Review of the Quality Assurance Framework of the National Atmospheric Emissions Inventory, Pollution Climate Mapping and Impact Pathway Models

Report prepared for Defra under Contract 21366

Hartley McMaster Limited
Broadway Chambers
1-5 The Broadway
St Peters Street
St Albans
Herts
AL1 3LH

Hartley McMaster Ltd
Analytical Management Consultancy

CONTENTS

EXECUTIVE SUMMARY

In March 2013 HM Treasury published a review of the quality assurance (QA) of Government analytical models. This report, known as the Macpherson Review, made a series of recommendations to extend best QA practice across the whole of government.

Defra is committed to following the recommendations of the Macpherson Review to ensure that models and their outputs are used correctly to support the department's business. As part of the rolling programme of implementing the Macpherson recommendations, Defra appointed Hartley McMaster to carry out an in-depth review and assessment of the QA policies and practices for three high-profile models that play a key role in informing policy and reporting on Air Quality:

- The National Atmospheric Emissions Inventory (NAEI).
- The Pollution Climate Mapping model (PCM).
- The Impact Pathway Methodology (IPM).

To carry out their work, the Hartley McMaster review team visited the offices of Ricardo-AEA (which as Inventory Agency leads the production of the NAEI, and is also responsible for PCM) and three contractors that develop parts of the NAEI model: the Centre for Ecology and Hydrology, Forest Research and Rothamsted Research. During these visits they interviewed staff responsible for the development and QA of the models and were shown evidence of QA materials that had been produced. IPM is owned by Defra so for this model they interviewed the member of Defra's staff who leads the team responsible for the model.

The review team found that the three models were well designed and well built by experienced professionals, who operate under very tight timescales for their delivery. The QA policies and practices adopted by the model builders were evolving during the review, and by the end of the review compared relatively well against three independent sets of best practice guidelines: the Intergovernmental Panel on Climate Change (IPCC) 2006 QA guidelines, the Department of Energy and Climate Change (DECC) QA guidelines, and the guidance within the Macpherson Review final report. Specifically, the two models classified as business critical by Defra – the NAEI and PCM – fully or partially adhered to all guidance that is considered mandatory by these three sets of guidelines; IPM, which is not classified as business critical, failed to follow some of the mandatory guidance, though the risk arising from this failure is relatively low.

These findings should give Defra confidence that the QA policies and practices used for the three models are largely fit for purpose. Despite this, modelling errors can still occur, and the team saw examples of these errors in their review. They noted that models had mostly been built by scientists rather than software specialists, so concepts such as formal specification, design, unit testing and acceptance testing were largely unfamiliar. Detailed checks had been carried out on the models, but these were nearly always scientific checks, for example the confirmation that outputs look sensible when presented as a time-series. Software checks, such as regression testing after models have been changed, were largely absent. By adding a software-style QA regime for these models, more can be done to reduce the risk of future errors. In this report the Hartley McMaster team makes the following recommendations, most of which apply across all business-critical models built or commissioned by Defra and DECC:

- **Recommendation 1: QA of the model templates must be distinct from QA of runs of the model.** This distinction between a template, which is an unpopulated database or spreadsheet, and a populated model, allows a more efficient and effective approach to testing and serves as the basis for many of the remaining recommendations.
- **Recommendation 2: All templates must be fully tested and peer reviewed**. Extensive and thorough validation of outputs was found to be performed on all models, but software-style verification testing of the empty template, using dummy data, was largely absent so must be added to ensure that errors are not being hidden by apparently sensible output data. Peer review adds a further layer of assurance that the models are operating as intended.
- **Recommendation 3: The specification of checks performed on the models must be brought to a common high standard.** All models were seen to have quality checks documented, but the checks were

sometimes vague or ambiguous. All checks must be described in a way that ensures two different checkers would perform the same actions.

- **Recommendation 4: Results of all checks must be recorded.** Although quality checks were documented for all models, there was at times no evidence that these checks had been performed. Records must be kept of checks that have been done, alongside those that should be done.
- **Recommendation 5: Checks and results must be documented in a clear and consistent way.** Across the models that were reviewed, checks and results were recorded in a variety of places, including in the QA Manual, in the User Guide and within the model itself. Checks and results must be recorded in a standard location for each model, and models must not be signed off until evidence is seen of the completion of all checks.
- **Recommendation 6: Model documentation must be improved.** Design documentation is important for continuity planning (to enable handover to an alternative contractor) and to support model testing, but was often missing or was captured within the model itself. The nature of the required design documentation must be mandated in future ITTs, in particular to address the possible conflict of interest whereby failure to produce documentation would make handover to an alternative contractor more difficult. In addition, ad-hoc requests for model outputs must be formalised in a specification to avoid misunderstandings about the nature of the data to be provided.
- **Recommendation 7: Robust version and change control must be introduced.** This control must cover templates, model runs, and reports that use the data output from the models.
- **Recommendation 8: Ongoing supervision of QA activities must be introduced.** This will confirm that QA Plans are being followed. Supervision should be through the contractors' internal audit function, who themselves should be supervised by the sponsoring Government department.
- **Recommendation 9: Full evidence of QA plans and checks must be collated and reviewed.** For the NAEI and similar multi-contractor models, responsibility for the overall QA of the assembled model may be unclear since individual contractors may have QA plans and records which are not shared with the contractor responsible for the ultimate assembly and delivery of the model. This main contractor – the Inventory Agency in the case of NAEI – must be given responsibility for collation and checking of all QA plans and records of the contractors that feed data into the model.
- **Recommendation 10: Departments must have emergency access to the model archive and associated documentation.** To ensure business continuity, it is important that Defra and DECC are be able to access completed model runs, including input data sets, the underlying templates and all design and user documentation.

Further details and discussion of these recommendations are contained in the body of the report.

1. INTRODUCTION

1.1 BACKGROUND TO THE REPORT

Quality Assurance (QA) is an essential part of the development, maintenance and use of models to support policy and decision making. Robust and appropriate QA reduces errors and gives decision makers greater confidence in using the model outputs to support their decisions.

Following the Macpherson Review of modelling in government, Defra initiated a review of the models that supported its business. As part of this, the Resource, Atmosphere and Sustainability Directorate has embarked on a strategic project to ensure robust and fit for purpose QA activities are in place to support the development, use and management of the analytical models being used to support its policy making.

Defra is committed to following the recommendations of the Macpherson Review to ensure that models and their outputs are used correctly to support the department's business. As part of the rolling programme implemented as part of the strategic programme, the following models were identified as ideal candidates for review, due to their key role in informing policy and reporting on Air Quality:

- The National Atmospheric Emissions Inventory (NAEI);
- The Pollution Climate Mapping model (PCM); and
- The Impact Pathway Methodology (IPM).

1.2 REVIEW OBJECTIVES

A Hartley McMaster team was appointed to:

- identify the QA policies and procedures in place for each of the models and how they are implemented;
- identify the QA policies and procedures in place for the interfaces between the high level models and between model components;
- seek out and review evidence of compliance with the existing policies and procedures;
- assess the suitability of those policies, procedures and activities in assuring the quality of the models and their outputs;
- assess whether the QA policies and procedures are fit for purpose with reference to published guidelines and model QA best practice; and
- make recommendations, where necessary, of ways in which the existing QA policies and procedures could be built upon, added to or improved to make the models more fit for purpose, meet modelling best practice and give greater assurance of the quality of the model's outputs.

1.3 OVERVIEW OF THE MODELS

The UK National Atmospheric Emissions Inventory (NAEI) is the reference standard for emissions of both greenhouse gases (GHG) and air quality pollutants in the United Kingdom and provides annual estimates of emissions to the atmosphere from UK sources at the highest level of disaggregation possible. At the front end the NAEI is made up of a series of linked Excel spreadsheets and Access databases. Outputs from the NAEI include air quality pollutants, greenhouse gases, regional pollutants contributing to acid deposition and photochemical pollution, persistent organic pollutants and other toxic pollutants such as heavy metals. A spatially disaggregated inventory is produced each year.

The mapped emissions are made available on the NAEI website[1] or through the online interactive Geographic Information System (GIS) tool[2]. The methodology for calculating air quality pollutant emissions is consistent

[1] http://naei.defra.gov.uk/

[2] http://naei.defra.gov.uk/data/gis-mapping

with the greenhouse gas inventory methodology. At the simplest level, historical emissions are calculated by combining an emission factor (for example, tonnes of a pollutant per million tonnes of fuel consumed) with an activity statistic (for example, million tonnes of fuel consumed). Commonly, activity data will consist of official national data sets such as population, Gross Domestic Product (GDP) and energy statistics. Projections of future emissions are also produced and reported separately as part of the inventory process.

Air quality outputs from the NAEI are used by Defra primarily to meet European Commission reporting requirements under the National Emissions Ceilings Directive (NECD, 2001/81/EC) and those of the United Nations Economic Commission for Europe (UN/ECE) under the Convention on Long-Range Trans-boundary Air Pollution (CLRTAP). The model outputs are incorporated into the UK's national air quality compliance assessments that are reported to the European Union (EU) under European Directives.

The Department of Energy & Climate Change (DECC) uses data from the NAEI to report the UK's greenhouse gas emissions to the United Nations Framework Convention on Climate Change, which also assesses progress towards emissions reduction targets under the Kyoto Protocol, and to the European Commission under the European Union Monitoring Mechanism. The greenhouse gas inventory is also used by DECC to assess the UK's progress toward its domestic emissions reduction targets under the UK Climate Change Act.

PCM is a collection of GIS based models used to estimate ambient airborne concentrations of key pollutants at background and roadside locations throughout the UK primarily to calculate population exposure, area and road length extents exceeding European limits and target values. PCM is not a full chemistry transport model; it is a collection of various model layers including interpolated measurements, dispersion models and emissions scenarios combined within GIS. It was specifically designed to fulfil criteria set down by the EU for modelling to augment the use of measurement stations.

IPM is an Excel based methodology for estimating the annual and cumulative impact of population exposure to airborne pollutants, both in life-years lost and economic impact, up to 100 years into the future. The Green Book, HM Treasury's guidance on appraisal of proposed policies, programmes and projects, recommends that the IPM approach is used in all cases where annualised impacts are estimated to be greater than £50m. Outputs from PCM are used to calculate population exposure to pollutants which are then used with health data and Life Tables produced by the Institute of Occupational Medicine to calculate the health impacts of short and long term exposure to air pollution using the methodology developed by the Committee on the Medical Effects of Air Pollutants (COMEAP). These health impacts are then monetised within Defra for use in Impact Assessments and other analyses.

NAEI and PCM are complex tools that have been developed by Ricardo-AEA in collaboration with Defra and DECC over successive compliance reporting contracts over the past decade. Third party models provide some of the inputs to the NAEI. Ricardo-AEA has developed a tool to estimate health impacts of air pollution using the IPM which they are contracted to provide to Defra.

1.4 The Work of the Review Team

This report presents the results of work undertaken by the review team between October 2014 and February 2015.

The review team:

- reviewed documentation on the QA policies and procedures in place for the three models and their sub-components;
- held face-to-face interviews with those responsible for QA of the models;
- conducted a series of on-site quality assurance audits of the overall QA and QA audit of the models;
- followed-up those audits with further clarification questions and requests for evidence of compliance with QA policies and procedures;
- considered published documents on model QA and specific QA guidelines to which the models are already subject in order to arrive at statements of best practice; and
- identified strengths and weaknesses of the existing approaches to QA and derived recommendations for improvement.

The on-site quality audits were core to the approach. Each was conducted in a similar fashion with the content structures to reflect the effective elements of quality assurance defined in the Macpherson Review and repeated in Figure 1. More details of the content are given in Appendix 1.

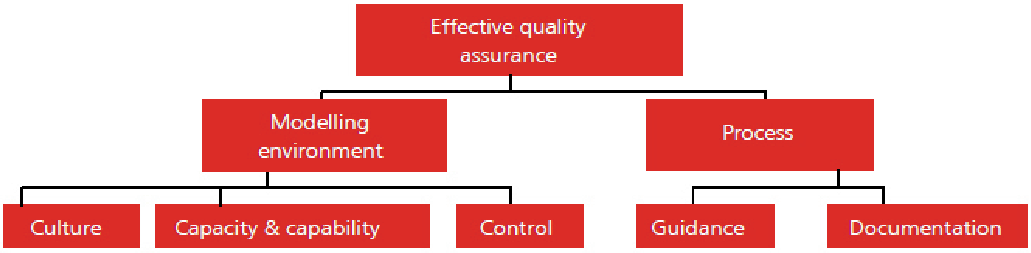


Figure 1 – Elements of effective quality assurance
Source: Review of quality assurance of Government analytical models, HM Treasury, March 2013

1.5 Structure of this Report

The rest of this report is structured as follows.

Section 2 describes best practice for Quality Assurance and Quality Control of models, with reference to three particular guides.

Sections 3, 4 and 5 describe the QA practices of the NAEI, PCM and IPM respectively, and compare them with the best practice guidance in Section 2.

Section 6 addresses a number of governance activities which cut across the three models.

Section 7 contains our finding concerning the modelling environment for the three models.

Section 8 considers QA issues arising at the interfaces between models.

Section 9 contains our recommendations.

2. BEST PRACTICE

There is no definitive and universally accepted statement of best practice for Quality Assurance and Quality Control of models. Instead, there is a range of guides that serve a variety of purposes to help modelling teams as they build and quality assure their models. For this review we have selected three guides that are particularly relevant for the air quality and greenhouse gas modelling work of Defra, DECC and their contractors:

1. 2006 IPCC Guidelines for National Greenhouse Gas Inventories, volume 1, chapter 6. Quality Assurance / Quality Control and Verification[3];
2. Review of quality assurance of Government analytical models by Sir Nicholas Macpherson[4]; and
3. DECC QA Guidance for Models[5].

It should be noted that this report predates the imminent publication of the Government Analytical Quality Assurance (AquA) Book, which will provide the official guidance on how to manage the QA of the development and use of models and other analytical activities. Each department will be able to use the AquA book to develop specific in-house guidance (as DECC have already done); by aligning with the AquA Book, the in-house standards will ensure that all aspects of the model lifecycle are covered by systematic quality processes.

Quality Assurance (QA) and Quality Control (QC) refer to activities that help ensure a model is error free and fit-for-purpose. In order to allow comparison of the guidance from these three sources, we group activities into the following categories taken from the Macpherson Review:

- Version control;
- Developer testing;
- Internal peer review;
- External peer review;
- Model audit;
- Quality assurance guidelines and checklists;
- Model documentation;
- Archiving;
- Governance;
- Transparency; and
- Periodic review.

In this section we compare the guidance from the three sources, category by category, providing where possible the precise statements of what is expected of modellers. In later sections we provide our view, by category, of whether a particular model satisfies the requirements stated in the guidance. We recognise that the guidance does not necessarily apply to the model in question – for example, Pollution Climate Mapping (PCM) is not governed by the IPCC Guidelines – however, it is useful to see how the QA/QC of models such as PCM performs against the best practice modelling guidance from authorities such as the IPCC.

Within the three sources of QA/QC guidance there is room for interpretation. In the words of Macpherson, "there is no shortcut or iron rule which can define the ideal type of QA for a given model". Macpherson provides a useful chart, reproduced below, which summarises which QA/QC activities should be used for which types of models.

---

[3] http://www.ipcc-nggip.iges.or.jp/public/2006gl/pdf/1_Volume1/V1_6_Ch6_QA_QC.pdf

[4] https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/206946/review_of_qa_of_govt_analytical_models_final_report_040313.pdf

[5] https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/384595/decc_qa_guidance_for_models.pdf

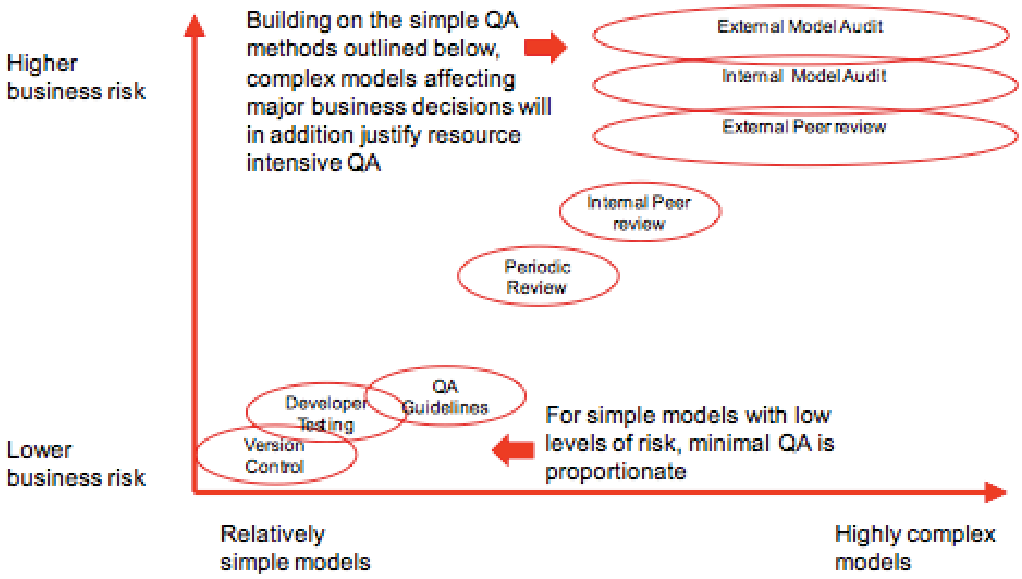


Figure 2 – Schematic showing indicative types of QA that might be expected given different levels of risk

Source: Review of quality assurance of Government analytical models, HM Treasury, March 2013

As Figure 2 shows, the amount of QA and QC undertaken for a model needs to be commensurate with the business criticality and level of risk inherent in the model. Furthermore, it should be sensitive to scope, cost and time constraints. The well-known quality triangle illustrates the trade-off between these four items.

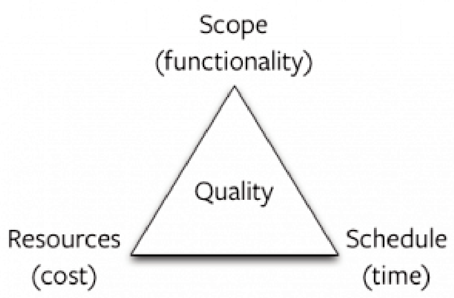


Figure 3 – The Quality Triangle

The projects reviewed in this report operate under tight timescales for delivery and this was taken into account in our findings. Such considerations may mean that best practice cannot always be followed.

The international reporting of greenhouse gas and air quality emissions, projections and maps imposes requirements on the method of measuring and modelling, including the approach to QA and QC. The guidelines are summarised in Table 1 below. Of these, the 2006 IPCC Guidelines are the most comprehensive, and serve as a reference for the European Monitoring and Evaluation Programme / European Environment Agency (EMEP/EEA) guidelines.

Report	Guidelines
GHG emissions	2006 IPCC Guidelines for National Greenhouse Gas Inventories
GHG projections	GHG Projection Guidelines (2012 EC publication)
GHG mapping	*No international reporting obligation hence no formal guidance*
AQ emissions	EMEP/EEA air pollutant emission inventory guidebook – 2013
AQ projections	EMEP/EEA air pollutant emission inventory guidebook – 2013
AQ mapping	EMEP/EEA air pollutant emission inventory guidebook – 2013
AQ measurement	*The requirement to measure air quality is defined in the 2008 EU Ambient Air Quality Directive. This directive assumes a regime based on measurement rather than modelling, so does not include guidance on how modelling should be carried out and what QA/QC procedures should be in place. Note that the Directive allows supplementary assessment through modelling, which is the approach the UK takes. The Forum for Air Quality Modelling (FAIRMODE) is currently considering model quality objectives with a view to developing standards in future.*
Local Authority CO_2 mapping	UK Statistics Authority QA Code of Practice

Table 1 – Applicable Guidelines

2.1 NOTE ON TERMINOLOGY: VERIFICATION VS. VALIDATION

The definition of the terms “verification” and “validation” used in this document mirror the definitions used by Macpherson, which themselves are taken from the international quality management system ISO9000. Verification is a process used to assess whether a model meets its specification. Validation is used to establish that a model meets its intended requirements. Verification is generally an internal process while validation reviews the output of the model to establish whether it is working correctly.

The DECC QA Guidance follows this definition, though methodology correctness, sensitivity testing, extreme values testing, re-performance testing and regression testing (sections 3.5.1, 3.5.3 to 3.5.6) are described as validation activities whereas we classify them as verification.

Confusingly, the IPCC guidelines’ definition of “verification” is similar to the Macpherson/ISO9000 definition of “validation”:

> *Verification refers to the collection of activities (…) that can help to establish its reliability for the intended applications of the inventory. For the purposes of this guidance, verification refers specifically to those methods that are external to the inventory and apply independent data, including comparisons with inventory estimates made by other bodies or through alternative methods.*

2.2 BEST PRACTICE COMPARISON: SUMMARY

	IPCC (note 1)	**Macpherson** (note 2)	**DECC**
Version Control			
Developer Testing: source data validation			
Developer Testing: model verification			
Developer Testing: model validation			
Developer Testing: model structure and clarity			
Peer Review: internal			
Peer Review: external			
Model Audit			
QA Guidelines & Checklists			
Model Documentation			
Archiving			
Governance			(note 3)
Transparency			
Periodic Review			

Notes:

1. The IPCC checks in Table 1 of the Guidance are optional. However, it is considered good practice to apply all checks to all parts of the inventory over an extended time period (IPCC section 6.6)
2. Mandatory vs Optional for Macpherson derived from Macpherson Chart 2C and elsewhere.
3. Describes governance within DECC.

Key:	
Mandatory	
Optional	
No guidance	

2.3 Best Practice Comparison: by Category

2.3.1 Version Control

	Text	Status	Reference
IPCC	Document the spreadsheet itself specifying its name, version, authors, updates, intended use and checking procedures so that it can be used as a data source of the derived results and referenced further on in the inventory process.	Optional	Box 6.4
Macpherson	[Models with high business risk] will be subject to basic version control processes. Systems in place to manage the development of the model and ensure any changes are captured.	Mandatory	Section 2.64 Table C.3
DECC	A clear system for version labelling. A version control log with clear documentation of the version history. It should be clear what has changed since previous versions and the distinction between a version and a scenario.	Mandatory	Section 3.2.6

2.3.2 Developer Testing: Source Data Validation

	Text	Status	Reference
IPCC	Cross-check a sample of input data from each category for transcription errors. Check for temporal consistency in time series input data for each category. The inventory compiler to confirm that national statistical agencies have implemented QC procedures equivalent to those in Table 6.1. Where possible, a comparison check of the national activity data with independently compiled activity data sources should be undertaken. National activity data should be compared with previous year's data for the category being evaluated.	Optional	Table 6.1 Table 6.1 Section 6.6 Section 6.7.2.1 Section 6.7.2.1
DECC	• Is a log containing key data characteristics (description, units, source etc.) available? • Has appropriate data been used? • Are the quality, characteristics, strengths and limitations of the data set fully understood and recorded? • Have data inputs been agreed and signed-off with the relevant approving body? • Is data plausible? Has input data been checked against primary reference for potential errors in copying / pasting / transforming? If required, have details on how the data have been	Mandatory	Section 3.6.1 Section 3.6.2

	imported/transformed or processed been recorded? • Have assumptions been fully understood and clearly recorded? • Are assumptions appropriate, applicable and logically coherent? • Are any limitations/caveats adequately described? • Are the quality, characteristics, strengths and limitations of the assumptions fully understood and recorded? • Have assumptions been agreed and signed-off with relevant stakeholders? • Are implicit assumptions also logged – for example an assumption of rational economic decision making?		

2.3.3 Developer Testing: Model Verification

	Text	**Status**	**Reference**
IPCC	• Reproduce a set of emissions and removals calculations. • Use a simple approximation method that gives similar results to the original and more complex calculation to ensure that there is no data input error or calculation error. • Check that emissions and removals data are correctly aggregated from lower reporting levels to higher reporting levels when preparing summaries. • Check that emissions and removals data are correctly transcribed between different intermediate products.	Optional	Table 6.1
Macpherson	[Models with high business risk] will be subject to developer testing. Individual testing of components of a model to ensure they are correctly coded and give the right result [Confirm] the logic flow within the model follows that defined at the model design stage, (at the level of individual units, multiple units or the complete code) Parallel builds – for complex, high-risk models there may be value in developing parallel builds to ensure cross-checking of results Sample testing of the range of validity of all input variables – this may not be possible for complex models, but parameter ranges of key variables should be tested. Input values outside the accepted ranges should also be included to test any exception and error handling within the model	Mandatory	Section 2.64 Table C.3 Table C.4

DECC	[Inspection to ensure formula correctness] [Check ease of use] [Review the code logic] [Check external links] [Implement auto-checks and error trapping] Review the model logic: • check the actual flow of data through the model against a stylised example of how data is perceived to be flowing through the model; • follow the evolution of the key model inputs across the model to understand how the data is transformed; and • check if the outputs of different scenarios (high, medium, low) are sensible. Sensitivity and scenario testing involves checking the response of the model to changes in variables. For some models, this will be a core part of their design, and it is critical to ensure that results make sense and are logical. Extreme values testing / model breaking is a test of the robustness of the model to values at the extreme limits of expected range. Re-performance testing involves implementing the model methodology in a completely new model. This may seek to completely replicate the functionality of the original in a shadow model, or may be a more simple calculation of key transformations from the original. Regression testing is used to give a model developer confidence that errors are not introduced when the model undergoes development. This can be done by using a set of standard inbuilt tests within the model which can be run after model development work. Additionally, a reference set of data, formulae or outputs can be compared before and after model development to confirm that changes made have only had implications in the intended areas.	Mandatory	Section 3.4.1 Section 3.4.2 Section 3.4.3 Section 3.4.4 Section 3.4.5 Section 3.5.1 Section 3.5.3 Section 3.5.4 Section 3.5.5 Section 3.5.6

2.3.4 Developer Testing: Model Validation

	Text	Status	Reference
IPCC	Check for temporal consistency in time series input data for each category. For each category, current inventory estimates should be compared to previous estimates, if available. If there are significant changes or departures from expected trends, re-check estimates and explain any differences. Significant changes in emissions or removals from previous years may indicate possible input or calculation errors.	Optional	Table 6.1 Table 6.1

Macpherson	Once the model is complete and has been subject to appropriate verification testing, a further validation testing phase should be conducted, and documented, to ensure the model is fit for the purpose.	Mandatory	Box 2.E
DECC	Does the model match historical results when using historical input data, to within an agreed tolerance level? This could also include a cross-check of model outputs against an alternative set of data or model.	Mandatory	Section 3.5.2

2.3.5 Developer Testing: Model Structure and Clarity

	Text	**Status**	**Reference**
IPCC	Check that units are properly labelled in calculation sheets.	Optional	Table 6.1
DECC	[Guidance on model and worksheet structure, formatting, formula clarity and robustness e.g. left-right, top-bottom logic rule; avoid hard-coded numbers; simplify formulas and break them into meaningful and easy to follow steps, use of error traps] Every input, table and output must be labelled and contain the correct units. Workbook comments should be clear and comprehensive As a general rule, each line of the code should be commented Named ranges (or tables), with an agreed naming convention, should be used extensively in the model	Mandatory	Sections 3.3.1, 3.3.2, 3.3.4, 3.3.7 Section 3.3.3 Section 3.3.5 Section 3.3.6 Section 3.3.8

2.3.6 Peer Review: Internal

	Text	**Status**	**Reference**
Macpherson	[Obtain] a critical evaluation from a third party independent of the development of the model, but from within the same organisation Independent review of model coding to ensure it meets the specification and is as free from errors as possible. This should be conducted by someone who is not part of the development team. Independent review of the verification testing results to ensure results are consistent with the model design specification. This should be conducted by someone who is not part of the development team.	Optional	Box 1.A Table C.3 Table C.3
DECC	It is essential that more detailed testing is conducted independently i.e. by someone who was not directly involved with the development. […] Ideally an individual who has the required skills in both the technique and the subject matter should be identified.	Mandatory	Section 3.1.2

2.3.7 Peer Review: External

	Text	Status	Reference
IPCC	Expert peer review consists of a review of calculations and assumptions by experts in relevant technical fields. This procedure is generally accomplished by reviewing the documentation associated with the methods and results, but usually does not include rigorous certification of data or references such as might be undertaken in an audit… Effective peer reviews often involve identifying and contacting key independent organisations or research institutions to identify the most appropriate individuals to conduct the review… The results of expert analyses from the UN Framework Convention on Climate Change (UNFCCC) processes … should only be considered as supplements to a nationally organised QA and review.	Optional	Section 6.8
Macpherson	Formal or informal engagement of a third party to conduct critical evaluation, from outside the organisation in which the model is being developed.[6]	Optional	Box 1.A
DECC	If appropriate, the project team (including your customer) should consider external QA	Optional	Section 3.1.2

[6] Table C.4 clarifies this "evaluation" to constitute model testing

2.3.8 Model Audit

	Text	Status	Reference
IPCC	[Internal or External] Audits may be used to evaluate how effectively the inventory compiler complies with the minimum QC specifications outlined in the QC plan. It is important that the auditor be independent of the inventory compiler as much as possible so as to be able to provide an objective assessment of the processes and data evaluated… In contrast to an expert peer review, audits do not focus on the result of calculation. Instead, they provide an in-depth analysis of the respective procedures taken to develop an inventory, and on the documentation available. It is good practice for the inventory compiler to develop a schedule of audits at strategic points in the inventory development. For example, audits related to initial data collection, measurement work, transcription, calculation and documentation may be conducted. Audits can be used to verify that the QC steps identified in Table 6.1 have been implemented, that category-specific QC procedures have been implemented according to the QC plan, and that the data quality objectives have been met.	Optional	Section 6.8
Macpherson	[Internal] A formal audit conducted within the organisation. This would need to be supported by full model specification and test documentation. [External] A comprehensive model- based audit would need to be supported by full model specification and test documentation, although a results-oriented audit might be a better alternative in a number of circumstances, particularly where there is regular updating and usage and "lower level" checks such as internal peer review are already in place. [External] The nature and extent of each of these types of QA may vary depending on what is appropriate for each model. An important example of this is external model audit, where there is a clear distinction between: • a comprehensive model-based audit which focuses on whether or not calculations are correct. This is likely to be resource-intensive but will probably only be needed once; and • a less detailed results-oriented audit which focuses on whether or not the results are reasonable. This should be quicker but is likely to be required each time the model is used.	Optional	Table C.4 Table C.4 Section 2.65

2.3.9 QA Guidelines and Checklists

	Text	Status	Reference
IPCC	The [QA/QC] plan should, in general, outline the QA/QC and verification activities that will be implemented and the institutional arrangements and responsibilities for implementing those activities. The plan should include a scheduled time frame for the QA/QC activities that follows inventory preparation from its initial development through to final reporting in any year. It is good practice for records of QA/QC activities to include the checks/audits/reviews that were performed, when they were performed, who performed them, and corrections and modifications to the inventory resulting from the QA/QC activity.	Mandatory	Section 6.5 Section 6.11.1
Macpherson	At the design stage, model design documentation […] should include the quality assurance strategy for the build and testing phases. At the test or 'deliver' stage, the documentation includes: a description of the tests run; the test results; any issues identified; and corrections made. Each business critical model should have clear documentation that sets out […] the quality assurance undertaken	Mandatory	Box 2.E Box 2.E Box 4.B
DECC	The QA plan should include: • an overview of QA activities you intend to carry out at each stage of the model cycle, with timings as appropriate; • a list of analytical professions, with potential names, who will be best placed to QA various aspects of the model; • details of the documentation you will provide as a record of QA activities, which should include a mandatory QA log as a bare minimum; and • an outline of the governance structure of the evidence you will be providing. As a minimum, and before a model is independently tested, it must be tested by the modeller. […] Documented evidence of this must be shown, both within the model itself (in terms of data and assumptions logs, model flow sheets etc.) and in the QA log for the model. Does the [QA] plan include a list of specific checks that need to be done on the model before outputs are shared? Is there evidence of QA processes carried out to date?	Mandatory	Section 2.5 Section 3.1.2 Appendix A Appendix A

2.3.10 Model Documentation

	Text	Status	Reference
IPCC	Examine the included intrinsic documentation to: • confirm that the appropriate data processing steps are correctly represented in the database; • confirm that data relationships are correctly represented in the database; • ensure that data fields are properly labelled and have the correct design specifications; and • ensure that adequate documentation of database and model structure and operation are archived.	Optional	Table 6.1
Macpherson	Each business critical model should have clear documentation that sets out the following: • the model's scope and specification; • the purpose, limitations and risks; • the quality assurance undertaken; • the identity of an appropriately senior model owner (Senior Responsible Owner or SRO) with overall responsibility to ensure the model is "fit-for-purpose", who will confirm the QA process is compliant and appropriate; that the model risks, limitations and major assumptions are understood by model users; and that the use for the model outputs are appropriate; and • that the model customer has understood the outputs and any major uncertainties, including the results of any sensitivity analysis. It is important that the design stage includes a clear understanding of the model structure and logic as well as the underlying assumptions, limitations, inputs required and outputs expected. The model SRO should at this stage check that the proposed design meets the organisation's requirements. They should check the assumptions, limitations, inputs and outputs to make sure they remain consistent with the intended use of the model, and discuss the most appropriate approach to QA.	Mandatory	Box 4.B Section 2.43
DECC	Documentation is critical for allowing the transfer of knowledge from developers to users, auditors and future developers. The scoping document for a model should provide an overview of the aims of the model, what function it is trying to fill and who the main stakeholders are. When testing the appropriateness of the specification documentation the following should be considered: • Does the model purpose and design match the specification document or does it need to be	Mandatory	Section 3.2 Section 3.2.1 Section 3.2.2

	updated? • Does it describe the main data inputs, calculations and outputs? • Is there a model map/logic diagram setting out how the model will achieve its purpose? The user guide should support independent use for a new model user who needs to run/operate the model and view outputs. The technical guide should explain the "nuts and bolts" of the model. This should be sufficiently clear to allow a model auditor or developer to understand how the model has been developed and to repeat the calculations if necessary. [For Visual Basic for Applications (VBA) or other code, it should be possible to] understand (via documentation) the purpose of the code.		Section 3.2.3 Section 3.2.4 Section 3.4.3

2.3.11 Archiving

	Text	Status	Reference
IPCC	Secure archiving of complete datasets, to include shared databases that are used in inventory development. This is particularly important for categories that rely on the multi-step development of emissions from a large set of primary data from outside sources.	Optional	Section 6.11.1

2.3.12 Governance

	Text	Status	Reference
IPCC	It is good practice for the inventory compiler to define specific responsibilities and procedures for the planning, preparation, and management of inventory activities	Optional	Section 6.4
Macpherson	There should be a single SRO for each model ("Model SRO") through its lifecycle, and clarity from the outset on how QA is to be managed. Key submissions using results from the model should summarise the QA that has been undertaken, including the extent of expert scrutiny and challenge. They should also confirm that the Model SRO is content that the QA process is compliant and appropriate, model risks, limitations and major assumptions are understood by users of the model, and the use of the model outputs are appropriate. The Accounting Officer's governance statement within the annual report should include confirmation that an appropriate QA framework is in place and is used for all	Mandatory	Rec 3 Rec 4

	business critical models. As part of this process, and to provide effective risk management, the Accounting Officer may wish to confirm that there is an up-to-date list of business critical models and that this is publicly available. This recommendation applies to Accounting Officers for Arm’s Length Bodies, as well as to departments.		
DECC	[Section 1.3 describes the accountability for QA within DECC across four roles: • Senior Responsible Owner (SRO) • Approving Body (AB) • Senior Analyst (SA) • Project Manager (PM)] The [clearance] statement should reflect the outputs of the QA process. The clearance decision […] should be based on the QA log and comments received from the testing process. The statement must include at a minimum: • the scope, type and level of QA that has been undertaken; • the key outstanding risks, uncertainties and issues with the model. Any significant remaining risks around quality must be clearly communicated to the approver; and • a viewpoint and explanation of whether the model and its outputs are fit for purpose. Approval/sign-off is the final agreement that the model may be used for its intended purpose, and must come after clearance Statement on QA that has and has not been done (and the associated risk),with a signature from someone suitably senior.	Mandatory	Section 1.3 Section 2.9 Section 2.10 Section 4.2

2.3.13 Transparency

	Text	Status	Reference
IPCC	There is sufficient and clear documentation such that individuals or groups other than the inventory compilers can understand how the inventory was compiled and can assure themselves it meets the good practice requirements for national greenhouse gas emissions inventories.	Optional	IPCC Introduction, Section 1.4[7]
Macpherson	Transparency is important because it facilitates effective scrutiny. Publishing all or some details of a model can therefore be a powerful quality assurance tool. Transparency can be a powerful tool […] as it allows the modelling team to harness the expertise of many third parties. Stakeholders often quoted external peer-review (whether through scientific publication or external model audit reports) as the gold standard of transparency.	Optional	Section 2.17 Section 2.50
DECC	[The guidance states that the model should be transparent, but no definition is provided]		

2.3.14 Periodic Review

	Text	Status	Reference
IPCC	[IPCC Guidelines do not explicitly address the need for periodic review, but given the model is intended for use on an ongoing basis, the importance of periodic review is implicit in the recommended QA/QC framework]		
Macpherson	Periodic review is an assessment of whether the model is fit-for-purpose when a model is being used on an ongoing basis or after a period of time has lapsed for a different use to that originally intended.	Optional	Section 3.23
DECC	Project reviews should include comparing the model with the original specification to complete the modelling cycle. The review can be used to agree with the customer that the model can be handed over and/or to determine timelines for further model development. Reviews should capture lessons learned, feedback and suggestions on model improvement. Future reviews of the model should be planned in for maintaining and upgrading the model. Reasons for this could include data refreshes, structural changes, changes in assumptions, different modelling techniques etc.		Section 2.12

[7] http://www.ipcc-nggip.iges.or.jp/public/2006gl/pdf/1_Volume1/V1_1_Ch1_Introduction.pdf

3. Current QA Practice: National Atmospheric Emissions Inventory

The UK National Atmospheric Emissions Inventory (NAEI) estimates emissions of a range of air quality pollutants and greenhouse gases in one over-arching database. To deliver these estimates, the NAEI team collects and analyses information from a wide range of sources, from national energy statistics through to data collected from individual industrial plants. Each year the latest set of data is added to the inventory and the full time series is updated to take account of improved data and any advances in the methodology used to estimate the emissions.

The compilation of the NAEI and reporting of air quality and greenhouse gas outputs is an annual cyclical process. The process begins in the summer after the year (n) for which the inventory is being compiled with requests for data being sent out to all data suppliers, followed by some pre-processing of complex datasets. Compilation of the NAEI takes place from September to November n+1 leading to delivery of the key air quality and greenhouse gas reporting commitments in December n+1 to February n+2, followed by supporting methodology reports. The inventory is finalised and locked down at the end of February n+2 to allow a range of subsequent deliverables to be completed over the spring and early summer of n+2 including spatial disaggregation of emissions, production of detailed road transport emissions and compilation of Devolved Administration and Local Authority inventories. Methodological updates are also undertaken before the next inventory cycle begins in the late summer.

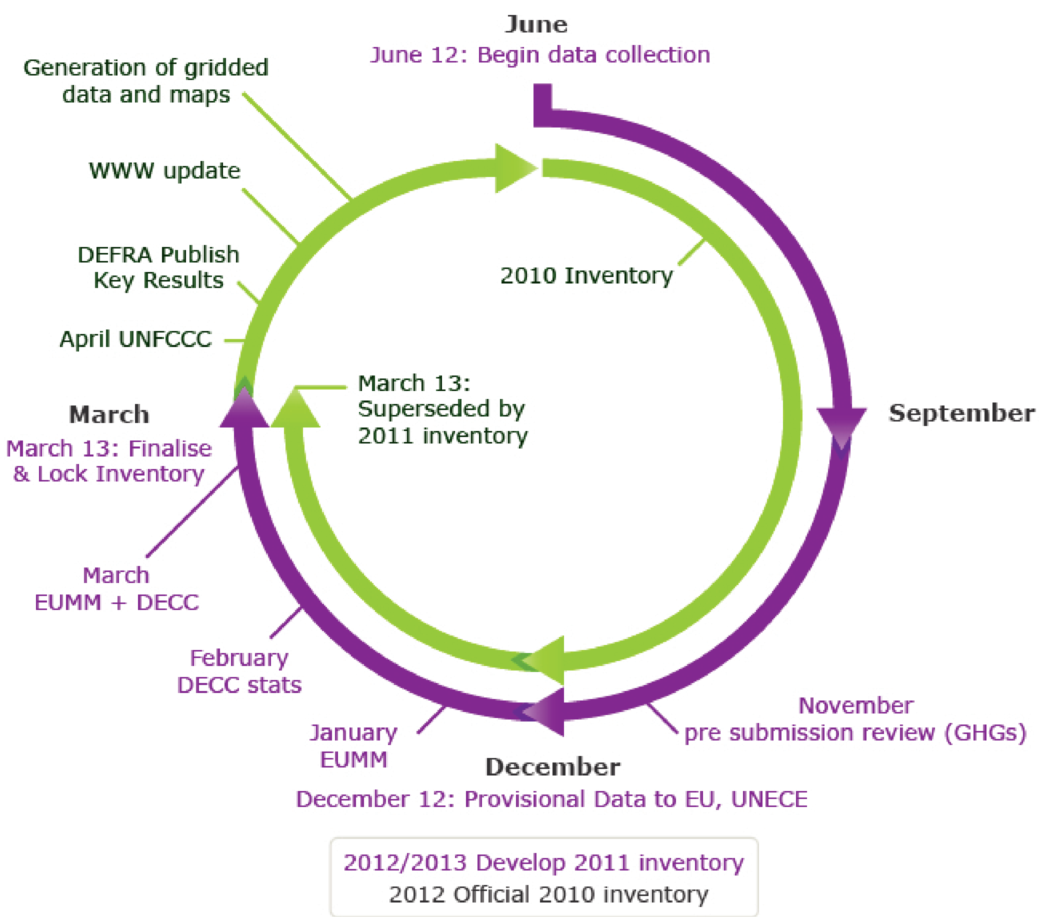


Figure 4 – Example Inventory Cycle
Source: NAEI website.

The work is jointly sponsored by Defra and DECC, and led by the Inventory Agency, Ricardo-AEA, supported by other contractors and subcontractors including Forest Research, Rothamsted Research and the Centre for Ecology and Hydrology.

3.1 Approach to QA

Ricardo-AEA describes a five-stage QA/QC process for the NAEI:

<u>Stages 1 and 2: Input Data Quality</u>

Whilst it is possible to maintain high standards of QA/QC on the processing and within systems managed directly by the Inventory Agency, the quality of the input data supplied can be variable. Meeting with data suppliers and the creation of data reporting templates and Data Supply Agreements for key data providers allows improved understanding of the data, and improved quality control. Quality audits are also regularly carried out to understand the QA/QC procedures data providers themselves utilise, and the extent to which this complies with similar inventory procedures.

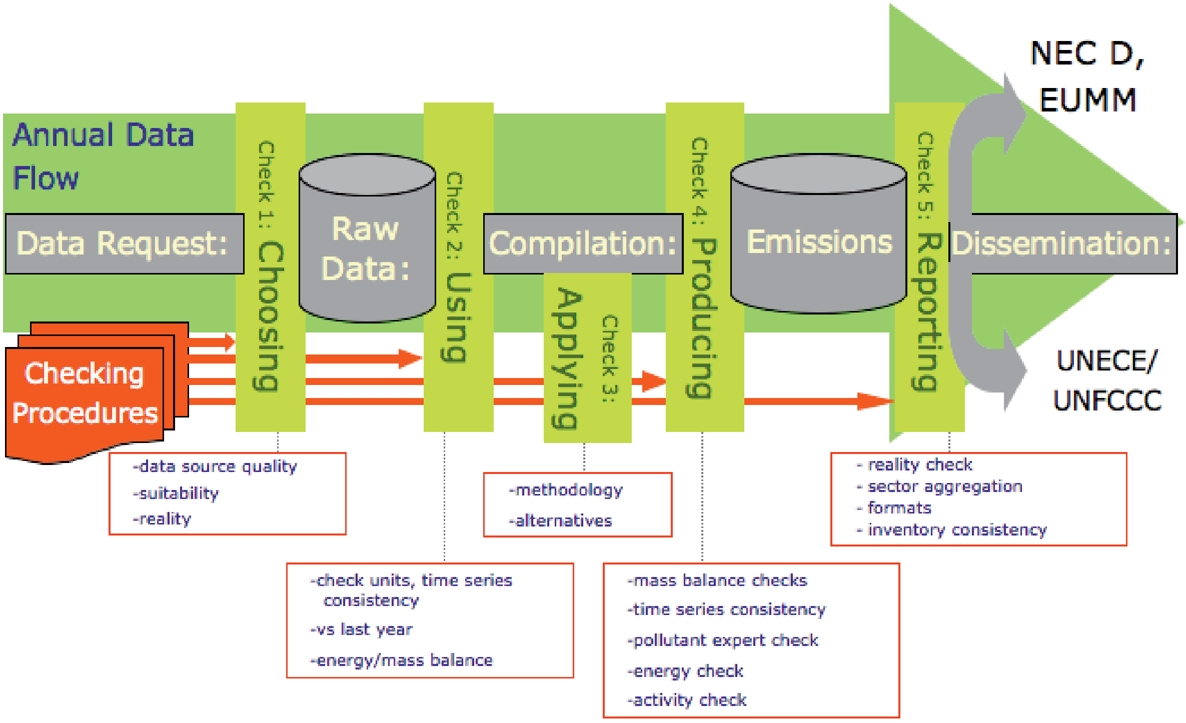


Figure 5 – NAEI Data Flow and QA/QC Checks
Source: Ricardo-AEA

<u>Stage 3: Spreadsheet Compilation</u>

There are a large number of QA/QC procedures which accompany this compilation stage. Each spreadsheet used for calculating estimates incorporates a QC sheet which includes key information including the unique identifiers and the spreadsheet version and Spreadsheet Reference Number, Spreadsheet Name, NAEI year, Status, Completion Date, Author, Approved by, Approval date, Description of contents, scope categories included, Activities, Pollutants, Years, a list of the data sources and reference materials, a colour-coding scheme for easy reference to data, calculations, checking cells; inter-dependencies: whether (and how) this spreadsheet interacts with other spreadsheets and results of QA/QC checks.

Although these spreadsheets vary considerably in their level of complexity there is a standardised procedure for completing the calculations:

1. The sheet is completed by the assigned compiler, and signed off as “final”.
2. The sheet is then checked by a second member of the team (there is defined guidance on the checks, which include methodology checks, logic checks, and inclusion of cross-checks and correct formatting). Any issues arising are addressed. The sheet is then assigned as “checked”.
3. There is then a second check by a Knowledge Leader (an experienced sector expert) or a Project Manager, with similar checks.

The sheet is then identified as being ready for uploading into the database.

A status spreadsheet known as the Mastersheet Summary Spreadsheet (see Appendix 3) links to all of the individual compilation spreadsheets and shows the progress, not only of the spreadsheet compilation, but also which data has been uploaded to the database.

Stage 4: Database Population

The central database is able to automatically upload data from the spreadsheets. However, as part of this upload there are a number of checks performed to ensure the data is complete, finalised and imported correctly. Once the system has checked that the individual calculation spreadsheets are finalised and up-to-date, the database then automatically uploads all output data from the spreadsheet into the compilation database. These systems ensure that the data, which is loaded from the spreadsheets into the database, is complete, and has been checked to standards as specified in the programme. There are then additional checks on the data in the database. Once all of these checks have been cleared, the database is then locked and no further changes are possible without permission from the project manager.

Stage 5: Reporting Emissions Datasets

Data extracted from the database typically requires formatting for formal submissions. In the case of the CLRTAP and UNFCCC/EUMM submission, a degree of automation has been incorporated into populating the required templates to minimise transcription errors. However, additional manual data entry and cross-checks are necessary and used to ensure that all data is correctly exported into the reporting templates. This ensures that the national totals agree with previously established data, and that the memo items are correctly reported.

3.2 QA Activities

NAEI	IPCC	Macpherson	DECC
Version Control	✓	✓	✓
Developer Testing: source data validation	✓✓		✓✓
Developer Testing: model verification	✓	✓	✓
Developer Testing: model validation	✓✓	✓✓	✓✓
Developer Testing: model structure and clarity	✓✓		✓
Peer Review: internal		✓	✓
Peer Review: external	✓✓	×	×
Model Audit	✓	✓	
QA Guidelines & Checklists	✓	✓	✓
Model Documentation	✓✓	✓	✓
Archiving	✓✓		
Governance	✓✓	✓	✓
Transparency	✓✓	×	
Periodic Review		✓✓	✓✓

Key:			
Mandatory	(blue)	Full adherence	✓✓
Optional	(yellow)	Partial adherence	✓
No guidance		No adherence	×

In the following sections we present our review findings and, where appropriate, a link to a recommendation that is fully described in Section 9 of this report. The links are shown as follows and are associated with the finding directly preceding them:
⇨ **Recommendation [number]**

The recommendations are present sequentially in Section 9.

3.2.1 Version Control

Mastersheets and most other spreadsheet models have a single version per year, identified by the inventory year, for example *Off-road_2011.xlsx*. Most did not create addition versions during development, instead working with a single file version and relying on daily automated backups to recover old versions.

There is no distinction made between the underlying spreadsheet or database templates, which often remain unchanged across many years, and the populated model, which is refreshed with new input data each year.
⇨ **Recommendation 1**

For spreadsheets, though the "date modified" was generally recorded, there was sometimes no record of what had changed between versions (i.e. between years). More complex models, such as databases, sometimes kept change control records in a separate document such as the User Guide (e.g. Final Users database).
⇨ **Recommendation 7a, 7b**

We saw two automated configuration control systems (Tortoise at CEH and Teamsource at Forest Research).

During the building of the model each year there is no locking or individual security protection for templates and populated models, even after sign-off; they are protected by the overall security arrangements for the inventory, with only authorised and trained personnel having access to a shared secured drive (designated the N: drive). Once the data outputs are submitted for the greenhouse gas inventory, all of the mastersheets are locked and changes cannot be made without requesting a password from the Project Manager.
⇨ **Recommendation 7e**

All changes to model outputs year-on-year are automatically identified, and the QC procedures for each mastersheet ensure that any differences are investigated, fully explained and allocated a "recalculation code". This ensures that changes to models that lead to changes in data outputs are fully documented.

3.2.2 Developer Testing: Source Data Validation

Source data validation in the NAEI is very thorough, using time-series checks (explaining changes year on year), explaining differences for a particular year across model iterations (for example, does the 2014 figure for 1999 match the 2015 figure?), comparing against alternative sources, and confirming that data has been imported correctly using internal consistency checks (e.g. the Aviation model). The team asks all data providers annually to provide information on their QA systems and specific QC routines, as well as data uncertainties and information on any management system standards adhered to. This information is not always provided.

3.2.3 Developer Testing: Model Verification

Internal consistency checks (using a "mass balance" or similar approach) and sense checks on outputs are common across the model components whilst verification checks to ensure a piece of code or query performs as intended are largely absent. The logic of some models, in particular the Master Spreadsheets, is so simple that verification testing is not needed; however, other modules (e.g. Final Users, Road Transport Emissions) contain complex queries or code which should be subjected to full testing. On two occasions (Agriculture and Off Road models) major rewrites uncovered calculation errors, so it is possible that there are calculation errors in other models that have remained undiscovered despite the internal consistency checks and sense checks[8].

The Aviation model is a good example of best practice in verification testing, with a Visual Basic script verified using a separate hand calculation and, separately, independent peer review of the code.
⇨ **Recommendation 2**

[8] Note that the interface issue between PCM and the Road Transport Emissions Projections arose because of the failure to test an apparently simple query.

3.2.4 Developer Testing: Model Validation

Each component of the NAEI, and the overall model, is subjected to time-series checks, trend checks, mass balance checks and against a variety of other sources, with significant variations explained.

3.2.5 Developer Testing: Model Structure and Clarity

The NAEI model is very large and complex, comprising around 100 component models, some of which are themselves made up of smaller spreadsheet or database models. It has been designed to allow flexibility as the formats of input data and the calculation methods are always evolving. The structure of the NAEI has evolved over many years; were it rebuilt today it would almost certainly be designed in a more integrated way. This decentralised structure, and in particular the large number of interfaces between its constituent parts, is a source of risk.

Individual spreadsheets within the NAEI are presented in a clear and consistent way. The NAEI colour-coding scheme is used to identify the sources of data for particular cells:

- Blue: content with no direct linkage (e.g. original data, copied and pasted input data). Data is labelled detailing its origin and the date it was last revised;
- Green: content that has been referenced from other places in the spreadsheet or other spreadsheets by cell referencing;
- Yellow: content that has been calculated or derived by a logical statement; and
- Pink: cross-checks; changes to red if check failed.

Access databases with embedded queries are used in many of the models but there were variations in the levels of clarity regarding their structure. In some the data structures were very simple, but others would benefit from improved clarity of the underlying data model and the queries used to process the data.

3.2.6 Peer Review: Internal

It is standard practice across the NAEI that individual component models are reviewed and approved by a Knowledge Leader who is independent of the development of the model. The Knowledge Leader is often a previous owner of the model so it could be argued that they are not truly independent. The alternative would be to choose someone who has less experience of the model (and, almost certainly, less experience of the science contained within the model). We do not believe this would result in a better QA/QC outcome since an increase in independence would be offset by a lesser ability to challenge the scientific assumptions behind the model.

Peer reviews by Knowledge Leaders focus on the science within the model and the reasonableness of the outputs when compared against a time series or other data sources. Knowledge Leaders do not review model logic in detail unless this is triggered by one of the output checks.
⇨ **Recommendation 2c**

The peer review of the Aviation module's Visual Basic script, conducted by a member of the Inventory Agency staff with software engineering experience, is an example of best practice.

3.2.7 Peer Review: External

A review of the NAEI model is carried out each year by the UNFCCC Expert Review Team. Other ad-hoc peer and bi-lateral reviews are documented in the National Inventory Report and the Informative Inventory Report. However, the UNFCCC and other reviews focus on the scientific methodology rather than the model logic and calculations. Both IPCC and Macpherson recommend a review of the model calculations be conducted by an external organisation. Given the nature of the models under review, and the other recommendations we are making in this report (specifically Recommendations 2 and 8), we do not believe an external review of model calculations is necessary.

3.2.8 Model Audit

Ricardo-AEA's internal audit function has traditionally audited business processes (e.g. project management, procurement) rather than individual projects. Going forward the intention is to introduce per model audits to check that models are following their own published QA/QC procedures. Models will be audited selectively and not every model every year.

⇨ **Recommendation 8b**

A one-off internal audit of the NAEI was carried out between June and August 2014. The scope of the work matched the scope of the review carried out by Hartley McMaster and documented in this report. There is some overlap in the recommendations. It was clear during the Hartley McMaster review that improvements in their approach to Quality Assurance and Control had been implemented by Ricardo-AEA as a result of this internal audit. Further improvements were also initiated as the Hartley McMaster audit progressed.

Ricardo-AEA has been subject to a number of third party audits and accreditations, including ISO 9001. None of these specifically evaluate the model itself (as recommended by Macpherson) or address how effectively the QC activities are being carried out (as recommended by the IPCC Guidelines).

⇨ **Recommendations 8a, 8c**

3.2.9 QA Guidelines and Checklists

The NAEI project maintains a wide variety of QA/QC materials, including:

- a QA Plan, summarising all QA activities for the model;
- QA Manuals for the larger and more complex modules, containing responsibilities for QA/QC activities and descriptions of tests to be done;
- a Mastersheet Summary Spreadsheet, used for tracking progress on the production of Mastersheets, including details of first and second level approvals; and
- a QA tab on all Mastersheets, containing a Progress & Checking section listing a selection of checks that have been performed, and their results.

We found these materials to be largely fit for purpose, though in some cases there were differences between the checks documented and those performed. Also, in models with multiple elements (e.g. Aviation) checks were recorded within individual elements rather than cascading down to the Mastersheet and so were not clearly visible to first and second checkers.

We recommend that for all component models:

- All checks are specified precisely. (⇨ **Recommendation 3**)
- The results of all checks are recorded. (⇨ **Recommendation 4**)
- Checks and results are documented in a clear and consistent way. (⇨ **Recommendation 5**).

The CARBINE model, built by Forest Research, has only recently been adopted for use within the NAEI and does not yet satisfy the QA requirements defined by the IPCC. This is a known issue within Defra and DECC.

⇨ **Recommendations 9a, 9b**

Where models had been handed over from a previous contractor (e.g. the Closed Coal Mine and Refrigeration models) QA materials were often missing, hence there was no evidence of QA/QC activities that had been performed on the model.

⇨ **Recommendations 8a, 8d**

3.2.10 Model Documentation

Comprehensive documentation has been published in the National Inventory Report (NIR), Information Inventory Report (IIR) and elsewhere, describing the NAEI's scientific methodology but, with a few exceptions (e.g. Aviation, Refrigeration) there is no documentation describing the design of the models themselves.

⇨ **Recommendation 6**

3.2.11 ARCHIVING
The NAEI adopts a consistent and clear file and workspace structure and nomenclature. With the exception of components created by other contractors, models and data are stored on a secure Ricardo-AEA server workspace that is backed up daily. At the end of each reporting cycle, all the database files, spreadsheets, on line manuals, electronic source data, records of communications, paper source data, output files representing all calculations for the full time series are frozen and archived on a central server. Electronic information is stored on hard disks that are regularly backed up. Paper information is archived in a Roller Racking system with a simple electronic database of all items references in the archive. Defra and DECC do not have access to these archives, or to copies of the models, in the event of an emergency.
⇨ **Recommendation 10**

3.2.12 GOVERNANCE
The Quality Plan for the NAEI defines roles, responsibilities and QA/QC activities for the whole model. Individual Quality Manuals for each NAEI component describe roles and activities in more detail, though these were still being developed as the Hartley McMaster review was being conducted.

The sign-off and approval for the use of NAEI data in a particular report is provided by a Knowledge Leader. We recommend this sign-off be at Director level for the contractor, and it should explicitly confirm that QA/QC checks are fit for purpose and have been completed.
⇨ **Recommendation 5d**

3.2.13 TRANSPARENCY
The IPCC guidelines state that documentation should exist to allow users to understand how the inventory was compiled. The National Inventory Report provides this information.

The Macpherson Review explains the importance of placing the completed model (rather than the results from the model) into the wider domain to allow scrutiny. Clearly this requirement is limited by commercial considerations (where the model IP is not Government owned) and by the need to ensure that the users of a model understand the scientific context; any decision to make the NAEI model fully transparent would need to be taken by Defra and DECC.

3.2.14 PERIODIC REVIEW
Macpherson recommends a periodic review to assess whether a model is fit-for-purpose when it is being used on an ongoing basis, after a period of time has elapsed or when it is used for a different purpose to that originally intended. The DECC guidance recommends regular reviews to compare against specification after changes have been made. The regular scientific peer review of the NAEI serves these needs.

3.3 RISK AREAS
The NAEI is a large and complex model, comprising around 100 component models. Some of these component models are simple NAEI Mastersheets, written in Excel, containing activity estimates and emission factors for a particular category of activity. Other component models are more complex, containing Access queries or VBA code, or series of spreadsheets performing more complex calculations.

For the NAEI the primary risks are:

- In the source data used to drive the model, some of which may not be subjected to rigorous QA/QC processes. Because of this, the source data is checked by the NAEI team using source apportionment, previous year comparisons, time-series consistency and, where possible, against alternative independent sources. Despite this, it is possible that source data may appear to be correct and pass all these tests, but still contain errors. For example, for many years CAA data feeding the Aviation model omitted certain flights, despite assurances from the CAA that the data was complete. There is no way this could have been known.
  ⇨ **Recommendations 6c, 9c**
- In the logic within a particular component model. For much of the logic there is no evidence of formal tests haven taken place. For simple component models this is not an issue, but for more complex models there is a risk that the logic contains errors.

- In the interfaces between the model elements e.g. when large blocks of data need to be cut and pasted from one element to the next. To address this, special checks have been set up to ensure the internal consistency of data that is carried across.
- In the communication of NAEI outputs to Defra and DECC. For example, in February 2014 it was noticed that the recently published “2012 UK GHG emissions” report contained different historic figures for landfill emissions than those published in DECC’s non-CO2 emissions projections report. After some investigation it transpired that the problem arose because of the use of the data within Defra and DECC, and not because of any failing in the NAEI version control: the two sets of data were taken from different inventory years, with significant changes to the calculated methane inventory in between.
  ⇨ **Recommendation 7d**

In addition to these modelling risks, there is a risk that the scientific approach produces erroneous results. This risk is addressed by the inclusion of checks against measured data from independent verification sites.

4. CURRENT QA PRACTICE: POLLUTION CLIMATE MAPPING

Pollution Climate Mapping (PCM) is a collection of models designed to report on the concentrations of particular pollutants in the atmosphere. These models are run by Ricardo-AEA on behalf of Defra. There is one model per pollutant (NO_X, NO_2, PM_{10}, $PM_{2.5}$, SO_2, benzene, ozone, As, Cd, Ni, Pb and B[a]p) each with two parts: a base year model and (for selected pollutants) a projections model. PCM provides outputs on a 1x1 km grid of background conditions plus around 9,000 representative road side values. PCM is also used for scenario assessment and population exposure calculations to assist policy developments and also provides model runs to support the writing of Air Quality Plans including Time Extension Notification (TEN) applications for PM_{10} and NO_X.

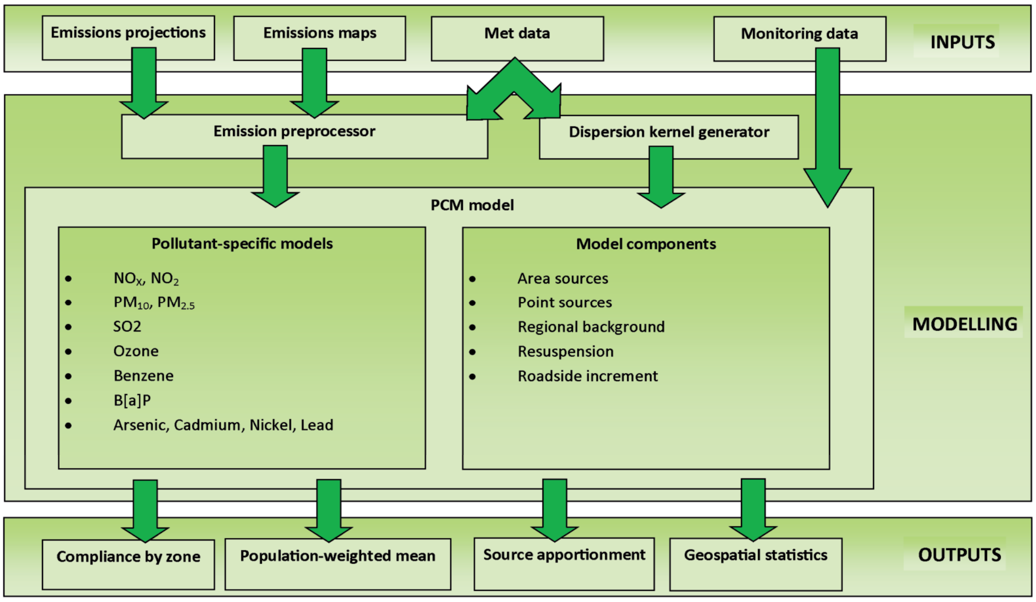


Figure 6 – Overview of Pollution Climate Mapping
Source: Ricardo-AEA

4.1 APPROACH TO QA

The three primary data sources are the NAEI, air quality measurement data collected under the Data Dissemination Unit (DDU) contract and published on Defra's UK-AIR website, and data from the UK Met Office. The first two sources are currently maintained by Ricardo-AEA and the input checks performed by the PCM team, and subsequent follow-ups between the project teams, constitute a useful source of "free" quality assurance of these two data sources.

The per-pollutant models are effectively templates, with input data refreshed each year. A range of checks are applied to these models, including checksums built into the models and automated checks to ensure data has been carried across correctly from the source databases. The builder of the model is expected to perform basic quality checks; additional checking of the populated model is done by a member of the team who is familiar with the model. To date these checks have not been documented and results have not been recorded, though a Checking Log has recently been created to capture these checks.

A detailed record of all model components and tasks, called the "Things to Do" list, is maintained and documents the team member and checker assigned to each, the status of the task, the deadline for the task to be complete and checked and any other relevant comments. During the modelling season (April to July) this list is updated regularly and is reviewed during weekly PCM team meetings.

A number of checks on the outputs are performed for all per-pollutant models, including:

- Calibration/Verification checks: model outputs are checked extensively during the calibration and verification stages. The model verification includes those national network sites used in the calibration (to demonstrate that the calibration has worked) and independent "verification sites" available to show the performance of the model overall. This is reviewed during model composition to identify any anomalies.

- Source apportionment, comparing each sector relative to each other and compared with the total concentration. This is reviewed during model composition to identify any anomalies.
- Previous year comparison: model outputs are checked against the previous year's outputs to check that models capture year to year changes.
- Historical model comparison: the current year's model results are also compared against long-term historical model results.
- Model intercomparison: the PCM models are actively compared against other models in regular model intercomparison work sponsored by Defra.
- Significant model evolutions are dry-run on the preceding year of modelling. This prevents unforeseen impacts propagating through to compliance reporting.

A continuous improvement log is maintained for the project, capturing improvements to be made in the science, modelling and approach to QA/QC. Larger improvement tasks need to be reviewed and approved by Defra prior to the start of any work.

The approach to QA is captured in a newly written QA/QC Manual.

4.2 QA Activities

PCM	IPCC	Macpherson	DECC
Version Control	✓	✓	✓
Developer Testing: source data validation	✓		✓✓
Developer Testing: model verification	✓	✓	✓
Developer Testing: model validation	✓✓	✓✓	✓✓
Developer Testing: model structure and clarity	✓✓		✓
Peer Review: internal		✓	✓
Peer Review: external	✓✓	×	×
Model Audit	✓	✓	
QA Guidelines & Checklists	✓	✓	✓
Model Documentation	✓✓	✓	✓
Archiving	✓✓		
Governance	✓✓	✓	✓
Transparency	✓✓	×	
Periodic Review		✓✓	✓✓

Key:			
Mandatory		Full adherence	✓✓
Optional		Partial adherence	✓
No guidance		No adherence	×

4.2.1 Version Control

The team maintains one version of each of the eight pollutant spreadsheets for each annual run of the model, with sub-versions created where necessary. A file naming convention is used which clearly identifies the component, year and sub-version. There is no change control sheet so the dates and nature of changes applied to a spreadsheet during each annual run are not recorded.

⇨ **Recommendation 7a**

There is no distinction made between the underlying per-pollutant templates, which often remain unchanged across many years, and the populated model, which is refreshed with new input data each year.

⇨ **Recommendations 1, 7c**

There is no locking or password protection of templates or populated models, so changes could be made after they are finalised and approved. The contractor is considering the introduction of spreadsheet locking, despite the impact on productivity during model development.

⇨ **Recommendation 7e**

4.2.2 Developer Testing: Source Data Validation

Data is compared against previous years' data, with significant changes challenged and explained. Additional detailed checks are done on the NAEI input data, in effect constituting acceptance testing of the NAEI model. Where possible, input data is also compared against an alternative independent data source. The team does not attempt to confirm that all providers of input data have implemented appropriate QC procedures, although the most significant data source – the emissions data from the NAEI – is known to have been subjected to appropriate levels of QA because of the close working relationship between the PCM and NAEI teams.

In the past there was a lack of formality in the sourcing of Road Transport Emissions Projections data from the NAEI team, though this has since been addressed.
⇨ **Recommendation 6c**

4.2.3 Developer Testing: Model Verification

PCM replicates the model process in both GIS and in spreadsheet tools to provide a completely independent check of the model mechanics.

4.2.4 Developer Testing: Model Validation

PCM model outputs are calibrated using measurements from real sites, and then verified using different "verification sites" which confirm that the model is performing correctly. In addition, each year's model is compared against long-term historical model results and against other models as part of Defra's Model Intercomparison Exercise.

4.2.5 Developer Testing: Model Structure and Clarity

The eight pollutant spreadsheets are presented in a clear and consistent way. The NAEI colour-coding scheme is used to identify the sources of data for particular cells:

- Blue: content with no direct linkage (e.g. original data, copied and pasted input data). These data should have metadata labelled above them detailing their origin and how they have been provided and the date they were last revised;
- Green: content that has been referenced from other places in the spreadsheet or other spreadsheets by cell referencing;
- Yellow: content that has been calculated or derived by a logical statement; and
- Pink: cross checks; changes to red if check failed.

The interaction between the pollutant spreadsheets and the various software components within PCM (GIS scripts, dispersion kernels and point source models) makes the overall model highly complex and difficult for a non-expert to follow. Higher complexity leads to higher risk, since the impact of the interplay of components is more difficult to understand.

Access databases with embedded queries are used in many of the models but there were variations in the levels of clarity regarding their structure. In some the data structures were very simple, but others would benefit from improved clarity of the underlying data model and the queries used to process the data.

4.2.6 Peer Review: Internal

Members of the PCM team with experience of models they are not responsible for are assigned checking activities in the newly introduced Checking Log. The checks are explicit and detailed. The checker is often a previous owner of the model so it could be argued that they are not truly independent. The alternative would be to choose someone who has less experience of the model (and, almost certainly, less experience of the science contained within the model). We are not convinced this would result in a better QA/QC outcome since an increase in independence would be offset by a lesser ability to challenge the scientific assumptions behind the model.
⇨ **Recommendation 2c**

4.2.7 Peer Review: External

Scientific peer review is carried out through published journal articles and Defra's Model Intercomparison Exercise. However, this work focuses on the scientific methodology rather than the calculations themselves. Both IPCC and Macpherson recommend a review of the model calculations, conducted by an external organisation. Given the nature of the models under review, and the other recommendations we are making in this report (specifically Recommendations 2 and 8), we do not believe an external review of model calculations is necessary.

4.2.8 Model Audit
PCM was subjected to a formal internal audit in 2014 after the interface issue was identified (see Section 8). Ricardo-AEA is planning the introduction of regular per model internal audits to check that models are following their own published QA/QC procedures. Models will be audited selectively and not every model, every year.

⇨ **Recommendation 8b**

Ricardo-AEA has been subject to a number of third party audits and accreditations, including ISO 9001. None of these specifically evaluate the model itself (as recommended by Macpherson) or address how effectively the QC activities are being carried out (as recommended by the IPCC Guidelines).

⇨ **Recommendations 8a, 8c**

4.2.9 QA Guidelines and Checklists
The need for a QA Manual (describing the overall approach to QA) and for a detailed Checking Log (specifying checks to be carried out) have been identified by the team, and both have been introduced in the last 12 months. The Checking Log is signed off per pollutant.

The team should ensure that within the Checking Log, or elsewhere in the QA/QC documentation, all checks are specified precisely.

⇨ **Recommendation 3**

4.2.10 Model Documentation
Comprehensive documentation has been published describing PCM's scientific methodology but it does not cover the design of the models themselves. In previous years the team wrote and maintained a selection of "recipe" documents designed to be a beginner's guide to each model in support of succession planning. These were labour intensive to revise and are now not actively maintained. Ricardo-AEA is considering reintroducing the recipe documents, subject to Defra approval.

⇨ **Recommendation 6**

4.2.11 Archiving
PCM adopts a consistent and clear file and workspace structure and nomenclature, which has been in place for over 10 years. Models and data are stored on a secure Ricardo-AEA server workspace that is backed up daily.

⇨ **Recommendation 10**

4.2.12 Governance
The PCM Quality Manual contains details of QA/QC roles for the project and sign off procedures for the model.

The sign-off and approval for the use of PCM data in a particular report should be at Director level for the contractor, and should explicitly confirm that QA/QC checks are fit for purpose and have been completed.

⇨ **Recommendation 5d**

4.2.13 Transparency
The IPCC guidelines state that documentation should exist to allow users to understand how the inventory was compiled. Although PCM is not subject to the IPCC guidelines, comprehensive documentation describing PCM's scientific methodology satisfies this purpose.

The Macpherson Review explains the importance of placing the completed model (rather than the results from the model) into the wider domain to allow scrutiny. Clearly this requirement is limited by commercial considerations (where the model IP is not Government owned) and by the need to ensure that the users of a model understand the scientific context; any decision to make the PCM model fully transparent would need to be taken by Defra.

4.2.14 Periodic Review
Macpherson recommends a periodic review to assess whether a model is fit-for-purpose when it is being used on an ongoing basis, after a period of time has elapsed or when it is used for a different purpose to that originally intended. The DECC guidance recommends regular reviews to compare against specification after changes have been made. The regular scientific peer review of PCM serves these needs.

4.3 RISK AREAS

PCM is a complex model, built from a number of elements across different platforms including:

- Spreadsheets (MS Excel);
- Database (MS Access);
- Dispersion model (ADMS 5.0);
- GIS (ArcInfo workstation); and
- Statistical models developed using the R statistical programming language.

Each element of the model is relatively simple and built using off-the-shelf software tools that can be relied upon to operate correctly. The primary risks are:

- In the source data used to drive the model, some of which may not be subjected to rigorous QA/QC processes. Because of this, the source data is checked by the PCM team using source apportionment, previous year comparisons, time-series consistency and, where possible, against alternative independent sources. Despite this, it is possible that source data may appear to be correct and pass all these tests, but still contain errors. See the table below for some examples of source data and the checks applied.

Data item	***Originator***	***PCM process/check***
Site classifications	*AURN & network providers*	*Sites coverage (modelling)*
Measured concentrations	*AURN & network providers*	*MAAQ monitoring summary (by zone and by station), time series analysis by station*
Met data	*Met Office*	*Dispersion kernels generation*
Area source grids	*NAEI*	*Area source grids*
Projections data	*NAEI*	*Projections data*
Point source emissions & stack parameters	*NAEI*	*Point sources modelling*
Road link emissions and ancillary data	*NAEI*	*Generation of cenlocs coverage and roadside modelling for individual pollutants*

Table 2 - Examples of data inputs and checks that are carried out

Source: Ricardo-AEA

- In the interfaces between the model elements e.g. when large blocks of data need to be cut and paste from one element to the next. To address this, special checks have been set up to ensure the internal consistency of data that is carried across.
- In the sequencing of application of these tools onto the data as it is transformed from input to output, for example the use of ARC Macro Language (AML) scripts to generate GIS grids. It is likely that any error would result the failure of the next step in the process since the data will be in the wrong format. If by chance the next step did not fail, it is likely that the results would be significantly out of line with previous year's results so would be detected by a range of output checks.

In addition to these modelling risks, there is a risk that the scientific approach produces erroneous results. This risk is addressed by the inclusion of checks against measured data from independent verification sites.

5. Current QA Practice: Impact Pathway Methodology

The Impact Pathway Methodology (IPM) is not a model per se, but an approach to valuing the consequences of changes to air quality on health. HM Treasury's Green Book recommends that IPM is used in all cases where the annualised impacts are estimated to be greater than £50m. Models created using IPM take data from the Pollution Climate Mapping model (PCM) to establish population exposure to pollutants; these exposures are then used with health data and the life tables produced by the Institute of Occupational Medicine to calculate the health impacts associated with long term exposure to air pollutants. Additional calculations are also carried out for the health impacts of short-term exposure and these can also be included, although the calculations are simpler and do not require the use of life tables. The health impacts are then monetised.

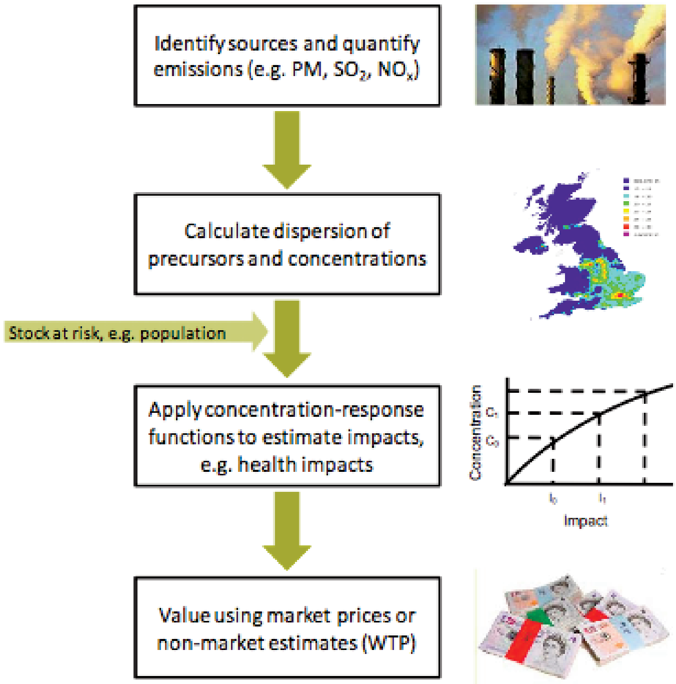


Figure 7 – Overview of Impact Pathway Methodology
Source: Impact pathway guidance for valuing changes in air quality, Defra, May 2013

For a typical IPM run, the tools used are:

- **PCM** (run by Ricardo-AEA) to convert projected emissions into population weighted pollutant concentrations. Two runs are required: one for the baseline and one for the scenario being assessed. This quantifies the exposure of people to changes in air quality;
- The **HealthCalcs Spreadsheet** (run by Ricardo-AEA) that includes the results from life table calculations for changes in life expectancy provided by the Institute of Occupational Medicine (IOM) and other data. This is used to quantify the health impacts associated with the change in pollutants, for example, using concentration-response functions that estimate the relationship between changes in air pollutants and changes in health outcomes; and
- A **Valuation Template** (usually run by Defra) to monetise the health impacts.

5.1 Approach to QA

PCM's approach to QA is covered in Section 4 of this document and is not repeated here, though note that two runs of PCM are needed and two Checking Logs would be completed, one for the baseline scenario and the other for the proposal under evaluation. The use of the HealthCalcs spreadsheet is covered by the same PCM QA/QC processes, with checks recorded in the Checking Logs. The Valuation Template is a relatively simple tool that takes inputs from the HealthCalcs spreadsheet to calculate economic impacts. It is owned by the

Defra Atmosphere & Industrial Emissions Evidence Team and comprises a single Excel worksheet containing 15 rows of data and calculations[9]. It was built in 2006 and there is no record of any QA/QC checks that might have been done at the time.

There are a variety of methods used by Defra to value the damage caused by air quality, such as the Impact Pathway Methodology (IPM), the damage costs approach and abatement costs approach[10]. The damage costs approach is used up to thirty times a year, though the full IPM methodology described in this document is applied much less frequently, and has not been used in the last two years. Air quality damage projects are led either by the Defra Atmosphere & Industrial Emissions Evidence Team, or by another government department or contractor, in which case the Defra Atmosphere & Industrial Emissions Evidence Team will QA the work. The output from a project is included in an Impact Assessment for a proposal. This Impact Assessment is reviewed to confirm "it represents a fair and reasonable view of the expected costs, benefits and impact". Implicit in this review is the confirmation that the underlying models are operating correctly.

The process for reviewing and approving an Impact Assessment involving the IPM is:

- Lead analyst uses the IPM method and produces evidence in an Impact Assessment.
- For Defra-led IPM projects, peer review of the Impact Assessment (within the department, and separately by a Defra Grade 7 economist from outside the Atmosphere & Industrial Emissions team). For projects led by other Departments, peer review is carried out by a Defra Atmosphere & Industrial Emissions Evidence team Grade 7 economist.
- Peer reviewer completes the IA Peer Review form. This asks a series of general questions. There are no standard checks on the approach or models used since each proposal is different.
- Approved by the Atmosphere & Industrial Emissions Evidence Team (for less than £20m impact) or Defra Deputy Director otherwise.
- Signed off by Defra Chief Economist.
- The quality of the evidence is assessed by the Regulatory Policy Committee.

Approvals are provided by email. There is no central log to record them.

[9] There are eight tabs but generally only one – the PM Chronic tab – is used since PM accounts for over 99% of the monetised impact.

[10] https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/197900/pb13913-impact-pathway-guidance.pdf

5.2 QA Activities

IPM	IPCC	Macpherson	DECC
Version Control	✗	✗	✗
Developer Testing: source data validation	✓		✓
Developer Testing: model verification	✗	✗	✗
Developer Testing: model validation	✗	✗	✗
Developer Testing: model structure and clarity	✓✓		✓✓
Peer Review: internal		✓✓	✓✓
Peer Review: external	✓✓	✓✓	✓✓
Model Audit	✗	✓	
QA Guidelines & Checklists	✗	✗	✗
Model Documentation	✓	✓	✓
Archiving	✓✓		
Governance	✓✓	✓	✓
Transparency	✓✓	✓	
Periodic Review		✓✓	✓✓

Key:			
Mandatory	(blue)	Full adherence	✓✓
Optional	(yellow)	Partial adherence	✓
No guidance		No adherence	✗

5.2.1 Version Control

The Valuation Template is effectively at version 1 and is stored on a Defra share drive. A version 2 will be created to include the impact of air pollution on productivity. The template is not locked or protected in any way, though it is archived.

⇨ **Recommendations 7a, 7e**

The Atmosphere & Industrial Emissions guidelines on version control are in the process of being established. In the meantime each project should maintain a configuration control document describing the template version and input data sets used (including PCM version and date).

⇨ **Recommendations 1, 7c**

5.2.2 Developer Testing: Source Data Validation

Source data from other models, notably PCM, is not tested as part of IPM since it is understood that a full set of quality checks have already been performed, though there is no formal confirmation that appropriate quality checks have been done. The Air Quality Evidence team sense checks the emission estimates.

There is a risk that assumptions behind the source data are not understood properly. This misunderstanding has resulted in previous errors in IPM (albeit for the noise modelling version).

⇨ **Recommendation 6c**

5.2.3 Developer Testing: Model Verification

The Valuation Template was built in 2006 and there is no record of any unit tests or other QA/QC checks that might have been done at the time. It has not been changed since then, though the impact of air pollution on productivity is currently being considered for addition into the template.

⇨ **Recommendation 2a**

5.2.4 Developer Testing: Model Validation

IPM has been compared to an alternative and broadly similar methodology, CAFE (Clean Air For Europe), in order to validate the approach. However there is no independent validation for individual proposals using IPM since there is rarely any meaningful model to compare against.

5.2.5 Developer Testing: Model Structure and Clarity

The Valuation Template is simple and clearly laid out, with text to guide the user incorporated into the spreadsheet itself.

5.2.6 Peer Review: Internal / External

A thorough independent peer review process for each run of IPM, which includes a review of the models behind the Impact Assessment, is described earlier in this section. This review may be internal (within Defra) or external (across Government departments) depending on the circumstances. This approach has been successful in catching errors in the past, for example the mis-transposition of numbers from PCM into the template.

The valuation template itself is widely used outside of Defra including by other government departments. This constitutes an effective peer review.

5.2.7 Model Audit

The use of the template by other departments and external contractors allows its critical evaluation so serves as an external audit (per the Macpherson definition). There has been no external audit of the QC of the valuation template or the IPM process. Given the simplicity of the template, we do not see this as necessary.

Internal or external audits of particular runs of the model, to develop individual Impact Assessments, are not needed given the oversight provided by the peer review process described above.

5.2.8 QA Guidelines and Checklists

No QA guidelines, checklists or similar materials exist for the Valuation Template. Despite its relative simplicity, it is good practice to create a QA Plan for this template, to be kept as part of the overall IPM documentation.

In addition to checking the Valuation Template, checks should also be made on runs of IPM, combining the operation of PCM, the HealthCalcs spreadsheet and the Valuation Template for a particular input data set. For PCM and HealthCalcs these checks are covering by the PCM Checking Log. For the master record of checks for an IPM run it is sufficient to record (on written confirmation from the PCM team) that these checks have been done. For all checks, the following principles apply:

- All checks are specified precisely (⇨ **Recommendation 3**)
- The results of all checks are recorded (⇨ **Recommendation 4**)
- Checks and results are documented in a clear and consistent way (⇨ **Recommendation 5**).

5.2.9 Model Documentation

An overview of the IPM approach is published on the Defra website and by HM Treasury as Green Book guidance. This overview describes the scientific and economic approach. There is no design or user documentation that describes the use of PCM, the HealthCalcs spreadsheet or the Valuation Template (for example, the processing steps).

⇨ **Recommendation 6d**

5.2.10 ARCHIVING

The Valuation Template is archived. Individual runs of IPM for particular Impact Assessments are stored in project folders on the Defra share drive.

⇨ **Recommendation 10**

5.2.11 GOVERNANCE

There is no QA governance of the Valuation Template. However, its simplicity and the governance around its use make additional governance unnecessary.

The peer review process defines clear roles for reviewers and approvers of each run of IPM. The completion of QC checks is reported upwards through the management structure.

5.2.12 TRANSPARENCY

The IPCC guidelines state that documentation should exist to allow users to understand how the inventory was compiled. Although IPM is not subject to the IPCC guidelines, documentation describing the IPM approach, held in the public domain, satisfies this purpose.

The Valuation Template is made available for use, and therefore detailed scrutiny, across Government. The IPM methodology is made available to the public via the Defra website.

Individual Impact Assessments are published and set out to fully explain the evidence base. The actual models behind them are not currently published; this would be challenging given the complexity of the PCM component and the ownership of the IP.

5.2.13 PERIODIC REVIEW

The use of the template by other Government departments constitutes a regular assessment of its fitness for purpose.

5.3 RISK AREAS

The risk of modelling errors in IPM is relatively low. The individual tools are low risk either because they are subject to detailed QA/QC activities (PCM and the HealthCalcs spreadsheet), or are very simple and subject to regular peer review (the Valuation Template). There is a small risk in manual transposition of output data from one model into another, though the comprehensive and detailed peer review process is likely to catch any errors. There is another risk on the interface between the elements resulting from the potential failure to communicate assumptions and data definitions; this can be addressed by formalising the request for data in a specification.

IPM has components for which the ownership rests with Defra, IOM and Ricardo-AEA. Advice has also been sought from organisations such as the Committee on the Medical Effects of Air Pollutants (COMEAP), Public Health England and the Interdepartmental Group on Costs and Benefits in its design. It would be possible and sensible to develop a single model, comprising a consistent set of spreadsheets under single ownership, to value the health impacts of air quality changes. This model could be extended to include the non-health impact pathways. Clearly any such change would have cost implications for Defra.

6. Additional Governance Activities

6.1 Ownership of IP

Most IP across the three models is owned by either Defra or DECC. There are complications when a model uses a component that has already been developed by a third party (e.g. Forest Research's CARBINE model) and is being used for purposes beyond those specified by Defra or DECC. In these cases the third party retains the IP for the model they have developed.

In some cases the contractor owns the IP for a component with Defra or DECC owning the IP for the output. For example, the IP for Mapping scripts and routines is owned by Ricado-AEA, whilst the IP for the outputs (in the form of GIS datasets and coverages) is owned by Defra.

Details of the NAEI IP ownership is referenced in Defra's 2011 ITT for the NAEI, section 1.5: "Some Intellectual Property (IP) rights are owned by Defra and some by AEA. A summary of the current standing of IP has been provided as part of the ITT documentation (Document Ref: 45321001/2009/CC7292/JMC to be provided separately)."

Defra owns all the IP for PCM and for the Valuation Template within IPM.

6.2 QA Contractual Arrangements

6.2.1 NAEI

The ITT for the NAEI states that the contractor should "compile and maintain a core NAEI database of activity data and emissions factors in accordance with the relevant international guidance for production of the AQP & GHG inventories". The relevant guidance is the IPCC 2006 guidelines (for greenhouse gases), and the EMEP/EEA air pollutant emission inventory guidebook (for air pollutants), both of which specify the required QA/QC practices.

In addition, the ITT contains the following section on QA/QC:

> *The Contractor must implement rigorous quality assurance and control procedures for the management of the data for all inventory activities including data submissions. The guidelines of the Intergovernmental Panel on Climate Change concerning determination of uncertainty and QA/QC shall be followed for all NAEI pollutants as far as is practicable. For National Statistics purposes the Contractor shall provide information for a Quality Assurance template on air emissions data and participate in the Quality Review process.*
>
> *The Contractor will develop a QA/QC plan to be agreed by the Authority and updated annually on 31 August (Deliverable 57).*

The ITT contains the following section on documentation. Note that this describes the documentation of the scientific methodology; there is no stated requirement to provide design documentation for the model code and logic.

> *Inventories must be fully traceable and transparent to reviewers, the Authority and other stakeholders. The Contractor will provide adequate data on basic assumptions and novel analyses to accompany all data provided to the Authority. All inventory material and alterations to assumptions, sources or emission factors must be clearly documented.*
>
> *The methodologies, emission factors and assumptions used in compiling the inventory should be continuously reviewed as new information, data sources and understanding of factors influencing emissions emerges. This is essential to ensure the quality of the inventories is maintained and improved wherever possible. Improvements can lead to changes in the historic time-series in reported emissions. It is vital for stakeholders to understand the causes and reasons for any changes. Where significant revisions of data or methodologies have been made, information will be provided on reasons for changes in publications including within the IIR and NIR.*

6.2.2 PCM

The contract for PCM mentions QA activities but is not explicit about how these should be undertaken and makes no reference to a dedicated QA plan.

6.2.3 IPM

IPM comprises three elements: the PCM and Healthcalcs spreadsheet are covered by the PCM contract (see 6.2.2) while the Valuation Template is generally run within Government departments so there is no contract involved.

6.3 Continuity Planning

The NAEI project team maintains two relevant documents:

- a risk assessment, by scientific area (e.g. F-gases, transport, waste), listing qualified staff and identifying those who need further development; and
- an inventory training plan, identifying where members of the team need to train other members on how to build and maintain a particular model component.

The training plan, rather than the risk assessment, flags where there are continuity risks. In some cases the plan seems to be reactive rather than proactive. For example, there appeared to be no cover for the member of staff who until recently maintained the NAEI database. The training of a replacement was triggered by his departure. We recommend that the staff continuity plan be included in the model documentation approved and signed off by the Senior Responsible Officer or senior executive in the contractor's organisation.
⇨ **Recommendation 5d**

The NAEI documents describe Ricardo-AEA owned components and not those owned and managed by other contractors. For the other contractors:

- Forest Research does not have a formal continuity plan, though there is partial cross-coverage within the team. The cross-coverage does not cover every part of the model so there are some continuity risks.
- Rothamsted Research and CEH do not have a formal continuity plan, though there is good cross-coverage within the team.

We recommend that the existence of appropriate staff continuity plans for all contractors working on a model should be confirmed as part of the approval and signed off by the Senior Responsible Officer or senior executive in the contractor's organisation.
⇨ **Recommendations 5d, 9a**

The PCM project team maintains a staff continuity plan as part of the risk assessment for project, detailing roles and staff identified as short and long term cover.

IPM maintains a formal staff continuity plan, with the team of four staff all proficient in its use. They observed that it is challenging for other teams to use the methodology.

One contractor noted that regular retendering of model development does not encourage continuity planning since the contractor may lose the contract and therefore waste their investment.

6.4 Communication of Model Outputs, Uncertainties and Risks

QA issues in the communication of model outputs are covered in Section 3.3 (for NAEI) and in Recommendation 7d.

Model uncertainties for the NAEI are reported as part of the standard UNFCCC annual reporting pack. For PCM the requirement for reporting model uncertainty is defined in the Air Quality Directive and described in the PCM QA/QC Manual. IPM takes its air quality data from PCM so is covered by the same approach.

Model risks (i.e. the risk that the model is not performing as specified) are in effect communicated through the various QA Manuals and Plans for the models under review. These documents describe the approach to model quality and the checks that are done to assure it. The lack of reporting of actual checks performed against the models, and the success or failure of these checks, means that the actual model risk is not being effectively

communicated to users.
⇨ **Recommendation 5d**

In line with the Macpherson recommendations, we suggest that when unavoidable time constraints prevent QA, it should be explicitly acknowledged and reported.

6.5 Continuous Improvement of Models and Working Practices

6.5.1 NAEI

The NAEI has an annual Inventory Improvement Programme, described in detail in the UK Informative Inventory Report, which covers the scientific methodology, the models and the working practices, including the QA/QC practices. Defra is responsible for improvements to the air quality parts of the inventory, and DECC for the greenhouse gas parts. An example of the improvement of QA/QC practices was the introduction of recalculation codes to explain year-on-year differences in the NAEI outputs (see Section 3.2.1).

In addition, some contractors have annual reviews of their QA/QC practices with Ricardo-AEA.

6.5.2 PCM

The PCM project maintains a continuous improvement log, categorised by pollutant, which is reviewed throughout the year and in particular each April prior to the annual run of compliance reporting. Significant changes require the review and sign-off by Defra. In addition, in April 2015 Ricardo-AEA will introduce an annual review to consider how to improve the model QA, covering for example new items added to the model which are not covered by existing QA/QC checks, or QA/QC activities that are no longer needed because of changes to the model.

6.5.3 IPM

The parts of the IPM process delivered by PCM are covered in the section above. Beyond that, there is no formal continuous improvement programme for the Valuation Template; given the simplicity of the template this would not be appropriate. Note that no method exists to review or update the Healthcalcs spreadsheet. It may not be consistent with latest COMEAP advice.

6.6 Security Markings

Contractors do not apply security markings (i.e. Official/Secret/Top Secret) to intermediate or final model deliverables. Output data from the NAEI is published in the form of official statistics, these having very limited release prior to their announcement. Defra and DECC expect contractors to use protective marking for sensitive information such as statistical releases but this is not currently enforced for the NAEI contract. The onus is on government officials to ensure contractors are following this policy where necessary.

7. MODELLING ENVIRONMENT

The Macpherson Review distinguishes *process requirements* and *modelling environment requirements* for effective QA. The bulk of this report addresses process requirements, covering the activities that organisations need to undertake to deliver reliable models. Modelling environment requirements are softer, and cover the organisational culture, the capacity (i.e. the amount of staff with the right skills and time available) and the capabilities of these staff. This section contains our observations on the modelling environment for the NAEI, PCM and IPM.

7.1 SCIENTIFIC VS. MODELLING FOCUS

We found a culture of scientific excellence amongst the contractors developing these models. There is great interest in, and care about, the nature of the science behind the models. The focus of QA/QC activities is on the science and does not always carry over to the testing and documentation of the models themselves.

Because the models have mostly been built by scientists rather than software specialists, concepts such as formal specification, design, unit testing and acceptance testing are unfamiliar. Detailed checks have been done on the models, but these are nearly always scientific checks, for example checking that outputs look sensible when presented as a time-series. Software checks, such as regression testing after model changes, are largely absent.

In many cases the models are very simple in terms of the calculations involved (e.g. combining an activity output volume with an emissions factor), albeit with many such calculations per model. Because of this, the degree of attention paid to the science and the understanding of data sources has not been brought to the development of the model templates themselves. Clearly this does not mean that templates have not been tested, but the QA emphasis has often been elsewhere and the collection of evidence reflects this.

7.2 PROJECT MANAGEMENT DISCIPLINE

The IPM methodology is owned and managed by Defra. The approach to running the IPM models appeared to be a business-as-usual activity. We would recommend more project management discipline, for example the creation of a project plan, a QA plan and other deliverables, as is done for NAEI and PCM runs.

As is noted elsewhere in this report, requests for model outputs from Defra and DECC to the contractors, and between contractor teams, are often informal emails. We recommend increased discipline in these requests, so that they should be formalised in a specification.

7.3 EVOLUTION OF THE QA/QC PROCESS

The NAEI and PCM contractors are constantly improving their quality processes. A number of improvements appeared during 2014 prior to our review, including:

- the NAEI QA/QC Manuals;
- the PCM QA/QC Manual;
- the PCM Checking Log, recording checks to be done and those completed;
- some User Guides, including Final Users user guide, were revised prior to our review; and
- the Aviation model unit testing, peer review and documentation.

8. Model Interfaces

In early 2014 a mistake in a dataset provided by the NAEI project to the PCM team led to an error in the reported NO_X emissions rate for specific census locations. The error was thoroughly investigated by Ricardo-AEA at the time and actions were taken to ensure it would not occur again. It was described as an "interface" issue because the error occurred at the interface of the NAEI and PCM projects.

In our work we examined this issue in case it was representative of other errors that might occur in one of the models under review, and that should be addressed by changes to the QA/QC approach. Our conclusions were:

- The root cause of the NAEI/PCM interface issue was an error in an Access query within mapping procedures taking data from the Road Transport database. This error had existed for a number of years so had not been discovered using the standard time-series checks used across the NAEI. In the event that such errors do not manifest themselves in visibly erroneous output data, they can only be discovered by validation testing or peer review. ⇨ **Recommendations 2a, 2c**
- The interface between projects such as NAEI and PCM, and between component models within the NAEI, is a genuine area of risk. For this reason, across the models under review there was significant focus by the contractors on detailed and thorough tests on delivery and receipt of data between models and components.
- We found examples of misunderstandings between data providers and data requesters on the nature of the data provided. For example, the CAA provided incomplete flight data to the Aviation model for a number of years, and an error was discovered in an IPM[11] peer review resulting from the misinterpretation of the units in source data. These errors have both been fixed, though similar undetected errors can only be avoided by improving the specification of the data to be provided. ⇨ **Recommendations 6c, 9c**

[11] An IPM model built for noise rather than air quality.

9. RECOMMENDATIONS

The review team found that the three models were well designed and well built by experienced professionals, who operate under very tight timescales for their delivery. The QA policies and practices adopted by the model builders were evolving during the review, and by the end of the review compared relatively well against best practice as defined by the IPCC 2006 QA guidelines, the DECC QA guidelines, and the guidance within the Macpherson Review final report. Specifically, the two models classified as business critical by Defra – the NAEI and PCM – fully or partially adhered to all guidance that is considered mandatory by these three sets of guidelines; IPM, which is not classified as business critical, failed to follow some of the mandatory guidance, though the risk arising from this failure is relatively low.

These findings should give Defra confidence that the QA policies and practices used for the three models are largely fit for purpose. Despite this, modelling errors can still occur, and the team saw examples of these errors in their review, such as the interface issue (section 8) and errors uncovered in the rewrite of the Agriculture and Off Road models (section 3.2.3). They noted that models had mostly been built by scientists rather than software specialists, so concepts such as formal specification, design, unit testing and acceptance testing were largely unfamiliar. Detailed checks had been carried out on the models, but these were nearly always scientific checks, for example the confirmation that outputs look sensible when presented as a time-series. Software checks, such as regression testing after models have been changed, were largely absent. By moving to a software-style QA regime for these models, more can be done to reduce the risk of future errors. In this report the Hartley McMaster team makes the following recommendations, which apply across all business-critical models built or commissioned by Defra and DECC unless stated otherwise.

In the list of recommendations, each has been assigned an owner, the recommended timing, the relative cost of implementation (high/medium/low) and the relative value delivered, in terms of the likelihood to reduce future errors (also high/medium/low). The cost and value assessments are impressionistic as this stage and will be firmed up in future discussions between Defra, DECC and the contractors.

Recommendation 1: QA of the model templates must be distinct from QA of runs of the model

Owner: Model Developer
Timing: three years
Cost=High, Value=High

The three models under review – the NAEI, PCM and IPM – share the characteristic that they are run repeatedly on different input data. The NAEI and PCM are run annually, whilst IPM is used on an ad-hoc basis for assessing the impact of policy proposals. Each model is made up of a set of components: Excel spreadsheets, Access databases and (for PCM) routines written in specialist software languages. These components are treated as templates for each run of the model: at the start of the run, data is cleared out, then new input data is imported and the processing steps are performed to generate end results.

Although the models are run at least annually, the underlying templates change less frequently. For example, many of the components of the NAEI have remained largely unchanged for 10+ years. For this reason, it is efficient from a quality perspective to separate out the QA/QC of the templates from the QA/QC of runs of the model:

- QA/QC of the templates must be performed each time a template is changed. It involves model verification i.e. testing the new functionality and regression testing pre-existing functionality.
- QA/QC of runs of the model must be performed each time the model is used. It involves source data validation, ensuring processing steps have been performed correctly, and validating the output data ("model validation"). Note that in contrast to the verification testing above, testing of the model runs is covered by existing QA/QC practices.

QA plans and manuals must reflect this distinction between QA of the templates and QA of runs of the model.

Recommendation 2: All templates must be fully tested and peer reviewed

Owner: Model Developer
Timing: three years
Cost=High, Value=High

This recommendation is in four parts, each of which is covered in detail below.

Recommendation 2a: Verification Testing

For each template, the following verification tests must be performed on creation and each time the logic within a template is changed, with results clearly documented:

- a regression test to ensure that the original functionality continues to operate as intended, with major routes through the code checked; and
- a test for the new functionality.

Both tests should define input values and expected output values for the model. The expected output values should be hand calculated without using the model (e.g. using a separate spreadsheet or paper working).

For the purposes of this recommendation, a template is a spreadsheet, database or software module that is used repeatedly. There is no distinction between simple templates such as NAEI Mastersheets and more complex templates such as the NAEI Aviation model. Both must be properly tested. However, the nature of the tests will be driven in part by the complexity of the template, with simpler templates generally requiring fewer tests.

Recommendation 2b: Acceptance Testing

It is standard practice in software development for a system to undergo formal acceptance testing by the customer to ensure it satisfies the documented requirements. In the case of the three models in this review the relevant Knowledge Leader or a similar senior officer within the contractor organisation must perform a surrogate acceptance test on behalf of Defra and DECC for each major template version release. In this case acceptance testing constitutes confirmation that the template performs to the scientific specification documented in the NIR, IIR or similar literature. Acceptance testing of the template must be documented alongside the verification testing described above. Note that this is different to, and independent of, the checks made by Knowledge Leaders on particular (usually annual) runs of the model.

Recommendation 2c: Independent Peer Review of Model Logic and Code

It is good practice for all templates to be reviewed by someone independent of the modelling team, either from within the organisation or outside. This review must be performed on creation and each time the template is changed, with results clearly documented. Specialist scientific knowledge is not needed; the reviewer must be an experienced modeller and must understand and challenge the model logic. The reviewer must not be the person who originally developed the model. It is possible to combine the roles of peer reviewer and verification tester, though it is preferable to keep these separate for more complex models.

Recommendation 2d: Scheduling of a Three-Year Programme of Verification, Acceptance Testing and Peer Review

This recommendation is specific to the NAEI, PCM, IPM and any other business-critical models that have not been fully tested.

It is not practical to test and peer review all the components of the NAEI, PCM and IPM immediately. Instead, a programme of work should be defined to complete this testing over a period of three years, according to these principles:

- any template undergoing major change must be tested at the time the change is made;
- other templates must be included in a rolling programme, prioritised by the relative impact of the model on the final outputs, the uncertainty associated with the model, and whether significant changes have recently been made to the model or the methodology; and

- these tests should be done "out of cycle" i.e. at a time when the templates are not needed for the building of a model. This ensures that resource and time constraints are avoided, and encourages the good practice of separating the testing of the templates from the testing of a model run.

The SRO for each model must be accountable for the completion of this work programme, and this risk and mitigation must be recorded in an appropriate corporate risk register.

Recommendation 3: The specification of checks performed on the models must be brought to a common high standard

Owner: Model Developer
Timing: 12 months
Cost=Low, Value=Medium

For some components it was found that QA materials failed to describe checks at all. In others, checks were described in language that left room for interpretation. The description of checks must be explicit and unambiguous. The acid test is that two different checkers, on reading the description, would perform exactly the same actions.

One reason offered for checks being unspecified or vaguely described was to allow the checker to use their own initiative in choosing what to check. This avoids the risk of group-think where the checks fail to test something that their author has overlooked. This risk can be addressed in two ways:

- by having an independent person write the checks; and
- by having the checker or peer reviewer carry out additional tests of their choice, beyond those specified in the QA materials.

Recommendation 4: Results of checks must be recorded

Owner: Model Developer
Timing: 12 months
Cost=Low, Value=Medium

Checks are documented in a variety of places, including the QA Manual, the User Guide, and on dedicated sheets within the model itself (see Recommendation 5). There was at times no evidence that these checks had been performed. Records must be kept of checks that have been done (alongside the list that should be done) with details of who did the check and when; these records must include details of any failures, the fixes applied and the results of rechecking. Results of checks must be recorded to allow other reviewers to confirm that they have completed successfully.

Recommendation 5: Checks and results must be documented in a clear and consistent way

Owners and timings below
Cost=Low, Value=Medium

This recommendation is in four parts, each of which is covered in detail below.

Recommendation 5a: Checks must be described once, in a standard location for the model

Owner (5a): Model Developer
Timing: 12 months

Across the models under review, QC checks were described in a number of different places, including:

- in the QA/QC manual (the most common occurrence);
- on a dedicated checking tab within the spreadsheet (e.g. the ChecksAndStepsToUpdate tab in the Uncertainties model);
- within the model itself (e.g. mass balance or time series checks);
- in the VBA code (e.g. NAEI core database); and
- in the User Guide (e.g. Aviation).

Sometimes checks were recorded in multiple places for the same model (e.g. Road Transport Emissions) with no alignment between the different checks.

Checks must be described once, in a standard location for the model. For example, the NAEI may record QC checks in the QA/QC manual, whilst PCM may adopt dedicated checking tabs. Recording checks and results in a clear and consistent way makes it easier for them to be reviewed by project or audit staff, and supports staff transitioning between models.

Recommendation 5b: The record of completed checks must be kept in a standard location for the model.

Owner (5b): Model Developer
Timing: 12 months

Records of the completion of these checks were also kept in a variety of places, including:

- most commonly, in the spreadsheet itself, next to the calculations;
- on the QA tab for the spreadsheet;
- in a separate document (e.g. Melmod); and
- higher level checks are recorded at the NAEI database level.

The record of completed checks must be kept in a standard location for the model. For example, the NAEI may record QC results on the QA tab, whilst PCM may record them in a separate document.

Recommendation 5c: Modules must not be signed off until evidence is seen of all the checks having passed, or if not, good reason is given.

Owner (5c): Model Developer
Timing: 12 months

Recommendation 5d: A robust hierarchy of checks must be created so that each sign off confirms that lower level checks have been completed successfully.

Owner (5d): Model Sponsor (Defra or DECC) and Model Developer
Timing: 12 months

Ultimately the final sign off, for the use of data in a particular report, confirms that someone independent of the modeller has:

- reviewed the checks that were performed on the constituent parts of the model and the model in total.;
- confirmed the checks are unambiguous and appropriate (i.e. the model hasn't changed in a way that makes the checks incomplete or redundant); and
- seen evidence that checks have been performed and have either passed, or any failures have been satisfactorily explained.

Final sign off for a business-critical model's output must be performed by the model Senior Responsible Officer. Models developed by an external contractor should also be signed off by a senior executive of the contractor's organisation, with the contractors' internal audit function reviewing these checks and results on an occasional basis.

The collation of this information for all business-critical models will allow the Department to adopt two Macpherson recommendations fully:

(i) the model SRO to confirm that the model's QA process is "compliant and appropriate" (Macpherson recommendation 3); and

(ii) the Accounting Officer to report that "an appropriate QA Framework is in place and is used for all business-critical models" (Macpherson recommendation 4).

Furthermore, this approach obviates the need for further external audits of business-critical models. By providing these assurances, contractors are auditing themselves.

Recommendation 6: Model Documentation must be improved

Owners and timings below
Cost=Medium, Value=Medium

The scope of models is clearly defined in the ITT and technical documentation (NAEI and PCM) and in the Impact Assessment (IPM). The amount of model design documentation was found to be variable across the models and components. To some extent this is understandable given the variation in complexity of the models, from simple NAEI Mastersheets, which perform a series of multiplications of activity levels and emission factors, to databases such as the Road Transport Emissions database. In some instances design documentation had been produced when a member of staff was leaving. This gives a good indication that prior to this, the documentation was not sufficient to describe the model to a new team member or for handover to an alternative contractor.

This recommendation is in four parts, each of which is covered in detail below.

Recommendation 6a: Design documentation must be produced for more complex models and components as part of the three-year programme described in Recommendation 2**.**

Owner (6a): Model Developer
Timing (6a): three years

A complex model is defined as one that does more than straightforward arithmetical calculations. For all complex models and components, design documentation should describe:

- how to run the model (effectively a user guide);
- any algorithms or queries;
- the structure of the model. For a large model with multiple components, such as the NAEI, this documentation should show how the different components work and interact with each other;
- the structure of the data within it (e.g. using a Logical Data Model);
- any user interfaces; and
- how navigation works in the model (e.g. clicking on X takes you to sheet Y).

As well as supporting maintenance and handover of the model, this design documentation is needed for the development of any test specifications (see Recommendation 2), since it describes how the logic in the model should work.

For a model containing multiple components, such as the NAEI, the structure of documentation should be standardised across the various contractors as far as practicable. An example of a standard suite of documentation is:

- Scope document;
- QA Manual describing the approach to QA and listing specific checks; and
- combined User Guide and Design document, as described above.

For all new models, the design must be completed and documented before implementation takes place.

Recommendation 6b: The structure of the documentation must be mandated in future ITTs.

Owner (6b): Defra and DECC
Timing (6b): at contract renewal

This will ensure consistency of approach across contractors, and will address the possible conflict of interest whereby failure to produce documentation would make handover to an alternative contractor more difficult.

Recommendation 6c: Regular or ad-hoc requests for model outputs, for use in reports or other models, must be formalised in a specification.

Owner (6c): Model Developer, Defra and DECC
Timing (6c): immediate

This is largely in place for the provision of key data from external sources (using Data Supply Agreements to define the content, timescales, format and assumptions) and for the passing of data sets between component modules in the NAEI (where the Mastersheets specify the format required) but needs to be added for requests between PCM, NAEI and IPM, and for Defra and DECC requests for ad-hoc datasets. The specification must include all assumptions and data definitions (including units) and should state acceptance criteria, preferably in the form of tests that must be run before delivery. The specification must be archived with other project documents. QA/QC must be carried out to the same standards that apply for the development of the core model.

Recommendation 6d: *(applies to IPM only)*: A simple user guide / design document must be produced describing the process steps for creating an IPM model.

Owner (6d): Defra
Timing (6d): 12 months

For each processing step, a check must be described, and space must be provided for sign-off that the check has been completed.

<u>Recommendation 7: Robust version and change control must be introduced</u>

Owners and timings below
Cost=Medium, Value=Medium

This recommendation is in five parts, each of which is covered in detail below.

Recommendation 7a: Version and change control must be provided for templates

Owner (7a): Model Developer
Timing: three years

All templates must include a cover sheet listing all versions and recording version number, date of release, owner, changes since the previous version, and, optionally, details of checks performed, by whom and dates (see Recommendation 5b).

The template name and version number constitute a unique identifier.

Recommendation 7b: Version and change control must be provided for runs of component models

Owner (7b): Model Developer
Timing: three years

A component model is a collection of one or more Excel, Access database or software templates which operate together to deliver part of a larger model. For example, the Aviation model within the NAEI contains a database and multiple spreadsheets, ultimately delivering emissions and fuel use from civil and military aviation. A run of a component model must be given a unique version identifier (e.g. Aviation-2012), with the following documented within the version:

- uniquely identified source data (referring to file names, versions, years etc); and
- version numbers of all templates used.

Recommendation 7c: Version and change control must be provided for runs of full models

Owner (7c): Model Developer
Timing: three years

A full model run (e.g. for the NAEI) will comprise data taken from multiple component models. This full model run must also have a unique identifier. This could be a year (e.g. NAEI-2012), a version (e.g. PCM v 6.2) or the name of an assessment (e.g. Impact Assessment 1010). This identifier allows the outputs to be described unambiguously in subsequent reports. For a particular model run, the contents of the run must be recorded using the uniquely identified components described in the recommendations above. This will allow the exact model to be re-run in the future and avoid any confusion or ambiguity over the meaning of the outputs.

Recommendation 7d: Version and change control must be provided for reports

Owner (7d): Defra and DECC
Timing: three years

Within Defra and DECC, modelling results used in reports must include explicit references to the unique identifier described above. This will avoid confusion over the provenance of particular figures, and in the event of an investigation will allow a drill-down into the exact source data and models used.

Recommendation 7e: A process must be created to ensure version and change control and sign-offs

Owner (7e): Model Developer
Timing: 12 months

Across the models under review it was generally possible to make changes to models and templates at any time, irrespective of their sign-off status. We saw one example of a model being changed after sign-off, with no evidence of a second sign-off. For business critical models, contractors and model developers within Defra and DECC must introduce processes or technology solutions to ensure that a change to a model or template triggers the completion of change control information and, if necessary, approvals and sign-offs.

Recommendation 8: Ongoing supervision of QA/QC activities must be introduced

Owners and timings below
Cost=Medium, Value=High

Ricardo-AEA has been subject to a number of third party audits and accreditations, including ISO 9001. These accreditations are not sufficient to ensure that QA/QC activities are in place for models.

This recommendation is in four parts, each of which is covered in detail below.

Recommendation 8a: Explicit QA/QC terms must be included in contracts

Owner (8a): Defra and DECC
Timing: at contract renewal

Contracts for model development must be explicit about the amount and nature of QA/QC activities to be performed. QA/QC activities come with a cost, so consideration should be given to what is appropriate, necessary and delivers value for money. For example, adherence to the DECC QA/QC guidelines could be mandated, or alternatively a light touch “best efforts” approach could be adopted. If the business criticality of the model demands robust QA/QC activities, but time or cost constraints prevent this happening, this must be explicitly acknowledged and reported within the Department. Contracts should also state that the handover of a model to a new contractor must include the transfer of QA plans and completed checklists so that the new contractor has evidence that checks have been carried out.

Recommendation 8b: Contractors’ internal audit functions must be responsible for ensuring that models are following their own published QA/QC procedures.

Owner (8b): Model Developer
Timing: immediate

Not every component needs to be checked every year, however over a period of five years the complete model must be reviewed.

Recommendation 8c: Contractor QA/QC activities must be made visible

Owner (8c): Defra and DECC
Timing: at contract renewal

For models developed by third-party developers, Defra and DECC must introduce measures via their contracts to get visibility of the QA/QC activities being performed by contractors. These measures do not need to be expensive or time-consuming; for example, a good approach would be a one-day review of the contractors’ audit plan, asking for evidence of auditing activity.

Recommendation 8d: A decision must be made on inherited models missing QA documentation

Owner (8d): Defra and DECC
Timing: immediate

For models that have passed from one contractor to another and which have no QA documentation, Defra and DECC need to make a decision whether to commission the quality assurance of the model, or accept the risk until the model is rebuilt.

Recommendation 9: Full evidence of QA/QC plans and checks must be collated and reviewed

Owners and timings below
Cost=Low, Value=Medium
Recommendation applies to the NAEI and other multi-contractor models

Examination of the NAEI ITT and the CEH tender for the Land Use, Land-Use Change and Forestry (LULUCF) sector reveals that there is no obligation for the developers of models "upstream" in the supply chain, such as CEH, Forest Research and Rothamsted Research, to provide QA plans or evidence of completed checks to Ricardo-AEA. The fact that CEH contracts directly with DECC for their part in the inventory compilation (with Forest Research subcontracting to CEH) and Rothamsted Research with Defra is significant; if instead they subcontracted to Ricardo-AEA, then Ricardo-AEA would be able to demand this information through their subcontract. The absence of these flows of QA plans and evidence leads to a breakdown of the overall QA/QC of the NAEI and the inability for the Inventory Agency to confidently assure the quality of NAEI deliverables.

This recommendation is in three parts, each of which is covered in detail below.

Recommendation 9a (short term): As part of their annual inventory compilation process, the Inventory Agency must ask all model owners for their QA plans and for evidence of completed checks.

Owner (9a): Inventory Agency
Timing: 12 months

These checks should be reviewed and insufficient QA/QC activities should be challenged. It seems likely that CEH, Forest Research, Rothamsted Research and other contractors will comply, despite the absence of a contractual requirement to do so.

Recommendation 9b (longer term): Clauses must be added to contracts with CEH, Forest Research, Rothamsted Research and other direct contractors to require the provision of QA plans and completed QA checklists to the Inventory Agency.

Owner (9b): Defra and DECC
Timing: at contract renewal

Recommendation 9c: Defra must consider extending the role of contractors to cover the review of QA plans and evidence of checks for key data suppliers.

Owner (9c): Defra
Timing: immediate

This will be limited to data suppliers under contract to Defra and DECC, such as the DDU suppliers, and could be implemented as contracts are renewed.

Recommendation 10: Departments must have emergency access to the model and associated documentation

Owner: Model Developer
Cost=Low, Value=Medium

Contractors are responsible for archiving each annual run of the NAEI and PCM. To ensure business continuity, it is important that Defra and DECC should be able to access these model runs, including input data sets, the underlying templates and all design and user documentation. Contractors must assure Defra and DECC that these archive copies would be made available under an arrangement equivalent to escrow to enable emergency access to these business critical models.

APPENDIX 1: MODEL QA AUDIT TEMPLATE

1. Introductions

2. Background
Who built the model?
Who runs the model to produce outputs?
Who uses the model outputs in their decision making?
What decisions do they make?
Who owns the IP?
What is your contractual relationship with the different users listed above (same org / contractor / sub-contractor)?

3. Modelling Process
What is your place in the model development process (base data provider / sub-model developer / compiler of all models etc)?
What processes are imposed on you by UN/EC/anyone else?
When does the model need to be re-run each year?
What changes year to year (process/model/data)?
Which parts of the process, model or source data are most risky?

4. Scope of QA
What is the scope of your QA responsibilities?
What do you assume others do?
How do you interact with others over QA?

5. QA Activities
Review of the QA Plan provided in advance, plus selected evidence.

6. Data & Assumptions
How do you confirm that the input data is complete, up to date and error free?
How do you know there were no errors in the transfer of the input data?
How do you know you have the correct version?
How would you know if the input data changed after it had been transferred?
Are you aware of / responsible for the QA of the input data?
How do you expose and agree assumptions and limitations with customers?
How do you record assumptions in the source data and track changes to these?

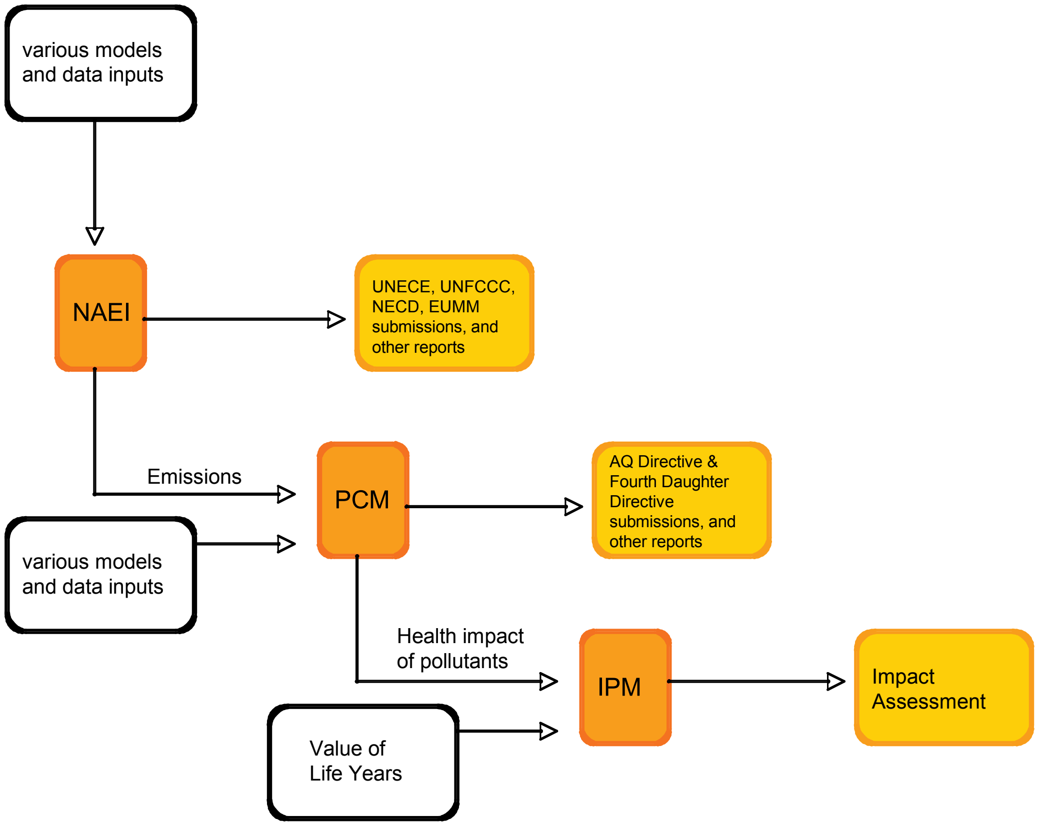
7. Error Tracking
How do you deal with issues or errors arising from QA activities?
Who gets to know about these errors, and how is this escalated to senior management and Defra/DECC?

8. Continuity
Do you have a staff continuity plan?

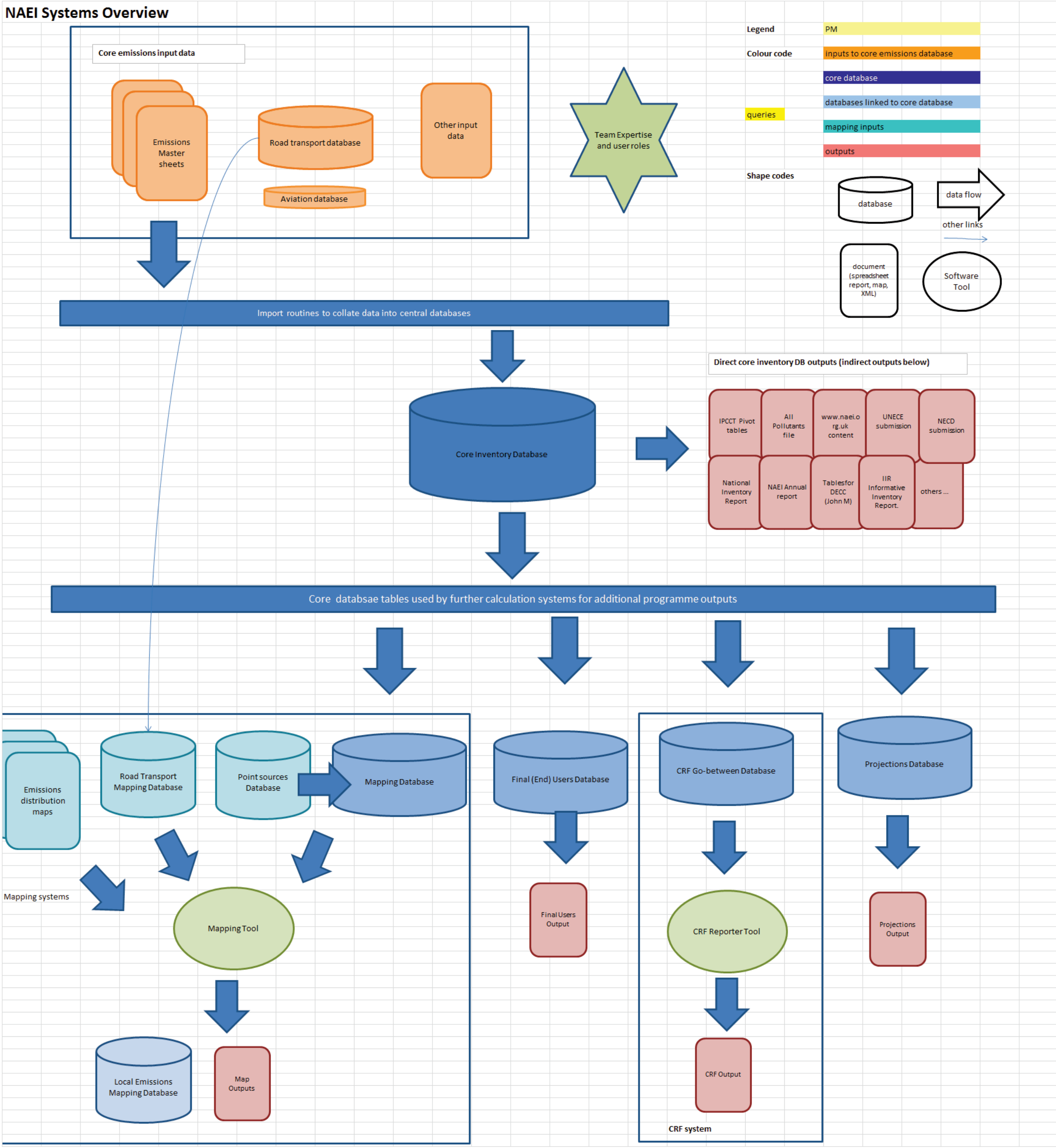
9. Continuous improvement
What do you do to continually improve the quality of the model?
What do you do to continually improve your quality processes?

10. Self assessment
What do you consider the strengths of your QA processes?
What are the weaknesses?
What are you plans for improvement?
If you had unlimited time/money/capability what changes would you make to ensure the quality of the model?
Is there anything else you want to mention?

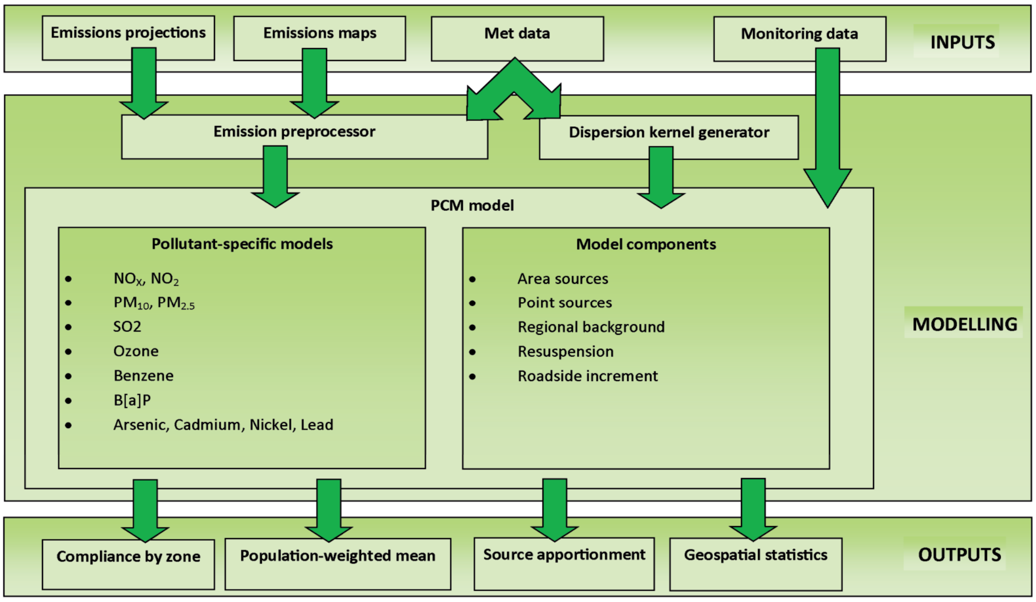
APPENDIX 2: MODEL MAPS



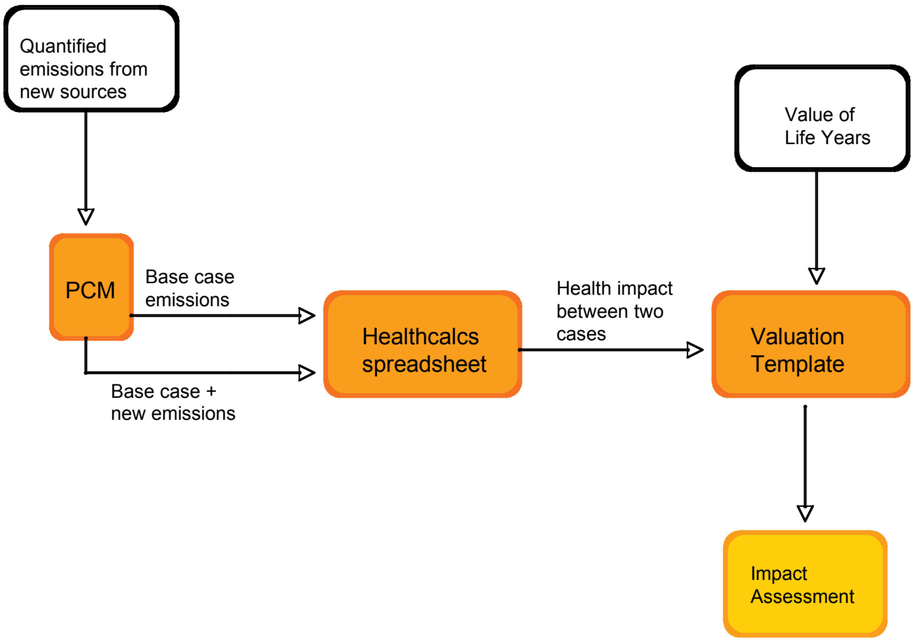
Interactions between NAEI, PCM and IPM
Source: Hartley McMaster analysis



Overview of NAEI
Source: Ricardo-AEA



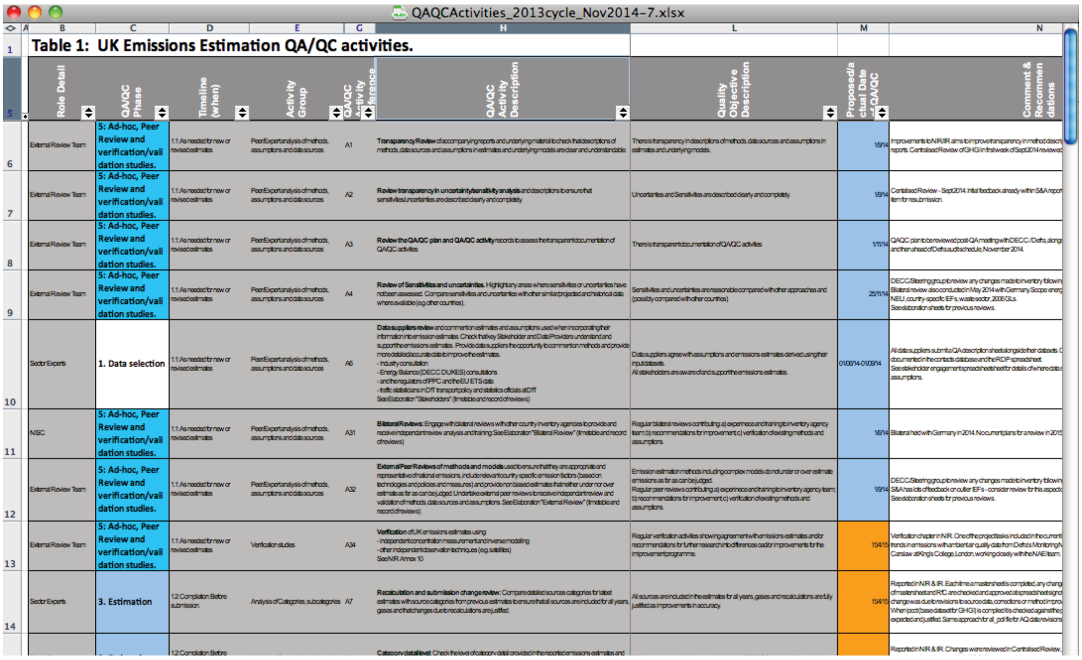
Overview of PCM
Source: Ricardo-AEA



Overview of IPM
Source: Hartley McMaster analysis

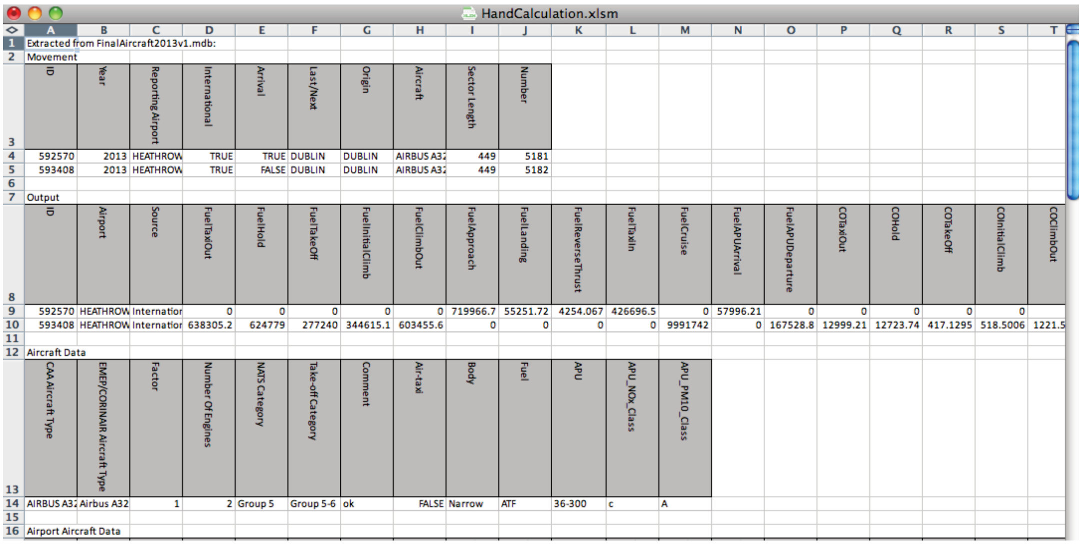
APPENDIX 3: SAMPLES OF EVIDENCE

3.1 NAEI QA/QC PLAN (EXTRACT)



Source: Ricardo-AEA

3.2 NAEI AVIATION MODEL HAND CALCULATION (EXTRACT)


HandCalculation.xlsm

Extracted from FinalAircraft2013v1.mdb:

Movement

ID	Year	Reporting Airport	International	Arrival	Last/Next	Origin	Aircraft	Sector Length	Number
592570	2013	HEATHROW	TRUE	TRUE	DUBLIN	DUBLIN	AIRBUS A32	449	5181
593408	2013	HEATHROW	TRUE	FALSE	DUBLIN	DUBLIN	AIRBUS A32	449	5182

Output

ID	Airport	Source	FuelTaxiOut	FuelHold	FuelTakeOff	FuelInitialClimb	FuelClimbOut	FuelApproach	FuelLanding	FuelReverseThrust	FuelTaxiIn	FuelCruise	FuelAPUArrival	FuelAPUDeparture	COTaxiOut	COHold	COTakeOff	COInitialClimb	COClimbOut
592570	HEATHROW	Internatior	0	0	0	0	0	719966.7	55251.72	4254.067	426696.5	0	57996.21	0	0	0	0	0	
593408	HEATHROW	Internatior	638305.2	624779	277240	344615.1	603455.6	0	0	0	0	9991742	0	167528.8	12999.21	12723.74	417.1295	518.5006	1221.5

Aircraft Data

CAA Aircraft Type	EMEP/CORINAIR Aircraft Type	Factor	Number Of Engines	NATS Category	Take-off Category	Comment	Air-taxi	Body	Fuel	APU	APU_NOx_Class	APU_PM10_Class
AIRBUS A32	Airbus A32	1	2	Group 5	Group 5-6	ok	FALSE	Narrow	ATF	36-300	c	A

Airport Aircraft Data

Source: Ricardo-AEA

3.3 NAEI AVIATION MODEL CODE WALKTHROUGH (EXTRACT)

FinalAircraft2013v1_Module1.bas

```
Option Compare Database
Option Explicit

Dim Stage_Length(16) As Integer

Dim RecordSetCruise0 As DAO.Recordset
Dim RecordSetCruise1 As DAO.Recordset

Function MakeOutput()

    Stage_Length(1) = 125
    Stage_Length(2) = 250
    Stage_Length(3) = 500
    Stage_Length(4) = 750
    Stage_Length(5) = 1000
    Stage_Length(6) = 1500
    Stage_Length(7) = 2000
    Stage_Length(8) = 2500
    Stage_Length(9) = 3000
    Stage_Length(10) = 3500
    Stage_Length(11) = 4000
    Stage_Length(12) = 4500
    Stage_Length(13) = 5000
    Stage_Length(14) = 5500
    Stage_Length(15) = 6000
    Stage_Length(16) = 6500

    Call MakeOutput2("Movement", "Output")
'   Call MakeOutput2("Movement 2013", "Output")

    Call MakeOutput2("Movement OT", "Output OT")
'   Call MakeOutput2("Movement OT 2013", "Output OT")

'   QA
'   Call MakeOutput2("Movement Extract", "Output")

End Function

Function MakeOutput2(strMovement As String, strOutput As String)

'   Correction Factors for default aircraft /engine
    Const NOx_fa_cf_app = 1.06146
    Const NOx_fa_cf_roll_100 = 0.9151593
    Const NOx_fa_cf_roll_85 = 0.9151593
    Const NOx_fa_cf_to_100 = 1.098785
    Const NOx_fa_cf_to_85 = 1.098785
    Const NOx_fa_cf_co = 1.136376

'   Movement
    Dim MyRecordSet As DAO.Recordset
    Set MyRecordSet = CurrentDb.OpenRecordset(strMovement)
    Dim Aircraft As String
    Dim Arrival As Boolean
    Dim International As Boolean
    Dim Number As Long
    Dim Reporting_Airport As String
    Dim Last_Next As String
    Dim Year As Integer
    Dim Sector_Length As Double

    Dim Year2 As Integer
    Dim Year4 As Integer

    SysCmd acSysCmdClearStatus

    Dim count As Long
    count = 0

    Dim RecordSet1 As DAO.Recordset
    Dim APU As String
    Dim APU_NOx_Class As String
    Dim APU_PM10_Class As String
    Dim Body As String
    Dim EMEP_CORINAIR_Aircraft_Type As String
    Dim NATS_Category  As String
    Dim Factor As Double
    Dim number_of_engines As Integer

    Dim RecordSet2 As DAO.Recordset
    Dim idle_fuel_flow_factor As Double
    Dim Fuel_7 As Double
    Dim Fuel_30 As Double
    Dim Fuel_85 As Double
    Dim Fuel_100 As Double

    Dim CO_7 As Double
    Dim CO_30 As Double
    Dim CO_85 As Double
    Dim CO_100 As Double
    Dim HC_7 As Double
```

NAEI Aviation database
Code review
Martin Peirce, Charles Walker
12/9/14

Missing executable lines amongst a lot of Dim lines means they get lost

Unclear variable names

Page 1 of 19

Source: Ricardo-AEA

3.4 PCM Checking Log for NO_x and NO_2 (extract)

00_MAAQ_checks_noxno2_2013.xlsx

	A	B	C	D	E	
1	Title	NOx and NO2 base year				
2	Date	22/7/14				
3	Originator	Sally Cooke				
4	Checker	John Stedman				
5				Blank cells represent no issues identified		
6	Item	Check	Status	Comments	Resolution	File location
7	Scripts	Originated by projections spreadsheet	JRS 22/07/2014			W:\aqd2013\projections\projections_na
8		Annotation updated and correct	JRS 22/07/2014			
9	Area source grids	Output file of check against NAEI year grids ge	JRS 22/07/2014			W:\aqd2013\nox\A1_nox_1c_checktestg
10	Variables	Updated	JRS 22/07/2014	a few unused variables to be deleted	Fixed	
11	Ancillary grids, coverages	Updated to use latest	JRS 22/07/2014			W:\master_utilities\basemaps\1_versio
12	Input data	Review cell referencing	JRS 22/07/2014			W:\aqd2013\nox\A_mapcalcs_nox2013.
13	Sites omitted	check completeness	JRS 22/07/2014			W:\aqd2013\nox\A_mapcalcs_nox2013.
14		review justification	JRS 22/07/2014			
15	BG calibration	Correctly calculated	JRS 22/07/2014			W:\aqd2013\nox\A_mapcalcs_nox2013.
16		Added in script	JRS 22/07/2014			
17	RS calibration	Correctly calculated	JRS 22/07/2014			W:\aqd2013\nox\A_mapcalcs_nox2013.
18		Added in script	JRS 22/07/2014	NB nox2013em has been saved to dbf as in	Added to Things to Do Differently list to be	
19	Source apportionment	Check looks sensible	JRS 22/07/2014			W:\aqd2013\nox\A_mapcalcs_nox2013.
20	Background modelling (mapc	Check spreadsheet concentrations at sites m	JRS 22/07/2014			W:\aqd2013\nox\A_mapcalcs_nox2013.
21	Roadside modelling (mapcal	Check spreadsheet concentrations at sites m	JRS 22/07/2014			
22	Model performance	DQO checks updated	JRS 22/07/2014			
23		Verification plots updated	JRS 22/07/2014			
24		Summary stats updated	JRS 22/07/2014	have to updated cell references for veri site	Fixed	
25	Zaggloms	Outputs referenced in spreadsheet	JRS 22/07/2014			W:\aqd2013\nox\1_summaryresults_no
26		All summaries updated	JRS 22/07/2014			
27		Excluded roads removed	JRS 22/07/2014			
28	Results check	Update comparison with previous year	JRS 22/07/2014	updated column J on sheet background to l	Fixed	W:\aqd2013\nox\A_checkingcomparisor
29	Maps	updated	JRS 23/07/2014			
30		labels updated	JRS 23/07/2014			
31	Spreadsheet checks (general	Cell ranges have been checked and are up to	checked			W:\aqd2013\nox\A_mapcalcs_nox2013.
32		Cell logic has been checked	checked			W:\aqd2013\nox\A_mapcalcs_nox2013.
33		Calculations have been reality checked	checked			W:\aqd2013\nox\A_mapcalcs_nox2013.
34						
35						
36	Pollutant-specific checks					
37	Rural NOx map	Correction for local sources	JRS 22/07/2014	see also checking recod on QA sheet		W:\aqd2013\nox\1_rural\1_ruralnoxcalc
38		Interpolation, check map looks sensible	JRS 22/07/2014	suggest update classes in png file so that y	Added to Things to Do Differently list to be	W:\aqd2013\nox\1_rural\
39		Aml scripts to generate rural NOx grid	JRS 22/07/2014	should have included ets points in correcti	Added to Things to Do Differently list to be	W:\aqd2013\nox\1_rural\
40	Primary NO2	Correct f-NO2 spreadsheet data used	JRS 22/07/2014			W:\aqd2013\nox\
41	Oxidant partitioning model	background calibration (measured data)	JRS 22/07/2014			W:\aqd2013\nox\1_oxidantpartitioningm
42		roadside calibration (measured data)	JRS 22/07/2014			W:\aqd2013\nox\1_oxidantpartitioningm
43		background verification (modelled data)	JRS 22/07/2014	updated Nox verification stats on summary	Fixed	W:\aqd2013\nox\1_oxidantpartitioningm
44		roadside verification (modelled data)	JRS 22/07/2014			W:\aqd2013\nox\1_oxidantpartitioningm
45		Correct coefficients in scripts	JRS 22/07/2014			W:\aqd2013\nox\
46	Hourly graph	Check selection of sites is concistent with modelling, data links are correct and graph looks sensible	JRS 24/07/2014			W:\aqd2013\nox\A_mapcalcs_nox2013.

Source: Ricardo-AEA

3.5 Impact Assessment Peer Review Form (extract)

PEER REVIEW OF IAs

The aim of the peer review of a full IA is to enable the Chief Economist to advise ministers, for the purpose of a consultation, that they can sign-off the statement in an impact assessment that:

'It represents a fair and reasonable view of the expected, costs, benefits an impact of the policy'
[The CE does not state that the benefits justify the costs as this is a judgement made by ministers]

Issues to consider when undertaking an IA peer review.

*Please indicate yes / no by placing **X** in the appropriate box, further comments are optional but please make them if you want to draw attention to areas of strengths or weaknesses.*

ASSUMPTIONS

Are the assumptions made:

	YES	NO
Clearly identified in all cases (could the peer reviewer list all the key assumptions if required)?	X	
Reasonable, credible, consistent and explained?	X	
Using appropriate ranges where there is uncertainty?	X	
Further Information:		

EVIDENCE BASE

	YES	NO
Where there are gaps in data, research or analysis, are these clearly identified?	X	
Is there a work plan in train to fill any such gaps? Is any evidence being sought through consultation?	X	
Has the impact assessment been developed as part of the policy cycle using a sound policy development model (eg. http://intranet/regulation-division/toolkit/frontpage.asp).	X	
Have any underlying models, whether they be scientific, econometric or spreadsheet models, been audited or otherwise peer reviewed.		X

Further Information:

The modelling is currently being peer reviewed in detail but the overall approach has been accepted as reasonable.

ANALYSIS

	YES	NO
Is the analysis proportionate, taking into account the likely magnitude of costs and benefits and sensitivity (i.e. has enough work been put in to justify significant change in policy)?	X	
Have the most appropriate techniques/mode of analysis been used?	X	
Has the HMT Green Book and other relevant guide-lines been useed appropriately and consistently?	X	
Has qualitative evidence been taken into consideration and is it clearly explained?	X	
Have costs and benefits been monetised, wherever this is feasible, and if not, quantified or discussed qualitatively?	X	
Have risk and uncertainty been fully taken into account when developing options?	X	
Further Information:		

PRESENTATION

	YES	NO
Are the figures and tables, clear, comprehensive and support the inferences drawn from them?	X	
Have all sources of data and research been clearly identified?	X	
Has the IA met its aim of explaining the case to its target audience?	X	

Further Information:

I believe that the presentation is suitable to this particular target audience. However in the next stage taking this to the public will require some tailoring.

Source: Defra

APPENDIX 4: LIST OF NAEI MASTERSHEETS

accidental fires	fuel transformation	rail
adipic	gas	refine
aerosol	gas oil	refineries
agriculture - PM	gasmain	primary aluminium - pah
aircraft	GHG - agriculture	road transport
base cation factors	GHG - LUCF	SCCP
black smoke	glass	secondary aluminium
bricks & ceramics	glue	secondary lead
combustion	halide factors	SO2 factors
carbon factors	Hg in products	SSF
cement	agricultural waste burning	sulphur - coal
chemical waste incineration	ink	sulphur - oils
chemicals - other	iron & steel	VOC factors
chemicals - PM & metals	lime	waste burning
chemicals - voc	metal factors	wood
cleaning	methane factors	shipping
CO factors	MSW incineration	bonfire night
coal	N2O factors	stored carbon
coating	nfm - PM & metals	peat
coating manufacture	NH3 factors	pbde
coke & ssf	nitric	OT and CD
coke ovens	NOx factors	Road transport fuel
construction	off-road	OTGasOil
consumer	offshore	InlandWaterways
copper alloys & semis	other incineration	Charcoal
crematoria	other oil	Biological waste treatment
dioxins	other p & s	IPCC Reference Approach_2013_EUMM
drink	other PI - VOC	Wastewater
electric arc furnaces	PAH	Mining
electricity	paint	Landfill
f gases	Part B - PM	petrochem
feedstock	PCB	agriculture - VOC
field	pesticides	PM2.5
fireworks	petrocoke	Anaesthesia - N2O
food	petrol	ipcc_GCV_&_conversion_table_2013
foot & mouth	power stations	gibraltar
foundry	primary aluminium	
fuel oil	quarrying	

Source: Ricardo-AEA