Final report for the Office of Communications (Ofcom)

WLR/LLU Charge Control 2017 Quality of Service model assessment

13 September 2017

James Allen, Mark Colville, Charles Khe

Ref: 2008669-373

www.analysysmason.com

Contents

Analysys Mason Limited
St Giles Court
24 Castle Street
Cambridge CB3 0AJ
UK
Tel: +44 (0)1223 460600
cambridge@analysysmason.com
www.analysysmason.com
Registered in England No. 5177472

1 Executive summary

1.1 Introduction

In order to set wholesale charges for Openreach products at an appropriate level, Ofcom needs to understand how changes to the quality of service (QoS) provided affect the efficient costs to Openreach of providing the relevant products. For this purpose, Ofcom and Analysys Mason jointly developed a model to help quantify the relationship between QoS targets, performance levels against those targets, and the resource levels needed to achieve them.[1] Ofcom relied upon outputs from this model, which we refer to in this report as the Ofcom resource performance model (or 'Ofcom RPM'), to inform the position set out in its consultation.[2]

Since the publication of Ofcom's consultation, Openreach has provided Ofcom with a copy of its own model, developed in collaboration with EY. This model, which we refer to as the 'Openreach model' and which was not available to Ofcom prior to the launch of the consultation, is considered by Openreach to be a superior tool for understanding the relationship between QoS and resource requirements. Openreach has relied upon outputs from this model, including a comparison to the Ofcom RPM in its response to Ofcom's consultation.

Analysys Mason has been commissioned by Ofcom to carry out an assessment of the Openreach model to support a further public consultation.

This document is our final report, covering our assessment of the model. It is not a standalone description of the model or a user guide and so should be read in conjunction with Ofcom's consultation document and with a description of the model methodology (which constituted a part of Openreach's response to Ofcom's previous consultation).

1.2 High-level description of the Openreach model

The Openreach model is an allocation model that simulates the field activities of Openreach's Service Delivery team and the impact of various supply- and demand-side levers on meeting job targets. The Openreach model simulates job arrivals and executions using discrete event simulation, where steps in the job servicing process all occur at discretised points in time. The model is run in a SimPy environment in Python 3, with input and output data read from and written into an Oracle database.

Job characteristics (e.g. task times, job visit rates) are assigned to individual jobs via random sampling from probability distributions generated from historical Openreach data. Jobs then enter

1 Ofcom initially developed a high-level model, which Analysys Mason reviewed and developed to increase its level of usability and functionality in a number of areas. The scope of our assignment also included producing outputs from the model for a range of scenarios of interest to Ofcom.

2 Ofcom, Quality of Service for WLR, MPF and GEA: consultation on proposed quality of services remedies, 31 March 2017. See https://www.ofcom.org.uk/consultations-and-statements/category-1/quality-of-service

the model by entering a queue that is ordered by the priority of the job, as determined by a job priority scoring framework. Jobs are then allocated to engineers, based on the minimum skill level required to complete the job, who then execute the job. The model collates job completion metrics on a weekly basis, which are batched together for the time frame of the simulation and provided as a model output.

The resources are expressed in full-time equivalents (FTE) terms and each level of resource will result in a certain level of performance against provision, care level 1 (CL1) and care level 2 (CL2) repair targets. These raw model outputs are then processed in various ways to calculate 'deltas'. These deltas are the percentage difference (delta) between the resources required over a specified period of time (e.g. one year) to deliver a specific level of performance compared to the resources required to deliver a (lower) *baseline* level of performance.

1.3 Findings and conclusions

We have carried out a detailed review of the Openreach model, including the input data, the code and the model outputs. Below we set out our key findings and present our conclusions on the extent to which the Openreach model is likely to be a reasonable approach to estimating the outputs of interest to Ofcom.

Methodology

The Openreach model takes a different and much more detailed approach to modelling the relationship between QoS and required resources compared to the Ofcom RPM. Similarly to the Ofcom RPM, some methodological issues need to be taken into account when considering the Openreach model.

There are a variety of areas where the Openreach model is more detailed than the Ofcom RPM including in relation to job durations, staff level fluctuations, staff skill levels, consideration of overtime, glass-ceiling[3] inputs and taking multiple visits per job into account. However, in some cases the effect on outputs is likely to be limited, whilst in others there are implementation complexities that make it unclear whether the extra detail is beneficial. The approach used to adjust the model outputs using the post-processing script in order to reflect staff overtime is a prime example.

At the same time, there are several areas where the Openreach model is less capable (or configured less capably) than the Ofcom RPM:

- Job-handling limitations, primarily relating to the handling of 'stress',[4] mean that the Openreach model may not adequately prioritise jobs to achieve improved performance. Indeed, the part of

---

3 I.e. jobs which will not be able to be completed within the required time frame to meet service level agreements independent of how many resources are available (e.g. because specialist equipment or civil works are required).

4 The stress factor is a measure of the shortfall or surplus in resource relative to the projected amount of resource required to complete jobs in the current workstack. The model can apply a different mode of operation in response to

the model that takes stress into account is turned off in the majority of scenarios reported by Openreach.

- Resource-handling limitations, primarily relating to the handling of each SOM[5] in isolation mean that 'loans'[6], which may improve performance, are not taken into account.
- There is no explicit mechanism within the Openreach model, beyond the basic prioritisation rules, to balance performance across different job types, meaning that resources required to achieve performance targets across all job types may be overestimated.
- Model configuration limitations mean that it is hard to guarantee that errors are not introduced in the running of the model.

Implementation and usability

The Openreach model is, broadly speaking, well constructed.

There is a bug which crashes a small percentage of runs. This somewhat reduces our confidence in the correctness of the design.

The model documentation is not sufficiently detailed, and there is no user interface (other than the development tools for Python and the chosen SQL database), making it difficult to install, configure and run. It is complex, slow and produces very large amounts of data (most of which is used once in post-processing before being discarded). It is also, in certain cases, unclear[7] how the actual input data reflects our understanding of the input parameters. For example, the values of the important glass-ceiling inputs in the model are not clearly linked to the "actual" or "raised" values in Openreach's consultation response (in fact, the "category distribution" ratios are almost identical in scenarios described by Openreach as having either "actual" or "raised" glass ceilings).

It is unlikely that Ofcom staff could make effective use of the model in its current form, if only due to the complexity, the relatively minimal documentation, and the time it takes to understand and configure the model and produce results, making sensitivity analysis very slow.

Outputs

We have been unable to reproduce in detail the outputs of the runs Openreach has performed, even starting from almost identical inputs. This may be due in part to the fact that the model will (due to

periods of high stress in relation to factors such as job prioritisation, shrinkage, customer appointment uptake and the clearing of tail jobs.

5 Areas in the Openreach network are divided between Senior Operations Managers, with each Senior Operations Manager having responsibility for Service Delivery within that area.

6 Loans is the term we use to denote Openreach engineers being appointed to jobs outside of their home SOM, in order to provide assistance in other SOMs where performance levels are struggling to meet targets. These are modelled as part of the Ofcom RPM, but are not taken into account by the Openreach model.

7 This is not to say that the implementation is necessarily incorrect.

the way in which it uses random numbers) give slightly different results on different detailed versions of Python, even if the random number seed and the PYTHONHASHSEED[8] are controlled.

Model runs with quite similar input sets, differing only in one or two input parameters (potentially with only relatively small changes in those inputs), can give quite widely varying results (in terms of the resource deltas to achieve a certain target performance level), especially if the target performance is close to the glass ceiling. Particular issues arise where the target performance level is not met for some job types (e.g. CL2). Figure 1.1 provides a summary of the results obtained from running four of the key high-level scenarios (HLS).[9]

Figure 1.1: Summary of results obtained from key High Level Scenarios (HLS) [Source: Analysys Mason (using Openreach model), 2017]

HLS	Key parameters defining scenario	Calculated delta for 90% performance (from our model runs, relative to postprocessing script baseline)	Target performance level actually achieved?
77	Actual 16/17 care level (CL) mix, Actual 15/16 glass ceiling (87.5%), 83% UG/BBUG skilling	38.6% (average over 2 runs)	No, neither run hits target, reaching [✂] and [✂] for CL2 in the worst GMs
112	Actual 16/17 CL mix, Openreach Raised glass ceiling (92.6%), 83% UG/BBUG skilling	11.0%	Yes
200	Actual 15/16 CL mix, Actual 15/16 glass ceiling (87.5%), 61% UG/BBUG skilling (Actual 15/16)	10.1%	No, [✂] CL1
203	Actual 16/17 CL mix, 'Alternative' Raised glass ceiling (i.e. 93.29%, due to 6.71% 'red' failures), 83% UG/BBUG skilling	18.4%	Close ([✂] CL2)

These results can be hard to interpret because of these significant variations in delta from apparently small changes in inputs. Some results are particularly difficult to explain (e.g. why HLS 203 has higher deltas than the otherwise similar HLS 213-17, described in Section 7.3). Very high deltas arise in some scenarios (e.g. HLS 77), which may be due to the glass ceiling being below the targeted performance (which means the target performance will not be achieved in all GM[10] areas).

This does not necessarily mean that the results are incorrect or that the model is unreliable, but rather that the sometimes counter-intuitive results and the difficulty of replicating outputs produced by Openreach means that Ofcom needs to treat the results with an appropriate level of caution.

8 A parameter within Python that determines how iteration over data structures known as dictionaries is carried out.

9 Note that more detailed scenario descriptions are provided in the equivalent table in Figure 8.1.

10 SOM areas are grouped together under the supervision of a General Manager. These groupings of SOM areas are referred to as GM areas, or GMs

Key sensitivities

Openreach has produced an analysis of why it thinks the two models differ in their calculation of resource deltas. This is presented in the form of a waterfall chart. However, this analysis is flawed in several ways.

Although in relation to 'minor fails',[11] the Ofcom RPM modelling does not actually function as Openreach seems to assume (due to the delay in reattempting failed jobs in the RPM), a revised approach to the same waterfall analysis does still show that the modelling of minor fails is important. However, it does not show factors other than minor fails and the glass ceiling (i.e. 'major fails' in the Ofcom RPM) to be causing material changes in the overall delta to reach 90% performance.

Furthermore, our analysis of the Ofcom RPM shows that the impact of loans, which are not considered within the Openreach model, is a third significant factor in determining the size of the deltas.

The use of steps in the modelled resource levels within the Openreach model means that deltas estimated using that model may have an additional variability of the order of 0.5%.

Summary

The Openreach model takes a different approach to the Ofcom RPM and can be considered complementary to it in some ways.

In our view, results from the Openreach model are likely to overestimate deltas for several reasons. In particular:

- The lack of balancing of resources across different job types (provision, CL1, CL2) means that CL1 often appears likely to more than meet its targets at a given level of resource uplift, which suggests that a different use of resources could hit the required CL2 performance at a lower level of total resource.
- The approach to appointment scheduling or job prioritisation in response to stress, because all response to stress is generally turned off in most scenarios
- Loans, according to need and resource availability, are not taken into account in the Openreach model.

On the other hand, it is possible that the overtime adjustments are leading to too low a delta.

It is useful to be able to test deltas coming from the Openreach model by comparing them to values produced by a similar but independent system, the Ofcom RPM. This is particularly the case when the Openreach model system is complex, slow and difficult to configure, which means that ad-hoc tests of specific parameters of interest are not easily conducted.

[11] I.e. jobs which fail at the first attempt, but which can potentially be re-attempted and completed within the required time frame.

The much simpler process of configuration (fewer input parameters) and the higher speed of running the Ofcom RPM mean that a variety of tests can be designed, run and analysed relatively quickly.

Some effects are dealt with naturally in one model; others are more naturally considered in the other. Loans, for example, are dealt with more naturally by the Ofcom RPM, whereas the impact of GM-to-GM[12] variation in glass ceiling (major fails) are dealt with more naturally within the Openreach model.

As noted above, the outputs (delta to reach 90% performance) we have obtained from the Openreach model vary quite widely, between 10.1% and 38.6% for apparently similar scenarios (see Figure 10.1). By way of comparison, the Ofcom RPM produces deltas ranging between 6.5% and 20.1%, depending on the assumption used for major fails (i.e. the level of the glass ceiling).

The Openreach model is difficult to use and generates some unexpected results. It is therefore unlikely that the model, as it currently stands, could be used in isolation to predict the required resource deltas. The Openreach model may help Ofcom, potentially in conjunction with the Ofcom RPM, to inform a range of potential resource uplifts that may be required to realise a given QoS increase.

12 SOM areas are grouped together under the supervision of a General Manager. These groupings of SOM areas are referred to as GM areas, or GMs

2 Introduction

In order to set wholesale charges for Openreach products at an appropriate level, Ofcom needs to understand how changes to the quality of service (QoS) provided affect the efficient costs to Openreach of providing the relevant products. In other words, Ofcom is looking to gain a better understanding of the costs associated with providing the relevant products to a specified minimum service level (MSL).

For this purpose, Ofcom and Analysys Mason jointly developed a model to help quantify the relationship between QoS targets, performance levels against those targets, and the resource levels needed to achieve them.[13] Ofcom relied upon outputs from this model, which we refer to in this report as the Ofcom resource performance model (or 'Ofcom RPM'), to inform the position set out in its consultation.[14] A report by Analysys Mason summarising the modelling approach used, the input data relied upon, limitations of the model and key outputs was published as an annex to Ofcom's consultation.[15]

Since the publication of Ofcom's consultation, Openreach has provided Ofcom with a copy of its own model, developed in collaboration with EY. This model, which we refer to as the 'Openreach model' and which was not available to Ofcom prior to the launch of the consultation, is considered by Openreach to be a superior tool for understanding the relationship between QoS and resource requirements. Openreach has relied upon outputs from this model, including a comparison to the Ofcom RPM in its response to Ofcom's consultation.

Analysys Mason has been commissioned by Ofcom to carry out an assessment of the model to support a further public consultation, which will include consideration of the Openreach model.

This document is our final report, covering our assessment of the model. It is not a standalone description of the model or a user guide and so should be read in conjunction with Ofcom's consultation document and with a description of the model methodology (which constituted a part of Openreach's response to Ofcom's March 2017 consultation.

The remainder of this document is laid out as follows:

- Section 3 provides an overview of the Openreach model
- Section 4 sets out our general comments on the Openreach model and its construction
- Section 5 analyses the run-to-run variations in the outputs of the Openreach model
- Section 6 discusses how the model has been used to produce a subset of the outputs relied upon by Openreach in its consultation response

13 Ofcom initially developed a high-level model, which Analysys Mason reviewed and developed to increase its level of usability and functionality in a number of areas. The scope of our assignment also included producing outputs from the model for a range of scenarios of interest to Ofcom.

14 Ofcom, Quality of Service for WLR, MPF and GEA: consultation on proposed quality of services remedies, 31 March 2017. See https://www.ofcom.org.uk/consultations-and-statements/category-1/quality-of-service

15 See https://www.ofcom.org.uk/__data/assets/pdf_file/0034/99646/Analysys-Mason-report.pdf

- Section 7 presents the results of additional scenarios that we have run using the Openreach model
- Section 8 assesses the limitations, biases and uncertainty of the Openreach model results
- Section 9 provides further relevant outputs from the Ofcom RPM
- Section 10 presents our conclusions on the Openreach model.

Confidential data within this report has been redacted and is indicated by the use of square brackets and the scissor symbol ‘[✂…]’.

3 Overview of the Openreach model design

Openreach provides a detailed description of the Openreach model in its response to Ofcom's consultation (primarily in Section 8), including a dedicated annex produced by EY (Annex 2).[16] We do not attempt to describe the Openreach model in the same level of detail in this section, but rather provide a high-level overview of the key features of the model design.

3.1 High-level model overview

The Openreach model is an allocation model that simulates the field activities of Openreach's Service Delivery team and the impact of various supply- and demand-side levers on meeting job targets, as summarised in Figure 3.1 below.

Figure 3.1: Supply and demand levers of the allocation model [Source: Openreach, EY, 2017]

Supply side	Demand side
Number of field engineers	Provision and repair job volumes
Service level definitions	Care level mix of these jobs
Job prioritisation rules	

The Openreach model simulates job arrivals and executions using discrete event simulation, where steps in the job servicing process all occur at discretised points in time. The model is run in a SimPy environment in Python 3, with input and output data read from and written into an Oracle database. An overview of the model process flow is presented in Figure 3.2 below.

16 See https://www.ofcom.org.uk/__data/assets/pdf_file/0013/105115/Openreach.pdf

Figure 3.2: High-level overview of the Openreach model [Source: Openreach, EY, 2017]

[✂]

Job characteristics (e.g. task times, job visit rates) are assigned to individual jobs via random sampling from probability distributions generated from historical Openreach data. Jobs then enter the model by entering a queue that is ordered by the priority of the job, as determined by a job priority scoring framework. Jobs are then allocated to engineers, based on the minimum skill level required to complete the job, who then execute the job. The model collates job completion metrics on a weekly basis, which are batched together for the time frame of the simulation and provided as a model output.

The resources are expressed in full-time equivalents (FTE) terms and each level of resource will result in a certain level of performance against provision, CL1 and CL2 repair targets. These raw model outputs are then processed in various ways to calculate 'deltas'. These deltas are the percentage difference (delta) between the resources required over a specified period of time (e.g. one year) to deliver a specific level of performance compared to the resources required to deliver a (lower) *baseline* level of performance.

Various aspects of the allocation model simulation process are discussed in more detail in Section 3.2.

3.2 Allocation model simulation process

This section discusses the discrete event simulation process in greater detail, including specific inputs and outputs.

Job entry

The actual number of jobs by product type, SOM[17] area and week is used to generate a set of daily job volume distributions. Individual jobs are assigned the characteristics listed below based on distribution sampling, which are themselves derived from historical Openreach data:

► *Categorisation*

Every possible combination of product, main fault location ("MFL"), care level and appointed/non-appointed status is assigned to each job, determined by a weekly frequency distribution.

► *Logged time*

The week, day and hour a job gets logged is determined by a distribution dependent on that job's care level. A uniform distribution determines the minute within each hour a job gets logged.

17 Areas in the Openreach network are divided between Senior Operations Managers, with each Senior Operations Manager having responsibility for Service Delivery within that area.

► *Appointment time agreed by customer*

A frequency distribution, based on a job's care level, the day of the week and the time of day that a job is logged, is sampled to determine the appointment time. Customers are implicitly assumed to accept the allocated appointment time.

► *Total visits*

A frequency distribution based on a job's MFL is sampled to determine the total number of visits required to complete each job. If a job requires more than one visit, the MFL categorisation is itself resampled for the second visit, with the result then applied for all subsequent visits.

Job allocation

The job allocation process is governed by both the specific characteristics of the job and by the engineer pool. Each job has an MFL which determines the minimum skill required for an engineer to execute the job. Jobs are allocated in priority order to the first available engineer with the lowest feasible skill level. Each SOM has its own pool of engineers from which resources are taken to complete jobs, and added back to when jobs are finished. Each pool's engineer resources (FTEs) are distributed across skill categories, which vary by week according to increases and decreases from hiring, attrition and shrinkage (e.g. sick leave). When reducing FTEs (e.g. at the end of the day, as engineers' shifts end), the model will remove available engineers first. If the reduction is not yet met, the model will remove the most recently allocated engineers from their allocated visit. Time spent on these jobs will be lost and the job will re-enter the queue. Skills [✂], as shown in Figure 3.3.

Figure 3.3: Engineer skill hierarchy [Source: Openreach, EY, 2017]

Skill	Definition
Broadband Underground ("BBUG")	Multi-skilled underground engineers with broadband skills
Underground ("UG")	Multi-skilled underground engineers
Customer Access Line ("CAL")	Basic repair and provision engineers
One Man Installer ("OMI")	Provision only engineers

While no distinction is made between sub-levels of experience within each skill level, the model can simulate scenarios which modify the skill mix in the pool, as well as any upskilling and the time at which the upskilling takes place. Once a visit has been allocated to an FTE resource, a frequency distribution is sampled to determine whether an additional resource is required to assist, and whether it is a skilled or an unskilled assist, where the skill for a skilled assist depends on the skill of the original resource allocated to the initial visit.

Job execution and output

A job's target completion time depends on the care level and/or the agreed appointment time. For provisions and care level 1 and 2 jobs, the end of the day of the agreed appointment date is used as the target completion time. For care levels 3 and 4 jobs, the end of the agreed appointment slot is used as the target completion time. If the appointment slot falls within the care level service level agreement (SLA), then the end of the SLA period is used instead.

A job in the queue can be either:

- a future job, where the target completion time has not yet passed and occurs a number of days into the future
- a due job, where the target completion time has not yet passed and falls on the same day as the simulation time, or
- a tail job, where the target completion time has passed.

Jobs enter the queue at the time at which they were logged, or if appointed, at the beginning of the appointment slot. A job is considered to be complete when all visits and assists have also been completed, at which point the model records whether the job passed or failed the target completion time. Incomplete jobs re-enter the queue as 'furthers' on completion of a previous visit instance and all associated assists; appointments only apply to the first visit. Travel and task times for the job are sampled from a frequency distribution, and are sampled independently for each visit, including assists. The task-time frequency distribution is dependent on the skill of the engineer allocated to the visit, product and MFL, with separate distributions used for furthers and completions. Inefficiencies due to "time-to-competency" are reflected in a task-time adjustment factor, defined by the week in which they occur until the next adjustment is defined.

Tail jobs (i.e. jobs that have missed their SLA and are overdue) are addressed at the end of each week. A distribution dependent on the number of days a job has been in the tail is sampled to determine whether a resource would have been allocated during that week to complete the job. The total time to complete any outstanding visits is based on the minimum skill required to complete the job. Upon completion of the job, it is marked as a fail.

Outputs are collated on a weekly basis at an SOM level, and include:

- volume of logged jobs, agreed appointments, passed completions and failed completions
- service performance split by product, MFL, care level and appointment status
- workstack (queue) length at the end of each week detailing the number of jobs against the number of days over/under their target completion time
- total available time by skill level split by time categories (idle, busy, lunch and end-of-day inefficiency)
- total busy time by skill level, split by time spent executing future, due and tail jobs during standard working hours
- total time clearing tail jobs outside of standard working hours
- projection by repair skill of available time and committed time

- stress factor.[18]

These outputs are written into both the Oracle database and into .csv files. The volume of data saved for each simulation is considerable: several MB per SOM per resource level.

Post-processing of outputs

As the outputs indicated above are only on a weekly intra-SOM basis and do not report overall performance by GM[19] or nationally, the outputs written into Oracle are processed further.

Openreach has defined various "high-level scenarios" (HLS) which run the model for a range of scenarios with consistent input parameters for each SOM and a set of resource levels intended to cover the range of performance of interest. These are numbered in a way that records the SOM and resource level: e.g. scenario 8211104 corresponds to HLS 82, SOM 111, and resource level 4.

► *SQL post-processing*

Openreach has used a number of different scripts to post-process the results within Oracle. Minor modifications need to be made to these scripts to handle each high-level scenario, usually replacing a high-level scenario number or (if the database is being run with different table names for the output tables) changing to the appropriate output table names.

"Smart resourcing uplift curve" is the only post-processing script used by the model runs Openreach relies on in its consultation response, apart from for the 'waterfall chart' (discussed in Section 6). The waterfall chart (Figure 39 in Openreach's consultation response) uses three different scripts:

- Resourcing uplift AM alignment.sql
- Resourcing uplift Raised GC.sql
- Smart Resourcing Uplift Curve.sql

These three scripts are quite different in important ways: they handle "fails" differently (in the first case), are relative to different baseline levels of resource, and "Smart resourcing uplift curve" also adjusts the resource level for a given performance to take into account the effects of overtime.

- Resourcing uplift AM alignment.sql is concerned with one particular way Openreach was trying to emulate the behaviour of the Ofcom RPM. We do not consider it further.

- Resourcing uplift Raised GC.sql does not take overtime into account and is relative to a baseline set by resource levels in each SOM that match the actual 15/16 performance in HLS 83.

---

18 The stress factor is a measure of the shortfall or surplus in resource relative to the projected amount of resource required to complete jobs in the current workstack. The model can apply a different mode of operation in response to periods of high stress in relation to factors such as job prioritisation, shrinkage, customer appointment uptake and the clearing of tail jobs.

19 SOM areas are grouped together under the supervision of a General Manager. These groupings of SOM areas are referred to as GM areas, or GMs

- Smart Resourcing Uplift Curve.sql adjusts the resource level for a given performance to take into account the effects of overtime and measures relative to the actual SOM resources in 15/16.

Overtime is an issue because it is used in the Openreach model to clear tail jobs. Adding resources leads to existing staff also working less overtime, which would be invisible in the results unless it was accounted for somehow. Openreach reports that in some circumstances if more resources are added less overtime is needed and the results changed in a way that was counterintuitive. To avoid this, Openreach uses a post-processing adjustment to the additional resources needed in each modelled scenario to ensure a more realistic level of overtime is implied (this adjustment reduces delta). However, changing the number of resources in post-processing is also not exactly equivalent to having that many in the modelling in the first place.

Excel post-processing

Openreach uses an Excel file, one per HLS to further process the outputs of the post-processing scripts. The results of the SQL post-processing lead to a single small table per HLS. This table can be exported from Oracle in a delimited format and imported into Excel, where it is pasted into a sheet in the workbook.

The 'delta' result is calculated based on the data from the SQL post-processing, for each GM in turn. Within each GM, the GM performance is calculated from SOMs based on the assumption that all SOM resources are incremented at the same time. The resulting national resource uplift is based on the sum of the GMs.

4 General comments on the Openreach model and its construction

In this section, we set out our views on various aspects of the Openreach model and its construction.

4.1 Coding and standard of construction

The code is generally tidy. However, there are few comments in the code. Additional comments would be useful (e.g. marking past changes, or marking areas where the code is correct but might be counterintuitive).

The database interaction is well constructed and has a very limited dependence on the choice of database: thanks partly to the use of the "sqlalchemy" library, it appears that EY can run near-identical code (with only the database connection string differing) against SQL Server, whereas Openreach uses Oracle (11g). We have used Oracle 12 SE2 for our own runs.

Within the code, there are occasional hints of previous design decisions that have been changed (e.g. unused variables). The most serious of these changes is that in various parts of the code and database GM_ID may mean GM_ID or it may mean SOM_ID if the user has chosen to use a more disaggregated configuration (i.e. it represents "chosen granularity of area ID" in some circumstances). If the model is to be used further in the future, it would be a good idea to clarify this point within the model and the supporting data tables, by renaming fields and variables so as to reflect their actual contents.

Some of the database fields have become 'denormalised' (i.e. one field contains multiple pieces of information), which is not best practice for relational databases. For example, ID for scenarios contains data relating to HLS, SOM (implying GM) and resource increment identifiers. Any query where the user is interested only in an individual SOM cannot directly access this and needs to write SQL code to extract it (which is also inefficient because there is no index on SOM).

4.2 Configuration control

We can see no evidence of configuration control of the model (e.g. there are no comments with version numbers in the various files, so it is not clear how any version control is achieved).

This does not mean that there is necessarily any issue with the model results, but is a potential risk to the quality of the results and the efficiency of the modelling process in future, especially if there are changes being made.

We recommend that some form of configuration control should be used in future. This could include version numbers in comments in each module and could also include comments on changes made to the code, and code version numbers recorded in the outputs.

4.3 Repeatability

As the model uses random numbers, it would be useful if the model would behave repeatably for testing purposes. For example, the ability to generate a known set of outputs would be a useful end-to-end test when installing the model.

An extensive attempt to get repeatability between our own model runs and those of Openreach and EY failed. This was because, although the model does set and manage a random seed for the random number generator used (provided by NumPy), it is known that it is also necessary to set PYTHONHASHSEED=0[20] in order that runs on a given version of Python are repeatable. This is because the model uses iteration over dictionaries[21] when assigning random inputs to specific parameters (e.g. the 'minute' at which a job arrived.)

However, iteration over dictionaries is not guaranteed to occur in the same order for different versions of Python (e.g. 3.4.1–3.6). Consequently, although the results of a given installation are repeatable, the model gives different results for different versions of Python. This is not an insuperable issue and is not likely to cause any bias in the results, but it does mean that it is not possible to have a simple end-to-end test of a new installation (one cannot simply say "run this scenario and you should get the following outputs").

In order to stay as close as possible to the Openreach configuration, we have used Python 3.4 for all the runs reported here, although it is not clear whether different deltas might be produced using a more recent version of Python (and if so, it is not clear, *a priori*, that one set of results would be more 'correct' than the other).

4.4 Known bug causing runs to fail

A small fraction of runs of the model (roughly 0.1%) will crash Python with an "index error" (the specific part of the code is to do with reduction in the number of engineers, likely to occur at the end of each modelled day). We presume that this is an 'edge case' that is not handled correctly in the event handlers which deal with the reduction in the number of engineers.

It is undesirable for there to be a known error[22] in a model which has not been addressed, even if it only comes into play rarely and is (by causing a crash) unlikely to affect the overall results (in that in all cases we have results from nearby resource levels above and below the value which fail and which give similar results). However, even if this bug is not likely to affect the overall results, its existence (and the fact that it has been allowed to persist in code used to support the consultation response) somewhat reduces our level of confidence in the correctness of the rest of the design (i.e. perhaps other cases are also handled incorrectly).

---

20 Note: uppercase is necessary.

21 A type of data structure in Python.

22 From discussions with Openreach, we understand that the company is aware of this bug.

We have not attempted to replace these results (e.g. by using an alternative random seed); we expect that the worst that will happen as a result is that the Excel using the outputs of the post-processing scripts might need to step to a slightly higher resource level (e.g. one more resource step) to reach a specified performance in that SOM.

4.5 Runs can be duplicated

The model is designed to be run in an unattended batch mode with each copy of the model picking up a list of jobs to be done, choosing the first in the list, marking in the database that it has started that job, and running that job. If many copies of the model are all running at the same time (which is the only way to use the model efficiently on a multi-core machine), occasionally two of them both do the same job, meaning that there can be duplicates in the results.

Although it is not clear whether any of the post-processing would be affected by duplicates, we have removed duplicates before post-processing (in case). There is therefore only a minor efficiency impact and no impact on the results.

4.6 Runs can be identical

The resource levels used within the simulation in a given SOM and week are a result of a process which takes a total resource level for that run in that SOM and week (derived from various tables) and divides them across different skill levels (according to ratios from another table), and across different times of day (corresponding to the result of shift patterns). Patently the number of engineers of each skill level at any given date and time of day has to be an integer.

Some runs for adjacent resource levels in a given SOM give identical results (with the same random seed and PYTHONHASHSEED=0). We assume that this is due to the way in which the resource levels are set to integers (i.e. the rounding involved means that some input resource levels which are similar (e.g. have one more resource overall in a given SOM) result in identical modelled resources by skill – the extra resource has been "rounded away"). The net effect is that the modelled resource is slightly less than the input requested, though it is that input level that is recorded as the resource level in the model output. There is therefore a small uncertainty in the X axis of the performance against resource level charts (the modelled resource level may be slightly different to that recorded, and will on average be slightly lower[23]). At a national level, this effect is likely to be small compared to the total resource and is not a significant source of error in the results (which will in any event fluctuate due to random effects).

This also creates a slight inefficiency (important because the model uses a lot of computer resources) in that runs are being done which are not useful.

23 This is about implementation choices: there are alternative ways to generate integer numbers of engineers which are close to the required distribution of skills.

4.7 Documentation

The documentation covers functional specification elements such as the desired inputs and outputs and the main design choices (such as discrete event simulation, level of granularity). It is concise and appears to be accurate (as far as it goes). It is however short for a model of this complexity. For example, it does not contain:

- exact details of minimum (or preferred) versions of Python and support libraries needed
- details of the software design such as:
  - approach to visit prioritisation
  - nature of events and event handling, including the more complex events such as “engineer stops work for the day” (which appears to be where the bug noted above exists)
  - database table design (e.g. required tables, field datatypes, indices, primary and foreign keys, constraints)
  - assumptions made about relationships between the tables, (e.g. if it is essential that two tables represent the exact same set of jobs, or not)
  - SQL post-processing procedures
  - Excel sheets
- detailed instructions for configuring the model for new sets of assumptions (e.g. “How to: guides”)
- hardware requirements (for the database, and for each of the Python calculations).

In this context, we note that Openreach has been helpful in addressing ad-hoc questions about use of the model, including those concerning the above areas.

4.8 Ease of use

The model is not easy to use end to end, even for relatively experienced modellers with good database and Excel skills. There is in effect no user interface other than a development environment for the chosen tools (Python, Oracle, Excel).

Installing the model took a considerable time because we were creating tables, loading data and modifying scripts to account for new schema and table names.

Using the model requires a combination of:

- generating input data tables
- configuring the model through setting many values in multiple additional input tables
- running the model in Python (likely requiring dozens of high-level scenarios, each of which may require 25–40 resource levels for each of 56 SOMs)
- dealing with run-time issues such as duplicated runs and runs which crash
- creating and running post-processing scripts in Oracle PL/SQL, and finally
- pasting values into a set of Excel sheets and interpreting the results.

Errors at any of these stages can be made, and there is little in the way of error checking that would detect mistakes. As such, it is definitely a tool for experienced users and would need to be significantly more automated to allow use by non-experts.

4.9 Speed and amount of computing resource required

The model is very slow to run. Running multiple high-level scenarios (for 56 SOM and 25–40 resource levels) requires very large amounts of computing resource to allow tractable use in a reasonable time. Each run (1 SOM, 1 resource level) takes hundreds of seconds (it gets slower as the database becomes larger – i.e. after a greater number of runs have taken place). We are running the Python code on two machines with 32GB of RAM and 12 cores each. With 18 copies of the model run in parallel over the two machines, we can calculate about 100 SOM/resource pairs an hour, which means that a single high-level scenario takes approximately 24 hours to run (assuming no crashes).

The database needed to support the runs is also large (multiple tens of GB) as each run creates very granular data regarding performance in each SOM in each week. The final outputs are much smaller, being single Excel sheets of a few MB.

This lack of speed is partly due to the choice of modelling approach (discrete event simulation looking at individual jobs) and tool (Python/SimPy). These choices are not inappropriate but they make the model harder and more expensive to use to generate understanding through sensitivity analysis.

4.10 The input data

The source of key input data to the model is not always clear. This is not to say it is necessarily incorrect, but rather that it is not always possible to understand whether this data is 'as intended', or indeed how raw data may have been processed in order for the key input data to be arrived at. The EY model documentation does not cover the input data or the processes that have been used to create it in any detail.

Openreach have provided us with a copy of AM_JOBDATA_FULL_BY_SOM, which is a processed version of the raw Openreach data; this table is the main input used to generate many of the distribution tables used by the model. It does not represent the original source of the information, which is a further set of tables recording the raw Openreach operational data, which we have not been provided. In some cases additional input tables are relied upon by these scripts e.g. TMP_AM_PD3_BUILD_GEA_LINKED_2

We have also been provided by Openreach with a document ("Allocation Model Data Inputs Documentation") containing a set of SQL scripts that have been used to generate AM_JOBDATA_FULL_BY_SOM (from its source tables) as well as the various tables directly used as model inputs (e.g. TBL_AM_CATEGORY_DISTNS) generated in turn from AM_JOBDATA_FULL_BY_SOM and in some cases additional input tables. Some of these scripts were modified by Analysys Mason and used to generate different inputs (e.g. the high-level scenarios that vary the CL mix).

Whilst analysis of the processing of the raw data to generate model inputs falls outside of the scope of this audit, we feel it is important to highlight areas where it is not obvious how the processed data corresponds to the desired input. One major example that we have come across relates to the glass-ceiling assumptions. Specifically, assumptions about the level of the glass ceiling need to be translated into the model input parameters.

Openreach describes each HLS as being based on either "actual glass ceiling" or "raised glass ceiling" assumptions. We had understood from Openreach's consultation response that the "actual glass ceiling" corresponded to an assumption of 87.5% (derived from 2015/16 data) and that the "raised glass ceiling" corresponded to an assumption of 92.6%.[24] However, upon inspection of the input data, the failure ratios of the relevant "category distributions" in fact appear to be almost identical:

- Those HLS described as "actual" and "actual 16/17 CL mix" have [✂] automatic fails (in TBL_AM_CATEGORY_DISTNS) depending on the GM.[25]
- Those described as "raised (16/17 CL mix)" have [✂] automatic fails depending on the GM – i.e. not obviously a "raised GC" at all.[26]

In neither case do the inputs appear to correspond directly to the glass-ceiling numbers presented in Table 9 of Openreach's consultation response: in both cases the automatic fails in the model inputs appear lower. However, we understand that these automatic fails may correspond only to the "red" items shown in Table 9 of the Openreach consultation response, with "amber" items taken into account implicitly within the model's other distributions. This means that it is possible to see how the input data for automatic fails may be consistent with our understanding of the "raised glass ceiling" assumption (92.6%) used by Openreach, but somewhat harder to see in relation to our understanding of the "actual glass ceiling" assumption (87.5%). It is therefore also possible that the "actual glass ceiling" is intended to be at a level higher than 87.5%.

This inconsistency notwithstanding, we note that the deltas calculated by the model under "actual glass ceiling" assumptions do appear substantially higher than using "raised glass ceiling" assumptions. Therefore, it is quite possible that both the input values and our understanding of the glass ceiling assumptions are correct, but the way in which this aspect of the model functions remains unclear to us.

24 See Table 9 of Openreach's consultation response.

25 The CL mix has not affected the overall ratios at all.

26 For completeness, on the same basis the "Raised (Analysys Mason)" glass ceiling inputs have [✂] automatic fails depending on the GM (i.e. the glass ceiling is higher, and automatic fails are lower).

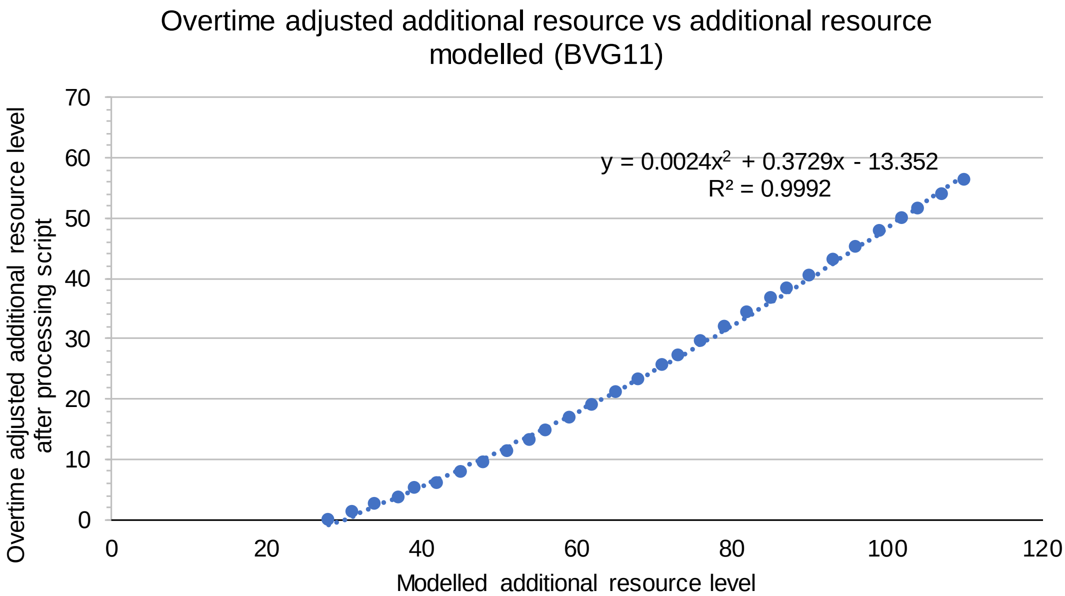
4.11 The SQL post-processing

The "Smart resourcing uplift curve" post-processing changes the modelled results by modifying the resource levels that give the performance modelled.

We have examined the resource levels before and after this adjustment, and believe that this adjustment is potentially removing too much of the additional resource (i.e. this is a case where the deltas reported by Openreach may be too low after the post-processing). We would expect the adjustment made to the resource levels to account for overtime to be a small percentage of the added resource (possibly of the order of the fraction of total worked hours that are overtime), and we would also expect it to be proportional to the incremental resource (i.e. over the baseline figure). While the adjustment does appear to be nearly linear, it has an offset (resource levels at low levels of performance are removed from the dataset[27]) which is unexpected, and a sizeable fraction of the modelled resource added is removed.

This is illustrated in Figure 4.1 below, which shows the modelled additional resource levels in a specific SOM and HLS before and after processing.

Figure 4.1: Example impact of 'Smart resourcing uplift' processing on modelled resource levels [Source: Analysys Mason, 2017]



To give an example of the overall effect on delta for 90% performance, we have processed HLS 203 outputs using two different post-processing scripts, one of which has the overtime correction turned off. These scripts use different baselines, so we use a self-baseline (i.e. we measure the difference in resource delta between achieving 80% performance and 90% performance in the same scenario). The self-baseline delta in the overtime adjusted case is 17%; in the non-overtime-adjusted case it is

27 This causes another inefficiency in that runs have been done which do not contribute to the results.

24.5%. This shows that the effect of the overtime adjustment is to reduce delta for 90% performance by a material amount.

4.12 The Excel post-processing workbook

The Excel workbook used for processing the results of the SQL is appropriate, but we have found five specific features that mean that the results need to be carefully interpreted:

- It is based on an assumption that repair CL2 jobs are the limiting factor for performance. Any scenario where this is not true will not have the correct delta calculated in the Excel post-processing workbook. Particular care is therefore needed if CL2 is not the limiting factor.
- If a GM does not meet the target performance for CL2, it deploys the lowest resource level that gets the highest performance reached. This means that the "UK Fld" sheet reports a delta result at the national level for a given level of performance even if this target performance level is not met in every GM.
- No interpolation is performed. However, due to the way the results are compared to a baseline value, the resulting error from the lack of interpolation would not always be biased,[28] and the resource steps modelled are quite small compared to the run-to-run variability (e.g. if the model is run with different random seeds).
- The Excel post-processing workbook always reports a minimum resource increment. This is unlikely to affect any of the results of interest (where resource increment is generally well above the minimum level), although it might suggest more resource than is actually required to provide the "baseline" level of performance (e.g. 80%).
- The reported deltas are measured relative to the baseline used by the SQL post-processing script in each case. Deltas with different baselines are therefore not directly comparable.

We have indicated in our results summary table in Figure 8.1 when either of the first two points above are relevant (CL1 performance <CL2 performance, or target not met in each GM).

By comparison, the Ofcom RPM Excel processing (based on "worst GM") reports the resource level at which all three job types (provisioning, CL1 and CL2) meet the target performance in each GM.

The use of "national" figures coming out of this calculation is comparable to the "worst GM" approach of the Ofcom RPM as long as the target level is met in each GM. As a result, attention should be paid to the "national" figures from the Openreach model (as long as the target is met in all GMs) and these should be compared to the "worst GM" figures from the Ofcom RPM.

28 Although this means that the resource estimate to hit the target performance level is potentially up to one modelled resource step too high, the net impact on the result is small in most cases because the same applies to the baseline value, assuming that it is also set by using the model (which would have the same issue, being on average half a unit of resource step too high). Sometimes the baseline value is not set using the model but is derived from actual 15/16 resources, in which case some bias may be introduced.

5 Run-to-run variation

In order to assess run-to-run variations in model outputs, we have looked at:

- run-to-run variation in Openreach's scenario with the 15/16 glass ceiling (HLS 77) – 2 runs
- run-to-run variation in the simplified scenario Openreach constructed to represent the Ofcom RPM at the start point of its waterfall chart (HLS 82) – 6 runs.

We discuss our analysis of each of the above in turn through the remainder of this section, before providing our conclusions on the run-to-run variation of the Openreach model.

5.1 HLS 77

We have run two copies of HLS 77 with different random seeds. In parallel with our work, Openreach and EY have also repeated a number of runs for a high-level scenario they call "calibrated allocation model". We believe this corresponds to HLS 77.

This scenario is described by Openreach as using the "actual 15/16 glass ceiling" (whose actual implementation we discussed in Section 4.10). Openreach has reported the deltas (relative to the baseline set in the post-processing script) as shown in Figure 5.1.

Figure 5.1: Openreach reported results for HLS 77 [Source: Openreach, 2017]

[✂]

Note that the above reported mean deltas differ slightly from the numbers presented by Openreach in Table 19 of its consultation response (e.g. [✂] vs. 24.9%). We believe that this difference is due to the numbers in Figure 5.1 comprising an average over different random seeds and because in the model runs underpinning the numbers presented in Openreach's consultation response, PYTHONHASHSEED was not being set to zero – an error which Openreach has subsequently corrected. We comment on the magnitude of this difference further below.

Our own model runs, based on two different random seeds for HLS 77, give the results shown in Figure 5.2 below.

Figure 5.2: Results for HLS 77 from Analysys Mason model runs [Source: Analysys Mason, 2017]

Repair MSL	Mean delta (2 runs)	Estimated sample standard deviation
87%	9.68%	0.30%
90% (not reached in all GMs)	38.61%	4.08%
93% (not reached in all GMs)	52.58%	3.55%

The 90% and 93% targets are not met in all GMs in our runs (and may not be met in Openreach's runs either; the company does not report whether this is the case). With a glass ceiling of 87.5% it should not be possible to meet the 90% and 93% performance targets, though as we note in Section 4.10 there is a lack of clarity as to how this glass ceiling assumption, including how it relates to the model inputs. The figures reported for 90% and 93% targets are not directly comparable because they are not reaching the target performance.

Without more runs of HLS 77 it is hard to tell whether the estimate of the standard deviation is an accurate one, but the difference between the 87% performance figures reported by Openreach and those we have calculated is significant if the results are normally distributed and if the standard deviation is indeed this small. In particular, a difference in the mean of over 1% implies being more than three standard deviations from the mean if Openreach's estimate of the mean is correct. There is obviously a substantial difference in the 90% and 93% performance levels, although as noted the figures we have obtained are not able to be directly compared to those obtained by Openreach due to the performance target not being achieved in all GM areas.

There are similar effects in the data reported by Openreach at the GM level, where the reported standard deviation is low compared to the difference between Openreach's originally reported results (in its consultation response) and the average resulting from Openreach's more recent ten runs. Of course, we would expect greater variability at the GM level, but to be more than three standard deviations from the mean is an unusual occurrence if the results were normally distributed, and to be seven standard deviations away is extremely unlikely (less than one in a trillion), suggesting a non-normal distribution.

5.2 HLS 82

As HLS 82 has various model features effectively turned off (e.g. jobs all have the same duration), it might be expected to show less run-to-run variation than HLS 77. This is indeed the case from our runs with six different random seeds, as shown in Figure 5.3 below.

Figure 5.3: National delta (relative to HLS 83 figures used by post-processing script) [Source: Analysys Mason, 2017]

Random seed	Performance level sought		
	83%	90%	93%
12345	1.99%	4.30%	5.78%
11111	2.01%	4.48%	5.82%
22222	2.01%	4.39%	5.75%
33333	2.07%	4.37%	5.82%
44444	2.01%	4.44%	5.82%
55555	1.90%	4.22%	5.70%
Mean	**2.00%**	**4.37%**	**5.78%**
Standard deviation	0.05%	0.09%	0.05%

What is clear in this scenario (and is even more clear if we examine the GM level data, not shown here for brevity) is that the lack of interpolation in the post-processing means that the "steps" in the resource levels modelled are having a small but material effect on the result at the GM level. This means that much of the variability from run to run (such as it is in this scenario) is being 'hidden within the resource steps' in the result. These steps are of different size in different GMs but are typically ~2% in GM resources, meaning that the resource estimates should be understood to have an additional variability of up to 2% at the GM level (which translate into steps of roughly 0.2% nationally). Not all GMs will be overestimated to the same degree, so the systematic error is likely to be 0.9% on average at a national level and vary over a range, with a 95% confidence interval of approximately 0.5%. The overtime-corrected post-processing, which measures delta relative to the actual level of resources in 15/16 may therefore overestimate delta by approximately 1% due to this effect (i.e. a quoted delta of 5% means 4%± 0.5%).

However, as similar effects will also apply to the baseline run where a baseline run is used, the estimated national deltas using baseline runs should be understood to have no systematic overestimation but a slightly wider confidence interval of 0.75% (i.e. a quoted delta of 5% means 5%± 0.75%).

5.3 Conclusions on run-to-run variations

The level of variability in national delta we have seen in the results for cases where the target performance is met is not excessive. On the other hand, cases we have calculated where the target performance is not met seem to have quite different results to those calculated by Openreach. We note that in such cases we are potentially not comparing like with like, and the resource requirements will be varying very strongly (as the curve is asymptotic to the glass ceiling).

If the results are assumed to be normally distributed, the estimated standard deviation in HLS 77 seems low, given the variation between our results, Openreach's original results, and the set of 10 results provided most recently by Openreach.

Under a similar assumption of a normal distribution, Openreach's original results at the GM level seem to be very different, although they are close at the national level. At the GM level, the results vary by more than five standard deviations from the average of the most recent set of 10. Given the asymptotic behaviour near the glass ceiling where the resources required for a given level of performance can increase very rapidly, it may be that the right approach is not to assume that the resource requirement results will be normally distributed (at least in this region). While this may indicate that many runs should be undertaken if results are to be used to support policy decisions, the sheer run time of the model even on very large multi-core PCs makes this a very expensive option in terms of both resources and elapsed time.

We think that the use of steps in the modelled resources means that estimated deltas may have an additional variability of the order of 0.5%.

6 Use of the model to reproduce a subset of the outputs relied upon in Openreach's consultation response

In this section, we consider the way in which the Openreach model has been used as part of Openreach's consultation response.

6.1 Subset of Openreach's scenarios considered

We have described in Section 4 how the model has been used by Openreach to generate results for HLS 77, the high-level scenario upon which the main resource deltas associated with achieving 87%, 90% and 93% performance appear to be based. In particular, we noted the difficulty in replicating these results and the potential for run-to-run variability.

The other main results presented in Openreach's consultation response upon which we focus in this report concern the 'waterfall chart' presented by Openreach in Figure 39 of its confidential response[29]. This waterfall chart summarises Openreach's analysis of why it thinks the Openreach model differs from the Ofcom RPM in its calculation of resource deltas. In order to analyse this waterfall chart, we have run and post-processed the following HLS used by Openreach to generate it:

- HLS 82
- HLS 100 (twice, using different post-processing scripts)
- HLS 106
- HLS 111
- HLS 107
- HLS 108
- HLS 109
- HLS 112
- HLS 77.

The Openreach waterfall chart scenarios use two different baselines: one for the "simplified model" (where the baseline is set by the resources needed to match SOM-level results in 15/16 in a particular HLS, 83), and one for the so-called "calibrated model"[30] (where the baseline is based on the actual SOM resource numbers in 15/16, and not on model outputs).

Openreach also uses different post-processing scripts consistent with these two baselines (HLS 82 and 100 use a script which takes a different approach to "fails" to try to emulate the way the Ofcom

[29] Discussion of this waterfall chart has been redacted by Openreach from its non-confidential consultation response.

[30] Note that the calibration is not so accurate as to result in the model outputting the *actual* SOM level performance at the *actual* SOM level resources.

RPM works); within the waterfall chart high level scenarios, only HLS 112 and HLS 77 use the overtime adjustment.

6.2 Discussion of the Openreach waterfall chart

Having re-run the scenarios used by Openreach to generate its waterfall chart, we obtain the results shown in Figure 6.1 for achieving 90% performance.

Figure 6.1: Scenario descriptions and run results for HLS used by Openreach to produce the waterfall chart [Source: Openreach, Analysys Mason, 2017]

HLS	Openreach scenario description (or equivalent for non-Openreach scenarios)	Glass ceiling	Glass ceiling (category distribution ID)	Care level mix	Skilling	Post-processing	Calculated delta for 90% performance (from our model runs) relative to relevant post processing script baseline
82	Openreach Simplified Allocation Model - uplift curve base	98.6% Provision, 97.0% Repair	2711	50/50 CL1/2	100% BBUG	Resourcing uplift AM alignment.sql	4.30%
100	Openreach Simplified Allocation Model - uplift curve 5% Repair Minor Failure	98.6% Provision, 97.0% Repair	2711	50/50 CL1/2	100% BBUG	Resourcing uplift AM alignment.sql	5.61%
100 (different post processing)	Openreach Simplified Allocation Model - uplift curve 5% Repair Minor Failure	Raised (16/17 CL Mix) – 92.6% Repair	2711	50/50 CL1/2	100% BBUG	Modified resourcing uplift	7.05%
106	A-M alignment Waterfall - Care Level Mix moving from 50/50 to 16/17 CL Mix	Raised (16/17 CL Mix) – 92.6% Repair	3311	16/17 CL Mix	100% BBUG	Modified resourcing uplift	7.81%
111	A-M alignment Waterfall - Using Actual Visit Rate	Raised (16/17 CL Mix) – 92.6% Repair	3311	16/17 CL Mix	100% BBUG	Modified resourcing uplift	27.22%
107	A-M alignment Waterfall - Including Assists	Raised (16/17 CL Mix) – 92.6% Repair	3311	16/17 CL Mix	100% BBUG	Modified resourcing uplift	28.72%
108	A-M alignment Waterfall - using actual Attendance Profile	Raised (16/17 CL Mix) – 92.6% Repair	3311	16/17 CL Mix	100% BBUG	Modified resourcing uplift	25.91%
109	A-M alignment Waterfall - using proposed 83% UG/BBUG Skill Mix	Raised (16/17 CL Mix) – 92.6% Repair	3311	16/17 CL Mix	83% UG/BBUG	Modified resourcing uplift	29.93%

HLS	Openreach scenario description (or equivalent for non-Openreach scenarios)	Glass ceiling	Glass ceiling (category distribution ID)	Care level mix	Skilling	Post-processing	Calculated delta for 90% performance (from our model runs) relative to relevant post processing script baseline
112	A-M alignment Waterfall - Actual TTs, Shrinkage, FTE & including non-MSL volumes	Raised (16/17 CL Mix) – 92.6% Repair	1911	16/17 CL Mix	83% UG/BBUG	Smart resourcing uplift (includes overtime adjustment)	10.99%
77	Calibrated Allocation Model - Year 3 uplift curve	Actual (16/17 CL Mix) – 87.5% Repair	1111	16/17 CL Mix	83% UG/BBUG	Smart resourcing uplift (includes overtime adjustment)	41.50%

We note, however, a number of problems with Openreach's analysis of the waterfall chart, as presented in its consultation response. We discuss these problems below.

6.2.1 Issues with Openreach's conclusions

The aim of the waterfall chart presented by Openreach in Figure 39 of its consultation response is to show the results of a series of high-level scenarios in which specific parameters vary in an additive way (change item 1, change item 1 and item 2, etc.) so as to measure the impact ("change in delta") resulting from each factor in turn (e.g. item 2). This could allow us to understand which of the changes from the simplified model result in the most material changes in delta; Openreach is using this analysis to argue that some of the simplifications in the Ofcom RPM affect the results in a material way.

To create a waterfall chart using the Openreach model, we need to run a specially designed set of high-level scenarios (for each SOM), using a range of resource inputs spanning (at least) the base case (which Openreach has set as the resources such that each SOM matches its real-world performance in 15/16) and the desired performance level. After calculating the delta in that high-level scenario, we can calculate high-level scenario to high-level scenario changes in delta.

Openreach has not provided us with all of the Excel workbooks it used to generate the waterfall chart. However, it has provided the inputs for the high-level scenarios used to create that chart, as well as examples of the post-processing scripts (which change as we move across the chart) and an example of the Excel workbook used for a single high-level scenario (and confirmed which output from that Excel workbook is relevant). We have also discussed with Openreach how the baseline resource levels were set.

Leaving aside any wider concerns with the model or the way in which the model has been configured, the reasons that the waterfall chart presented by Openreach does not provide solid evidence can be summarised as follows:

- incorrect/inconsistent baselines are used
- some steps apply the overtime adjustments, and some do not
- an insufficiently wide range of resources have been tested in some of the high-level scenarios
- responses to stress and prioritisation are not used
- the simplified model of Openreach may not be a good representation of the Ofcom RPM.

We consider each of these reasons in turn below.

Incorrect/inconsistent baselines are used

Openreach does not calculate delta relative to the correct baseline in each scenario (it does so for the first and last scenario, but not for those in between). This is important because some changes to model parameters may affect both the resources required for the baseline and for the desired performance level; any change that does alter the baseline (i.e. resources to get SOM results as per 15/16) will have its delta incorrectly calculated.

In particular, Openreach should apply those changes that were true in 15/16 (like skills, shrinkage, task times, travel, rostering, visits per job, etc.) first, and things that were not true in 15/16, like 16/17 care level mix, later in the waterfall. This would allow measurements of delta to be relative to the ‘most similar’ 15/16 baseline. Not following this approach means that the effect of some of these 2016/17 changes might be over- or under- estimated.

Overtime adjustments are not applied consistently

The last two steps in the waterfall chart use a different post-processing script, which includes the overtime adjustments. If this is the right thing to do, then it should be done at all steps, or there is a risk it will mean that delta (or changes in delta) are being over or underestimated for various steps within the waterfall chart.

Insufficient ranges of resources are tested

The range of resource levels tested by Openreach does not reach the target 93% performance level in several of the scenarios. As noted above, in this case, the Excel workbook used by the Openreach post-processing script reports either:

a) the maximum resource level tested, if the glass ceiling has not been reached; or

b) the minimum resource to reach the glass ceiling (i.e. the point beyond which the results do not improve), if there is such a level.

Either ‘a’ or ‘b’ means that some of the results are reaching different levels of performance, and the deltas are therefore not comparable. This in turn means that the changes in delta from one scenario to the next are not meaningful.

Outcome ‘a’, when it occurs, means that the results for those high-level scenarios (those which needed to be run with more resources to hit target performance) do not inform us of the true delta resulting from those scenarios' input changes. Any that hits the limit gives an incorrect delta; any further high-level scenario in the waterfall would consequently also have an incorrect change in delta relative to that result. This ‘running into the resource limit’ occurs in the middle of the Openreach chart, and persists until the last two scenarios, meaning that the right-hand side of the chart is almost meaningless.

If we change the target performance level (say to 90%), the resources may have been sufficient and we may not hit the glass ceiling but:

- the ordering issue is still of concern (we would prefer Openreach to do all the changes that are true in 2015/6 first)
- it is still necessary to use a baseline that is appropriate to each run.

Stress and prioritisation are not used

Aspects which might reduce delta are turned off in the scenarios used for the waterfall chart. There are two principal issues:

- Stress is not configured to be used to change appointment distribution or job prioritisation in the waterfall plot high-level scenarios. We presume that an accurate modelling of the system response under stress would improve repair performance (which is a limiting factor) because we know that under stress Openreach does change the way it appoints provision jobs. It is not clear whether this would reduce delta (i.e. the additional resource to reach a specified high level of performance), but we feel it is likely that it would do so.

- All but the last two high-level scenarios used in the waterfall calculations have a very limited use of prioritisation, only prioritising appointed jobs (i.e. they are using SCORING_FRAMEWORK_ID=11). It is unclear whether using a different scoring framework would reduce delta for those runs where it is not active, but it seems highly likely that putting some priority on jobs due today (over those due later) would result in lower deltas than not doing so. What is more uncertain is the size of this effect. The resource balancing within the Ofcom RPM is different to the prioritisation used in the Openreach model, and also different to the simplified configuration of that model that Openreach is using within much of the waterfall chart. A lack of prioritisation in the Openreach model will mean jobs are simply addressed in order of arrival, which is not the case in the Ofcom RPM because resources are balanced in that model between separate queues for CL1 and CL2.

The simplified model of Openreach may not be a good representation of the Ofcom RPM

We have already commented on the (lack) of use of prioritisation in the simplified version of the Openreach model which Openreach uses to try to mimic the behaviour of the Ofcom RPM.

Another over-simplification is to assume that the “minor fails” in the Ofcom RPM are best represented as 1.05 visits per job in the Openreach model. Multiple visit jobs in the Openreach model are different to minor fails in the Ofcom RPM because minor fails in the Ofcom RPM can only be reattempted the next day (whereas a second job visit in the Openreach model can in principle be done as soon as the relevant resource can be found).

As a result of constructing a “straw man” version of the Ofcom RPM, the specific impact of moving to more realistic assumptions may be being overestimated. This deficiency of the waterfall chart is not easy to remedy without running different scenarios (e.g. using prioritisation) to test whether these effects are material, which would be time consuming.

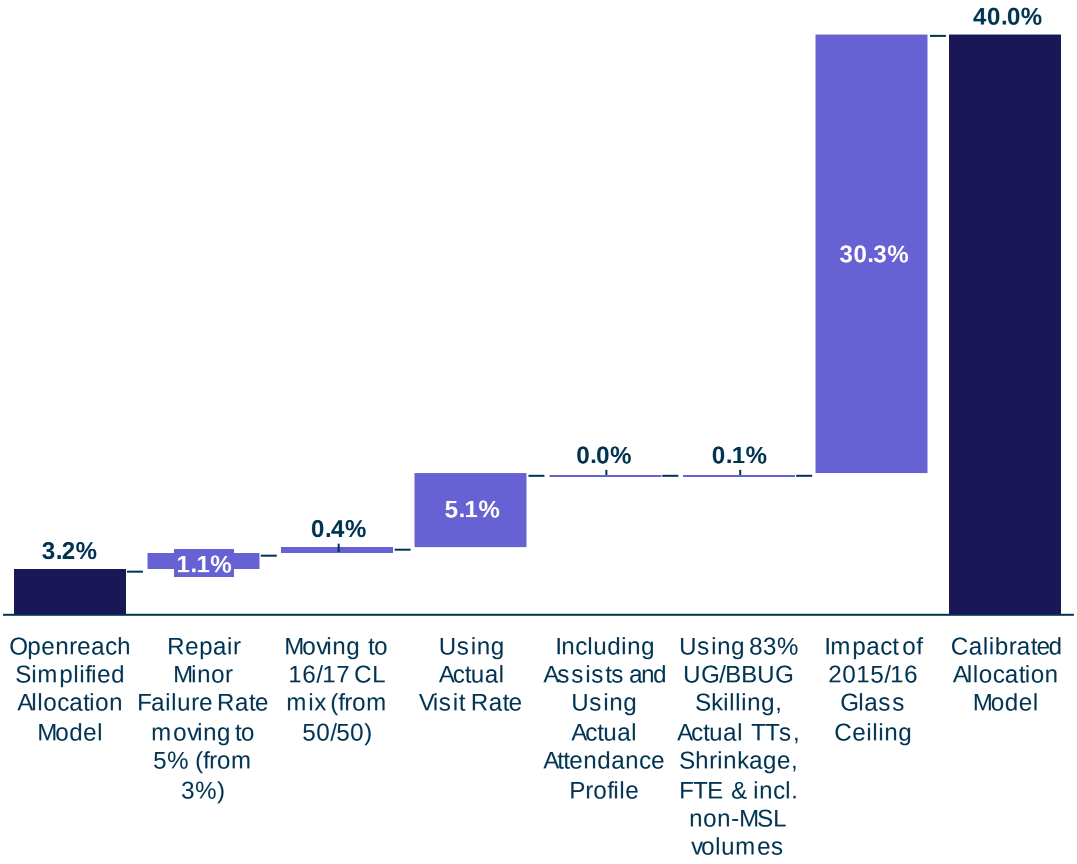
6.2.2 Revised waterfall chart analysis

If we assume that each scenario can be used to generate its own 80% performance baseline, then we can estimate delta for 90% performance in each scenario that reaches this performance level. In other words, we report delta in that HLS as relative to the resources required to give 80% performance in the same HLS.

Scenarios that do not meet the target 90% performance level for CL2 in each GM are combined with subsequent steps (i.e. we do not look at the impact of the parameter changed in that HLS in isolation). This ensures that each step has a valid and comparable data point on which to base its estimated step in delta.[31]

The resulting waterfall chart is illustrated in Figure 6.2, with the details underlying the chart shown in Figure 6.3 below.

Figure 6.2: Revised waterfall chart [Source: Analysys Mason (using Openreach model), 2017]



31 Note that we ignore any cases where the CL1 target is missed at the 80 % performance level. If we were to correct for this then the calculated deltas may decrease somewhat for the first few steps in the waterfall. However, it is unlikely that the effect would be material enough to alter our conclusions.

Figure 6.3: Details underlying the revised waterfall chart [Source: Analysys Mason (using Openreach model), 2017]

HLS	Openreach scenario description (or equivalent for non-Openreach scenarios)	Glass ceiling	Step in delta	Delta relative to 80% performance in the same scenario	Calculated delta for 90% performance (from our model runs) relative to post-process script baseline	Worst GM CL2 performance
82	Openreach Simplified Allocation Model - uplift curve base	98.6% Provision, 97.0% Repair	3.2%	3.2%	4.3%	[✂]
100 (different post processing)	Openreach Simplified Allocation Model - uplift curve 5% Repair Minor Failure	Raised (16/17 CL mix) 92.6% repair	1.1%	4.3%	7.0%	[✂]
106	A-M alignment Waterfall - Care Level Mix moving from 50/50 to 16/17 CL Mix	Raised (16/17 CL mix) 92.6% repair	0.4%	4.7%	7.8%	[✂]
111	A-M alignment Waterfall - Using Actual Visit Rate	Raised (16/17 CL mix) 92.6% repair	5.1%	9.8%	27.2%	[✂]
108	A-M alignment Waterfall - including assists and using actual Attendance Profile	Raised (16/17 CL mix) 92.6% repair	0.0%	9.8%	25.9%	[✂]
112	A-M alignment Waterfall - using 83% UG/BBUG skill mix Actual TTs, Shrinkage, FTE & including non-MSL volumes	Raised (16/17 CL mix) 92.6% repair	–0.1%	9.7%	11.0%	[✂]
77	Calibrated Allocation Model - Year 3 uplift curve	Actual (16/17 CL Mix) – 87.5% Repair	30.3%	40.0%	41.5%[32]	[✂]

32 Note that we also have another run of high level scenario 77 with a different random seed and slightly different results, but for simplicity we show the result for just one run here

6.3 Conclusions on the waterfall analysis

Several conclusions can be drawn from this revised waterfall analysis:

- It is apparent that delta for 90% performance is strongly dependent on the modelling of ‘minor fails’ (which relates to “visits per job” using the Openreach model terminology).
- We already knew that glass-ceiling effects were going to be important and this seems to be borne out by HLS 77, which has a much higher delta for 90% performance than the step between any other scenarios. This higher step in delta might relate to its changed glass ceiling (i.e. major fails) relative to HLS 112. High deltas indicate we must be close to the glass ceiling; the fact that not all GMs reach the target performance indicates we may be above the glass ceiling (and that we would need to interpret that delta very carefully).
- Some of the results (estimated changes in delta) are small relative to the run-to-run variability of the model and accuracy limits imposed by the lack of interpolation in the post-processing (which limits resolution to one modelled step in resources). Many of the steps in the waterfall give small changes in results (say 1%) and, are likely not to be useful estimates of the impact of those changes, given random fluctuations will apply to both the measured and baseline values. All they tell us is that the impact of some of these factors (applied in a specified order) is comparable to the typical uncertainty in model result. We could reduce the variability using multiple sets of model runs for a high-level scenario (e.g. multiple random seeds), but this is too time consuming given the current model run-time.

7 Additional scenario results

In this section, we report the results of a number of additional scenarios that we have run using the Openreach model. The aim of these scenarios is to investigate the behaviour of the model in relation to certain parameters of interest to Ofcom.

7.1 HLS 200: Clean 15/16 baseline

Estimating delta in resources is always relative to some baseline figure.

As we have noted in our discussion of the Openreach waterfall analysis, the calibration data point needs to be representative of the situation in 15/16 because it is actual 15/16 performance that we can measure. The Ofcom RPM also takes this approach, measuring resource deltas relative to 15/16 model resources needed to hit the actual performance (80%) in the 15/16 12-day FAD case.

Accordingly, we have constructed a scenario to provide a clean 'baseline' for 15/16 performance using the actual 15/16 care level mix, 15/16 glass ceiling (87.5%), and 15/16 skill levels. We call this HLS 200. We use the Openreach post-processing that allows for overtime ("Smart Resourcing Uplift Curve").

The 80% performance resource level for HLS 200 implies a resource delta of 1.3% on actual resource levels in 15/16, whilst a further 8.8% uplift in resources is required to achieve 90% performance. The results are summarised in Figure 7.1 below.

Figure 7.1: Results of clean 15/16 baseline scenario [Source: Analysys Mason (using Openreach model), 2017]

Performance level	Scenario 200 resource level relative to same baseline as HLS 77 (15/16 actuals)	Relative to 'clean' baseline of 80% performance
80%	1.3%	0% (by definition)
90%	10.1% (Note: CL1 does not hit 90% target at this level)	8.8%

7.2 HLS 203: Representative case using 16/17 care level mix, alternative raised glass ceiling ('red' elements only)

We have also run a scenario using the 16/17 care level mix, higher skill levels (83% UG/BBUG) and an alternative raised glass-ceiling input, specified by Ofcom. This 'alternative raised glass ceiling' includes only the "red" items shown in Table 9 of the Openreach consultation response, with the exception of clear rejects, which are assumed to be abolished and therefore contribute 0% to the major fails. This results in major fails of 6.71% and a corresponding glass ceiling of 93.29%.

We call this HLS 203; it is similar to Openreach's HLS 112 except for the changed "glass ceiling". The deltas relative to the 15/16 actuals and relative to the clean baseline are shown in Figure 7.2 below.

Figure 7.2: Model results for HLS 203 [Source: Analysys Mason (using Openreach model), 2017]

Performance level	HLS 203 resource level relative to same baseline as HLS 77 (15/16 actuals)	Relative to clean baseline 80%
80%	1.4%	0.1%
90%	18.4%	17.1%

We do not show higher performance levels as they are not reached in all GMs.

However, the delta here is materially higher than for 90% performance in the near-equivalent Openreach HLS 112 (for which the 90% delta in our calculation is 11.0% relative to the post-processing script baseline) and the various runs with different care level mixes (see below), all of which suggest that delta for 90% performance should be of the order of 10%.

7.3 HLS 213-217: Effect of varying care level mix

We have run variants of HLS 203 (16/17 CL mix, alternative raised glass ceiling ("red only"), 83% UG/BBUG skilling) with different CL mixes (same numbers of jobs as HLS 203, but with jobs assigned at random to CL1 or CL2 to achieve the required CL mix). As in the HLS 200 and 203 cases above, the post-processing includes the overtime adjustment. We call these scenarios HLS 213 to 217; results are as shown in Figure 7.3 below.

Figure 7.3: Model results for HLS 213-217 (varying care level mix) [Source: Analysys Mason (using Openreach model), 2017]

HLS	CL1:CL2 ratio	National resource delta for 90% performance in worst GM (relative to baseline of Openreach post-process script)	National resource delta for 90% performance in worst GM (relative to 80% performance in the 'clean baseline')
213	30:70	11.2%	9.9%
214	40:60	11.2%	9.9%
216	60:40	10.8%	9.6%
217	70:30	10.5%	9.3%

As the overall confidence intervals on these values are of the order of 1%, in effect there is no measurable change in delta if the care level mix changes. This is counter-intuitive. We would expect increasing the fraction of CL2 (which has a tighter time constraint, and which partly relies on Saturday morning working by a fraction of the workforce) makes it more difficult to hit 90% performance, and this expected effect was present in the Ofcom RPM results presented in Ofcom's consultation (as well as Openreach's distribution model that was used to inform charge controls for the current period).

The calculated deltas would in our view be likely to be overestimates (for reasons we discuss in Section 8). Although many of these factors may be similar across the different CL mixes, some may be varying for different CL mixes (e.g. the extent to which CL1 overperforms may be varying as the care level mix changes).

This is an example of a sensitivity analysis that may be of interest to Ofcom where both the Ofcom RPM and the Openreach model are capable of modelling the effect but give quite different results.

We also note that the results are also inconsistent with those of HLS 203 which ought to lie within the range (as the 16/17 CL mix is of the order of 50–60% CL1), particularly as these scenario inputs are otherwise very similar to HLS 203. It is hard to understand how this disagreement with HLS 203 arises: one possibility is that in HLS 203 CL1 and CL2 jobs are correlated in some way that is lost by the assignment of jobs to a random care level in the data pre-processing for these scenarios. We know that there are such correlations which will be lost, for example there is in 16/17 much more CL1 in areas where unbundlers are present (i.e. because other operators purchase MPF from Openreach at CL1), which means that a uniform care level mix means more use of CL2 in urban areas and more use of CL1 in rural. But similar effects would also be present in the Ofcom RPM in the varying care level mix cases.

8 Assessment of limitations, biases and uncertainty in the results vis-à-vis the Ofcom RPM

In this section, we provide an assessment of the limits and biases in the Openreach model as well as the level of errors and uncertainty in the results in relation to the Ofcom RMP.

- We begin in Section 8.1 by describing some of the limitations and potential biases inherent in the Openreach model and compare these to the assumptions and limitations made within the Ofcom RPM. Our analysis concludes with a summary of the potential biases contained within the Openreach model.
- In Section 8.2 we then go on to consider the level of uncertainty that we are able to observe in the results of the Openreach model through analysis of some of the model outputs that we have obtained. We conclude by identifying the key input parameters, which are most likely to drive the model results.

8.1 Modelling assumptions, limitations and biases

A number of issues of concern were noted in the Openreach model, mostly around job and resource handling, but also around running the model, which are described in greater detail below. We begin by considering aspects in which the Openreach model is more detailed than the Ofcom RPM before looking at ways in which the Openreach model is less capable (or configured less capably) than the Ofcom RPM.

8.1.1 Aspects in which the Openreach model is more detailed than the Ofcom RPM

There are a variety of areas where the Openreach model is more detailed than the Ofcom RPM including:

- job durations
- FTE staff level fluctuations
- staff skills
- consideration of overtime
- glass-ceiling inputs varying by GM area
- multiple visits per job.

The last of these is certainly having an effect on outputs, as the waterfall analysis presented in Section 6 highlighted, although the assumption made in the Ofcom RPM is different to that modelled by Openreach. Our amended version of Openreach's waterfall analysis shows that the remainder of these factors are unlikely to be significant in modifying the results.

Including overtime is reflective of reality but gives rise to complications in the post-processing and may contribute to model outputs being difficult to interpret.

In particular, using the post-processing to account for changes to overtime may not be the best approach, because post-processing is simply not equivalent to changing the resource levels in the model (if for no other reason than if the resources are modelled as present, a different number of jobs can be served, which will affect the performance).

Alternative approaches might be:

- modifying the way in which overtime is modelled (e.g. turning off overtime when using the model to estimate resource deltas, or extending the list of types of jobs that will be served with overtime if insufficient overtime is being used, noting that this might not be how Openreach would work in practice today), or
- calculating a metric of total cost which incorporates the costs of resources and the different costs incurred by using overtime (if less is used, less cost will be incurred).

8.1.2 Ways in which the modelling is less capable (or is configured less capably) than the Ofcom RPM

Four principal aspects of the Openreach model are less capable than the Ofcom RPM:

- Job-handling limitations, primarily relating to the handling of 'stress', mean that the Openreach model may not adequately prioritise repair jobs to achieve improved performance.
- Resource-handling limitations, primarily relating to the handling of each SOM in isolation, mean that 'loans',[33] which may improve performance, are not taken into account.
- There is no explicit mechanism within the Openreach model, beyond the prioritisation rules, to balance performance across different job types, meaning that resources required to achieve performance targets across all job types may be overestimated.
- The complexity of the process of model configuration means that it is hard to guarantee that errors are not introduced in the running of the model.

We consider each of these aspects in turn below.

Job-handling limitations

Some aspects of job handling are opaque or implicit within the Openreach model. There is no explicit diary for provisioning jobs, which means the impact of increased resource on provisioning (i.e. that there will be more slots available for appointments sooner) will be poorly modelled.

In the Openreach model, there is no account taken of the 'no-show' element of failed jobs, and it is unclear if these jobs are taken out of the statistics. We suspect that they are not, which will make a material difference when the model runs close to the glass ceiling. It is however likely that this only applies to appointed jobs, and is therefore less of a factor for fault repair than was previously assumed within the Ofcom RPM.

33 Loans is the term we use to denote Openreach engineers being appointed to jobs outside of their home SOM, in order to provide assistance in other SOMs where performance levels are struggling to meet targets. These are modelled as part of the Ofcom RPM, but are not taken into account by the Openreach model.

Frames jobs are mixed into the result statistics, which is a different approach from the Ofcom RPM and will result in lower deltas for achieving similar performance levels while also overstating performance relative to the Ofcom RPM for a given level of resource. In other words, a 90% target including frames jobs only requires a lower target of the non-frames jobs to be met.

The main job-handling limitation we are concerned about though is the approach to modelling stress. The impact of stress is modelled through a step change in various parameters rather than being gradual. However, if Openreach believes that the model calibrates better to the real world with the impact of stress incorporated, it is unclear why it has been turned off for the scenarios presented. While it is true that deprioritising Broadband Boost and SFI under stress might be overly flattering to WLR/MPF/NGA deltas, one might nonetheless expect provisioning to also be deprioritised under stress.

The modelling of stress and in particular how this affects modelling of provisioning appointing therefore differs in some important ways from the Ofcom RPM. In the Ofcom RPM, there is in effect an appointment book for the future days provisioning jobs, and incoming provisioning jobs are ‘booked’ into these future slots (and then have priority and will be resourced on the relevant day).

In the Openreach model, the ‘stress’ factor can be used to change the distribution of appointment times (relative to the time a job arrives). However, firstly this stress capability is not actually used (i.e. it is configured to be off) in the runs Openreach is using to provide quantitative output in its consultation response. And secondly, even if it were used, the stress factor is a binary one (i.e. the appointment time distributions for a given scenario, job type, etc., are selected from two options depending on the current level of stress).

In the Ofcom RPM, the current level of demand (i.e. the length of the queues), the expected performance level and the ‘bias factors’ all combine to determine how many provision appointments will be offered on future days, and this can vary adaptively over time according to the degree of demand for resources (noting that the effect is medium term, as changing future levels of diary slots has no effect on repair resource availability for several days). Under greater repair stress, the Ofcom RPM will make larger adjustments to the future number of provision appointment slots.

Resource-handling limitations

The Openreach model calculates one SOM at a time, so it does not know how many resources are spare and can be loaned to a neighbouring SOM.

The modelling of ‘loans’ is only possible in the Openreach model via a modified version of the FTE assumptions in a given SOM and week (meaning it is not driven by the availability of additional resource in a given SOM on a given day). We note as an aside that changing the model to allow loans to be modelled based on staff availability would be a very major task to implement within the model as the model is not running each SOM for each day at the same time (instead it runs them in series). The Ofcom RPM by contrast was designed to allow loans to be modelled.

Also, the quantisation of resource levels being tested might lead to a slight overestimate to resource levels. Resource levels are tested in fixed increments (defined by the set of scenario runs used). This will tend to slightly overestimate resource requirements as there is no interpolation between different resource level results, as is done in the Ofcom RPM. However, the actual impact on deltas is unclear.

Balance of performance between job types

The Openreach model tends to result in both provisioning and CL1 significantly outperforming CL2, at high performance levels. Unlike the Ofcom RPM (which has 'bias factors') there is no mechanism that has been used to push the model towards spending more of its resources on CL2. One possibility is that the existing prioritisation mechanism could be re-configured to do this (e.g. give CL2 jobs that are due tomorrow more priority than overdue jobs). Openreach have explained the prioritisation they use, but have not explained why these choices of prioritisation parameters were made.

Model configuration complexity

The Openreach model set-up makes it prone to human error. There is no user interface to run the model or manage database entries.

It is not clear how some of the input parameter values are derived. Inspection of the category distributions used to implement the glass ceiling (major fails) shows that the ratios of the fails to 'normal' CL1 and CL2 jobs are not that different in the scenarios which Openreach has described as 'raised glass ceiling 16/17 CL mix' and 'Actual 16/17 CL mix'.

The Openreach model can be used with some scenario runs having failed (e.g. due to the bug, or missing data). Given that a scenario needs thousands of runs (40 resource levels for 56 SOMs) for just one random seed, it is not obvious how these errors are reviewed or eliminated by Openreach before using the results. By comparison, the Ofcom RPM stops with an error and writes no output file in various circumstances, making it obvious to the user that action is required by failing to produce results for all resource levels for that scenario.

Finally, there are no signs of configuration control on the code of the scripts used in scenarios of importance. While this is to some extent understandable since it is an experimental/test system, it would be advisable for this to be formalised now that Openreach is relying on the results.

8.1.3 Other considerations

We note a number of considerations relating to the modelling approach which may affect the results of the Openreach model, but do not obviously bias them in any particular direction.

Job type distributions of different kinds may not be taking correlations into account, which might exist in the real data.

It is also not obvious whether any rounding effects exist as a result of the sampling, which may affect actual jobs per hour or ratios of jobs of different types (it is clear that sometimes rounding affects the number of engineers with specific skills, as some runs with slightly different levels of resource give identical outputs).

When setting the care level mix to 100% CL1, the Openreach model assumes no Saturday working, whereas the Ofcom RPM assumes the current level of Saturday working and allows CL1 handling on Saturday in that case. Neither is wrong, they are just different.

8.1.4 Summary of potential biases within the Openreach model

Results from the Openreach model are likely to overestimate delta in the light of several of the effects noted above. In particular:

- The lack of balancing of resources across different job types (provision, CL1, CL2) means that CL1 often appears likely to more than meet its targets at a given level of resource uplift, which suggests that a different use of resources could hit the required CL2 performance at a lower level of total resource.
- The approach to appointment scheduling or job prioritisation in response to stress, because all response to stress is generally turned off in most scenarios.
- Loans, according to need and resource availability, are not taken into account in the Openreach model.

On the other hand, it is possible that the overtime adjustments are leading to too low a delta.

8.2 Summary of results for additional scenarios, for 90% performance

Figure 8.1 below summarises some of the most interesting, and difficult to explain, results that we have been obtained from our model runs. In all cases, the “Smart” script has been used to adjust for overtime.

Figure 8.1: Summary of results obtained from key HLS [Source: Analysys Mason (using Openreach model), 2017]

HLS	Openreach scenario description (or equivalent for non-Openreach scenarios)	Glass ceiling	Glass ceiling (category distribution ID)	Care level mix	Skilling	Calculated delta for 90% performance (from our model runs) relative to post-process script baseline	Target performance level actually achieved?
77	Actual 16/17 CL Mix	Actual for 15/16 – 87.5% Repair	1111	Actual 16/17	83% UG/BBUG	38.6% (average over 2 runs)	No, neither run hits target, reaching [✂] and [✂] for CL2 in the worst GMs
112	Raised 16/17 CL Mix	Raised (Openreach) – 92.6% Repair	1911	Actual 16/17	83% UG/BBUG	11.0%	Yes
200	Analysys Mason 15/16 clean baseline	Actual for 15/16 – 87.5% Repair	311	Actual 15/16	Actual 15/16 (61%)	10.1%	No, [✂] for CL1
203	Analysys Mason	Raised (Analysys Mason – i.e. 93.29% Repair, due to 6.71% red failures)	3911	Actual 16/17	83% UG/BBUG	18.4%	Close ([✂] CL2)

.

From these results, we make the following main observations:

- HLS 112 has a very much lower delta than HLS 77. The scenarios are identical apart from a change to the glass ceiling. In principle, this change could have a significant effect. In particular, for HLS 77 the glass ceiling, which we understand is intended to be 87.5%, is lower than the 90% target performance level, which means it would be impossible to achieve and could lead the model to produce very high deltas. However, our analysis of the category distribution file suggests that the change to glass-ceiling inputs is actually (surprisingly) minimal. This difference in delta is therefore very tricky to understand.

- HLS 200 also produces a very much lower delta than HLS 77 as a result of a (small) change in care level mix and a change to skilling. The behaviour in HLS 200, whereby CL1 becomes a limiting factor, is odd. This could be driven by skilling. The huge fall in delta may be, at least in part, explained by CL1 failing to hit target in HLS 200. If CL1 is forced to hit target then delta greatly increases to ~28%. However, this is still materially lower than the delta for HLS 77, which itself does not hit target for CL2.

- HLS 203 having a higher delta than HLS 112 is counter-intuitive since the only change is a (further) raise of the glass ceiling for HLS 203, which should lower the delta. However, we noted in Section 7.3 that when care level mix is changed either side of 16/17 actuals from HLS 203, we get a delta that falls to around 11% (i.e. similar to HLS 112) in all cases.

The results that we have obtained suggest a high level of uncertainty in the outputs of the model. Model runs with quite similar inputs can give quite widely varying results, especially if the target performance is close to the glass ceiling. These results can be hard to interpret because of the significant variations in delta from apparently small changes in inputs. Whilst this uncertainty and difficulty of interpretation do not necessarily mean that the model is producing incorrect results, it does mean that the results should be treated by Ofcom with an appropriate level of caution.

Our analysis suggests that there are likely to be two key parameter sets within the Openreach model that have the most significant effect on the size of the deltas:

- the glass-ceiling assumptions (or in Ofcom RPM’s terminology the proportion of jobs that are major fails)
- assumptions relating to multiple visits – i.e. the distribution of the number of visits required to complete each job (modelled through the minor fail rate in the Ofcom RPM).

Furthermore, we expect that the impact of loans, which are not considered within the Openreach model, is also significant in determining the size of the calculated deltas. We explore the impact of these parameters further in Section 9, through further consideration of the Ofcom RPM.

9 Further outputs from the Ofcom RPM

In this section, we provide outputs from some further runs of the Ofcom RPM in order to test scenarios relating to the three key parameters identified in Section 8, including the impact of ‘loans’, which are not taken into account within the Openreach model but potentially have a significant effect on deltas.

9.1 Using the Ofcom RPM to test various additional scenarios of interest

We have used the Ofcom RPM to consider:

- the impact of revised assumptions for major and minor failure rates
- whether the Ofcom RPM could be used to estimate the impact of variation in the major failure rate across GMs (the “worst GM glass ceiling” effect), and
- the impact of loans.

We have assumed:

- a higher level of major fails based on the Openreach glass-ceiling analysis “red” items
- slightly higher minor fails than the previous 5% case (based on the visits per job statistics and some assumptions about how many of these will be “on the day” fails requiring a visit on the next day) – 5.45%.
- a new target performance level of 90% (compared to the previous 93%).

We continue to use the 80% 12-day FAD baseline from the calculations provided in our previous report on the Ofcom RPM.[34] This is an approximation; we could, in principle, re-calculate the 80% baseline for a 12-day FAD based on a revised minor fails assumption and 15/16 major fails and CL mix.

90% performance with 5.45% minor fail and varying levels of major fail

The waterfall analysis discussed in Section 6 showed that the use of ‘actual visit rates’ within the Openreach model had a significant effect on deltas. The approach taken to model multiple visits within the Ofcom RPM is different: the minor failure rate is the relevant lever.

Openreach data within the model (table TBL_AM_VISITS_PER_JOB_DISTNS) suggests that approximately [✂] of jobs require two visits, a further [✂] require three visits and a further [✂] require four or more visits. The minor fails in the Ofcom RPM are those jobs that cannot be completed on the day that they are first attempted. However, it is important to distinguish between ‘on-the-day failures’ and ‘failures at the first attempt’. As part of its waterfall analysis, Openreach

[34] See https://www.ofcom.org.uk/__data/assets/pdf_file/0034/99646/Analysys-Mason-report.pdf

appears to characterise the minor fails within the Ofcom RPM as corresponding to the latter, when in fact they correspond to the former.

From the Openreach data, it is possible to estimate the proportion of 'on-the-day failures' and use this as an input assumption for minor failure rate in the Ofcom RPM. A proportion of jobs requiring two visits will be on-the-day failures, particularly those that are first attempted towards the end of the day (meaning that there is no time for a further attempt to be made on the same day). The same applies for jobs requiring three visits, though the proportion which will be attempted early enough to avoid becoming an on-the-day failure is likely to be materially lower. Jobs requiring four (or more) visits are almost certain to be on-the-day failures.

Making different assumptions about the proportions of each category of jobs that end up as on-the-day failures we derive a range of between 4.7% and 6.2% of jobs that are likely to be on-the-day failures.[35] We take the mid-point of this range, 5.45%, as an estimate of minor fails to use within the Ofcom RPM for our additional model runs.

With loans turned on, and a 5.45% minor fail rate, the resource delta for 15/16 to raise an 80% performance level to hit a 90% target (at the baseline CL mix) depends strongly on the major fail rate assumed, as shown in Figure 9.1 below.

Figure 9.1: Resource delta required for 90% performance under different assumptions on major fail rate [Source: Analysys Mason (based on Ofcom RPM), 2017]

Major fail rate	Delta for 90% performance relative to resources required for 80% performance in 15/16 from our previous report
5%	6.5%
6.71%	20.1%

These results show that small differences in the assumed major fail rates will have a very significant impact on the modelled delta (i.e. when we are close to the glass ceiling, the results are very sensitive to the inputs).

We test two major fail rates: 5% and 6.71%. Our choice of 6.71% is based on the "red" elements of the glass ceiling Openreach is arguing for in its consultation response (i.e. the 'alternative' raised glass ceiling). Our 5% value is based on [✂].[36]

At roughly 20% delta, the 6.71% major fail case delta is notably higher than the previous delta results for the 93% performance case (set out in our previous report). We also note that the 20%

35 [✂].

36 Care levels 6 and 7 denote the jobs that the Openreach model classifies as 'automatic fails' for care levels 1 and 2 jobs respectively (i.e. the jobs for each care level that are part of the glass ceiling).

delta is higher than the deltas we have produced from using the Openreach model with the ‘alternative’ raised glass ceiling.[37]

Higher level of major fails/worst GM glass ceiling

Openreach argues that the glass ceiling (major fails) is materially different in different GM areas.

In reality, the additional resources to deal with this issue would be targeted on the worst-performing GMs. However, the Ofcom RPM assumes that additional resources (before loans) are spread at the same ratio as existing resources across the GMs; if the GMs have very different major fail rates and performance this means we are not getting the right ‘bang per buck’ from additional resources (i.e. not all additional resources are being used in the way that maximises the increase in performance).

As a result, it would not be effective to estimate the impact of this effect by increasing the major fails across the board. One option would be to make the major fail percentages vary by GM in the Ofcom RPM, which would mean some minor changes to the code to enable this variation.

This an example of an issue that is currently dealt with more naturally in the Openreach model because it models the resources needed to hit the target performance in each GM (and already sets a different glass-ceiling value for each GM via the input data).

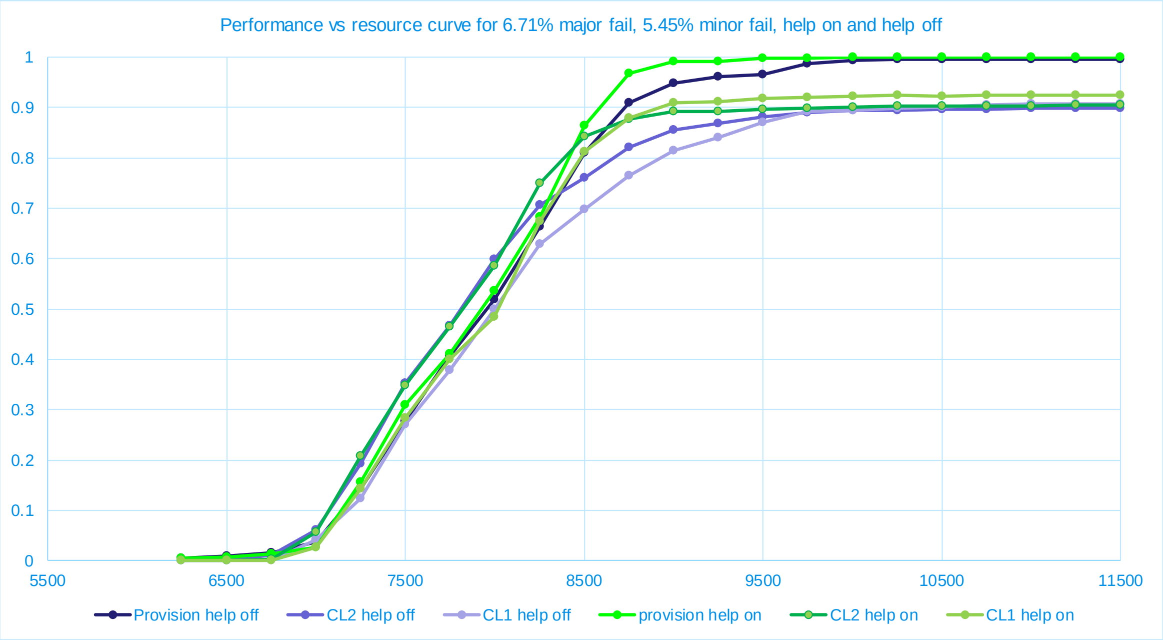
Impact of turning off loans

With loans turned on, the resource delta for 15/16 to raise an 80% performance level relative to a 12-day FAD case using a 6.71% major fail rate is 20.1%.

Figure 9.2 below shows a comparison between the performance versus resource curves for provision, CL1 and CL2 repair jobs in the two cases with loans turned on and off.

37 In this context, we note that the dataset sent to us by Openreach (category distribution IDs “3911-4057”) appears to have only 4.18–5.70% (varying by GM, average 5.05%) automatic fails in a similar manner to the discrepancy in glass-ceiling input data we noted for Openreach’s glass-ceiling scenarios in Section 4.10. If the glass ceiling were set at too high a level, this might explain why the Ofcom RPM produces a higher delta with 6.71% major fails.

Figure 9.2: Impact of turning loans off [Source: Analysys Mason (using Ofcom RPM), 2017]



With loans turned off (purple lines), the deltas are substantially increased. For example, at this level of major fails, a 90% target (at the baseline CL mix) is not reached at the highest resource level tested (11 500, equivalent to approximately 40% delta). It is possible that this materially higher delta could be reduced by changing the major fail rate or other model parameters such as bias factors, but it does show that loans may be critical to include in accurate modelling of delta at the levels of performance being modelled.

Whilst the exact parameters that should be used to model loans (criteria for lending resources, level of inefficiency in doing so, etc.) are open to debate, it is certainly true in practice that Openreach does use loans as part of its service delivery.

We note that the increase in delta with loans turned off is dominated by just a couple of GMs. Part of this effect might be due to the way in which the Ofcom RPM allocates additional resources: the lodge loan facility may be (in effect) partly balancing resources across GMs if there is some inefficiency in the initial (fixed) allocation.

9.2 Conclusions gained from further use of the Ofcom RPM

It is useful to be able to test deltas coming from the Openreach model by comparing them to values produced by a similar but independent system, the Ofcom RPM. This is particularly the case when the Openreach model system is complex, slow and difficult to configure, which means that ad-hoc tests of specific parameters of interest are not easily conducted.

The much simpler process of configuration (fewer input parameters) and the higher speed of running the Ofcom RPM mean that a variety of tests can be designed, run and analysed relatively quickly.

Some effects are dealt with naturally in one model, others are more naturally considered in the other. Loans, for example, are dealt with more naturally by the Ofcom RPM, whereas the impact of GM-to-GM variation in glass ceiling (major fails) are dealt with more naturally within the Openreach model.

10 Conclusions

In this section, we draw conclusions on the extent to which the Openreach model is likely to be a reasonable approach to estimating the outputs of interest to Ofcom.

Methodology

The Openreach model takes a different and much more detailed approach to modelling the relationship between QoS and required resource compared to the Ofcom RPM. Similarly to the Ofcom RPM, some methodological issues need to be taken into account when considering the Openreach model.

There are a variety of areas where the Openreach model is more detailed than the Ofcom RPM including in relation to job durations, staff level fluctuations, staff skill levels, consideration of overtime, glass-ceiling inputs and taking multiple visits per job into account. However, in some cases the effect on outputs is likely to be limited, whilst in others there are implementation complexities that make it unclear whether the extra detail is beneficial – the approach to modelling overtime using the post-processing being a prime example.

At the same time, there are several areas where the Openreach model is less capable (or configured less capably) than the Ofcom RPM:

- Job-handling limitations, primarily relating to the handling of ‘stress’, mean that the Openreach model may not adequately prioritise repair jobs to achieve improved performance. Indeed, the part of the model that takes stress into account is turned off in the majority of scenarios reported by Openreach.
- Resource-handling limitations, primarily relating to the handling of each SOM in isolation, mean that ‘loans’, which may improve performance, are not taken into account.
- There is no explicit mechanism within the Openreach model, beyond the basic prioritisation rules, to balance performance across different job types, meaning that resources required to achieve performance targets across all job types may be overestimated.
- Model configuration limitations mean that it is hard to guarantee that errors are not introduced in the running of the model.

Implementation and usability

The Openreach model is, broadly speaking, well constructed.

There is a bug which crashes a small percentage of runs. This somewhat reduces our confidence in the correctness of the design.

The model documentation is not sufficiently detailed, and there is no user interface (other than the development tools for Python and the chosen SQL database), making it difficult to install, configure and run. It is complex, slow and produces very large amounts of data (most of which is used once in post-processing before being discarded). It is also, in certain cases, unclear[38] how the actual input data reflects our understanding of the input parameters. For example, the values of the important glass-ceiling inputs in the model are not clearly linked to the “actual” or “raised” values in Openreach’s consultation response (in fact, the “category distribution” ratios are almost identical in scenarios described by Openreach as having “actual” or “raised” glass ceilings).

It is unlikely that Ofcom staff could make effective use of the model in its current form, if only due to the complexity, the relatively minimal documentation, and the time it takes to understand and configure the model and produce results, making sensitivity analysis very slow.

Outputs

We have been unable to reproduce in detail the outputs of the runs Openreach has performed, even starting from almost identical inputs. This may be due in part to the fact that the model will (due to the way in which it uses random numbers) give slightly different results on different detailed versions of Python, even if the random number seed and the PYTHONHASHSEED are controlled.

Model runs with quite similar input sets, differing only in one or two input parameters (potentially with only relatively small changes in those inputs), can give quite widely varying results (in terms of the resource deltas to achieve a certain target performance level), especially if the target performance is close to the glass ceiling. Particular issues arise where the target performance level is not met for some job types (e.g. CL2). Figure 10.1 provides a summary of the results obtained from running four of the key HLS.[39]

Figure 10.1: Summary of results obtained from key HLS [Source: Analysys Mason (using Openreach model), 2017]

HLS	Key parameters defining scenario	Calculated delta for 90% performance (from our model runs) relative to post-processing script baseline	Target performance level actually achieved?
77	Actual 16/17 CL mix, Actual 15/16 glass ceiling (87.5%), 83% UG/BBUG skilling	38.6% (average over 2 runs)	No, neither run hits target, reaching [✂] and [✂] for CL2 in the worst GMs
112	Actual 16/17 CL mix, Openreach Raised glass ceiling (92.6%), 83% UG/BBUG skilling	11.0%	Yes
200	Actual 15/16 CL mix, Actual 15/16 glass ceiling (87.5%), 61% UG/BBUG skilling (Actual 15/16)	10.1%	No, [✂] for CL1

38 This is not to say that the implementation is necessarily incorrect.

39 Note that more detailed scenario descriptions are provided in the equivalent table in Figure 8.1.

HLS	Key parameters defining scenario	Calculated delta for 90% performance (from our model runs) relative to post-processing script baseline	Target performance level actually achieved?
203	Actual 16/17 CL mix, ‘Alternative’ Raised glass ceiling (i.e. 93.29%, due to 6.71% red failures), 83% UG/BBUG skilling	18.4%	Close ([✂] CL2)

These results can be hard to interpret due to these significant variations in delta from apparently small changes in inputs. Some results are particularly difficult to explain (e.g. why HLS 203 has higher deltas than the otherwise similar HLS 213-17, described in Section 7.3). Very high deltas arise in some scenarios (e.g. HLS 77), which may be due to the targeted performance being above the glass ceiling.

This does not necessarily mean that the results are incorrect or that the model is unreliable, but rather that the sometimes counter-intuitive results and the difficulty of replicating outputs produced by Openreach mean that Ofcom needs to treat the results with an appropriate level of caution.

Key sensitivities

The waterfall analysis produced by Openreach as part of its consultation response is flawed in several ways.

Although in relation to ‘minor fails’, the Ofcom RPM modelling does not actually function as Openreach seems to assume (due to the delay in reattempting failed jobs), a revised approach to the same waterfall analysis still shows that the modelling of minor fails is important. However, it does not show factors other than this and the glass ceiling (‘major fails’) to be causing material changes in the overall delta to reach 90% performance.

Furthermore, our analysis of the Ofcom RPM shows that the impact of loans, which are not considered within the Openreach model, is a third significant factor in determining the size of the deltas.

The use of steps in the modelled resource levels within the Openreach model means that deltas estimated using that model may have an additional variability of the order of 0.5%.

Summary

The Openreach model takes a different approach to the Ofcom RPM and can be considered complementary to it in some ways.

In our view, results from the Openreach model are likely to overestimate delta for several reasons. In particular:

- The lack of balancing of resources across different job types (provision, CL1, CL2) means that CL1 often appears likely to more than meet its targets at a given level of resource uplift, which suggests that a different use of resources could hit the required CL2 performance at a lower level of total resource.
- The approach to appointment scheduling or job prioritisation in response to stress, because all response to stress is generally turned off in most scenarios.
- Loans, according to need and resource availability, are not taken into account in the Openreach model.

On the other hand, it is possible that the overtime adjustments are leading to too low a delta.

It is useful to be able to test deltas coming from the Openreach model by comparing them to values produced by a similar but independent system, the Ofcom RPM. This is particularly the case when the Openreach model system is complex, slow, and difficult to configure, which means that ad-hoc tests of specific parameters of interest are not easily conducted.

The much simpler process of configuration (fewer input parameters) and the higher speed of running the Ofcom RPM mean that a variety of tests can be designed, run and analysed relatively quickly.

Some effects are dealt with naturally in one model; others are more naturally considered in the other. Loans, for example, are dealt with more naturally by the Ofcom RPM, whereas the impact of GM-to-GM variation in glass ceiling (major fails) are dealt with more naturally within the Openreach model.

As noted above, the outputs (delta to reach 90% performance) we have obtained from the Openreach model vary quite widely, between 10.1% and 38.6% for apparently relatively similar scenarios (see Figure 10.1). By way of comparison, the Ofcom RPM produces deltas ranging between 6.5% and 20.1%, depending on the assumption used for major fails (i.e. the level of the glass ceiling).

The Openreach model is difficult to use and generates some unexpected results. It is therefore unlikely that the model, as it currently stands, could be used in isolation to predict the required resource deltas. The Openreach model may help Ofcom, potentially in conjunction with the Ofcom RPM, to inform a range of potential resource uplifts that may be required to realise a given QoS increase.