Environmental Science & Technology

ARTICLE

pubs.acs.org/est

Quantifying Remediation Effectiveness under Variable External Forcing Using Contaminant Rating Curves

James W. Kirchner,*[,†,‡] Carrie M. Austin,[§] Alexandra Myers,[§,∥] and Dyan C. Whyte[§]

[†]Department of Earth and Planetary Science, University of California, Berkeley, California 94720-4767, United States

[‡]Swiss Federal Research Institute WSL, CH-8903 Birmensdorf, Switzerland and Department of Environmental Sciences, Swiss Federal Institute of Technology (ETH), CH-8092 Zürich, Switzerland

[§]California Environmental Protection Agency, San Francisco Bay Regional Water Quality Control Board, 1515 Clay Street, Suite 1400, Oakland, California 94612-1482, United States

[∥]Department of Geography, University of California, Berkeley, California 94720-4767, United States

S Supporting Information

ABSTRACT: Remediation efforts are typically assessed through before-and-after comparisons of contaminant concentrations or loads. These comparisons can be misleading when external drivers, such as weather conditions, differ between the pre- and postremediation monitoring periods. Here, we show that remediation effectiveness may be better assessed by comparing pre- and postremediation contaminant rating curves, which permit "all else equal" comparisons of pre- and postremediation contaminant concentrations and loads under at any specified external forcing. We illustrate this approach with a remediation case study at an abandoned mercury mine in Northern California. Measured mercury loads in the stream draining the mine site were a factor of 1000 smaller after the remediation than before, superficially suggesting that the cleanup was 99.9% effective, but rainstorms were weaker and less frequent during the postremediation monitoring period. Our analysis shows that this difference in weather conditions alone reduced mercury loads at our site by a factor of 73−85, with a further factor of 12.6−14.5 being attributable to the remediation itself, implying that the cleanup was 92−93% (rather than 99.9%) effective. Our results illustrate the need to account for external confounding drivers when assessing remediation efforts, particularly in systems with highly episodic forcing.

INTRODUCTION

Abandoned mines are significant sources of water pollution throughout the world.[1] On U.S. federal lands alone, thousands of kilometers of streams are threatened by pollution from tens of thousands of abandoned mines, with cleanup costs estimated in the tens of billions of dollars.[2] The effectiveness of these remediation efforts is typically assessed, if at all, through before-and-after comparisons of contaminant concentrations or loads. Concentrations and loads can fluctuate dramatically, however, as storm events trigger episodic erosion of tailings piles, transport of sediment-bound contaminants, and releases of acid mine drainage.[1,3−5] Thus, delivery of pollutants to downstream waters will not only depend on mine site conditions (and thus on the effectiveness of remediation efforts) but also on external drivers such as the magnitude and frequency of storm events. Because these external drivers can differ substantially between pre- and postremediation monitoring periods, it is often difficult to tell how much of any measured change in concentrations or loads is attributable to the cleanup itself and how much is due to differences in the external drivers.[6]

Here, we present a method for quantifying the effectiveness of remediation efforts while explicitly accounting for changes in external drivers. We previously proposed[7] that remediation effectiveness could be visualized by comparing contaminant rating curves, to show how the contaminant of interest responds to external forcing (such as stream discharge, as a proxy measure for storm intensity) under both "before" and "after" conditions.[8,9] Here, we make this approach more explicitly quantitative, by using these rating curves to estimate the cumulative contaminant loads that would be observed if the external drivers were identical in the "before" and "after" periods. Comparing these loads gives a quantitative estimate of the effectiveness of the remediation efforts and corrects for the confounding variation in the external drivers. We illustrate this approach with a remediation case study at an abandoned Hg mine in Northern California.

Study Site. The Gambonini mine, located approximately 45 km north of San Francisco, is similar to many other mercury mines in the California Coast Range.[10] The Gambonini mine is the largest industrial source of Hg pollution to Tomales Bay,

Received: May 1, 2011
Accepted: August 9, 2011
Revised: August 8, 2011
Published: August 09, 2011

Table 1. Pre- and Postremediation Water Quality Samples at Gambonini Mine Flume[a]

	preremediation	postremediation
Mercury Grab Samples	$n = 18$	$n = 10$
total Hg (ng/L)	167 000 ± 58 000 (490−1 000 000)	740 ± 290 (58−2600)
filtered Hg (ng/L)	66 ± 7 (31−140)	61 ± 9 (29−110)
particulate Hg (μg/g)	66 ± 8 (21−136)	15 ± 3 (6−32)
TSS (mg/L)	2600 ± 1100 (19−19 800)	63 ± 30 (1−303)
Isokinetic Sediment Samples	$n = 16$	$n = 11$
TSS (mg/L)	1170 ± 310 (5−3700)	105 ± 42 (5−420)

[a] Values given as mean±standard error and range.

which is otherwise nearly pristine. Mercury-laden mine waste accumulates at the Walker Creek delta in Tomales Bay, roughly 20 km downstream from the Gambonini mine, resulting in elevated methylmercury in biota.[11] From 1964 to 1970, open-pit mining was used to extract approximately 170 000 kg of Hg from a high-grade localized cinnabar ore body at the Gambonini mine.[11,12] Overburden was dumped on the steep slope below the pit, and the processed ore was dumped into an adjacent ravine. By 1990, the steep ephemeral creek draining the area was incising through the toe of the waste pile and causing mass failure (Supporting Information, Figure S1). The waste pile also showed abundant evidence of surface erosion. More information about the mine and its history is available elsewhere.[7,12]

We intensively monitored the small stream draining the mine site during the El Niño winter of 1998, using continuous measurements of turbidity[13] as a proxy[14,15] for water-column concentrations of sediment-bound mercury. These monitoring efforts[7] revealed substantial Hg loads and triggered emergency Superfund designation for the Gambonini mine, with approximately 3 million dollars allocated for remediation work. The remediation design (Supporting Information, Figure S1) included cut-and-fill regrading, installation of a drainage system, and revegetation with native plants, with the goal of stabilizing the primary mine waste deposit and reducing the erosion of Hg-laden sediment.[12]

To quantify the effectiveness of the remediation, we remeasured Hg loads from the mine site in the winter of 2005, employing the same methods we used in 1998. Our measurements indicated that Hg loads in 2005 were a factor of 1000 lower than the preremediation loads we measured in 1998. However, rainfall in 2005 was also lower by a factor of 3, with markedly reduced storm frequency and intensity. Rainfall, particularly during intense storms, is a strong driver of erosion and downstream transport of the mining waste. Properly assessing the effectiveness of the remediation requires determining how much of the 1000-fold reduction in Hg loads is attributable to the remediation itself and how much is attributable to the difference in rainfall between the two years.

METHODS

Continuous Monitoring of Discharge and Suspended Sediment. We continuously monitored the small stream draining the Gambonini mine during winter storm seasons before and after the remediation (January−February of 1998 and 2005, respectively). This stream collects all the stormwater and seepage from the mine and the waste pile. We constructed a flume roughly 800 m downstream from the waste pile to continuously record stream discharge and turbidity. The mine and the waste pile comprised about 7% of the 0.7 km^2 catchment area draining to our gauging station. The metal cut-throat flume[16] was designed to accommodate a large range of flows and to pass bedload, both essential requirements for measuring discharge in flashy, steep, headwater streams like ours. At the flume, rainfall was recorded by a tipping bucket rain gauge, discharge was measured from water levels in two stilling wells, and turbidity was measured using an optical backscatter sensor (OBS). All of these measurements were digitally recorded at 5−15 min intervals. We calibrated the flume using manually measured velocities spanning a range of flow depths and calibrated the OBS readings using total suspended solids (TSS) samples from an isokinetic depth-integrated sampler[7] (Supporting Information, Figure S3; eqs S1−S5). During preremediation storm events, turbidity sometimes exceeded the measurement range of the OBS. Under these conditions, TSS concentrations were strongly correlated with discharge measurements ($r^2 = 0.93$) and could be inferred directly from them (Supporting Information, Figure S3b, eq S3).

Mercury Sampling and Analysis. Stream Hg samples were collected in triplicate and composited to minimize the effects of short-term fluctuations in concentrations. Samples were immediately placed on ice, and within 24 h, samples were composited, filtered, and oxidized with BrCl in the lab. Stream samples were analyzed for total, dissolved (filter size 0.45 μm), and particulate Hg concentrations using ultraclean sample handling techniques,[17] $SnCl_2$ reduction, dual gold amalgamation, and cold vapor atomic fluorescence detection.[18] Recovery from a total of 18 matrix spikes ranged from 90.5% to 109%, and recovery for certified standard reference materials ranged from 94% to 114.1%. The minimum detection limits for total Hg were 2−3 orders of magnitude, or more, below our sample concentrations.

Continuous Hg Inferred from Turbidity Monitoring. In both the pre- and postremediation monitoring periods, precipitation and stream discharge were highly episodic. TSS and Hg concentrations varied by orders of magnitude, making simple averages unreliable as estimates of the long-term loading to receiving waters (Table 1, Figure 1). Accurately measuring cumulative loads in episodic systems such as ours requires continuous monitoring.[7] Particulate Hg dominated total loads during both monitoring periods, and total Hg was strongly correlated with TSS concentrations ($r^2 = 0.96$ and 0.85 in 1998 and 2005, respectively; Supporting Information, Figure S2). We therefore developed calibration equations relating dissolved, particulate, and total Hg concentrations to TSS concentrations (Supporting Information, eqs S6−S11) and used these to infer a continuous proxy record of Hg loads from the high-frequency instrumental measurements of turbidity. Mercury loads inferred by this technique closely match ($r^2 = 0.97$) manually measured instantaneous loads, spot sampled across 5 orders of magnitude (Supporting Information, Figure S4). We calculated cumulative Hg loads by integrating the time series of continuous proxy measurements over both monitoring periods.

Modeling of Hg Loads under Different Streamflow Regimes. Accurate assessments of remediation effectiveness require before-and-after comparisons of contaminant loads under comparable levels of external forcing (represented in our case by streamflow, as a measure of effective precipitation). To facilitate these comparisons, we combined our calibration equations expressing Hg as a function of TSS, described above, with rating

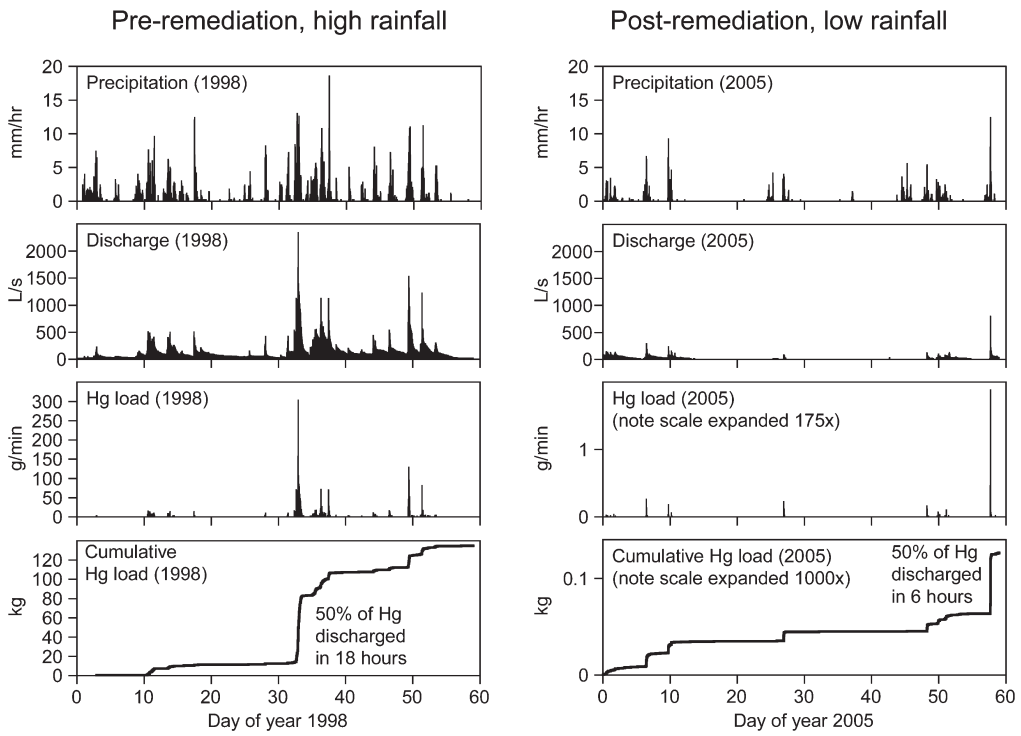

Figure 1. Response of stream discharge and mercury loads to rainfall events in 1998 (before remediation, left column) and 2005 (after remediation, right column). Much less rain fell in 2005 than in 1998, and stream discharge was much lower. Sediment and mercury loads were much lower in 2005 than in 1998, due to the effects of both site remediation and reduced precipitation intensity. Mercury discharges were highly episodic, with 50% of the two-month total load occurring in a single storm in each year.

curves expressing TSS as a function of discharge under pre- and postremediation site conditions. Suspended sediment concentrations are highly stochastic, varying greatly even for a single given value of discharge. Therefore, we divided the continuous TSS measurements for each monitoring period into a series of discharge bins, calculated the average TSS for each bin, and fitted quadratic rating curves (Supporting Information, eqs S12 and S13) to the binned averages. If there were substantial hysteresis in the TSS-discharge relationship, it could be incorporated in the analysis by using separate rating curves for rising and falling flows, but our measurements indicate only minor hysteresis (Supporting Information, Figure S5b).

The resulting rating curves (Figure 2b, Supporting Information, Figure S5) will not necessarily agree with individual instantaneous TSS values, but they will accurately reflect the average TSS, making them suitable for estimating cumulative loads. These TSS rating curves, combined with our Hg−TSS calibration equations, yield Hg rating curves expressing average concentrations as a function of discharge, under both pre- and postremediation conditions (Supporting Information, eqs S14 and S15, Figure 3). These Hg rating curves can then be combined with either historical or hypothetical discharge records to estimate cumulative Hg loads in any given rainfall regime. These estimated loads agree with those calculated from the continuous monitoring data to within 2 and 7% in the pre- and postremediation periods, respectively, verifying the accuracy of this approach.

RESULTS AND DISCUSSION

Postremediation Monitoring Showed Markedly Lower Hg Loads But Also Lower Rainfall. Rainfall was much lower in 2005 compared to 1998, with markedly lower storm frequency and intensity. Figure 1 illustrates the effects of the lower rainfall intensity in 2005 on stream discharge and pollutant loads. Whereas rainfall was 3 times lower in 2005 than in 1998, discharge was nearly 8 times lower in 2005, because less intense and less frequent rainfall allowed for proportionally greater infiltration and evapotranspiration. Sediment and Hg loads were highly episodic, with individual high-intensity storms discharging roughly half the Hg load in each monitoring period (Figure 1, Table 2). Sediment and Hg loads were much lower in 2005 than in 1998 (note the changes in scale between 1998 and 2005 in Figure 1). In comparison to the preremediation cumulative loads, over 180 times less sediment and 1000 times less Hg were discharged after remediation (Table 2). However, this simple before-and-after comparison of contaminant concentrations and loads does not distinguish the effects of site cleanup from the effects of the markedly lower rainfall in 2005. How much of the decrease is attributable to the site remediation, and how much is attributable to the differences in weather between the two years?

Distinguishing Weather and Site Remediation Effects on Hg Loads. Estimating the effects of the site remediation requires normalizing for the differences in weather between the pre- and postremediation monitoring periods. In other words, we should not compare the measured preremediation (1998) mercury loads with the loads that were actually observed postremediation (with markedly lower rainfall) but instead compare them with an estimate of what mercury loads would have been under postremediation site conditions with 1998's intense rainfall. Likewise, we should compare the measured postremediation mercury loads, not with the actual measured loads in 1998, but with an estimate of the mercury loads that would have occurred under

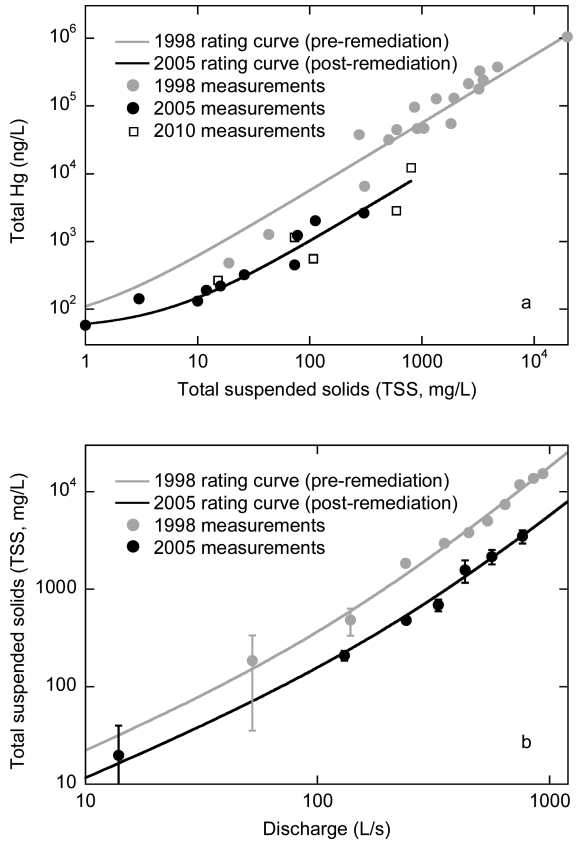

Figure 2. Separation of contaminant rating curves into their two component parts: (a) the dependence of total Hg concentrations on total suspended solids (TSS), and (b) the relationship between TSS and discharge. Total Hg concentrations converge at low TSS because dissolved Hg concentrations are similar in pre- and postremediation periods. Hg concentrations for a given TSS were lower by a factor of 4−6 under postremediation conditions (panel a), and TSS for a given discharge was lower by a factor of 2−3 (panel b). As a result, Hg for a given stream discharge was lower by a factor of 8−18 under postremediation conditions (Figure 3).

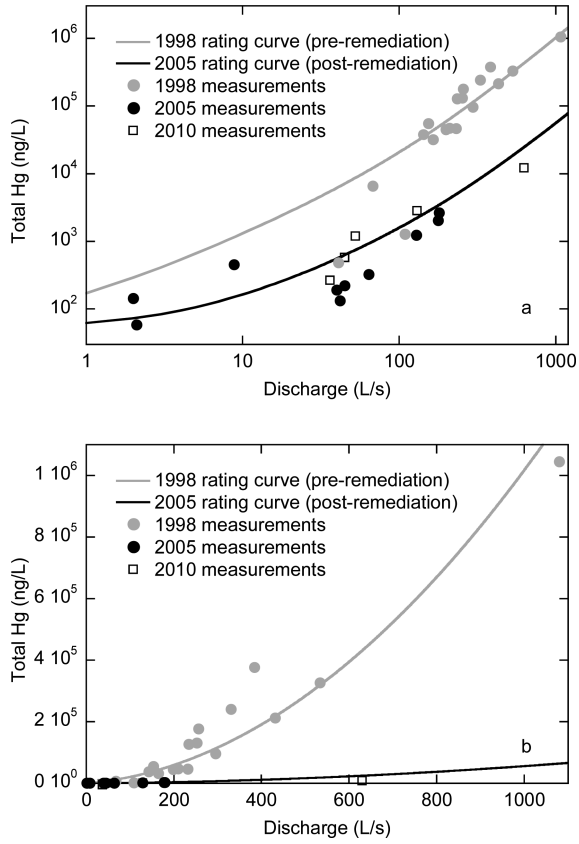

Figure 3. Before-and-after contaminant rating curves relating Hg concentrations and stream discharge on logarithmic (a) and linear (b) axes. Preremediation Hg concentrations and rating curve (gray dots and line) reflect baseline conditions at the mine site. Postremediation Hg concentrations and rating curve (black dots and line) lie well below the preremediation data. Additional samples taken in 2010 (open squares) confirm the general trend of the postremediation rating curve but were not used to generate it. Rating curves were not fitted to the plotted data but instead were calculated from the dependence of Hg concentrations on TSS (Figure 2a, Supporting Information, Figure S2) and the dependence of TSS on discharge (Figure 2b, Supporting Information, Figure S5). The distance between the two rating curves shows the decrease in Hg concentrations at any given discharge and thus measures remediation effectiveness for any given hydrologic forcing.

preremediation site conditions with 2005's scanty rains. These hypothetical combinations of site conditions and weather conditions can be modeled straightforwardly using the contaminant rating curves shown in Figures 2 and 3, together with the corresponding discharge time series (preremediation rating curves with 2005 discharge records and postremediation rating curves with 1998 discharge records).

The results of this analysis are shown in Table 3. We estimate that if 1998's very wet weather had also been repeated in 2005, the cumulative Hg load under postremediation conditions would have been 9.3 ± 2.4 kg, a factor of 14.5 smaller than the 135 ± 20 kg measured under preremediation conditions (Table 3). Therefore, we infer that the remediation reduced Hg loads by (135 − 14.5)/135, or 93 ± 2%, relative to what they would have been under 1998's wet weather conditions. We likewise estimate that if 2005's dry weather had also occurred in 1998, the cumulative Hg load under preremediation conditions would have been 1.59 ± 0.53 kg, a factor of 12.6 larger than the 0.126 ± 0.022 kg measured under postremediation site conditions and rainfall, implying a 92 ± 3% reduction in Hg loads due to the remediation (Table 3). The estimated effect of the remediation is the same, within uncertainties, under either weather regime.

Using the same load estimates, we can also calculate the consequences of different weather patterns for cumulative loads from the site. Reading across the rows of Table 3, we estimate that under preremediation site conditions, the difference between wet-weather (1998) and dry-weather (2005) conditions would account for a factor of 85 ± 31 difference in Hg loads. Under postremediation site conditions, the difference in weather accounts for a factor of 73 ± 23 difference in Hg loads; again, the two estimates agree within their estimated uncertainties.

We could have used the rating curves to model all four combinations of pre- and postremediation site conditions and weather in Table 3, rather than comparing modeled loads for the two counterfactual cases with proxy measurements of loads for the two cases where these are available. The former approach would be preferable if there were large discrepancies between modeled loads and the measured loads, because it would avoid misinterpreting these discrepancies as effects of weather or site conditions.

(Such large discrepancies might also lead one to question the validity of the modeled loads in the first place.) In the present case, however, the discrepancies are small and both approaches yield similar results.

Measured Cumulative Loads as Indicators of Cleanup Effectiveness. In these particular monitoring periods at this particular site, cumulative Hg loads are affected more by the pre- and postremediation differences in weather conditions than by the results of the remediation. This observation implies that if the sequence of weather conditions had been reversed, that is, if the wet winter of 1998 had come instead in 2005, measured Hg loads would have been markedly higher after the remediation than before, by a factor of 5.9 ± 2.5 (this result can be obtained by comparing the upper right and lower-left elements of Table 3). Such a circumstance could lead to the false conclusion that the cleanup efforts had been counterproductive. This example illustrates the risks involved in measuring remediation effectiveness only through changes in contaminant loads, particularly when those loads are strongly influenced by confounding changes in external drivers.

Table 2. Precipitation, Discharge, and Loads Recorded at Gambonini Mine Flume during Pre- and Postremediation Monitoring Periods[a]

	preremediation 1/1/98–2/28/98	postremediation 1/1/05–2/28/05
cumulative precipitation (mm/2 months)	890	240
max 24 h precipitation (mm/24 h)	100	37
max 1 h precipitation (mm/h)	17	12
cumulative discharge (m^3/2 months)	697 000	89 100
max 24 h discharge (m^3/24 h)	73 100	7800
max 1 h discharge (m^3/h)	6500	1700
cumulative sediment loads (kg/2 months)	2 343 000	12 500
max 24 h sediment loads (kg/24 h)	1 214 000	6330
max 1 h sediment loads (kg/h)	198 000	4640
cumulative Hg loads (kg/2 months)	135	0.126
max 24 h Hg loads (kg/24 h)	68.3	0.0620
max 1 h Hg loads (kg/h)	11.1	0.0452

[a] Results rounded to two or three significant figures. Preremediation discharges and loads differ from those reported by Whyte and Kirchner[7] due to improved flume calibrations.

Concentrations in Grab Samples as Measures of Cleanup Effectiveness. Due to the practical difficulties in measuring contaminant loads over extended periods, cleanup effectiveness is often assessed by before-and-after comparisons of contaminant concentrations in grab samples. If contaminant concentrations are strongly influenced by external drivers, however, they will depend on the particular times the grab samples are taken and the external forcings that prevail at those times. We can illustrate the effects of external drivers and sampling times by estimating the grab sample concentrations that we would have measured if the wet winter of 1998 had come after the remediation instead of before it. To do this, we rescaled each grab sample measurement (Table 1 and Supporting Information, Tables S1 and S2) by the ratio between the rating curves for the pre- and postremediation conditions, for the particular discharges associated with each of the sampling times. This calculation preserves a realistic degree of variability around the rating curves and thus is more realistic than modeling the grab sample concentrations directly from the rating curves themselves. The results of this calculation (Table 4) show that, if the sequence of the weather conditions had been reversed, the benefits of the cleanup would not have been visible in the concentrations measured in our grab samples.

Contaminant Concentrations on Sediment as Measures of Cleanup Effectiveness. Total Hg at our site is dominated by particulate Hg, which is the product of the TSS concentration (mg of sediment per L of water) and the Hg concentration in the suspended sediment itself (μg of Hg per g of sediment). TSS concentrations reflect the rate of erosion and transport of sediment and therefore vary strongly with rainfall forcing, whereas Hg concentrations in the suspended sediment mostly reflect the mix of sediment sources, which is less sensitive to weather conditions. Hg concentrations in the suspended sediment show a factor of 4.4 ± 1.0 or 6.9 ± 1.0 decrease following the cleanup, depending on whether one assumes the actual or inverted sequence of wet and dry years. Because sediment Hg concentrations are relatively insensitive to rainfall forcing, they indicate that the remediation is beneficial regardless of whether wet conditions precede (Table 1) or follow (Table 4)

Table 3. Effects of Remediation and Weather Conditions on Sediment and Mercury Loads[a]

Hg Loads (kg/2 months)	precipitation conditions	
site conditions	wet year (1998)	dry year (2005)
preremediation	135 ± 20 (measured)	1.59 ± 0.53 (modeled)
postremediation	9.3 ± 2.4 (modeled)	0.126 ± 0.022 (measured)
reduction attributable to remediation	93% ± 2% (factor of 14.5)	92% ± 3% (factor of 12.6)
Sediment Loads (metric tons/2 months)	**precipitation conditions**	
site conditions	wet year (1998)	dry year (2005)
preremediation	2393 ± 347 (measured)	28.1 ± 9.4 (modeled)
postremediation	950 ± 224 (modeled)	12.5 ± 1.9 (measured)
reduction attributable to remediation	60% ± 11% (factor of 2.6)	55% ± 16% (factor of 2.2)

[a] Measured 1998 sediment and Hg loads differ from those reported by Whyte and Kirchner[7] due to improved flume calibrations.

Table 4. Water Quality Measurements Expected If Wet Weather Had Followed, Rather Than Preceded, Site Remediation[a]

	preremediation site conditions with dry year (2005) precipitation	postremediation site conditions with wet year (1998) precipitation
Mercury Grab Samples	$n = 10$	$n = 18$
total Hg (ng/L)	9800 ± 4300 (240−38 000)	10 000 ± 3200 (44−57 000)
filtered Hg (ng/L)	65 ± 9 (31−120)	61 ± 6 (29−130)
particulate Hg (μg/g)	76 ± 8 (24−110)	11 ± 1 (1−23)
TSS (mg/L)	150 ± 75 (2−770)	890 ± 330 (9−6200)
Isokinetic Sediment Samples	$n = 11$	$n = 16$
TSS (mg/L)	260 ± 110 (10−1100)	430 ± 110 (2−1300)

[a] Values given as mean±standard error and range.

remediation of the site. But, for the same reason, sediment Hg concentrations do not reflect changes in sediment loads resulting from erosion control measures (Figure 2b) and thus underestimate the benefits of the remediation.

Contaminant Rating Curves as Measures of Cleanup Effectiveness. In contrast to contaminant concentrations or loads, contaminant rating curves (Figures 2 and 3) have the distinct advantage that they allow all-else-equal comparisons of pre- and postremediation contaminant concentrations at any specified level of external forcing. Thus, it does not matter whether the external forcing is comparable in the pre- and postremediation periods.

From Figure 3a, one can directly see that for any specified discharge between 10 and 1000 L/s, postremediation concentrations of total Hg are a factor of 8−18 lower than preremediation concentrations at the same discharge, implying that the remediation has reduced Hg loads by 87−95%. Figure 2 shows that this reduction in contaminant loads has two distinct components. Under postremediation conditions, expected Hg concentrations are lower (by a factor of 4−6) at any TSS concentration between 10 and 1000 mg/L (Figure 2a), and average TSS concentrations are lower (by a factor of 2−3) at any specified discharge between 10 and 1000 L/s (Figure 2b). Multiplying these factors together directly yields the 8- to 18-fold reduction in Hg concentrations shown in Figure 3. Most importantly, these conclusions can be drawn on an all-else-equal basis, whether or not the external forcing is similar in the pre- and postremediation monitoring periods.

Changes in site conditions resulting from the remediation are reflected in the differences between the pre- and postremediation rating curves. The downward shift in the relationship between total Hg and suspended solids concentrations, for example, (Figure 2a; Supporting Information, Figure S2) reflects decreases in Hg concentrations in the eroded mine waste (because the most contaminated waste was removed), and reductions in the amount of mine waste eroded relative to cleaner sediment coming from the rest of the catchment. Likewise, the downward shift in the sediment rating curve (Figure 2b; Supporting Information, Figure S5a) reflects the reduced erodibility of the waste pile and stream bed across a range of hydrologic forcing.

Value of Continuous Proxy Data. Contaminant rating curves are typically constructed directly from spot measurements of contaminant concentrations and discharge, for example, by fitting regression curves directly to the gray and black dots in Figure 3. Such rating curves are relatively simple to construct, their data requirements are modest, software is available[19] to automatically estimate them and translate them into loads, and in many cases they should yield reasonable estimates of remediation effectiveness. Our analysis presented above is more complex; first, TSS is measured continuously with an optical sensor, then binned averages of these measurements are used to estimate a TSS rating curve (Figure 2b), then this is transformed into a Hg rating curve (Figure 3) using the relationship between Hg and TSS in grab samples (Supporting Information, Figure S3). Is this extra effort worthwhile, or would conventional rating curves work just as well?

To test the conventional rating curve approach, we fitted power-law rating curves directly to log−log plots of the 1998 and 2005 measured values of total Hg concentrations and discharge. We corrected these curves for log retransformation bias[20] and combined them with the 1998 and 2005 discharge records to model cumulative Hg loads (Supporting Information, Table S7). For the preremediation case (1998), the conventional rating curve approach gives a total Hg load of 435 ± 225 kg (relative standard error: 52%), compared to 135 ± 20 kg (relative standard error: 15%) obtained from our continuous proxy measurements. For the postremediation case (2005), the conventional rating curve approach gives a total Hg load of 0.066 ± 0.021 kg (relative standard error: 32%), compared to 0.126 ± 0.022 kg (relative standard error: 18%) obtained from our continuous proxy measurements. In other words, the relative uncertainties are roughly 2−3 times larger in the conventional rating curve approach than in our proxy measurement approach. The conventional rating curve approach also overestimates the preremediation Hg load by more than 3-fold, underestimates the postremediation load by nearly 2-fold, and overestimates the effectiveness of the remediation by large factors (compare Tables 3 and S7 of the Supporting Information).

Why does this happen? In highly episodic streams like ours, TSS can vary by orders of magnitude at any given flow. In such streams, rating curves constructed from a relative handful of grab samples will be highly variable, depending on whether the samples were collected at times with above- or below-average sediment concentrations. Log−log rating curves can compound this problem, because points at low concentration and discharge (which make a trivial contribution to the total load) can have significant leverage on the high-concentration, high-discharge end of the curve and thus a large influence on the estimate of total load. Where streams are less episodic, however, or where more comprehensive sets of grab samples are available, the conventional rating curve approach may be a viable alternative to the more complex proxy measurement method described here.

Smarter Analyses, More Data, or Both? Changes in external forcing can confound assessments of remediation effectiveness based on simple before-and-after comparisons of contaminant concentrations or loads. The pitfalls in such assessments are particularly apparent in our case study for several reasons: because mercury loads at the Gambonini mine are highly sensitive to episodic forcing, because that forcing differed dramatically between the pre- and postremediation monitoring periods, and because those monitoring periods themselves were brief, with limited numbers of direct measurements. The problems

outlined here could have been alleviated somewhat through more extensive sampling that spanned multiple years, thereby encompassing wider ranges of natural variability. The fact that many remediation projects involve even less before-and-after monitoring than we conducted at the Gambonini mine highlights the need for much bigger investments in such measurement programs. Even extensive sampling programs, however, are likely to capture different levels of pre- and postremediation external forcing. For example, USGS gauging records collected 5 km downstream of the Gambonini mine show that a runoff event like the 1998 storm that dominates the preremediation Hg loads has occurred only once, lasting only one day, in the dozen years since the remediation. Thus, even with more and better sampling, methods like those outlined here would still be needed to correct for differences in external forcing and accurately quantify remediation effectiveness.

Society invests significant resources in remediation efforts. Justifying these investments requires not just monitoring, but also quantitatively assessing the resulting data[21,22] "to evaluate remediation effectiveness, to determine whether clean-up goals have been met, and to assess which remediation strategies are most effective".[9] Accurate assessments require separating remediation effects from confounding factors, such as variations in external drivers. As our case study shows, when external drivers differ significantly between pre- and postremediation monitoring periods, conventional comparisons of contaminant concentrations and loads can be highly misleading measures of remediation effectiveness. Even when external drivers change, however, contaminant rating curves can show remediation effects clearly and intuitively. Before-and-after comparisons of contaminant rating curves therefore merit serious consideration as tools for assessing the environmental benefit of remediation projects.

■ ASSOCIATED CONTENT

Supporting Information. Additional text detailing calibration equations, calculation procedures, and error propagation methodology, seven tables of primary data and modeled loads, and six figures showing site conditions, calibration curves, and rating curves. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: +41 44 739 26 55; fax: +41 44 739 27 25; e-mail: james.kirchner@wsl.ch, kirchner@berkeley.edu.

■ ACKNOWLEDGMENT

Our field sampling campaigns would not have been possible without the help of many individuals, most notably Jill Marshall, John Replogle, and Elysha Anderson. We thank two anonymous reviewers for their comments on the manuscript. This work was partially supported by the California Water Quality Control Board cleanup and abatement fund and by NSF grant EAR-9357931 to J.W.K.

■ REFERENCES

(1) Macklin, M. G.; Brewer, P. A.; Hudson-Edwards, K. A.; Bird, G.; Coulthard, T. J.; Dennis, I. A.; Lechler, P. J.; Miller, J. R.; Turner, J. N. A geomorphological approach to the management of rivers contaminated by metal mining. *Geomorphology* **2006**, *79*, 423–447.

(2) *Federal land management: Information on efforts to inventory abandoned hard rock mines*; U.S. General Accounting Office: Washington, D.C., 1996; http://www.gao.gov/archive/1996/rc96030.pdf.

(3) Rytuba, J. J. Mercury from mineral deposits and potential environmental impact. *Environ. Geol.* **2003**, *43*, 326–339.

(4) Ferrara, R.; Maserti, B. E.; Breder, R. Mercury in abiotic and biotic compartments of an area affected by a geochemical anomaly (M. Amiata, Italy). *Water, Air, Soil Pollut.* **1991**, *56*, 219–233.

(5) Mighanetara, K.; Braungardt, C.; Rieuwerts, J.; Azizi, F. Contaminant fluxes from point and diffuse sources from abandoned mines in the River Tamar catchment, UK. *J. Geochem. Explor.* **2009**, *100*, 116–124.

(6) Runkel, R. L.; Bencala, K. E.; Kimball, B. A.; Walton-Day, K.; Verplanck, P. L. A comparison of pre- and post-remediation water quality, Mineral Creek, Colorado. *Hydrol. Processes* **2009**, *23*, 3319–3333.

(7) Whyte, D. C.; Kirchner, J. W. Assessing water quality impacts and cleanup effectiveness in streams dominated by episodic mercury discharges. *Sci. Total Environ.* **2000**, *260*, 1–9.

(8) Gelinas, S. L.; Tupling, R. Monitoring remediation--have mine-waste and mill-tailings removal and flood-plain restoration been successful in the High Ore Creek Valley? In *Integrated Investigations of Environmental Effects of Historical Mining in the Basin and Boulder Mining Districts, Boulder River Watershed, Jefferson County, Montana*; Nimick, D. A., Church, S. E., Finger, S. E., Eds.; U.S. Geological Survey: Denver, CO, 2004; pp 461−473.

(9) Unruh, D. M.; Church, S. E.; Nimick, D. A.; Fey, D. L. Metal contamination and post-remediation recovery in the Boulder River watershed, Jefferson County, Montana. *Geochem.: Explor. Environ., Anal.* **2009**, *9*, 179–199.

(10) Rytuba, J. J. Mercury mine drainage and processes that control its environmental impact. *Sci. Total Environ.* **2000**, *260*, 57–71.

(11) Johnson, B. E.; Esser, B. K.; Whyte, D. C.; Ganguli, P. M.; Austin, C. M.; Hunt, J. R. Mercury accumulation and attenuation at a rapidly forming delta with a point source of mining waste. *Sci. Total Environ.* **2009**, *407*, 5056–5070.

(12) Smelser, M. G.; Whyte, D. C. Remediation of the Gambonini mercury mine, Marin County, California. In *Engineering Geology Practice in Northern California*; Ferriz, H., Anderson, R., Eds.; California Division of Mines and Geology: Sacramento, CA, 2001; pp 629−640.

(13) Gippel, C. J. Potential of turbidity monitoring for measuring the transport of suspended solids in streams. *Hydrol. Processes* **1995**, *9*, 83–97.

(14) Grayson, R. B.; Finlayson, B. L.; Gippel, C. J.; Hart, B. T. The potential of field turbidity measurements for the computation of total phosphorus and suspended solids loads. *J. Environ. Manage.* **1996**, *47*, 257–267.

(15) Stubblefield, A. P.; Reuter, J. E.; Dahlgren, R. A.; Goldman, C. R. Use of turbidometry to characterize suspended sediment and phosphorus fluxes in the Lake Tahoe basin, California, USA. *Hydrol. Processes* **2007**, *21*, 281–291.

(16) Kilpatrick, F. A.; Schneider, V. R. Use of flumes in measuring discharge. In *U.S. Geological Survey Techniques of Water-Resources Investigations, Book 3*; U.S. Government Printing Office: Washington, D.C., 1983; pp 55.

(17) *Method 1669: Sampling Ambient Water for Trace Metals at EPA Water Quality Criteria Levels*; United States Environmental Protection Agency: Washington, D.C., 1996; http://www.epa.gov/waterscience/methods/method/inorganics/1669.pdf.

(18) *Method 1631, Revision E: Mercury in Water by Oxidation, Purge and Trap, and Cold Vapor Atomic Fluorescence Spectrometry*; United States Environmental Protection Agency: Washington, D.C., 2002; http://www.epa.gov/waterscience/methods/method/mercury/1631.pdf.

(19) Runkel, R. L.; Crawford, C. G.; Cohn, T. A. Load Estimator (LOADEST): A FORTRAN program for estimating constituent loads in streams and rivers. In *U.S. Geological Survey Techniques and Methods, Book 4*; U.S. Government Printing Office: Washington, D.C., 2004; pp 69.

(20) Ferguson, R. I. River loads underestimated by rating curves. *Water Resour. Res.* **1986**, *22*, 74–76.

(21) Kondolf, G. M.; Micheli, E. R. Evaluating stream restoration projects. *Environ. Manage.* **1995**, *19*, 1–15.

(22) Finger, S. E.; Church, S. E.; Nimick, D. A. Evaluating the success of remediation in the boulder river watershed. In *Integrated Investigations of Environmental Effects of Historical Mining in the Basin and Boulder Mining Districts, Boulder River Watershed, Jefferson County, Montana*; Nimick, D. A., Church, S. E., Finger, S. E., Eds.; U.S. Geological Survey: Denver, CO, 2004; pp 497−502.