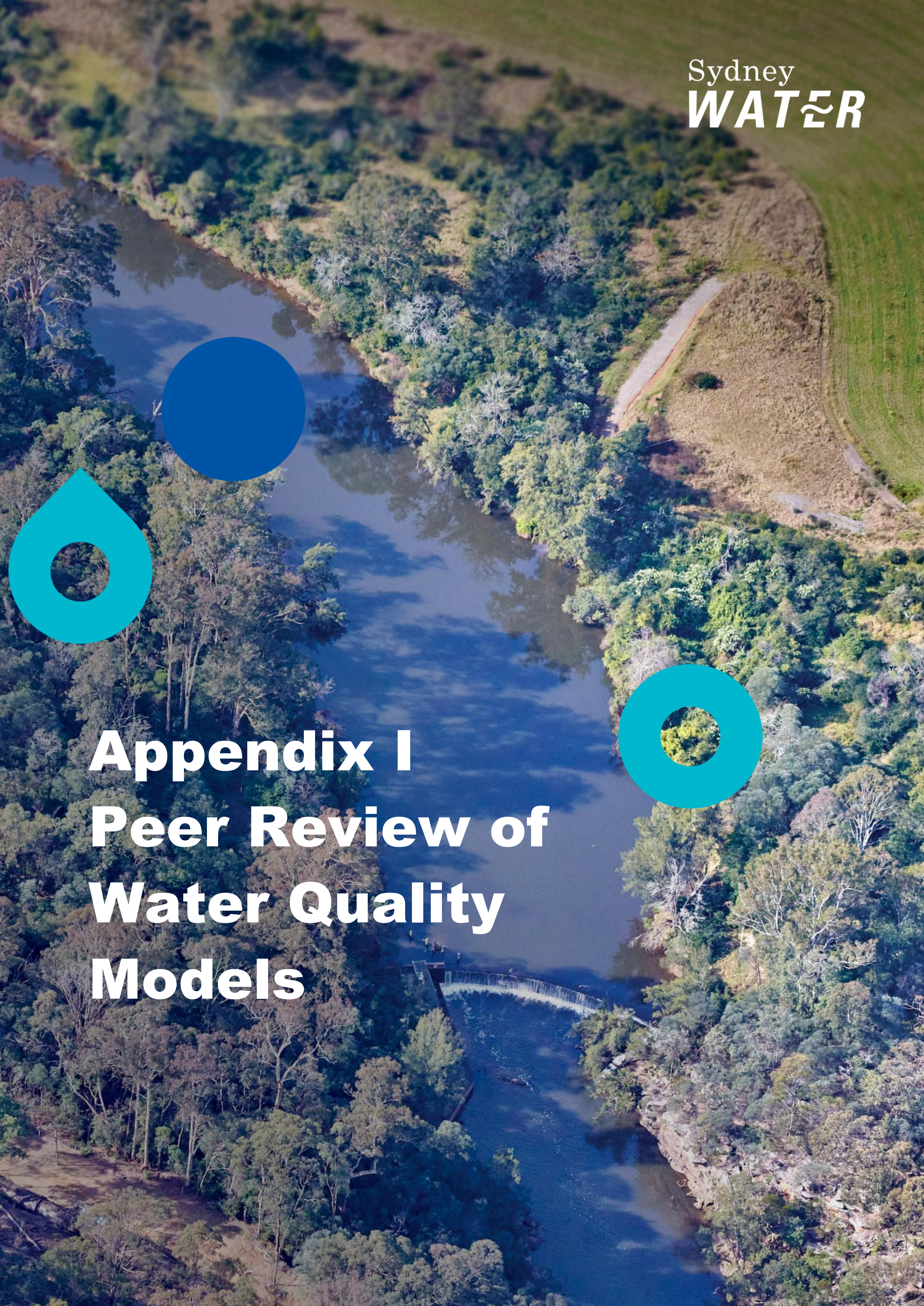

Sydney WATER

Appendix I Peer Review of Water Quality Models

12th March 2021

WRL Ref: WRL2020058 BMM L20210312

COMMERCIAL IN CONFIDENCE

Mr Paul Dunne
Senior Water Resources Engineer
Integrated System Planning, Liveable City Solutions
Sydney Water, Level 10, 1 Smith Street, Parramatta NSW 2150

UNSW SYDNEY
Water Research Laboratory
School of Civil and Environmental Engineering

By Email: paul.dunne@sydneywater.com.au

Dear Paul,

South Creek STP Environmental Impact Assessment Peer Review of Water Quality Models

1. Introduction

I, Brett Miller, am the Principal Engineer for Hydraulics and Modelling at the UNSW Water Research Laboratory. I was engaged by Sydney Water to review the calibration of the Hawkesbury Nepean and South Creek hydrodynamic and water quality modelling as documented in *Beling E., Kumandur K., and Hipsey M: "Hawkesbury Nepean and South Creek TUFLOW FV and AED2 Model Calibration Report", Final Draft 25/1/2021* ("the 2021 report").

This modelling exercise has been undertaken to update the calibration of the Hawkesbury Nepean modelling system as developed in 2014 while also developing a hydrodynamic and water quality model of South Creek.

The model input and output files have not been reviewed directly.

This 2021 calibration report makes reference to the previous calibration *report "Water Quality Modelling of the Hawkesbury-Nepean River System - Hawkesbury-Nepean River and South Creek Model", Final Calibration Report, 24 February 2014.* I have not reviewed the 2014 report in detail, but have referred to it where necessary to understand aspects of the 2021 report.

The catchment modelling report *"Hawkesbury-Nepean Source Model Update and Calibration", December 2020* is understood to have been previously reviewed by others. I have not reviewed this 2020 report except where necessary to understand the boundary conditions to the 2021 report.

The calibration report states that one driver for model development was the Environmental Impact Statement for the Upper South Creek Advanced Water Recycling Centre. I have also received a specific list of scenarios are to be considered with the South Creek model.

2. Summary of Findings

From my review of the calibration report and my understanding of the scenarios that are to be assessed, I consider that the model is suitable for running these scenarios. <u>However, scenario results should only be compared against each other as relative changes and not used for considering absolute water quality concentrations (such as comparison with a water quality target).</u> The amount of relative change between scenarios should be compared to the sensitivity analysis (discussed below) when considering the effectiveness of any particular scenario. If the relative change between scenarios is less than observed in the sensitivity analysis, then additional studies may be required to confidently distinguish between the scenarios.

All water quality models are highly sensitive to the boundary conditions. I recommend that the catchment modelling (SOURCE), the sewer modelling (MOUSE) and the wastewater treatment plant (WWTP) modelling not be referred to in the report as "Supplementary Models". While each of these models has its own calibration report, the very minimal discussion in this receiving water quality calibration report does not adequately inform the reader of the uncertainties in these boundary condition models.

For the catchment runoff, I agree with the approach of using a single dry weather concentration and single event mean concentration for each catchment type. This is the simplest method supported by available data. However, considerable uncertainty remains in each adopted constituent concentration. The calibration report does not present any sensitivity analysis of the resulting estuary water quality to the catchment runoff concentrations. A comprehensive sensitivity analysis should be undertaken and the results of the sensitivity analysis included in the calibration report.

The standard method of separating datasets into a calibration period and a verification period has not been undertaken. Further, statistical analyses of model results against data have not been presented. As such, it is not possible to quantify the water quality model's ability to reproduce an independent dataset. Without this quantification of model performance, I recommend that the sensitivity analysis be comprehensive.

The 2021 calibration report relies heavily on information previously documented in the 2014 calibration report. However it is apparent that there has been ongoing development of the catchment, hydrodynamic and water quality models. This 2021 calibration report would benefit from having complete documentation of all assumptions, parameters, calibration and verification within a single report. Given the complexity of the system, the amount of development undertaken and the roadmap of future development, a single, stand-alone document would ensure that all model stakeholders and users of the model outputs are aware of the confidence, assumptions and limitations of the models. This observation is provided as a suggestion and has not influenced the current calibration review.

3. General Comments

The suite of model software tools adopted in this calibration report are appropriate for the scenarios and conditions to be modelled. I consider these modelling tools to be in accord with industry best practice.

I consider that the mesh resolution, the model domain, and the location and types of boundary conditions are appropriate.

The SOURCE model has been used to provide a timeseries of catchment runoff. The concentrations of water quality constituents have been prescribed as either dry weather concentrations (DWC) or event mean concentrations (EMC). I consider this approach to be appropriate, but the report would benefit from presenting the uncertainty in each of the adopted values and documentation of a sensitivity analysis in the water quality results.

The TUFLOW-FV model has predicted flow velocities, depths, temperature and salinity at each node at each timestep. The water quality model AED2 has modelled: dissolved oxygen, silica, nitrogen, phosphorus, carbon, suspended sediments, planktons and pathogens in various forms (particulate and dissolved) and suspended or in the bed sediments. I consider this to be a suitable hydrodynamic and water quality framework for this study.

As with many water quality models, several of the water quality constituents do not have field measurements for calibration or verification, but are assumed to be necessary for the cycling of other measured constituents (such as nutrients). I presume that the 2014 investigations identified the range of water quality constituents and interactions necessary to adequately reproduce measured water quality. On the basis of the available field measurements, I do not believe that the water quality modelling warrants any more complexity.

The report does not explicitly state the timestep used for each of the models. The 2014 report discusses the implications of using a catchment model with a daily timestep to provide boundary conditions for a sub-daily timestep water quality model. This is particularly relevant for short discharge events and warrants further discussion in the report.

Modelling of South Creek and the Hawkesbury Nepean in separate hydrodynamic and water quality models has the advantages of greater modelling speed for scenarios in South Creek. However, the potential disjoint between the models has only discussed the implications to the Hawkesbury Nepean. The calibration report should explicitly state that any changed in water quality predicted by the Hawkesbury Nepean model cannot influence scenarios run in the South Creek model. The lower part of the South Creek model is in the tidal pool and with the disjoint there is no way that WQ in the lower South Creek can be influenced by water quality in the Hawkesbury Nepean, which might occur in the prototype.

The Hawkesbury Nepean model has sections of Manning's 'n' roughness up to 0.07. The South Creek model has considerable amounts of the mesh with Manning's 'n' roughness greater than 0.09. This is a very high value that is presumably required to maintain water in the steeper upstream areas of the model during dry periods. In reality the watercourse may either dry or become a series of riffles and pools. While I do not object to this approach, the implications to water quality during dry periods should be discussed.

4. Specific Comments on the Calibration

4.1 Overview

The calibration report does not state a calibration period and a validation period. This is a major shortcoming in the approach. It is a standard approach in modelling to separate datasets and use one part to tune (calibrate) model parameters and another part to independently verify the models performance.

Statistical analysis of model results against data has not been presented. As such, quantitative terminology that infers statistical analysis such as "high correlation" should not be used in the discussion. While I agree that a visual inspection of model timeseries versus measurements will often provide greatest insight into processes, I recommend that statistical analysis also be considered to quantify which model regions and constituents can have the greatest confidence placed upon them. Ideally the statistical analysis would be completed separately for the calibration period and a validation period.

The report uses a zonal analysis and a transect analysis to visually assess the performance of the modelling in different ways. A clearer set figures of for the locations of each graph would greatly assist reading the report. The report lacks a plot presenting the transect distances for the South Creek model.

I appreciate the zonal data analysis method provides an assessment of the variability within a "zone". However if the model is being verified over a zone and not to individual sampling points, it raises the question of whether the model needed to be run at the current resolution. This is an observation and does not require action, but coarsening the model may provide opportunities to improve run-times.

The report should explain that the medians in the transect plots are the temporal medians over a period of time, whereas the medians in the zonal plots are the spatial medians across the zone.

On all zonal plots, it would be useful to include the representative distance from the estuary to allow for easier comparison with the transect plots.

It would be useful to include on the transect plots and zonal timeseries plots some indication of dry, moderate or wet periods to allow for better understanding of the inflow conditions.

4.2 Hawkesbury Nepean – Comparison of Model with Measurements

My observations are as follows:

- Water level comparisons show that the model generally underestimates the tidal range above Wisemans Ferry. This presumably also corresponds to an underestimation of the tidal exchange and hence mixing.
- The flow and flow frequency comparisons were not presented in this calibration report but rather referred to the 2014 report. Given that the catchment modelling has been updated, I believe that this comparison should be re-run and presented.
- Observations in the salinity transect plots demonstrate that the estuary responds quickly to inflow events. Including up to three months of data on a single transect plot creates considerable scatter in both data and models making it difficult to assess the model performance.
- Salinity timeseries in the zonal analysis demonstrates that the model is reproducing the saline dynamic of the estuary generally well. The model appears to overestimate saline intrusion events in Zone 2 indicating that either the model is underestimating inflows or overestimating dispersion in the middle of the estuary.
- The modelled temperature matches the measurements well.

- The temperature transect plots all show a drop in temperature at around 150km which should be explained. This variability is also noted in other constituents. Model stability should be investigated.
- The modelled temperature in Zone 2a-Box 6 during 2012/2013 shows an excessively large range that is not evident at other locations. This plot and the model stability should be checked.
- There are few suspended sediment measurements to compare with model predictions. The model generally seems to be overpredicting suspended sediment. Like temperature, there appears some model instability in suspended sediment at 150km.
- Turbidity measurements are well reproduced by the model in some periods and overestimated by the model in others.
- Dissolved oxygen (DO) measurements appear to be well reproduced throughout the model. DO has been presented as mg/L, whereas presentation as percentage of saturated DO often assists in understanding oxygen demand. Like temperature, the DO predictions about 150km are unusual.
- The ammonium transect plots show a considerable range of measurements at 210km which is presumable a point source. However, the model does not reproduce this range or peak in data, indicating that a source is potentially not adequately included as a boundary condition.
- The ammonium measurements have considerable variability in Zone SC that is not reproduced in the model. This would indicate that there was a difference between South Creek actual discharges and predicted discharges.
- The range of oxidised nitrogen measurements are not reproduced by the model. This most likely indicates that actual point source discharges are not being adequately represented in model boundary conditions.
- The model is representing the general peak in oxidised nitrogen in the middle of the estuary. The zonal plots indicate that the model is generally underestimating the total oxidized nitrogen.
- The total nitrogen measurements and model predictions follow the same trends as the ammonium and total oxidised nitrogen.
- The model appears to overestimate total phosphorus (TP) and filterable reactive phosphorous (FRP) throughout the entire estuary. This would indicate that the boundary condition concentrations are too high. The transect plots indicate model peaks in FRP at key locations (e.g. 120km, 130km 210km) which indicate that point source FRP concentrations may be overestimated in the model.
- The Chlorophyll-a measurements and model predictions both show considerable variability and are generally not well matched. The measurements show peaks about 100km which are outside of the range of model predictions during many time periods. The zonal timeseries plots show a poor to average match with the data.
- There were very few e-coli measurements and as such the e-coli model cannot be considered verified.
- The enterococci measurements were limited and located around the 140km to 180km reach. Neither the timeseries not the transect plots showed good match between model and data. I note that Zone 3 Box 4 shows significantly greater variability in model predictions than Zone 3 Box 4, while the measurements appear very similar. This warrants further uncertainty discussion.

4.3 South Creek – Comparison of Model with Measurements

My observations are as follows:

- There are no observations in Eastern Creek or McKenzie Creek to assess the model performance in those areas. These areas should be considered unverified.
- Both the Source and TUFLOW models reproduce the timeseries of flow by approximately matching the peak discharges in each event. However, the flow frequency curves show significant differences between modelled flows and measured flows in the lowest 20%. This may be particularly relevant for scenarios concentrating on dry weather periods and as such warrants further discussion.
- The model predictions match the salinity measurements well, noting that both are below 0.5ppt at all times.
- The model predictions match the temperature measurements well.
- The total suspended solids model predictions show a poor to average comparison with the measurements. Different trends were observed in TSS for the different periods which may correspond to land-use changes in those periods.
- Dissolved oxygen (DO) predictions match well with measurements, noting that there is very limited temporal variability and that the seasonal variability is directly related to water temperature. Presentation of the model and data as percentage of saturated DO often assists interpretation of oxygen demands.
- The ammonium measurements fall within the range of the model predictions, as also does oxidised nitrogen and total nitrogen. The tidal zone measurements show variability greater than the median prediction, however this may pertain to the phase of the tide when measurements were taken.
- The measurements of Total Phosphorus and Filterable Reactive Phosphorous are matched well by the model.
- The Chlorophyll-a measurements and model predictions both show considerable variability and are generally not well matched. In particular, the 2014/2015 and 2017/2018 model predictions overestimate the measurements. While the model range still covers the measurements, the model median is several times higher.
- There were no e-coli measurements for comparison with model predictions.
- There were very few enterococci measurements. The model significantly overpredicted enterococci in the non-ephemeral zone.

4.4 Calibration Report Discussion and Conclusions

The observations made in report sections 4.2.1 and 4.2.2 discussing model performance are reasonable, but I recommend that the discussion concentrate more on the confidence in the model. Water quality modelling is inherently complicated with many more parameter and source terms than can be accurately determined from typical data collection programs. With only one final calibrated scenario presented (for each model) it is very difficult for the reader to interpret what processes are being affected by various boundary conditions and rates. Presentation of a comprehensive sensitivity analysis would greatly assist with this discussion and interpretation of how much confidence can be placed on model predictions.

5. Summary

Thank you for the opportunity to undertake this review. Should you have any queries, please contact me on B.Miller@wrl.unsw.edu.au or 0414 385 491

Yours sincerely,

Brett Miller
Principal Engineer – Hydraulics and Modelling

4th November 2021

WRL Ref: WRL2020058 BMM L20211104

UNSW
Water Research
Laboratory

Mr Paul Dunne
Senior Water Resources Engineer
Integrated System Planning, Liveable City Solutions
Sydney Water, Level 10, 1 Smith Street, Parramatta NSW 2150

By Email: paul.dunne@sydneywater.com.au

Dear Paul,

South Creek STP Environmental Impact Assessment
Peer Review of Water Quality Models

I, Brett Miller, am the Principal Engineer for Hydraulics and Modelling at the UNSW Water Research Laboratory. I was originally engaged by Sydney Water to review the calibration of the Hawkesbury Nepean and South Creek hydrodynamic and water quality modelling as documented in:

> *Beling E., Kumandur K., and Hipsey M: "Hawkesbury Nepean and South Creek TUFLOW FV and AED2 Model Calibration Report", Final Draft 25/1/2021.*

This modelling exercise was undertaken to update the calibration of the Hawkesbury Nepean modelling system as developed in 2014 while also developing a hydrodynamic and water quality model of South Creek.

I provided my findings in a letter report dated the 12th March 2021.

Subsequent to my letter, Sydney Water undertook additional investigations and reporting. I was supplied with the updated calibration report:

> *Beling E., Kumandur K., Hipsey M., and Huang P.: "Hawkesbury Nepean and South Creek TUFLOW FV and AED2 Model Calibration Report", August 2021*

The concerns raised and suggestions made in my letter of the 12th March 2021 have been addressed. I am pleased to have been given the opportunity to provide the initial and this subsequent review. In my opinion, the calibrated model is suitable for running the scenarios that are to be considered for the Environmental Impact Statement for the Upper South Creek Advanced Water Recycling Centre.

The calibration report includes the necessary information to help interpret all scenario model predictions.

The quality of the calibration has been possible due to the extensive data collection program. The inclusion of statistical analysis of the calibration and verification has strengthened the report. Tables 4.2 to 4.6 of the report provide a useful summary of the confidence that can be placed on each parameter in each part of the waterway.

I recommend that the statistical analysis of the calibration and verification in Section 4 should also report:

- The equations used for each of the four statistical measures.
- The number (n) of "samples" vs "model" data points that were used in each period, parameter and waterway zone.
- Definition of what quantitative measures comprised "poor", "acceptable" and "accurate".
- Referencing of statistical and modelling papers as to why these values were adopted.

In my previous review I included bullet points of my observations of the comparison between model and measurements. These comments are largely still valid, but they have been addressed or included in the calibration report discussion.

Sensitivity analysis has been undertaken to assess the effect of varying the boundary conditions including catchment runoff upon concentrations within the waterway. This has identified that the absolute value of concentrations in the waterway are highly dependent upon these uncertain boundary conditions.

My previous review recommended that scenario results should only be compared against each other as relative changes and not used for considering absolute water quality concentrations. The sensitivity analysis now allows for predicted absolute water quality concentrations to be interpreted within a sensitivity range. As such I now consider that scenario results can be interpreted both relative to each other and as absolute concentrations within the sensitivity range.

Other specific comments on the report are listed below as bullet points.

- Section 2.3.2 discusses the interface between the two models. However this is the first time in the report that the Hawkesbury Nepean and South Creek models are identified as separate models. There should be an introductory sentence about them being separate before discussing the interfaces.

- Section 3.1.2 presents Mannings "n" values. The report states that the high values of Mannings "n" are only on the overbank, however this isn't apparent in Figure 3.3 which has high values but does not plot overbank areas. For example, the whole reach from the South Creek junction to Penrith is plotted as n= 0.07. While this is not necessarily a problem, the report should state why the values are so high and what they physically represent.

- In Figure 3.11 it is difficult to interpret the start and finish of each of the boxes and zones. Upon request, I was supplied with a table of chainages which should also be included in the report. Figures 4.1 to 4.7 show similar information as Figure 3.11, but are a better presentation style. You could consider removing or updating Figure 3.11.

Thank you for the opportunity to undertake this review. Should you have any queries, please contact me on B.Miller@wrl.unsw.edu.au or 0414 385 491

Yours sincerely,

Brett Miller
Principal Engineer – Hydraulics and Modelling