**Referee 3: Mirjam Blokker**

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.

**Comment 1:** Section 3,2 is water quality assessment, and the trials that were done are some sort of a risk assessment to see how much of a turbidity response is caused by a certain increase in shear stress (due to an increase in flow). This is a controlled flow increase test, if you like. It resembles the standardized RPM (resuspension potential method, Vreeburg and Schaap 2004), except that it was not controlled to the same increase in flow each time the test was done. I believe the term “conditioning test” should be avoided. It is a risk assessment, not a cleaning action. Also, the PODDS explained shear stress conditioning, to avoid high future turbidity responses, is something very different. Hence, I would avoid using “conditioning” in this paper.

**C1 Ans:** The distinction between a risk assessment and cleaning action is interesting. The network intervention of the RPM is a cleaning action – removing material from the network due to increase in hydraulic forces. How the data is then interpreted can be a part of a risk assessment. Interpretation of the RPM may be complicated since it uses a fixed velocity criterion, which then exerts a diameter dependent force and also takes no cognisance of normal, daily or recent hydraulic conditions. To overcome this, we used an excess shear stress criterion. However, due to the invasive cleaning application, both pipe roughness ($k_s$) and diameter changed to such an extent that the applied excess shear could not be systematically achieved.

Any increase in hydraulic forces above those normal experiences will remove material and hence have a cleaning action. When such an increase is small in magnitude and duration, it is a long way from removing all material and hence has a conditioning effect – removing all material with adhered strength up to the imposed force. This is exactly consistent with the PODDS based shear stress conditioning concept.

**Comment 2:** It is not clear to me why the 6 trials should best be compared (in fig 5) by dividing the turbidity by the product of shear stress and pipe wall area. I would like to see fig 5 also for the clean turbidity*Q data, and a better explanation of why this division of tau is valid or could be valid.

**C2 Ans:** The volumetric turbidity calculation by multiplying the volume of water and measured turbidity gives the mass flux effect for a specified period. Based on the PODDS concept, material layers are held in various strength profiles adhered on the complete pipe surface. This volume will be a complex function of the surface affected area and the imposed shear force. Hence when pipe area and imposed force are not constant, normalisation is required. Arguably the change in the area here was small and had little effect, but was included for completeness. The invasive cleaning produced a substantial change in roughness and hence imposed excess shear stress. Therefore, it was vital that the results were normalised by this in order to be comparable.

**Comment 3:** Why does table 1 not contain the results of trial 3, 4 and 5? Is it possible to find some sort of correlation between ks and turbidity response (corrected for shear if you like)? Could pressure and flow data indicate over time the diameter reduction and thus indicate the growth of the loose material (plus biofilm)? It would be worthwhile to check this briefly and discuss, without being able to prove this based on only one trunk main.

**C3 Ans:** Unfortunately, practical constraints where such that the data necessary for detailed hydraulic calibration was not collected for all events.

Previous work (Boxall et al., 2003)[1] has suggested that the change in roughness (~0.01mm) corresponding to notable turbidity (~ 10NTU) response for a 3inch 1.6km cast iron pipe is significantly less than the accuracy of the hydraulic calibration possible here. Hence while the above is a very worth concept and one we would wish to explore, however, it was not feasible here.

**Comment 4:** If for Table 1 I add diameter and two times the roughness, I approximately get the assumed diameter of 228 mm. In the calibration test, is the sum of D and ks limited to this? If yes, please mention this. If not, would it be a good idea to do so?

**C4 Ans**: As with the original work (Boxall et al. 2004)[2], the paired value (roughness and diameter) solution is constrained by the original pipe diameter. Comment to this effect has been added to the paper.

**Comment 5**: Fig 5 suggest turbidity response after 12 months was similar to pre cleaning, whereas the ks was not yet increased to the same amount.

**C5 Ans:** This result is covered in the manuscript, that the invasive cleaning was effective in removing significant amounts of historical accumulations. However, the remaining material was not necessarily mobilised easily through invasive cleaning exercise and hence represent a discolouration risk. After 12 months this weaker material risk had returned, but the stronger material apparently responsible for the initial roughness height had not.

**Comment 6:** Fig 4b shows that PODDS was able to simulate the measured data quite well. Since the text says that the max of 1 NTU was not expected, I assume the PODDS result could not have been generated before the trial results were available. It would be helpful to clarify this in the text. I am wondering, what would PODDS have predicted based on the data of fig 4a? This could indicate what the actual results of the cleaning were. Could you use PODDS to predict for each trial what the turbidity response would be for a set controlled flow increase? Thus mimicking the test under the same conditions, and then compare the results. In which case the division by shear stress would not be needed.

**C6 Ans:** We are not sure where the text says max 1.0 NTU was not expected after invasive cleaning. It was assumed that there could be low response due to the pipe wall cleaned with the invasive application. Did you mean that the response of 1.0 NTU turbidity target of conditioning trial?

This is an interesting question regarding mimicking the response. PODDS is a semi-empirical model with model parameters requiring calibration, although previously calibrated parameters have shown to be transferable. Before the simulation, we tested the model response using previously calibrated parameters and created a scenario profile. However, depending on the network conditions, the measured response can be varied e.g. flow estimation at service reservoir outlet, flow fluctuation of the standpipe. In practice, the target turbidity response of 1.0 NTU for flow conditioning trial has a certain buffer for not exceeding the threshold limit. During the trial (figure 4a), the response was recorded up to 2.0 NTU, although as long as the response was below the regulator limit of 4.0 NTU it was OK.

Mimicking was not effectively possible from pre to post cleaning as the PODDS model cannot track simultaneously accumulated layer mobilisation-accumulation processes. It is possible to simulate the discolouration response for a set of flow increase from trial 3 to 6 taking information from figure 4b model parameter. This is our next step to evaluate the quality performance so that we can avoid extensive fieldwork. However, it is still good to have some degree of fieldwork information so that we can accept and validate the performance assessment. The mimicking is effectively possible with the newly developed Variable Conditioning Discolouration (VCD) model by Furnass, (2015)[3] which can track both mobilisation and regeneration for a long term. However, to simulate this process in VCD model, we do require both continuous flow, upstream-downstream turbidity response which is unavailable for this study.

**Comment 7:** Fig 3: shear stress during ice pigging must be much higher, but is not easy to calculate. Instead, I would leave this part out to avoid confusion. What happened around 27 September (downstream pipe break?)? Caption should say 2015.

**C7 Ans:** We agree with the reviewer that estimation of shear stress due to the ice plug formation is hard, if not impossible.

A planned night time downstream network flushing was undertaken from 22nd September till 23rd September (2 nights). Also, there was a downstream burst event recorded as starting on 24th September and continuing until 25th September (figure 1). Unfortunately, we did not have any turbidity loggers deployed on this trunk main which could have captured the mobilisation response due to the burst event. These events could have mobilised accumulated material as it was higher than trial 2 event and potentially influence trial 3 responses. These event details have been added to the paper.

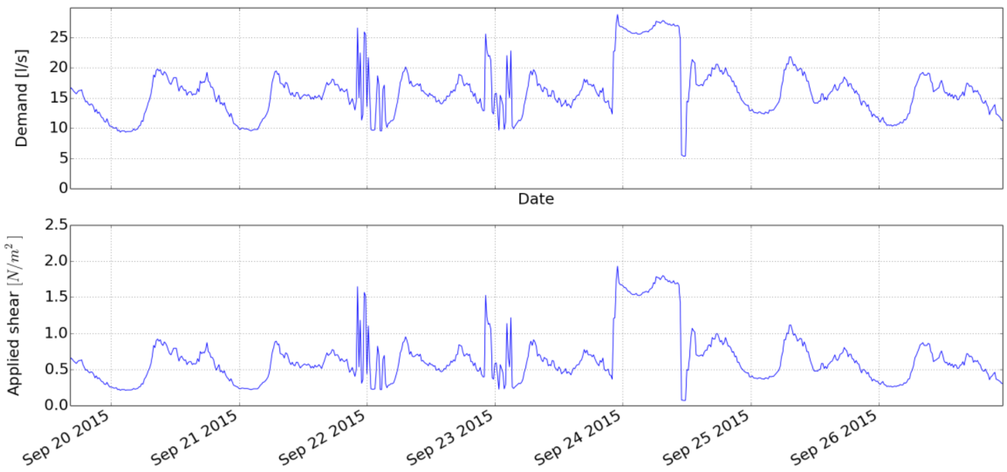


*Figure 1: Flow and shear stress data from referee comment section*

**Comment 8:** I do not understand how asset deterioration (for other than cast iron pipes) would lead to water quality issues. I can see that if no cleaning is done, time will cause more particulate accumulation, but this does not relate to the age or the condition of the pipe.

**C8 Ans**: We have used deterioration in a broad sense, i.e. to encompass material accumulation at the pipe wall, whether this is from the bulk water in all pipe types or to also include corrosion processes within cast iron pipes. The amount of material accumulated at the pipe wall is a prime indicator of the asset condition.

**Comment 9:** This is one of several studies that "suggests" a temperature dependence. The reference to Sharpe's thesis is very limited. The biofilm explanation is not substantiated with this particular AC trunk main study.

**C9 Ans**: We agree that this work contributes to the body of research that suggests temperature dependence, but does not definitively prove it. Sharpe's thesis is a very relevant work rigorously proving this link, and that it is biofilm-dependent, further publications from this work are not yet in print so cannot be referenced. Further references that suggest temperature dependence have been added in the paper including e.g. Blokker and Schaap, (2015)[4]. Regarding AC main biofilms, some published papers are stating how pipe material influence biofilms structure. There is no reason or evidence in previous research to suggest biofilms within AC pipe would be fundamentally different to other pipes. Hence the suggestion made is reasonable. However, this is a single study and hence not possible to confirm anything without investigating similar test profile on varying pipe material.

**Comment 10**: There are quite some grammar mistakes and typos that need to be looked at. For a conference paper the limited number of references was ok, but I would like to see an introduction with some more references added, in order to place this work more in perspective.

**C10 Ans**:

- We will correct the grammar and typing mistakes.

- The literature review and references have been revisited and updated in response to the reviewer comment; however, the changes are limited by the strict paper length constraints.

References stated in this discussion papers:

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. J. Boxall, A. Saul, J. D. Gunstead, and N. Dewis, ‘Regeneration of Discolouration in Distribution Systems’, in *World Water & Environmental Resources Congress 2003*, Philadelphia, Pennsylvania, United States, 2003, pp. 1–9.

3. W. R. Furnass, ‘Modelling both the continual accumulation and erosion of discolouration material in drinking water distribution systems’, Doctorate of Philosophy, University of Sheffield, Sheffield, 2015.

4. E. J. M. Blokker and P. G. Schaap, ‘Temperature Influences Discolouration risk’, in Computing and Control for the Water Industry (CCWI), United Kingdom, 2015, vol. 119, pp. 280–289.