**Referee 1: Anonymous**

First of all, the authors would like to thank the reviewer for taking valuable time to review and for the critical assessment of the paper.

1) **Comment 1:** Grammatical and language issue

C1 Ans: We will correct the stated grammatical errors and language problems the reviewer stated in the comments sections.

**Comment 2 :** The text also contains logical errors or incomplete descriptions making it ambiguous or hard to follow (again something co-authors should have picked up on). Examples include:

- Page 2, Line 5: "Discolouration is the water quality issue most apparent to customer, causing the highest contact rates worldwide."

Ans: We are not sure what is unclear in this sentence.

Regarding contact rate term, we think the rate term is correct i.e. number of customer contacts per 1000 population per year. Otherwise, the contact numbers are not comparable.

- Page 4, Line 8: "minimising visual dissimilarities and errors between downstream simulated and measured pressure"

Ans: Wording of the sentence has been ammended.

"minimising visual dissimilarities and maximising correlation coefficient ($R^2$) between downstream simulated and measured pressure."

- Page 4, Line 9: "While pipe roughness alone can produce accurate simulation of observed pressure". How does the pipe roughness do this? :

Ans: To conduct a hydraulic modelling calibration, it is a standard practice to change only pipe roughness to reduce the difference between measured and simulated pressure – i.e. increase roughness until headloss is sufficient to match the observed pressures. However, this can result in unrealistically large roughness values and erroneous velocities and travel times that are particularly important for water quality simulation. Boxall et al., (2004)[1] showed that 1 mm pipe roughness value effectively reduces pipe diameter by 2 mm, matching both pressure and travel time data, suggesting the importance of changing both pipe roughness and diameter simultaneously.

- Page 4, Line 10: "inaccurate representation of velocities which can be significant for quality application can persist as the above is an indeterminate problem space.":

Ans: This sentence refers to the necessity to simulate velocities and hence travel times for water quality accurately. For a given imposed flow, various combinations of diameter and roughness can produce similar headloss and pressure – an indeterminate problem space. Each of these paired values has a unique velocity. It is important that the correct pairing is selected to simulate water quality effects.

The full sentence we agree however is unwieldy so has been edited to "While pipe roughness ($k_s$) alone can be modified to produce an accurate simulation of observed pressure, inaccurate representation of velocities, which can be significant for the quality applications, can persist. This is because hydraulic calibration is an indeterminate problem space where various combinations of diameter and roughness can produce similar headloss and pressure.

- Page 4, Line 23: "Thus, from hydraulic model optimisation, a seven fold reduction in roughness height was found after the invasive cleaning." This is inaccurate. When a reduction is made, it has to be stated relative to the original value. A reduction of one fold means that the value was reduced by 100 %, i.e. to zero:

Ans: We think it is correct. The same explanation is given to the Oxford dictionaries "https://en.oxforddictionaries.com/definition/sevenfold".

2) **Comment 3:** Nothing is mentioned in the paper on the possibility of leakage from the pipe and new leaks forming during the testing period. How would leakage have affected the results?

C3 Ans: The following has been added to the Results and discussion chapter:

The trunk main studied had no known leakage, as assessed through night line analysis. The effect of any unknown background leakage would have been manifest in the flow data that was used as an input to the model. The night line was not observed to change from the start to the end of the monitoring period (other than due to known operational changes) suggesting no new leakage occurred during the study.

**3) Comment 4:** The calibrated pipe roughness values include minor losses at bends and joints. Why weren't these incorporated in the model and how will they likely impact on the results?

C4 Ans: We had developed the hydraulic model as realistically as possible from industry records and local operation knowledge. During the model construction, minor losses were incorporated as EPANET loss coefficient inputs, determined from the EPANET manual Table 3.3 (p-32)[2]. A comment to explain this has been added to the manuscript. These values were not considered as calibration variables, and where fixed throughout the simulations so effects would have been constant.

**4) Comment 5**

- Acronyms should be used with discretion in publications since unfamiliar acronyms serve to obfuscate rather than clarify the text. I suggest removing ‘TOTEX’, ‘PODDS’ and ‘SR’

Ans: PODDS (Prediction of Discolouration in Distribution Systems) term has been using in the academic literature since 2001. The PODDS model theory drives the discolouration risks assessment for this case study. A change of SR to ‘service reservoir’ and TOTEX to ‘total expenditure’ has been added to the manuscript.

- Be consistent with the use of capital or small letters when referencing figures: ‘1A’ not ‘1a’.

Ans: All figure referencing has been changes to small letters in the manuscript i.e. 1a

- Page 2, Line 6: What does ‘international accepted’ research mean? Why not simply state that research was conducted and the findings are...

Ans: The PODDS model has published simulated discolouration responses from the UK and other countries e.g. Australia (Boxall and Prince, 2006)[3], Portugal (Husband and Boxall, 2016)[4] etc. So we think this is conceptually correct to include the international status, however agree ‘accepted’ is not appropriate and we have changed this to ‘validated’.

- Page 2, Line 31: “mostly residential with consistent demand across the year”. Do you mean that there was not seasonal variation in demand?

Ans: Yes.

- Page 4, Line 1: “two (2) ATI NephNet turbidity loggers were used with a 1 second sampling interval to ensure data validation and confidence.” There is no need to repeat the written ‘two’ with a number ‘2’. Using two loggers does not automatically ensure data integrity. Describe how this was done.

Ans: (2) has been removed from the manuscript.

The below explanation has been added to the paper:
The ATI turbidity loggers were calibrated under laboratory conditions and using two loggers to ensure that the collected data was consistent. The spot check of these instrument outputs was tested via HACH handheld logger which was calibrated against formazin turbidity standard samples.

- PEST calibration software”. This software was developed for a watershed model. An explanation of the method and how it was applied to the pipe roughness problems is required:

Ans: PEST is a model independent calibration software (Doherty, 2005)[5] and has been extensively tested for various watershed models. PEST calibration ability was also tested for EPANET model calibration (Méndez et al., 2013)[6]. The model was previously integrated into

MODFLOW, a groundwater modelling software as well. These new references have been added to the manuscript appropriate section.

The below explanation of the method has been added to Chapter (3.1):
The estimated boundary condition of pipe roughness and the diameter using Boxall et al. (2004)[1] concept has been applied to the PEST in conjunction with the EPANET model to determine the best possible solutions comparing simulated and measured downstream pressure.

- Page 5: An explanation of the 'operational circumstances' that lead to the changes in consumption pattern should be provided. How are these expected to have influenced the results?

Ans: An explanation of the operational circumstances reasons has been added to the discussion chapter (3.2):
From 27 August 2015 till 16 October 2015, a few new properties were connected to the downstream distribution zone fed from the investigated trunk main during additional repair work in a neighbouring network. Demand was increased by about 3 l/s during this process which can be confirmed from the continuous night line profile for over two months.

- Page 6, Line 4: "To avoid regulatory turbidity limit (4.0 NTU), shear stress was reduced stepwise to 1.135 $N/m^2$". As I understand this test, the shear values were generated by flushing the pipe through a hydrant. Why would the turbidity limit then apply, or was consumers simultaneously connected to the system?:

Ans: The reviewer is correct that these operations were undertaken by opening fire hydrants; however, this was only to achieve additional flow and associated shear stress. The majority of the flow was due to the downstream demands of the associated network, and hence the regulatory limit applied.

- Page 7, Line 7: "Benefits expected due to invasive cleaning included an improvement in hydraulic capacity and a reduction in discolouration risk, as well as improve asset resilience and pipe life span." It is not clear how invasive cleaning would improve the 'resilience' and 'life span' of a pipe:

Ans: Resilience encompasses many factors, including hydraulic capacity which was clearly improved here. Life span has been removed.

- It will be useful to have a table with the test parameters on which Figure 5 was based to allow the reader to get a better understanding of the variations observed:

Ans: An explanation of material release rate and accounting variables was in the manuscript Page 6: line 22-25 and Page 7: line 1-5. Addition to this, a table of test parameters has been added to the manuscript:

*Table 1: Test parameters for material release rate calculations*

| Parameters | Unit | Trial 1 | Trial 2 | Trial 3 | Trial 4 | Trial 5 | Trial 6 |
|---|---|---|---|---|---|---|---|
| Diameter, D | m | 0.215 | 0.2274 | 0.2274 | 0.2274 | 0.2274 | 0.22404 |
| $K_s$ | mm | 6.82 | 1.05 | 1.05 | 1.05 | 1.05 | 2.28 |

Though the $k_s$ and diameter values were unknown in trial 3, 4 and 5, paired values ($k_s$ and D) were assumed to be equal to trial 2.

To calculate imposed excess shear stress addition to the above table, discolouration material density ($\rho$) and gravity (g) was used at 1100 Kg/m$^3$ (Boxall et al., 2001; Ryan et al., 2008)[7,8] and 9.81 m/s$^2$.

<u>References stated in this discussion papers:</u>

1. J. B. Boxall, A. J. Saul, and P. J. Skipworth, ‘Modeling for Hydraulic Capacity’, Journal - American Water Works Association, vol. 96, no. 4, pp. 161–169, Apr. 2004.

2. L. . Rossman, ‘Epanet 2: User Manual’, USEPA, Cincinnati, OH, User Manual EPA/600/R-00/057, Sep. 2000.

3. J. Boxall and R. Prince, ‘Modelling discolouration in a Melbourne (Australia) potable water distribution system.’, J Water SRT - Aqua, vol. 55, pp. 207–219, 2006.

4. S. Husband and J. Boxall, ‘Understanding and managing discolouration risk in trunk mains’, Water Research, vol. 107, pp. 127–140, Dec. 2016.

5. J. Doherty, ‘PEST - Model-Independent Parameter Estimation’, Watermark Numerical Computing, Brisbane, Australia, User Manual, 2005.

6. M. Méndez, J. A. Araya, and L. D. Sánchez, ‘Automated parameter optimization of a water distribution system’, *Journal of Hydroinformatics*, vol. 15, no. 1, pp. 71–85, Jan. 2013.

7. J. Boxall, P. J. Skipworth, and A. Saul, ‘A novel approach to modelling sediment movement in distribution mains based on particle characteristics’, in *ater Software Systems: v. 1: Theory and Applications (Water Engineering & Management)*, Hertfordshire, UK: Research Studies Press, 2001, pp. 263–273.

8. G. Ryan *et al.*, ‘Particles in Water Distribution System: Characteristics of particulates Matter in Drinking Water Supplies’, CRC, Australia, Researh Report 33, 2008.