Numerical comparability of ISO 7027-compliant nephelometric turbidity sensors commonly used in New Zealand

Prepared for Envirolink

March 2021

Prepared by:
Rob Davies-Colley
Stephan Heubeck
Ron Ovenden
Alex Vincent
George Payne
Andrew Hughes

For any information regarding this report please contact:

Rob Davies-Colley
Principal Scientist - Water Quality
Catchment Processes
+64-7-856 1725
rob.davies-colley@niwa.co.nz

National Institute of Water & Atmospheric Research Ltd
PO Box 11115
Hamilton 3251

Quality Assurance Statement		
[signature]	Reviewed by:	Dr Murray Hicks
[signature]	Formatting checked by:	Juliet Milne
[signature]	Approved for release by:	Dr Neale Hudson

Cover photo: Adding a highly turbid slurry of the layer clay mineral, kaolinite, to the turbidity sensor testing tank. [Rob Davies-Colley, NIWA]

Recommended citation:
Davies-Colley, R., Heubeck, S., Ovenden, R., Vincent, A., Payne, G., Hughes, A. (2020) Numerical comparability of ISO 7027-compliant turbidity sensors commonly used in New Zealand*. NIWA Client Report 2020-282HN.* 61 p.

Contents

Tables

Figures

Executive summary

Nephelometric turbidity, a relative measure of light scattering, is widely used as an index of 'water cloudiness' or haze, and as a surrogate for water clarity or suspended particulate matter (SPM) concentration. Turbidity is a very convenient metric, being simple to measure, and sensors are relatively inexpensive and widely available. However, despite correct calibration to formazin (the most widely used calibration standard), different turbidity sensors can produce different numerical outputs on the same natural water containing natural suspended particulate matter (SPM). This issue has long been recognised, and turbidity sensor design standards have been developed to improve the comparability of sensors. The ISO 7027 standard[1], specifying 90° scattering of near-infra-red (NIR) radiation in the 830-890 nm range, has been adopted by many turbidity sensor manufacturers. In New Zealand, the National Environmental Monitoring Standard (NEMS) for turbidity (NEMS 2016) requires use of ISO 7027 sensors for high frequency field-based monitoring[2]. The NEMS *Turbidity* also recommends validation of field measurements with ISO 7027-compliant sensors. These recommendations appear to expect the response of different ISO 7027-compliant sensors to be closely numerically comparable.

Field experience from a NIWA site on the Mataura River with multiple ISO 7027-compliant (or near-compliant) turbidity sensors operating side-by-side suggested appreciable numerical difference in responses over hydrograph events. A preliminary laboratory tank experiment with natural river SPM (Hughes et al. 2019) also suggested considerable divergence in numerical response of the same make-models of turbidity instruments.

MBIE Envirolink funding was sought by Marlborough District Council on behalf of the regional sector to conduct a new experimental comparison of turbidity sensors. The goal was to rigorously test comparability of a wider range of ISO 7027-compliant turbidity sensors on different (optically contrasting) fine particulate matter suspensions. The following ISO 7027 field sensors (widely used by regional councils) were tested in triplicate:

- YSI EXO2 Sonde-turbidity.

- Hach Solitax t-line sc.

- WTW VisoTurb.

- Observator Analite NEP5000.

Additionally, two ISO 7027-compliant laboratory bench turbidity instruments, in which samples are contained in cylindrical cuvettes, were also tested: Hach TL2310 LED and Hach 2100Q-is.

The day before the 3-day campaign of tank experiments, all of the tested sensors were calibrated on freshly-made formazin standards in a ratio progression: 1,000, 250, 100, 25 and 10 FNU – so as to put all sensors on the same basis for comparison on test suspensions over this range of turbidity. The

---

[1] The latest version of ISO 7027 (ISO 7027-1:2016) specifies both nephelometry and *diffuse* attenuation (referred to as "turbidimetry") as valid turbidity measurement approaches (with attenuation reporting in formazin attenuation units – FAU). Note that (collimated) *beam* attenuation is fundamentally different, being an absolute metric (in SI units: m^{-1}).

[2] NEMS (2016) referenced ISO 7027:1999 which only specified nephelometers. ISO 7027-1:2016 was released shortly after NEMS (2016) Henceforth throughout this document, the term ISO 7027 refers to ISO 7027: 1999. We expect that a forthcoming update of the NEMS *Turbidity* will address ISO 7027-1:2016.

calibrations were re-run four days later, after all the tank experiments – to check for stability of calibrations (and any sensor drift).

Three different suspensions were used for the tank tests, which were performed over a wide range of concentrations:

- River stormflow SPM (mostly fine-medium silt; tested over an 81-fold concentration range).

- Kaolinite clay (81-fold concentration range).

- Phytoplankton (9-fold concentration range).

The river silt was concentrated from suspended sediment in storm-flows in a natural stream). Laboratory-grade kaolinite comprising plate-shaped particles) was used to provide a contrasting test suspension. (Kaolinite is a layer 1:1 alumino-silicate that is a common suspended constituent of rivers throughout NZ.) A suspension of phytoplankton, relevant to use of turbidimeters in standing waters such as lakes, was obtained from a eutrophic pond. All three test suspensions were carefully characterised as to suspended matter mass concentration and optical concentration (beam attenuation measured on a WETLabs C-Star transmissometer). Particle size distributions (PSD), measured on an EyeTech laser time-of-transit (obscuration time) instrument, showed all three suspended materials to be mostly in the coarse clay to medium silt range, with particle size modes around 5 μm.

Field sensors were tested (in triplicates of the same make-model) in a 170 L laboratory recirculation tank. For each test material we started with stock suspensions at high concentrations and tested the sensors over a ratio progression of successively lower concentrations. Water samples were taken at the beginning and end of measurements at each concentration, for measuring light beam attenuation and suspended particulate matter (SPM) mass concentration. The results from our experiments were plotted as calibrated turbidity in Formazin Nephelometric Units (FNU) versus optical concentration (light beam attenuation).

We found that different ISO 7027-compliant turbidity sensors respond proportionally and with fairly good reproducibility between sensors of the same make-model (average coefficient of variation for the higher concentrations typically better than about 10% at higher concentrations). This suggests reasonable numerical 'interchangeability' of sensors of the same-make model – which is important if a sensor has to be swapped out part-way through a field campaign.

However, the numerical output of the different make-models of sensor diverged appreciably and ranged more than two-fold in FNU units in all three test suspensions. The divergent output of different sensors, despite identical calibration to formazin, may be attributed to small differences in design features such as spectral emission of light source, spectral sensitivity of detector, detector angle (and angle-range) and beam configuration; all combined with the very different optical properties (light scattering as a function of angle) of natural suspensoids compared to formazin. Our experimental results suggest that even very subtle differences (e.g., differences, within tolerance, in the manufacture of components) between sensors compliant with the same international standard (ISO 7027) can combine with particle optics to influence turbidity sensor response.

Our finding that turbidity measurements are instrument-dependent, even for sensors compliant with the same international design standard, has important ramifications. In particular, reporting of turbidity relative to formazin as if it were an absolute quantity (e.g., as has been applied on some resource consent conditions) is best avoided. Furthermore, caution is needed when using discrete State of the Environment (SoE) turbidity datasets because measurements (historically in NTUs) may

not be entirely comparable, having likely been made on different instruments in different laboratories at different times. However, turbidity can be used as a *relative* measure (e.g., to indicate a change upstream versus downstream of a point source discharge). Field turbidity provides a valuable proxy for several sediment-related variables of interest in water quality (e.g., suspended sediment concentration (SSC), visual clarity) *if sufficient local calibration data are collected.*

1   Introduction

1.1   Background

Nephelometric turbidity quantifies the scattering of light in a fine particle suspension measured relative to that of a standard suspension, typically of formazin (e.g., Davies-Colley and Smith 2001). This metric is widely used to quantify 'water cloudiness' or haze in water, and to indicate water clarity and suspended particulate matter (SPM) concentration (Kitchener, Wainwright et al. 2017). Turbidity is a very convenient metric, being simple to measure, and both laboratory and field sensors are relatively cheap and widely available. Turbidity can be analysed on both discrete samples contained in cylindrical cuvettes by hand-held or laboratory bench turbidimeters, or at high frequency ('continuously') within streams, lakes, estuaries and marine waters using *in situ* sensors. Nephelometers measuring side-scattering of light (centred on 90°) are probably most common, but back-scattering sensors are also widely used, and, less often, scattering at other angle ranges is detected.

Despite its advantages, nephelometric turbidity is only a *relative* measure of light side-scattering – measured relative to formazin standards. Formazin is an *arbitrary* reference in the sense that many other stably light-scattering materials could have been employed. Turbidity units such as NTU (nephelometric turbidity units) and FNU (formazin nephelometric units) are *informal* units and cannot be universally converted to 'proper' scientific (SI) quantities. Furthermore, turbidity, as a relative measure of light scattering through a restricted angular range (e.g., 90 ±2.5° in the ISO 7027 standard) does not itself affect aquatic organisms or human use of waters. For this reason, turbidity is best used as a *proxy* for quantities which *do* affect aquatic habitat and water use – most typically visual clarity and SPM concentration.

It has long been known that different turbidity sensors vary in their numerical response in natural suspensions, despite identical calibration to formazin (Austin 1973; McCluney 1975). More recently, Rymszewicz et al. (2017) reported a five-fold range in output of 12 different turbidity sensors of various design principle. They also showed that virtually any turbidity sensor, irrespective of wavelength used, light scattering angles detected and numerical output, should be suitable as a proxy for suspended sediment concentration in sediment load calculations.

Numerical differences between turbidity sensors have been attributed to differences in optical design (e.g., scattering angle and angular range, operating wavelength of light source, spectral sensitivity of detector) (McCluney 1975). To address this issue, manufacturers now design their turbidity sensors to a particular standard (e.g., ISO 7027 and US EPA Method 180.1) with the expectation that this achieves numerical comparability. Many field-type sensors are designed to the ISO 7027 standard for nephelometry. (Note that the latest release of ISO 7027 also endorses turbidity measurement by diffuse attenuation, sometimes referred to as "turbidimetry" – with reporting in formazin attenuation units, FAU – see footnotes 1 and 2 of the Executive Summary.) Until recently, laboratory bench sensors were generally designed to the US EPA Method 180.1 standard (specifying side-scattering of visible light), however, ISO 7027, specifying side scattering centred on 90° of near infra-red (NIR) radiation in the 830-890 nm range of wavelengths, is becoming more prevalent for all sensor types, including laboratory instruments. One reason to favour the ISO 7027 standard is that interference by ubiquitous humic substances is negligible for instruments operating at NIR wavelengths because humics absorb NIR much more weakly than visible light (Kirk 2011).

The New Zealand (NZ) National Environmental Monitoring Standard (NEMS) for Turbidity (NEMS 2016) requires that ISO 7027 compliant sensors be used in high-frequency field-based monitoring of turbidity, and recommends validation of field measurements with an ISO 7027 compliant sensor – apparently on the expectation that the response of different ISO 7027 sensors is closely comparable.

However, recently-published research (e.g., Rymszewicz et al. 2017) showed that the response of different turbidity sensors, including side-scattering meters operating in the NIR), can vary appreciably in the same suspension. Recent field experience with multiple ISO 7027-compliant (or near-compliant) turbidity sensors deployed in parallel at the NIWA monitoring site on the Mataura River, suggests appreciable numerical divergence in response, with a two-fold range in output over hydrograph events for the following sensors (in order of FNU output): Observator Analite > Forest Technologies DTS-12 > EXO2 sonde turbidity – calibration problems or drift seem highly unlikely to explain this diversity of response (Evan Baddock, NIWA-Dunedin, pers. comm.). A preliminary laboratory tank experiment, using river stormflow sediment at different concentrations (Hughes et al. 2019), also suggest considerable divergence in numerical turbidity readings from these same ISO 7027 sensors.

The present experimental campaign was prompted by a discussion of the pending review of the NEMS *Turbidity* with the NEMS Steering Committee on 20 September 2019. As a result, an MBIE Envirolink Large Advice Grant (2044-MLDC155 lodged by Marlborough District Council) was sought to improve and expand the earlier experiments with a wider range of *in situ* turbidity sensors (commonly used in NZ), a wider range of suspensoids, and more robust experimental protocols.

1.2 Scope

This study comprised a series of laboratory tank experiments designed to test comparability of different ISO 7027-compliant turbidity sensors commonly used in NZ by regional councils and other agencies for measuring turbidity *in situ* or on discrete water samples. After careful calibration on the *same* series of formazin (turbidity standard) suspensions, the sensors were tested in three different 'natural' suspensions: river stormflow silt, kaolinite (layer clay) and eutrophic pond water (laden with phytoplankton).

The main objective of this study was **to compare the numerical outputs of turbidity sensors designed to the ISO 7027 standard that are commonly used in NZ.**

A subsidiary objective was to test the comparability of ISO 7027 compliant turbidity sensors of the *same make and model* – effectively to test sensor reproducibility and therefore 'inter-changeability' (in terms of response to a standard formazin suspension).

2   Methods and materials

2.1   Approach

This comparison of turbidity sensors was inspired by the work of Rymszewicz et al. (2017) and builds on a preliminary comparison reported by Hughes et al. (2019). Our approach was to run several ISO 7027-compliant sensors in the same test tank at different concentrations of natural suspensions. Ideally the sensors would have been run *simultaneously* (side-by-side) in the test tank, but side-by-side deployment in preliminary work (Hughes et al. 2919) was questionable because of potential 'over-crowding', disrupting tank mixing with mutual interference of sensors. Instead we chose to run the sensors *sequentially* (in triplicate), while checking that (1) no mutual interference occurred (with triplicate sensors spaced about 150 mm apart) and (2) no change in test suspensions occurred during the sensor deployments.

For the experiment we gathered a range of different ISO 7027-compliant turbidity sensors that are commonly used throughout NZ by regional councils and other agencies. The six make-models tested included four different field-type (*in situ*) sensors (in triplicate) plus two different cuvette instruments. The field sensors were deployed in a 170 L recirculating tank designed for calibrating optical backscatter sensors (another type of turbidimeter) in coastal waters. Water samples taken from the test tank over a 100-fold ratio progression of dilution (and thus concentration) were used for verifying concentrations and testing in the cuvette turbidity instruments. The sensors were tested in three strongly (optically) contrasting suspended materials: river stormflow silt, kaolinite (layer clay) and phytoplanktonic algae (in eutrophic pond water). Our approach was broadly similar to that of Rymszewic et al. (2017), except that we compared only ISO 7027 compliant turbidity sensors (on a wider range of suspensoids), and underpinned all turbidity measurements with an absolute metric – the beam attenuation coefficient (beam-c) as an 'absolute' measure of optical concentration.

2.2   Turbidity sensors

Field *in situ* sensors (tested in triplicate) were:

- EXO turbidity sensors (as deployed on YSI EXO2 sondes).

- Hach Solitax t-line sc.

- WTW VisoTurb.

- Observator Analite NEP5000.

Cuvette turbidity instruments tested were**:**

- Hach TL2310 LED (NIWA water quality laboratory benchtop instrument).

- Hach 2100Q-is (hand-held cuvette instrument).

These instruments are compared and contrasted in Table 2-1. Note that warmup time is typically stated by turbidity sensor manufacturers, but was not relevant to our experiments because all sensors were kept 'on' and running continually for all calibrations and tank tests.

In preliminary work (Hughes et al. 2019), Forest Technology Systems DTS-12 sensors were also tested. DTS-12 sensors were *not* included in the present study because this make-model does not fully conform to the ISO 7027 standard[3].

2.2.1   Hach TL2310 LED

The Hach TL2310 LED turbidimeter is a laboratory bench instrument, using cylindrical cuvettes (Figure 2-1). For the comparison experiments we used the Hach TL2310 LED instrument located in the NIWA-Hamilton Water Quality Laboratory. This turbidimeter is validated every working day against manufacturer-supplied StablCal^{TM} gel standards (ranging between <0.1 – 1,000 FNU) and is fully recalibrated every 3 months.
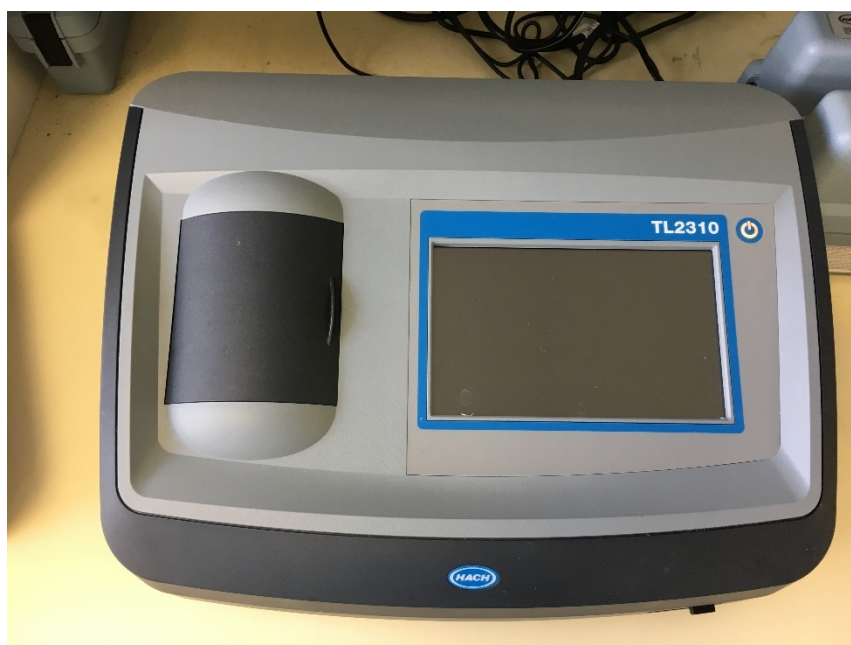
Figure 2-1:   **Hach TL2310 LED turbidimeter.**   This unit is in the NIWA-Hamilton Water Quality Laboratory.

---

[3]The DTS-12 laser diode has a peak wavelength of 780 nm, slightly outside the range of 830-890 nm specified in the ISO 7027 standard.

Table 2-1: **Turbidity sensors tested.** All six instruments are near-infrared nephelometers complying with the ISO 7027 standard. Field *in situ* sensors were tested in triplicate.

Make-model	Serial numbers	Tank rod numbers	Nominal range	Cleaning system	Display and control	Signal sampling	Notes (including calibration)
Field (*in situ*) instruments							
EXO turbidity 'smart sensor'	port "1"	EXO2 sonde on a single rod (with 3X turbidity)	0-4,000 FNU (0-1,000 at high precision)	Rotary brush (controlled by handheld)	Sonde hand-held (logger)	Sampling twice per second	NIWA equipment. Calibrated EXO turbidity sensors all loaded on one YSI EXO2 sonde
	port "3"						
	port "6"						
Hach Solitax ts-line sc	1844051	1	0-4,000 FNU	Brush unit (operated via controller)	2 X Hach SC200 Controllers	Controller screen updated every second	NIWA equipment (used without special calibration)
	1516088	2					
	1440664	3					
WTW Visoturb 700	18411145	4	0-4,000 FNU	Ultrasonic	1 X WTW Controller	Controller screen updated every second	Provided by Environment Southland (used without special calibration)
	12491426	5					
	18411139	6					
Observator Analite NEP5000	109931 (Adr0)	7	0-5,000 FNU	Brush unit (operated by Power Off/On)	Ecologger with Lap-top running Starlog V4	Sampling every 60 s using SDI-12	NIWA equipment. Calibrated and programmed by IS
	110693 (Adr1)	8					
	108195 (Adr2)	9					
Cuvette instruments							
Hach TLD2310LED	2018030C0015	N/A	0-1,000 FNU	N/A	LED screen		NIWA WQ Laboratory bench instrument
Hach 2100Qis	18070C009415	N/A	0-800 FNU	N/A	LED screen	Not stated in manual	NIWA portable (battery-powered) cuvette instrument

2.2.2 Hach 2100Q-is

The Hach 2100Q-is turbidimeter is a portable (battery-powered) unit used for analysing samples in cylindrical cuvettes (Figure 2-2). A NIWA-owned Hach 2100Q-is instrument was used for the experiments. Prior to analysing samples, the calibration procedure (as outlined in the manufacturer's manual) was followed. This involved selecting the 'calibration mode' and using the supplied StablCal^{TM} gel standards (ranging from 20 to 800 FNU) to internally recalibrate the sensor response.

Figure 2-2: **Hach 2100Q-is portable cuvette turbidimeter.** One of the cylindrical cuvettes (with black screw top) used to contain water samples is shown to the left of the instrument.

2.2.3 YSI turbidity sensors (on EXO sondes)

YSI EXO Sondes are multi-variable water quality measuring instruments with on-board logging or external readout. The EXO2 Sonde (Figure 2-3) collects data from up to six user-replaceable sensors. These sensors are all in the same plane such that a rotary brush system can clean all the sensors simultaneously. A variety of water quality variables can be measured by specific sensors, including turbidity. The calibration procedure for these units is carried out through the KorEXO software and any necessary adjustments in the calibrations are programmed into the 'smart' sensors themselves. Three turbidity sensors were loaded onto the one EXO2 sonde body for the duration of the comparisons reported here. These sensors are the smallest of all the field-type sensors tested at only 16 mm diameter, and their ellipsoid-shaped sampling volume is within about 5 mm from the optical window. Calibrations were performed in distilled water and user-mixed formazin suspensions ranging from 0 to 1,000 FNU. The brush wiper was activated to remove any potentially interfering air bubbles before recording turbidity data.

Figure 2-3:   Mounting the EXO2 Sonde in the test tank. Three turbidity sensors were fitted to the same sonde body (blue cylinder). The sonde is taped (pink fluoro tape) to an aluminium rod for mounting the sensors (at the correct height and orientation) using retort clamps to fix the rod to a stage above the tank. The tank contained pond water (suspended phytoplanktonic algae) when the photograph was taken.

2.2.4   Hach Solitax

The Hach Solitax ts-line sc turbidity sensor has a nominal 0-4,000 FNU range. The sensors each have a brush wiper unit which was operated at the start of each deployment via the SC4200 controller unit (Figure 2-4). Three NIWA-owned sensors (and two Hach SC200 controllers) were used in this work. The units were used as factory-calibrated. The manufacturer's manual suggests use with at least 30 mm clearance from the optical face. Sensor output (at one reading per second) was read on the controller unit and logged onto SD cards.

Figure 2-4:   Hach Solitax sensor with controller unit. Note the brush wiper unit that wipes across the face of the senor to keep the optical windows clean.

2.2.5   WTW VisoTurb

The WTW VisoTurb 700 IQ turbidity sensor has a nominal 0-4,000 FNU range (Figure 2-5). Instead of a brush wiper unit, these sensors have an ultrasonic cleaning system, but this was not operated in our experiments because fouling was not expected. Instead the WTW sensors were manually brushed in an attempt to be consistent with other sensors with integral brush units. The manufacturer provides graphs suggesting a clearance of order 50 mm to avoid offsets due to reflections at low turbidity (say < 10 FNU). We were loaned three WTW VisoTurb sensors by Environment Southland as well as one WTW controller (accepting two of the sensor units). We therefore deployed two of the three sensors together in the test tank, both connected to the one controller, but, to avoid having to restart (and wait for warm-up), measurements with the third sensor were conducted later (in a beaker) using samples taken at each concentration. We verified (by swapping around the WTW sensors) that the results were indistinguishable within experimental error in the test tank versus a laboratory beaker. The WTW controller had no on-board logging facility so we simply read the turbidity response (updated every second) on the controller display and recorded seven readings over at least two minutes. These sensors were used as supplied (without prior calibration).

Figure 2-5:   WTW VisoTurb turbidity sensor. (Storage endcaps removed).

2.2.6 Observator Analite NEP5000

The Observator Analite NEP 5000 turbidity sensor is an SDI-12 compatible *in situ* nephelometer with a 0-5,000 (nominal) FNU range (Figure 2-6). Three Observator sensors were provided by NIWA-Instrument Systems, after calibration to formazin standards and setting to auto-range in the 0-3,000 FNU range (for which they use the 0-100 and 100-1,000 FNU ranges). The manufacturer suggests use with a 50 mm setback from the optical face. The Observator sensors each have a small brush wiper unit which was operated at the start of each deployment by switching the power supply momentarily off, then on again. Data from the Observator sensors was logged by a NIWA Ecologger, and 1 minute-averaged sensor output was recorded by a laptop computer using Starlog V4 software.

Figure 2-6:   Observator Analite NEP 5000 turbidity sensors. A housing (shroud on bottom right) protects the optical face of this sensor with its small rotary brush unit. The three sensors were connected to the NIWA Ecologger (shown) outputting to a laptop computer running Starlog V4.

2.2.7 C-Star beam transmissometer

Because there is no absolute (light scattering) standard for turbidity, we referenced turbidity sensor response in our test suspensions to 'optical concentration' (in absolute units of beam attenuation), as well as mass concentration. Therefore, the light beam attenuation coefficient (beam-c) was measured on each sample (diluted as appropriate) as a reference and check of dilution accuracy. The beam attenuation coefficient (m^{-1}) measures total light (beam) attenuation by both absorption and scattering (at *all* angles) of both dissolved and particulate constituents. Beam attenuation has the advantage that it can be measured very precisely (CV of replicates better than 3%), much more so than TSS. Beam attenuation is closely (inversely) related to visual clarity measured as visual range of a black body (Zanevald and Pegau 2003).

We used a green light (530 nm) WET Labs C-Star beam transmissometer of 250 mm path length, set up in the laboratory in flow-through mode (Figure 2-7). The beam transmissometer could not be deployed directly in the test tank because it would have over-ranged in all but the lowest concentrations used. Instead we measured beam attenuation on test tank samples after appropriate (volumetric) dilution.

To standardise the C-Star instrument, very pure water in the light beam path was used to set 100% transmittance[4].

Figure 2-7: **WET Labs® C-Star beam transmissometer set up in flow-through mode for laboratory measurements of beam attenuation coefficient (beam-c).** L to R: C-Star sensor with flow-tube fitted (vertical mounting), peristaltic pump, diluted sample (stirred by magnetic stirrer), pure water storage flask, and laptop computer running Terra Term web 3.1). The C-Star emits a beam of green light (530 nm) and measures the fraction transmitted through the water sample.

2.3 Local calibration of sensors to formazin

All sensors compared here were 'locally calibrated' by sequential measurements in the *same* formazin standards (treated like another sample suspension) over the 10-1,000 FNU range – which encompasses the turbidity in our natural suspensions except for the very lowest concentrations. These measurements provided relationships for each sensor between formazin reference value and raw sensor output, and these calibration equations were then used to adjust the raw sensor output to calibrated FNUs. This local calibration was performed both before and after deployment in the test tank in 'natural' suspensions. The rationale here was two-fold:

- ▪ To put all sensors on a common basis for comparison.

- ▪ To check that drift of response had not affected any of the sensors during the tank experiments.

Sensor responses were recorded on freshly-made formazin standards at 1,000, 250, 100, 25 and 10 FNU. The standards were all made at 1 L volume from stock 4,000 FNU formazin (Catalogue 246149, expiry date February 2022) using volumetric laboratory glassware (volumetric flasks of 1 L, 250 mL and 25 mL and a 10 mL pipette were used). Measurements before the experiment were made on 6th June 2020. The formazin standards were then stored dark and refrigerated for the duration of the tank experiments – which ran for three days from 7-9th June 2020 inclusive. The formazin standards were reused (after warming to room temperature) to check sensor calibrations on 10th June 2020.

---

[4] For even greater accuracy, reference readings in air (clearer than the clearest laboratory water) may be used to standardise the instrument (with account taken of the light reflections at the two optical windows plus the light attenuation of water itself).

The calibration runs on 6th June were used to prepare calibration equations for providing turbidity in FNU relative to formazin on the same basis for all sensors. The sensor calibrations were then *validated* (on 10th June 2020 after completion of the test tank experiments) on the *same* formazin standards to check for drift – on the reasonable assumption (proven as shown below) that the standards (stored refrigerated and dark) were stable for this period. Kitchener et al. (2017) present a table showing that stability of formazin standards varies strongly with concentration, with standards above 20 FNU stable for at least one month. On that basis, the stability of only our lowest (10 FNU) standard over the five days of storage could be in doubt – and as we show, there was no evidence of change in that standard.

Sensor output in each standard was recorded *without any change in sensor setting*. That is, no attempt was made to adjust sensor output to account for any discrepancies compared to standards. The measurements on these standards were simply used to *numerically* correct the response of each sensor on natural suspensions to formazin nephelometric units (FNU) using the calibration equations.

Ideally the calibration on fresh formazin standards would have been conducted in the same 170 L laboratory recirculation tank as used for test suspensions. However, that would have required very large volumes of formazin standards. Such large amounts of standard would have been very expensive, and disposal of the formazin, made with toxic reagents, would have been problematic. In any case, particles in formazin suspensions have little tendency to settle, so homogeneity was easily achieved with a laboratory magnetic stirrer. Instead we used relatively small volume (1 L) standards in which we measured the response of all the sensors separately, and sequentially, checking that we had at least 50 mm clearance from container surfaces to avoid reflections affecting output. In practice a laboratory retort stand was used to clamp all sensors in position, ensuring their optical faces were at least 50 mm clear of the base of the laboratory container (Figure 2-8).

Figure 2-8:   Calibration to formazin standards before measurements on natural suspensions. A WTW VisoTurb sensor is being calibrated, using the EXO2 sonde base as a beaker, and with magnetic stirring (blue stirrer unit). The WTW controller is the ivory-coloured box (with readout) to the right of the magnetic stirrer.

2.4   Experimental tank and sampling

The laboratory tank used in this experiment (Figure 2-9) was made of clear acrylic sheet-plastic supported by a steel frame, has a total volume of 170 litres (not including about 2 litres volume in the hosing). The 'testing' part of the tank has a uniform cross-section of 384 mm square X 1,000 mm height. Recirculation is achieved by a diaphragm pump, while a paddle wheel constantly agitates the tank contents.

We explicitly checked for stability of optical character and turbidity at each concentration step by:

- repeating deployment of the YSI EXO2 sonde after testing all the other field-type sensors, and

- testing samples taken at the beginning (initial samples) and again at the end (final samples) of each test concentration using the cuvette turbidity instruments.

To obtain a wide range of concentrations, we started with each stock suspension at high concentration and worked towards lower concentrations with three-fold dilutions at each step. Each three-fold dilution was achieved by removing two-thirds of the tank volume and replacing this volume with degassed tap water. The working volume of the tank in the current experiments was 163.3 L (at a marked level). The level corresponding to removal of two-thirds of the contents was calculated: (163.3 X 2/3 L)/(384 mm)^2 = 108,867 cm^3/1474.56 cm^2 or 73.8 cm drop in level. So, at each concentration decrement, two-thirds of the tank contents was discarded (after sampling for cuvette tests, and mass and beam-c concentration measurements) by running the suspension to waste until the level had dropped by 73.8 cm. The test tank was then refilled to the 163.3 litre level with Hamilton City tap water from a 1 m^3 storage tank. The tap water had been held in this tank to degas for several days prior – in an effort to prevent microbubble formation in the test suspension or on sensor surfaces.

The tap water also provided a thermal buffer during experimental runs. To further buffer sensors against temperature effects, all sensors were contained in a chilly bin of tap water between deployments in the test tank. The temperature sensor on the EXO2 sonde recorded a warming of the test tank due to pumping and agitation during a test run at each concentration – by about 1.5°C over the approximately 60 minutes required for sequential sensor deployments. In practice we matched test tank and storage temperatures within about 2°C, and thus expected to avoid temperature artefacts.

At each concentration of test suspensions, after 5 minutes recirculation and agitation to achieve full mixing, small samples (initial samples) of about 80 mL of tank contents were collected from the sampling port (tap near the two-thirds volume level) (Figure 2-9) for measurement of turbidity on cuvette instruments. Final samples at that concentration were also taken after all field turbidity sensors had been tested, including a re-run with the EXO2 sonde turbidity sensors. Final samples were taken from the sampling port in duplicate: (1) a 1 L bottle for laboratory measurement of total suspended solids (TSS) and volatile suspended solids (VSS) content (APHA methods), plus (2) a 2 L bottle for re-measurement of cuvette turbidity and for measurement of beam-c (optical concentration).

Additionally, samples of each of the three test suspensions were used for particle sizing as described in the next section.

2.5 Suspended particle mixtures

2.5.1 Approach and rationale

Our approach in choosing suspensions to test was to cover a range of materials and associated optical (light-scattering) characteristics commonly encountered in NZ's diverse rivers and lakes. The rationale here was to extend the comparison of turbidity sensors to more than one (type of) suspension as in the work of Rymsewicz et al. (2017) and in our own preliminary work (Hughes et al. 2019) in which only river stormflow sediment was tested. Accordingly, we sought different materials that are also encountered in routine use of both field (in situ) and cuvette-type turbidity sensors. Different types of suspended particle, such as equidimensional mineral particles versus plate-shaped minerals (clays, micas), and mineral particles versus organic particles (e.g., algal cells), are expected to have different scattering functions (angular dependence of scattering, Kirk 2011) and so vary in ratio of side to total light scattering. We expected that the relative turbidity sensor output might vary with changed light scattering conditions in suspensions of different materials.
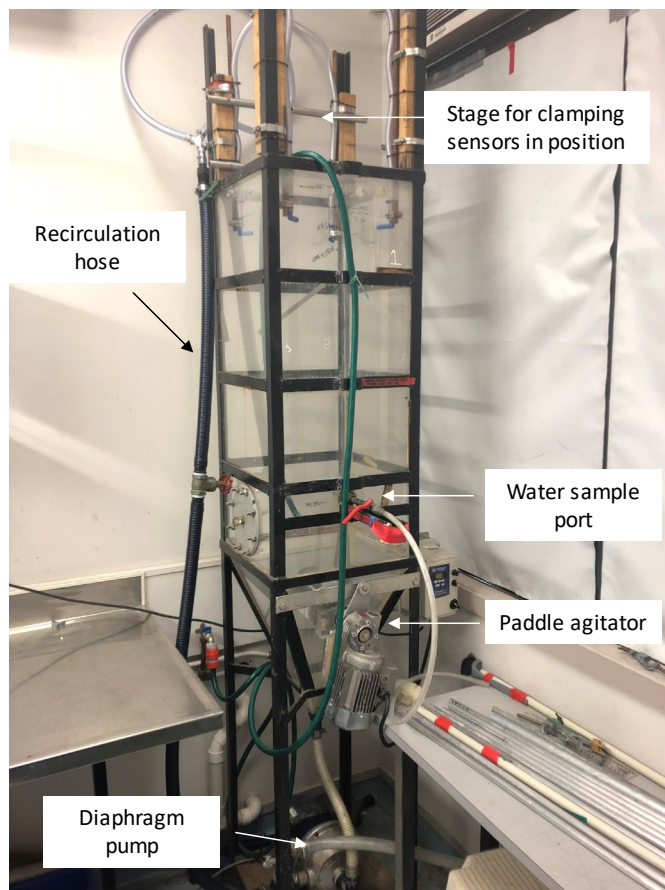
Figure 2-9:   **Laboratory recirculation tank used for turbidity sensor testing.** Tank volume is 170 L. Note the block-out blinds on the window to the right of the tank – to exclude (NIR radiation in) sunlight that might affect NIR nephelometers.

The primary application of many *in situ* turbidimeters in NZ is as a proxy for suspended sediment concentration in rivers, particularly over the flood events that typically dominate catchment loads of sediment and other particulate or particle-associated contaminants such as total phosphorus and faecal microbes (Davies-Colley and Hughes 2019). Therefore, we used sediment, mainly in the silt

size range, derived from river stormflows. This material was captured using a Phillips-type (time-integrating) sampler (Phillips et al. 2000) set at a high stage (a stormflow level expected, from the nearby water level recorder, to be overtopped about 12 times per year). Phillips-type samplers do not representatively capture relatively slow-settling layer clay minerals (Phillips et al. 2000) which are common in rivers in catchments with clay-rich soils in some parts of NZ. Therefore, we also tested our range of turbidity sensors on laboratory-grade kaolinite (a 1:1 alumina:silica layer clay) which is common in NZ soils and many rivers especially at baseflow. Turbidity sensors are commonly used in standing waters as well as rivers, and in eutrophic lakes phytoplankton (suspended algae – i.e., *organic* rather than mineral particles) often dominate light attenuation. Accordingly, we also tested our range of turbidity sensors on phytoplankton-laden water derived from a eutrophic pond.

The three different test suspensions are listed in Table 2-2, with details of preparation and characterisation given below.

Table 2-2: **Experimental suspensions.** Source and notes on preparation are given. Summary particle size data are for samples taken at the lowest concentration (end-of-run) after several hours' recirculation in the test tank at different concentrations.

Material	Source	Preparation	Sizes (μm) (cross-section-area weighted)			
			Mode	Percentiles		
				10	50	90
River silt	Mangaotama Stream, Whatawhata	Integrated stormflow sample from Phillips sampler, successively decanted (3X) to remove sand and coarser silt	5	1.2	4.8	11.6
Kaolinite clay	Laboratory grade kaolinite (dry powder)	100.8 g mass added to 1 L of tap water and stirred, wet-sieved at 63 μm to remove impurities, made up to 163.3 L	Multiple modes	0.93	3.4	9.3
Organic particles (green algae)	Eutrophic pond water	Wet-sieved at 63 μm to remove zooplankton and duckweed	6.5	3.3	6.7	13.9

2.5.2 Particle sizing of test suspensions

Particle size distributions (PSD) were determined for the test suspensions on a sample taken from the tank at the *highest* concentration (START sample, after mixing), and on a sample taken from the tank at the *lowest* concentration at the end of the dilution series (END sample). The rationale was to check whether tank mixing had caused any nett breakup of particles (or aggregation of primary particles), potentially changing optical character, and thus turbidity, during the test run. Particle sizes were measured *on individual particles* with an EyeTech® laser streaming instrument (time-of-transit principle[5]). Samples were run both before and after ultrasound treatment (which disperses aggregates) to give an indication of the degree of aggregation of particles in test suspensions.

---

[5] The time-of-transit (laser obscuration) principle is a completely different sizing principle from laser diffraction instruments – which attempt to best-fit observed angular distributions of scattering to assumed PSDs.

2.5.3 River silt

Stormflow suspended material was obtained from the Mangaotama Stream at the Whatawhata Research Station (WRS) near Hamilton. The catchment is dominated by steep to hilly topography, comprised of Mesozoic sedimentary sandstones and mudstones (greywacke and argillite) with strongly weathered yellow-brown earth soils. Patches of overlying volcanic ash remain in the more gently-sloping parts of the catchment. The catchment above the sampling site (PW5 hydrometric site; Hughes et al. 2020) is of mixed land use, including sheep-beef pasture, pine plantation, and native riparian plantings. Davies-Colley & Hughes (2020) showed that under low flow clear-water conditions at this and other WRS sites, suspensoids comprise mainly fine layer clays with very high optical cross-section per unit mass concentration (up to 3 $m^2 \, g^{-1}$), whereas silts with much lower optical cross-section (averaging about 0.15 $m^2 \, g^{-1}$) dominate stormflows.

Stormflow sediment was collected from the hydrometric station on the Mangaotama Stream (Site PW5). The Phillips sampler (100 mm diameter cylindrical body) was fixed in the channel, cylindrical axis horizontal, and with its entry orifice (4 mm diameter) facing upstream into the water current. The sampler orifice was positioned 15 cm above the baseflow level, based on the long-term continuous level record for this small stream, so as to intercept only larger flood events (occurring, on average, about 12 times per year). The Phillips sampler is essentially an isokinetic sampler that efficiently samples and then traps suspended particles in the silt and sand size ranges, while clay-sized materials, particularly plate-shaped particles, have a settling velocity too low to be trapped efficiently and mostly exit the sampler.

The muddy sand sediment collected in the Phillips sampler was subjected to successive decantation in three stages to remove most material coarser than about 30 μm (i.e., sand (> 63 μm) and coarse silt (32-63 μm)). These coarse size fractions have relatively minor effects on light scattering, and thus turbidity, because their specific surfaces are low compared to smaller particles (Davies-Colley and Smith 2001). By removing coarse silt as well as sand we expected to reduce maximum settling velocity (proportional to diameter squared according to Stokes Law) sufficiently to avoid any experimental artefacts with fast-settling particles in tank tests.

Figure 2-10 shows the PSD of the resulting stormflow silt at the START and END of the tank testing run. The cross-sectional area-weighted PSDs are given being most relevant to optical effects. The PSDs have strong modes at around 5 μm (fine silt), a significant fine 'tail' of clay-sized material and a 'shoulder' of coarser material (medium silt). There was little change in PSD between the first test-tank sample (high concentration at START of dilution series – red curve) and the last test-tank sample (at the END of the dilution series – blue curve) during the 'exposure' of the material to several hours of tank pumping and agitation – confirming that there was little or no tendency for breakup of aggregates during the experiment. Ultrasound exposure had minimal effect on the PSDs (data not shown), suggesting that most particles in the experimental mix were not aggregates.
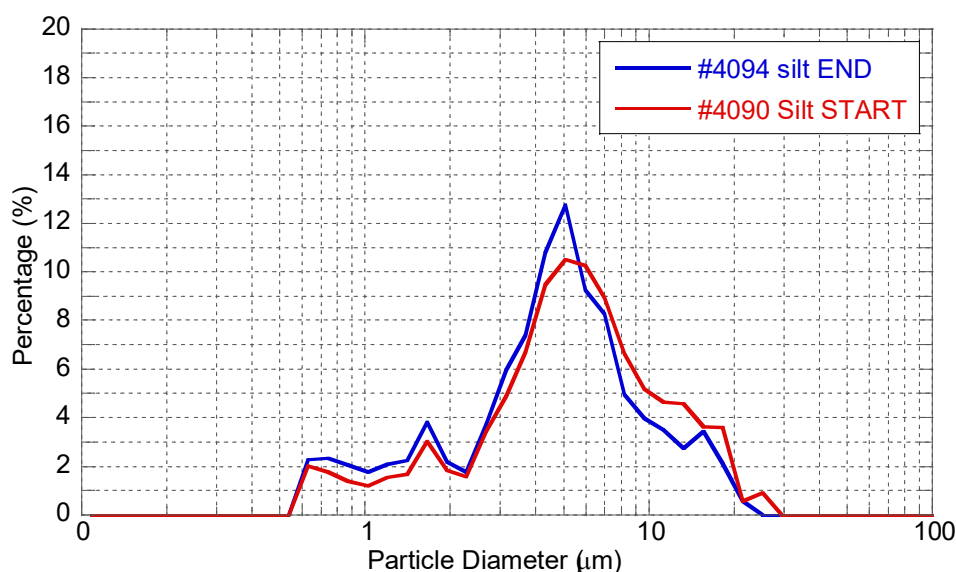
Figure 2-10: River stormflow silt particle size distribution. The red curve is for the high concentration stock river silt at the START of the experiment, and the blue curve is for the low concentration sample at the END of the test tank runs. Both curves are cross-section area-weighted PSDs for EyeTech® particle sizing before ultrasound treatment.

2.5.4 Kaolinite clay

Kaolinite is a layer clay mineral (1:1 alumino-silicate) that is a common constituent of NZ soils and thus commonly present in rivers. Fine, plate-shaped kaolinite particles settle very slowly and scatter light intensely, particularly near 90°, because of their high cross-sectional area to volume ratio (Gibbs 1978). A weighed dry mass (100.8 g) of commercial laboratory-grade kaolinite was added to degassed tap water in a 10 L bucket and the resulting suspension sieved through a 63 μm sieve to remove any relatively rapidly-settling aggregates or sand-sized contaminants[6]. The resulting sieve-passing kaolinite suspension, contained in a 10 L bucket, was then poured into the partly filled test tank (Figure 2-11) and made up to volume (163.3 L) with degassed tap water. Expected mass concentration was therefore known: 100.8 g/163.3 L = 617 mg/L and agreed closely (well within typical experimental error) with measured TSS of 611 mg/L (611 g m^{-3}).

Figure 2-12 shows the PSDs for START (high concentration – red curve) and END (low concentration – blue curve) in samples of the kaolinite suspension taken during the tank tests. The PSD is quite broad, extending across much of the coarse clay and fine-medium silt range, with multiple modes (major mode at 4-5 μm). A small change in PSD during the tank tests, with loss of some (~ 10% by cross-sectional area) initial material around 30-40 μm and small compensatory gains distributed across the medium silt to clay range, likely indicates some dis-aggregation. Ultrasound exposure produced a large shift (fining) of the kaolinite PSDs (Figure 2-12), quite unlike the river silt which lacked sensitivity to ultrasound. Apparently, the kaolinite suspension was appreciably aggregated, and these aggregates were mostly stable (or at least in dynamic equilibrium) to agitation and pumping in the test tank, despite being susceptible to disruption by ultrasound.

---

[6] A very small amount of gritty contaminant was retained by the sieve, estimated at << 1%.

Figure 2-11:   Adding kaolinite to the test tank. The tank was partly filled with degassed tap water, and after addition of the kaolinite slurry the tank volume was topped up to the operating level (red line, 163.3 L).
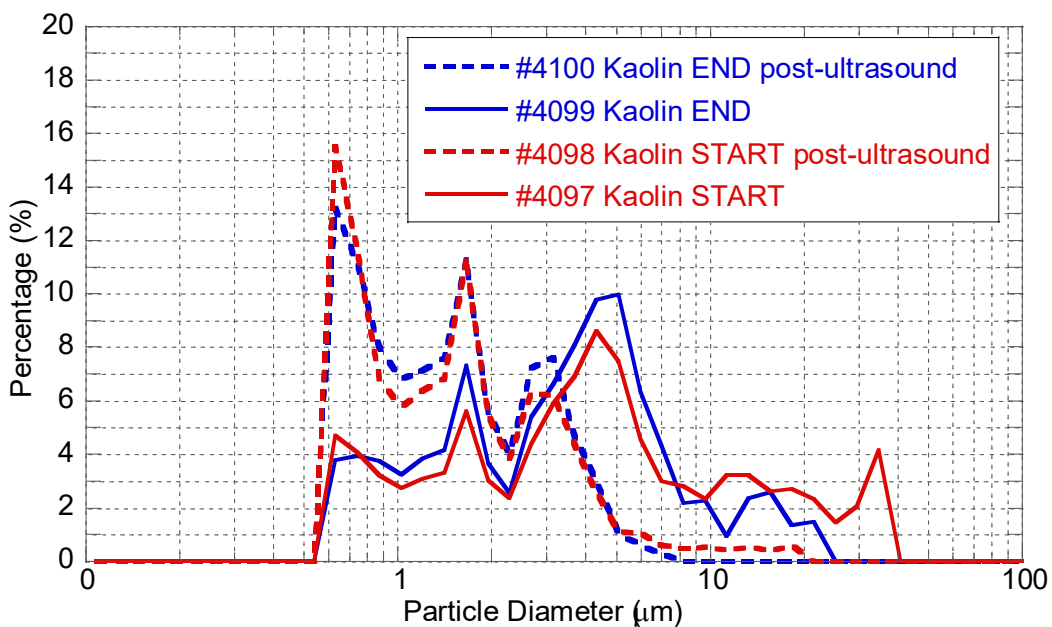
Figure 2-12:   Kaolinite particle size distribution. Continuous curves are PSDs from EyeTech® particle sizing before ultrasound treatment, and the dashed curves are for the corresponding samples after ultrasound treatment. All curves are cross-sectional area-weighted PSDs. The red curves are for the high concentration stock kaolinite suspension at the START of the experiment, and the blue curves are for the low concentration sample at the END of the test tank runs.

2.5.5   Phytoplankton

We collected about 180 L of water, fairly heavily-laden with phytoplankton, from a eutrophic pond near the Cambridge Wastewater Treatment Plant, by pumping with a small bilge pump into two 50 L barrels. Back at the laboratory, this phytoplankton slurry was delivered to the test tank with the same bilge pump. A 63 μm sieve, topped by a 125 μm pre-sieve, was used to remove larger particles, including zooplankton and duckweed, before the sieve-passing water actually entered the test tank. Microscopic examination by a NIWA technician expert in algal screening, revealed a diverse assemblage of phytoplankton in the pond water. Green algae, including some colonial types, dominated (particularly *Acutodesmus*, *Desmodesmus*, and *Scenedesmus* spp.), but some cyanobacteria were seen (especially *Aphanizomenon gracile*) and a range of other types (e.g., dinoflagellates, euglenoids, Cryptophyceae) were present at relatively low abundance.

Figure 2-13 shows PSDs for START (high concentration – red curve) and END (low concentration – blue curve) of the tank tests with pond water. A fairly simple unimodal distribution (mode at about 6-8 μm) probably represents mainly green algae (which typically have cells about this size). A fine tail may be cyanobacteria and/or clay and the coarse 'shoulder' probably represents colonial green algae observed microscopically. There was a slight shift in PSD through the tank runs, suggesting possible break-up of some colonies of green algae (a minor component), but the PSDs were insensitive to ultrasound (data not shown).
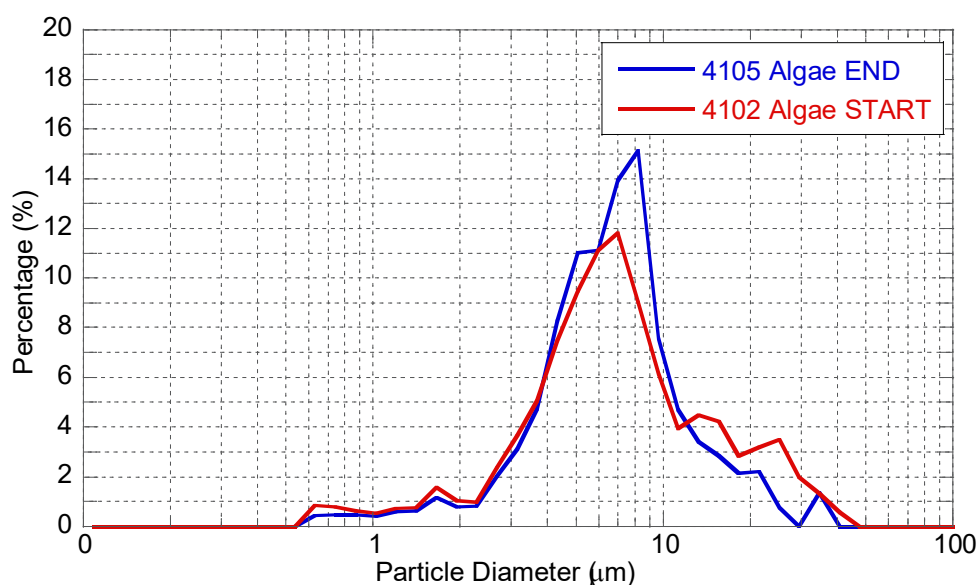
Figure 2-13:   Phytoplankton particle size distribution. The red curve is for the high concentration stock pond water at the START of the experiment, and the blue curve is for the low concentration sample at the END of the test tank runs. Both curves are cross-sectional area-weighted PSDs for EyeTech® particle sizing before ultrasound treatment.

In order to compare the PSDs of the three test suspensions, in Figure 2-14 we plot cumulative particle size distributions at both the START (Panel A) and END (Panel B) of each test run for each of the three suspended materials, and both before and after ultrasound dispersion (us). The river stormflow silt (brown curves) was intermediate in size, shifted little in PSD through the 81-fold dilution series, and was only slightly aggregated as inferred from ultrasound treatment. The phytoplanktonic algae (green curves) were coarsest in size and shifted somewhat (more) through the 9-fold dilution series, apparently due to break-up of coarse material that, as mentioned earlier, may

have been colonies observed microscopically. In contrast, the kaolinite was comparatively fine and seemed to be appreciably aggregated. A coarse shoulder at around 30 μm was progressively broken up during the 81-fold dilution series, and ultrasound produced a major shift (fining) of the PSDs (e.g. START median shifting from 4 to 1.3 μm).
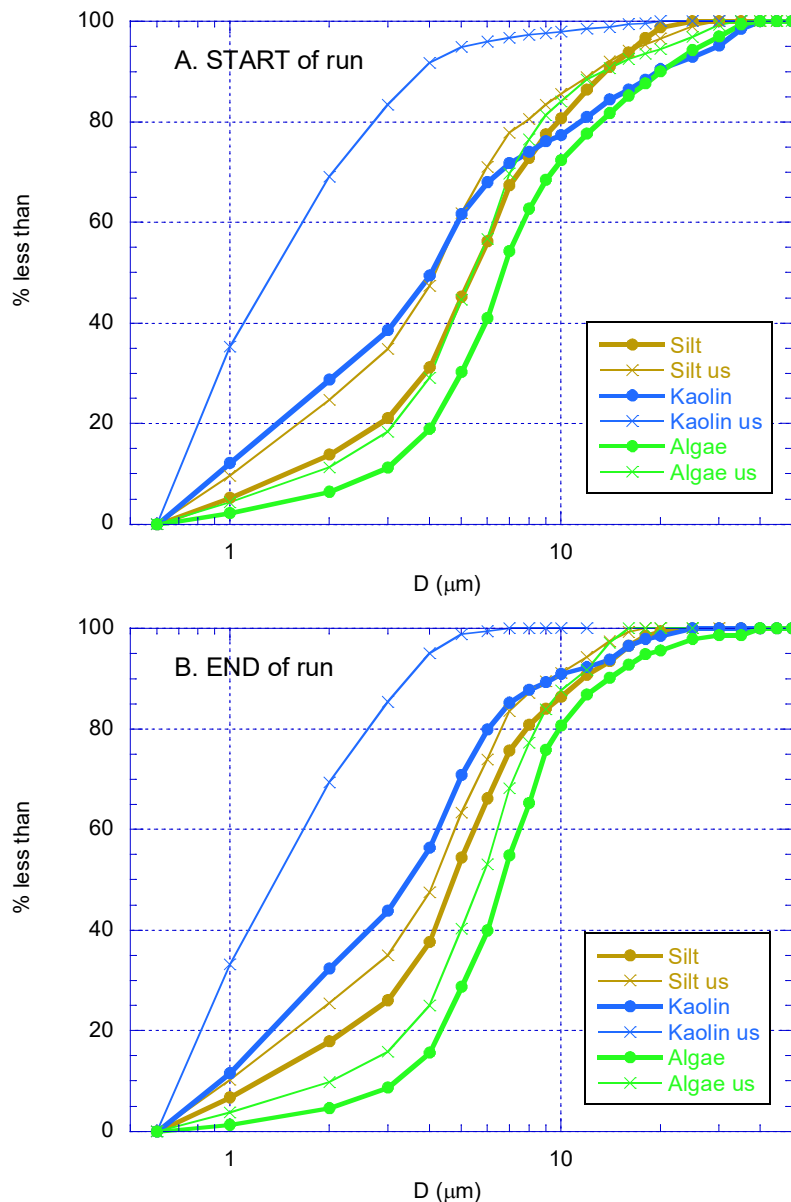
Figure 2-14: Cumulative particle size distributions for three different suspended materials. Panel A: PSDs at the START of the dilution series; Panel B. PSDs at the END of the dilution series. In each case the PSD as measured untreated (solid dots and lines) can be compared with after ultrasound treatment (X's and thin lines).

2.5.6 Concentrations of suspensions

Water samples were taken at the start (initial samples) and end (final samples) from the tank at each concentration for turbidity measurements on the two cuvette instruments. Additionally, the final samples (2 L and 1 L volume) were used for measurement of turbidity with the third WTW VisoTurb sensor (because only two sensors could be connected to the two ports on the single WTW controller that we had available). These final samples were used to measure concentrations of suspended materials – both mass concentration and optical concentration.

Mass concentration of suspensoids (and their organic content) at each concentration decrement was measured by standard filtration through GF/C filters followed by ignition (TSS/VSS measurement APHA 2540 D). However, we used a lower ignition temperature (400°C for 6 hrs) than standard (550°C for 2 hrs) to avoid spurious weight loss due to clay de-hydroxylation (Grove and Bilotta 2014).

The beam attenuation coefficient was measured at 530 nm to quantify optical concentration. The WETLabs C-Star beam transmissometer, operated in flow-through mode as described in sub-section 2.2.7, was used for the measurements referred to henceforth as beam-c.

Beam-c in the dilution water (degassed Hamilton City tap water) was very low at 0.17 m^{-1} (referenced to pure water). Using a semi-empirical equation derived by Zanevald and Pegau (2003) that relates visibility to beam attenuation at 530 nm (accounting for attenuation by water itself), the corresponding visual clarity would be approximately 20 m.

Data on suspensoid character is collated in Table 2-3, and can be summarised as follows:

- The *river silt* had a starting mass concentration of 1,060 g m^{-3}. This material was 5.8% (by mass) volatile – which is fairly typical for river stormflows.

- The *kaolinite* had a starting measured mass concentration of 611 g m^{-3} (comparing well with the concentration of 617 g m^{-3} calculated from the measured dry weight added to a known volume of water) and was only 1.3% volatile.

- The *pondwater* solids had a starting mass concentration of 59 g m^{-3} and was 79% volatile, with an associated chlorophyll-*a* content of 0.51 g m^{-3} – consistent with mostly algal biomass.

Table 2-3: **Laboratory concentration measurements on test suspensions.** Measurements by the NIWA-Hamilton Water Quality Laboratory include turbidity by TL2310 bench instrument; TSS and VSS by standard gravimetry on GF/C filters (APHA 2540D), and chlorophyll-a (Chla) by spectrophotometry on acetone extracts of GF/C filters (APHA 10200H). Beam-c was measured as described in Section 2.3.7. (Only the starting mass concentrations were measured for kaolinite; other concentrations (in parentheses) were calculated from the dilution factor.) Note that the turbidities were as measured by laboratory technicians separately from our measurements (on the same instrument: TL2310 LED) as part of the comparison. The laboratory turbidities correlated closely with ours but were somewhat higher.

Suspensoid sample	Turbidity (FNU)	TSS (g m^{-3})	VSS (g m^{-3})	Chl*a* (mg m^{-3})	Beam-c (m^{-1})
River silt					
Silt-starting suspension	984	1,060	62.0		553
Silt -1/3 dilution	296	350	23.8		183
Silt -1/9 dilution	93	116	6.0		59.4
Silt -1/27 dilution	27.8	38.5	<0.5		19.3
Silt -1/81 dilution	11.1	13.2	<0.5		6.57
Kaolinite					
Kaolinite starting suspension	556	611	7.8		277
Kaolinite -1/3 dilution		(204)			98.5
Kaolinite -1/9 dilution		(68)			32.7
Kaolinite -1/27 dilution		(22.6)			11.2
Kaolinite -1/81 dilution		(7.5)			3.91
Pond water					
Pond water - starting suspension	31.9	58.8	46.7	508	47.1
Pond algae -1/3 dilution	10.7	18.7	14.6	158	15.0
Pond algae -1/9 dilution	3.6	6.8	5.0	39.6	4.6

2.6   Laboratory experiments

The field-type turbidimeter sensors were deployed in the test tank at the level of the sampling port (Figure 2-9). In order to position the sensors and orient them appropriately they were all taped (using fluoro-pink duct tape) to alloy rods that were, in turn, clamped to a horizontal rod above the tank using retort clamps. Between deployments the sensors were stored in a chilly bin of tap water to maintain temperature close to that in the test tank (Figure 2-15). Temperatures of test tank contents and the chilly bin were monitored using the temperature sensor on the EXO2 sonde. This monitoring showed that that temperature of sensors and test tank suspensions were consistent within about 2°C throughout the experiments.

Figure 2-15:   Storage of field turbidimeters in a chilly bin of tap water between test-tank deployments.
Keeping sensors close to tank temperature was expected to avoid artefacts due to temperature changes. The sensors were individually taped (pink duct tape) to alloy rods – which allowed sensors to be positioned at the correct depth (opposite the sampling port) and orientation within the test tank. Retort clamps were used to fix the rods to a stage above the tank.

Field-type (*in situ*) sensors were positioned sequentially in the test tank (in triplicate) at the depth of the sampling port (Figure 2-16). Before any data were collected, the brush wipers of the EXO2 sonde, Hach Solitax and Observator sensors were operated to remove bubbles that could cause measurement artefacts. With the WTW sensors we could not tell visually when their ultrasonic cleaning system was operating, and, in any case, fouling was not expected over the time scale (few hours) of our experiments, so the ultrasonic cleaning was not operated. Instead, the WTW sensors were simply lowered just below the water surface and their optical faces brushed manually with a soft brush to remove any bubbles (and to be more nearly consistent with other sensors incorporating brush units) before being lowered to the sampling port level and clamped in position.
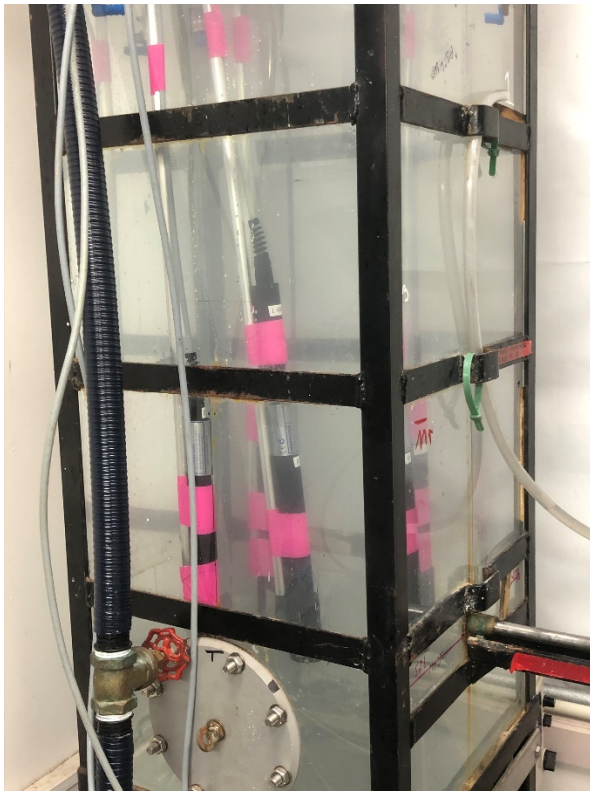
Figure 2-16: Positioning sensors inside the test tank. Observator sensors are shown (in triplicate) positioned (using retort clamps) with sensing heads arranged to use all the available tank space at the sampling port level. The photograph was taken with kaolinite at the lowest measured concentration in the tank. (There are only three sensors in the tank – as seen through the window on the left: the image of sensors through the wall on the right is a refraction effect).

Warmup time for the sensors was not a concern because all sensors were kept 'on' constantly throughout each day of the experimental campaign. Data were read off the controller units over at least 2 minutes (as well as being logged where this was feasible) before replacing with the next make-model of sensor. Sensors were deployed in the same order at each concentration increment, namely: (1) EXO2 sonde, (2) Hach Solitax, (3) WTW VisoTurb, (4) Observator Analite. The EXO2 sonde measurements were repeated at the end of the series to check for optical changes in the suspension at that dilution step and to monitor temperature change in the test tank compared to that in the storage tank.

At each concentration of the test suspensions initial samples were obtained, after mixing but before sensor deployment, and final samples after all sensor deployments. These samples were stored dark in a refrigerator until measurements (on the following day) on the two cuvette-type turbidity instruments (Hach TLD2310LED and Hach 2100Q-is). Prior to measurement, samples were re-warmed to room temperature in a bath of running tap water, to avoid condensation on cuvette surfaces, which confounds nephelometric measurements. Measurements were also made of beam attenuation on the WETlabs C-Star beam transmissometer operated in flow-through mode, which provided both an absolute reference and a check on dilutions of stock suspensions (sub-section 2.2.7).

2.7 Data collation and analysis

Turbidity sensor response data recorded by loggers were obtained where feasible, otherwise manual records of response (seven spot readings taken over at least 2 minutes) were used. We verified agreement of manually recorded and logged data where both were available, and on that basis felt justified in using manual data throughout for consistency. Sensor response data were reduced to means and standard deviations for subsequent analysis and plotting.

Because of the ratio progression of both formazin standards and concentrations of natural test suspensions, we originally envisaged plotting all data on logarithmic (ratio) scales. Logarithmic scales are preferred to achieve symmetry of data distributions as well as to cope with variance increasing with magnitude, and non-linearity. However, as shown below, the sensor responses were all so nearly linear as to remove one major reason for log-transforming. More importantly, the diversity of sensor response turned out to be much better displayed on linear scales, being difficult to assess on logarithmic scales. For these reasons we used linear scales and supplemented plots of all data (all five concentrations) with 'zoom' plots to better resolve the lower three data-points.

We first plotted responses of sensors versus concentrations of the freshly-made formazin standards. The resulting curves provided local calibrations for all sensors for comparison on the same basis. Potential drift in the response of the sensors over the four days of the experiment was checked by comparing calibration measurements repeated after the test tank experiments (on 10 June) with those measured (on 6 June) before the test tank experiments commenced (running from 7-9 June inclusive). Stationarity of the calibration formazin stocks between 6 and 10 June was checked by looking for any consistent response change across all sensors.

Turbidity outputs reported in calibrated FNU were then plotted against optical concentration (beam-c) to permit a direct comparison of different sensors – both different units of the same make-model (in triplicate) and different make-models. We used optical concentration (beam-c) as the reference rather than mass concentration (TSS) because of the better precision of the optical metric, although as we show in the results, beam-c and TSS are very closely linearly correlated, so the findings are not sensitive to choice of reference.

MS EXCEL® was used for data collation and calculation of statistics. The exploratory data analysis (EDS) package DataDesk® (Velleman 2012) was used initially for graphical data exploration, and MatLab® , Kaleidograph® and EXCEL® for curve fitting and publication-quality graphics.

3 Results

3.1 Optical concentrations (and stability) of test suspensions

The three different types of test suspension – river silt, kaolinite, and algae (in pond water), were diluted three-fold at each concentration step. This procedure was validated by examining correlations between (the inverse of) dilution factor, measured mass concentrations (TSS, VSS, and Chl$a$ for the algal-laden pondwater) and measured optical concentration (beam-c) given in Table 2-3. The Pearson's moment correlations were all very high for log-transformed variables, exceeding 0.999 for dilution factor vs TSS vs beam-c (data not shown). Even for the volatile suspended solids concentration and the chlorophyll-a pigment (in pond water algae), correlations exceeded 0.995.

Figure 3-1 shows the relationship between optical and mass concentration for the three test suspensions. The power law fits to the data had exponents very close to unity, i.e., relationship was very near-linear. Interestingly, the two mineral suspensions plotted very similarly, but the algae plotted appreciably higher. This shows that the two mineral suspensions had rather similar optical cross-sections (ratio of beam-c to TSS) of 0.52 m^2 g^{-1} for silt and 0.45 m^2 g^{-1} for kaolinite. The higher optical cross-section of algae (0.80 m^2 g^{-1}) probably reflects the lower density of algae which more than compensates, as regards total light attenuation (absorption plus scattering of light), for their lower refractive index.

We used beam-c as the reference concentration for its high precision (about 3% relative standard deviation of repeated measurements), better than that typically attained for TSS (about 10%). However, the correlation of beam-c and mass concentration was so close and near-linear (Figure 3-1) (Table 3-1) that the results presented below are not sensitive to this choice.

All three of our test suspensions (river silt, kaolinite and pond algae) appeared to be fairly optically stable for the duration of the tank tests. Evidence for this includes:

- Very high linear correlation (and good numerical agreement) of concentrations calculated from dilution factor versus mass and optical concentrations measured directly.

- Minimal change in particle size distributions at end versus start of dilution series with test suspensions (e.g., Figure 2-10), with the partial exception of kaolinite

However, closer inspection of Figure 3-1 reveals that the line for kaolinite is slightly skewed relative to that for river silt, and the power exponent is correspondingly lower (0.97 compared to 1.01) and significantly lower than unity (Table 3-1), consistent with slight non-linearity. A possible explanation is the shift in PSD (progressive fining) for kaolinite from start to end of the 81-fold dilution series, as shown in Figure 2-12. Such fining would be expected to increase optical cross-section slightly with progressive dilution of the kaolinite.
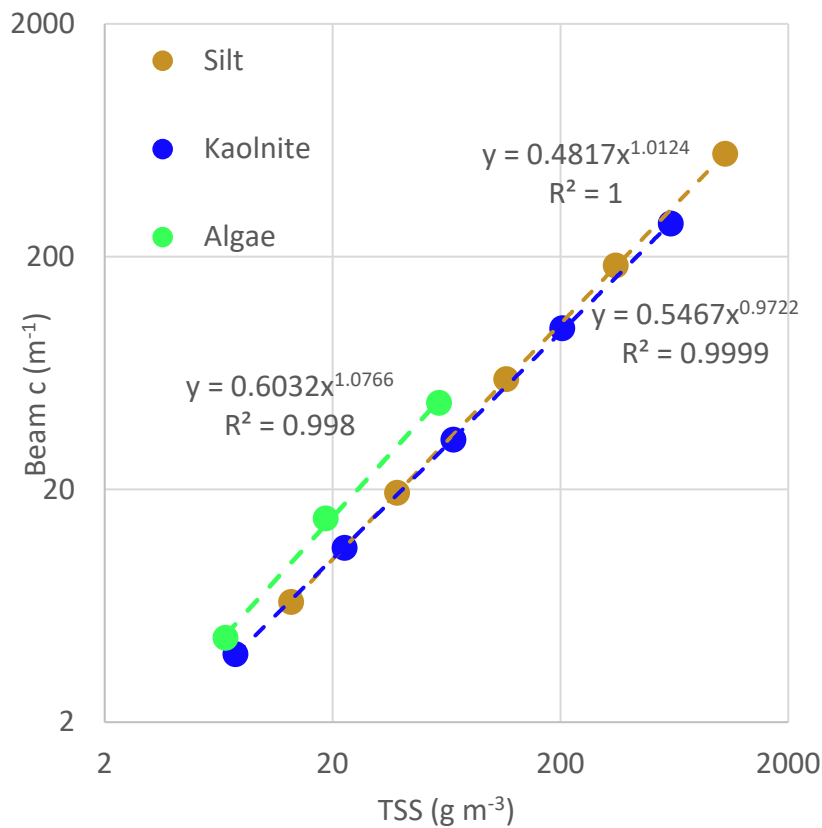
Figure 3-1: **Optical concentration versus mass concentration of test suspensions.** Optical concentration (beam attenuation coefficient) was measured on final samples at each dilution step by beam transmissometer. Mass concentration was measured on final samples of river silt and pondwater (algae) at each concentration, and as calculated from dilution factor and starting concentration of kaolinite.

Table 3-1: **Statistics of fitting lines for optical concentration versus mass concentration.** The p-values are for the t-test of slope of lines in Figure 3-1 versus unity.

	Slope (exponent)	Confidence interval	t-test p-value
River silt	1.0124	1.0056-1.0192	<0.001
Kaolinite	0.9722	0.9558-0.9886	<0.001
Algae	1.0766	0.458-1.696	0.029

3.2 Formazin calibration of sensors and sensor drift

Sensor response to the standard suspensions of formazin are shown in Appendix A, where readings on the standards both *before* and *after* the tank tests are given. The changes in response, before versus after, were usually of small magnitude, and unlikely to be of practical significance, although t-tests showed that some of the sensors (two of the Hach Solitax and one of the WTW sensors) had suffered a statistically significant change in response line (slope). The change in response of the Observators is probably not meaningful because initial data for the high standard (1,000 FNU) are missing due to over-ranging – prior to re-adjustment of the A/D conversion (counts per volt) on the Ecologger.

Response of all tested sensors was highly linear over the full range. This is good evidence that the formazin standards were stable over the four days of (dark, refrigerated) storage – because, had there been any drift in standards, this is most likely to have affected the low-end concentrations (say 10 and 25 FNU) (Kitchener et al. 2017).

Linear equations for the lines of best fit of formazin standard concentration (*F*) versus sensor response (*R*) were used as calibrations applied to sensor response in test suspensions:

$$F = aR + b.$$

where a and b are line-fitting parameters. (Refer Appendix A.)

3.3 Sensor response in test suspensions

Appendix B tabulates the response of turbidity sensors in test suspensions at different concentrations (Table B-1). Also given in Appendix B (Table B-2) are the statistics of response across the same make-model, with the coefficient of variation (CV) used to quantify variance. Turbidity sensor response is plotted on linear scales which serve better to illustrate the diversity of numerical response of different make-models of turbidity sensor to the same suspension.

3.3.1 Agreement within a make-model of turbidity sensor

Table 3-2 collects the equations of the lines-of-best-fit for triplicates of the same make-model of field-type sensor and shows that the response curves were very similar in slope although less so as regards (typically small) intercepts. Table 3-3 collates the p-values from the t-tests that tested for differences in the regression slopes between the triplicates of each field turbidity sensor make-model. Despite the small differences in magnitude of the regression slopes, there were four pairings (involving two sensors EXO6, and Observator7) where these differences were statistically significant. Table 3-4 shows the agreement between triplicates of the same-make-model, measured as coefficients of variation (CV), for the higher concentrations studied. The agreement is appreciably weaker in a relative sense at lower concentrations.

Agreement between triplicate units of the same make-model of field-type (*in situ*) sensor was generally good (within 10%) – at least at relatively high concentrations (Table 3-4). We would expect good agreement within a make-model if manufacturing tolerances for sensor components are much more precise than demanded by the ISO 7027 standard.

The CVs provide a measure of the reproducibility within a make-model. Fairly good reproducibility implies that a sensor is 'inter-changeable' – in the sense that, if a sensor has to be replaced, using the same make-model should avoid a large step-change in the turbidity record.

Table 3-2: **Linear equations quantifying average response of field-type sensors in three different suspensions.** Slope, Y-intercept and R2 are given for each of four field sensors (averages across a triplicate) in each of three different suspensions.

Sensor	Silt (n = 5)			Kaolinite (n = 5)			Algae (n = 3)		
	Slope	Y intercept	R²	Slope	Y intercept	R²	Slope	Y intercept	R²
EXO1	0.93	1.36	1.00	1.28	-3.93	1.00	0.39	0.16	1.00
EXO3	0.91	-0.33	1.00	1.20	-5.19	1.00	0.43	-0.77	1.00
EXO6	0.95	-4.51	1.00	1.27	-6.79	1.00	0.43	-4.73	1.00
Hach1	1.66	-8.24	1.00	1.65	-0.78	1.00	0.55	9.32	1.00
Hach2	1.75	-14.36	1.00	1.76	-12.54	1.00	0.59	-3.64	1.00
Hach3	1.75	-14.57	1.00	1.72	-8.62	1.00	0.55	1.83	1.00
WTW4	1.83	-24.39	1.00	1.75	-9.92	1.00	0.51	3.12	1.00
WTW5	1.83	-19.95	1.00	1.84	-5.43	1.00	0.55	2.75	1.00
WTW6	1.92	-31.97	0.99	1.90	-18.28	0.99	0.58	-2.55	1.00
Observator7	1.93	-3.39	1.00	2.06	-5.74	1.00	0.78	1.18	1.00
Observator8	1.83	-7.72	1.00	2.10	-1.37	1.00	0.81	-0.10	1.00
Observator9	1.85	-1.56	1.00	2.41	-7.22	1.00	0.83	1.52	1.00

Table 3-3: **P-values from paired sample t-tests of regression slopes between the different units of each turbidity sensor model (α = 0.05).** Bold values indicate statistically significant difference between the regression slope pair tested.

Sensor pairs tested	p-value
EXO1 vs EXO3	0.3551
EXO1 vs EXO6	**0.0363**
EXO3 vs EXO6	**0.0039**
Hach1 vs Hach2	0.1348
Hach1 vs Hach3	0.1745
Hach2 vs Hach3	0.8776
WTW4 vs WTW5	0.9447
WTW4 vs WTW6	0.4276
WTW5 vs WTW6	0.4160
Observator7 vs Observator8	**0.0338**
Observator7 vs Observator9	**0.0015**
Observator8 vs Observator9	0.6370

Table 3-4: **Variation in response of field-type turbidity sensors at higher concentrations.** Coefficients of variation (CV) of the triplicates of four different field-type sensors characterise the within-sensor variation (reported as SD for standard deviation).

Sample name	Beam-c (m^{-1})	EXO Mean (FNU)	SD	CV (%)	Hach Mean (FNU)	SD	CV (%)	WTW Mean (FNU)	SD	CV (%)	Observator Mean (FNU)	SD	CV (%)
Silt	183	174.1	3.4	2.0	282.6	16.9	6.0	268.8	9.0	3.3	326.5	20.0	6.1
	553	512.8	9.7	1.9	945.0	22.3	2.4	1,018.7	25.4	2.5	1,032.3	29.6	2.9
Kaolinite	98.5	111.5	7.0	6.3	144.0	2.7	1.9	146.7	11.5	7.8	204.6	18.8	9.2
	277	342.2	11.8	3.4	472.4	10.1	2.1	503.3	19.7	3.9	604.0	52.2	8.6

3.3.2 Response of different make-models of turbidity sensor in test suspensions

Figure 3-2 shows the formazin-calibrated sensor response (averaged for each triplicate of make-model) versus beam-c in river silt, Figure 3-3 shows the same sensors' averaged response in kaolinite, and Figure 3-4 shows their averaged response in pond water (phytoplanktonic algae). All of the turbidity sensor response data are well-fitted by simple linear equations, albeit strongly weighted to the higher concentrations. The 'zoom' plots show that there is some 'cross-over' of lines at lower concentrations (~10 FNU or lower), which derives from fitting formazin calibration lines on linear scales (which intrinsically weights data at higher turbidities more strongly).

For each suspension there is appreciable variation in response (more than 2-fold) between different make-models of turbidity sensor. This divergence in response is unexpected for sensors all designed according to an international standard – a standard that is intended to achieve comparability in numerical output.

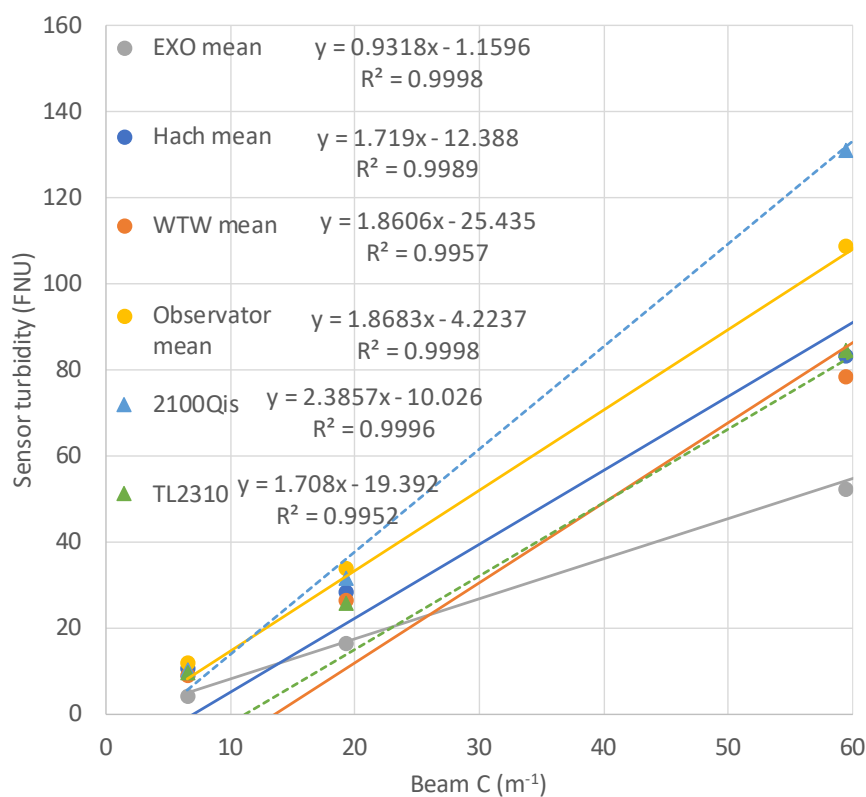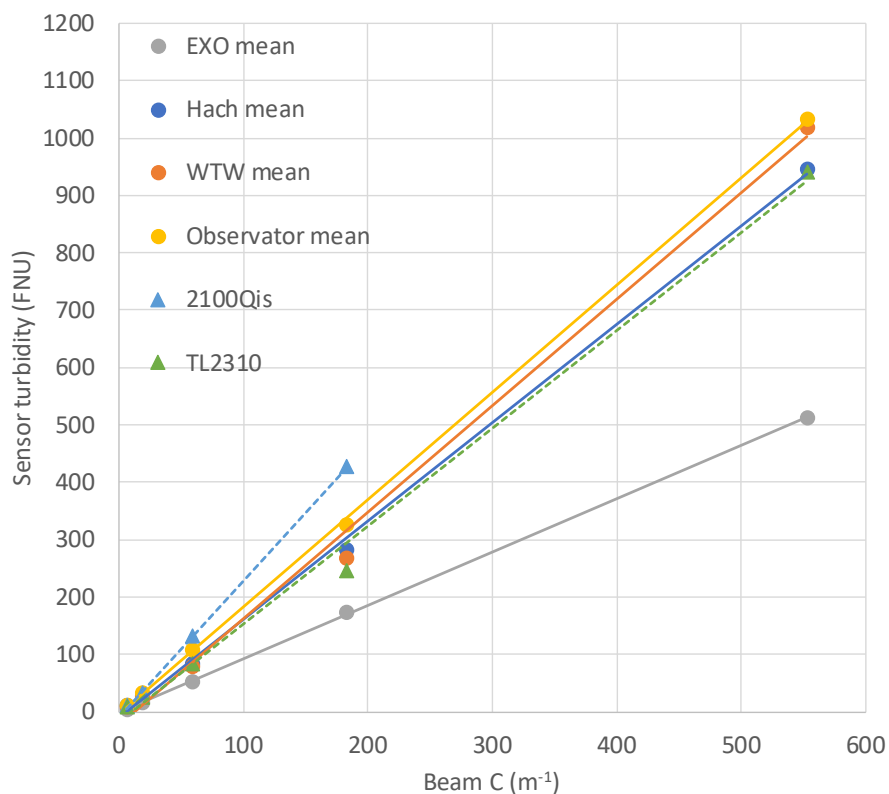**Figure 3-2:** **Turbidity sensor response in river stormflow silt.** Upper panel: all data, Lower panel: 'zoom' on lower three data-points for each sensor. Data for field-type sensors are shown as solid points fitted with continuous lines; cuvette sensor data are shown as triangles fitted with dashed lines. Response (corrected to formazin) is averaged across each make-model triplicate of field sensors.
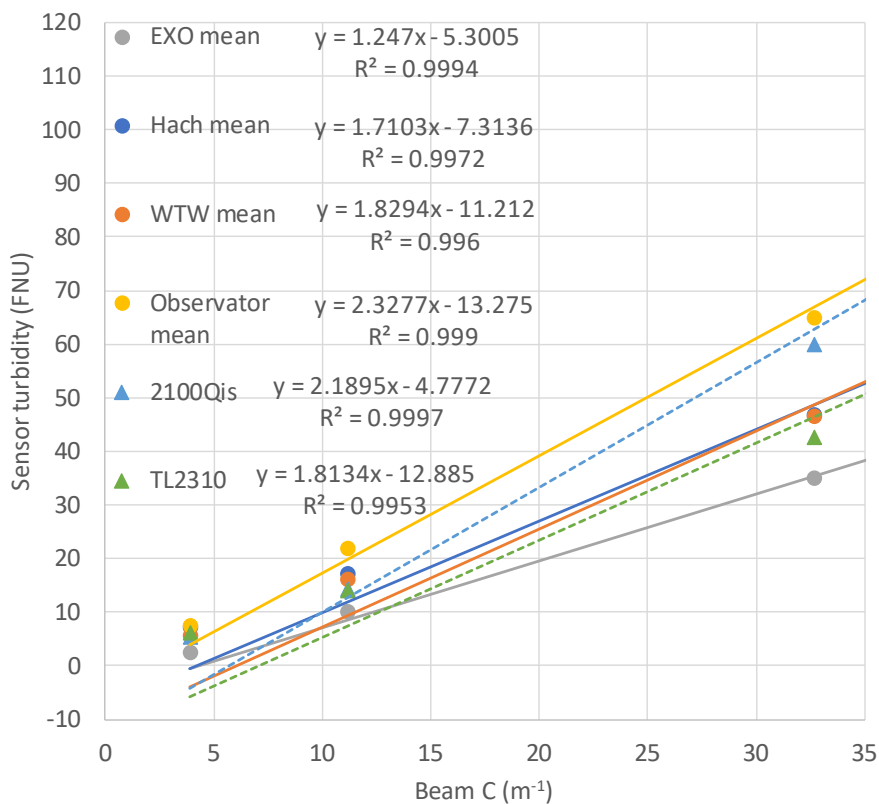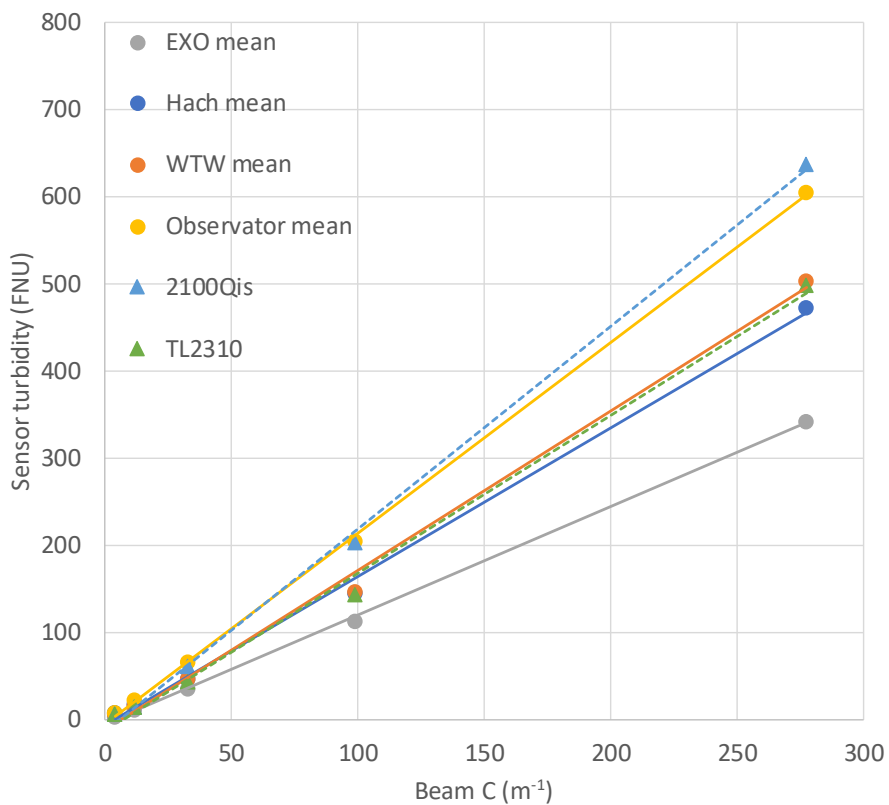
Figure 3-3:   Turbidity sensor response in kaolinite. Upper panel: all data, Lower panel: 'zoom' on lower three data-points for each sensor. Data for field-type sensors are shown as solid points fitted with continuous lines; cuvette sensor data are shown as triangles fitted with dashed lines. Response (corrected to formazin) is averaged across each make-model triplicate of field sensors.

Figure 3-4: Turbidity sensor response in pondwater. Data for field-type (in situ) sensors are shown as solid points fitted with continuous lines; cuvette sensor data are shown as triangles fitted with dashed lines. Response (corrected to formazin) is avera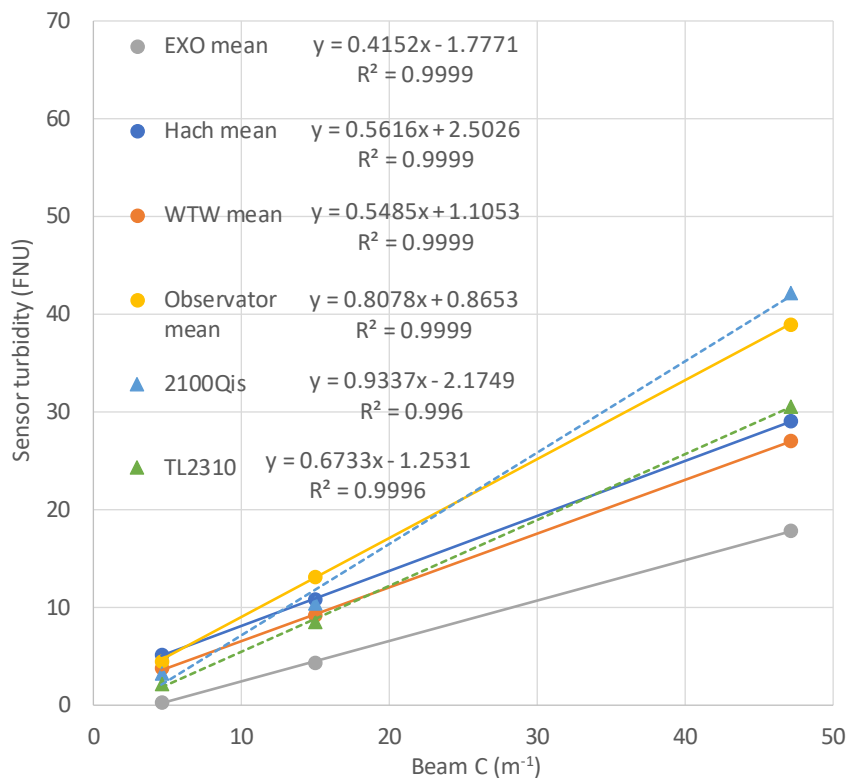ged across each make-model triplicate of field sensors. (Colour coding of different make-models is as for Figures 3-2 and 3-3).

The relative numerical response of the different sensors is fairly similar for the three test suspensions, with the EXO sensors consistently low and the Hach 2100Q-is consistently high, but with some variation in relative position of the other sensors between test suspensions. The order of the response for the three suspensions is as follows:

Stormflow silt: Hach 21OOQ-is > Observator > WTW > Hach Solitax > Hach TL2310LED > EXO

Kaolinite: Hach 21OOQ-is > Observator > WTW > Hach TL2310LED > Hach Solitax > EXO

Pond water: Hach 21OOQ-is > Observator > Hach TL2310LED > Hach Solitax > WTW > EXO.

Results for t-tests for differences in the regression slopes of the different sensors showed that the regression slopes are significantly different between most sensor pairings for all three suspensions (Table 3-5, Table 3-6, Table 3-7). These tables show that the EXO sensor is consistently significantly lower, and the Hach 2100Q-is cuvette instrument mostly significantly higher, than other sensors. The other sensors tend to 'cluster' in an intermediate range of responses, and this clustering is quite 'tight' for river stormflow silt, but less so for kaolinite and pond water.

Table 3-5: **P-values from paired sample t-tests of regression slopes between each turbidity sensor model (stormflow silt suspension).** The regression slopes for the field sensors (EXO, Hach Solitax, WTW and Observator) were based on the mean response of the three units of each model tested. Bold values indicate a statistically significant difference in regression slopes (α = 0.05).

Sensor	EXO mean	Hach Solitax mean	WTW mean	Observator mean	Hach 2100Qis	Hach TL2310LED
EXO mean	X	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**
Hach Solitax mean		X	0.1197	**0.0062**	**<0.0001**	0.8891
WTW mean			X	0.9192	**0.0011**	0.1723
Observator mean				X	**<0.0001**	0.0623
Hach 2100Qis					X	**0.0003**
Hach TL2310LED						X

Table 3-6: **P-values from paired sample t-tests of regression slopes between each turbidity sensor model (kaolinite suspension).** The regression slopes for the field sensors (EXO, Hach Solitax, WTW and Observator) were based on the mean response of the three units of each model tested. Bold values indicate a statistically significant difference in regression slopes (α = 0.05).

Sensor	EXO mean	Hach Solitax mean	WTW mean	Observator mean	Hach 2100Qis	Hach TL2310LED
EXO mean	X	**0.0002**	**0.0002**	**<0.0001**	**<0.0001**	**0.0003**
Hach Solitax mean		X	0.2110	**0.0001**	**0.0025**	0.2885
WTW mean			X	**0.0022**	**0.0185**	0.8762
Observator mean				X	0.2868	**0.0023**
Hach 2100Qis					X	**0.0176**
Hach TL2310LED						X

Table 3-7: **P-values from paired sample t-tests of regression slopes between each turbidity sensor model (pond algae suspension).** Note, the regression slopes for the field sensor (EXO, Hach Solitax, WTW and Observator) were based on the mean of the response from the three units of each model used.

Sensor	EXO mean	Hach Solitax mean	WTW mean	Observator mean	Hach 2100Qis	Hach TL2310LED
EXO mean	X	**0.0020**	**0.0036**	**0.0004**	**0.0128**	**0.0033**
Hach Solitax mean		X	0.2209	**0.0009**	**0.0243**	**0.0165**
WTW mean			X	**0.0011**	**0.0228**	**0.0146**
Observator mean				X	0.1675	**0.0126**
Hach 2100Qis					X	**0.0500**
Hach TL2310LED						X

4 Discussion

In this this section we discuss the findings of our experiments in relation to (1) the comparability of ISO 7027-compliant turbidity sensor outputs and (2) the reproducibility of outputs from the same make-model of sensor. We then discuss the findings in the context of high frequency turbidity monitoring as a proxy for SPM concentration or visual water clarity (or related water quality variables of interest), with the aim of informing the next iteration of the NEMS *Turbidity*.

4.1 Sensor comparability

Our laboratory tank experiments show that different turbidity sensors compliant with the ISO 7027 standard output appreciably different numerical FNU values on the *same* fine sediment suspensions. There was about a two-fold numerical range across the sensors tested in the different test suspensions. This is appreciably less than the five-fold range reported by Rymszewicz et al. (2017) for a wider range of sensors (most not ISO 7027-compliant), but still 'problematically' large. The Hach 2100Q-is handheld sensor consistently returned the highest FNU values while the YSI EXO2 sonde sensors consistently returned the lowest FNU values. Numerical FNU values from the Hach Solitax, WTW Viso Turb, Observator Analite and laboratory Hach TL2310 sensors were intermediate and sometimes (notably for river silt) clustered fairly closely such that slopes of response lines were not all significantly different.

The variation in sensor response occurred despite calibration with the *same* freshly-made formazin standards. That is, we must regard the response differences as a real phenomenon and of practical concern. Previous authors have attributed differences in turbidity sensor response to differences in spectral emission of light source, spectral sensitivity of detector, light scattering angle (and angle-range) detected, and beam configuration – which combine with the very different light scattering properties of natural suspended particulate matter compared to formazin particles (Davies-Colley and Smith 2001; Rymszewicz et al. 2017).

Our results suggest that even very subtle differences in sensor design can combine with differences in optical characteristics of SPM to influence sensor response. Evidence for the influence of suspensoid optical characteristics (angular distribution of light scattering) is seen in the slightly different ratios of response of the tested sensors in the three test suspensions.

4.2 Sensor reproducibility

Our laboratory tank experiments showed that triplicates of field-type sensors agreed fairly closely, at least at higher concentrations of test suspensions (within 10% CV). This fairly good reproducibility has implications for sensor inter-changeability, particularly where interest is in the high range values occurring during runoff events. In this context, if a field sensor has to be swapped out due to loss or malfunction part-way through a field campaign it is ideally replaced with the *same*-make-model (with equivalent calibration) – with the reasonable expectation of a closely-similar response. That is, it should usually be safe to assume that the surrogate calibration data for the 'old' field sensor (e.g., to SSC) can be re-used with the new sensor, until such time as a surrogate calibration can be confirmed for the new sensor.

If substitution with a *different* make-model is unavoidable, the 'old' calibration data is not useless. The sensors we tested all responded very nearly linearly over the range < 1,000 FNU, so it is reasonable to expect that the responses of the 'new' and 'old' sensors will be in the same ratio despite numerical differences. This linearity can be exploited by estimating the ratio of response of

the two sensors from a plot of SSC or visibility versus field turbidity. The ratio of response is simply the slope of the calibration line after the sensor change with that of the original calibration line. If the monitoring targets the low range, which might be the case, for example, when monitoring versus attribute values for visual clarity in the NPS-FM (2020: Table 8), then the results presented here are ambivalent in regard to sensor interchangeability, because of greater relative variability among sensors (up to around 100%, Table B-2).

We note that if field turbidity sensors are locally calibrated on an ongoing basis to absolute quantities of interest such as SSC or beam-c, problems with step-change in response arising from swap-out of sensors should be avoided so long as the new calibration can be established quickly. Often, however, there is a time lag with building a calibration, and many months may pass before runoff events provide samples adequately covering the high range. The record becomes vulnerable to further instrument failure over this period, and there may be an urgent demand for useful results (e.g., calibrated SSC record for compliance purposes). In this context, the flexibility to interchange sensors of the same make-model remains important.

4.3   Implications for monitoring turbidity as a proxy for SPM concentration or visual water clarity

4.3.1   Sensor selection

Despite the differing numerical response (FNU values) on the same fine sediment suspension, there was a very strong linear correlation between the output of the different ISO 7027-compliant sensors – at least over the range tested. Therefore, all of the sensor make-models tested should be satisfactory for *in situ* continuous monitoring as a proxy for SPM concentration or water clarity or related water quality variable of interest (e.g., total phosphorus, *E. coli*), provided that there is site-specific 'calibration' to the water quality variable(s) of interest.

Because there is no 'best' turbidity sensor, and accuracy of turbidity is undefined, the numerical output of the chosen sensor seems irrelevant. The choice of sensor can be made on purely pragmatic grounds, including: sensor range, stability, vulnerability to fouling, power draw, reliability, and cost. The upper end of the measurement range is often important for SPM monitoring given that the majority of sediment is transported under storm flow conditions. While sensors with outputs of up to 3,000–5,000 FNU may be needed to derive robust catchment sediment load estimates, a range of 0–1,000 FNU may suffice for establishing a continuous record of visual water clarity. (A 1,000 FNU turbidity corresponds, very roughly, to a visibility of < 10 mm.)

4.3.2   Sensor calibration

Our experimental findings suggest that turbidity sensor calibration to beam attenuation coefficient (beam-c) provides an alternative approach to monitoring for sensor drift (and thus drift in local calibration, whether to beam-c or SSC). Beam-c can be measured discretely[7] both *in situ*, using on-the-spot air-calibration, or on water samples submitted to the laboratory for testing using a beam transmissometer operated in flow-through mode. For this, the most practical approach is to

---

[7] Measuring beam-c 'continuously' in the field is not usually practical, at least in rivers, because transmissometers are very vulnerable to fouling (much more so than turbidity nephelometers) and have a limited dynamic range (60-fold for transmissometers versus several 100-fold for nephelometers)

measure visual water clarity in the field (which can be accurately converted to beam-c)[8] as the on-site check against field turbidity sensor drift, backed up by laboratory beam-c measurement (measured on a water sample within 48 hr of collection). These calibrations, and checks, should normally be made monthly combined with sensor cleaning, but it is important to obtain some data under stormflow conditions (say with special visits three times per year or by use of auto-samplers), in order to extend the calibration to high values of beam-c (high turbidity and low clarity).

However, it should be noted that changes in sediment suspension properties (e.g., particle size, shape, composition, colour) can potentially cause a shift in relationship between field turbidity and beam-c – as evident, for example, from comparing this relationship for algae-rich pond water in Figure 3-4 with those relationships for the kaolinite and river silt mixtures in Figures 3-2 and 3-3. This creates ambiguity in whether an observed fluctuation or drift-trend has an environmental or instrument cause. Where change in sediment properties is to be expected, for example with major land use change or construction, a reference (dip mode) nephelometer, ideally of the same type as the field instrument, could be used to distinguish a real change (due to change in nature of SPM) in the beam-c-turbidity calibration versus drift of the field nephelometer – providing this check monitoring spans the period of change. This approach differs somewhat from the approach for monitoring turbidity sensor drift given in the NEMS *Turbidity*, which makes no assumptions of material changes but always includes regular field checks of the in-situ sensor against either an ISO 7027 portable sensor or on water samples taken to the laboratory and measured on an ISO 7027 bench turbidimeter. This resolves the potential ambiguity issue, but there are toxicity and stability concerns with using formazin that need to be carefully managed in a laboratory environment.

Thus, there are advantages and disadvantages of both the beam-c and NEMS *Turbidity* approaches for detecting instrument drift. Deciding which approach is better will be advanced by more information on the variability of beam-c vs turbidity relationships in natural suspensions, and perhaps also by the availability of a non-toxic and stable material reference replacing formazin. We note that AMCO Clear, a non-toxic, relatively stable synthetic polymer, has recently been certified by both the US EPA and ISO as an alternative reference material to formazin.

In the meantime, NIWA is currently preparing a Standard Operating Protocol (SOP) for NIWA field team staff regarding operation of field turbidimeters with local calibration (and checking for drift) using field visual clarity backed up by laboratory beam-c measurement (Dr Lucy McKergow pers. comm.). We recommend laboratory beam-c measurements are made with a beam transmissometer operated in flow-through mode with suitably (volumetrically) diluted water samples as in the current work (Section 2.2.7). Beam-c has the advantage that it is almost immune to step changes (caused by instrument change) of the type that can confound statistical summaries and trend-testing.

As regards discrete laboratory turbidity, our experimental finding of divergent numerical response for cuvette turbidimeters suggests that NZ's national discrete turbidity dataset, measured on different nephelometers in different laboratories at different times, may be compromised (not fully comparable) to an unknown extent. This has long been suspected. For example, Davies-Colley and Smith (2001) reported a 30% average shift in turbidity (NTU) for a wide range of NZ rivers (but

---

[8] Visual clarity (black disc visibility, y_{BD}) is inversely related to light beam attenuation at 550 nm, $c(550)$:

$y_{BD} = 4.8/c(550)$, where the beam attenuation coefficient at 550 nm (peak sensitivity of the human eye) can be estimated with good accuracy from measurements at 530 nm (e.g., a green-light WETLabs C-Star transmissometers) by the semi-empirical equation: $c(550) = 0.9*c(530) + 0.081$. Visual clarity can be measured with good precision (repeatability of about 5% standard deviation) and reproducibility of 7% *in situ* or on a sample contained in a SHMAK clarity tube (see NEMS *Water Quality*). Beam attenuation is capable of measurement on water samples in the laboratory with even better precision, of around 3% repeatability.

varying between individual rivers) with change in laboratory bench nephelometer from the 'traditional' Hach 2100A to a newer Hach 2100AN instrument. In their study of reproducibility of water quality measurements on six diverse rivers in the Wellington region, Davies-Colley et al. (2019) showed that turbidity, measured on two different Hach nephelometers (in two different laboratories), agreed only weakly – in marked contrast to the good agreement of paired visual clarity measurements.

5   Conclusions

The key finding from this study is that different ISO 7027-compliant turbidity sensors have numerically different responses to the same fine particle suspensions – despite identical calibration to formazin. We found a two-fold numerical range in FNU values across the different sensors used in our experiments; a wider range of sensors, combined with a wider range of suspensions, might plausibly further widen the range of responses. Note, however, that none of the sensors tested should be considered 'best' or better than the others: there is *no* 'correct' (absolute) turbidity.

The difference in the outputs of the different sensors is attributed to subtle differences in design combined with the very different light-scattering properties of natural suspended particulate matter compared to particles in formazin. Because of the optical difference between formazin and natural SPM, setting nephelometric turbidity sensors to read correctly on formazin does not guarantee their identical response in natural waters.

The difference in sensor numerical outputs also highlights that turbidity should not be applied as an absolute standard (e.g., in resource consent conditions)[9]. However, turbidity can be used as a *relative* measure (e.g., to indicate a change upstream versus downstream of a point source discharge).

Despite the different (and arbitrary) responses between different sensors, their FNU outputs were strongly linearly correlated. That is, all the tested sensors responded linearly to suspensions of river silt, kaolinite and phytoplankton. This linearity of response means that if nephelometers are *locally* calibrated to an absolute metric of interest (e.g., SSC or visual clarity), these quantities, inferred from turbidity records, can be compared on the same basis.

Field turbidity sensors are useful tools in water quality and related fields as a local proxy for sediment-related variables of interest – usually measures of suspended particulate matter (SPM) or of water clarity. Therefore, frequent (and on-going) local calibration of field turbidity to SPM concentrations or light beam attenuation is normally desirable.

---

[9] We note that the National Policy Statement for Freshwater Management (NPS-FM 2020) gives an attribute table (Table 8) to protect aquatic life from SPM damages, *only for visual clarity*, not turbidity.

6 Recommendations

We recommend that the findings from this report are used to inform the review of the NEMS *Turbidity*. In particular, that:

1. The scope of the NEMS *Turbidity* is extended to provide a basis for high frequency turbidity measurements as a proxy for key water quality variables, including visual water clarity, as well as SSC.

2. Beam attenuation is added to the NEMS *Turbidity* as a calibration metric – in view of its high precision, low cost and accurate interconversion with an important 'target' metric: visual clarity.

3. Research is needed on better methods for detecting field turbidity sensor drift. Beam attenuation should be suitable to detect drift except where a major shift in SPM character occurs, and so offers an alternative to the use of formazin or equivalent reference materials.

7   Acknowledgements

8　Glossary of abbreviations and terms

Absorption
(of light) Capture of light **photons** and transformation of their energy into another form, ultimately heat.

Attenuation
(of a light beam) Reduction in beam radiance due to both *scattering* and *absorption* of **photons**.

Beam attenuation coefficient
The sum of the *absorption* coefficient and *scattering* coefficient. Quantifies total attenuation of a light beam and controls *visual clarity*.

Calibration
(in metrology) Comparison of measurement values delivered by a sensor with a calibration standard.

Clay mineral
Hydrous aluminium phyllosilicates formed as rock weathering products and commonly found in soils and waters. Many clay minerals form hexagonal sheets similar to micas, for example the kaolin group, including kaolinite.

Drift
Shift of the output of a sensor from *calibration* over time. Sometimes referred to as 'non-stationarity'.

Formazin
An intensely scattering material used as a convenient (although arbitrary) material standard in *nephelometry* (relative measurement of light scattering). Produced by the reaction of hexamethylenetetramine with hydrazine sulphate in water to give colloidal particles of a heterocyclic polymer.

Nephelometry
Measurement of light *scattering* relative to that of a standard material. (An 'informal' measurement; not a proper SI measure of light scattering).

NIR
Near infra-red radiation. **Nephelometers** using light-emitting diodes (LEDs) emitting NIR are preferred for use in natural waters because of their insensitivity to light absorption by ubiquitous humic materials.

Photon
Quantum of electromagnetic radiation, including light.

Phytoplankton
Autotrophic (self-feeding) components of the plankton in waters, including cyanobacteria and (eukaryotic) algae such as the green algae.

Scattering function
Relative **scattering** as a function of *angle* of trajectory of scattered versus incident photons.

Scattering
(of light) Change in trajectory of light **photons** due to interactions with particles or molecules in the medium (e.g., water).

Silt
Granular material of a particular size range between clay(size) and sand. ISO 14588 grades silt between 2 and 63 μm – which is the definition we use.

SPM
Suspended particulate matter. (Particles of insoluble materials suspended in water).

SSC	Suspended sediment concentration. The mass of **SPM** per unit volume in a (whole) water sample as collected, operationally, by a glass-fibre filter.
TSS	Total suspended solids. The mass of **SPM** per unit volume in a volumetric subsample (assumed representative) of the whole water sample. (TSS should ideally equal SSC, but can differ in practice if fast-settling particles or floatables are not sub-sampled representatively).
Turbidity	Cloudiness or haze in a fluid caused by light *scattering* from suspended particles. Turbidity can be considered a measure, relative to **formazin** or other reference suspension, of light scattering through a specified angular range.
Visual clarity	(Visibility) Visual range in a medium (e.g., water). Visibility in the *horizontal* direction is completely controlled by the ***beam attenuation coefficient.***
Validation	Verifying, using on-site measurements or laboratory measurements on water samples, that measurement values delivered by a sensor are accurate.
VSS	Volatile suspended solids. The volatile content (loss of mass on ignition at a defined temperature and time) of **SPM** per unit volume in a volumetric subsample (assumed representative) of a water sample. Index of organic content.

9 References

Austin, R.W. (1973) Problems in measuring turbidity as a water quality parameter. *USEPA seminar on methodology for monitoring the marine environment October 16-18, 1973*, Seattle, WA.

Davies-Colley, R., Hughes, A. (2020) Sediment-related water quality of small hill-country streams near Whatawhata, New Zealand. Response to integrated catchment management (ICM). *New Zealand Journal of Marine and Freshwater Research*: 1-25. 10.1080/00288330.2020.1761840

Davies-Colley, R., Milne, J., Heath, M.W. (2019) Reproducibility of river water quality measurements: inter-agency comparisons for quality assurance. *New Zealand Journal of Marine and Freshwater Research*, 53(3): 437-450. 10.1080/00288330.2019.1585886

Davies-Colley, R.J., Hughes, A.O. (2019) Continuous turbidity records as a proxy for water quality at NIWA benchmark sites. *NIWA Client Report 2019220HN*. 60 p.

Davies-Colley, R.J., Smith, D.G. (2001) Turbidity, suspended sediment, and water clarity: A review. *Journal of the American Water Resources Association*, 37(5): 1,085-1,101.

Gibbs, R.J. (1978) Light scattering from particles of different shapes. *Journal of Geophysical Research*, 83C1: 501-502.

Grove, M.K., Bilotta, G.S. (2014) On the use of loss-on-ignition techniques to quantify fluvial particulate organic carbon. *Earth Surf. Process. Landforms*, 39: 1,146–1,152. DOI: 10.1002/esp.3509

Hughes, A., Davies-Colley , R., Heubeck, S. (2019) Comparability of ISO 7027 compliant turbidity sensors. *NIWA Client Report 2019125HN*. 26 p.

ISO (1999) *ISO 7027:1999. Water quality — Determination of turbidity*. International Organization for Standardization (ISO). 10 p.

ISO (2016) *ISO 7027-1:2016. Water quality — Determination of turbidity*. International Organization for Standardization (ISO). 9 p.

Kirk, J.T.O. (2011) *Light and photosynthesis in aquatic ecosystems*. Cambridge University Press, New York. 649 p.

Kitchener, B.G.B., Wainwright, J., Parsons, A.J. (2017) A review of the principles of turbidity measurement. *Progress in Physical Geography: Earth and Environment*, 41(5): 620-642. 10.1177/0309133317726540

McCluney, W.R. (1975) Radiometry of water turbidity measurement. *Journal of the Water Pollution Control Federation*, 47: 252-266.

NEMS (2016) National Environmental Monitoring Standard - Turbidity: measurement, processing and archiving of turbidity data. *National Environmental Monitoring Standard*: 74 p. http://www.lawa.org.nz/media/16586/nems-turbidity-recording-2013-06.pdf

Phillips, J., Russell, M., Walling, D. (2000) Time-integrated sampling of fluvial suspended sediment: a simple methodology for small catchments. *Hydrological processes*, 14(14): 2,589-2,602.

Rymszewicz, A., O'Sullivan, J.J., Bruen, M., Turner, J.N., Lawler, D.M., Conroy, E., Kelly-Quinn, M. (2017) Measurement differences between turbidity instruments, and their implications for suspended sediment concentration and load calculations: A sensor inter-comparison study. *Journal of Environmental Management*, 199: 99-108. https://doi.org/10.1016/j.jenvman.2017.05.017

Velleman, P.F. (2012) DataDesk: an interactive package for data exploration, display, model building, and data analysis. *WIREs Compuational Statistics*, 4(4): 407-414. https://doi.org/10.1002/wics.1208

Zanevald, J.R.V., Pegau, W.S. (2003) Robust underwater visibility parameter. *Optics express*, 11: 2997-3009. http://www.opticsexpress.org/abstract.cfm?URI=OPEX-11-23-2997

Appendix A   Formazin calibration plots and lines

All sensors (both cuvette instruments and field sensors) were 'locally calibrated' by sequential measurements in the *same* formazin standards (treated like another sample suspension) over the 10-1,000 FNU range. Further information can be found in Section 2.3 of this report. Shown below are the calibration data (Table A-1, Table A-2) and plots of the before calibration data (Figures A-1 to A5) The calibration plots are given as formazin standard versus sensor response (contrary to the convention of putting the independent variable on the X-axis) because the calibration equations are formazin standard value versus sensor response. Equations of calibration lines (before tank tests) are given in Table A-3, and a table of t-tests of before versus after calibration lines is given in Table A-4. Two of the Hach sensors differed 'significantly' in calibration after the tank tests and one of the WTWs, while all three of the Observator lines had shifted 'significantly'. However, the apparent calibration shifts were small in magnitude despite the t-test results (which, in any case are very strongly driven by the high – 1,000 FNU – point), so we chose to ignore these shifts and just use the 'before' calibration lines from Table A-3.

Table A-1: **Calibrations before tank tests (8 June 2020).** All data are numerical responses in 'raw' FNU.

Standard	EXO1	EXO2	EXO6	Hach1	Hach2	Hach3	WTW4	WTW5	WTW6	Observator7	Observator8	Observator9	2100Qis	TL2310
10 FNU	9.2	9.8	10.2	8.3	10.2	12.2	11.6	13.3	No data	10.4	10.2	10.3	10.6	9.66
25 FNU	24.1	24.6	25.1	21.6	26.3	30.8	27.3	30.6	27.3	25.62	25.25	25.3	26.3	23.8
100 FNU	93.5	98.5	103.8	93.7	106.0	121.0	103.5	117.0	102	103.78	101.7	102.72	104	94.9
250 FNU	261.1	268.3	261.1	236	270	306	255	274	279	258.78	247.31	261.14	254	244
1,000 FNU	993.4	1,023.4	984.2	1,036	1,022	1,235	1,062	1,050	1,030	No data	No data	No data	No data	950

Table A-2: **Calibrations after tank tests (12 June 2020).** All data are numerical responses in 'raw' FNU.

Standard	EXO1	EXO2	EXO6	Hach1	Hach2	Hach3	WTW4	WTW5	WTW6	Observator7	Observator8	Observator9	2100Qis	TL2310
10 FNU	9.9	9.9	9.6	11.6	12.1	14.4	15.8	18.5	14.8	10.6	10.4	10.6	10.9	9.7
25 FNU	24.7	24.8	24.6	25.0	27.2	32.3	32.5	39.8	33.3	25.1	24.3	24.9	26.2	23.6
100 FNU	99.3	100.5	98.7	97.5	104.1	123.6	102.9	122.3	108.3	98.5	98.9	98.8	102.0	94.6
250 FNU	258.3	256.1	257.6	241.0	263.3	310.3	246.1	284.7	258.1	252.7	244.4	264.1	252.0	246.0
1,000 FNU	978.4	1,038.0	974.9	1,009.3	977.6	1,198.1	994.7	1,063.9	995.3	949.7	906.1	875.7	No data	944.0
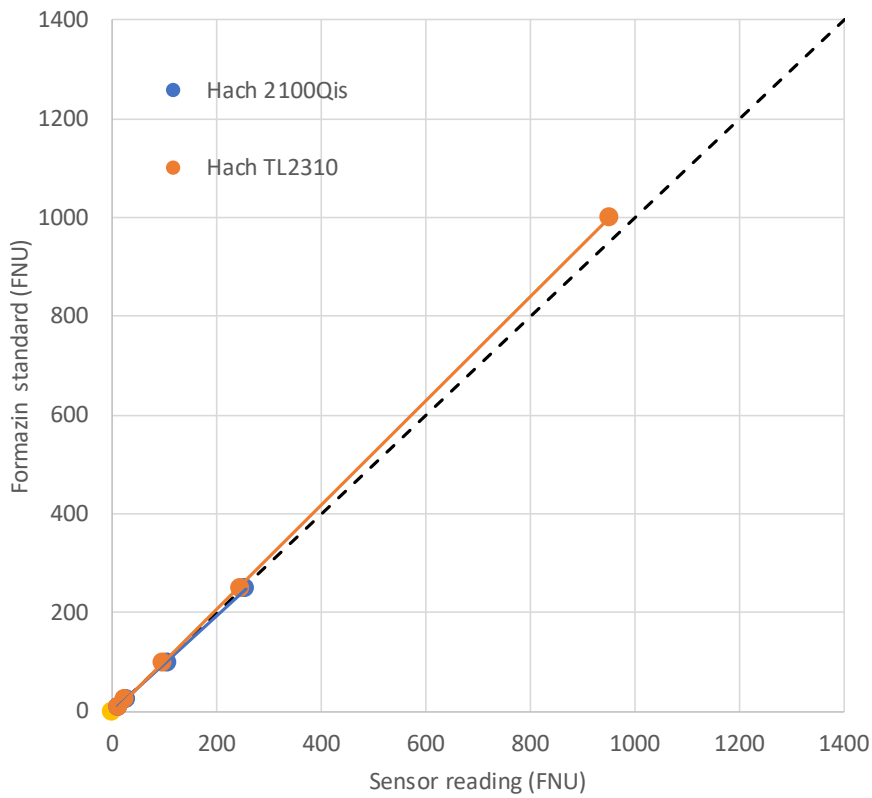
Figure A-1: **Cuvette turbidity instrument response in Formazin standards.** (Before tank tests).
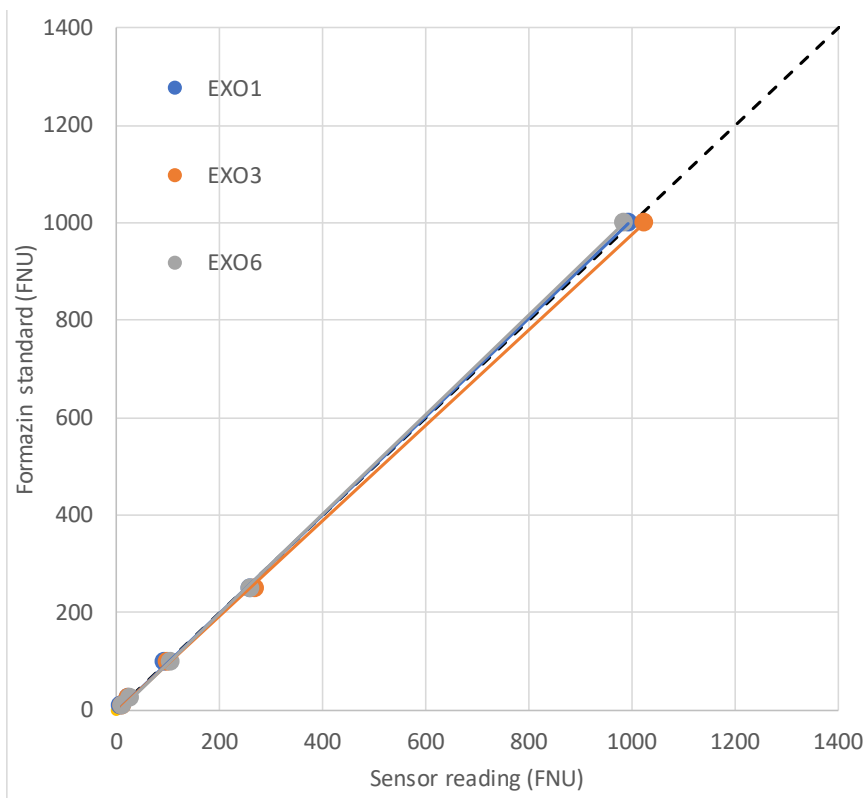
Figure A-2: **EXO Sonde Turbidity response in Formazin standards.** (Before tank tests).
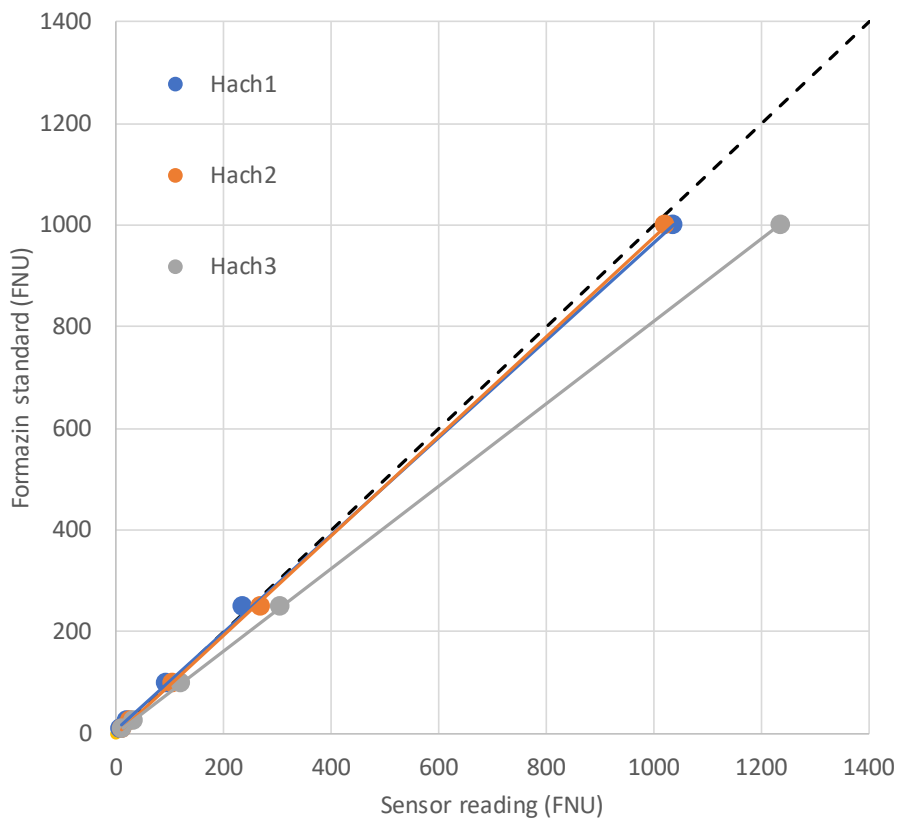
Figure A-3: **Hach Solitax response in Formazin standards.** (Before tank tests).
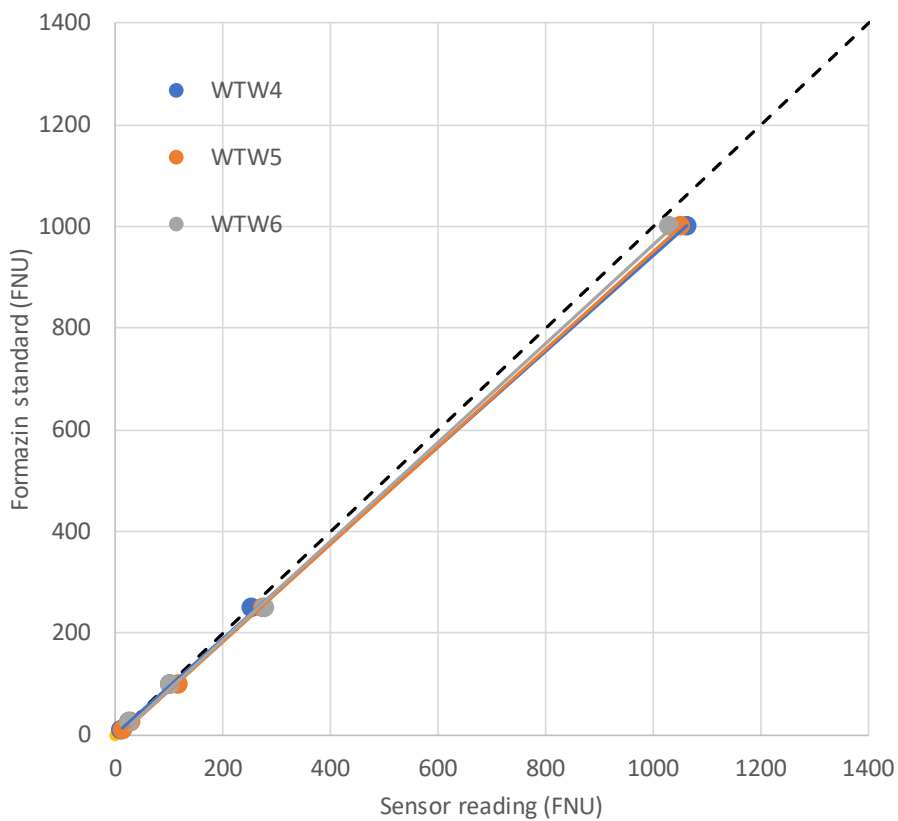
Figure A-4: **WTW VisoTurb response in Formazin standards.** (Before tank tests).
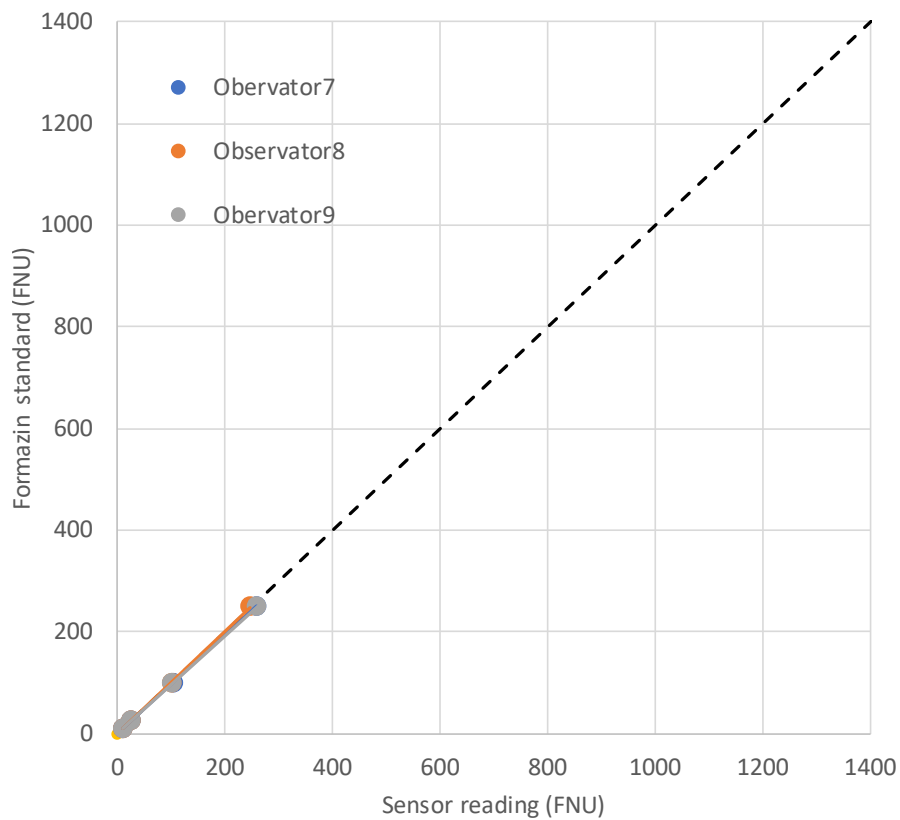
Figure A-5:   Observator Analite response in Formazin standards. (Before tank tests). Note that the high point (1000 FNU) over-ranged the A/D setting which was subsequently re-set for tank tests and re-calibrations.

Table A-3: Linear equations quantifying response of field-type sensors to Formazin standards. Slope, Y-intercept and R^2 are given for both the cuvette turbidity instruments and for each of four field sensors (averages across a triplicate) in Formazin standards in the range of 10-1,000 FNU. Data are for before tank tests with the sensors.

Sensor	n	Slope	y intercept	R^2
EXO1	5	1.00	-0.58	1.00
EXO3	5	0.98	-1.06	1.00
EXO6	5	1.02	-4.92	1.00
Hach1	5	0.96	9.20	1.00
Hach2	5	0.98	-4.16	1.00
Hach3	5	0.81	1.08	1.00
WTW4	5	0.94	2.31	1.00
WTW5	5	0.96	-7.49	1.00
WTW6	5	0.97	-6.45	1.00
Observator7	4	0.97	0.03	1.00
Observator8	4	1.01	-1.04	1.00
Obervator9	4	0.96	0.81	1.00
Hach 2100Qis	4	0.99	-1.16	1.00
Hach TL2310	5	1.05	-1.47	1.00

Table A-4:   T-test results for before versus after calibrations. Bold indicates statistically significant differences (alpha = 0.05). Note that the 'before' calibration lines for the three Observators (also the Hach 2100Qis) were missing the high (1000 FNU) datapoints, so the comparison of before (*N* = 4) versus after (*N* = 5) calibration lines is compromised.

Sensor	Slope (before)	Slope (after)	p-value
EXO1	1.0048	1.0229	0.2117
EXO3	0.9759	0.9619	0.1544
EXO6	1.0182	1.0265	0.5397
Hach1	0.9595	0.9899	0.0639
Hach2	**0.9800**	**1.0271**	**0.0130**
Hach3	**0.8092**	**0.8368**	**0.0004**
WTW4	**0.9411**	**1.0111**	**0.0001**
WTW5	0.9579	0.9512	0.4770
WTW6	0.9739	1.0124	0.0637
Observator7	**0.9657**	**1.0554**	**0.0003**
Observator8	**1.0122**	**1.1080**	**0.0014**
Observator9	**0.9557**	**1.1473**	**0.0027**
Hach 2100Qis	0.9867	0.9958	0.2398
Hach TL2310	1.0529	1.0598	0.4344

Appendix B   Responses of turbidity sensors in different test suspensions

Table B-1 gives response (in calibrated FNUs) of sensors in different test suspensions.

The agreement between triplicates of the same-make-model, measured as coefficients of variation (CV), were shown in Section 3.3.1 to mostly be fairly good at the higher concentrations studied. The agreement is appreciably weaker in a relative sense at lower concentrations as shown in Table B-2. Any variations between different make-models in their response to Formazin standards and their subsequent calibration linear equations amplifies any large variations seen in the lower concentrations and the resulting average sensor value and its CV.

Table B-1: Response of turbidity sensors (after calibration) in different test suspensions. Values are given for both the cuvette instruments and each of four field sensors in each of three different suspensions All values shown are in (calibrated) FNU units except beam-c (m^{-1}).

Sample name	Beam c (m^{-1})	Exo1	Exo2	Exo6	Hach1	Hach2	Hach3	WTW4	WTW5	WTW6	Observator7	Observator8	Observator9	2100Qis	TL2310
Silt11	6.6	5.7	5.3	1.4	17.0	4.8	9.9	10.4	10.6	5.3	11.6	10.7	13.5	10.2	9.7
Silt37	19.3	18.5	16.4	13.8	34.2	23.1	27.6	26.8	27.5	24.6	35.9	30.1	35.3	31.6	25.7
Silt111	59.4	54.0	50.9	51.7	85.5	80.8	83.6	78.3	83.0	73.7	113.5	108.9	104.0	131.1	84.6
Silt333	183.0	178.0	172.8	171.5	265.8	299.5	282.7	261.2	278.7	266.4	340.1	303.5	335.8	427.1	246.0
Silt1000	553.0	512.1	503.5	522.9	919.7	953.2	962.0	1,001.9	1,006.4	1,047.9	1,066.1	1,010.8	1,020.1	-	939.8
Kaolinite7	3.9	4.2	3.5	-0.4	13.7	1.2	6.3	7.3	7.4	2.3	7.1	6.7	8.3	5.4	6.3
Kaolinite23	11.2	11.9	10.9	7.6	23.7	10.8	17.0	16.3	19.5	12.6	20.9	20.8	24.1	14.3	14.0
Kaolinite70	32.7	38.4	33.1	33.7	50.2	43.9	46.2	43.7	55.5	40.4	63.1	62.9	69.0	59.9	42.5
Kaolinite230	98.5	114.1	103.6	116.8	143.8	146.8	141.4	140.9	160.0	139.2	183.1	213.3	217.5	203.1	142.8
Kaolinite700	277.0	352.5	329.3	344.8	461.1	480.5	475.7	481.1	510.1	518.6	570.6	577.2	664.2	636.2	497.6
Algae3	4.6	1.9	1.4	-2.6	12.0	-0.8	4.3	5.6	5.5	0.1	4.4	3.7	5.3	3.2	2.1
Algae10	15.0	6.0	5.4	1.6	17.4	4.9	10.2	10.6	10.8	6.2	13.4	12.0	14.0	10.4	8.5
Algae30	47.1	18.4	19.3	15.6	35.2	24.0	27.8	27.3	28.9	24.8	37.8	38.1	40.7	42.2	30.5

Numerical comparability of ISO 7027-compliant turbidity sensors commonly used in New Zealand

Table B-2: **Variation in response of field-type turbidity sensors.** Coefficients of variation of the (triplicates of) four different field-type sensors characterise the within make-model variation.

Sample name	Beam c (m^{-1})	EXO mean (FNU)	SD	CV (%)	Hach mean (FNU)	SD	CV (%)	WTW mean (FNU)	SD	CV (%)	Observator mean (FNU)	SD	CV (%)
Silt11	6.57	4.1	2.4	57.1	10.6	6.1	57.9	8.8	3.0	34.7	11.9	1.4	11.9
Silt37	19.3	16.2	2.3	14.5	28.3	5.6	19.9	26.3	1.5	5.7	33.8	3.2	9.4
Silt111	59.4	52.2	1.6	3.0	83.3	2.4	2.8	78.3	4.7	6.0	108.8	4.7	4.4
Silt333	183	174.1	3.4	2.0	282.6	16.9	6.0	268.8	9.0	3.3	326.5	20.0	6.1
Silt1000	553	512.8	9.7	1.9	945.0	22.3	2.4	1,018.7	25.4	2.5	1,032.3	29.6	2.9
Kaolinite7	3.91	2.5	2.5	100.2	7.1	6.3	88.8	5.7	2.9	51.1	7.4	0.8	11.5
Kaolinite23	11.2	10.1	2.3	22.5	17.1	6.5	37.7	16.1	3.5	21.5	21.9	1.9	8.7
Kaolinite70	32.7	35.1	2.9	8.2	46.8	3.2	6.8	46.6	7.9	17.1	65.0	3.5	5.3
Kaolinite230	98.5	111.5	7.0	6.3	144.0	2.7	1.9	146.7	11.5	7.8	204.6	18.8	9.2
Kaolinite700	277	342.2	11.8	3.4	472.4	10.1	2.1	503.3	19.7	3.9	604.0	52.2	8.6
Algae3	4.6	0.2	2.5	1,097.8	5.2	6.4	124.1	3.7	3.2	84.6	4.5	0.8	18.8
Algae10	15	4.3	2.4	55.1	10.8	6.3	57.8	9.2	2.6	28.2	13.1	1.0	7.7
Algae30	47.1	17.8	1.9	10.8	29.0	5.7	19.6	27.0	2.1	7.6	38.9	1.6	4.1