

# SDG Audit of Monitoring Point Weightings and Cancellation Minutes

## Recalibration of rail industry performance regime

Rail Delivery Group

12 January, 2018

# Introduction

## Audit remit

- To undertake an Audit of the calculation of Monitoring Point Weightings and Cancellation Minutes that were completed by PwC in Phase 1 of the Recalibration work.
- The methodology for Phase 1 had been agreed with the Recalibration Working Group. Therefore, we are Auditing how the calculations have been completed against an agreed process – we are not auditing the agreed process itself.
- For the Monitoring Point Weightings Audit, there are three spreadsheet models for each TOC: A, B and C. We were only provided with model C, so are unable to check the details within Models A and B.
- For the Cancellation Minutes Audit, For each TOC, there are two STATA code files along with a spreadsheet model. Most of the calculations and processing has been done within the STATA Code.

## Audit terminology

- We assess the calculations of each formula to review the process from inputs to calculated values to outputs.
- We have also categorised the Audit findings into three groups:
- **High Priority:** where we think that there is an error in either the inputs, calculations, or processes that will mean that the result calculated is not consistent with the agreed methodology.
- **Medium Priority:** where we think that because of the way the model/code is set up there is a medium-to-high risk of an error occurring due to a lack of robustness in either the inputs, calculations or processes.
- **Low Priority:** where we think that the approach used could be improved to increase the robustness of the models, for example, to ensure any future changes are entered correctly. These are typically defined as Modelling Best Practice (MBP) improvements.

## Summary – Monitoring Point Weightings

- **High Priority Issues:** We have identified errors that will need to be fixed by the Phase 1 consultants.
- **Medium Priority:** There are places in which the formulas are not correct in the Greater Anglia model. We have tested these and although they do not impact on the results in the Greater Anglia model, could impact on the results in other TOC models, creating a high risk of error occurring. More detailed checks will be needed on these on a TOC-by-TOC basis.
- **Low Priority:** We have identified a large number of places where the models do not follow the principles of MBP. Whilst each of these only creates a low risk of error, the high volume of them means that the overall risk is more substantial. Some of the models are large and take a very long time to calculate, in part, due to the computationally inefficient way in which the calculations are written.

## Summary – Cancellation Minutes

- **High Priority Issues:** The treatment of trains that split and join appears to not be done correctly (see SWR slide). Also errors within the Monitoring Point Weightings process may also affect the CMs
- **Medium Priority Issues:** There are places in which the STATA code doesn't completely align with the documentation. There are also a number of areas where we need further clarification.
- **Low Priority Issues:** We have identified a small number of places where the coding does not follow the principles of MBP. Each of these only creates a low risk of error.
- The coding is relative efficient in its operation and the programmes run within a reasonable timeframe.
- The coding is appropriately adapted to deal with some non-standard conditions (e.g. the Saturday Peak on Wales and Border HL05).

# Calculation Checks and Types of Issues Identified - MPWs

## Calculation checks - MPWs

Check	Status	Reference (Sheet, Range, Cells)			Comment
MPs are correctly identified within the model. (Note we have not seen a list of Mps outside of the model)	To Confirm	[TOC Name]	Column	AI	Some service patterns with a high volume of demand are not captured.
Direction of travel is correctly assigned	To Confirm	[TOC Name]	Column	AB	In the vast majority of cases it would appear that this is correctly assigned. We have identified specific cases for each TOC which need clarification.
Peak and Off-Peak are correctly assigned	To Confirm	[TOC Name]	Column	AC:AD	In nearly all cases it would appear that this is correctly assigned. We have identified specific cases for each TOC which need clarification.
Calculation is based using alighters (i.e. not boarders or on-train demand)	Pass	[TOC Name]	Column	AK:AL	
Calculation assigns previous stations to the MPs	Pass	[TOC Name]	Column	AK:AL	
The data referenced covers all relevant rows in CP6 MPWs sheet	Pass	CP6 MPWs	Column	J:N	
Daily totals are correct	To Confirm	CP6 MPWs	Column	T:Y	Totals exclude cases where last station is not an MP
The data referenced covers all relevant rows in [TOC Name] sheet	Pass	CP6 MPWs	Column	T:Y	
Annual totals are correct	To Confirm	CP6 MPWs	Column	AA:AF	Depends on treatment of Bank Holidays, especially when dealing with only one recalibration year
MP totals are normalised across the data within that Service Group + Peak Type combo	Pass	CP6 MPWs	Column	L	Consistent with guidance - this average across all demand in two years. (Rather than a straight average of 15/16 and 16/17).
Totals for each Service Group add up to 1	Pass	CP6 MPWs	Column	L	
The data referenced covers all relevant rows in [TOC Name] sheet	Pass	Station Index	Column	AC	



# High Priority issues

## High Priority

- **The high priority issue(s) we have identified so far:**
- Some services that have a large number of passengers travelling on them are not monitored anywhere on the route.
- The assignment of route into Forward and Reverse does not look correct in some cases. This will lead to the incorrect MPWs (and therefore CMs as well)
- The assignment of peak type into Peak and Off-Peak does not look correct in at least one case. This will lead to the incorrect MPWs (and therefore CMs as well)
- Also, we need to understand the methodology for assigning demand on services for which the last station is not a Monitoring Point.

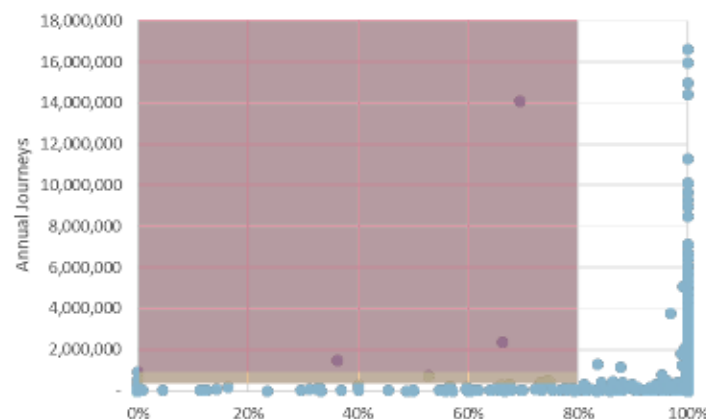
## High Priority: List of Issues

Priority	Category	Issue	Reference
High	Operational	There are some services patterns with a high number of people travelling on them that are not monitored anywhere on the route.	
High	Calculations	There appears to be many cases where the Direction has not been corrected assigned. This causes there to be at least one case where the Peak has not been corrected assigned	[TOC Name]: Column AB (AC)
High	Calculations	Check - what happens if the last station is not a Monitoring Point. Should the alighters here go on the previous MP.	[TOC Name]: Column AK:AL
High	Calculations	The MPWs and CMs results looks at large changes in absolute terms, but not relative terms. For example, North Fambridge in EB02 Peak and Off Peak has seen a halving of MPW between CP5 and CP6, but as it is only -0.02 less than CP5, it is not highlighted. What is the reason for this decrease? Did the CP5 figure consider all passengers on the branch, rather than only those that had alighted by North Fambridge?	

## Monitored demand

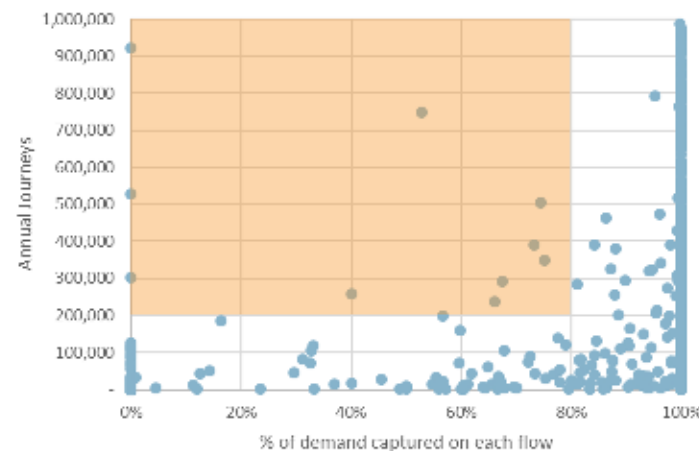
- There are a few flows where there is a high level of demand on the flow but only a small amount of this is captured at the selected MPs.
- Flows in the “Red zone” (demand > 1 million and monitored demand <80%)

Flow: Direction	"Monitored demand (%)"	Total demand (annual)	Operator
BIS-SRA: Forward	36%	1,459,328	Greater Anglia
WAT-WAT: Reverse	69%	14,084,200	SWR
NCL-LIV: Forward	66%	2,344,212	TPE



- Flows in the “Orange zone” (demand > 200,000 and monitored demand < 80%)

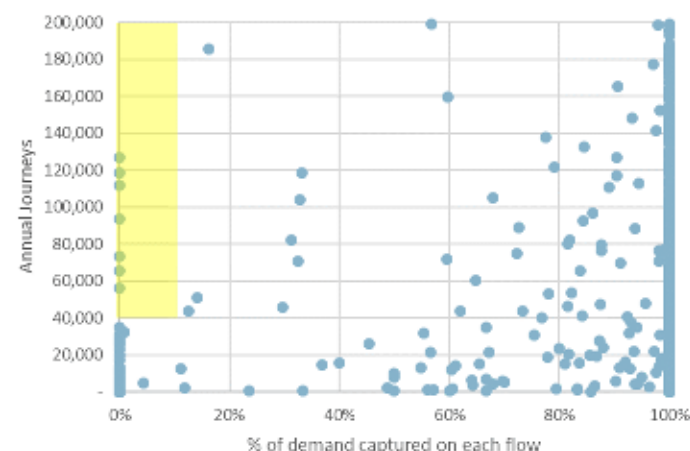
Flow: Direction	"Monitored demand (%)"	Total demand (annual)	Operator
MER-BYI: Reverse	74%	505,388	ATW
LST-BXB: Reverse	0%	527,020	Greater Anglia
LST-SMN: Reverse	73%	390,000	Greater Anglia
LST-WTM: Reverse	75%	350,220	Greater Anglia
NRW-SHM: Reverse	67%	291,980	Greater Anglia
BHM-BHI: Forward	0%	302,952	CrossCountry
AYR-EDB: Reverse	66%	238,680	ScotRail
GLC-GLC: Forward	0%	923,260	ScotRail
GLT-EDB: Forward	53%	747,500	ScotRail



## Monitored demand

- There are a few flows with a demand greater than 50,000 people but which are not monitored
- Flows in the “Yellow zone” (demand > 50 and monitored demand < 10%)

Flow: Direction	"Monitored demand (%)"	Total demand (annual)	Operator
MYB-WRU: Reverse	0%	93,600	Chiltern
SMN-WIC: Forward	0%	126,724	Greater Anglia
SSD-CBG: Forward	0%	65,520	Greater Anglia
DMF-NCL: Reverse	0%	112,060	ScotRail
STR-KMK: Reverse	0%	56,160	ScotRail
TWB-SGL: Forward	0%	118,560	ScotRail
WEY-BCU: Forward	0%	73,320	SWR



- A key reason for demand being un-monitored is that there is no MP at the end of the route. PwC and RDG to confirm how demand beyond the final MP should be treated.

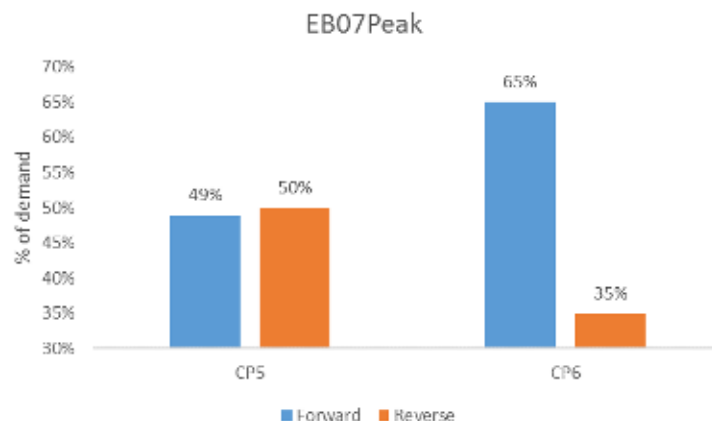
## Impact of un-monitored demand on weightings

- A Liverpool Street to Broxbourne service (TSC 920) has no MP on it, as Cheshunt is only a MP on TSC 912. This means that demand in the Reverse direction is not accounted for on any of the Liverpool Street to Broxbourne services.

					CP5	CP6
EB07Peak	912	Forward	LONDON LIVERPOOL STREET	LONDON LIVERPOOL STREET (912F)	5%	33%
EB07Peak	912	Reverse	CHESHUNT	CHESHUNT (912R)	5%	12%
EB07Peak	920	Forward	LONDON LIVERPOOL STREET	LONDON LIVERPOOL STREET (920F)	44%	32%
EB07Peak	920	Reverse	HERTFORD EAST	HERTFORD EAST (920R)	45%	23%

London Livrpl St	1826	1826	EB9200
Bethnal Green	1829	1829	EB9200
Hackney Downs	1833	1833	EB9200
Tottenham Hale	1839	1839	EB9200
Ponders End	1844	1844	EB9200
Brimsdown	1846	1846	EB9200
Enfield Lock	1849	1849	EB9200
Waltham Cross	1851	1851	EB9200
Cheshunt	1854	1854	EB9200
Broxbourne	1901	1901	EB9200

- EB07 Peak (West Anglia Inners) is very unbalanced between Forward and Reverse



# Directions

- There are four different types of error that have been made on assigning the direction of the flows:
  - (a) Flows does not have any direction assigned (e.g. CLT-WON on Greater Anglia)
  - (b) Flows have Reverse assigned for both A to B and B to A (e.g. MAN-SHR and SHR-MAN on ATW) *(we note that this is OK in the situation where both directions are Forward and there is no flow specified in the PEARS direction dataset).*
  - (c) Flows have two directions assigned (e.g. NOT-STP on East Midlands)
  - (d) Flows as “Reverse” even though they should be “Forward” (e.g. CET-LST on Greater Anglia)
- The directions are assigned in Model B for each TOC, which was not in scope for our Audit. We have highlighted diagnostics for identifying an error but PwC will need to confirm whether this is comprehensive and that no other Direction errors exist.
- As a result of the direction issues, on some occasions the peak type allocation is therefore not correct (e.g. CET-LST on Greater Anglia).

TOC	Direction	W	Reverse	Peak
1635 CET-LST 1746	W	Reverse	Peak	
1630 LST 1720	W	Reverse	Peak	

## Assigning demand to MPs

- North Fambridge result looks odd. It has a large percentage reduction 57%. The model only checks for large absolute reductions but this could miss out issues where the weighting is small but has changed significantly
- Checks built-in to the model:

Monitoring Point Location	CP5 MPW	Estimated CP6 MPW	Change (CP6-CP5)
NORTH FAMBRIDGE	0.044	0.0191	-0.02
NORTH FAMBRIDGE	0.0517	0.0234	-0.03

- Our checking process – looking at both absolute and relative changes

	CP5	CP6	Abs	%
STRATFORD (943F)	26%	26%	-1%	-2%
LONDON LIVERPOOL STREET (943F)	24%	20%	-4%	-17%
NORTH FAMBRIDGE (943R)	4%	2%	-2%	-57%
SOUTHEND VICTORIA (943R)	24%	20%	-4%	-17%
WICKFORD (943R)	21%	33%	11%	53%



# Medium Priority issues

## Medium Priority

- We have identified the following medium priority issues. These involve a risk of an error being created through structural issues in the model.
- Some formulas do not extend down the full range of data. Could lead to missing information on some TOCs.
- There does not appear to be any specific treatment for Splitting and Joining services. An error could occur if the location where the first portion of the service terminates is not a Monitoring Point.
- There are hard-coded values in the middle of calculation columns. There is no process to check that this has been done correctly.
- Some formulas change in the middle of calculation columns. There is no information to explain why this is the case or that it is correct.

## Medium Priority: List of Issues

Priority	Category	Issue	Reference
Medium	MBP-Hard coding	The service codes that are "All Trains" are hard-coded into the cells rather than being a lookup.	[TOC Name]: Column AD
Medium	Calculations	Formula in cell does not extend to full range in the model. This gives an incorrect value	[TOC Name]: Column AM1, AN1
Medium	MBP-Hard coding	There are hard-coded values in the middle of calculations. On Greater Anglia, this is particularly apparent in cells AS37397 this should be manual entry but isn't.	[TOC Name]: Column AS:AT
Medium	MBP-Formatting	Formulas change part-way down (on multiple occasions). Not clearly marked in Column AT. Whilst an attempt has been made to mark this more clearly in Columns AU and AV - the marking of changes doesn't fully line-up with the changes made	CP6 MPWs: AT9 (etc.)
Medium	Operational	The number of bank holidays varies between 2015/16 and 2016/17, should these not be standardised for a representative year?	CP6 MPWs: AM4:AR4
Medium	Calculations	The formula '=INDEX('Greater Anglia'!\$B\$5:\$R\$15000, MATCH(\$C5,'Greater Anglia'!\$R\$5:\$R\$15000,0),1)' doesn't cover the full range of data in the Greater Anglia sheet. Whilst we do not think this affects the results - it is inconsistent with the text in D1 which states this formula should cover the whole range.	Station Index: Column D
Medium	Calculations	The formula '=INDEX('Greater Anglia'!\$B\$5:\$R\$13925,MATCH('Station Index'!\$AB5,'Greater Anglia'!\$R\$5:\$R\$13925,0),1)' doesn't cover the full range of data in the Greater Anglia sheet. Whilst we do not think this affects the results - it is inconsistent with the text in D1 which states this formula should cover the whole range.	Station Index: Column AC

## Bank Holidays

- For TOCs that only use 2016/17 as recalibration there could be a small discrepancy if it is assumed that there are 10 Bank Holidays a year for CP6 – rather than the average of 8.

Weekdays per year	Saturdays per year	Sundays per year	Bank Holidays 2015/16	Bank Holidays 2016/17	Which Timetable do Bank Holidays follow? (press Ctrl+s after selecting)
261	52	52	6	10	Saturday

- Most operators do not operate services on Christmas Day and Boxing Day, so the total days per year should probably add up to 363, not 365 as shown. Some operators also reduce their level of service on the weekdays between Boxing Day and New Years Day.
- Also, operators may vary their bank holiday timetable depending on the bank holiday, for example GA operate a Sunday timetable on New Years Day, but Saturday timetables on other days. Note that the Franchise Agreement for GA requires a Sunday level of service.

# Low Priority issues

## Low Priority

- **The low priority issues we have identified so far:**
- Models are not structured in a MBP way. User inputs are scattered through the sheets.
- Models are not formatted in a MBP way. It is difficult to see what cells are for inputs, calculations and outputs.
- Formulas within a table change part-way through and these changes are not clearly marked
- Many of the calculations within the sheets are not required and obscure the flow of the information
- There is no “error trapping”, so it is difficult to know whether any error in any cell is acceptable
- There are multiple tables in some sheets, for which the rows and columns do not align with other tables in that sheet
- Diagnostic tests are often not testing the right information. There are some obvious diagnostics that are not tested for.
- There is no error-checking in the model (e.g. check that all MPWs sum to 1 for each service group).

## Low Priority: List of Issues (1)

Priority	Category	Issue	Reference
Low	MBP-Separation of inputs/outputs	Model sheets are not organised in a logical progression from Inputs > Calculations > Outputs. See "Data Flows" for more details	
Low	MBP-Formatting	The Headings are not always shown clearly. This makes it difficult to see what the columns are referring to. For example, columns X and Y appear to be showing Dep and Arr - in fact they are departure service code and arrival service code	[TOC Name]: Row 3
Low	MBP-Formatting	Formatting changes at an intermediate point within a spreadsheet. There does not appear to be any reason for this change in formatting	[TOC Name]: Row 15878
Low	MBP-Formatting	Calculations change across columns without any formatting to show that there is a difference in formula between one column and the next	[TOC Name]: Column AD:AE etc.
Low	MBP-Formatting	Formula is more complex than it needs to be	[TOC Name]: Column AG
Low	MBP-Separation of inputs/outputs	First cell is an input. All the other cells are formulae	[TOC Name]: Column AJ
Low	Calculations	Refers to instances when the Monitoring Point (column AI) is 3, however, this column can only contain a 0, 1 or 2. Clarify why formula has 'OR(\$AI5=1,\$AI5=3)' when the equivalent formula in the adjacent cell is '\$AI5=1'	[TOC Name]: Column AL
Low	MBP-Formatting	The row of calculations in "CP6 MPWs" is offset by one compared with that in "Station Index"	CP6 MPWs All Rows
Low	MBP-Formatting	Columns B:I are formatted as inputs. They are in fact calculations (more specifically they are a calculation feed).	CP6 MPWs B:I
Low	MBP-Formatting	The formula changes in the middle of the calculations	CP6 MPWs N
Low	Inputs	There were 366 days in 2015/16	CP6 MPWs: AA:AC
Low	MBP-Model design	Even within one sheet - the calculation move from right-to-left in some places and left-to-right in others	CP6 MPWs: J:AF
Low	MBP-Separation of inputs/outputs	The inputs for how Bank Holidays are treated should be put into an dedicated Input Sheet. It is hard for the user to find this Input here	CP6 MPWs: AM4:AQ4
Low	MBP-Model design	Columns are full of ERROR text. These columns are not required anyway	CP6 MPWs: P:Q



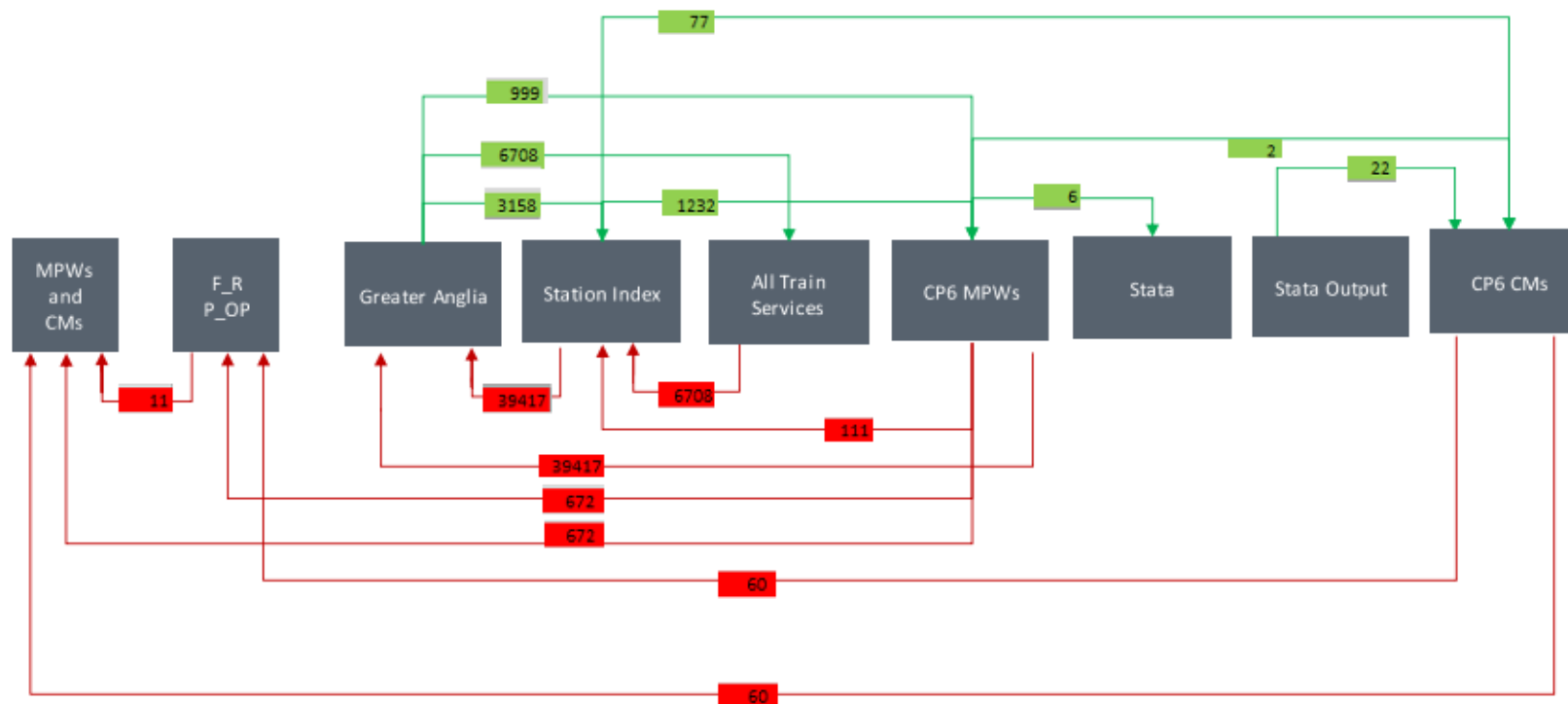
## Low Priority: List of Issues (2)

Priority	Category	Issue	Reference
Low	MBP-Separation of inputs/outputs	The 'Station Index' sheet has multiple tables contained within it. The layout of the sheet is confusing as the information along a single row refers to many different items. For example, Row 62 contains information on Birmingham Snow Hill and then Lowestoft. It could easily be placed across multiple sheets. In the middle of this data, is a table on CP5 results which doesn't seem to fit with the rest of the items in this sheet.	Station Index
Low	MBP-Model design	Column has no dependents within the rest of the model. Unsure of the purpose of this column	Station Index: Column F
Low	MBP-Model design	This table could be moved as it hinders the visual flow of information. The table to the left and right are both at a station level. It relates to CM whilst the rest of the sheet refers to MPWs.	Station Index: Columns H:P
Low	MBP-Formatting	This column is a calculation but it is formatted as an input	Station Index: Columns P
Low	MBP-Formatting	Error trapping would make this cleaner to read	Station Index: Columns D:E, Z:AA
Low	Text	Description in AA3 points to Column Q, which is blank	Station Index: Columns AA3
Low	MBP-Model design	Column has no dependents within the rest of the model. Unsure of the purpose of this column	Station Index: Column AD
Low	MBP-Model design	Some sheets within the model have data entered starting in Row 4 others have it starting in Row 5	
Low	MBP-Model design	1.5 parameter is crucial - yet it is not displayed clearly in an input sheet	CP6 CMs: S4
Low	Calculations	S8 Model C Greater Anglia 01092019, tab "CP5 MPWs", column AT, calculates the number of trains per weekday at each location (by direction) based on records in tab "Greater Anglia", column AI. Column AI records a 1 at locations at which trains originate in some cases (e.g. Wickford), when there are clearly not going to be any passengers alighting at that point. Should these locations not be ignored? This column does not appear to feed any other spreadsheets, therefore it may not be an issue.	CP6 MPWs: AT
Low	MBP-Separation of inputs/outputs	The Results model has calculations within it. This should all be hard-pasted values. Creates a risk of differences being introduced between Model C and the Results model	"Greater Anglia Draft Results" model



# Model design

- The model is not structured in a linear fashion. We have created a model map (below) which shows flows of information from left-to-right (green arrows) and right-to-left (red arrows).
- In addition, inputs are not clearly separated from outputs. It is not clear what is an Input sheet and what is a Calculation or Output sheet.



## Formatting

- The final column in this table is a calculation but it is shown as an input

Monitoring Point Location	CP5 MPW	CP6 MPW
STRATFORD	0.2634	0.2577

- The format for calculations is not clear and is inconsistently applied

Monitoring Point Location	CP5 MPW	2015/16 CP6 MPW	2016/17 CP6 MPW	2015/16 and 2016/17 CP6 MPW	Adjustments where necessary	Estimated CP6 MPW
3 STRATFORD	0.2634	0.2616	0.2539	0.2577	N/A	0.2577
7 LONDON LIVERPOOL STREET	0.2366	0.1935	0.1977	0.1956	N/A	0.1956
3 NORTH FAMBURIDGE	0.044	0.0198	0.0184	0.0191	N/A	0.0191
3 SOUTHEND VICTORIA	0.0422	0.0088	0.0045	0.0084	N/A	0.0084

- Changes in formulas have been marked up in some places but not other

71	82	52
18	0	0
18	0	0
3	0	0
16	0	0
19	0	0

## Formatting (con't)

- Calculations indicate that a manual intervention is required but does not check to see if formulas or manual intervention are used

0	0	0	0	0
0	69	70	69	70
0	0	0	0	0
1 Manual	Manual		0	0
1 Manual	Manual		0	0
1 Manual	Manual		14	20

- One row relates to different things (Marks Tey, EB07 then Marks Tey)

1 MARKS TEY	0.0051	0.0164	Greater Anglia Inners	EB	EB07	Peak	EB07Peak	30	MKT	#N/A	MKT	Marks Tey
-------------	--------	--------	-----------------------	----	------	------	----------	----	-----	------	-----	-----------

- There is no error-trapping – it is difficult to know if these errors are intended

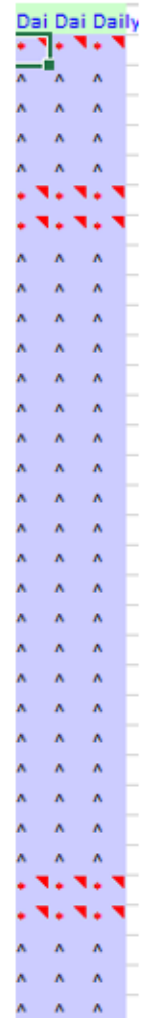
#N/A	WON	WON	Walton On Naze	1
CLT	#N/A	CLT	Clacton	1
COL	#N/A	COL	Colchester	1
#N/A	SDA	SDA	Stafford London	1

- Many of the formulae rely on hard-pasted values

```
'=IF(NOT(OR($Y5="891",$Y5="893",$Y5="894",$Y5="895",$Y5="896",$Y5="897",$Y5="899")), $AC5, "All Trains")
```

## Formatting (con't)

- We used OAK (Operis Analysis Kit) software to analyse the structure of the model.
- OAK maps each cell to state whether it is based on previous cells above (denoted by “^”) or if a different formula is entered (denoted by a “\*”).
- Models should be structured in a way that doesn't involve multiple changes of formula within a single column. This is not the case in the Greater Anglia model (see right).



## Model size and calculation duration

- Some of the spreadsheet models are large in size, which combined with the computationally inefficient calculations means that the models take a long time to recalculate. This can make the models less stable (i.e. they crash more frequently) and it is harder to make updates to the information.
- On opening the ScotRail model, it took 10 minutes\* to complete just 25% of its calculations. The extended time for calculation makes the model harder to use and also more difficult to Audit. (\* Note, this time will vary according to the specification of the machine used – but it is an example of how long it could take.)

# Calculation Checks and Types of Issues Identified - CMs

# Calculation checks

## Calculation checks

- List of checks of the STATA programmes. The left-hand column is an index of the steps taken.

Code	STATA Program	Pass or Fail	Comments	PWC Report Reference
A	Generate a list of origin stations by: <ul style="list-style-type: none"> <li>Day of the week</li> <li>Direction</li> <li>Peak/off peak/all trains</li> <li>ID assigned for each unique train</li> </ul>	Pass		
B	Generate a list of destination stations by: <ul style="list-style-type: none"> <li>Day of the week</li> <li>Direction</li> <li>Peak/off peak/all</li> <li>ID assigned for each unique train</li> </ul>	Pass		
C	Generate a list of origin – destination pairs, based on merging (A) and (B) above.	Pass		
D	Calculate the number of trains between each origin and destination by day of the week, direction and peak/off peak/all trains.	Pass		
E	Calculate the operational hours for trains between the origin and destination by: <ul style="list-style-type: none"> <li>Inserting 3 hours if 'peak'.</li> <li>Inserting the difference between the first and last train (minus 3 hours) if 'off peak'</li> <li>Inserting the difference between the first and last train if 'all trains'.</li> </ul>	Pass	<i>[Checks for other TOCs - The three hours are only removed if the total number of operational hours is above 15 hours. Are there any circumstances where there could be a peak service and a total operating day of 15 hours or fewer?]</i>	
F	Keep origin-destination pairs if: <ul style="list-style-type: none"> <li>Based on the correct direction for the origin-destination pair...</li> <li>Origin is in the opposite direction and is an MP.</li> <li>Destination is in the correct direction and is an MP.</li> </ul>	Pass		



## Calculation checks

Code	STATA Program	Pass or Fail	Comments	PWC Report Reference
G	The operational hours are divided by the number of trains minus 1 and are rounded to give the CM. Anything with under two trains is dropped.	Pass	No account is taken of gaps in service etc which might give strange rounding. For example, if the rounded CM is 16, should this really be 15 minutes based on the timetable?	Page 10 Bullet 1
H	Weighting figures for weekdays, Saturdays and Sundays are calculated based on alighters per day type / alighters for all days of the <u>week</u> . <sup>34</sup>	To Confirm	This is affected by the bank holiday calculation, in the same way as for MPWs	
I	For each combination of origin – destination, direction, the sum of weightings is calculated (i.e. weekday + Saturday + Sunday if the train runs every day = 1)	Pass		
J	For each combination of origin – destination, direction, day, peak/off peak/all trains, an adjusted <u>daytype</u> weighting is calculated, based on <u>dotw</u> weight / sum of weightings	Pass	This presumably stops the weighting being affected by no trains running on a <u>particular day</u> of the week	
K	For each origin – destination, direction, day, peak/off peak/all trains, a weighted CM figure is calculated based on the rounded CM figure from (G) and the weighting calculated in (J).	Pass		
L	A total CM figure (based on the sum of weighted CM values calculated in (K)) is calculated for each combination of: Origin Destination Direction Peak/off peak/all trains	Pass		
M	For each origin – destination, direction and peak/off peak/all trains combination, the total number of trains per week is calculated and the total number of operating hours per week is calculated.	Pass		

## Calculation checks

Code	STATA Program	Pass or Fail	Comments	PWC Report Reference
N	The service groups by origin-destination, direction and peak/off peak/all trains are merged into the data.	To confirm	Need to check what happens when origin – destination, direction and peak/off peak/all trains is in more than one service group.	Page 10 Bullet 3
O	The MPW figure for the origin station (based on the opposite direction) and the MPW figure for the destination station (based on the correct direction) are multiplied to create a new MPW based weight.	Pass		Page 10 Bullet 4
P	For each service <u>group</u> the total of the weights in (O) is calculated.	Pass		
Q	For each origin – destination, direction and peak/off peak/all trains, the weight as a proportion of the total weight for the service group is calculated.	Pass		
R	The total weighted CM calculated in (L) is multiplied by the weight calculated in (Q) to give a weighted CM for that <u>particular origin</u> – destination, direction and peak/off peak/all trains.	Pass		
S	The final CM figure is calculated based on the sum of the weighted CM figures for each service group, multiplied by 1.5.	Pass		Page 11 Bullet 5 & 6
T	The process is repeated from (G) but with any CM value above a certain value removed.	To confirm	Note that this value changes depending on the operator. GA used 120 minutes, LOROL uses 60 minutes	

# High Priority issues

## High Priority

- **The high priority issue(s) we have identified so far:**
- We do not think that the CM Stata code is correctly dealing with trains that split and join and is erroneously creating flows between two separate portions of the service. (See next slide)
- There appears to be an error in how service code 209 is coded in Stata between peak and off-peak.
- The CM results may be affected by any of the errors identified in the calculation of Monitoring Point Weightings. For example, trains which do not pass an MP en-route will not be included in the calculation. Also, where the direction information is wrong in the MPW calculations this will affect the CMs.
- Need to confirm whether the calculation methodology takes into consideration any increases or reductions in frequency between the May 2017 timetable and the timetable that will be operating at the start of Control Period 6 (i.e. Dec 18 timetable).

## Splitting and joining

- We believe that splitting and joining is not dealt with correctly in the Cancellation Minutes code. On a service with two portions: Bristol Temple Meads to Salisbury and Exeter to Salisbury (joining at Salisbury to continue to London. This gives a Bristol TM to Honiton flow – even though there is no flow. Can PwC confirm whether this is correct.

station1	station2	peakoff	W
Bristol Temple M	Basingstoke	Off-Peak	700
Bristol Temple M	Clapham Junction	Off-Peak	700
Bristol Temple M	Honiton	Off-Peak	700
Bristol Temple M	London Waterloo	Off-Peak	700
Bristol Temple M	Salisbury	Off-Peak	1332
Bristol Temple M	Woking	Off-Peak	700
Bristol Temple M	Yeovil Junction	Off-Peak	700

- This could affect any TOC with splitting/joining

	Joins	Splits
ATW	Yes	Yes
EMT	None	Yes
Great Western	None	Yes
Greater Anglia	Yes	Yes
Merseyrail	Yes	Yes
ScotRail	Yes	Yes
SWT	Yes	Yes
TPE	None	Yes
West Coast	Yes	Yes

# Medium Priority issues

## Medium Priority Issues (1)

- Some trains change from being peak trains to off-peak trains en-route. Journeys from stations in the off-peak section to those in the peak section are omitted completely.
- Trains which change service group en-route may be similarly affected. [To check further]
- 15/16 and 16/17 are combined in the Excel spreadsheet based on the total alighting figures across the entire TOC for each year. This differs from the methodology used to combine the 15/16 MPWs and 16/17 MPWs which combines them based on the alighting figures at each individual station.
- The STATA methodology does appear to follow PWC's agreed methodology, with the exception of the removal of anything with a service frequency greater than 120 minutes, the differences for each TOC are noted in the documentation except for Arriva Rail London. Also, it is ambiguous as to whether the Hull Trains data has a 120 minutes threshold or not (the coding and the documentation are not aligned). See next slide

## Service Interval Thresholds

TOC	Drop if CM >	Comments
ARL	60	Inconsistent with documentation
ATW	120	Pass
c2c	120	Pass
Chiltern	120	Pass
CrossCountry	120	Pass
East Coast	180	Pass
East Midlands	120	Pass
Grand Central	No limit	Pass
Great Western	120	Pass
Greater Anglia	120	Pass
Hull Trains	No limit	Code appears to exclude at 120. But documentation suggests otherwise
London Midland	120	Pass
Merseyrail	120	Pass
Scotrail	120	Pass
South West Trains	120	Pass
TPE	120	Pass
Virgin West Coast	120	Check appropriateness for Euston to Bangor



## Medium Priority Issues (2)

- **Loop Services:** Between MPs, the MPWs are taken for the opposite direction (in the case of station 1) and the actual direction (in the case of station 2). In the Merseyrail do file there is a small bit of code which stops direction1 being the opposite direction, if station 1 = station 2 (presumably this only affects loop services). It is not clear why the inversion of direction 1 does not happen in this situation and why other operators with loop services (e.g. SWT) are not treated in the same way.
- **Weighting by Days:** For some reason Hull Trains uses Chiltern Days.dta instead of the Days.dta created from the Excel spreadsheet.

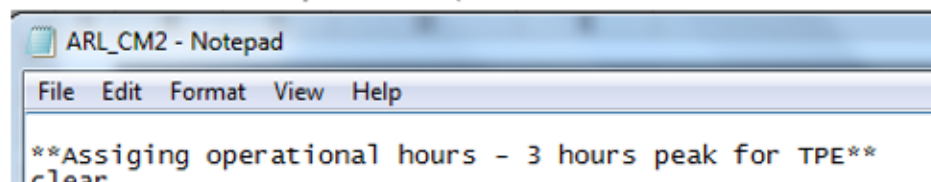
## Medium Priority Issues (3)

- A number of ‘simplifying’ have been made in the code. We think that the documentation should be more explicit about these simplifications and the impact they could have. These include:
  - Methodology assumes that passengers can take other trains between station pairs within the same TOC but in a different service code, should a train be cancelled. This does not take into consideration the speed of trains. For example, between Reading and Paddington, would a passenger take a relief line train if their main line train was delayed?
  - Only a single TOCs timetable is considered at any point. E.g. Doncaster to King’s Cross on Grand Central does not consider any East Coast services on the same route.
  - No account is taken for the regularity of the service interval. E.g. services departing at xx:00 and xx:05 are treated the same as services departing at xx:00 and xx:30.
  - The intervals are not weighted by demand across the day – e.g. a 90 minute gap at 9pm is weighted the same as a 90 minute gap at 10am
  - The documentation doesn’t show many of the detail changes that are required for each TOC. This may make it harder to use the code in any future runs.

# Low Priority issues

## Low Priority

- The low priority issue(s) we have identified so far:
- There are a large number of constants required in the STATA code. Some of these constants needs to be changed for each TOCs (e.g. the number of trains, number of service groups, peak timings, threshold interval for excluding services). These constants are not defined at the top of the code – but appear sporadically throughout the code. This creates the risk that, if any of these were to change it is not as straightforward as it could be to identify where in the code all the alterations are required.
- The TOC spreadsheet model has a number of Modelling Best Practice issues (see the review of the Monitoring Point Weightings).
- In some models (e.g. GWR) there are multiple options – it is not made clear which is used in the final results.
- It would be clearer if there was a separate model for MPWs and CMs.
- Some of the annotation is not updated (see below text for TPE in ARL code).



```
ARL_CM2 - Notepad
File Edit Format View Help
**Assigning operational hours - 3 hours peak for TPE**
clear
```

# TOC Specific Issues Identified

(Covering both CM and MPWs)

# Arriva Rail London

- **High Priority**

- The methodology for passengers that both board and alight within the East London Line (ELL) core needs to be confirmed. Stations within the ELL core are not given a Service Group, it shows as #N/A, however the passenger alighting in the core are counted in the first MP after the core. This seems OK, if they boarded before the core and alighted after the core but not if they board and alight in the core. PwC to confirm approach and update the documentation and the modelling (if appropriate).
- Willesden Junction to Clapham Junction is not capturing any demand in the F direction. Is this intentional?

- **Medium Priority**

- On the CMs, the interval threshold is 60 minutes rather than 120 minutes. The documentation should be updated to reflect this.

- **Low Priority**

- All the Modelling Best Practices issues that have been identified (see earlier slides)
- In the sheet with the MOIRA loadings in it. Cells AM1 and AN1 do not extend over the full range of data.
- There are many cells that return an error in the spreadsheet. Including columns AS and AT in ARL sheet.

# Arriva Trains Wales

- **High Priority**

- MAN-SHR and SHR-MAN are both shown in the R direction. Is this correct?
- PwC to check that the splitting and joining issue in the Stata code does not create any errors on this TOC.

- **Medium Priority**

- Manual column AS:AT in ATW sheet doesn't do anything. It is not clear why this has been set to manual. PwC to confirm.

- **Low Priority**

- All the Modelling Best Practices issues that have been identified (see earlier slides)
- The “All Train Services” sheets are full of cell returning errors.
- In the sheet with the MOIRA loadings in it. Cells AM1 and AN1 do not extend over the full range of data.

## c2c

- **High Priority**
  - FST to SRY showing as both a F and R flow
  - FST to GRY showing as both a F and R flow
- **Medium Priority**
  - No medium priority issues found
- **Low Priority**
  - All the Modelling Best Practices issues that have been identified (see earlier slides)



# Chiltern

- **High Priority**
  - No high priority issues found
- **Medium Priority**
  - No medium priority issues found
- **Low Priority**
  - All the Modelling Best Practices issues that have been identified (see earlier slides)

# CrossCountry

- **High Priority**
  - We recognise that all EH01 services codes at Birmingham NS have been combined into an “All” category. We need clarity on how this will be dealt with in the regime – presumably the system will need an assigned service code for each MP. PwC to confirm.
- **Medium Priority**
  - No medium priority issues found
- **Low Priority**
  - All the Modelling Best Practices issues that have been identified (see earlier slides)

# East Coast

- **High Priority**
  - No high priority issues found
- **Medium Priority**
  - No medium priority issues found
- **Low Priority**
  - All the Modelling Best Practices issues that have been identified (see earlier slides)
  - The “All Train Services” sheets are full of cell returning errors.

# East Midlands

- **High Priority**
  - NOT-STP showing as both a F and R
  - PwC to check that the splitting and joining issue in the Stata code does not create any errors on this TOC.
- **Medium Priority**
  - No medium priority issues found
- **Low Priority**
  - All the Modelling Best Practices issues that have been identified (see earlier slides)
  - In the sheet with the MOIRA loadings in it. Cells AM1 and AN1 do not extend over the full range of data.

# Grand Central

- **High Priority**
  - No high priority issues found
- **Medium Priority**
  - No medium priority issues found
- **Low Priority**
  - All the Modelling Best Practices issues that have been identified (see earlier slides)

# Great Western

- **High Priority**
  - PwC to check that the splitting and joining issue in the Stata code does not create any errors on this TOC.
- **Medium Priority**
  - No medium priority issues found
- **Low Priority**
  - All the Modelling Best Practices issues that have been identified (see earlier slides)
  - In the sheet with the MOIRA loadings in it. Cells AM1 and AN1 do not extend over the full range of data.
  - The “All Train Services” sheets are full of cell returning errors.

# Greater Anglia

- **High Priority**

- The Greater Anglia model has a higher number of Direction and Monitored Demand issues (see previous “High Priority” slides).
- Not sure what columns AO and AP in Greater Anglia sheet are doing. Seems to give a strange result in cell AF15781

- **Medium Priority**

- The assumption on Cambridge North in tab "CP6 MPWs" only assume abstracted demand from Cambridge and no additional demand, is this correct? Cambridge North is only an MP in the Down direction and not the Up direction, is this correct? In addition, why is Cambridge North only in EB06 and not EB05 (which trains also stop at Cambridge North)? The abstraction at Cambridge, alongside the reduction already calculated for CP6 results in the MPW reducing by nearly one half compared with CP5.
- The MPWs and CMs results looks at large changes in absolute terms, but not relative terms. For example, North Fambridge in EB02 Peak and Off Peak has seen a halving of MPW between CP5 and CP6, but as it is only -0.02 less than CP5, it is not highlighted. What is the reason for this decrease? Did the CP5 figure consider all passengers on the branch, rather than only those that had alighted by North Fambridge?

- **Low Priority**

- All the Modelling Best Practices issues that have been identified (see earlier slides)

# Hull Trains

- **High Priority**
  - No high priority issues found
- **Medium Priority**
  - No medium priority issues found
- **Low Priority**
  - All the Modelling Best Practices issues that have been identified (see earlier slides)



# London Midland

- **High Priority**

- In the STATA code, the “209” adjustment for LM adds 30 minutes to both the peak and off-peak. Should it add 30 minutes to the peak and reduce the off-peak by 30 minutes

- **Medium Priority**

- Sheet "CP6 MPW" cells M59 and M83 say "Adjusted" but it is not clear where this adjustment has been made

- **Low Priority**

- All the Modelling Best Practices issues that have been identified (see earlier slides)
- The “All Train Services” sheets are full of cell returning errors.

# Merseyrail

- **High Priority**
  - PwC to check that the splitting and joining issue in the Stata code does not create any errors on this TOC.
- **Medium Priority**
  - No medium priority issues found
- **Low Priority**
  - All the Modelling Best Practices issues that have been identified (see earlier slides)

# ScotRail

- **High Priority**

- 0% of GLC-GLC demand is not captured. Is this correct?
- Quite a few same direction in both ways. Understand that direction is F unless R triggered but some of these are both R
- PwC to check that the splitting and joining issue in the Stata code does not create any errors on this TOC.

- **Medium Priority**

- V. high F/R splits on HA02 and HA03. PwC to confirm they are happy with this/

- **Low Priority**

- All the Modelling Best Practices issues that have been identified (see earlier slides)
- In the sheet with the MOIRA loadings in it. Cells AM1 and AN1 do not extend over the full range of data.

# South Western Railway

- **High Priority**

- We believe that splitting and joining is not dealt with correctly in the Cancellation Minutes code. On a service with two portions Bristol Temple Meads to Salisbury and Exeter to Salisbury (joining at Salisbury to continue to London. This gives a Bristol TM to Honiton flow – even though there is no flow. Can PwC confirm whether this is correct.

station1	station2	peakoff	W
Bristol Temple M	Basingstoke	Off-Peak	700
Bristol Temple M	Clapham Junction	Off-Peak	700
Bristol Temple M	Honiton	Off-Peak	700
Bristol Temple M	London Waterloo	Off-Peak	700
Bristol Temple M	Salisbury	Off-Peak	1332
Bristol Temple M	Woking	Off-Peak	700
Bristol Temple M	Yeovil Junction	Off-Peak	700

- In the Stata Code, the SWR loop services are not treating in the same way as the Merseyrail loop services (i.e. with directions reversed to account for the loop), which means that this may not be calculating correctly. PwC to confirm.

- **Medium Priority**

- Manual column AS:AT in “SWT” sheet doesn't do anything. Is this correct?

- **Low Priority**

- All the Modelling Best Practices issues that have been identified (see earlier slides)
- In the sheet with the MOIRA loadings in it. Cells AM1 and AN1 do not extend over the full range of data.
- The “All Train Services” sheets are full of cell returning errors.

# TransPennine Express

- **High Priority**
  - No MP at Liverpool on Newcastle to Liverpool services
  - Some of the direction/peak issues that applied to Greater Anglia also apply here.
  - PwC to confirm that no changes as a result of the impact of the new Ordsall Curve are in scope
  - PwC to check that the splitting and joining issue in the Stata code does not create any errors on this TOC.
- **Medium Priority**
  - No medium priority issues found
  - We recognise that TPE have asked for Trains minus 1 rather than trains. But we don't have insight into why there are able to make this change as it affects the denominator in the calculation. PwC to confirm
- **Low Priority**
  - All the Modelling Best Practices issues that have been identified (see earlier slides)

# Virgin West Coast

- **High Priority**
  - PwC to check that the splitting and joining issue in the Stata code does not create any errors on this TOC.
- **Medium Priority**
  - No medium priority issues found
- **Low Priority**
  - All the Modelling Best Practices issues that have been identified (see earlier slides)