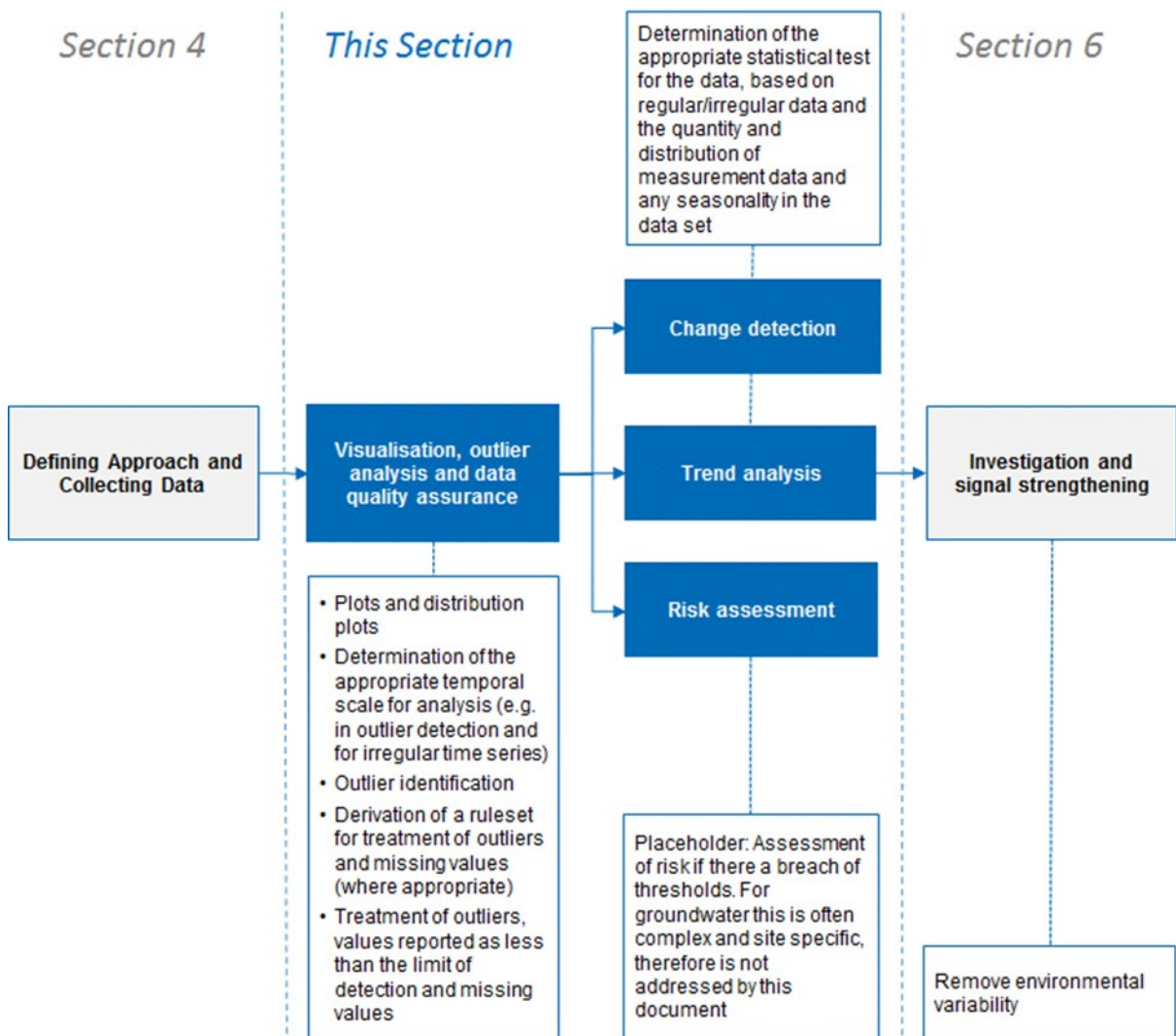
Figure 5.1   Process for statistical analysis of data for determining a baseline and/or change in groundwater quality at OOG sites
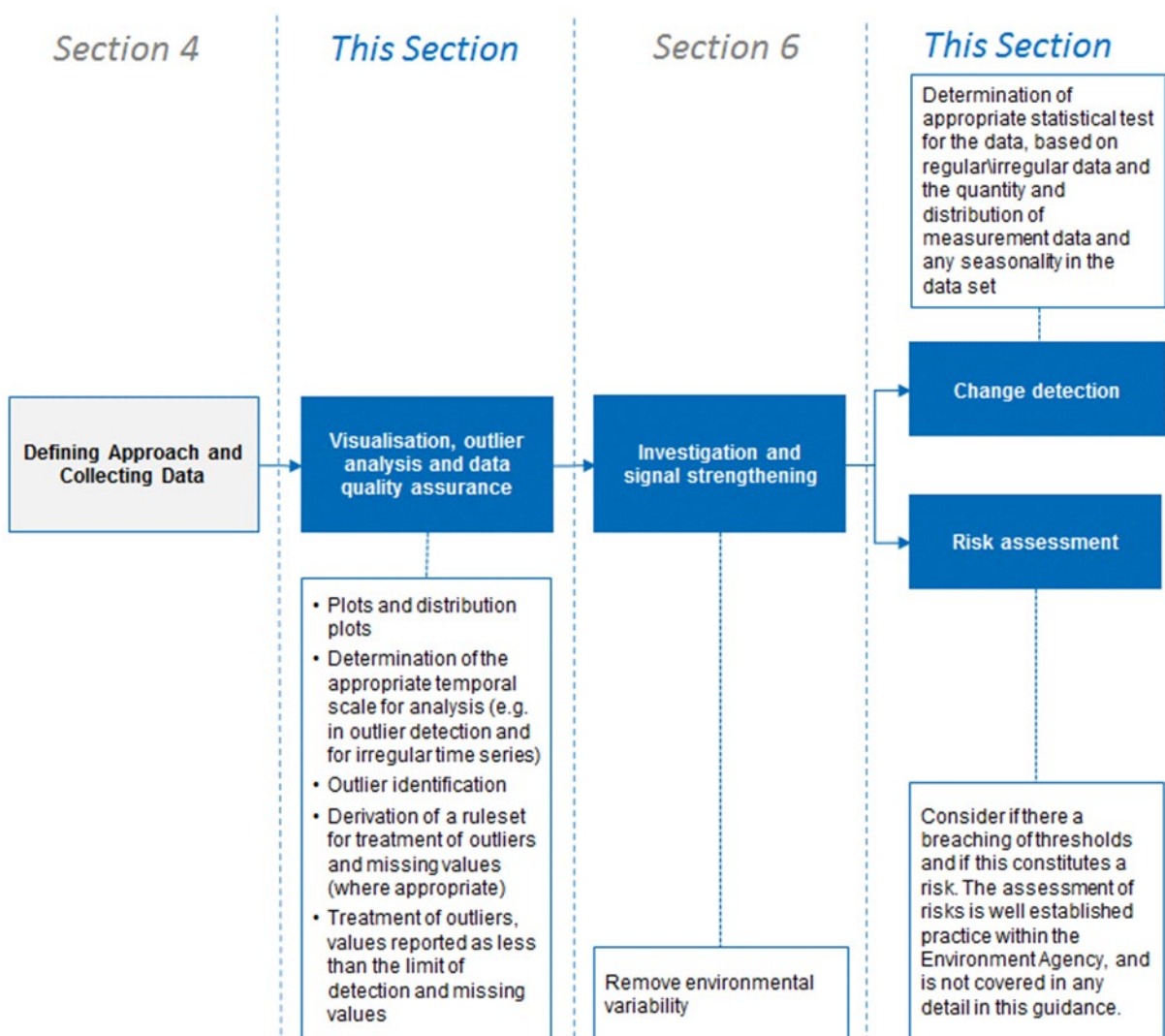
Figure 5.2   Process for statistical analysis of data for determining a baseline and/or change in air quality at OOG sites

5.2   Statistical analysis to establish the baseline

A decision tree showing the process of analysing the data to establish a baseline condition is presented in Diagram 2 (Figure 5.3).This is followed by Notes on its interpretation. This assumes that:

- ongoing QA/QC checks are conducted during data collection

- the duration of monitoring was specified during the monitoring design stage

Once the specified duration of monitoring is achieved, the data should be plotted and any outliers removed as necessary. Some interpretation of the data should be made in terms of:

- how variable each contaminant–site combination is

- whether the frequency and duration of the monitoring appears to have captured the background environmental variability (that is, that the baseline has been adequately defined)
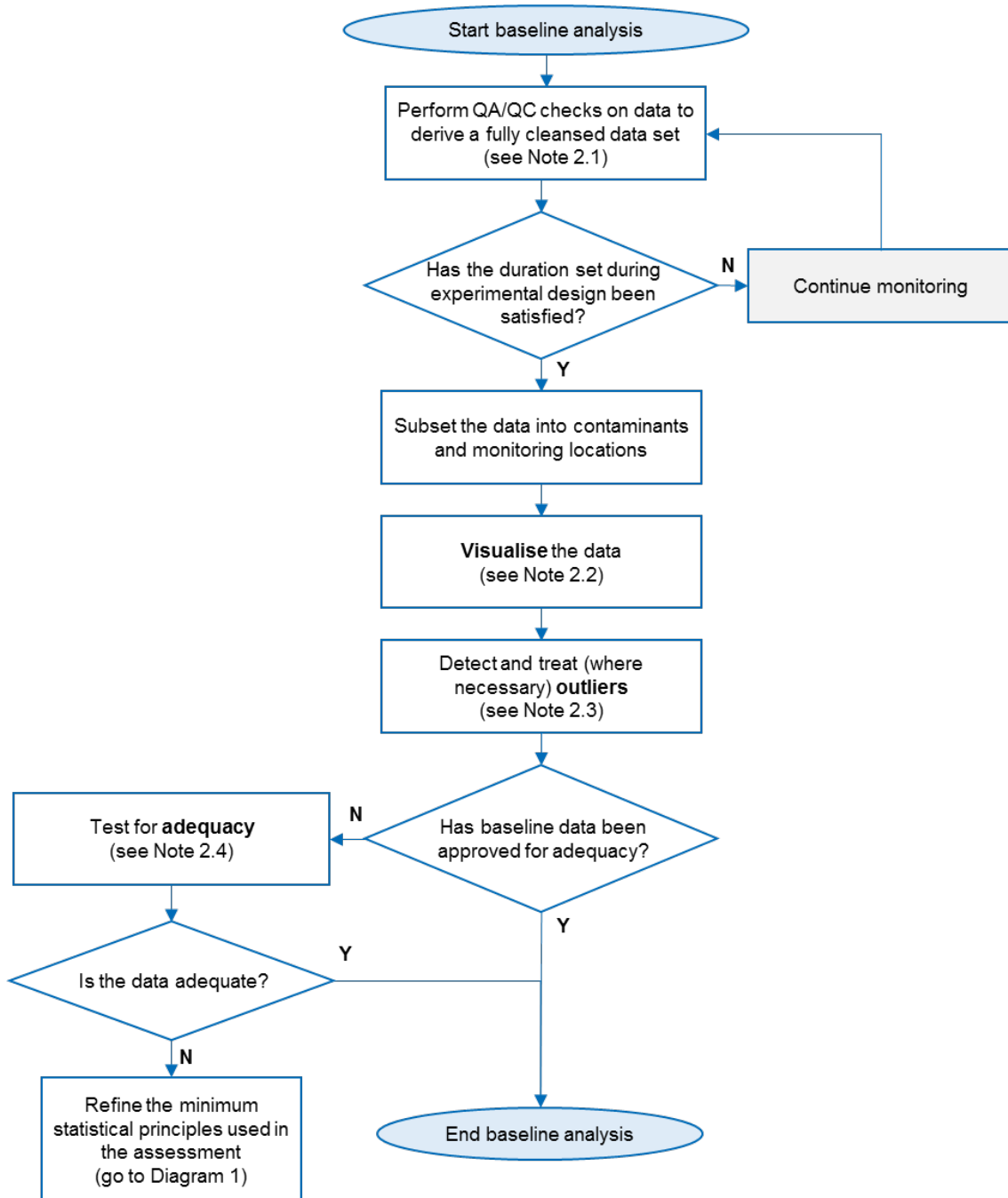
Figure 5.3   Diagram 2: Decision tree for statistical analysis of data for baseline establishment at OOG sites

Note 2.1: QA/QC checks

Laboratory procedures for QA/QC checks are well documented and are not detailed within this document. The only exception is a discussion on the limit of detection.

The Environment Agency recommends taking half of the limit of detection for values that are recorded as below this limit. Although this document makes no recommendation about modifying the limit of detection, it is noted that for formal change detection[7] methods it would be preferable that, for this activity, values below the limit of detection are excluded from the analysis except in cases where a large proportion of the observations (>40%) fall into this category. This recommendation is made on the basis that the methods of change detection use observed variance in the data and this may be unfairly biased where a single substitution value is used. Under the null hypothesis that there is no change, the removal of these values would not affect the results.

The output of this step should be a fully 'cleansed' dataset (that is, a dataset which is ready for use having had poor quality data removed or rectified). Outliers may or may not have been corrected for; cases where there has been no outlier correction are addressed in the next stage of the process (see Notes 2.2 and 2.3).

Note 2.2: Data visualisation

Exploratory data analysis involves the use of statistical techniques to identify patterns that may be hidden in a group of numbers. Descriptive statistics allow the characteristics of the underlying distribution of a dataset to be quickly described through a simplified set of values.

In the first instance, a simple time series plot of the data should be used to highlight the changes through time. It is recommended that plots are drawn over the same time period to allow direct comparison between observations to see if there are visual correlations of high and low values.

A box plot (also called a box and whisker diagram) is a standardised way of displaying the distribution of data and allows the ready comparison of multiple datasets. It uses the median, the approximate quartiles, and the lowest and highest data points to convey the level, spread and symmetry of a distribution of data values. It can also be easily refined to identify outlier data values (see Note 2.3), where outlier data values are defined as to be beyond the whiskers; some statistical software and other packages (for example, Microsoft® Excel) now default to this view.

General characteristics such as the symmetry of the distribution, the location of the central value and the spread of the observations are immediately apparent, and can be seen in the example shown in Figure 5.4. The approximate symmetry of the 2 box and whisker plots in this example show that the data are approximately normal.

---

[7] Referred to elsewhere as systematic change detection.
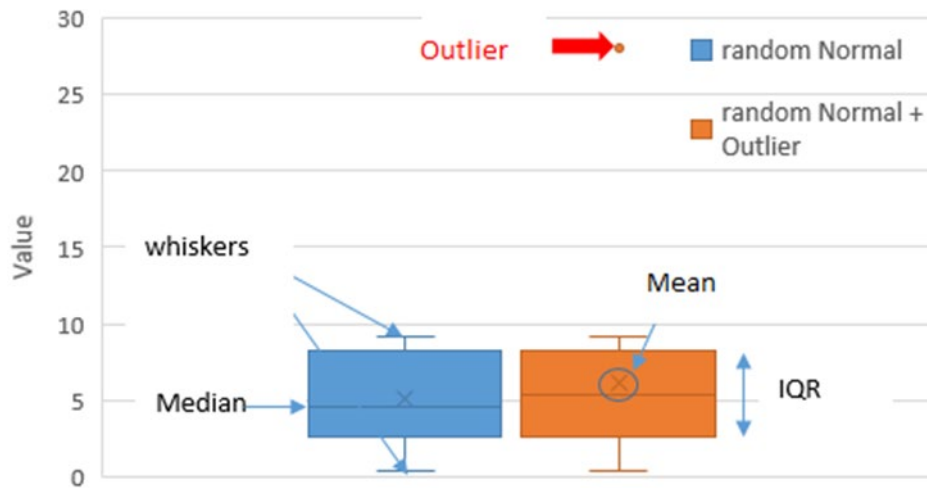
Figure 5.4   Example box plot

Care should be taken to check what default assumptions are taken in the specification of outliers (for example, whiskers in the statistical software language R default to 1.5 times the interquartile range; see Note 2.3). In Excel 2016, the default is based on the standard deviation.

Using the box plot technique, data can be visualised and the distribution plotted by:

- month

- month and year

- season

- season and year

- year

There are no fixed rules on which type of plot should be used. Some suggestions about which may be the most appropriate – based around the frequency of data collection – are given in Table 5.1.

- Where there are multiple observations per month (less than weekly), it is recommended that observations are plotted by month.

- Where observations are recorded monthly, observations should be plotted seasonally.

- Where there is more than one year of observations, by year plots could also be plotted to provide a visual interpretation of the inter- and intra- month and year variation.

Examples of the application of the different resolutions of plots are provided in the case studies in Section A.3 of the Annex.

Table 5.1 Recommended plot resolution for box plots

Frequency of observation	Plot type					
	Month	Season	Year	Month/ year	Season/ year	Full time period
Hourly	✓*		✓	(✓)		
Daily	✓*		✓	(✓)		
Weekly	✓*	✓	✓	(✓)	(✓)	
Fortnightly	✓	✓*	✓	(✓)	(✓)	
Monthly		✓*	✓		(✓)	
Bimonthly		✓*	✓		(✓)	
Quarterly			✓*			
Annually						✓*

Notes: (✓) multiple years only
* Recommended resolution of outlier analysis. This should be transferred to its bracketed counterpart where there are multiple years of data and there appears to be a significant difference between those years.

Note 2.3: Outlier detection

An outlier is defined as 'an observation which deviates so much from the other observations as to arouse suspicions that it was generated by a different mechanism' (Hawkins 1980). There are 2 main reasons why the identification of potential outliers is important.

- The outlier may be the result of sampling error and thus indicate erroneous data which should be removed from the analysis.

- The outlier may indicate true anomalies in the data which are of scientific interest and therefore robust statistical techniques need to be considered to investigate these further.

A default is suggested (and has been used in the case studies) of 1.5 multiplied by the interquartile range either side of the upper and lower quartile. This can be expressed as values that lie below the lower limit or above the upper limit of the outlier limits as defined below for the set of values x;

$$\text{Outlier Lower Lim}(x) = Q25(x) - 1.5(Q75(x) - Q25(x)) \qquad \text{(Equation 5.1)}$$

$$\text{Outlier Upper Lim}(x) = Q75(x) + 1.5(Q75(x) - Q25(x)) \qquad \text{(Equation 5.2)}$$

The suggested default aligns with the default used in many software packages and what is often shown on box plots. R will auto-generate and report outliers based on Equations 5.1 and 5.2 without the need to explicitly calculate them. Other options include a range approximating the 95% confidence interval or 99% confidence interval assuming a normal distribution (approximately 2 or 3

times the standard deviation respectively). These options will generally identify fewer observations for further analysis as potential outliers, but implicitly assume that the data are normally distributed.

In the statistical guidance provided on the analysis of landfill monitoring data (Environment Agency 2002), it is suggested that the multiple outlier test is adopted. This adopts the assumption of normality, recursively applying the algorithm on values outside of the confidence interval. This routine was available to users using the Environment Agency's Test Data Facility at the time of writing the 2002 report. This document does not stipulate which outlier test should be adopted; users should choose what they believe to be the most appropriate.

It should be acknowledged that outliers are not necessarily invalid data points; they may well be valid and the most important, information-rich part of the dataset. Under no circumstances should they be automatically removed from the dataset. Outliers may deserve special consideration: they may be the key to the phenomenon under study or the result of human error.

Note 2.4: Adequacy

This step is necessary to test that the conclusions are not statistically different with different frequencies or duration of data collection. The question is: Can you be confident that there is no seasonality, or if there is, can it be fully characterised?

Hopefully, the data received will cover a full seasonal cycle and/or no unexpected change will be seen. However, the examples within the case studies demonstrate that this is not always the case (see Section A.3 of the Annex). In Case Study 1, visual analysis of the data showed that there was an apparent increase in PM_{10} concentrations over a monitoring period with a high frequency of measurements but a short duration (7 months). In Case Study 1, the baseline establishment was based on frequencies of data collected approximately every 6 months.

To test for adequacy, it is first necessary to identify an appropriate distribution. The assumption should be that the data are normally distributed; the data should first be tested to see if this assumption is justified (see Note 3.1). Where it is not justified, the suitability of different distributions should be assessed. A guide to selecting the appropriate distribution for testing against is shown in Figure 5.5.
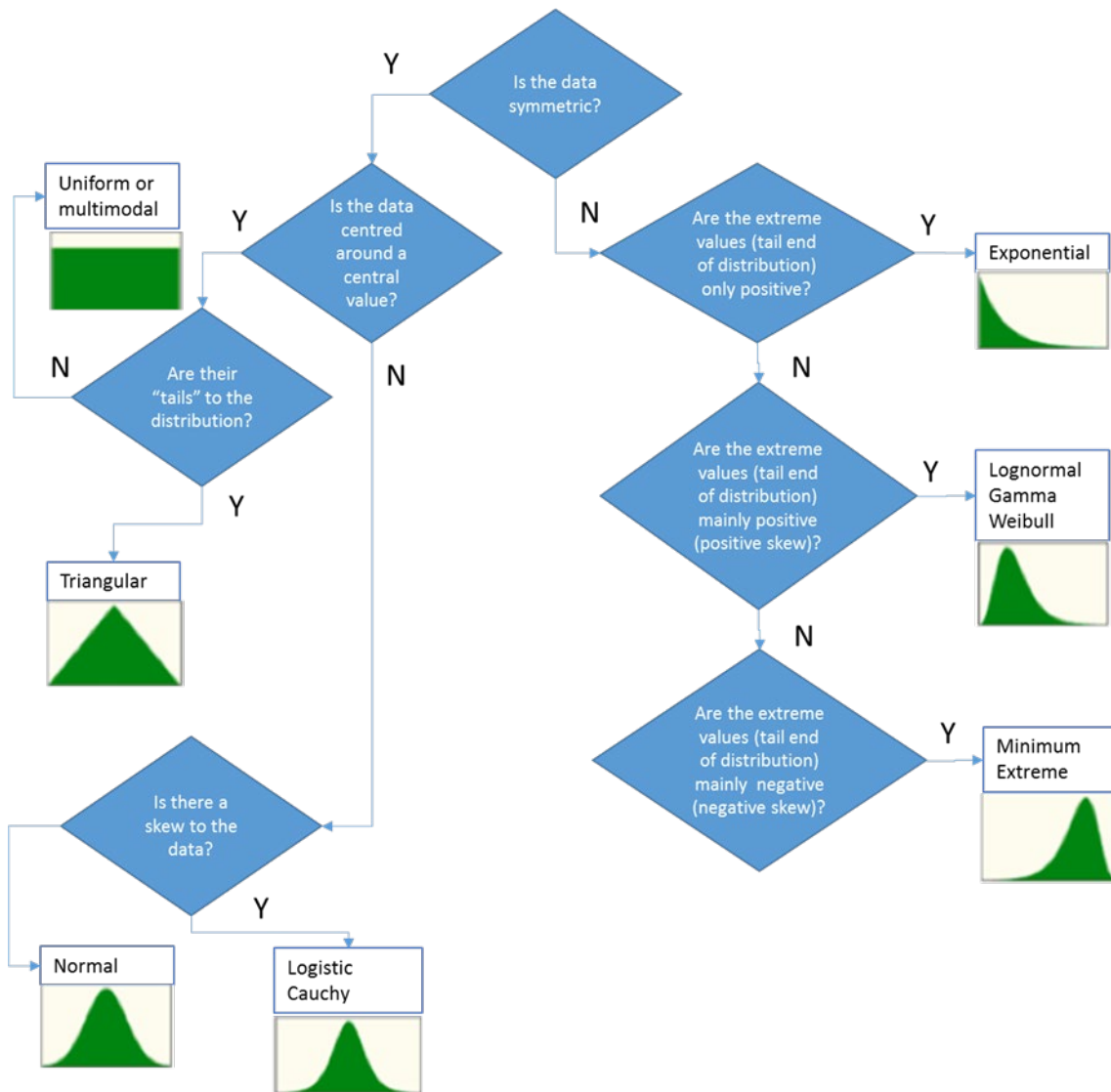
Figure 5.5   Guide to selecting type of data distribution

Once an appropriate distribution has been selected and fitted to the sample dataset, multiple simulations can be run with the data using different frequencies and durations of analysis. The results can be used to determine the sensitivity of the output to the choice made in frequency or duration set at the monitoring design stage.

This type of analysis could also be used to test what a statistically robust survey frequency and duration would be when a highly intensive dataset is available over a long enough time period. The case studies in the Annex provide examples of how data may be assessed to determine their adequacy for baseline establishment.

The discussion below is based on an example that looks at what difference a reduction in the monitoring duration or frequency might make to the outcomes of the adequacy assessment. The example uses data provided in Case Study 4, where baseline borehole observations of methane were tested for adequacy for 2 sites and were determined to be adequate. A time series plot of the observations provided is shown in Figure 5.6.
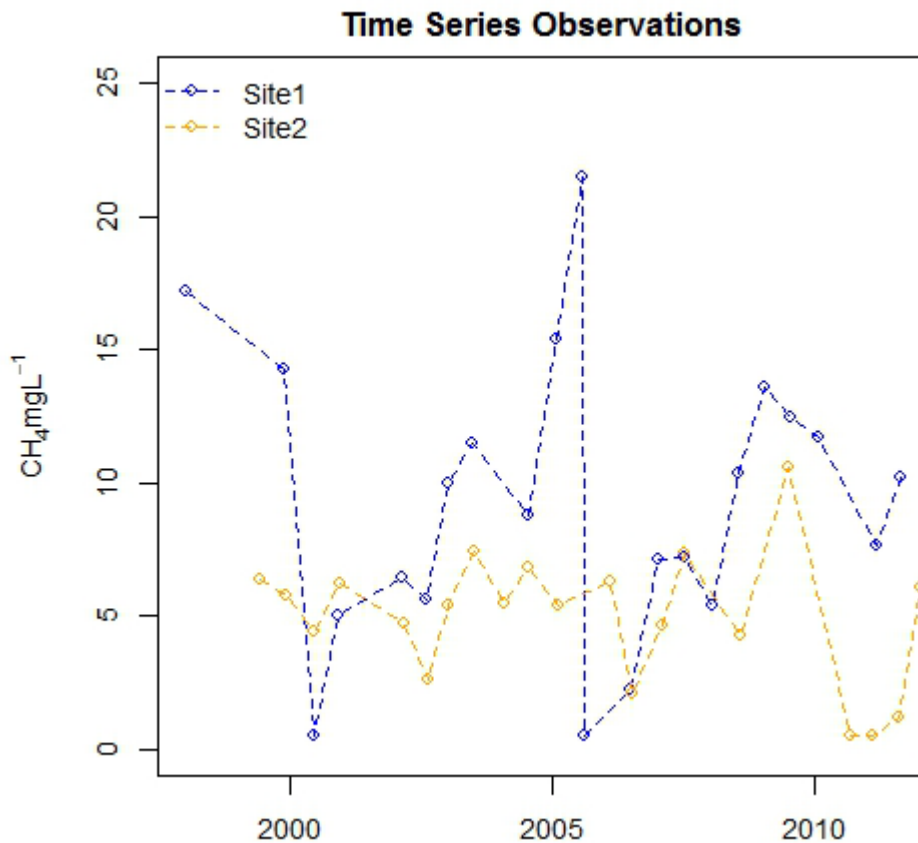
Figure 5.6   Dissolved methane observations at 2 baseline monitoring wells, 1998 to 2012

Source: National groundwater monitoring network

The frequency of observations in the dataset was restricted to approximately 2 observations per year, but occurred over a long time period (14 years). Visual analysis, of the data (see Case Study 4 in Section A.3 of the Annex for further detail of the assessment) did not appear to show any seasonality.

Below it is shown how the data can be used to investigate if a reduction in frequency of observation/duration of baseline monitoring might affect the conclusions reached – what if there was only half the number of observations or if there were twice the number of observations? If the consensus of conclusions reached is independent of the number of data points, there can be confidence that the data are adequate for use as a baseline.

The plots that follow utilise the functionality of the 'fitdistrplus' package in the statistical software language R. This functionality can be used to assess what distribution may best represent the observed data at Site 1 in Case Study 4.[8] A less powerful but more generic assessment method for assessing the appropriateness of a data distribution was given in the case study and it is not necessary to use R or the fitdistrplus package. Nevertheless, it is presented

---

[8] Useful functions include: fitdist(<data>,<distribution name>), which allows the user to generate to fit a named distribution to the data set of interest; and gofstat(list(<fitted distribution 1>,..< fitted distribution n>)) which allows the user to compare and contrast the fit statistics of the 1 … n fitted distributions (for example, normal, log normal, gamma and Weibull).

here as a useful tool. Five distributions were tested for suitability – normal, lognormal, a gamma distribution fitted by maximum likelihood estimation (gamma MLE), a gamma distribution fitted by matching moment estimation (gamma MME) and the Weibull distribution.[9] The 3 most plausible are plotted for Site 1 in Figure 5.7. Tables 5.2 and 5.3 show statistics for Site 1 reported through the 'gofstat' function in R.
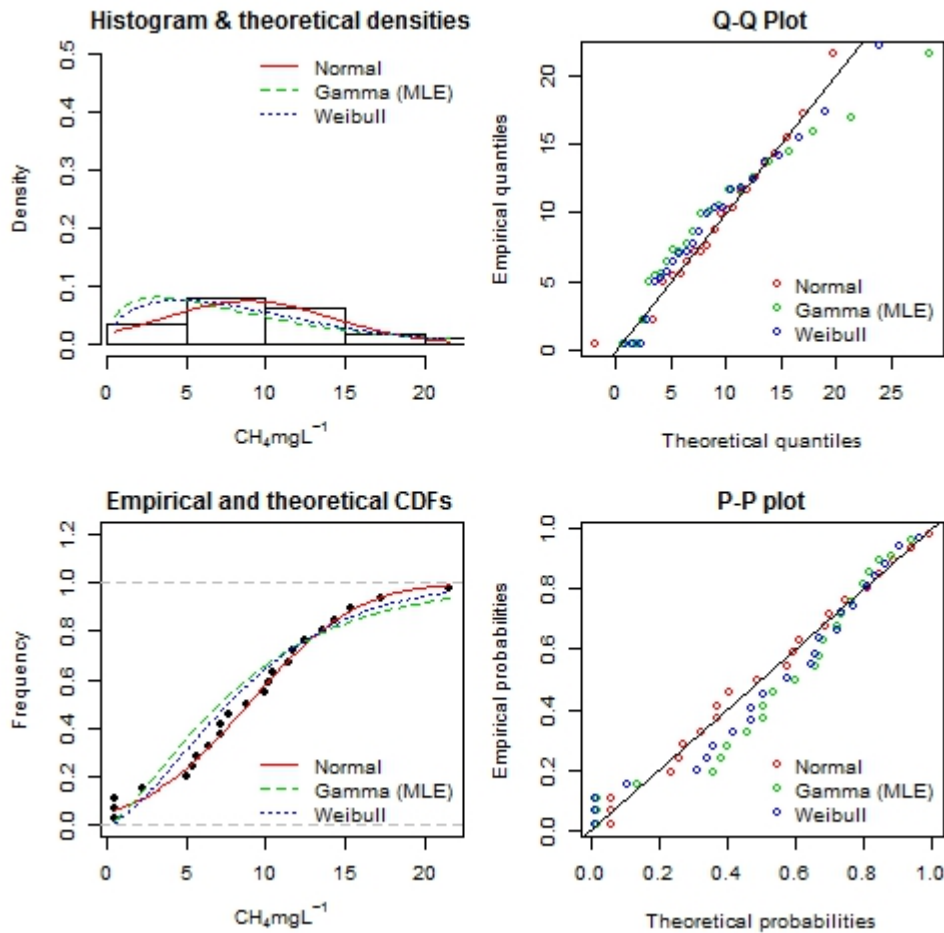
Figure 5.7   Visualisations associated with different theoretical distributions associated with the empirical measurements of dissolved methane at Site 1, Case Study 4

Table 5.2   Goodness-of-fit statistics, Site 1, Case Study 4

	Normal	LogNormal	Gamma (MME)	Gamma (MLE)	Weibull
Kolmogorov–Smirnov statistic	0.072	0.248	0.129	0.183	0.137
Cramér–von Mises statistic	0.018	0.326	0.077	0.169	0.097

---

[9] Readers should not concern themselves with the definition of these terms – they are simply used to represent a range of appropriate distributions for testing/to demonstrate proof of the method.

	Normal	LogNormal	Gamma (MME)	Gamma (MLE)	Weibull
Anderson–Darling statistic	0.180	2.019	1.681	1.105	0.855

Table 5.3   Goodness-of-fit criteria, Site 1, Case Study 4

	Normal	LogNormal	Gamma (MME)	Gamma (MLE)	Weibull
Akaike's Information Criterion	146.513	157.171	154.152	148.404	146.376
Bayesian Information Criterion	148.784	159.442	156.423	150.675	148.647

The statistics indicate that application of the theoretical normal distribution has the lowest probability of non-rejection of difference from the hypothesised distribution, using all 3 of the methods listed. At the 95% confidence level, for example, a difference from the hypothesis that the data followed a normal distribution using the Cramér–von Mises statistic would not be rejected, but the hypothesis that the data followed a gamma distribution would be (Table 5.2). Using the goodness of fit criterion, the minimum deviation was achieved through application of the normal and Weibull distributions (Table 5.3). However, the statistics and visual fits indicate that the normal distribution is the most acceptable for Site 1 and the decision would be taken to proceed on that basis. This is the same result as that from the generic assessments made in the case study.

Using the sample mean and standard deviation, 100 time series of 21 observations from a normal distribution were generated in R. This could also be performed in Excel using the NormInv and Rand commands. A further 100 time series were then generated using the sample mean and standard deviation from only the first 12 observations to represent a scenario in which the data had been collected less frequently/over a shorter duration. With no seasonality, this scenario could be assumed to represent a single year of sampling with monthly measurements.

Figure 5.8 shows a box plot of the simulated results using this new scenario. Here the blue points represent the observed concentrations, and the box and whiskers show the range of simulated data. A CUSUM test performed on these simulated datasets (see Note 3.4) resulted in no instances of change being detected using only 12 observations. In simulations from the full 21 observations, only one simulation resulted in a change. It is reassuring that both give the same result in terms of accuracy.

Visual interpretation of Figure 5.8 provides an explanation of the result and the context for the implications of the finding. The first half of the dataset showed less variability than the second half, with the lowest and highest values measured in the latter half of the monitoring period. This resulted in measurement 17 representing a higher concentration than any data point simulated (at any time reference) in the reduced data scenario. The lower variability observed in the first 12 measurements gave the dataset false precision; had the site gone operational with a baseline of this shorter length,

there would be a greater likelihood of falsely detecting change in the operational data.

This analysis provides an example of the importance of having sufficient data. Confidence in the baseline was overemphasised, with a change more likely to be falsely detected in the operational stage.
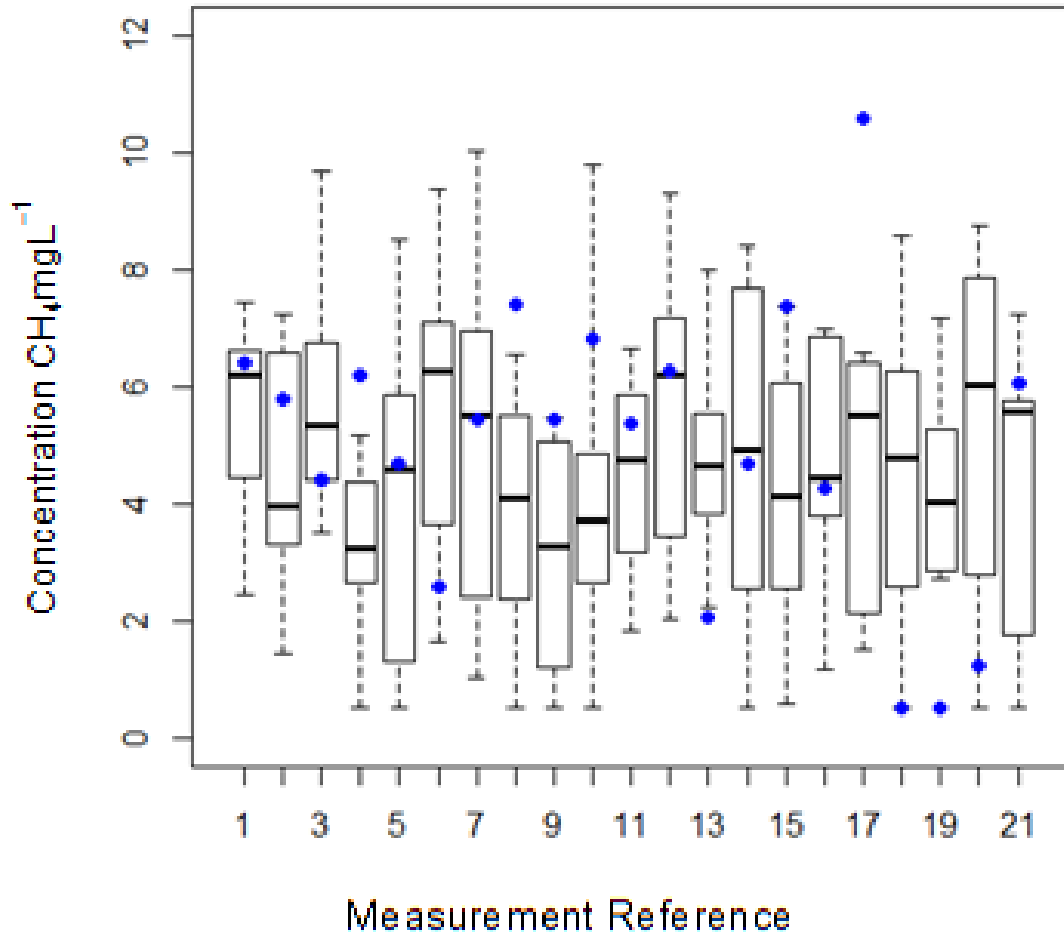
Figure 5.8   Box plot of simulated measurements using observations 1 to 12, Site 1, Case Study 4

Note: Blue points represent all measured values.

5.3   Statistical analysis to determine change/cause to trigger further data investigation

Once a baseline level is established, monitoring continues into the operational stage. Then the task for statistical analysis is to determine:

- whether or not a change from baseline levels has occurred i.e. change detection (this section)

- whether that change is due to OOG operations i.e. change attribution (Section 6)

The approaches to change detection for air quality and groundwater quality are shown in Diagrams 3-AQ and 3-GW respectively (Figure 5.9 and Figure 5.10). The first few steps are associated with checking the data quality and are identical to those employed in the baseline analysis. As most statistical tests have an underlying assumption of normality, the decision trees guide users through asking this question and send them on one of two alternative routes of assessment – one for data that satisfy this assumption and one that does not.

Note 3.1: Test for normality

There are several formal tests for assessing whether data satisfy the assumption that their distribution is normal (that is, testing for 'normality'). There are subtle differences between these tests according to whether the data can be assumed to be representative of an overall population (>50 observations) or are from a sample dataset (<50 observations), as well as the suitability for different levels of skew. The tests assess whether there is statistical evidence to suggest the assumption of normality can be rejected. They do not test whether a normal distribution can be accepted or is a better descriptor than other distributions of the underlying data. There is concern that if a decision on normality is based on test statistics only, then normality may be rejected falsely (type I error). It is therefore recommended that the primary decision over whether a distribution is normal is based on a visual assessment using some standard plots. These should be supported with a consensus from the results of the hypothesis tests using a combination of statistical assessments.

Visual assessment

Plotting a histogram of the data is recommended as a first step. This is an easy and accessible way of visualising the data; such plots are common across statistical and non-statistical (for example, Excel) software packages. Normal (or Gaussian) distributions have 2 main parameters – location (mean) and scale (standard deviation). On a histogram, a normal distribution appears as a bell-shaped curve, with the highest frequency (height of the bell) centred around the mean, and the spread (affected by the standard deviation) equal on either side. Different combinations of mean and standard deviation can lead to thin and tall, or fat and wide, normal distributions.

The second recommended step is a quantile–quantile plot (Q-QPlot). A normal Q-Q Plot reshapes the data to show what proportion of the dataset fits the expected theoretical proportion of the normal distribution model, based on the sample's mean and standard deviation. A normal distribution is thicker around the mean and thinner as you move away from the mean; consequently a normally-distribution should appear as a straight line on a Q-Q plot, with ~68% of the points 1 standard deviation away from the mean, 95.4% of the points 2 standard deviations from the mean and 99.7% of the points three 3 standard deviations from the mean. Q-Q plots should be available in most statistical software packages. Q-Q plots in Excel are not straightforward single line commands, but can be performed using the RANK, COUNT and NORM.S.INV commands.
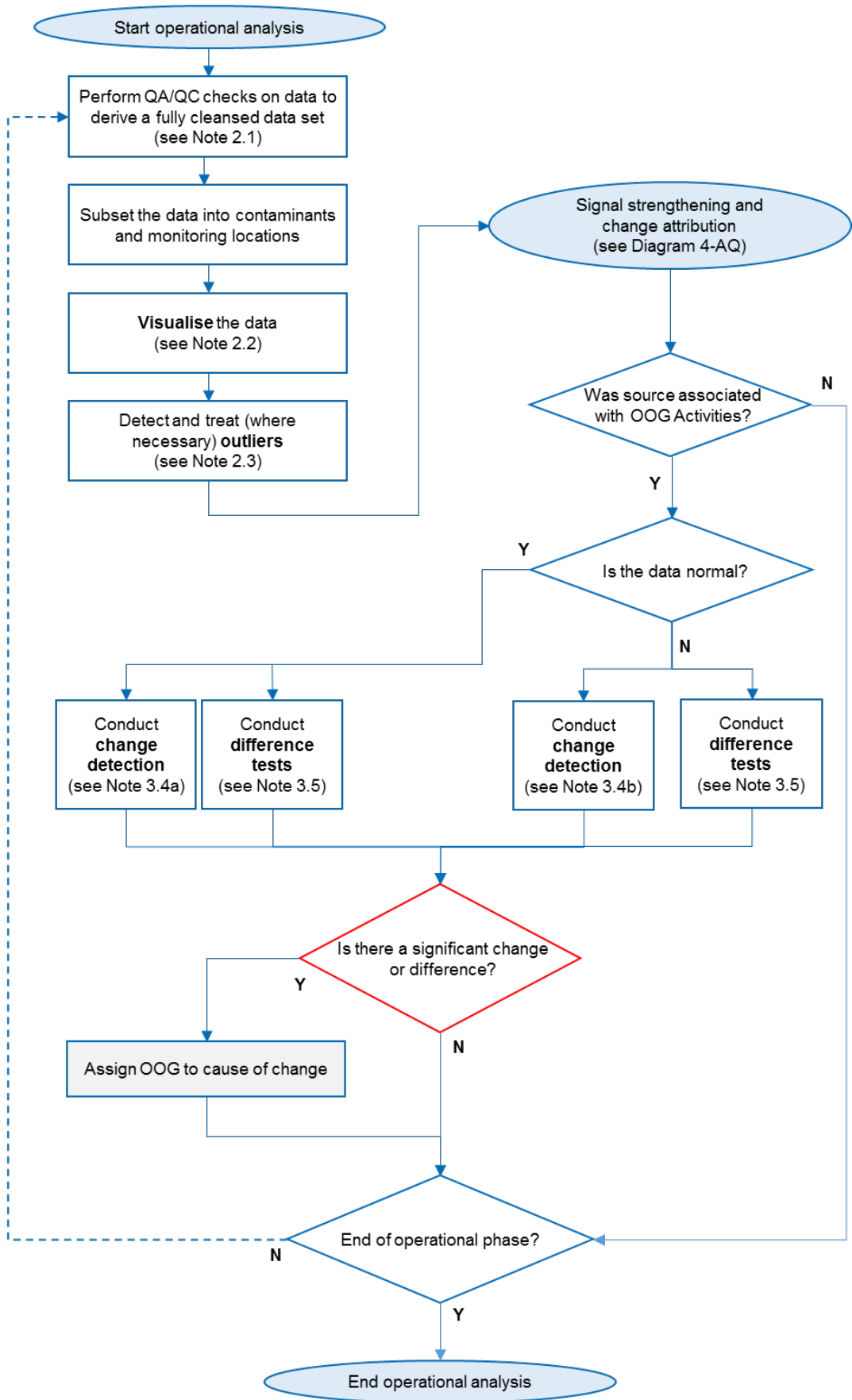
Figure 5.9   Diagram 3-AQ: Decision Tree for statistical analysis of data for determining change in air quality for OOG activities
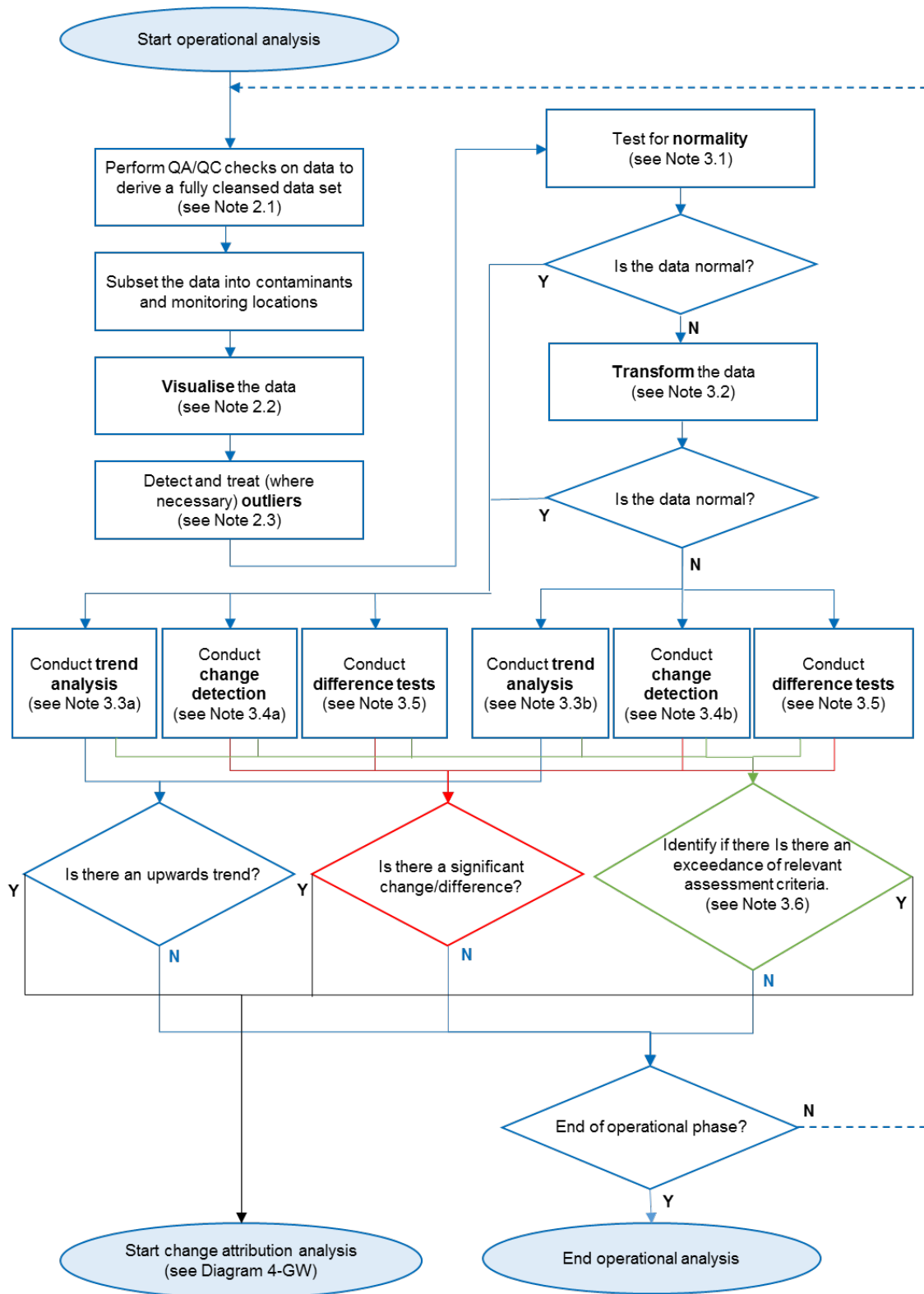
Figure 5.10 Diagram 3-GW: Decision tree for statistical analysis of data for determining change in groundwater quality for OOG activities

Statistical assessment

Four statistical tests are widely used for checking normality:

- Kolmogorov–Smirnov/Lilliefors test (Smirnov 1936, Kolmogorov 1956)

- Anderson–Darling test (Anderson and Darling 1952)

- Cramér–von Mises test (Anderson 1962)

- Shapiro–Wilk test (Shapiro and Wilk 1965)

These tests are well known for their simplicity and availability in most statistical software including SAS, PASW (formerly SPSS), STATA, Minitab and R). Other tests are available but are not as commonly employed.

It is recommended that users obtain access to statistical software. Where this is not available, the tests can be employed in software tools such as Excel where the appropriate formulae are applied.

Sample size is an important factor that can influence the outcome of these statistical tests. For example, the Shapiro–Wilk test requires the sample size to be between 3 and 50 (Shapiro and Wilk 1965). This test has since been adapted to allow safe application to sample sizes of 50–5,000 (D'Agostino 1971) and has been updated as a modification in some statistical languages (for example, R).[10] The Anderson–Darling test is the recommended empirical distribution function (EDF) test by Stephens (1986). Compared with the Cramér–von Mises test (as second choice), it gives more weight to the tails of the distribution. Other authors prefer the Shapiro–Wilk test as it is the most sensitive normality test in rejecting the null hypothesis of normality at the smallest sample sizes, at all levels of skewness and kurtosis[11] (Ahad et al. 2011). When only one normality test is implemented, use of the Shapiro–Wilk test is recommended to test the normality of data.

In the general case, performing a combination of an Anderson–Darling test, Shapiro–Wilks test and Cramér–von Mises test is recommended. The null hypothesis to which the statistical tests for normality pertain is defined as follows:

- null hypothesis (H_0) – the distribution is normal

- alternative hypothesis (H_1) – the distribution is not normal

In hypothesis testing, a statistical test is applied to determine if the null hypothesis can be rejected (that is, if it is possible to reject that the distribution is normal). The general method to do this would be to assess whether the calculated *p* value is less than the critical threshold (typically 0.05, which is equivalent to a 95% confidence level).

Each statistical test will create a test statistic for comparison against the associated distribution and most software languages will create an associated *p* value. Assuming that the null hypothesis is true (that is, that the data are normally distributed), the *p* value tells us the probability that the data seen in the sample are purely random chance. Hence, for a *p* value with a significance of 0.05, there is only a 5% probability of seeing the same kind of data from this

---

[10] Users should check the implementation of this assumption in their software language of choice.

[11] Kurtosis is a measure of the 'peakedness' of a distribution.

process, assuming that the process in question does represent normally distributed data. With such a low chance of normality, the original hypothesis that the data do come from a normal distribution would be rejected. Most statistical studies set the 'significance' level at which the default hypothesis (that this data comes from a normal distribution) is rejected at p values of 0.05 (5%) or less.

Examples of the application of visual assessment and corroboration of the results through the use of statistical tests are shown in the case studies in Section A.3 of the Annex. Example applications of visualisations and hypothesis testing are considered above.

Note 3.2: Transforming the data

Data from real-time observations are often skewed, meaning that assumptions of standard statistical tests of normality are often violated. If the distribution of the continuous data is non-normal, then transformation (taking for example log10 or natural log (ln) of the data) can often be used to make the data normal and thus increase the validity of the statistical analyses (for example, regression analysis). Testing if the application of a logistic transformation can lead to the non-rejection of normality, where the untransformed data does not satisfy this assumption, is recommended.

Care should be taken to ensure that any transformation does not exacerbate the problem of skewness. It is also not guaranteed that transformation of the data will ensure a better approximation of the normal distribution. In this case, the user will be guided through a suitable non-parametric (distribution free) analysis, corresponding to the branches of the decision tree in Fig. 5.4 that do not assume a normal distribution.

Note 3.3: Trend analysis

The Water Framework Directive and its daughter Groundwater Directive require all EU Member States to:

- monitor and assess the quality and quantity of European waters on the basis of common criteria
- identify and reverse trends in groundwater pollution

Trend assessment by OOG operators is considered an important component in the protection of groundwaters.

Before undertaking trend analysis, it is important to recognise that any data from a series of real-time observations may have the following components:

- trend component
- seasonal component
- cyclical component
- irregular component

In the example shown in Figure 5.11, the observed data contain an underlying upwards trend, annual seasonal variability and background variability in the measurements (irregular component).
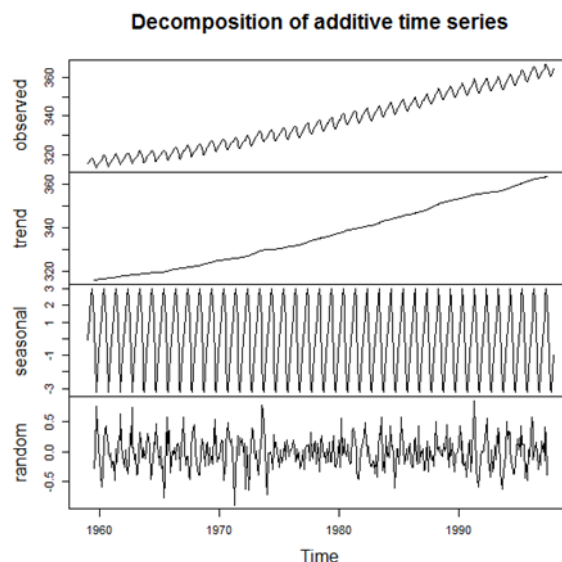
Decomposition of additive time series

Figure 5.11 Example component breakdown of a time series

A common example of where a cyclical component may also exist would be the approximately 2–7 year return period cycle of the El Niño Southern Oscillation in climate science or the 11-year cyclic effects of the Sunspot Index in the oceanic sciences. To date, no cyclic components have been identified that are relevant to baseline surveys for OOG activity. However, there may be cyclical components in baseline survey data due to recurring variations in the emissions and impacts of off-site (non-OOG) activities e.g. diurnal and weekly variations in nearby road traffic. It is important to note the difference between longer term cycles (~2–7 year return period) and more frequent cycles (for example, diurnal, weekly and seasonal); the latter are discussed in Section 2.2 and Note 4.8-AQ.

It is also important to ensure any observed trend cannot be attributed to seasonal and/or cyclical components; the ideal length of time series for continuous observations should therefore be >10 years (European Commission 2009). If the time series is shorter than this, seasonal and cyclical influences should be considered when drawing final conclusions. For OOG activity, it is anticipated that time series analysis will begin within this time frame (i.e. when series are <10 years long) and so care should be applied in interpretation.

In the majority of cases, in accordance with the Water Framework Directive guidelines on trend analysis (European Commission 2009), the appropriate test will be a linear regression. This assumes that the assumption of normality is satisfied.

This document does not give full details of trend assessment, as detailed guidance on recommended methods already exists (European Commission 2001, 2009). Instead some observations are made on points of difference/clarification for the specific circumstances of OOG operations.

One observation concerns the treatment of values below the limits of detection or quantification. Note 2.1 recommends that, for change detection, the test

should be performed without the inclusion of these values, or ideally the test should be repeated with and without them to assess any possible implications ('statistical influence') of the assumption. For trend detection, it is recommended that values are included in the analysis using the standard assumption of half of the limit of quantification as detailed in Smedley and Brewerton (1997).

Note 3.3a: Trend analysis where the assumption of normality is satisfied (parametric tests)

All statistical software packages (and some non-statistical software such as Excel) allow easy access to regression testing. See the instructions for trend determination in European Commission (2009) for details of how to perform this analysis. As well as assessing the trend during different lifecycle stages, the confidence with which that trend can be assessed should also be calculated.

Note 3.3b: Trend analysis where the assumption of normality is not satisfied (non-parametric tests)

The guidance on groundwater status and trend assessment provided in European Commission (2009) concluded that the generalised linear regression test[12] based on the LOESS (LOcal regrESSion) smoother should be the recommended method for assessing statistically significant (monotonic) trends at groundwater body level. This conclusion was based on the finding that, for extensibility and the power to detect trend, the linear methods (based on a linear model) outperformed non-parametric methods (Mann–Kendall test). Linear methods can also be considered to be more accessible in terms of software implementation and ease of use. In general, where violations of underlying assumptions are made, results are relatively insensitive to change. For completeness it is recommended that the adoption of non-parametric methods should be considered; a linear approach can be defaulted to at this point.

There is no explicit testing for seasonality within the decision tree (Diagram 3-GW, Figure 5.10). However, should a non-parametric test be chosen, the effects of seasonality should be taken into account in method selection. It is recommended that seasonality is assessed through:

1. a visual assessment using year-on-year plots

2. the fitting of a sinusoidal model to the data ('seasonal harmonic analysis') where required

An example year-on-year plot for colour measurements (surface water) is shown in Figure 5.12 to illustrate the utility of this visualisation in determining seasonality.

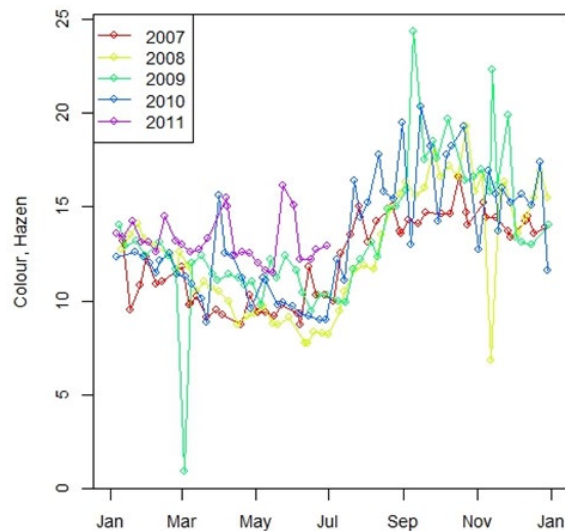[12] Analysis of variance (ANOVA) test

Figure 5.12 Demonstration of year-on-year plot for visual assessment of seasonality

It is recommended that either a Mann–Kendall test or a Seasonal Mann–Kendall test (as appropriate) is used for non-parametric tests of trend. The Seasonal Kendall test, which is a variation on the Mann–Kendall test, can be used to test for a monotonic trend of the variable of interest when the data collected over time are expected to change in the same direction (up or down) for one or more seasons (for example, months). A monotonic upward (or downward) trend means that the variable consistently increases (decreases) over time, but the trend may or may not be linear. The presence of seasonality implies that the data have different distributions for different seasons (for example, months) of the year.

These tests are available in a variety of software languages. Care should be taken in the application of irregular time series in applying these techniques. Averaging across months or seasons is recommended where this is the case, because although there are modified extensions to the Seasonal Mann–Kendall test, they can be statistically advanced. Further details of the application of the various tests can be found in the guidance on groundwater status and trend assessment (European Commission 2009).

Note 3.4: Change detection

Formal methods of change detection originate from the industrial area of 'quality control', whereby samples taken at regular intervals from the output of an industrial production process are analysed with the aim that the output meets a required standard or specification. There is a parallel with quality control in an industrial setting to groundwater quality and air quality protection in OOG operations in that:

> 'samples can be taken at regular intervals from the surrounds of an OOG production site and analysed with the aim that the output meets a target standard'.

This standard could be a fixed air quality or water quality goal (for example, to ensure compliance with an environmental permit). However, a better metric

might be to ensure no change from current conditions, with that value set as the baseline.

Two common forms of change detection (or change point detection) are the adoption of the control (Shewhart) chart and the cumulative sum (CUSUM) chart.

The basis of the Shewhart chart is the simple plot of successive sample means, with often a simultaneous parallel chart of some measure of spread. The chart's purpose is to provide a quick visual indication in any trends (for example, drift in the mean) or an increase in the mean without the need for advanced statistics. Due to the limited nature of statistical inference, however, use of this approach is not proposed for OOG operations.

Note 3.4a: Change detection where the assumption of normality is satisfied (parametric tests)

CUSUM is a method of change detection that is widely applied and accessible in a range of statistical tools. In Excel, the 'sum of the differences from target mean' can be calculated and the appropriate test statistic applied. CUSUM is, however, only one tool and only shows where a difference in the mean has occurred. A multi-faceted approach is therefore proposed to assessing whether change has taken place. Such an approach also considers changes in the variance – which can often be more sensitive to detecting true change than the mean.

A large number of modifications have been developed to hone formal change detection techniques to account for multiple change points, calculation with prescribed confidence, and to consider not just mean and variance changes but also trend changes, and changes in stage and amplitude. It is not possible to consider all of these here, with each software language having its own deviations on method and extensions to method. To indicate the difference in power of these approaches, however, the effect of using the CUSUM test statistic versus implementing the Pruned Exact Linear Time (PELT) algorithm is demonstrated. Both tests were run on the same data, and both assess change in the mean, plotted using the R 'changepoint' package (Figure 5.13). Although both find the same first 'changepoint',[13] the PELT method finds a second transition.

---

[13] A changepoint is a point in time at which a change occurs in one or multiple things.
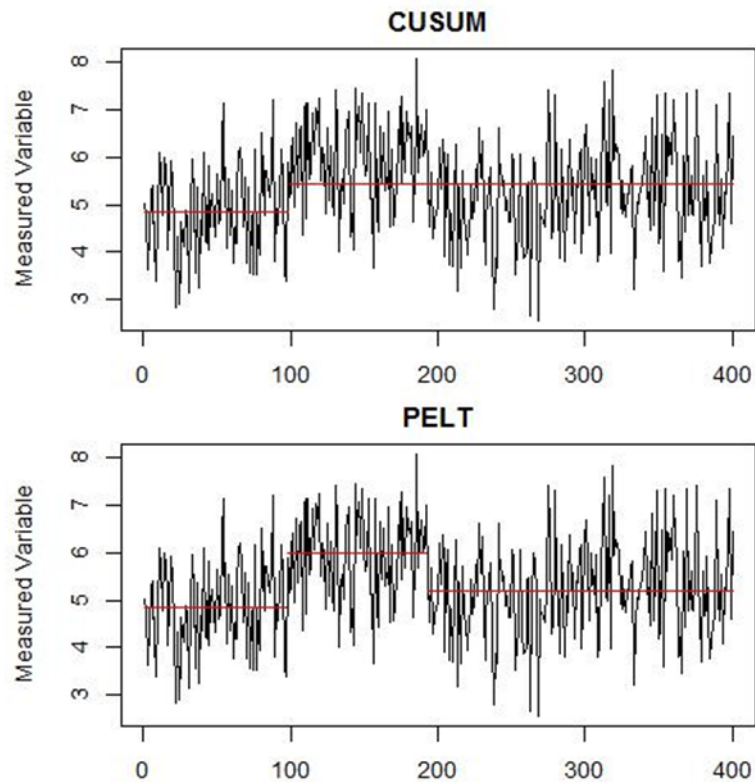
Figure 5.13 Application of CUSUM and the PELT algorithm to change point detection

In R, the changepoint package implements various mainstream (including CUSUM) and specialised changepoint methods for finding single and multiple changepoints within data. A changepoint method tries to identify if and when a change took place by looking at different components. Useful functions in R include cpt.mean(), cpt.var() and cpt.meanvar(). Cpt.mean is concerned with trying to identify if and when a change in the mean took place; cpt.var is similarly concerned with the variance, and cpt.meanvar, the combination of the mean and the variance. Examples of the application of these methods in determining change are included in the case studies in Section A.3 of the Annex.

This document does not assess confidence in the result of change detection. There are different ways of assessing confidence, one of which is to adopt a similar approach to that shown in Note 2.4 on adequacy testing, whereby multiple simulations are performed using the appropriate distribution (in this case the normal distribution).

Note 3.4b: Change detection where the assumption of normality is not satisfied (non-parametric tests)

A number of tests have been developed for use in the non-parametric case. As with trend analysis, the violation of certain statistical principles may lead the user to select a non-parametric test as best practice. However, since these can be more advanced, a simpler test (such as CUSUM) may be more appropriate.

The most widely available non-parametric test for change detection is Pettitt's test, which is a ranked method on means and is the counterpart to CUSUM.

The singular spectrum analysis method, available as an R package, is less established as a method but would appear highly suitable in that is a non-parametric method targeting not just the mean, but also other descriptors of variation. The technique decomposes a time series into the sum of trend, periodics and noise. It can be used to detect changes in mean, variance of noise, amplitude of periodics, frequency of periodics and coefficients of linear recurrent formulae.

Note 3.5: Change difference

Tests to detect a systematic change in mean value can be assessed between different locations and time periods, for example, between upgradient and downgradient boreholes, or upwind and downwind monitoring stations or baseline and operational fixed time point measurements (where appropriately time matched). In general, this provides a way of quantifying whether an observed difference in mean values is likely to occur by chance in the same time and space.

The main use of this method is in detecting persistent differences in mean quality through time between a specific combination of groups of boreholes or a specific combination of air quality monitoring stations.

In the medium to long term following the onset of operations, a change in concentration may well manifest itself in a relatively abrupt shift in the mean after a period of stability, rather than a steady drift in concentration. The former of these 2 types of change is most likely a candidate for formal change detection techniques as described in Note 3.4. For the latter, a more standard 'change difference' approach is likely to be appropriate.

Note 3.6: Groundwater assessment criteria

The regulatory background and groundwater quality assessment criteria for groundwater are described in detail in the literature review in Section A.1 of the Annex. The two most important objectives for groundwater quality under the Water Framework Directive can be summarised as follows:

- To prevent or limit the input of pollutants into groundwater and to prevent the deterioration of the status of groundwater

- To reverse any significant and sustained upward trend in the concentration of any pollutant

Under the Water Framework Directive, 'pollution' is defined as the direct or indirect introduction of substances into land or water as a result of human activity that may cause harm to human health, aquatic ecosystems or terrestrial ecosystems dependent on the aquatic ecosystem. Therefore there is a reliance on receptor-based standards for the assessment of compliance with groundwater quality objectives.

Guidance on determining hazardous and non-hazardous substances, and appropriate standards and thresholds, and how these should be used to determine risk to groundwater receptors is given in UKTAG (2013), Environment Agency (2017) and JAGDAG (2017).

5.4 Updating the monitoring design following statistical analysis

There are several mechanisms by which the monitoring design can be improved using the outputs from the statistical analysis.

The first mechanism is to question whether the data are adequate (Guideline Note 2.4). Where the data are found not to be adequate, then the frequency of sampling could be increased or the duration for baseline establishment could be extended from that originally proposed in order to increase the likelihood of detecting true change. Should the monitoring campaign not be amended, then the statistical considerations would still be updated as a result of the adequacy finding; this could lead to the operator to decide to monitor more frequently during the operational stage.

The second mechanism is to implement an adaptive monitoring approach (see Section 3). Here the monitoring campaign would be updated during the operational stage using the outputs of the analysis. Examples include:

- scaling down the monitoring of correlated variables

- increased monitoring of contaminants that had breached an assessment limit, or had shown a change or trend that could be attributed to OOG activities

6 Stage 5: Signal strengthening and change attribution

6.1 Air quality

Diagram 4-AQ (Figure 6.1) outlines the flow for signal strengthening and change attribution in air quality for OOG activities. It is is supported by Notes 4.1-AQ to 4.8-AQ.

Note 4.1-AQ: Collate explanatory variable data

The conceptual model should be used to identify additional sources and pathways that may explain the observations at the site. Several strands of explanatory data can be used to inform the change attribution stage. Examples include:

- non-target source information

- local meteorological data (wind direction, speed and temperature)

Comparison of satellite images and photo-panorama around the monitoring site with baseline imagery could be used to identify any changes in land use (for example, new agricultural buildings, roads and housing).

The operator should also consider if there are other activities in the area that may be emitting similar contaminants that are not included in the conceptual model. The conceptual model should be updated accordingly, accessing the data where possible.

Note 4.2-AQ: Conduct wind sector analysis

Wind sector analysis considers where air masses have come from (for example, wind sectors north, north-east and east) and calculates the measured concentration recorded during the time that the wind direction and speed were within each relevant sector. For surveys conducted with longer averaging times than the meteorological datasets, it is possible to allocate a particular trajectory to a sector based on assumptions about the proportion of time it is in that sector, for example, assuming a measurement corresponds to a westerly wind direction ($67.5°–112.5°$) if it spends at least 50% of its time in that sector.

Air quality observations can be time matched and assigned to the wind speed and direction, and polar plots used to visualise the resulting wind and pollution roses. This analysis can be performed using the purpose-built package OpenAir in R. Examples of this analysis are shown in Case Study 2 in Section A.3 in the Annex.

Note 4.3-AQ: Perform conditional analysis to increase signal strength

As a first step, it is necessary to identify appropriate rulesets. These should be informed by visualisations of the concentration data, other explanatory data and the outputs of the wind sector analysis as well as from expert judgement. Appropriate rulesets should be derived that represent strong and frequent impacts from target source(s) and filtering applied accordingly. Examples of this analysis are shown in Case Study 2 in Section A.3 in the Annex.
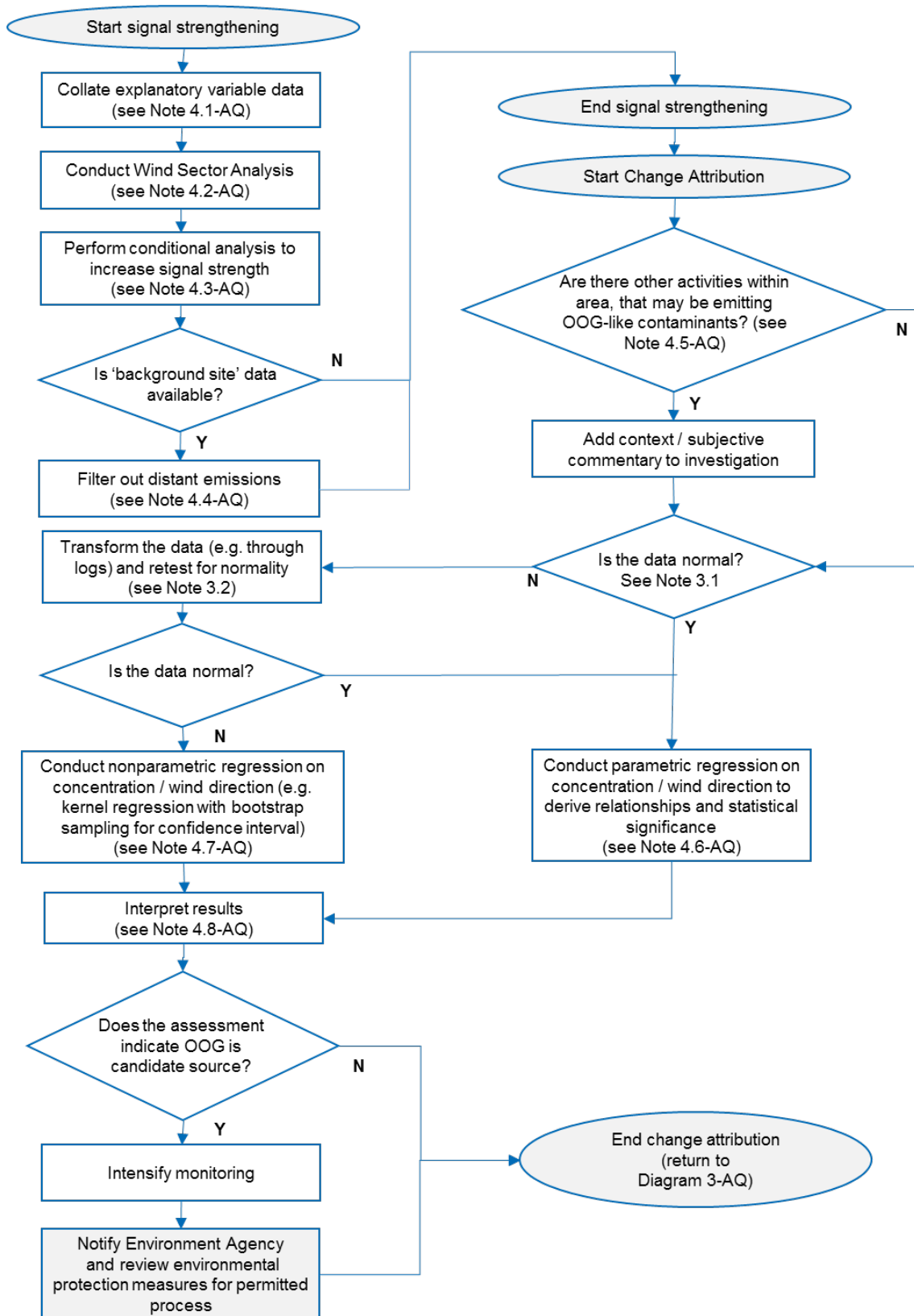
Figure 6.1   Diagram 4-AQ: Decision tree for signal strengthening and change attribution in air quality for OOG activities

Note 4.4-AQ: Filter out distant emissions

To strengthen the source of the OOG activity signal, filtering out any distant emissions responsible for background concentrations is recommended. This can be done through paired observations of time averaged data for upwind and downwind monitoring points, where for example a daily averaged value during a period of consistent wind direction would represent average concentrations from the associated wind direction. It is not possible to perform this filtering process where concentrations cannot be attributed to a wind direction with confidence (for example, where there is no onsite, close proximity weather station).

Modified values for a wind direction (D) specific time averaged reference period t (for example, daily average for N winds on 1 January 2001) at the downwind monitoring location (AQdown) are evaluated from the upgradient measurements at station AQup and the difference from the 'typical' or 'average' value at that location, AQup(typ)[D], according to Equation 6.1:

$$\widehat{AQ_{down}}[t, D] = AQ_{down}[t, D] - AQ_{up}[t, D] \quad \text{(Equation 6.1)}$$

Note 4.5-AQ: Other activities

To attribute observed variability to other sources, it is necessary to identify which of the monitored variables may have originated from non-OOG activities. Examples of potentially contributing sources of VOCs, NOx, $PM_{2.5}$, carbon monoxide and methane include:

- road traffic
- agriculture
- domestic combustion
- construction activity
- industrial processes

This step involves interpreting the data collated in in the first step (Note 4.1-AQ). It is anticipated that this activity will largely be informed by expert judgement, the literature and best practice.

Note 4.6-AQ: Application of parametric techniques for change attribution (linear regression)

See Note 4.4-GW.

Note 4.7-AQ: Application of non-parametric techniques for change attribution (PCA/kernel estimator methods)

Principal component analysis (PCA) is a technique used to make multivariate data easier to understand. PCA can be used to display visually as much as possible of the total variation in the data in a few dimensions (for example, in the creation of biplots). It works by a linear decomposition of a set of correlated

variables, reducing the number of variables and removing correlation between them.

PCA projects observations from a p-dimensional space with p variables to a k-dimensional space (where k < p). PCA dimensions are also called axes or factors. If the information associated with the first 2 or 3 axes represents a sufficient percentage of the total variability of the scatter plot, the observations could be represented on a two-dimensional or three-dimensional chart, thus making interpretation much easier.

General rulesets exist that can be helpful in determining the number of principal components to retain in explaining the total variance of the dataset, which are often known as 'stopping rules or criteria'. There are a number of objective rules for 'stopping' based around either confidence intervals or the average values of statistical tests, but it is not necessary to understand these for the purposes of attribution.

An example of how PCA may be applied is in evaluating:

- how much of the variance in an observed contaminant can be explained by the presence of other contaminants at background sites and downwind of the facility

- how these differ/align to the expected signature from plant operations

PCA is widely available in different statistical software packages (for example, Statistica). As a popular and widely used technique, add-ons have also been developed in Excel for its use (for example, XLSTAT-Base or Multibase). Using commands such as STANDARDIZE, MMULT and COV, it is also possible to calculate the eigenvalues associated with the variables and their principal components. An example of the use of PCA is given in Case Study 3 in Section A.3 of the Annex.

Note 4.8-AQ: Interpret attribution test results

This step has been added as a separate component to that of the statistical analysis for attribution, as it may require expert knowledge and judgement that incorporates supplementary explanatory factors (qualitative and quantitative). This is separate to the statistical analysis up to this point, as it contextualises the outputs in terms of the site location, activities and air quality ambient levels/dispersion.

This part of the process will require interpretation of measured data in the light of experience of the factors that affect measured levels of airborne pollutants. Relevant aspects may include:

- consideration of pollutant ratios to identify contributing sources

- consideration of diurnal, weekly or seasonal variability

- evaluation in the light of modelled datasets such as site-specific modelling studies or Defra gridded air quality datasets

6.2   Groundwater

Diagram 4-GW (Figure 6.2) outlines the flow for signal strengthening and change attribution in groundwater quality for OOG activities. It is supported by Notes 4.1-GW to 4.6-GW.

Note 4.1-GW: Collate explanatory variable data

The conceptual model should be used to identify additional sources and pathways for change attribution analysis. As a minimum, these data should be the flow data from site, but could also include details of background measurements from non-OOG monitoring of the aquifer, or information about other natural or anthropogenic contributions (qualitative or otherwise). The operator should also consider if there are other activities within the area that may be emitting similar contaminants that are not included in the conceptual model. The conceptual model should be updated accordingly and the data accessed where possible.

Where it is possible to identify the cause of a specific element of variability in the dataset, then it should in theory be possible to subtract this from the overall measure of variability observed for a particular determinand or monitoring point. In this way the total level of variability exhibited for the location or determinand can be substantially reduced, making it potentially possible to obtain an earlier warning of developing trends (Environment Agency 2002). However, this may not be straightforward in practice, given the complications of irregular and infrequent time series coupled with the complexities of spatiotemporal differences in borehole measurements. Care should therefore be applied in the interpretation.
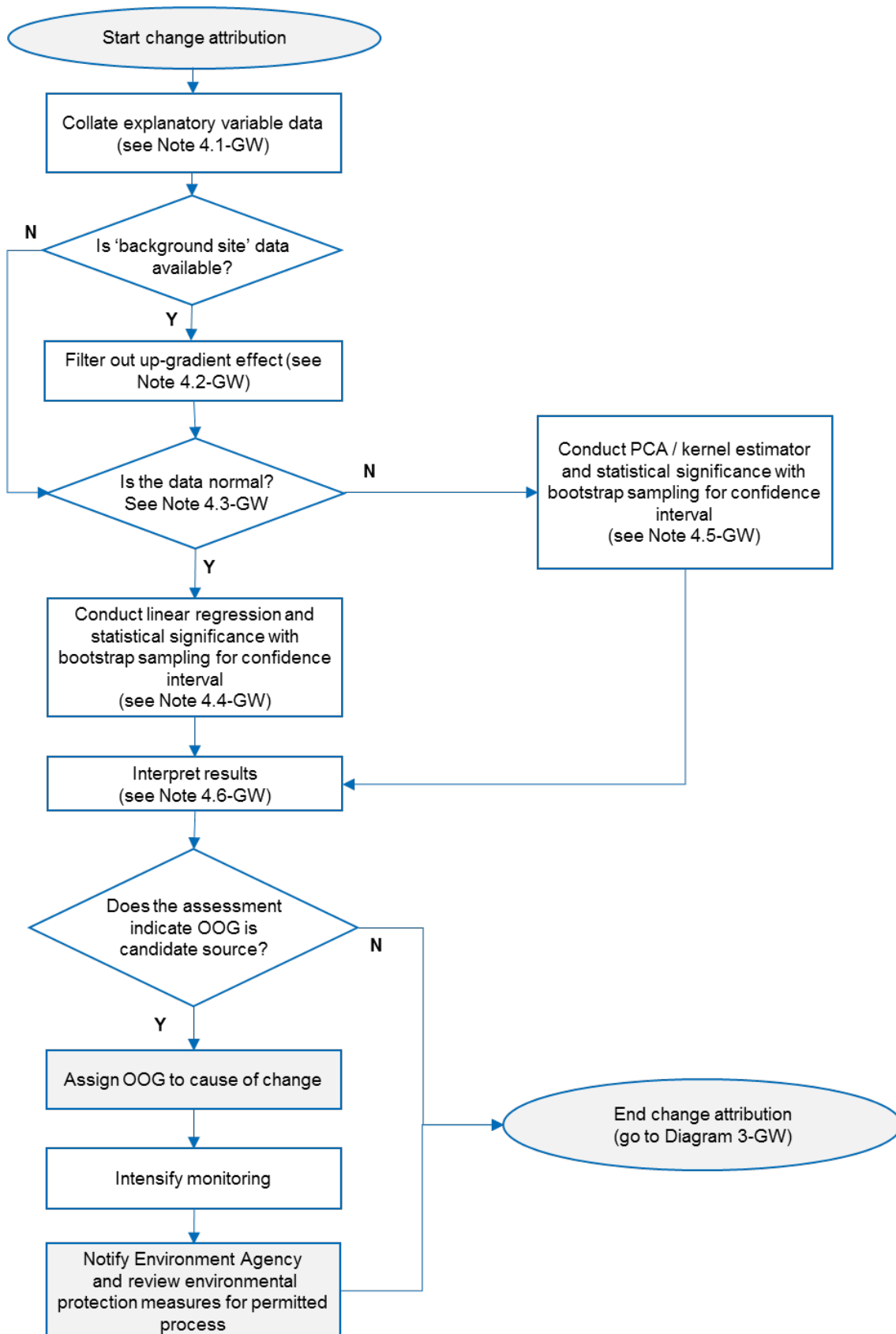
Figure 6.2   Diagram 4-GW: Decision Tree for statistical investigation of groundwater quality for OOG activities

Note 4.2-GW: Filter out upgradient variability

To strengthen the signal of any effects of OOG operations, environmental and other anthropogenic inputs detectable in the aquifer upgradient of the OOG facility should be filtered out from the measured data as a first step. This may be of use in separating out the effects of fluctuating background and/or seasonal conditions in addition to other external activities affecting the monitoring data from those that can be attributed to the OOG operation.

However, this process may not be as simple as it may first appear. The guidance provided for the interpretation of landfill monitoring data similarly recommends using a filtering approach (Environment Agency 2002). It promotes a time-paired value difference approach in which 'corrected' values for a downgradient borehole (Bdown) are evaluated from the upgradient measurements at borehole (Bup) and the difference from the 'typical' or 'average' value at that location, Bup(typ), according to Equation 6.2:

$$\widehat{B_{down}}[t] = B_{down}[t] + \left(B_{up}(\text{typ}) - B_{up}[t]\right) \quad \text{Equation 6.2}$$

It can be assumed that an appropriate value for Bup(typ) is the annualised average at that site such that Bup(typ) = (B_up). However, it is also assumed that:

- the samples are taken at regular time points and be matched at both sites

- no lag effect is applied to represent travel time for contaminants between upgradient and downgradient locations

It may therefore be appropriate to use time averaged values (for example, monthly or seasonally) and to include some requirement to account for travel time effects where distances are 'large' and residence times 'long'.

Furthermore, as there may be missing observations, a ruleset should be developed for how to deal with this in the analysis. Further research is required in this area – with application of data – to understand how this may best be applied to measurements for OOG – especially where measurements may be infrequent and unmatched (that is, sampled at different times). No specific recommendation is therefore made on rulesets that should be applied and, on this basis, it is assumed that operators will not apply filtering at the current time.

In cases where there is more than one upgradient borehole location, this follows the guidance given for the interpretation of landfill monitoring data (Environment Agency 2002). This makes the presumption that multiple locations have been chosen because collectively they represent the best description of upgradient quality. On 'each sampling occasion', therefore, the mean concentration across all the control boreholes would be calculated in the representation of Bup(typ) and Bup in Equation 6.2.

Note 4.3-GW: Normality checks

By now, the data should have been assessed for the appropriateness of the underlying distributional assumption and should be in 1 of 3 forms:

- raw data to which the assumption of normality cannot be rejected at the appropriate significance level;

- log-transformed data for which the assumption of normality cannot be rejected at the appropriate significance level, although this assumption can be rejected for the untransformed data;

- raw data, on which the assumption of normality on both the raw data and log-transformed data was rejected in both circumstances

If the data have been further processed since the previous normality check (that is, upgradient data were available and the fingerprint of these data has been removed; see Note 4.2-GW), it would be prudent to retest the data (see Notes 3.1 and 3.2).

Note 4.4-GW: Application of parametric techniques for change attribution (linear regression)

The most frequently used and easily accessible method of change attribution is in the application of multiple linear regression methods. These attribute combinations of potential sources of contamination to explain that observed.

An ordinary least squares (OLS) regression with one regressor (independent variable) has the form:

$$Y = a_0 + a_1 x + \varepsilon \qquad \text{Equation 6.3}$$

There are both practical and theoretical reasons to use OLS estimators of a_0 and a_1. OLS is the dominant method used in practice and has become the common language for regression analysis. Where the underlying assumptions on distribution are satisfied, the regression R^2 and adjusted R^2 tell us whether the regressors are good at predicting or explaining the values of the dependent variable in the sample of the data at hand (assuming that the relationship can be defined in a linear fashion).

Adjusted R^2 is preferred to R^2, as it considers the degrees of freedom in estimating the parameters. However, R^2 or adjusted R^2 do not tell us whether:

- an included variable is statistically significant

- the regressors are a true cause of the movements in the dependent variable

- there is omitted variable bias

- the most appropriate set of regressors has been chosen

For this an ANOVA is recommended. This is not covered here as a basic statistical test; consulting any basic data analysis textbook is recommended.

While correlation techniques do not involve an implicit assumption of causality, regression techniques do and therefore tend to convey a more powerful message.

Note 4.5-GW: Application of non-parametric techniques for change attribution (PCA/kernel estimator methods)

See Note 4.7-AQ.

Note 4.6-GW: Interpret attribution test results

This step has been added as a separate component to that of the statistical analysis for attribution as it may require expert knowledge and judgement that incorporates supplementary explanatory factors (qualitative and quantitative) that are separate to the statistical analysis in isolation.

References

AHAD, N.A., YIN, T.S., OTHMAN, A.R. AND YAACOB, C.R., 2011. Sensitivity of normality tests to non-normal data. *Sains Malaysiana*, 40 (6), 637-641.

AMEC, 2014. *Technical support for assessing the need for a risk management framework for unconventional gas extraction*. Final report. Brussels: European Commission.

ANDERSON, T.W., 1962. On the distribution of the two-sample Cramér-von Mises criterion. *Annals of Mathematical Statistics*, 33 (3), 1148-1169.

ANDERSON, T.W. AND DARLING, D.A., 1952. Asymptotic theory of certain 'goodness-of-fit' criteria based on stochastic processes. *Annals of Mathematical Statistics*, 23 (2), 193-212.

API, 2009. *Hydraulic fracturing operations – well construction and integrity guidelines*. API Guidance Document HF1. Washington DC: American Petroleum Institute.

API, 2010. *Water management associated with hydraulic fracturing.* API Guidance Document HF2. Washington DC: American Petroleum Institute.

BRANTLEY, S.L., YOXTHEIMER, D., ARJMAND, S., GRIEVE, P., VIDIC, R., POLLAK, J., LLEWELLYN, G.T., ABAD, J. AND SIMON, C., 2014. Water resource impacts during unconventional shale gas development: the Pennsylvania experience. *International Journal of Coal Geology*, 126, 140-156.

COUNCIL OF CANADIAN ACADEMIES, 2014. *Environmental impacts of shale gas extraction in Canada*. Ottawa: The Expert Panel on Harnessing Science and Technology to Understand the Environmental Impacts of Shale Gas Extraction.

D'AGOSTINO, R.B., 1971. An omnibus test of normality for moderate and large size samples. Biometrika, 58 (2), 341-348.

DECC, 2013. *Onshore oil and gas exploration in the UK: Regulation and best practice. England.* London: Department of Energy and Climate Change.

DEFRA, 2016. *Local Air Quality Management. Technical Guidance (TG16).* London: Department for Environment, Food and Rural Affairs.

DEFRA AND ENVIRONMENT AGENCY, 2016. *Groundwater risk assessment for your environmental permit* [online]. London: Department for Environment, Food and Rural Affairs. Available from: https://www.gov.uk/guidance/groundwater-risk-assessment-for-your-environmental-permit [Accessed 9 January 2018].

ENVIRONMENT AGENCY, 2002. *Techniques for the interpretation of landfill monitoring data. Guidance notes.* Final R&D Technical Report P1-471. Unpublished.

ENVIRONMENT AGENCY, 2003. *LFTGN02: Guidance on monitoring of landfill leachate, groundwater and surface water.* Bristol: Environment Agency.

ENVIRONMENT AGENCY, 2014. *Hydrogeological risk assessment report template* [online]. Bristol: Environment Agency. Available from: https://www.gov.uk/government/publications/hydrogeological-risk-assessment-report-template [Accessed 9 January 2018].

ENVIRONMENT AGENCY, 2016. *Onshore oil & gas sector guidance, Version 1*. Bristol: Environment Agency. Available from: https://www.gov.uk/government/publications/onshore-oil-and-gas-exploration-and-extraction-environmental-permits [Accessed 9 January 2018].

ENVIRONMENT AGENCY, 2017. *Hazardous substances to groundwater: minimum reporting values* [online]. Bristol: Environment Agency. Available from: https://www.gov.uk/government/publications/values-for-groundwater-risk-assessments/hazardous-substances-to-groundwater-minimum-reporting-values [Accessed 9 January 2018].

EPA, 2016. *Final Report 4: Impacts and mitigation measures (part 1 of 3).* Environmental Impacts of Unconventional Gas Exploration and Extraction (UGEE) Joint Research Programme. Dublin: Environmental Protection Agency. Available from: http://www.epa.ie/pubs/reports/research/ugeejointresearchprogramme/ [Accessed 12 January 2018].

EUROPEAN COMMISSION, 2001. *Statistical aspects of the identification of groundwater pollution trends and aggregation of monitoring results – Working Group 2.8 Statistics.* Technical Report No. 1. Luxembourg: Office for Official Publications of the European Communities.

EUROPEAN COMMISSION, 2009. *Common Implementation Strategy for the Water Framework Directive (2000/60/EC). Guidance Document No. 18. Guidance on groundwater status and trend assessment*. Technical Report - 2009 – 026. Luxembourg: Office for Official Publications of the European Communities.

EUROPEAN COMMISSION, 2010. *Common Implementation Strategy for the Water Framework Directive (2000/60/EC). Guidance Document No. 26. Guidance on risk assessment and the use of conceptual models for groundwater.* Technical Report - 2010 - 042. Luxembourg: Office for Official Publications of the European Communities.

EUROPEAN COMMISSION, 2014. Commission Recommendation of 22 January 2014 on minimum principles for the exploration and production of hydrocarbons (such as shale gas) using high-volume hydraulic fracturing (2014/70/EU). *Official Journal of the European Union*, L 39 (8.2.2014), 72-78.

GROUND WATER PROTECTION COUNCIL AND ALL CONSULTING, 2009. *Modern shale gas development in the United States: a primer.* Washington DC: US Department of Energy.

HAWKINS, D., 1980. *Identification of Outliers*. London: Chapman and Hall.

HIGHWAYS AGENCY, TRANSPORT SCOTLAND, WELSH ASSEMBLY GOVERNMENT AND DEPARTMENT FOR REGIONAL DEVELOPMENT NORTHERN IRELAND, 2007. *Design Manual for Roads and Bridges. Volume 11 Environmental Assessment. Section 3 Environmental Assessment*

Techniques. Part 1 HA 207/07, Air Quality. Norwich: TSO. Available from: http://www.standardsforhighways.co.uk/ha/standards/dmrb/ [Accessed 12 January 2018].

JAGDAG, 2017. *Methodology for the determination of hazardous substances 2014-16* [online]. Edinburgh: Joint Agencies Groundwater Directive Advisory Group. Available from: http://www.wfduk.org/stakeholders/methodology-determination-hazardous-substances-2014 [Accessed 9 January 2018].

KOLMOGOROV, A., 1956. *Foundations of the Theory of Probability*, 2nd edition. New York: Chelsea Publishing Company

SHAND, P., EDMUNDS, W.M., LAWRENCE, A.R., SMEDLEY, P.L. AND BURKE, S. 2007. *The Natural (Baseline) Quality of Groundwaters in England and Wales* British Geological Survey RR/07/06 and Environment Agency NC/99/74/24

SHAPIRO, S.S. AND WILK, M.B., 1965. An analysis of variance test for normality. *Biometrika*, 52 (3/4), 591-611.

SMEDLEY, P.D. AND BREWERTON, L.J., 1997. *The Natural (Baseline) Quality of Groundwaters in England and Wales series*. R&D Project Record W6/i722. Bristol: Environment Agency.

SMIRNOV, S., 1936. Beschreibung einer neuen Acartia-Art aus dem Japanischen Meer nebst einiger Bemerkungen über die Untergattung Euacartia Steuer. *Zoologischer Anzeiger*, 114: 87-92.

STEPHENS, M.A., 1986. Tests based on EDF statistics. In *Goodness-of-Fit Techniques* (ed. R.B. D'Agostino and M.A. Stephens). New York: Marcel Dekker.

UKOOG, 2015a. *Guidelines for the establishment of environmental baselines for UK onshore oil and gas, Issue 1*. London: UK Onshore Oil and Gas.

UKOOG, 2015b. *UK onshore shale gas well guidelines: exploration and appraisal phase*, *Issue 3*. London: UK Onshore Oil and Gas. Available from: http://www.ukoog.org.uk/images/ukoog/pdfs/ShaleGasWellGuidelinesIssue3.pdf [Accessed 15 January 2018].

UKTAG, 2012. *Paper 11b(i) Groundwater chemical classification for the purposes of the Water Framework Directive and the Groundwater Directive.* Edinburgh: UK Technical Advisory Group on the Water Framework Directive. Available from: https://www.wfduk.org/tagged/groundwater [Accessed 12 January 2018].

UKTAG, 2013. *Phase 3 Review of Environmental Standards 2012/13.* Edinburgh: UK Technical Advisory Group on the Water Framework Directive. Available from: https://www.wfduk.org/resources/category/environmental-standards-202 [Accessed 12 January 2018].

USEPA, 2015. *Assessment of the potential impacts of hydraulic fracturing for oil and gas on drinking water resources (external review draft).* EPA/600/R-15/047a. Washington DC: US Environmental Protection Agency.

VENGOSH, A., JACKSON, R.B., WARNER, N., DARRAH, T.H. AND KONDASH, A., 2014. A critical review of the risks to water resources from unconventional shale gas development and hydraulic fracturing in the United States. *Environmental Science and*

76

List of abbreviations

ANOVA   analysis of variance

AURN   Automatic Urban and Rural Network

EPA   Environmental Protection Agency [Ireland]

GC-MS   gas chromatography–mass spectrometry

HGV   heavy good vehicle

NMVOCs   non-methane volatile organic compounds

NOx   nitrogen oxides

NORM   naturally occurring radioactive materials

OLS   ordinary least squares

OOG   onshore gas and oil

PAHs   polycyclic aromatic hydrocarbons

PCA   principal component analysis

PELT   Pruned Exact Linear Time [algorithm]

PM   particulate matter

QA   quality assurance

QC   quality control

UKOOG   UK Onshore Oil and Gas

USEPA   US Environmental Protection Agency

VOCs   volatile organic compounds

Annex A: Supplementary information

See separate file.

Would you like to find out more about us or your environment?

Then call us on

03708 506 506 (Monday to Friday, 8am to 6pm)

Email: enquiries@environment-agency.gov.uk

Or visit our website

www.gov.uk/environment-agency

incident hotline

0800 807060 **(24 hours)**

floodline

0345 988 1188 **(24 hours)**

Find out about call charges (https://www.gov.uk/call-charges)

Environment first

Are you viewing this onscreen? Please consider the environment and only print if absolutely necessary. If you are reading a paper copy, please don't forget to reuse and recycle.