

Comparisons of Turbidity Data Collected with Different Instruments

Report on a Cooperative Agreement Between the California Department of Forestry and Fire Protection and USDA Forest Service--Pacific Southwest Research Station (PSW Agreement # 06-CO-11272133-041)

Additional Funding was Provided by the State Water Resources Control Board's Surface Water Ambient Monitoring Program.

Jack Lewis, US Forest Service, Pacific Southwest Experiment Station
Rand Eads, RiverMetrics LLC
Randy Klein, Hydrologist

ABSTRACT

Different turbidity measurement devices do not necessarily produce compatible data, even when calibrated to the same standards. A variety of devices have been and will continue to be used for measuring turbidity in different watershed studies. It would be of great benefit if data from different locations could be compared through post-processing. The main objectives of this study are to quantify differences among several turbidity measurement devices and to determine the magnitude of the potential errors associated with attempts to standardize turbidity data. If a relationship between two measurement devices is insensitive to the suspension being measured, then conversions between different devices can easily be made once the relationship is established. Otherwise, conversions will introduce errors that could influence conclusions and two sensors might rank the same set of samples differently.

This investigation compares eight turbidity measurement devices that are commonly used to measure turbidity in field or laboratory settings. The devices are compared in 24 different sediment samples, each mixed to seven different concentrations, from ten watersheds in northern coastal California. A mixing apparatus was designed and constructed for this study to keep samples suspended for measurement by *in situ* sensors. The average of three measurements was recorded for each combination of device and sediment mixture.

Readings of the same sediment mixture by different sensors commonly differed by up to a factor of two, and in the most extreme case, by a factor of three. As expected, the relationships of turbidity to suspended sediment concentration depended strongly on the sediment as well as the sensor because of differences in particle size, shape and composition.

Comparing relationships between sensor readings in various sediment mixtures, we found considerable dependence upon sediment type for most sensor pairings. For a given turbidity by sensor A, turbidity by sensor B can vary by a factor of up to two or more across sediment samples for some sensor pairings. Relationships between sensors are

often curvilinear and tend to diverge as turbidity increases. Somewhat less dependency on sediment type was observed between sensors that conform to the same measurement standard (backscatter, EPA Method 180.1, or ISO 7027).

Various relationships were considered to evaluate the error associated with assuming a fixed relationship for converting turbidity readings among sensor types. Among parametric models considered, log-log regression most often best described the relationships, followed by quadratic regression. The form of the best relationship depends upon the particular sensor pairing. This report presents statistics for the best relationship found for each sensor pairing. Average errors associated with assuming these fixed relationships varied from 2.0 to 18.3% (mean 9.1%), while maximum errors varied from 12.5 to 83.0% (mean 39.2%). It is difficult to generalize because errors depend on the particular sensors and sediments involved. Maximum errors are generally associated with extreme textures. In applying the relationships reported here, it will be imperative to evaluate the robustness of conclusions to the expected magnitude of errors, considering the particular sediments under study.

INTRODUCTION

Different turbidity instruments yield different results for the same sample, even when calibrated to the same standards (Gray and Glysson, 2003; Anderson, 2004). Davies-Colley and Smith (2001) recommend abandonment of turbidity monitoring in favor of a more reproducible parameter—beam attenuation. However attenuation devices are non-linear and work best when turbidity is less than about 100 attenuation units (personal communication; John Downing, President, D&A Instruments Co., Port Townsend, WA, March 2007). For the time being, turbidity instruments based on the principle of light-scattering are well-entrenched in the practice of forestry-related stream water quality monitoring (Harris et al., in press). They are likely to remain so unless advancements in other technologies such as laser scattering and transmissometry or beam attenuation make them more suitable and cost-effective for *in-situ* stream deployments.

The U.S. Geological Survey has recently prepared guidance for collection and reporting of turbidity data (Anderson 2004). Recognizing the broad array of dynamic turbidity sensors (*in situ*, submersible sensor) and static (bench top and portable) meters, the report states:

“The use of consistent procedures and instruments within and among projects or programs for which turbidity data will be compared over space and time is crucial for the success of the data-collection program”

However, the multitude of manufacturers and different entities monitoring water quality with different purposes and budget constraints inevitably has resulted in the use of different instrumentation for measuring turbidity in different studies.

Practitioners collecting turbidity data may not be aware of the incompatibility of results from different types of sensors, potentially leading to erroneous comparisons of turbidity

data sets. There is potential for combining regional turbidity data for use in broad-scale analyses if the data are directly comparable or can be made so through post-processing. The potential benefit of combining regional turbidity data sets is the attainment of larger sample sizes, representing a wider range of environmental conditions and disturbance regimes, and providing a stronger basis for drawing meaningful conclusions.

Preliminary analyses of data from multiple sensors deployed simultaneously at the same field location suggest that strong relationships can be developed for converting one sensor's data to equivalent values for another sensor. However, it is not known how consistent such conversions are for different suspended sediment particle sizes and compositions, hence among streams and rivers or between years at the same location. The main objectives of this study are to quantify differences among several turbidity measurement devices and to determine the magnitude of the potential errors associated with attempts to standardize turbidity data. It is expected that such conversions can be done with definable error limits that may permit some uses of the data and prohibit others.

If relationships can be developed among different sensors and those relationships vary significantly according to sediment composition, then the interpretation of data becomes problematic as illustrated in Figure 1, where the relationship between turbidity sensors A and B differ for streams X and Y. If current conditions correspond to the points shown on the graph, then sensor A tells us stream Y is more turbid while sensor B tells us stream X is more turbid. Although neither sensor can be considered the 'standard', one must necessarily be chosen as such before we can agree on which stream is more turbid. Since the two relationships in Figure 1 are well separated, it is easy to find conditions that are ranked differently by the two sensors. If relationships do not differ greatly from stream to stream, then the rankings are less likely to depend on the sensor. Therefore, it is very important to establish how sensitive these relationships are to the medium being measured and to consider not only the specific relationships among sensors, but also the degree to which relationships are likely to shift among different sediments.

METHODS

Sensors

Four sensors dominate continuous turbidity data collected in northern coastal California¹:

- The OBS-3, an analog backscatter sensor made by D&A Instruments Co.
- The DTS-12 digital sensor (formerly OBS-12), made by Forest Technology Systems, Inc.
- Models 6026 and 6136 made by YSI (Yellow Springs Instrument Co.)

The OBS-3 is no longer manufactured and has been replaced by another backscatter sensor, the OBS-3+. Other turbidity instruments in common use are the Hach 2100P, a

¹ The use of trade, product, industry or firm names is for descriptive purposes only and does not imply endorsement by the authors

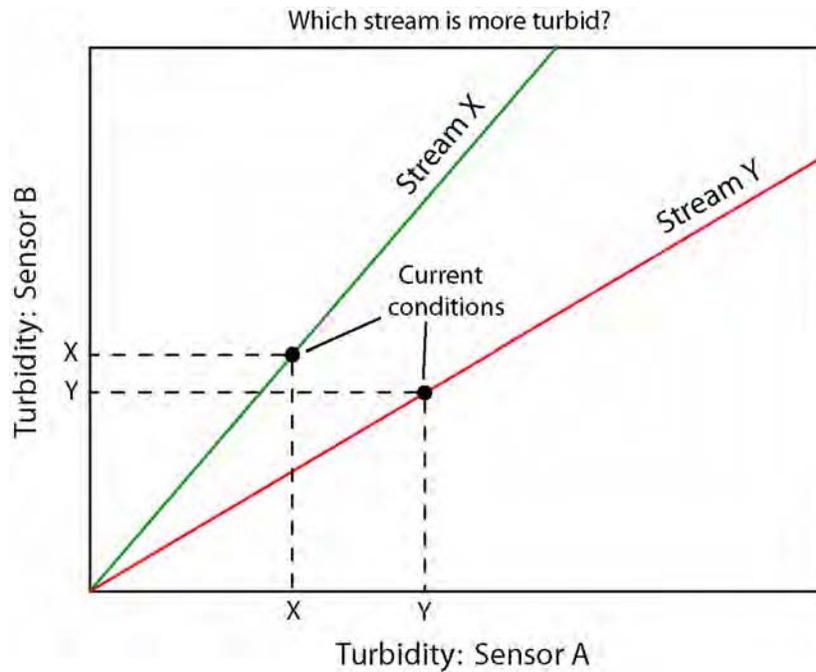


Figure 1. If relationships between sensors depend strongly on the stream being measured, then turbidity rankings of streams may differ by sensor.

portable meter often used to measure grab samples in the field, and Hach laboratory models 2100N and 2100AN. The sensors and meters included in this study are summarized in Table 1. The Analite NEP395, made by McVan Instruments Co., is not

Table 1. Turbidity sensors and meters included in experiment

Sensor	Light source	Detector angle(s)	Range of scattering angles	Method	Upper limit, units ^{1,2}	Use	Wiper
OBS-3	865 nm	N/A	140-160 ³	N/A	2000 FBU	<i>in situ</i>	No
OBS-3+	850 nm	N/A	110-140 ³	N/A	2000 FBU	<i>in situ</i>	No
DTS-12	780 nm	65	85 -160 ³	N/A	2000 FNU	<i>in situ</i>	Yes
NEP395	860 nm	90	75-105 ⁴	ISO 7027 ⁵	1000 FNU	<i>in situ</i>	Yes
YSI 6026	860 nm	90	75-105 ⁴	ISO 7027 ⁵	1000 FNU	<i>in situ</i>	Yes
YSI 6136	860 nm	90	75-105 ⁴	ISO 7027 ⁵	1000 FNU	<i>in situ</i>	Yes
2100P	tungsten	90,180 ⁶	0-90	EPA 180.1 ⁷	1000 NTRU	portable	No
2100AN ⁸	tungsten	45, 90, 135, 180 ⁶	0-135	EPA 180.1 ⁷	10000 NTRU	static	No

¹ Units are defined according to light source and detection angle (Anderson, 2005).

² Some of these sensors can be calibrated or set for alternate ranges.

³ Angles specified for turbidity-free water and include refraction effects (reported by John Downing, D&A Instruments Co.).

⁴ Angles specified for air (determined from ISO 7027 definition)

⁵ International Organization for Standardization, method 7027.

⁶ Turbidity derived as ratio of light scattering at different angles by multiple detectors.

⁷ U.S. Environmental Protection Agency, method 180.1 (ratiometric).

⁸ Hach 2100AN has several options for light source and detection angle, however this experiment utilized only the settings shown in this table.

widely used in California at this time but incorporates a combination of features that the authors felt warranted its inclusion in this study. Not all of the sensors had been calibrated prior to our obtaining them, so we read each sensor in 0, 500, 1000, and 2000 formazin standards at the beginning and end of the experiment. Since none of the sensors changed appreciably, the average of the two calibration data sets was used to compute a single calibration equation for each sensor. No corrections were necessary for the DTS-12 or Hach 2100P. Linear equations were adequate for the YSI 6136 and NEP395, and quadratic equations were needed for the OBS-3, OBS-3+, YSI 6026 and Hach 2100AN. These equations were used to mathematically post-correct each sensor's data so that all sensors would have returned the same turbidity in formazin standards.

The digital sensors (DTS-12, NEP395, YSI 6026 and YSI 6136) each use their own algorithms for determining turbidity from a series of rapid measurements. The DTS-12 and NEP395 report either the mean or median turbidity, while the YSI sensors use a proprietary filtering algorithm. We recorded the median turbidity from the DTS-12 and NEP395 sensors and the filtered YSI readings. For the OBS-3 sensors, we used our own algorithm that records the median of 60 measurements taken at half-second intervals.

Sediments

Sediment samples were collected from 10 watersheds located in the northern part of California's Coast Ranges: Garcia River, Navarro River, and Caspar Creek in Mendocino County; Bull Creek in southern Humboldt County; Elk River, Jacoby Creek, and Freshwater Creek in the Humboldt Bay area; Prairie Creek, Lost Man Creek, and Larry Damm Creek in Redwood National Park (Figure 2). These watersheds all have soils derived predominantly from late Mesozoic or Cenozoic sedimentary rocks, including the Franciscan Formation and other marine or continental sedimentary deposits. Fine suspendable material was targeted for sampling, including

1. Streamside landslide toe material
2. Colluvial or residual streambanks
3. Road inboard ditches, especially at culvert inlets
4. In-channel alluvial deposits (fine material in backwater deposits, typically near or within log jams or overbank flood deposits)

Sediment descriptions are provided in Table 2. The samples include a wide distribution of textures (Figure 3), with 0-50% clay, 10-80% silt and 0-90% sand. The organic content of each sample was determined by loss on ignition. Samples were burned in a muffle furnace for 4 hours at 550° C. Percent organics fell in a narrow range from 2.5% to 6.6%, except sample PRU1, which had 17.1% organic content (Table 3). Particle size distribution by volume was also determined by laser diffraction using a Micromeritics Saturn Digisizer 5200 (Table 4).

Mixing apparatus

A mixing apparatus was devised for suspending the sediments during measurement by the *in situ* sensors. The apparatus consisted of a stand for suspending a turbidity sensor and a variable-speed electric drill fitted with a paint-mixing paddle in an 11.4-liter (12-quart) feed bucket (Figure 4a). Each turbidity sensor was fitted with a mounting bracket

Table 2. Sediment samples used in experiment

ID	Watershed	Texture (see Fig 3)	Description
GAR2	Garcia R	a - loam	road surface runoff deposit
GAR4	Garcia R	b - sandy loam	alluvium from overflow channel
HHB1	Freshwater Cr	c - clay loam	roadside flood plain deposit
HHB2	Freshwater Cr	d - sandy loam	alluvial deposit near top of streambank under bridge
HHB3	Freshwater Cr	e - clay loam	residual streambank, B horizon of redwood forest soil
HHB4	Freshwater Cr	f - clay loam	alluvial or residual material from road culvert inlet
JBW1	Jacoby Cr	g - sandy loam or sandy clay loam	alluvial backwater created by large woody debris
JBW2	Jacoby Cr	h - sand	high flow backwater pool
KRW2	NF Elk R	i - silt loam	bladed material from road ditch
LDC1	Larry Damm Cr	j - silt loam	dark brown material from alluvial backwater
LLM1	Little Lost Man Cr	k - clay loam or silty clay loam	colluvial mixture of old alluvium and regolith from 5 feet above channel
LLM2	Little Lost Man Cr	l - loam	eroding alluvial streambank
LMC1	Lost Man Cr	m - silt, silt loam	goeey grey, mottled parent material in streambank
LMC2	Lost Man Cr	n - sand	alluvial backwater from abandoned pool
MBU1	Bull Cr	o - clay or silty clay	landslide deposit
NAV2	Navarro R	p - silty clay loam	fine wet mud skimmed off surface of alluvial backwater
NAV3	Navarro R	q - silty clay	landslide deposit
NFC1	Caspar Cr	r - clay	landslide toe deposit
NFC2	Caspar Cr	s - sandy clay or sandy clay loam	streambank
NFC3	Caspar Cr	t - sand	road ditch
PRU1	Upper Prairie Cr	u - sandy loam	channel margin on streambed, high in organics
SFM1	SF Elk R	v - silt loam	flood deposit, top of bank
SFM2	SF Elk R	w - silt loam	unmottled colluvium from toe of landslide deposit
SFM3	SF Elk R	x - silty clay	mottled colluvium from toe of landslide deposit

(Figure 4b) that could be quickly attached or released from the stand with a wing nut. The mixing paddle was positioned near the bottom of the bucket and the drill speed was set as high as possible, approximately 400 rpm, without creating observable bubbles (Figure 4c). The flat side of the bucket helped create a turbulent mix (Figure 4d) that prevented the segregation of large particles to the outside as would be expected in a round centrifuge.

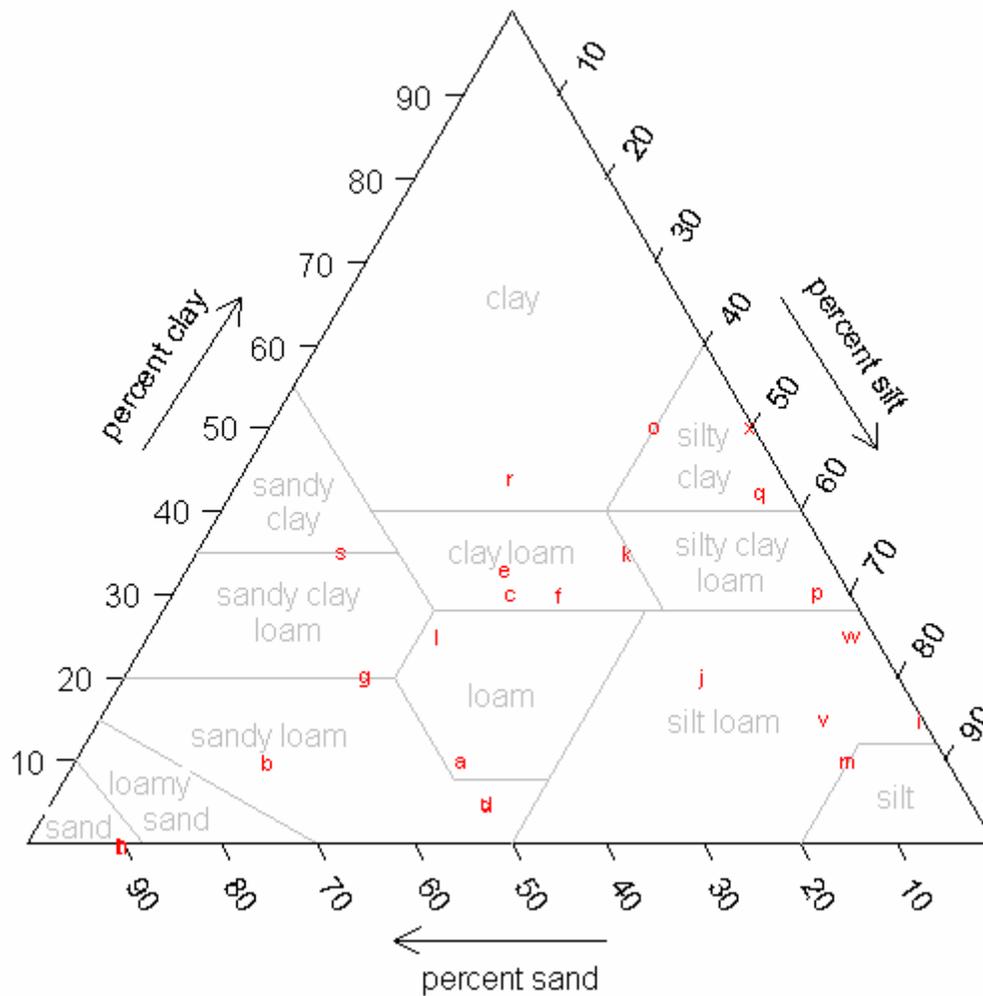


Figure 3. Texture-by-feel of sediment samples listed in Table 2

Experimental procedure

Samples were initially wet-sieved to remove gravel and sand particles larger than 0.5 mm (Figure 5a) and diluted to a volume of about eight liters. While sand particles up to 2 mm may be suspended in rivers, it would have been difficult to keep such particles suspended in the mixing apparatus, and the contribution of medium and coarse sand to turbidity is generally unimportant when finer particles are also present (Foster et al., 1992). The OBS-3 was mounted first for setting the nominal turbidity levels (targets) because, in contrast to several of the sensors, it could return instantaneous readings. The wet-sieved slurry was stirred and aliquots were added to the mixing bucket (Figure 5b) as necessary to reach targets of 25, 50, 100, 200, 400, 800, and 1200 turbidity units (Figure 5b). The highest nominal level was originally set at 1600 units, but when it was discovered that only the OBS-3, OBS-3+, and Hach 2100AN could read those mixtures, we lowered the

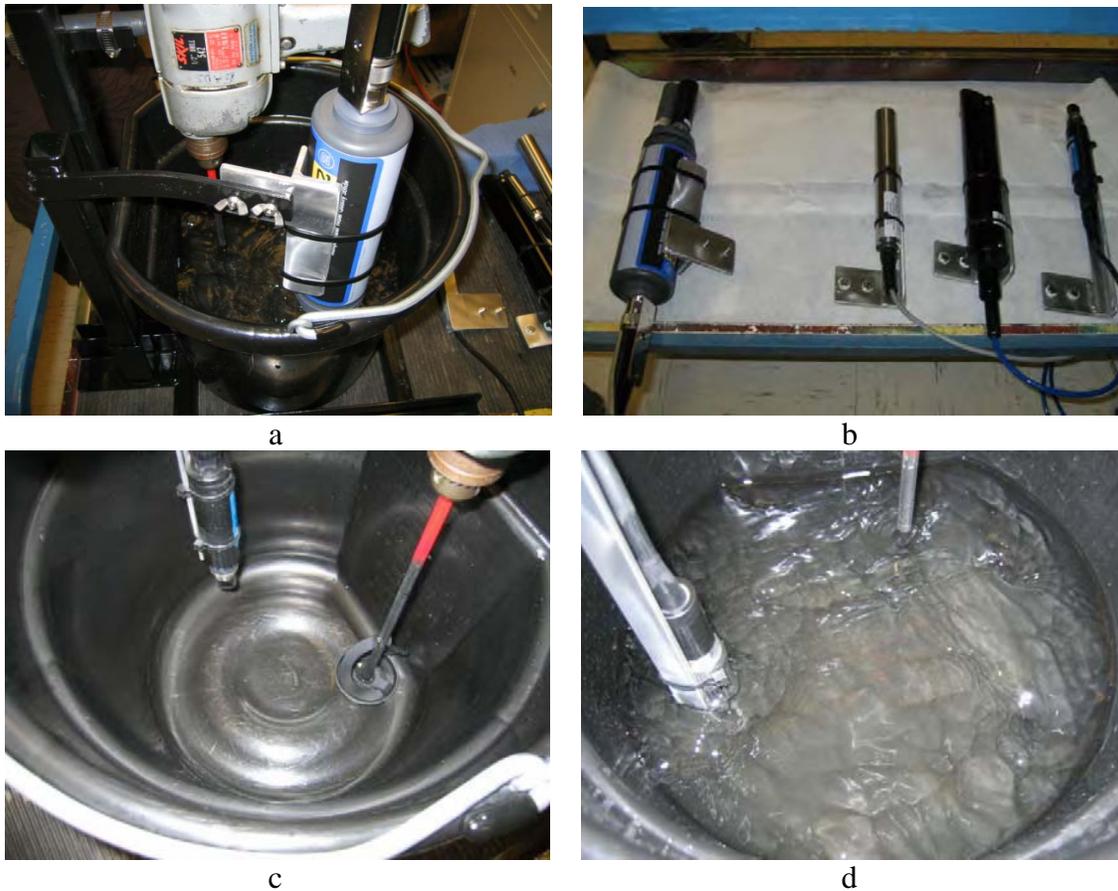


Figure 4. Sediment mixing apparatus. (a) Mixing bucket, drill and sensor stand, (b) sensors and mounting brackets, (c) locations of mixing paddle and sensor, (d) measuring in the mixed sediment suspension.

highest target to 1200 turbidity units.

All sensors were connected to a Campbell CR10X data logger. After each target reading was reached, the sensors were placed in the mixing bucket, one at a time. As soon as three readings were obtained from a sensor, it was removed and the next sensor was mounted. The last sensor to be read was always the OBS-3, so that the turbidity at start and end of each run could be compared. There were no indications of a decline in turbidity during any of the runs, indicating that steady-state suspensions were achieved.

Between readings, depth-integrated subsamples were extracted from each mixture for measurement by the Hach 2100P and 2100AN meters (Figure 5c). With signal averaging off on both meters, each sample was agitated by inverting it three times, the sample was placed in the meter, and the first displayed reading was recorded (Figure 5d). The procedure was repeated three times for each meter. Subsamples were also collected from the mixing bucket for determination of suspended sediment concentration (SSC) at the 25, 200, and 1200 turbidity targets. SSC was determined gravimetrically following vacuum filtration through one-micron glass membrane filters. SSC and organic content are shown in Table 3. The three replicate measurements recorded for each turbidity

sensor or meter were averaged and the calibration corrections were applied before subsequent analysis.



Figure 5. (a) wet-sieving to 0.5 mm, (b) adding sediment from the slurry to the mixing bucket, (c) subsampling from the mixing bucket for (d) Hach measurements.

YSI Corrections

A temperature sensor must be mounted with the YSI turbidity sensors in a multi-parameter sonde for YSI turbidity readings to be properly temperature-compensated. We discovered this procedural requirement only after runs had been completed without a temperature sensor on five sediments (GAR2, LMC1, LMC2, NFC1, and SFM3).

Therefore, a temperature sensor was installed, the YSI sensors were recalibrated in formazin and, after all the other sediments were completed, a second run was performed

Table 3. *Suspended sediment concentration (SSC) and organic content of samples at nominal levels of 20, 200, and 1200 turbidity units, as measured by the OBS-3.*

Sample	SSC (mg/L) @20	SSC (mg/L) @200	SSC (mg/L) @1200	Organics (%)
GAR2	110	1244	7459	4.5
GAR4	108	1093	9271	2.7
HHB1	124	958	6932	4.5
HHB2	139	1614	11280	3.2
HHB3	95	671	4198	3.9
HHB4	73	700	4403	5.5
JBW1	134	1119	6602	3.8
JBW2	155	1281	11750	3.3
KRW2	63	568	3716	4.6
LDC1	86	1386	7665	6.1
LLM1	43	431	2722	5.5
LLM2	75	978	6475	4.8
LMC1	48	556	4876	4.5
LMC2	76	825	5373	4.2
MBU1	62	499	3189	3.4
NAV2	33	299	1907	6.6
NAV3	21	259	1985	5.0
NFC1	30	463	2818	4.1
NFC2	130	663	7201	3.6
NFC3	131	1255	8417	2.5
PRU1	84	726	6084	17.1
SFM1	114	912	8073	3.8
SFM2	76	735	5982	3.1
SFM3	57	466	4042	3.6

on the five sediments that had been measured without the YSI temperature sensor. Readings were taken only with the OBS-3 and the two YSI sensors. After averaging the three replicate measurements and applying the calibration corrections, a small adjustment was made to each value to correct for the fact that the second mix did not have precisely the same turbidity as the first mix when all the other sensors were read. YSI turbidity corresponding to OBS-3 readings recorded in the initial mix were interpolated from cubic splines relating OBS-3 and YSI turbidity in the second mix (Figure 6).

Table 4. Particle size distributions below 500 μm , as determined by laser diffraction using a Micromeritics Saturn Digisizer 5200.

Sample	% by volume in size class with specified upper boundary					
	500 μm	250 μm	125 μm	62.5 μm	15 μm	3 μm
GAR2	10.0	17.2	24.9	30.5	10.8	6.6
GAR4	1.8	20.2	30.4	30.5	10.2	6.9
HHB1	0.4	8.9	18.6	34.6	25.4	12.0
HHB2	0.3	14.1	30.1	31.5	15.8	8.3
HHB3	1.6	12.1	15.2	26.1	27.2	17.7
HHB4	1.3	6.3	14.5	32.8	30.3	14.9
JBW1	3.8	14.0	19.2	28.9	21.7	12.5
JBW2	5.8	26.2	27.2	21.1	12.4	7.4
KRW2	0.0	0.4	3.6	38.8	38.0	19.1
LDC1	1.4	20.3	26.9	29.9	14.8	6.7
LLM1	1.9	7.7	7.9	26.0	34.0	22.5
LLM2	1.9	7.1	10.8	31.2	32.5	16.6
LMC1	4.5	11.1	13.9	27.8	28.6	14.1
LMC2	13.3	28.4	19.5	17.8	13.8	7.2
MBU1	1.3	5.7	6.0	20.9	38.5	27.7
NAV2	0.0	0.8	2.7	15.6	47.8	33.1
NAV3	0.1	2.5	6.3	19.9	36.5	34.6
NFC1	6.5	10.5	11.3	23.6	26.4	21.7
NFC2	8.3	17.2	17.3	24.4	20.6	12.2
NFC3	26.9	29.1	22.2	12.2	4.9	4.7
PRU1	7.9	24.1	25.1	25.3	12.5	5.1
SFM1	1.1	8.7	22.1	40.2	18.4	9.4
SFM2	0.0	0.0	2.4	40.3	38.8	18.6
SFM3	0.0	0.3	7.6	37.3	35.4	19.4

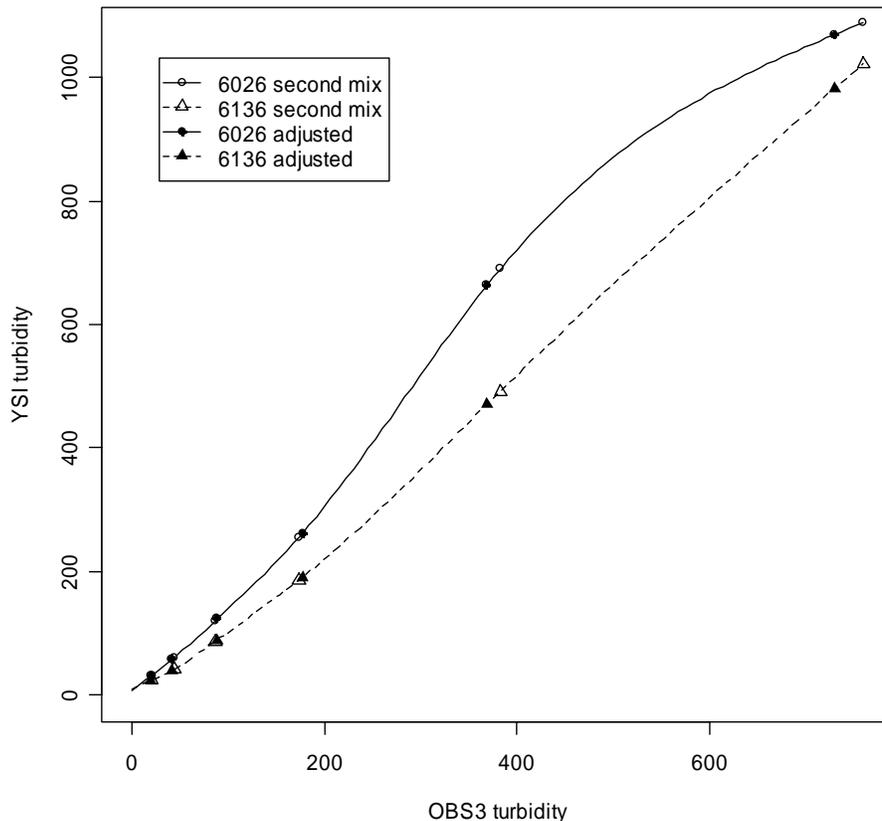


Figure 6. YSI turbidity in second mix was converted to equivalent turbidity in first mix (shown as "adjusted" in legend) by interpolation from the relation between YSI and OBS-3 turbidities, illustrated for sample LMC2.

RESULTS

Figure 7 shows that, for all sensors, the relationship of turbidity to SSC depends strongly on the sediment. This follows naturally from the fact that turbidity is highly dependent on particle size distribution, as well as other sediment characteristics. In each frame of the figure, the NAV2 and NAV3 sediments have the highest turbidity for a given SSC, while HHB2 and JBW2 are among the lowest. Both NAV2 and NAV3 have less than 10% sand, while HHB2 and JBW2 have less than 10% clay.

Figure 8 shows that the sensors may vary by roughly a factor of two in the turbidity reported for a given sample. The largest ratio between any two sensor readings was 3.0 for sample SFM2 at a concentration of 5980 mg/L, in which the Hach 2100AN read 3103 NTU while the OBS-3+ read 1021 FBU. Such differences are not unexpected because the sensors have different measuring characteristics, including wavelength, scattering angles measured, number of detectors, aperture angles of the cones of light emitted and detected, and volume measured. In most cases the backscatter sensors, OBS-3 and OBS-3+ report the lowest turbidity values, while the NEP395 and YSI 6026 report the

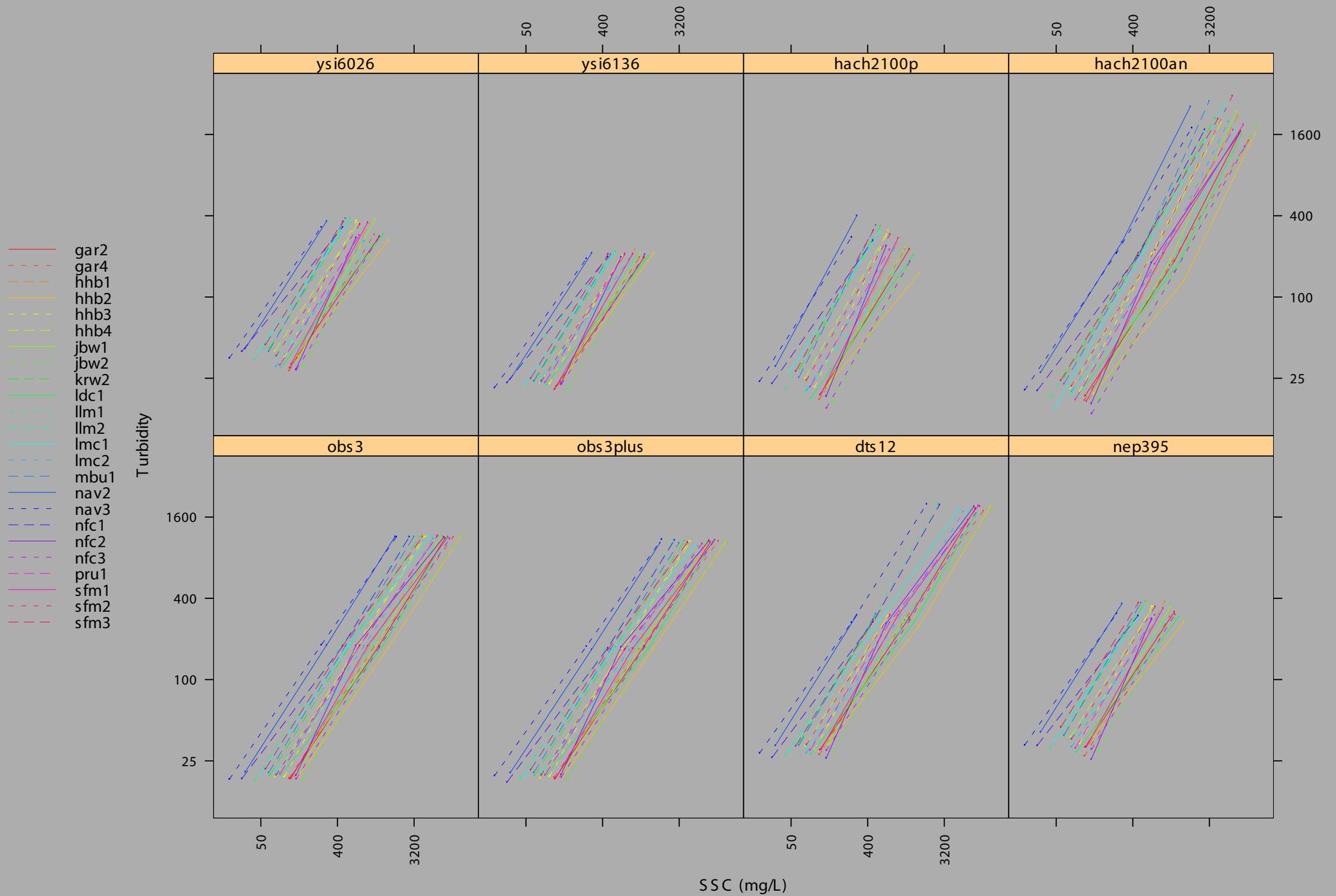


Figure 7. Turbidity is plotted as a function of SSC for each sediment by sensor. Each sediment was sampled at nominal turbidity levels of 25, 200, and 1200 units as measured by the OBS-3 sensor. The YSI 6026, YSI 6136, NEP395, Hach 2100P, and some of the DTS-12 readings are off-scale at the 1200 level and are not shown.

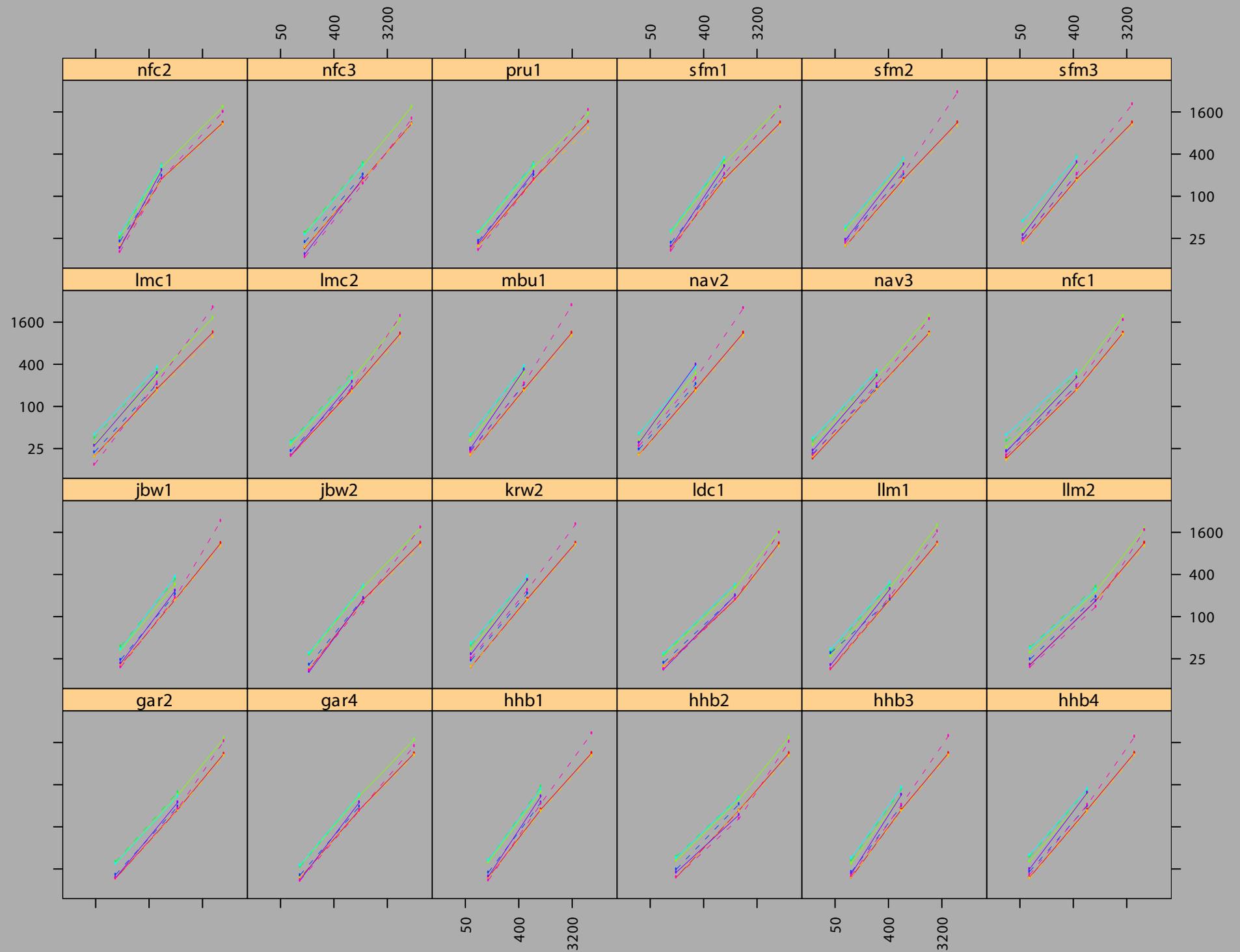


Figure 8. Turbidity is plotted as a function of SSC for each sensor by sediment. Each sediment was sampled at nominal turbidity levels of 25, 200, and 1200 units as measured by the OBS-3 sensor. The YSI 6026, YSI 6136, NEP395, Hach 2100P, and some of the DTS-12 readings are off-scale at the 1200 level and are not shown.

highest. At higher SSCs, above the operating range of the NEP395 and YSI 6026, the DTS-12 and the Hach 2100AN report the highest turbidity.

Figures 9 and 10 show the relationships among all pairs of sensors for each sediment type. Figure 9 shows the complete data set for each sensor pair, including out-of-range data. The x scale in each column is fixed to the range of data in that column and the y scale in each row is fixed to the range of data in that row, but x scales vary among rows and y scales vary among columns. Therefore, the line of perfect agreement ($y=x$) appears to have different slopes because of the variation in x and y scales. There are no Hach 2100P data at nominal turbidity values of 800 or 1200 for sediments other than NFC1 because the meter reported out-of-range data for these mixtures. The NEP395 and both YSI sensors reached a plateau at or near their maximum values in samples at a nominal turbidity of 800. That is, when the OBS-3 was reading 800 or above these sensors were generally reading above 1000 and were beyond their ranges of sensitivity.

In Figure 10, out-of-range data have been eliminated, and all x and y scales have been fixed with limits of 0 to 1250 turbidity units, so sensor ranges and differences in slope can be more easily compared. The line of perfect agreement ($y=x$) is again shown on each plot for reference. There is considerable variation among sediment samples for most sensor pairings. The curves all tend to diverge as turbidity increases. Absolute differences in turbidity are greatest at high turbidity values. Divergence appears to be smaller for certain pairings, e.g. OBS-3/OBS-3+ and NEP395/YSI 6026. Magnification is necessary to readily perceive the variability among curves for those pairings with a limited range, e.g. OBS-3+/YSI 6026.

We can quantify the errors that might occur if sensor readings were standardized based on a unique relationship for any sensor pairing. By looking at the deviations of data points from an assumed relationship, we can express the errors as a percentage of the predicted value for any given nominal turbidity. However, the errors are dependent upon the form of the assumed relationship, which could take many forms.

We initially assume a simple ratio can be applied to convert one sensor's readings to equivalent values of another. Since all the curves appear to approach the origin and most are linear, a ratio model appears reasonable at first glance for most of the scatterplots in Figure 10. Many different estimators can be used to estimate a ratio. The three simplest are linear regression without an intercept, ratio of means, and mean of ratios. If errors are normally distributed for a given x , then each of these is the best linear unbiased estimator (BLUE) under specific circumstances. Linear regression with no intercept is BLUE when the error variance is independent of x . The ratio of means estimator, $\sum y_i / \sum x_i$, is BLUE when the error variance is proportional to x . The mean of ratios estimator, $(\sum y_i / x_i) / n$, is BLUE when the error variance is proportional to x^2 . Since the relationships diverge linearly in Figure 10, the error variance should be roughly proportional to x^2 , implying the mean of ratios estimator is the best choice among these three estimators. This estimator is the equivalent of a least squares regression estimator with observations weighted inversely proportional to x^2 . Therefore, large observations receive less weight than small observations and, for relationships that are nonlinear, the

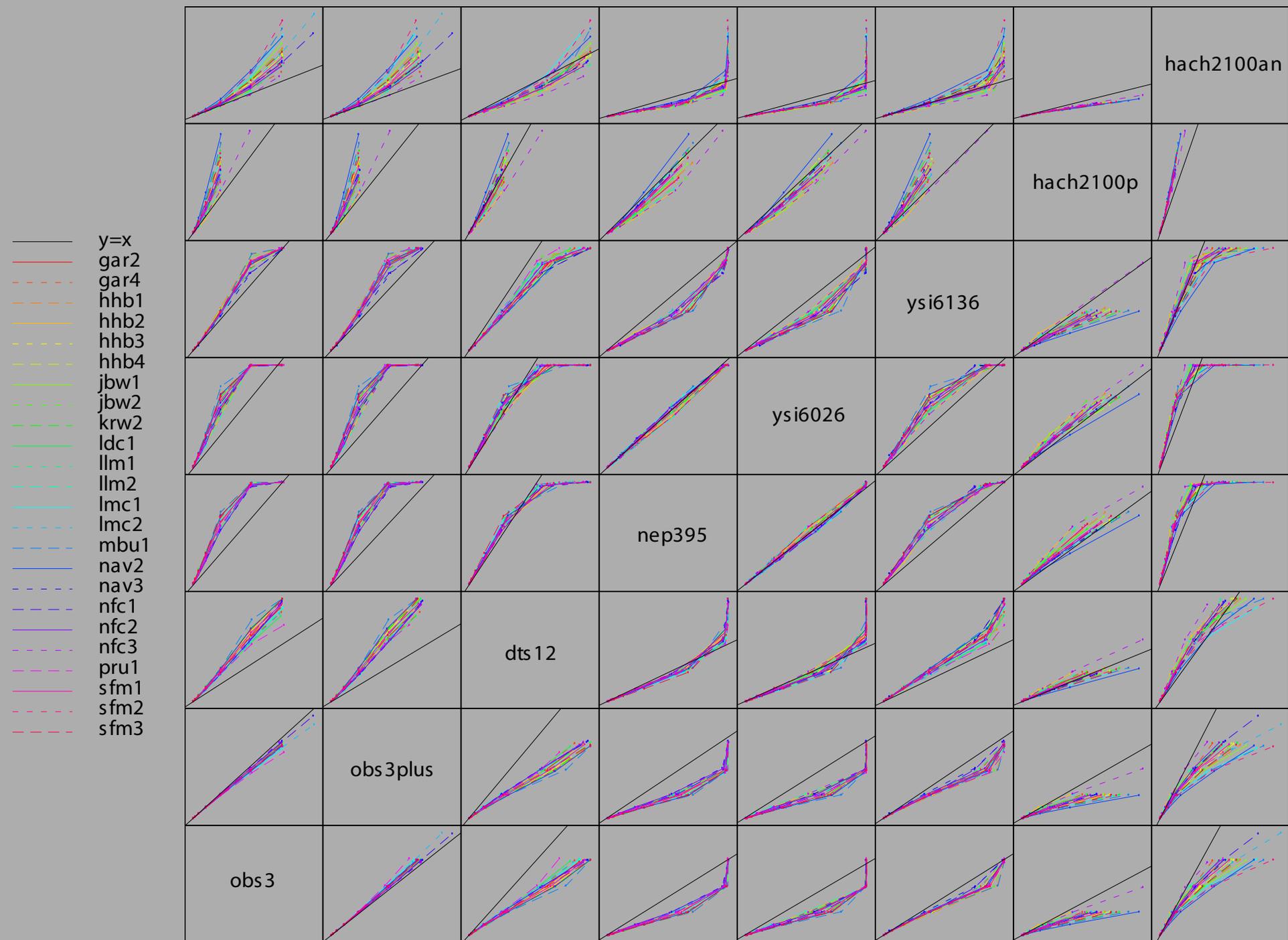


Figure 9. Scatterplot matrix, showing the relationships among turbidity readings for all pairs of sensors. Each line represents one sediment type. Out-of-range data have *not* been omitted, except for the Hach 2100P. The x variable for a given plot is the sensor whose name appears on the diagonal in the same column. The y variable for a given plot is the sensor whose name appears on the diagonal in the same row. The x and y scales are set by the range of data for the sensors in the corresponding column and row.

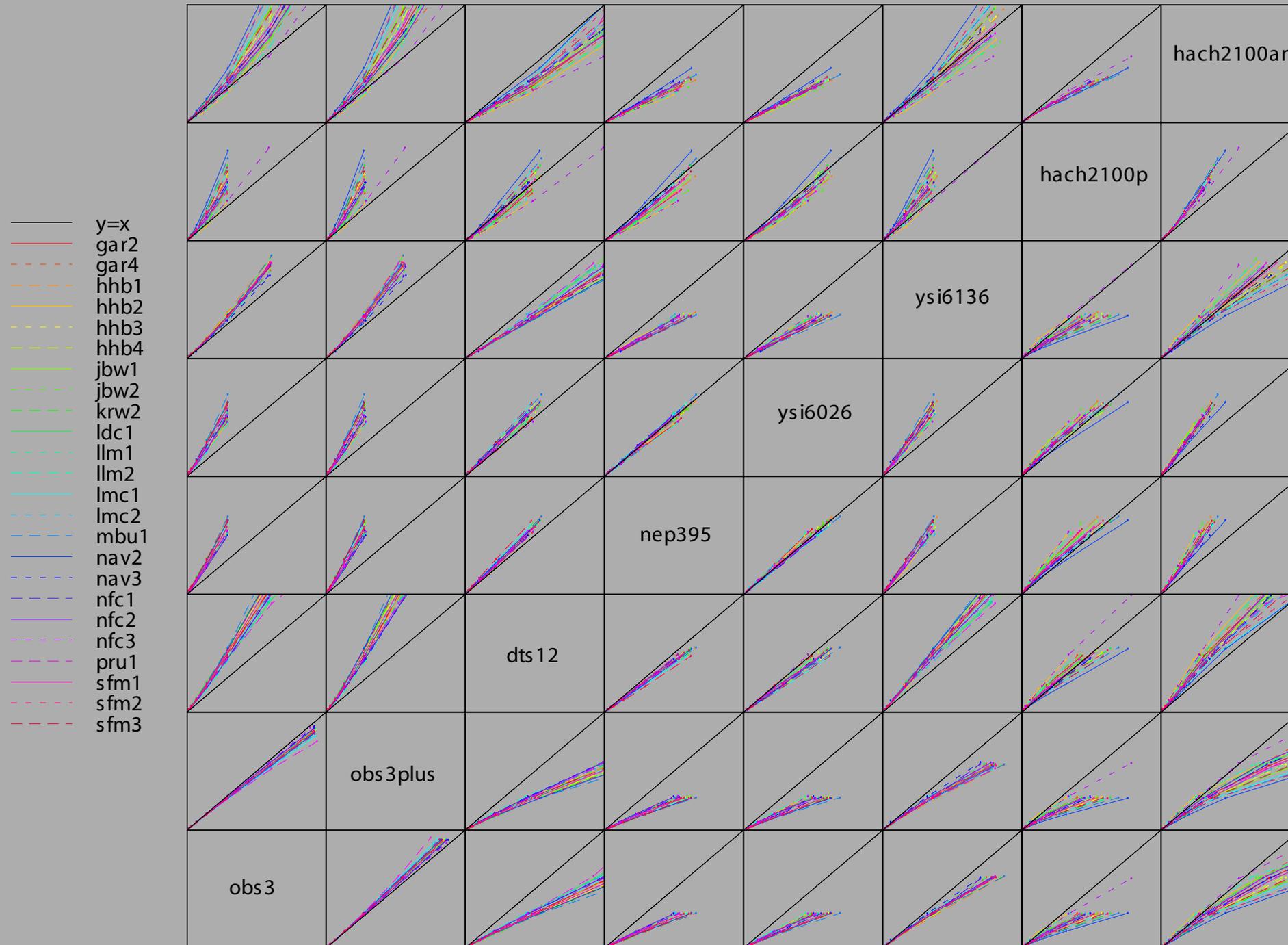


Figure 10. Scatterplot matrix, showing the relationships among turbidity readings for all pairs of sensors. This figure is similar to Figure 9 except that out-of-range data have been omitted and the ranges of all x and y scales are fixed at 0 to 1250. As a result, 29 DTS-12 points and 28 Hach 2100AN points above 1250 are not displayed.

mean of ratios is most accurate at low turbidity levels. Figure 11 shows the mean and maximum errors. Errors are expressed as the absolute value of the percentage of the measured value, for each sensor pairing at each nominal turbidity level. Each sensor pairing appears twice to allow for prediction in both directions. For each graph, the predictor sensor name is in the same column and the predicted sensor name is in the same row. The number in each graph is the estimated ratio. While the mean error is generally small, averaging 12% across all sensor pairings, the maximum error often exceeds 50% and occasionally even exceeds 100%. The mean of the maximum error across all sensor pairings is 54%. These large errors occur partly because of lack of fit and partly due to high variance.

Errors due to lack of fit can be reduced by choosing better prediction models. We considered six types of models.

1. *ratio*. Ratio estimated as the mean ratio of y to x .
2. *lme*. Ratio estimated using a linear mixed-effects model in which the error is assumed proportional to a power of x , and the power of x is optimized using a restricted maximum likelihood method (Pinheiro and Bates, 2000).
3. *lm*. Simple linear regression model solved by unweighted least squares.
4. *quad*. Quadratic regression model solved by unweighted least squares.
5. *logxy*. Power function solved by unweighted least squares on log-transformed data. For this analysis we apply the Baskerville (1972) bias correction when back-transforming predictions. The anti-log of the prediction is multiplied by $\exp(0.5s^2)$, where s is the residual standard error from the regression.
6. *loess*. Locally weighted non-parametric regression (Cleveland, 1993). A loess model with appropriate smoothing parameters is assured of passing near the mean of y for any given x . This method is the most flexible but requires the coordinates of the curve for prediction. We do not intend for practitioners to use loess for prediction, but include it as a "gold standard", indicative of the best performance that might be obtained for a parametric model with 1 or 2 parameters.

Figure 12 shows the six types of models fitted to the problem of converting DTS-12 data to Hach 2100AN data. Each cluster of data roughly represents a nominal turbidity level. The 800 and 1200 levels run into one another in the upper right portion of the graph, as do the 25 and 50 levels in the lower left. The linear model (*lm*) misses most of the points at the 25, 50 and 400 levels, while the *ratio* and *lme* models under predict most of the data at the 800 and 1200 levels. Therefore, a non-linear model is needed. The quadratic (*quad*) and *loess* model are quite similar except above 1500 FNU. The *loess* model has the lowest mean error (11.0%), followed by *logxy* (11.7%) and *quad* (13.9%). The maximum error is minimized in this case by *logxy* (48%).

The best model for each predictor and response sensor was chosen using the following procedure.

1. Eliminate all out-of-range data (i.e. NEP395 and YSI 6026 data above 1000 FNU, YSI 6136 data above 1100 FNU, DTS-12 data at 2046 FNU, and Hach 2100P data above 1000 NTU). This eliminated all YSI 6026, NEP395 and Hach 2100P

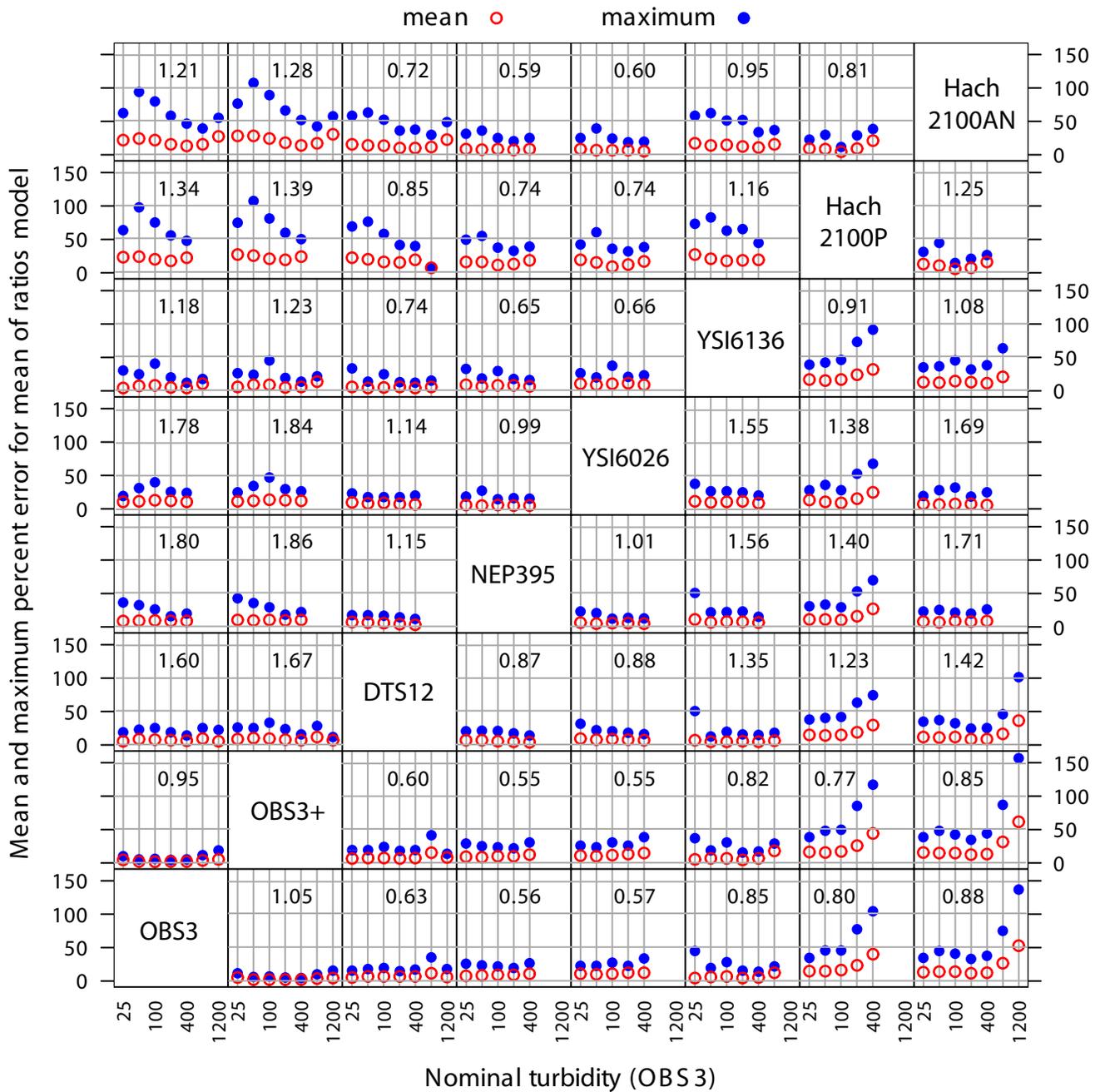


Figure 11. Mean and maximum errors for ratio model relating all pairs of sensors. The number displayed in each graph is the estimated ratio, based on the mean of ratios model. For each graph, the predictor is the sensor whose name is on the diagonal in the same column and predicted value is for the sensor whose name is on the diagonal in the same row.

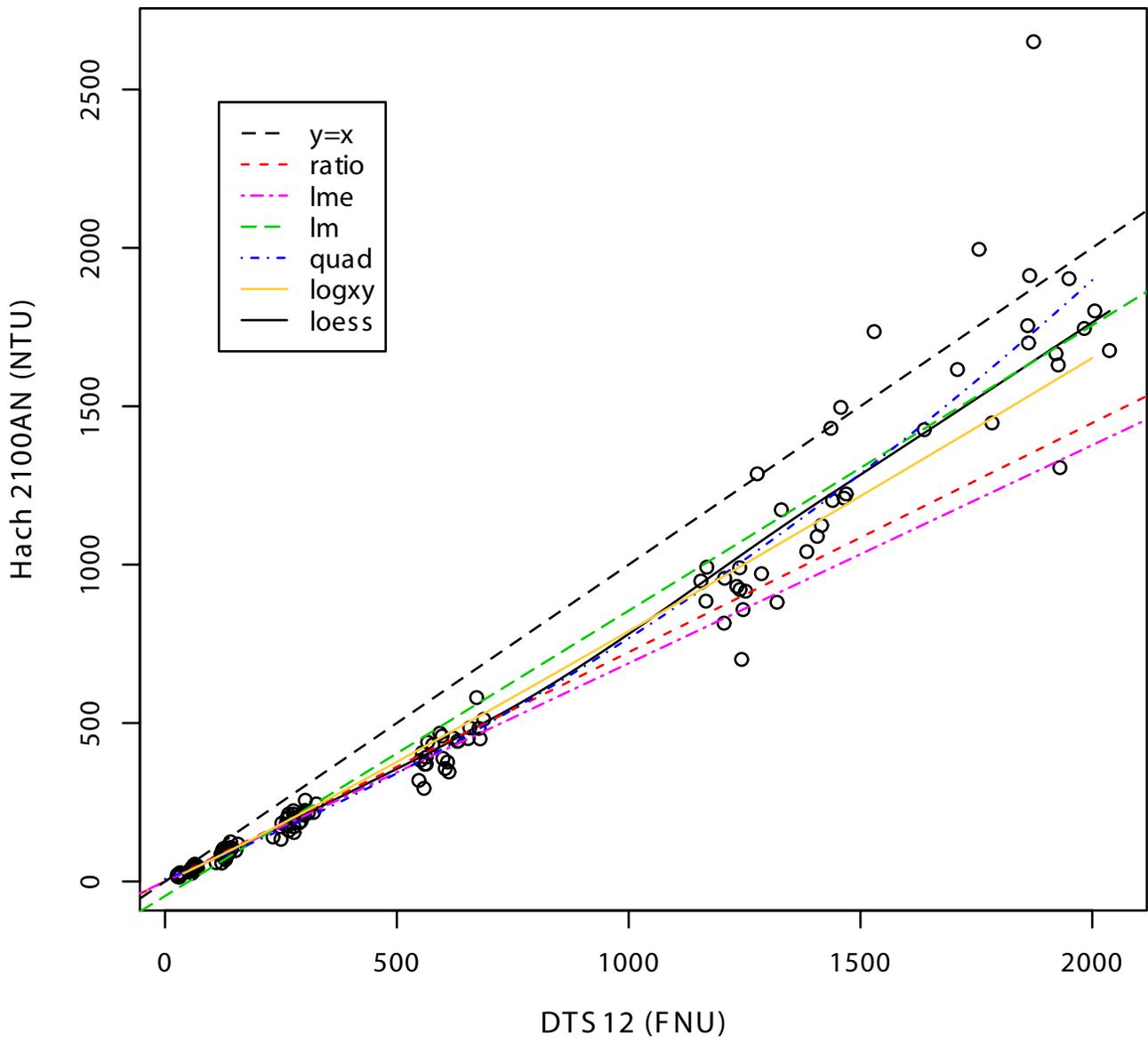


Figure 12. Six models for converting DTS-12 data to Hach 2100AN. Data for nominal turbidity levels from 25 to 1200 are included for all 24 sediment types.

- data at nominal turbidity levels of 800 and 1200; and YSI 6136 data at the nominal turbidity level of 1200.
2. Fit all six models to the remaining data.
 3. Compute the absolute value of the difference between the predicted and measured value for each model, and express it as a percentage of the measured value. Denote this error quantity $|e|$.
 4. Find the mean of $|e|$ for each sensor pairing at each nominal turbidity value.
 5. Compute the mean (over all nominal turbidity levels) of the mean errors for each sensor pairing. Denote this quantity $\overline{|e|}$.
 6. Compute the maximum of $|e|$, over all nominal turbidity levels, for each sensor pairing. Denote this quantity $|e|_{\max}$.
 7. Select the "mini-mean" model, i.e. the model that minimizes $\overline{|e|}$, the grand mean of the percentage errors $|e|$.
 8. Compare the mini-mean model to the "mini-max" model, which minimizes $|e|_{\max}$.

Table 5 shows the mean of the two criteria $\overline{|e|}$ and $|e|_{\max}$ for each model and the mean ranking of each model, considering all sensor pairings. With all four criteria, the *loess* model performs best, followed by the *logxy* and *quad* models. For individual sensor pairings, the rankings vary, but in general, *loess* performs the best (as expected) and *logxy* is the best of the parametric models. According to the continuous criteria ($\overline{|e|}$ and $|e|_{\max}$), *logxy* performs nearly as well as *loess*, while the gap between *logxy* and the other models is more substantial.

Table 5. Comparison of model types.

Criterion	<i>ratio</i>	<i>lme</i>	<i>lm</i>	<i>quad</i>	<i>logxy</i>	<i>loess</i>
mean of $\overline{ e }$	11.8	11.8	22.4	10.8	9.4	9.1
mean of $ e _{\max}$	54.2	52.0	98.6	47.2	40.2	38.4
mean rank of $\overline{ e }$	4.30	4.25	5.25	3.36	2.37	1.46
mean rank of $ e _{\max}$	3.98	3.62	5.30	3.50	2.57	2.02

In comparing the mini-mean and mini-max models, we find that for 29 of 56 pairings, the same model is selected by both criteria. For the remaining pairings, $\overline{|e|}$ averages 0.82% higher (absolute difference) for the mini-max model than for the mini-mean model, while $|e|_{\max}$ averages 3.59% higher (absolute difference) for the mini-mean model than for the mini-max model. In other words, neither the choice nor the performance of the model is very sensitive to the selection criterion. Although *loess* ranks highest, it is not the best of the six models for every sensor pairing; otherwise, its mean rank would be 1.00.

Table 6 shows the model selected for each pairing of predictor and response sensor. Loess models were excluded from consideration because of the awkwardness of applying

Table 6. Best model and associated errors for all turbidity sensor pairings. Best model was selected by minimizing the average percent error $\overline{|e|}$. Corresponding errors from loess model are shown for comparison.

Predictor	Response	Model	b ₀	b ₁	b ₂	$\overline{ e }$	$ e _{\max}$	loess $\overline{ e }$	loess $ e _{\max}$
OBS-3	OBS-3+	logxy	1.0635	0.9789		2.0	14.0	1.9	12.9
OBS-3	DTS-12	logxy	1.3728	1.0302		6.5	30.2	6.5	28.7
OBS-3	NEP395	quad	1.0057	1.6794	0.000700	8.6	32.5	8.5	32.6
OBS-3	YSI 6026	quad	0.4637	1.6803	0.000620	10.9	36.6	10.9	33.6
OBS-3	YSI 6136	quad	-0.2307	1.1022	0.000295	4.8	35.7	4.0	32.9
OBS-3	2100P	quad	-2.2871	1.1576	0.001614	17.0	72.2	17.0	75.1
OBS-3	2100AN	logxy	0.6521	1.1176		15.0	63.7	14.4	64.2
OBS-3+	OBS-3	logxy	0.9416	1.0212		2.0	12.5	1.9	11.7
OBS-3+	DTS-12	logxy	1.2917	1.0517		6.5	29.7	6.2	27.1
OBS-3+	NEP395	quad	-1.8398	1.7843	0.000743	9.5	31.2	9.3	35.2
OBS-3+	YSI 6026	quad	-3.7934	1.8203	0.000553	11.8	45.8	11.5	40.4
OBS-3+	YSI 6136	quad	-2.0520	1.1528	0.000405	6.2	37.4	4.9	37.0
OBS-3+	2100P	logxy	0.6690	1.1626		18.3	83.0	17.8	82.8
OBS-3+	2100AN	quad	-0.3120	0.9640	0.000879	14.8	61.5	15.4	74.2
DTS-12	OBS-3	logxy	0.7537	0.9673		6.3	27.0	6.8	27.3
DTS-12	OBS-3+	logxy	0.8039	0.9475		6.2	28.1	6.6	29.2
DTS-12	NEP395	lm	-1.6869	1.1664		5.3	19.2	5.3	18.2
DTS-12	YSI 6026	lm	-0.9514	1.1486		8.0	25.4	7.9	22.8
DTS-12	YSI 6136	lme		0.7421		5.3	33.5	5.2	30.7
DTS-12	2100P	lme		0.7550		15.9	56.1	18.0	60.3
DTS-12	2100AN	logxy	0.5048	1.0648		11.7	48.1	11.0	51.3
NEP395	OBS-3	logxy	0.6946	0.9582		8.6	21.8	8.3	21.2
NEP395	OBS-3+	logxy	0.7326	0.9416		9.0	21.5	8.8	21.6
NEP395	DTS-12	logxy	0.9453	0.9845		5.3	22.9	5.2	22.8
NEP395	YSI 6026	quad	0.6884	0.9835	0.000003	5.3	27.0	5.2	26.3
NEP395	YSI 6136	logxy	0.7637	0.9674		7.3	29.6	7.0	27.7
NEP395	2100P	logxy	0.4180	1.1121		11.3	40.8	11.8	48.0
NEP395	2100AN	quad	-0.5610	0.5982	0.000002	8.0	36.1	8.1	36.2
YSI 6026	OBS-3	logxy	0.6820	0.9640		10.7	28.5	10.2	28.1
YSI 6026	OBS-3+	logxy	0.7207	0.9470		11.1	35.0	10.6	34.4
YSI 6026	DTS-12	logxy	0.9257	0.9908		8.0	33.5	7.8	33.5
YSI 6026	NEP395	logxy	0.9794	1.0063		5.4	22.0	5.1	21.2
YSI 6026	YSI 6136	logxy	0.7470	0.9738		9.5	37.2	9.1	33.7
YSI 6026	2100P	logxy	0.3959	1.1241		8.4	41.7	8.9	44.1
YSI 6026	2100AN	lm	-1.0338	0.6086		6.6	38.8	6.4	37.1
YSI 6136	OBS-3	quad	-0.3548	0.9052	-0.000154	4.3	51.4	4.1	49.2
YSI 6136	OBS-3+	quad	-0.1163	0.8775	-0.000186	4.7	45.2	4.7	44.6
YSI 6136	DTS-12	quad	-1.8351	1.3965	-0.000061	5.4	49.0	5.4	48.3
YSI 6136	NEP395	lm	-2.3246	1.5986		7.5	47.6	7.5	46.7
YSI 6136	YSI 6026	logxy	1.4423	1.0152		9.9	34.7	10.0	35.0

YSI 6136	2100P	logxy	0.6130	1.1370		16.7	81.7	16.7	84.4
YSI 6136	2100AN	logxy	0.7423	1.0487		13.1	54.8	13.1	58.8
2100P	OBS-3	quad	-0.6040	0.8888	-0.000510	13.7	51.9	14.1	43.8
2100P	OBS-3+	quad	-0.0345	0.8587	-0.000511	14.2	56.2	14.6	41.5
2100P	DTS-12	logxy	2.2233	0.8716		12.4	38.0	12.1	38.0
2100P	NEP395	logxy	2.3597	0.8871		9.9	32.2	9.8	29.6
2100P	YSI 6026	logxy	2.3934	0.8817		7.5	29.5	7.7	28.8
2100P	YSI 6136	logxy	1.7804	0.8554		13.9	42.1	13.8	41.9
2100P	2100AN	quad	3.3057	0.8219	-0.000253	4.6	21.2	4.5	23.1
2100AN	OBS-3	quad	3.8041	0.9396	-0.000186	11.3	43.9	12.6	45.5
2100AN	OBS-3+	quad	7.0692	0.8754	-0.000180	13.2	51.9	13.3	48.6
2100AN	DTS-12	logxy	2.0323	0.9300		11.1	65.6	10.1	52.4
2100AN	NEP395	lme		1.6996		7.7	26.2	7.6	25.1
2100AN	YSI 6026	logxy	1.9212	0.9719		6.5	31.1	6.3	28.1
2100AN	YSI 6136	logxy	1.4540	0.9402		12.2	47.9	11.6	41.2
2100AN	2100P	quad	0.6413	1.0944	0.000978	4.6	33.5	5.0	33.5
Grand mean						9.13	39.19	9.07	38.36

them for prediction. The table shows the coefficients of the selected models, so this table can be used for prediction. For the *logxy* model, b_0 represents the bias-corrected constant and b_1 represents the exponent of the model expressed as a power function (i.e. predictions are calculated as $y = b_0 x^{b_1}$). In general, both the mean and maximum criteria are close to their "optima" as expressed by *loess*, indicating that lack of fit is unimportant and most of the error is due to variation in y for a given x . Figure 13 shows the mean and maximum value of $|e|$ for the Table 6 models at each nominal turbidity level. The mean error tends not to vary much with turbidity and never exceeds 25%. The maximum error is more variable, and for some sensor pairings, notably when conversions are being made to Hach equivalents, errors of 50% or more are common. The magnitude of error does not seem to vary in a systematic way with turbidity, and it depends strongly on the particular sensor pairings and types of sensors. For example, conversions between DTS-12 and NEP395 can be made in either direction with average errors of 5.3% and maximum error no greater than 23%. Conversions within a group of ISO 7027 sensors (NEP395, YSI 6026, and YSI 6136), backscatter sensors (OBS-3, OBS-3+), or EPA 180.1 meters (Hach 2100P, Hach 2100AN) can in general be made with less error than conversions between groups. Conversions between the two OBS sensors, made by the same manufacturer, can be made in either direction with mean error of 2.0% and maximum error no greater than 14%. Conversions between the two Hach meters can be made with mean errors of 4.6% in either direction and maximum error of 33%. Conversions between the two YSI sensors can be made with mean errors of about 10% and maximum error of 37%.

DISCUSSION AND CONCLUSIONS

These results confirm the well-known fact that raw turbidity readings from different sensors and meters are not comparable. Readings of the same sediment mixture by

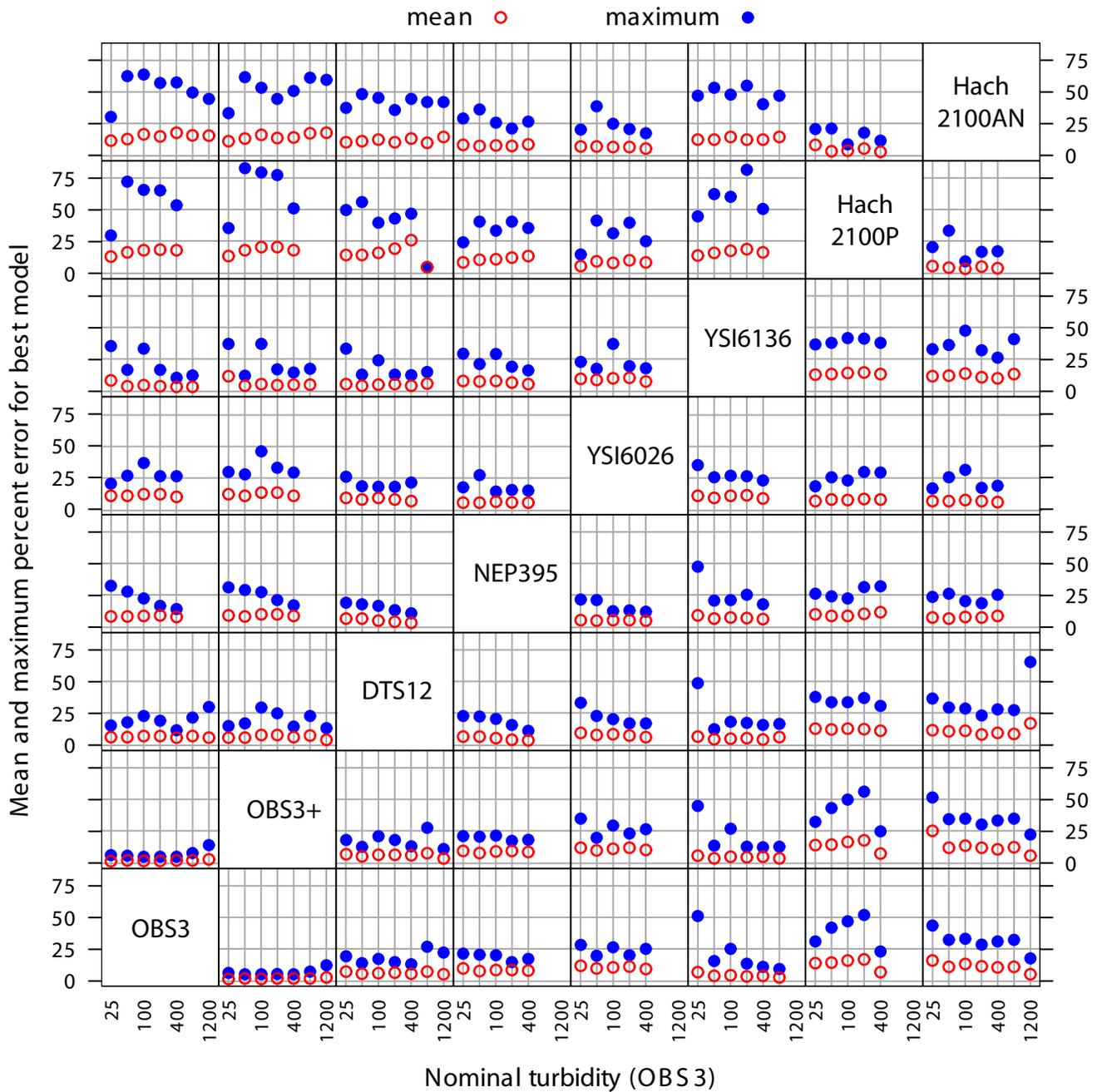


Figure 13. Mean and maximum errors for the best model (see Table 6) relating each pair of sensors. For each graph, the predictor is the sensor whose name appears on the diagonal in the same column and the predicted turbidity is for the sensor whose name appears on the diagonal in the same row.

different sensors differed by up to a factor of three. The problem of converting values between instruments is not trivial because the conversion depends on the sediment being measured, which of course varies in space and time. To minimize the error of converting between instruments, a representative set of samples would need to be measured by both sensors at all locations during the time period being compared. Since that approach is impractical for most monitoring, the next best approach would be to convert using average relationships such as those shown in Table 6. The errors associated with such conversions may be small or large and depend on the particular sensors and sediments being considered. Given the variability shown in Figures 9 and 10, the dilemma presented in Figure 1 can be real for some sensor pairings. More specifically, turbidity rankings may not be consistent among sensors unless data differ by at least a factor of two after standardization to a common instrument.

Even sensors that conform to the same standards (backscatter, EPA Method 180.1 or ISO 7027) do not necessarily give similar turbidity readings. However, relationships between sensors of the same design were more consistent, for different sediments, than relationships between sensors that used different methods. Conversions between sensors in the same grouping can be made with less error. Conversion of *in situ* sensor readings to laboratory readings is thus prone to relatively large errors unless the laboratory meter is set to use the same method as the *in situ* meter. For example, the Hach 2100AN may be set to conform to EPA Method 180.1 or ISO 7027, by specifying white or infrared light, Ratio ON or OFF, and Backscatter OFF. In this study, we used only Ratio mode with white light. Our results suggest that different settings might have improved the consistency of the Hach 2100AN output relative to the backscatter and ISO 7027 sensors that we tested. However, that does not aid in comparisons of field sensors that conform to different standards. For example, direct conversions between an OBS-3 and DTS-12 sensor appear to be more reliable than converting both to Hach 2100AN equivalents. Conversions between the OBS-3 and DTS-12 can be made in either direction with mean error of about 6.5% and maximum error of 30%. Maximum errors are associated with textural extremes such as PRU1 (<5% clay) or KRW2 (<5% sand).

The maximum errors we report may be unusual for actual suspended sediments of our sampled watersheds, which usually represent a mix of sediments, rather than isolated samples from colluvial or alluvial deposits. While the samples gathered for this study represent a wide range of textures, they probably include a relatively limited range of mineralogies, being confined to sedimentary formations of the northern California coastal belt. Therefore, application of the Table 6 relationships may be subject to errors similar to the reported maxima or even greater, if applied in, say, granitic watersheds with coarse sediments or micaceous minerals. The average errors reported in Table 6 and Figure 13 are probably acceptable for many applications, but any given watershed may not be 'average' as defined by the samples we collected, so it will always be important to consider the characteristics of the sediments as well as the particular sensors involved. Further investigations may permit quantification of the role of sediment particle size distribution and organic composition in the variability of relationships between sensors.

References

- Anderson, C.W., 2004. Turbidity (version 2.0): U.S. Geological Survey Techniques of Water-Resources Investigations, book 9, chap. A6, section 6.7, 64 p.; accessed October 28, 2004, at <http://pubs.water.usgs.gov/twri9A6/>
- Baskerville, G.L. 1972. Use of logarithmic regression in the estimation of plant biomass. *Canadian Journal of Forestry* 2:49-53.
- Cleveland, W.S. 1993. *Visualizing Data*. Summit, New Jersey: Hobart Press. 360 pp.
- Davies-Colley, R.J. and D.G. Smith. 2001. Turbidity, suspended sediment, and water clarity: A review. *Journal of the American Water Resources Association* 37: 1085-1101.
- Foster, I.D.L., R. Millington, and R.G. Grew. 1992. The impact of particle size controls on stream turbidity measurement; some implications for suspended sediment yield estimation. In: *Erosion and Sediment Transport Monitoring Programmes in River Basins* (ed. by J. Bogen, D.E. Walling & T.J. Day). Proc. Oslo Symp., August 1992, 51-62. IAHS Publ. 210.
- Gray, J.R., and G.D. Glysson, 2003. Proceedings of the Federal Interagency Workshop on Turbidity and Other Sediment Surrogates, April 30–May 2, 2002, Reno, Nevada: U.S. Geological Survey Circular 1250, 56 p.; accessed May 27, 2005, at <http://pubs.water.usgs.gov/circ1250/>
- Harris, R.R., K. Sullivan, P.H. Cafferata, J.R. Munn, and K.M. Faucher, in press. Applications of turbidity monitoring to forest management in California. *Environmental Management*.
- Pinheiro, J.C., and D.M. Bates, 2000. *Mixed-effects models in S and S-Plus*. New York: Springer-Verlag. 528 pp.