Water Resources Research

RESEARCH ARTICLE
10.1029/2019WR026389

Measuring and Modeling Gravel Transport at Caspar Creek, CA, to Detect Changes in Sediment Supply, Storage, and Transport Efficiency

P. W. Richardson[1], J. W. Wagenbrenner[1], D. G. Sutherland[1], and T. E. Lisle[1]

[1]USDA Forest Service, Pacific Southwest Research Station, Arcata, CA, USA

Key Points:
- We reconstructed a 55-year record of annual gravel yields from weir pond deposits at Caspar Creek, CA
- Reconstructed gravel yields compare well to modeled gravel yields and to a continuous record of bed load transport rates
- Comparing reconstructed yields to predictions reveals a decrease in transport associated with increased downed wood and sediment storage

Supporting Information:
- Supporting Information S1

Correspondence to:
P. W. Richardson,
paul.richardson@usda.gov

Citation:
Richardson, P. W., Wagenbrenner, J. W., Sutherland, D. G., & Lisle, T. E. (2020). Measuring and modeling gravel transport at Caspar Creek, CA, to detect changes in sediment supply, storage, and transport efficiency. *Water Resources Research*, *56*, e2019WR026389. https://doi.org/10.1029/2019WR026389

Received 20 SEP 2019
Accepted 9 APR 2020
Accepted article online 14 APR 2020

T. E. Lisle retired.

Abstract We developed a technique for reconstructing annual gravel yields and generated a 55-year record of gravel transport for the North Fork catchment of the Caspar Creek Experimental Watersheds in Northern California. The technique relies on field data collection including annual surveys of weir pond volumes and suspended sediment measurements, as well as an accounting for settling of suspended sediment and organic matter in the pond. We compared these annual yields to gravel yields predicted by the Wilcock two-fraction bed load transport model, which we calibrated from measured values at Caspar Creek. We considered three velocity-discharge relationships and found that values of hydraulic variables measured during storms produced the best fit between reconstructed and predicted annual gravel yields when years with large disturbances were excluded. We also compared predicted gravel transport rates to bed load transport rates measured from 1988 to 1995 with bed load pit samplers. We found that the calibrated model predictions agreed well with the field-measured bed load transport rates. To investigate the role of supply and storage on gravel transport, we compared the reconstructed gravel yields to predicted gravel yields and found that increased occurrence of landslides and headcut erosion in the 1990s and early 2000s did not lead to an increase in gravel yields. Instead, input of large downed wood in the 1990s created storage space and decreased bed load delivery to the weir pond.

Plain Language Summary Bed load sediment yields are important for understanding catchment responses to natural and human caused disturbances, but are difficult to measure. We demonstrate a method that uses pond sediment accumulations to quantify annual bed load yields at Caspar Creek, CA. We also use these data and data collected during more focused storm-based sampling to show that a common bed load model provides relatively accurate estimates of gravel transport at Caspar Creek. Comparing annual gravel yields to predicted gravel yields reveals that gravel transport decreased in the 1990s and 2000s after a large number of trees blew down and increased the wood content in the channel. The approach presented here can be applied to other sites where similar data are available, such as other ponds or small reservoirs.

1. Introduction

1.1. Motivation and Background

Understanding how sediment supply, storage, and transport change through time and respond to land management activities is critical for assessing present and possible future stream habitat conditions. Channel bed composition, morphology, and mobility—which are influenced by the delivery of lithic material to the channel and the rate at which different grain sizes are transported, stored, and organized along the bed—influence anadromous salmonid spawning conditions (e.g., Kondolf & Gordon, 1993; Lisle & Lewis, 2009; Montgomery et al., 2011) and are important to a wide variety of other lotic organisms including primary producers, macroinvertebrates, and other fish (e.g., Merz & Chan, 2005; Reice & Url, 1985; Wood & Armitage, 1997). Bed load transport can also create hazards during large storms (Rinderer et al., 2009), cause long-term changes to stream morphology that can increase flooding risks (Davies & McSaveney, 2011), and incur large financial losses due to damages to infrastructure and property (Badoux et al., 2014). Unfortunately, long-term bed load studies, which are critical for assessing sediment supply, storage, and changes in transport, are rare due in part to difficulties in measuring bed load transport rates or yields (e.g., Bunte et al., 2004; Bunte & MacDonald, 1995; Duck & McManus, 1994; Hean & Nanson, 1987; Wilcock, 2001; Wilcock et al., 2009; Yager et al., 2015). In particular, measuring bed load yields for the duration required to assess the effects of supply and storage is difficult as many common sampling techniques are

labor intensive and not well-suited to long-term studies. Promising indirect measurement techniques such as fiber optic systems (Bray & Dunne, 2017; Selker et al., 2006), morphological analysis of the streambed and bars (e.g., Lane et al., 1995, 2003), geophones and seismic monitoring (e.g., Barrière et al., 2015; Burtin et al., 2011; Hsu et al., 2011; Rickenmann et al., 2012; Roth et al., 2017; Turowski et al., 2011), and bed load tracers (e.g., Klösch & Habersack, 2018; Phillips & Jerolmack, 2014; Schmidt & Ergenzinger, 1992) are under development and may prove useful for developing long-term bed load transport records. However, as promising as these new techniques may be for future studies, they are of little use for reconstructing historical transport records.

Multiple bed load transport models have been developed to aid the task of predicting and understanding changes in transport rates (e.g., Bakke et al., 2002; Parker, 1990; Parker et al., 1982; Parker & Klingeman, 1982; Wilcock, 2001; Wilcock & Crowe, 2003). Long-term records of bed load transport rates would help users determine which models are best suited for predicting transport for gravel-bed streams in different environments. By comparing long-term bed load measurements with accurate predictions and focusing on differences between the two, the consequences of land disturbance can be assessed and changes in sediment supply, storage, and transport can be detected. A site that is well-suited for comparing long-term measurements with bed load predictions is the Caspar Creek Experimental Watersheds in the Northern California Coast Ranges (Figure 1a).

Monitoring at Caspar Creek began in 1962 and a rich data set has been created with the primary purpose of investigating the consequences of timber harvesting, but the data set is also well-suited to address many other questions relevant to small (<5 km^2), forested catchments in temperate, rainfall-dominated climates. We investigated bed load transport for the 4.79 km^2 North Fork catchment of Caspar Creek for a 55-year period and focus on the responses of bed load transport rates, supply, and storage to forest landscape disturbances. Old-growth coast redwood trees were harvested from the 1860s to 1904, and the catchment was partially and experimentally clear-cut from 1985 to 1992, with about half the watershed harvested (Henry, 1998). Following the timber harvest and road construction, landslide occurrence increased in harvested areas of the North Fork (Reid & Keppeler, 2012). In the early to mid-1990s, wind storms caused a substantial increase of large downed wood in the main stem of the North Fork (Reid & Hilton, 1998). The ability to assess how these events influenced bed load transport has been hampered by the difficulty in measuring and modeling bed load yields. Previous efforts have been made to assess changes in bed load transport at Caspar Creek using the weir pond volumes (Lewis, 1998; Lisle & Napolitano, 1998), but with limited success due to the challenge of deciphering changes in bed load yields from the increases in settled suspended sediment owing to the increased post-harvest suspended sediment yields. There is a clear need for approaches to reconstruct bed load yields and for modeling that is consistent and relatively accurate at predicting yields for undisturbed and managed conditions.

For the time being, assessing deposits trapped behind impoundments (e.g., Duck & McManus, 1994) may harbor the highest potential for reconstructing long-term bed load records. Many experimental field sites include sediment traps or weirs capable of trapping most coarse material at the catchment scale and are regularly surveyed to track the volume of trapped sediment (e.g., Cafferata & Reid, 2013; Grant & Wolff, 1991; Renard et al., 2008; Wagenbrenner & Robichaud, 2014). An impediment to determining bed load from deposition in ponds arises when assessing how much of the trapped material is bed load relative to settled suspended sediment or organic material. Previous studies have made efforts to assess sediment trap efficiency—the amount of sediment trapped behind an impoundment relative to the amount that passes the impoundment (e.g., Brune, 1953; Heinemarm, 1981; Rausch & Schreiber, 1981; Verstraeten & Poesen, 2001), but these values include all sediment and are not specific to bed load.

A large weir pond at the outlet of the North Fork catchment has been surveyed annually since 1962. Although insignificant amounts of suspended sediment may settle in smaller weir ponds (Hiraoka et al., 2015), our observations at Caspar Creek and other experimental catchments suggest that a significant portion of pond deposits are composed of settled suspended sediment. We suspect this is also the case for many similarly sized weir ponds at other locations. Lisle and Napolitano (1998) treated the total North Fork weir pond volume as bed load and did not find significant changes in bed load transport following timber harvesting of the North Fork. However, their analysis of channel cross-sections along the main stem of the North Fork showed aggradation during the period following harvesting, which they attributed to

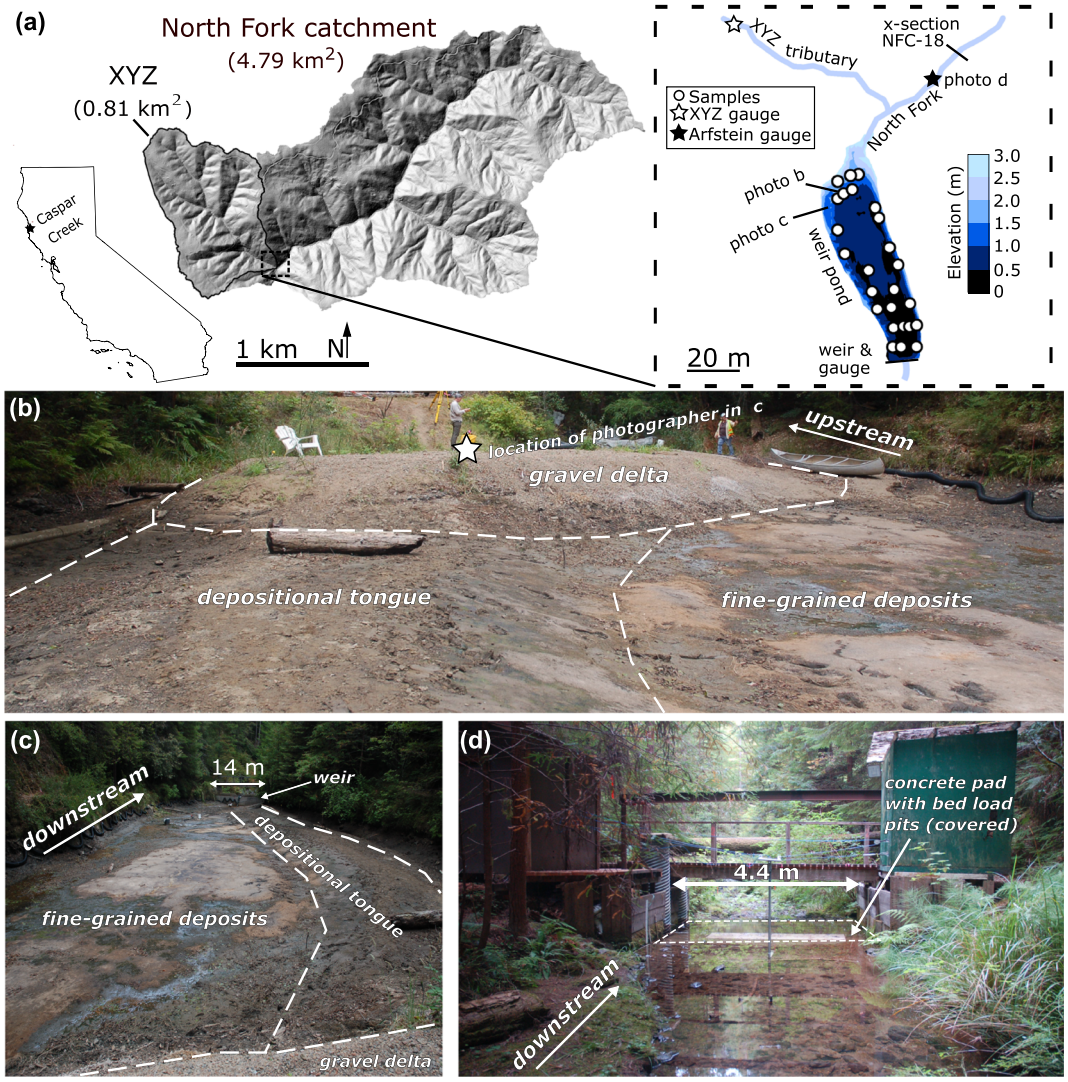

Figure 1. (a) Shaded relief map of the North Fork catchment of the Caspar Creek Experimental Watersheds created from LiDAR data gridded to 1 m. The XYZ subcatchment is outlined in black. Inset of California (left) with location of study site marked. The boxed inset (right) shows details near the North Fork weir pond including a digital elevation map of the pond surface created from 2018 survey data and photograph locations (b)–(d). (b and c) Photographs of the drained North Fork weir pond in preparation for a sediment excavation during summer 2018. Approximate boundaries for different depositional zones are marked in each photograph. (d) Photograph looking downstream to the Arfstein (Arf) gauging station during low summer flow. The engineered cross-section, concrete pad, and bed load pits are visible below the bridge.

increased downed wood in the channel. Hassan et al. (2014) also examined storm-based bed load transport rates during and after timber harvesting of the North Fork measured with Birkbeck-style bed load pits from 1988 to 1995. The bed load pits filled during some storms resulting in an incomplete transport record, but the results did reveal a decrease in bed load transport rates from 1988 to 1995, which were attributed to decreased sediment supply, sediment deposition upstream of downed wood, and bed armoring (Hassan et al., 2014).

1.2. Objectives and Approach

We had three primary objectives for this study: (1) reconstruct a long-term record of gravel transport for the North Fork of Caspar Creek from annually surveyed weir pond deposits and estimates of settled suspended sediment and organic matter, (2) calibrate a bed load transport model for the North Fork to predict annual

gravel yields, and (3) compare reconstructed and predicted gravel yields to detect changes in transport conditions subject to variations in sediment supply and storage.

We began by developing a technique to reconstruct annual bed load and gravel yields from previously collected data and new measurements to produce a 55-year record of bed load yields for the North Fork of Caspar Creek. Next, we assessed discharge through the Arfstein reach of the North Fork, which is immediately upstream of the weir pond, to assemble a consistent discharge data set for the gravel transport model (GTM). We then applied a two-fraction bed load transport model (Wilcock, 2001) and focused on predicting gravel yields. The theoretical basis of the GTM justifies an extrapolation of predicted bed load transport rates for flows greater than those sampled during bed load measurements. Because predicted gravel transport rates may be sensitive to the velocity-discharge relationship, we investigated three velocity-discharge relationships and considered the consequences of each on our results.

We compared the transport rates predicted from the calibrated GTMs to both the reconstructed annual gravel yields and to bed load transport rates measured during 13 storm events from 1988 to 1995 with Birkbeck-style bed load pit samplers installed at the Arfstein station (Figure 1), affording us two opportunities to directly test the quality of our gravel reconstruction. We concluded the study by calculating departures between the predicted and reconstructed gravel yields to detect changes in sediment transport from 1962 to 2017 and investigated the role of known disturbances on sediment supply, storage, and transport efficiency.

2. Site Description, Field Measurements, and Methods

2.1. Site Description

This analysis focused on the North Fork catchment (NFC) of the Caspar Creek Experimental Watersheds (Figure 1). The experimental watershed is located ~10 km south of the town of Fort Bragg and ~7 km inland from the Pacific Ocean. The catchment is incised into a flight of Pleistocene marine terraces (Muhs et al., 2003) and underlain by the Coastal Belt of the Franciscan Complex which includes a variety of Eocene or younger sedimentary rocks including greywacke, shale, limestone, argillites, and conglomerates (Evitt & Pierce, 1975; Hahm et al., 2019; Langenheim et al., 2013). Specifically in the North Fork catchment, the predominate rock types are greywacke and shale (Lisle & Napolitano, 1998). When exposed near the surface, rocks from the Coastal Belt weather quickly and shales disaggregate easily (Hahm et al., 2019).

Streamflow is driven by wet-season rainfall, which primarily occurs between October and April with an annual average of 1,170 mm (Cafferata & Reid, 2013). Snow is rare and the proximity to the Pacific Ocean and coastal fog in the summer helps maintain moderate temperatures throughout the year. Air temperature ranges from a mean of 7 °C in January to 15 °C in July (Keppeler & Brown, 1998).

A compound 120° v-notch and rectangular weir was installed in 1962 at the outlet of the North Fork catchment (Figure 1). Stage has been measured nearly continuously at 10-min intervals since November 1962. The hydrologic year (HY) is defined at Caspar Creek as 1 August through 31 July. Discharge is calculated from the stage using a calibrated rating equation. Most emphasis in prior Caspar Creek studies has been on accurately monitoring storm events, which are events for which the stage exceeds 0.61 m (2 ft). At this stage, flow overtops the v-notch and activates the rectangular weir, and this stage corresponds to a discharge of 0.69 m^3/s. We focused on $Q_{NFC} > 0.69$ m^3/s because discharge below the storm threshold is unlikely to be important for gravel transport in the North Fork. The weir pond is ~70 m long and its width varies between ~14 m at its narrowest point and ~24 m at its widest. The weir pond captures bed load, some settled suspended sediment, and some organic material. Weir pond excavations have occurred periodically over the 55-year period, averaging every 5 to 10 years.

A delta, composed primarily of gravel, is formed during storm events at the upstream end of the weir pond and a depositional tongue extends from the downstream end of the delta toward the weir (Figure 1). The remainder of the pond is primarily fine-grained lithic deposits and organic material. The XYZ tributary joins the main stem of the North Fork ~26 m upstream of the weir pond delta (Figure 1). The main stem of the North Fork and the XYZ tributary near the confluence with the North Fork have gravel-bed channels. The North Fork is a single-thread, plane-bed channel with pools and bars that are forced by obstructions (e.g., large wood) and irregular boundaries. The North Fork typically experiences perennial flow while little to

no flow is measured in the XYZ flume during summer. The Arfstein station is located on the main stem of the North Fork ~43 m upstream of the weir pond and ~17 m upstream of the XYZ confluence. Four Birkbeck-style bed load pits are installed in a concrete pad that spans the channel at the Arfstein gauge (Figure 1), and bed load transport was measured using these pits from 1988–1995 (Hassan et al., 2014).

2.2. Reconstructing Annual Bed Load Yields

We solved for the annual bed load mass in the weir pond using conservation of mass and volume. The volume of the weir pond can be subdivided into individual components according to

$$V_p = V_b + V_s + V_o, \tag{1}$$

where V is annual volume and subscripts denote the pond sediment (p), bed load (b), settled suspended sediment (s), and organic matter (o). The sum of V_s and V_o can be approximated as

$$V_s + V_o = \frac{M_s + M_o}{\rho_{s,o}}, \tag{2}$$

where M_s is the annual settled suspended sediment mass, M_o is the annual organic mass, and $\rho_{s,o}$ is the dry density of mixed organic matter and settled suspended sediment.

The mass of the weir pond deposits can also be subdivided into individual components according to

$$M_p = M_b + M_s + M_o, \tag{3}$$

where M_p is the pond mass and M_b is the bed load mass. We estimated the annual organic mass according to

$$M_o = f_o \rho_p V_p, \tag{4}$$

where f_o is the mass fraction of organic matter in the pond and ρ_p is the dry pond density. The pond and bed load mass are calculated according to

$$M_p = \rho_p V_p, \tag{5a}$$

and

$$M_b = \rho_b V_b. \tag{5b}$$

By combining equations 1–5b, we solved for the bed load mass, M_b,

$$M_b = \frac{\rho_b V_p - \frac{\rho_b}{\rho_{s,o}}\left(1 + \frac{f_o}{1-f_o}\right) M_s}{1 + \frac{\rho_b}{\rho_{s,o}}\left(\frac{f_o}{1-f_o}\right)}. \tag{6}$$

For the model comparison, we focused on the gravel portion of the bed load (M_g), which we estimated as

$$M_g = f_{dg} M_b, \tag{7}$$

where f_{dg} is the gravel fraction (≥2 mm) of the delta. We determined f_{dg} from analysis of delta bulk samples collected ~5 cm below the armored surface during summer 2018. Our workflow for the gravel yield reconstruction is diagrammed in Figure 2 and the values in equations 6 and 7 are summarized in Table 1.

In addition, once M_b is known, we can solve for the pond density by combining equations 3–5a, so that

$$\rho_p = \frac{M_b + M_s}{V_p(1 - f_o)}. \tag{8}$$

Weir pond depositional surfaces have been surveyed each summer. Accumulated volumes were estimated by differencing surveyed sediment surfaces for consecutive years (see supporting information Text S2 for additional information regarding the sediment pond volumes).

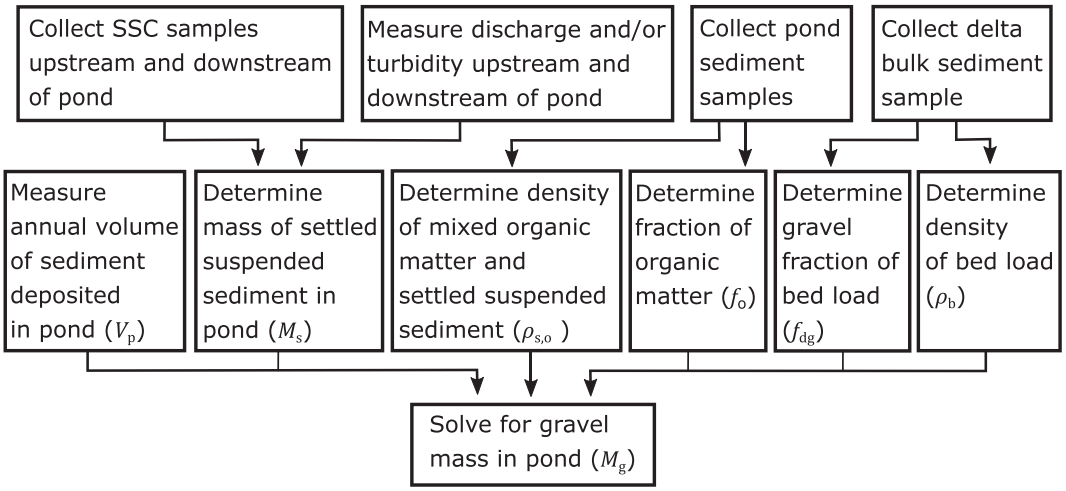

Figure 2. Diagram showing suggested workflow to solve for the annual mass of gravel in weir ponds. In some cases, it may be possible to estimate necessary values such as the mass of settled suspended sediment (M_s) instead of measuring it directly.

Suspended sediment concentration (SSC, mg/L) has been measured from samples collected at the North Fork weir gauging station since fall 1962 using three different techniques, and annual yields have been calculated at the weir since HY1963. We reassessed previous estimates of annual suspended sediment loads for HY1963 to HY1975 (Text S1 and Figure S2). Storm loads have been calculated at the weir and the Arfstein station since HY1986 in preparation for the first experimental harvest of the North Fork while storm loads exist for HY2000 onward for XYZ. When suspended sediment storm loads exist for gauging stations upstream of the weir pond (2000–present), we separately accounted for the mass of settled suspended sediment that occurs during and between storms. We calculated the storm mass of settled suspended sediment according to

$$M_s^{\text{storm } i} = M_{\text{Arf}}^{\text{storm } i} + M_{\text{XYZ}}^{\text{storm } i} - M_{\text{NFC}}^{\text{storm } i}, \quad (9)$$

where i denotes the storm number, $M_{\text{Arf}}^{\text{storm } i}$ is the suspended sediment mass measured at the Arfstein station, $M_{\text{XYZ}}^{\text{storm } i}$ is the suspended sediment mass measured at the XYZ station, and $M_{\text{NFC}}^{\text{storm } i}$ is the suspended sediment

Table 1
Description of Values Used to Reconstruct Annual Gravel Yields for the North Fork of Caspar Creek

Value	Frequency of measurement	Description
Pond sediment volume (V_p)	Annual	The surface of the pond sediment was surveyed annually during the summer and accumulation volumes were calculated by differencing surveyed surfaces for consecutive years.
Bed load density (ρ_b)	Once[a]	Sediment from the weir pond delta was collected to determine a characteristic density of bed load deposited in the pond.
Mixed organic matter and settled suspended sediment density ($\rho_{s,o}$)	Once[a]	Sediment samples from multiple locations were collected to determine a characteristic density of mixed organic matter and settled suspended sediment in the pond.
Mass of settled suspended sediment (M_s)	Storms or annual	When possible (HY2000 onward), we calculated M_s directly by differencing suspended sediment storm yields upstream and downstream of the weir pond. Prior to HY2000, we used an estimated settling fraction to calculate the settled suspended sediment mass from the North Fork suspended sediment yield measured at the weir.
Fraction of organic matter in pond sediment (f_o)	Once[a]	Sediment samples from multiple locations were collected to determine a characteristic fraction of organic matter in the pond.
Gravel fraction of the delta (f_{dg})	Once[a]	We collected a bulk sediment sample from the weir pond delta to determine a characteristic gravel fraction.

[a]More frequent measurements may reduce uncertainty.

mass measured at the weir. The total mass of storm-transported, settled suspended sediment that occurred during a given year was $M_s^{\text{storm annual}}$, the sum of $M_s^{\text{storm } i}$ for all storms (i) for a given year.

We estimated the annual interstorm settling yield, $M_s^{\text{interstorm}}$, according to

$$M_s^{\text{interstorm}} = f_s\left(M_{\text{NFC}}^{\text{total}} - M_{\text{NFC}}^{\text{storm annual}}\right), \tag{10}$$

where $M_{\text{NFC}}^{\text{total}}$ is the total annual yield of suspended sediment measured at the weir and includes both interstorm and storm yields, $M_{\text{NFC}}^{\text{storm annual}}$ is the annual yield of suspended sediment for storm events measured at the weir, and f_s is the settling fraction, which we estimated as

$$f_s = \frac{M_s^{\text{storm annual}}}{M_{\text{NFC}}^{\text{storm annual}}}. \tag{11}$$

In this case, f_s equals the fraction of suspended sediment that settled in the weir pond during storm events relative to the fraction of suspended sediment measured at the weir during storm events. Finally, we calculated the total annual settled suspended sediment mass as

$$M_s = M_s^{\text{storm annual}} + M_s^{\text{interstorm}}. \tag{12}$$

For years prior to HY2000, which was the first hydrologic year that suspended sediment yields were measured at the XYZ gauge, we reconstructed the mass of settled suspended sediment ($M_s^{'}$) and substituted $M_s^{'}$ for M_s in equation 6. We estimated $M_s^{'}$ according to

$$M_s^{'} = f_s M_{\text{NFC}}^{\text{total}}, \tag{13}$$

which enabled us to estimate the total mass of settled suspended sediment that occurs during storms and the interstorm period at once. We assumed that the settling fraction was the same for storms and interstorm periods.

We measured bulk densities of bed load ρ_b and mixed settled suspended sediment and organics $\rho_{s,o}$ from pond sediment samples. We collected the sediment samples along 10 approximately evenly spaced transects perpendicular to the long axis of the pond. The average number of samples collected on each transect was 5, but ranged between 1 and 7 with some samples collected near the surface and others up to depths of ~1 m, for a total of 46 samples (Figure 1a). Bed load density was determined from sediment samples collected from the delta while the density of the mixed settled suspended sediment and organics was determined from the remaining sediment samples. The sediment samples were collected by carefully pushing a coring sleeve of known volume into the sediment until the sleeve was flush with the sampling surface and completely filled with no voids. The samples were oven-dried and the sediment density was determined from the dry mass and original sample volume.

We estimated f_o from the same sediment samples used to measure density. The dried samples were separated into coarse (≥2 mm) and fine (<2 mm) fractions. To determine the organic mass of the coarse fraction, we separated the organic material from the lithic material, recorded the masses of each group, and calculated an organic fraction from these masses. To determine the organic mass of the fine grain fraction, we completed loss on ignition analysis (Burt, 2014). We then used the organic fraction and sample masses of the coarse and fine fractions to determine an average organic fraction for each sample. Since the number of samples we collected varied for each non-delta transect, we calculated the mean density of mixed settled suspended sediment and organics ($\rho_{s,o}$) for each of the eight transects and calculated $\rho_{s,o}$ as the mean of these transects. We calculated an average organic fraction (f_o) for each transect and estimated f_o as the average of the 10 transects.

We estimated the uncertainty in annual gravel yields using a Monte Carlo method and propagated the uncertainty through equations 6 and 7. When possible, we used standard deviations and mean values to create distributions for the Monte Carlo uncertainty analysis. In each case, we assumed that the uncertainty was described reasonably well with a normal distribution. For ρ_b, $\rho_{s,o}$, f_o, we treated the reported standard errors, which represent the variability for the total pond, as the annual standard deviations for each respective pond value. For V_p and f_g, which had unknown variance, we assumed that one standard deviation was 20% of the

mean value. For the years for which we reconstructed the settling fraction, we used the standard deviation of the annual setting fraction to create the distribution for M_s. For the years for which the settling fraction was calculated directly from measured suspended sediment yields (HY2000–HY2017), we assumed a standard deviation of 20% of the mean value because the variance for the suspended sediment storm yields, which we used to calculate the settling fraction according to equation 11, has not been quantified. We solved equations 6 and 7 10^6 times with values randomly selected from the distributions and calculated the standard deviation for each annual value of M_g from the resulting distribution.

2.3. Discharge Through the Arfstein Reach

The Arfstein (Arf) gauging station was installed on the main stem of the North Fork in 1983 along a natural channel reach. Originally, the walls of the gauged cross-section were plywood. In 1987, a concrete pad was added and bed load pits were installed. The concrete pad was extended in 1988. In 1999, the plywood walls were replaced with cinder blocks and wooden boards (Figure 1d). Currently, the gauged section has a rectangular geometry that is 4.42 m wide (Figure 1d), and the prior geometry was similar. A series of rating curves were developed from velocimeter measurements at the gauging station, and these were used through HY1999. In HY2000, the rating curves were abandoned in favor of a differencing approach where the discharge at the Arfstein station (Q_{Arf}) was estimated as the difference between the discharge measured at the weir (Q_{NFC}) and the discharge measured in a fiberglass Montana flume on the XYZ tributary (Q_{XYZ}) (Figure 1):

$$Q_{Arf} = Q_{NFC} - Q_{XYZ}. \tag{14}$$

For consistency with the recent approach, we reconstructed discharge through the Arfstein reach for the period before HY2000 (Q'_{Arf}) using the mean fraction of Q_{Arf} to Q_{NFC} for HY2000 to HY2017. To test our reconstruction, we regressed Q'_{Arf} against Q_{Arf} for HY2000 to HY2017.

2.4. Estimating Flow Velocity Through the Arfstein Reach

We estimated the mean flow velocity for the zone of active bed load transport (U_{bed}) with the 10-min discharge record through the respective study reaches, according to a power law relationship,

$$U_{bed} = \alpha Q^{\beta}, \tag{15}$$

where α and β are calibrated parameters. We defined the zone of active bed load transport as the portion of the channel cross-section for which bed load transport is expected to occur for a typical storm event, which is less than or equal to the wetted channel width as it excludes perennially vegetated bars.

No long-term measurements of U were available for the Arfstein reach, so we considered three methods for calibrating equation 15 and determined which method best predicts annual gravel yields and continuous bed load transport rates. We used (1) flow stage and discharge measured at the Arfstein gauging station between HY2000 and HY2017 in conjunction with hydraulic geometry of the engineered Arfstein cross-section (engineered cross-section method); (2) HEC-RAS 5.0.6 modeling software (Brunner, 2016) to predict flow through a natural cross-section (natural x-section method); and (3) field measurements of channel width and water depth collected during storm events between 1978 and 1980 (monitored storms method).

For each approach, we calibrated equation 15 by linearly regressing log-transformed U_{bed} against log-transformed Q using a least-squares approach to determine α and β. In addition, velocity measurements were made with a Price AA current meter periodically between HY1984 and HY1993 from the Arfstein bridge (Figure 1d), which allows for an independent assessment of the velocity-discharge relationship through the engineered Arfstein cross-section.

For the engineered cross-section method, we binned the 10-min record of U_{bed} estimated using the geometry of the engineered cross-section, measured stage, and discharge from equation 14 into 15 bins and regressed the log-transformed U_{bed} against log-transformed Q_{Arf} to develop a rating curve.

The natural cross-section, XS NFC-18, is located ~16 m upstream of the engineered Arfstein cross-section, immediately upstream of a small pool, and immediately downstream of a riffle reach. We used measured stage and discharge as a downstream boundary condition at the engineered cross-section and solved a 1-d steady flow problem to determine mean cross-sectional channel velocity at XS NFC-18. We assumed

commonly used contraction and expansion coefficients of 0.1 and 0.3, respectively, to model the loss in energy due to changes in channel geometry (Brunner, 2016). We estimated a value of Manning's n for the Arfstein engineered cross-section from hydraulic geometry, stage, and discharge data and used this value to inform our choice of an appropriate value of Manning's n for the natural cross-section. We used the geometry at the Arfstein gauging station, surveyed cross-section data collected at the natural cross-section in summer 2015, and a rating curve determined for HY2000 to HY2017 through the engineered Arfstein cross-section to reconstruct U_{bed} through the natural cross-section. We used HEC-RAS 5.0.6 to estimate flow velocity for subcritical flows and to estimate U_{bed} over the observed range of Q from HY2000 to HY2017. We exported modeled values of U_{bed} binned in 0.25 m^3/s increments of Q_{Arf} and estimated α and β from equation 15 using the same log-log regression approach as for the engineered-reach method.

For the monitored storms method, we recorded mean water depth (h) and wetted channel width (w) during seven storm events through the Arfstein reach between HY1978 and 1980. These measurements were made prior to the installation of the Arfstein station, concrete pad, and wood banks. We calculated U_{bed} as

$$U_{bed} = c\left(\frac{Q'_{Arf}}{h^{*}w}\right), \tag{16}$$

where c was a correction factor that accounts for the difference in mean flow velocity of the wetted channel cross-sectional area (U_{chan}) and U_{bed}. We used HEC-RAS 5.0.6 to estimate U_{bed}/U_{chan} for the 2-year return flow for the natural cross-section and set c to this U_{bed}/U_{chan}.

2.5. Data Availability Statement

Discharge, SSC, pond volumes, and bed load transport rate data used in this study are publicly available (Richardson et al., 2020a, 2020b). Additional data including annual suspended sediment yields, reconstructed bed and gravel yields, predicted gravel yields, velocity meter data at the Arfstein station, grain size distributions of the streambed near the Arfstein reach and also for the North Fork delta, profile elevation for the Arfstein reach, cross-section data for NFC-18, water depth and stream width measurements made during storms, and North Fork pond sediment sample data are also publicly available (Richardson & Wagenbrenner, 2020).

3. Predicting Gravel Transport Rates and Calculating Gravel Departures

Although sand is trapped in the weir pond, we focused our analysis on gravel, in contrast to sand which may have been transported in suspension, as we are confident that gravel was delivered as bed load. We modeled gravel (≥2 mm) transport with the two-fraction bed load transport model developed by Wilcock (2001) and implemented it in Mathworks MATLAB 2018a (MATLAB, 2018). We chose this model because it can be calibrated from field measurements and is well-suited for modeling bed load transport for gravel-bed streams (Wilcock, 2001). Given that much of the accuracy of a bed load transport model depends on the calibration instead of the model itself (Wilcock et al., 2009), we focused on predictions from a single transport model and tested three different velocity-discharge relationships as described in section 2.4. With appropriate calibration, we would expect that other bed load transport models (e.g., Bakke et al., 2002; Parker, 1990; Parker et al., 1982; Parker & Klingeman, 1982; Wilcock, 2001; Wilcock & Crowe, 2003) could produce comparable results.

We made minor modifications to the bed load transport model derived by Wilcock (2001). Briefly, transport rate of the gravel fraction per unit bed width (q_g, kg/s/m) is modeled according to

$$q_g = \frac{\left(f_g \rho_s \left(\tau'/\rho\right)^{1.5}\right)}{\left(\rho_s/\rho - 1\right)g} W^*_g, \tag{17}$$

where f_g is the gravel fraction exposed at the surface of the channel bed, ρ is water density, ρ_s is sediment density, g is gravitational acceleration, τ' is the stress imparted by the flow on sediment grains (i.e., skin friction), and W^*_g is a nondimensionalized transport function of the form

$$W_g^* = \begin{cases} 11.2\left(1-0.846\frac{\tau_{rg}}{\tau'}\right)^{4.5} & \tau_{rg}/\tau' \geq 1 \\ 0.0025\left(\frac{\tau'}{\tau_{rg}}\right)^{14.2} & \tau_{rg}/\tau' < 1. \end{cases} \quad (18)$$

The reference shear stress for gravel transport τ_{rg} is the shear stress for which a small, consistent, and observable amount of grain motion occurs. τ_{rg} corresponds to the transport rate for $W_g^* = 0.002$ (Wilcock, 2001; Wilcock et al., 2009). Following the derivation of Wilcock et al. (2009), we calculated the total boundary shear stress (τ_o) by combining Manning's equation,

$$U = \frac{R^{2/3}\sqrt{S}}{n}, \quad (19)$$

where R is hydraulic radius, n is Manning's n which describes channel roughness, and S is bed slope, with the simplified form of the Saint-Venant equation for steady uniform flow,

$$\tau_o = \rho g R S, \quad (20)$$

where ρ is density of water, g is gravitational acceleration, and solved for τ_o:

$$\tau_o = \rho g S^{1/4}(nU)^{3/2}. \quad (21)$$

Equation 20 is a reasonable approximation of the total boundary shear stress for steady uniform flow (Wilcock et al., 2009)—an assumption made here.

We partitioned τ_o and calculated τ' by substituting the Strickler relationship (Wilcock et al., 2009), which relates grainsize (D, m) to the roughness of the sediment on the bed (n_D, s/m$^{1/3}$),

$$n_D = 0.040 D^{1/6}, \quad (22)$$

into equation 21, yielding

$$\tau' = 0.008\rho g(SD)^{1/4}U^{3/2}. \quad (23)$$

For our modeling effort, we explicitly focused on the cross-sectional width for which most bed load transport occurs (w), as is informed by field observations of the bed and banks. The active zone of bed load transport is typically bounded by perennial vegetation, bedrock banks, or stable, large downed wood. We included unvegetated sediment bars that are likely submerged during storm events. Because we focused on the zone of active bed load transport, we substitute U_{bed} for U in equation 23. Following Wilcock et al. (2009), we estimate D as $2D_{65}$ where D_{65} is the 65th percentile of the bed surface grainsize. We calibrated D_{65} from the measured grainsize distribution of the bed surface (Wilcock & Kenworthy, 2002). Substituting U_{bed} for U and $2D_{65}$ for D into equation 23 yields

$$\tau' = 0.0095\rho g(SD_{65})^{1/4}U_{bed}^{3/2}. \quad (24)$$

We calculated the annual gravel yield (Ω_g, kg/year) as

$$\Omega_g = w\int_{t=1}^{m} q_g dt, \quad (25)$$

$t = 1$ is the first measurement of each hydrologic year, m is the last measurement of the year, and equation 25 is solved discretely at 10-min intervals, which corresponds to the interval at which discharge is measured at Caspar Creek. We considered three different values of U_{bed} estimated from the three velocity-discharge relationships defined in section 2.4.

We calibrated equation 24 for the Arfstein reach. D_{65} was estimated from a pebble count along the Arfstein reach and S was calculated from a total station survey of the water surface elevation through the Arfstein reach. For each of the three Q-U_{bed} relationships, we calibrated a single value of τ_{rg}. For each year, we found the value of τ_{rg} that was required to make the predicted annual gravel yields match the reconstructed gravel

yields within a tolerance of one ton. We considered values of τ_{rg} starting at $\tau_{rg} = 0$ and increased τ_{rg} in increments of 0.1 until the tolerance condition was met. For each velocity-discharge approach, we then calculated a single value of τ_{rg} as the mean of the annual values of τ_{rg} required to make the predicted and reconstructed gravel yields match.

We relied on the Shields Number to partially assess our model calibration. The Shields Number (τ^*_{rg}) for the reference shear stress for gravel transport is calculated according to

$$\tau^*_{rg} = \frac{\tau_{rg}}{(\rho_s - \rho)gD_{50}}, \tag{26}$$

where ρ is fluid density, ρ_s is sediment density, g is gravitational acceleration, and D_{50} is the median grain size. Flume experiments suggest that $\tau^*_{rg} \sim 0.03–0.04$ when the surface sediment is mostly gravel (Wilcock, 2001). Also, by comparing the calibrated value of τ_{rg} against values estimated from the Shields equation, we were able to assess the efficiency of bed load transport relative to ideal conditions observed in flume studies (Wilcock, 1998).

We used the Nash-Sutcliffe efficiency (NSE) (Nash & Sutcliffe, 1970) to assess the ability of the model to predict the reconstructed gravel yields. NSE is a normalized metric suitable for comparing model predictions against measurements by comparing the magnitude of residual variance between the predicted and measured values to the measured data variance (Nash & Sutcliffe, 1970). We compared the predicted and reconstructed gravel yields for all years and, in a second comparison, for a subset of years when no major disturbances occurred. Specifically, for the second comparison, we excluded years when timber harvesting occurred, the 5 years after harvesting—which was the period of elevated sediment response after harvesting in the South Fork of Caspar Creek (Lewis, 1998)—and years for which large annual landslide volumes (>1,000 m^3) are known. Large landslides have often led to elevated sediment yields (Cafferata & Reid, 2013; Lewis, 1998; Reid et al., 2017; Rice et al., 1979). The primary second-growth logging period for the North Fork catchment occurred from the spring of 1989 (after the last HY1989 storm) until January 1992. The XYZ subcatchment was harvested from the spring of 1985 (after the last HY1985 storm) until spring 1986. We treat HY1986 and HY1990 to HY1992 as harvest years. We treat HY1987 to HY1989 and HY1993 to HY1997 as potential harvest response years.

By comparing predicted yields to reconstructed yields, it may be possible to assess how sediment supply, storage, and transport efficiency influence bed load transport rates and yields. To assess variations between predicted and reconstructed gravel yields, we calculated an annual departure:

$$\text{Departure} = \Omega_g - M_g. \tag{27}$$

A positive departure indicates overpredicted gravels yields while a negative departure indicates underpredicted yields. We calculated departures when the difference between the predicted gravel yields and reconstructed gravel yields was greater than one standard deviation of the reconstructed gravel yields determined from the Monte Carlo uncertainty analysis.

4. Results

4.1. Reconstructed Annual Bed Load and Gravel Yields

The average annual settling fraction f_s, which is the fraction of suspended sediment mass that settled in the pond relative to the mass of suspended sediment measured at the weir, was 0.28 ± 0.05 (mean ± s.e.) (Figure 3a). For HY2000 to HY2017, the median interstorm suspended sediment yield was 7% of the annual yield and decreased rapidly as annual suspended sediment yields increased. NSE = 0.93 for the comparison of predicted and measured annual settled suspended sediment masses for HY2000 to HY2017 (equation 14) (Figure 3b); NSE = 0.91 for individual storm yields for the same period (Figure 3b).

We measured the bulk density of the pond bed load deposits (ρ_b) as 1.45 ± 0.04 g/cm^3 (mean ± s.e.). The mean density of the mixed settled suspended sediment and organic matter ($\rho_{s,o}$) was 0.86 ± 0.07 g/cm^3 (mean ± s.e.). The organic fraction (f_o) was 0.09 ± 0.02 (mean ± s.e.) for the 10 transects. From analysis of

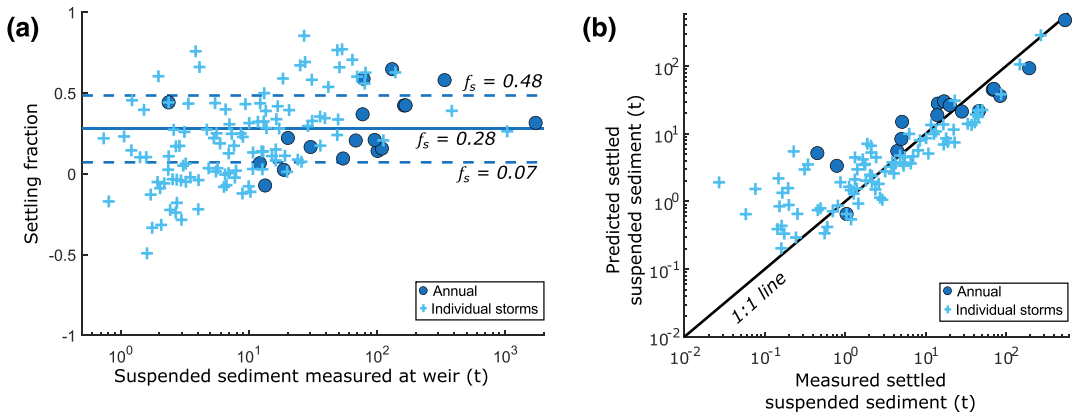

Figure 3. Annual and individual storm suspended sediment that settled in the North Fork weir pond for HY2000 to HY2017. (a) Settling fraction f_s versus suspended sediment measured at the weir for annual values ($M_s^{\text{storm annual}}$) and individual storms ($M_s^{\text{storm } i}$). The solid line shows the mean annual settling fraction and the dashed lines mark the one standard deviation bounds for the annual results. (b) Predicted settled suspended sediment versus measured settled suspended sediment for annual values ($M_s^{\text{storm annual}}$) and individual storms ($M_s^{\text{storm } i}$). The predicted values are predicted by solving equation 11 for $M_s^{\text{storm annual}}$ using $f_s = 0.28$. Negative values, which indicate pond sediment erosion, are not shown in (b).

bulk sediment samples collected from the delta, we determined that the gravel fraction of the delta (f_{dg}) was 0.68.

Thirteen of the 55 reconstructed bed load yields were negative values. This occurred for two reasons. The first reason was that six of the pond surveys yielded negative volume change, which always resulted in negative bed load yields. Negative pond volumes likely occurred due to pond erosion, sediment compaction, decay of organic material in the pond, survey errors, or a combination of these causes (Rice et al., 1979). The other reason for negative bed load yields was that in some years, the total accumulated mass minus the mass of settled suspended sediment and organic matter was negative. Because bed load is unlikely to be transported beyond the weir, we assigned a value of 0 t to negative bed load yields for subsequent analysis.

For HY1963 to HY2017 and excluding years for which pond volumes were negative, bed load accounted for 46% of the mass of material deposited in the pond, while gravel accounted for 31% on average. For the same period, bed load accounted for 22% of the total annual mass of transported material to the pond while gravel accounted for 15%. The corresponding mean annual sediment trap efficiency, which is the percentage of mass trapped in the pond relative to the total transported mass, was 42%.

The three largest bed load yields occurred in HY1965 (974 t), HY1966 (1,055 t), and HY1974 (805 t). These years were also associated with the largest increases in pond sediment volume, but not in the same order. HY1974 had the greatest increase in volume (2,039 m^3) followed by HY1966 (1,184 m^3) and HY1965 (1,006 m^3). The largest suspended sediment yield (3,928 t) on record occurred during HY1974, which contributed to the largest pond sediment volume, but this year produced only the third largest bed load yield (Figure 4). Thus, a disproportionate suspended sediment yield in HY1974 accounts for the exceptional volume of deposition in the pond.

For years prior to HY2000, the uncertainty in the settled suspended sediment mass was most often the largest contributor to the total uncertainty in the Monte Carlo uncertainty analysis. In contrast, for HY2000 onward, the assumed uncertainty of the pond volume was most often the largest contributor to the total uncertainty. One standard deviation from the Monte Carlo analysis ranged from 25–881% for the nonzero annual gravel yields, with a median of 80%. The highest relative uncertainties typically occurred for years with low gravel yields and low total uncertainty (Figure 4).

We used equation 8 to calculate pond sediment density ρ_p for each year. Excluding years when the pond volume was negative, the mean pond density for the 55-year record was 1.19 ± 0.10 g/cm^3 (mean ± s.e.). HY1981 produced an unrealistic pond density of 5.83 g/cm^3, which occurred because the surveyed pond

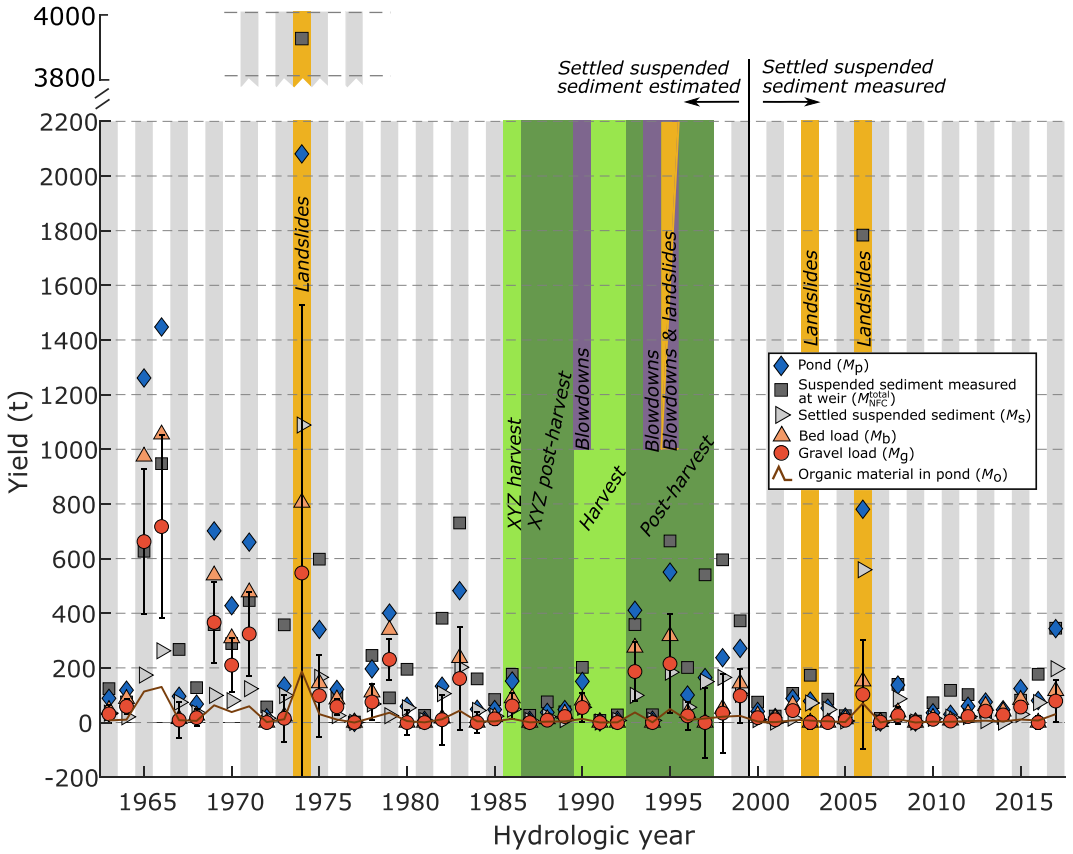

Figure 4. Summary of annual bed load, suspended load, and uncertainty data for the North Fork weir pond. Error bars are gravel load uncertainties and represent one standard deviation calculated from our Monte Carlo error analysis. Years with major disturbances are marked with yellow, green, or purple bars. The HY1974 suspended sediment yield measured at the weir (M_{NFC}^{total}) is plotted above the scale used for the other data.

volume (1.4 m^3) was small relative to the reconstructed settled suspended sediment yield (7.2 t). The next two highest estimates of ρ_p were 1.62 t/m^3 (HY2007) and 1.34 g/cm^3 (HY2014). When the HY1981 outlier and the years with negative pond volumes were excluded, the mean ρ_p was 1.10 ± 0.02 g/cm^3 (mean ± s.e.), which also equaled the median ρ_p when just the negative pond volumes were excluded. We suggest that ρ_p= 1.10 g/cm^3 is the better of the two long-term means reported here and used that value in subsequent analysis and discussion.

4.2. Discharge Through the Arfstein Reach

We found that a fraction of 0.85 ± 0.03 (mean ± s.d.) of the 10-min discharge >0.69 m^3/s measured at the North Fork weir for HY2000 to HY2017 flowed through the Arfstein reach (Figure S1). The relationship between Q_{Arf} and Q_{NFC} exhibited considerably less variability above the storm threshold of Q_{NFC} of 0.69 m^3/s than below it (Figure S1). The regression of reconstructed discharge (Q'_{Arf}) against measured discharge (Q_{Arf}) for HY2000 to HY2017 resulted in a R^2 of 0.999 (Figure S1b). Discharge through the Arfstein reach varied from nearly 0 m^3/s during low flows that occurred some years to a peak reconstructed discharge of 7.3 m^3/s in January 1974.

4.3. Estimated Flow Velocity Through the Study Reach

Of the three velocity-discharge relationships for the Arfstein reach (Figure 5, Table 2), the engineered cross-section approach produced the highest velocity per discharge for the range of Q_{Arf} during the study period (HY1963–HY2017). For Q_{Arf} below about 3.5 m^3/s, the monitored storm method produced the lowest

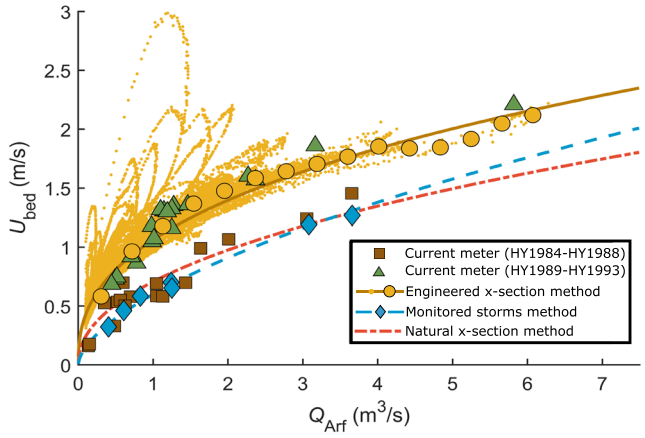

Figure 5. Velocity-discharge relationships through the Arfstein reach. For the engineered cross-section method, the small data points are calculated from the 10-min discharge record and the large data points are the binned results for the 10-min discharge record. The log-transformed regressions are plotted as curves for each velocity-discharge calibration method.

estimates of U_{bed}. For Q_{Arf} greater than about 3.5 m^3/s, the monitored storm method produced values of U_{bed} that were intermediate between the engineered cross-section and natural cross-section methods.

For the engineered cross-section method, we excluded $Q_{Arf} < 0.1$ m^3/s in our calibration as U_{bed} was prone to high uncertainty at low flows (Figure 5). For the engineered cross-section method, the mean flow velocity through the Arfstein station cross-section equaled U_{bed} as the entire width of the channel experiences bed load transport. The estimated U_{bed} for the engineered cross-section approach was similar to velocities measured for HY1989 to HY1993 with a current meter, which were higher than the velocimeter measurements from HY1984 to HY1988 (Figure 5). Ultimately, we found a good fit for the binned values of log-transformed U_{bed} against log-transformed Q_{Arf} ($R^2 = 0.969$) for the engineered cross-section method.

For the natural cross-section method, we used the reach-average slope (0.014), binned Q_{Arf} and U_{bed} values from the engineered cross-section method, and the Arfstein station channel width (4.42 m) to estimate a mean value of Manning's n of 0.036 ± 0.004 (mean ± s.d.) using equation 19. Because XS NFC-18 is across a natural reach, we assumed the value at the high end of the range (0.04), which is within the typical range of Manning's n values for mountain streams (Chow, 1959). The regression of the binned values of log-transformed U_{bed} estimated from HEC-RAS against log-transformed Q_{Arf} also yielded a good fit ($R^2 = 0.995$). Large pieces of downed wood are often found on the banks of channels in the North Fork and interact with high flows. For XS NFC-18, a log was oriented parallel to the flow direction on the left bank, and we used the log as the left boundary for defining the active bed load channel width. The opposing bank was an undercut bedrock ledge. The active width for bed load transport through this cross-section appeared similar to other sections in the Arfstein reach.

For the monitored storm method, we primarily relied on field measurements from seven monitored storms to estimate U_{bed} according to equation 16. Using HEC-RAS, we estimated c as 1.07 for the 2-year return flow through XS NFC-18, and we assumed this to be representative of a natural cross-section in the Arfstein reach. $R^2 = 0.987$ for the regression of log-transformed U_{bed} against log-transformed Q'_{Arf}.

4.4. Predicted Gravel Yields

We considered three different model formulations for predicting gravel transport, with each formulation relying on and named for a different velocity-discharge relationship (Table 2). Across the three approaches, predicted transport rates were most similar for discharges between about 2.5 and 3.5 m^3/s, which brackets the 2-year return-period flow (3.03 m^3/s) (Figure 6). For the 2-year return flow, the predicted gravel transport rates ranged from 58 to 62 kg/min/channel width (all reported transport rates are for the entire width of the channel unless otherwise noted). None of the velocity-discharge relationships consistently predicted the highest gravel transport rate per unit discharge. For $Q_{Arf} < 0.88$ m^3/s and 2.74 m^3/s $< Q_{Arf} <$ 3.49 m^3/s, the engineered cross-section GTM predicted the highest gravel transport rates. For 0.88 m^3/s $< Q_{Arf} <$ 2.74 m^3/s, the natural cross-section GTM led to the highest transport rates,

Table 2
Summary of Parameters for Equation 15 and Width of Active Bed Load Transport (w). Values Reported as Mean ± Standard Error

Method	α	β	w (m)	Notes
Engineered cross-section	1.06 ± 0.03	0.39 ± 0.02	4.42	Calibrated from geometry, stage, and discharge estimated with equation (14) for HY2000 to HY2017.
Natural cross-section	0.71 ± 0.01	0.46 ± 0.01	3.62	Calibrated for XS NFC-18, a natural reach immediately upstream of the Arfstein gauging station, using data from HY2000 to HY2017.
Monitored storms	0.60 ± 0.01	0.60 ± 0.03	3.62	Calibrated from seven sets of field measurements made during storm events that occurred during HY1978 to HY1980. w is assumed to be the same as for XS NFC-18.

Table 3
Summary of Values for Bed Load Model Calibration and Assessment

Variable or parameter	Value	Notes
Sediment density ($\rho_{sediment}$)	2.6 g/cm^3	Average density of coastal belt of Franciscan complex (Bailey et al., 1964).
Channel slope (S)	0.014	Mean gradient through Arfstein reach measured over a distance of ~90 m and calculated from water surface elevation data collected during low flow conditions (summer 2018).
D_{50}	1.7 cm	Calculated from pebble count of bed surface through Arfstein reach (1993).
D_{65}	2.3 cm	
Gravel fraction at bed surface (f_g)	0.91	Determined from pebble count of bed surface through Arfstein reach (1993). f_g reflects the fraction ≥4 mm as the pebble count did not differentiate grains <4 mm.

although they were only slightly greater than the engineered cross-section values. For $Q_{Arf} > 3.49$ m^3/s, the monitored storms GTM predicted the highest rates.

For Q_{Arf} less than 0.69 m^3/s (the storm threshold), the predicted transport rates were very small for all of the calibrated models (<10^{-3} kg/min/channel width), but predicted transport rates increased rapidly as discharge increased (Figure 6). At 7.3 m^3/s, the highest recorded discharge, the monitored storms model predicted a transport rate of 894 kg/min/channel width, which was 23% higher than the engineered cross-section GTM (726 kg/min/channel width) and 62% higher than the natural cross-section GTM (553 kg/min/channel width). All of the models overpredicted the yields for HY1974 (Figures 7b and S3), which was the stormiest year on record and when a major landslide (3,306 m^3) delivered much sediment directly to the North Fork channel (Rice et al., 1979), and the year of the greatest annual suspended sediment load (3,928 t).

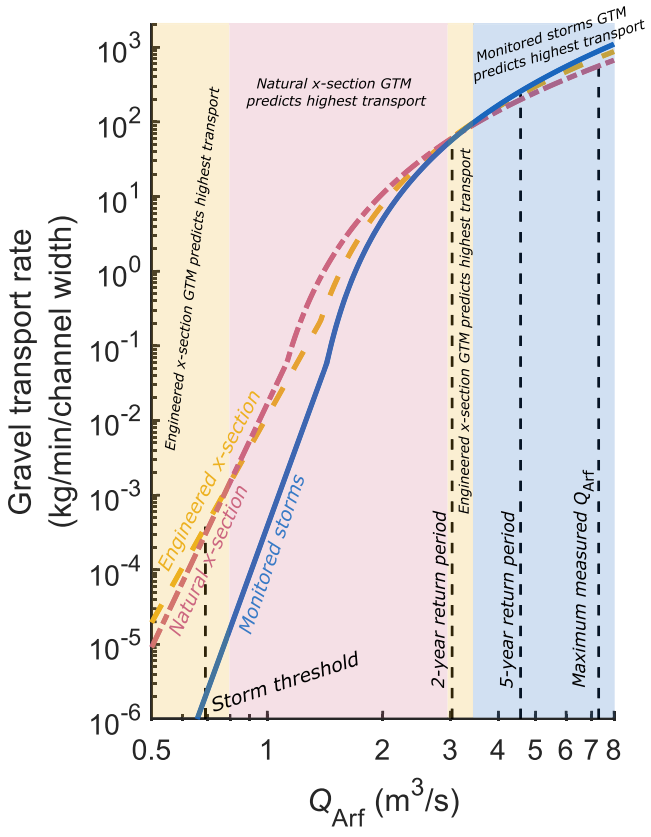

Figure 6. Comparison of transport rates for the three gravel transport model (GTM) formulations: the engineered cross-section GTM, the natural cross-section GTM, and the monitored storms GTM. Two-year and 5-year return-period flows were determined from the HY1963 to HY2017 annual maxima at Arfstein reach and a log-Pearson Type III distribution. The shading color corresponds to the model that predicted the highest transport rate for a particular range of Q_{Arf}.

A comparison of reconstructed and predicted annual gravel yields resulted in low values of NSE for all three GTMs (Figures 8, S4, 7, and S3, Table 4). When years with major disturbances (section 3) were excluded, NSE for reconstructed and predicted gravel yields increased substantially (Figures 8, S4, and Table 4), and the monitored storms methods produced the highest NSE (0.74) of the three models (Table 4). HY1974 substantially contributed to the misfit between the predicted and reconstructed yields. When HY1974 was excluded from the annual yields, NSE increased substantially for all of the models, and specifically for the monitored storms GTM, NSE increased from 0.06 to 0.47. Because the monitored storms method produced the best fit for years excluding major disturbances, we focus on the results for this GTM from here forward.

We compared our calibrated values of τ_{rg} with values expected from theory and flume experiments (e.g., Wilcock, 1998, 2001). D_{50} of the bed through the Arfstein reach was 1.7 cm (Table 3), which produced a range of τ_{rg} of 8.0–10.7 Pa using equation 26 for $\tau^*_{rg} = 0.03$–0.04, which is a typical range of τ^*_{rg} for gravel-bed streams (e.g., Wilcock, 1998, 2001). The best-fit model resulted in τ_{rg} of 8.0–16.4 Pa (Table 4), producing a reasonable match to values expected from flume experiments (e.g., Wilcock, 1998, 2001).

4.5. Comparison of Calibrated Model Results With Continuous Bed Load Transport Rates

Comparing gravel transport rates predicted by the calibrated models to those measured by continuous bed load pit sampling from HY1988 to HY1995 indicates the accuracy of the models within the range of variability of transport rates for a given flow magnitude (Figure 9). The largest

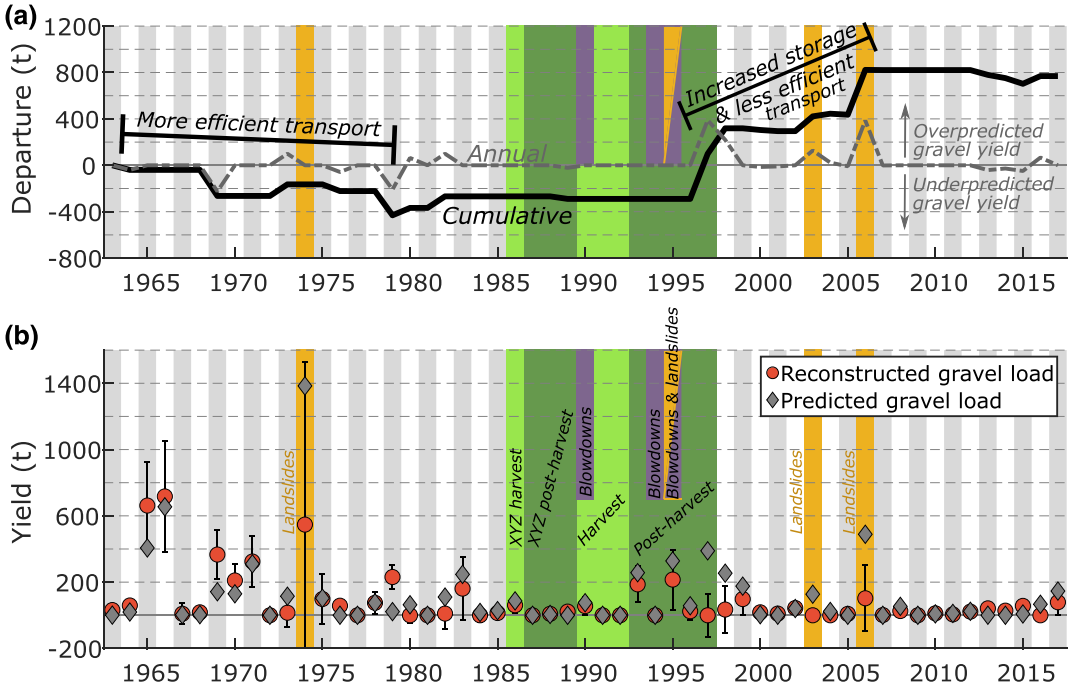

Figure 7. (a) Annual and cumulative departures for the gravel yields (departure = predicted gravel yield − reconstructed gravel yield). Departures greater than one standard deviation of the reconstructed yields are shown. (b) Predicted annual gravel yield (Ω_g) and reconstructed annual gravel yield (M_g). The uncertainties in the reconstructed yields are shown as one standard deviation and calculated from the Monte Carlo uncertainty analysis. Years with major disturbances are marked with yellow, green, or purple color bars and are described in (b). The results for the engineered cross-section approach and the natural cross-section approach are shown in Figure S3.

flow measured while the bed load pits were active (3.19 m^3/s) was only slightly above the 2-year return period discharge (3.03 m^3/s). Given that limitation, however, a wide range of transport rates per skin friction were available to compare GTMs to measured transport rates. For transport rates greater than about 0.3 kg/min/channel width, all of the GTMs predicted transport rates that fell within the range of transport rates measured with the bed load pits (Figures 9 and S5). The monitored storms GTM underpredicted gravel transport for rates below about 0.3 kg/min/channel width. By decreasing τ_{rg} by 20%, predicted rates better matched bed load pit transport rates at low transport rates (below 0.3 kg/min/channel width).

In general, transport efficiency, which is represented in the relationship between transport rate and skin friction, decreased from HY1988 to HY1995. Bed load transport efficiency is a transport rate relative to a theoretical limit or other reference value for a given parameter of impelling force, e.g., skin friction or stream power (Bagnold, 1966; Gomez, 2006). Relatively high transport rates indicate high efficiency. All else being equal, decreased sediment supply, bed coarsening (often due to bed armoring), or increased sediment storage capacity could lead to a decrease in transport efficiency. Hassan et al. (2014) suggested that bed coarsening, increased downed wood, and reduced sediment supply were responsible for the decrease in bed load transport rates per shear stress measured from HY1988 to HY1995 for the North Fork.

The bed load pit samplers collected sand, gravel, and some finer sediment. For all samples for which the grain size was determined, gravel accounted for at least 50% of the total sample and was often greater than 75% (Figures 9 and S5). Samples collected for the highest transport rates typically had gravel fractions >75%. The discrepancy between the predicted gravel yields and the measured bed load transport rates at low skin friction values may be because the pit sampler transport measurements included sand, especially as lower transport rates often had higher fractions of sand than those at high transport rates. At low discharges, where the mismatch between the predictions from the monitored

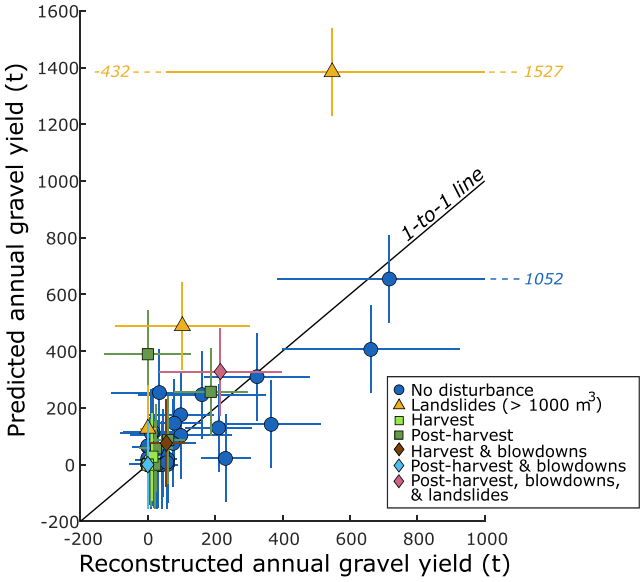

Figure 8. Predicted annual gravel yields (Ω_g) versus reconstructed annual gravel yields (M_g) for the monitored storms GTM. The uncertainties in the reconstructed yields are shown as one standard deviation and calculated from the Monte Carlo uncertainty analysis; the uncertainties in the predicted yields are shown as one RMSE and calculated from the residuals between Ω_g and M_g for all years (Table 4). The results for the engineered cross-section approach and the natural cross-section approach are shown in Figure S4. Where the uncertainty bars extend beyond the plot *x*axis limits, the uncertainty bars are dashed and the value written next to the dashed line.

storms GTM and the measured transport rates occurred, sand was likely preferentially transported relative to gravel (e.g., Ferguson et al., 1996; Hassan et al., 2014; Jackson & Beschta, 1982), which may explain the discrepancy.

4.6. Comparison of Reconstructed and Predicted Gravel Yields

By comparing predicted gravel yields to reconstructed gravel yields, we identified changes in transport behavior from HY1963 to HY2017. Plotting the departures that are greater than one standard deviation for the reconstructed gravel yields revealed four primary trends (Figures 7 and S3). From HY1963 to HY1979, the departures were mostly negative, reflecting the underpredicted yields during this period and suggesting an increase in gravel transport efficiency to the weir pond. From HY1980 to HY1996, the departures were small. From HY1997 to HY2006, the departures were primarily positive, reflecting overpredicted yields and decreased gravel transport efficiency to the weir pond. From HY2007 to HY2017, the departures were again small. In general, there was a shift from mostly negative departures at the beginning of the study to mostly positive departures toward the latter half of the study.

For HY1963 to HY1979, the engineered cross-section and the natural cross-section GTMs produced larger underpredictions in gravel yields (669 and 714 t cumulative departure, respectively; Figure S3) relative to the monitored storms GTM (431 t cumulative departure; Figure 7). For HY1997 to HY2006, the cumulative departures for the three models were similar and ranged from 1,084 t for the natural cross-section GTM to 1,153 t for the engineered cross-section GTM (Figures 7 and S3). Although these cumulative departures for HY1997 to HY2006 were less than twice the maximum annual gravel yields, they were 12 to 13 times the average annual gravel yield of 87 t.

5. Discussion

5.1. Assessment of the Bed Load Reconstruction

Reconstructing long-term yields is significantly benefited from temporally consistent and accurate data. As the reader might expect, changes in data collection at Caspar Creek have occurred over the 55-year period investigated here. However, these changes—like the switch in suspended sediment measurements from fixed-stage samplers to pumped samplers in HY1976—did not hamper our ability to reconstruct bed load yields and assess the quality of the reconstruction. These changes did require us to reassess some prior data and analyses, in particular annual suspended sediment yields from the beginning of the study. Although it may be tempting to infer bed load yields from surveyed pond sediment volumes alone, high variability in suspended sediment transport and, by extension, settled suspended sediment in the pond can complicate the assessment. In particular, years with high suspended sediment loads often corresponded to years with large, infrequent storms or periods following major disturbances, further challenging a simpler approach than used here. The occurrence of the largest increase in volume of pond sediment (HY1974) but third greatest bed load yield confirms our concern that bed load yields cannot always be inferred directly from pond volumes. The technique that we derived for reconstructing bed load yields from a combination of annual increment of pond sediment volume, settled suspended sediment, and organic matter was designed to address these challenges. Given the relative robustness of our approach to different sampling methods and catchment responses, we suggest that our reconstruction technique can be adapted to other sites.

For years prior to HY2000, we assumed a constant settling fraction, but our analysis suggests that this fraction likely varied each year (Figure 3). Estimating a single settling fraction is not unlike assigning a single trap efficiency value to a reservoir, which is often done (e.g., Brune, 1953; Heinemarm, 1981; Verstraeten & Poesen, 2001). In practice, it is best to directly account for the mass of settled suspended sediment as we

Table 4
Gravel Transport Model Results

Velocity-discharge calibration	τ_{rg} (Pa)	All years		Years excluding major disturbances	
		NSE	RMSE	NSE	RMSE
Engineered cross-section	16.4	0.21	142	0.69	90
Natural cross-section	8.2	0.38	125	0.62	100
Monitored storms	8.0	0.06	155	0.74	83

Note. The Nash-Sutcliffe efficiency (NSE) and root-mean-square error (RMSE) are shown for each modeling approach for all years and years excluding major disturbances.

did for HY2000 to HY2017. Our Monte Carlo analysis suggested that the uncertainty in predicting the mass of settled suspended sediment usually dominated the total error when large suspended sediment yields occurred. For years after HY2000, the uncertainty in the settled suspended sediment mass was greatly reduced due to our direct accounting of the mass of settled suspended sediment, which increased our confidence in the gravel yields from HY2000 onward. However, few other studies will have the data necessary to directly calculate the deposited fraction of suspended sediment.

Some opportunities exist for comparing our results to previous results from Caspar Creek, and we consider them here. A pond density of 1.185 g/cm^3 for the Caspar Creek weir deposits was used in prior reports (e.g., Lewis, 1998; Rice et al., 1979), but few details exist on how this estimate was originally made. Reassuringly, our mean pond density for the 55-year record of 1.10 ± 0.02 g/cm^3 (mean ± s.e.) was only slightly less than the prior value. Similarly, an earlier study estimated the gravel fraction to be 26% of the pond sediment (Napolitano, 1996). This percentage reflected the average for transport events that occurred shortly before the estimate was determined. For HY1963 to HY2017, we determined that 31% of the pond sediment was gravel. Beyond temporal variability, an explanation is that Napolitano (1996) used gravel fractions collected along a long-axis transect of the pond and did not account for variable sediment thickness, such as in the delta. We found that the delta deposits, which were mostly gravel, were considerably thicker than the remaining deposits such as in the tongue (Figure 1). The increased thickness of gravel in the delta likely led to an underestimate of gravel using Napolitano's (1996) approach. Despite the difference in gravel pond volume, using values reported by Napolitano (1996), gravel was 14% of the total annual yield, which was very similar to our long-term averaged value of 15%.

Cafferata and Spittler (1998) used values reported in Napolitano (1996) to estimate that 20% of the total suspended sediment settles in the weir pond. Using a different approach, Lewis (1998) estimated that 40% of the total suspended sediment settles in the weir pond. We calculated that the annual fraction of suspended sediment that settles in the pond relative to that entering the pond ranged from −15% to 40% and averaged 19%.

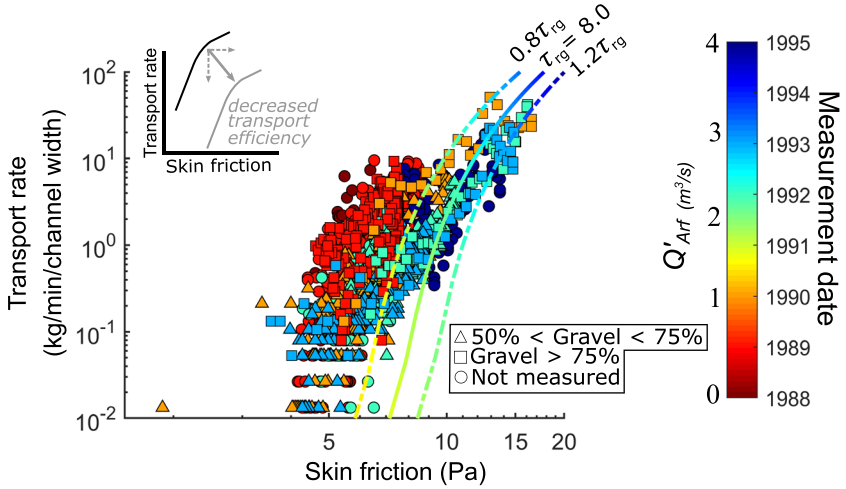

Figure 9. Predicted transport rates versus skin friction at the Arfstein reach for the monitored storms method. We calculated skin friction (τ') with equation 24 for each velocity-discharge relationship using 0.85 * Q_{NFC} to estimate Q through the Arfstein reach (Q'_{Arf}). The modeled transport relationship for the best-fit value of the reference shear stress for gravel transport (τ_{rg}) is plotted as a solid line, and the transport relationships for a 20% change in τ_{rg} are plotted as dashed lines. The color bar indicates Q'_{Arf} for the gravel model (lines) and the measurement date for the bed load pit measurements (symbols). The symbol shape indicates the gravel percentage when measured. Measured transport rates are for a channel width of 4.42 m at the Arfstein station. The inset shows the expected shift of transport rates downward or to the right for a decrease in sediment transport efficiency. The results for the engineered cross-section approach and the natural cross-section approach are shown in Figure S5.

Our mean value agrees closely with the value reported by Cafferata and Spittler (1998) while the upper annual estimate for our range matches Lewis' (1998) estimate.

5.2. Assessment of GTM for Caspar Creek

Choosing an appropriate cross-section for which to model bed load transport is an important decision as spatial variability can lead to a wide range of predicted transport rates (Wilcock et al., 2009). Unfortunately, practitioners are often limited to modeling transport to cross-sections where velocity-discharge data exist or bed load transport rates have previously been measured. Of the three velocity-discharge relationships that we investigated, the velocity-discharge relationship through the engineered cross-section predicted the highest velocity for a given discharge (Figure 5). However, somewhat counterintuitively, the engineered cross-section method did not lead to the highest predicted transport rates due to the higher τ_{rg} required to calibrate the GTM (Figure 6).

We suggest that the GTM based on the monitored storms velocity-discharge relationship was the best approach of the three that we evaluated. Unfortunately, field measurements of stream width and depth for a wide range of flow conditions are often lacking in other studies. Using the available discharge, stage, and cross-section geometry for a natural cross-section and determining the velocity-discharge relationship using HEC-RAS yielded fairly accurate gravel transport predictions and required less data than our preferred method. If data are incomplete or if a monitored cross-section is deemed unsuitable, as was the case of our partially engineered Arfstein station cross-section, we demonstrated that reconstructing a velocity-discharge relationship for a nearby cross-section is a viable approach.

When τ_{rg} was calculated excluding the disturbance years, the comparison between predicted and reconstructed yields improved dramatically. One explanation for this result is that large landslides led to decreased bed load yields because they delivered coarse hillslope material to the channel, and this material was more difficult to transport. However, bed coarsening due to landslides is an unlikely explanation for the overpredicted yields, especially since some of the slides occurred a large distance from the weir pond. For example, the 1974 slide was ~2 km upstream of the weir pond. Barring the possibility that large landslides decreased transport, we consider an alternative explanation. Large landslides typically occurred during stormy years when large departures in absolute terms were also more likely to occur. For example, it is possible that large downed wood from earlier blowdown events trapped significant quantities of sediment in HY2006 causing the departure that year. It may have been coincidental that a major landslide also occurred in HY2006. Yet another explanation is that the quality of the annual gravel yield reconstruction was lowest for years when large landslides occurred (Figure 4). High suspended sediment yields increased the uncertainty in the reconstruction especially for years prior to HY2000 when we estimated the settling fraction of suspended sediment (Figure 4). Even for years when the fraction of suspended sediment that settled was known, inaccuracy in the density of combined suspended sediment and organic material would have a larger influence on gravel yield accuracy relative to years with lower suspended sediment yields. Given these different factors, the concurrence of large landslides and stormy years, which were both more likely to experience departures, contributed to the large departures instead of landslides being directly responsible for the overpredicted yields.

Our calibrated reference shear stress τ_{rg} for the monitored storms GTM agreed well with gravel transport rates measured in laboratory flumes for mixed sand and gravel beds (e.g., Wilcock, 1998). $\tau_{rg} = 8.0$ Pa corresponds to a Shields Number for the reference shear stress for gravel transport τ^*_{rg} of 0.030 when calibrated using values measured along the Arfstein reach of the North Fork (Table 4). Under conditions when supply is limited, we would expect higher values of τ_{rg} to calibrate the GTMs. The chosen velocity-discharge relationship has a large influence on the interpretation of the calibrated value of τ_{rg}. The engineered cross-section GTM required a considerably higher τ_{rg} of 16.4 Pa, which corresponds to τ^*_{rg} of 0.061, which is higher than expected for a pure gravel or mixed sand-gravel bed for idealized transport conditions (Wilcock, 1998). High values of τ^*_{rg}, like the value that we estimated for the engineered cross-section GTM, may indicate low sediment supply.

In a compilation of 45 gravel-bed streams and rivers, Mueller et al. (2005) found that τ^*_{rg} ranged from 0.01 to 0.12, with a median of 0.04 and decreased with downstream distance. Our estimates of τ^*_{rg} are within

the reported range by Mueller et al. (2005). Lamb et al. (2008) reported a wide range of critical Shields stress (τ_c^*) estimated from field and flume data. Estimates of τ_c^* should be slightly less than τ_{rg}^* when calculated for the same site and conditions as τ_c^* is the threshold of motion for bed grain motion as opposed to τ_{rg}^* which describes a small amount of transport. From analysis of the flume and field data, Lamb et al. (2008) estimated $\tau_c^* = 0.15 S^{0.25}$ where S is the channel gradient. Using the North Fork channel slope (0.014), τ_c^* is 0.052, which is slightly less than τ_{rg}^* for the engineered cross-section GTM as expected since τ_c^* should be less than τ_{rg}^*. Although the estimate of τ_c^* using the approach of Lamb et al. (2008) is 73% higher than the calibrated τ_{rg}^* for the monitored storms GTM. This discrepancy, in addition to the large variability in τ_c^* reported in Lamb et al. (2008) even for a specific channel gradient, should encourage a calibration approach where τ_{rg}^* or τ_c^* is calibrated against measured transport rates or yields whenever possible.

The range of τ_{rg}^* that we found for the different GTMs highlights an important issue: the chosen cross-section and resulting velocity-discharge relationship has a substantial impact on the calculated τ_{rg}^* and interpretation of bed load transport behavior. Velocity-discharge relationships can vary considerably even for a single reach (Yager et al., 2018) and should encourage a careful assessment of estimated values of τ_{rg}^*. Since the North Fork engineered cross-section was designed to convey sediment efficiently, it is not surprising that we inferred sediment supply limits for the engineered cross-section. However, supply limitations for the engineered cross-section do not imply supply limitations elsewhere. $\tau_{rg}^* = 0.030$ for the monitored storms GTM and $\tau_{rg}^* = 0.031$ for the natural cross-section GTM are in good agreement with ideal transport conditions for a mixed sand-gravel bed (Wilcock, 1998). We suggest that supply limits are not typical along the North Fork main stem based on the good agreement between τ_{rg}^* from flume experiments and the calibrated values of τ_{rg}^* for the monitored storms and natural cross-section GTMs.

5.3. Comparison of Reconstructed Yields, Bed Load Predictions, and Continuous Transport Rates

We found that the predicted bed load transport rates compared well to transport rates measured with the bed load pits. Importantly, the model results which we compared in Figures 9 and S5 were calibrated from the reconstructed yields and field measurements and were independent of the short-term continuous record of transport rates. This agreement between independent observations suggests remarkable consistency between our reconstructed gravel yields, model-predicted gravel yields, and measured bed load transport rates. However, we also recognize that a wide range of measured bed load transport rates per skin friction (Figure 9) provided some leniency in comparing our predictions with measurements. Our predicted transport rates matched measured bed load transport rates best during the latter half of the bed load transport monitoring phase, when transport efficiency was lowest.

Although the monitored storms method produced the best match to the reconstructed yields (Table 4), a weakness of this approach is that it appears to underestimate transport rates at low skin friction (Figure 9). The mismatch at low skin friction may have been due to the discrepancy between predicting gravel transport rates and measuring all bed load transport rates, including sand and finer sediment, in the bed load pits. However, even this possible mismatch did not lead to a large difference in annual yields, as transport at the relatively low discharges associated with the low skin friction values did not contribute substantially to annual loads.

One reason why the monitored storm method and natural cross-section method required calibration with lower values of τ_{rg} than those predicted by the Shields equation (equation 26) is that we assumed all the bed load deposited in the weir pond had been transported through the Arfstein reach. However, the XYZ tributary joins downstream of the Arfstein gauging station, and therefore probably contributed some of the bed load in the weir pond. We assumed that the bed load contributed from the XYZ tributary was minor and chose not to differentiate gravel transported to the pond from the XYZ tributary from gravel transported through the Arfstein reach. The XYZ catchment is 17% of the total North Fork catchment area and previous field observations suggested that bed load transport from the catchment was relatively minor. If the XYZ catchment contributed a gravel load that was proportional to its catchment area, the gravel load transported through the Arfstein reach would be 17% less than the gravel yield in the weir pond. Explicitly accounting for the bed load contribution from the XYZ catchment would have increased the best-fit values of τ_{rg} for the

three GTMs. In that case, the monitored storms method and natural cross-section method would produce comparisons that still better-matched the values of τ_{rg} predicted from the Shields equation. However, the best-fit τ_{rg} for the engineered cross-section method would also increase, further overpredicting the value of τ_{rg} expected from the Shields equation. Considering how accounting for the XYZ catchment would influence best-fit values of τ_{rg} offers further support that the engineered-reach GTM was less suitable for describing the transport conditions through the Arfstein reach, although the best-fit value of τ^*_{rg} for the engineered-reach GTM was still within the measured range of values reported by Mueller et al. (2005).

5.4. Comparison of Caspar Creek Sediment Accumulation Yields to Other Sites

Several other sites around the world have reported sediment yields from impoundments, including weir ponds, without explicitly accounting for organic matter and settled suspended sediment (e.g., Johnson & Rothacher, 2019; Leeks & Marks, 1997; Troendle & Olsen, 1994). Dendy and Champion (1978) compiled annual sediment accumulations for over 1,750 ponds and reservoirs from data that were available in 1975 (including data for the North Fork of Caspar Creek from 1963 to 1970).

In this compilation, for 41 watersheds in Northern California and coastal southern Oregon, accumulation volume normalized by contributing area ranged from 0.05 to 354 m^3/km^2/year. For 1963 to 1970, Caspar Creek yields ranged from 14 to 247 m^3/km^2/year, which suggests that Caspar Creek accumulation rates are within the observed range for Northern California and southern Oregon. However, the accumulation volume for HY1974 for the North Fork was 426 m^3/km^2/year, which is higher than all of the yields reported for Northern California and southern Oregon by Dendy and Champion (1978). Many of the accumulation rates reported for weir ponds in other parts of the United States were larger than the HY1974 accumulation volume (Dendy & Champion, 1978). As previously mentioned, a large landslide (3,306 m^3) reached the North Fork channel in HY1974 causing elevated yields (Rice et al., 1979), which is likely why the HY1974 yield was elevated relative to the yields reported for other regional sites by Dendy and Champion (1978).

Ultimately, the exceptionally long record of accumulation rates at Caspar Creek increased the probability that an annual accumulation volume for the North Fork would be outside the range reported by Dendy and Champion (1978) and serves as strong evidence for the value of long, continuous records. Of the Northern California and southern Oregon sites, the North Fork and South Fork weir ponds at Caspar Creek were the best represented in Dendy and Champion (1978) which included 8 years of annual records whereas the other Northern California and southern Oregon sites typically only had 1 or 2 years of sediment accumulation data. If the 39 other sites had records of equal length (55 years) to the North Fork weir pond, we expect that the North Fork of Caspar Creek would not seem exceptionally high as very large and very small transport yields would be likely to occur given a similar observation record.

5.5. Detecting Changes in Sediment Supply, Storage, and Transport Efficiency

A decrease in gravel-transport efficiency, expressed as downward shifting transport curves, was measured at the bed load pits from HY1988 to HY1995 (Figure 9). Decreased gravel-transport efficiency is also reflected in the positive departures between the modeled and reconstructed annual gravel yields from HY1993 to HY2006, although the departures greater than one standard deviation only occurred in HY1997, HY1998, and HY2006 (Figure 7). It is important to distinguish transport capacity from transport efficiency to interpret year-to-year variations in gravel transport rates and yields. Regarding transport capacity, Mackin (1948) recognized that minor departures from equilibrium could occur around an average transport condition existing over a "period of years," but that the variations did not necessarily indicate a change in transport capacity. We conclude that year-to-year fluctuations in bed load yield from the North Fork of Caspar Creek are due in part to changes in transport efficiency. Longer-term changes in transport trends over a "period of years" in conjunction with knowledge of equilibrium conditions would be necessary to assess if there was a change in transport capacity.

One explanation for the departures that we measured between HY1997 and HY2006 is that sediment supply to the channel decreased. This seems unlikely in consideration of observations indicating that sediment supply to the North Fork main stem increased in the 1990s and 2000s. Landslide occurrence along roads increased by a factor of three in harvested areas relative to unharvested areas and by an order of magnitude for moderate to large landslides (98 to 4900 m3) in harvested areas after harvesting (Reid & Keppeler, 2012). Of the five largest recorded landslides in the North Fork catchment since 1962, four occurred during or after

1995 (Reid et al., 2017). Only one large landslide was recorded before 1985, when an improved inventory of landslides began. Nonetheless, the increased occurrence of landslides witnessed in the late 1990s and early 2000s likely contributed to increased sediment supply. Moreover, Reid et al. (2010) concluded that much of the sediment transported along the North Fork was produced by headcut erosion of tributary channels, which increased in the 1990s due to increased discharge following harvesting. Both increased landslide occurrence and headcut extension point to increased sediment supply during the 1990s and early 2000s, which makes a reduction in sediment supply to the channel an unlikely explanation for the decrease in transport efficiency observed at the bed load pits from HY1988 to HY1995 (Figure 9) and for the positive departures in annual gravel yields from HY1997 to HY2006 (Figure 7).

A stronger argument can be made that increased wood loading during the 1990s after logging led to a decrease in transport capacity and positive departures in gravel yield as much of the bed load bound for the weir pond was deposited behind newly formed jams and wood steps and around form roughness created by other wood in the channel. Implementation of best management timber harvesting practices during the North Fork harvest, specifically, watercourse protection zones, limited direct disturbance to the channel (Cafferata & Reid, 2013; Lewis, 1998). Despite the lack of direct impact of logging operations, indirect increases of wood input occurred following a series of major blowdowns in the winters of 1990, 1994, and 1995 (Reid & Hilton, 1998). These blowdowns substantially increased the quantity of large downed wood along the North Fork main stem, which consequently increased the potential for sediment storage.

Many studies have shown that wood inputs commonly increase sediment storage (e.g., Collins et al., 2012; May & Lee, 2004; Ryan et al., 2014; Wohl & Scott, 2017). This increase in sediment storage behind jams likely resulted in decreased downstream sediment supply and bed armoring. For alluvial streams, decreased sediment supply leads to coarsening (i.e., armoring) of the bed surface (Dietrich et al., 1989), and, in turn, coarsening leads to decreased transport efficiency due to higher entrainment thresholds. Year-to-year variations in gravel transport may be due to temporary storage and release of sediment behind numerous wood jams. We expect that the opening and closing of "wood valves" alternatively create pulses and deficits of gravel load that spread downstream and produce seasonal and annual variations in gravel transport rate.

Lisle and Napolitano (1998) investigated bed load response to North Fork harvesting for HY1990 to HY1996 and, although they did not detect changes in bed load yields at the North Fork weir pond, they did find evidence of aggradation along the main stem associated with large wood deposits that occurred following the blowdowns. Our GTMs overpredicted gravel yields for HY1993 to HY1999, but the difference between the reconstructed gravel yields and the predicted gravel yields was only greater than one standard deviation of the reconstructed gravel yields for HY1997 and HY1998 (Figure 7). Our overpredicted gravel yields for HY1993 to HY1999 are consistent with the findings of channel aggradation reported by Lisle and Napolitano (1998). We find it most likely that increased storage due to the input of downed wood was also the primary cause for the positive departures for HY1997 to HY2006.

The increase in wood loading and storage capacity beginning in the 1990s also offers an explanation for why the reconstructed gravel yields near the beginning of the operation of the weir were substantially underpredicted, since at that time storage capacity was presumably less and transport efficiency was greater as less wood was in the channel. Consequently, the North Fork main stem experienced more efficient gravel transport for some of the years prior to the blowdowns in the 1990s. This is consistent with the findings of Leeks and Marks (1997) who noted the importance of large downed wood at reducing bed load sediment delivery. Gravel yields were also slightly underpredicted for HY2013 to HY2015, which were relatively dry years, and overpredicted for the relatively wet HY2017 (Figure 7). Whether aggradation and storage continues is yet to be determined, but we expect that future measurements will yield the answer.

6. Conclusions

We developed an approach for reconstructing annual gravel yields from sediment reservoir surveys, measurements of suspended sediment, and recently collected field measurements. We used this technique to reconstruct a 55-year record of gravel transport for the North Fork of Caspar Creek, CA. We then compared reconstructed annual gravel yields to annual gravel yields predicted from a field-calibrated GTM using three different velocity-discharge relationships to estimate skin friction. The model that included the velocity-discharge relationship using measurements during storms produced results that best matched the

reconstructed annual gravel yields. We found that all three calibrated models were able to match the bed load transport rates measured with pit samplers just upstream of the North Fork weir at the Arfstein station for the moderate storms during the 1988 to 1995 sampling period. By comparing reconstructed and predicted gravel yields from 1962 to 2017, we were able to identify a period of relatively high gravel transport efficiency during an early phase of the study and a period of low efficiency and increased sediment storage during a period about 30 years into the study. The shift to lower gravel transport efficiency was likely the consequence of wood inputs that led to increased sediment storage and aggradation upstream of the Arfstein reach and decreased sediment supply to the Arfstein reach. The decrease in sediment supply may have led to coarsening of the bed surface through the Arfstein reach. Future bed load studies at Caspar Creek and elsewhere would be strongly benefitted by the collection of data necessary to determine grain size distributions of the bed surface in addition to transported bed load. We expect that our bed load reconstruction approach may prove useful for reconstructing bed load yields at other sites where suspended sediment yields are known and large weir ponds or impoundments trap bed load. Ultimately, the approach presented here should contribute to improved evaluation of long-term trends in gravel loads and allow better development of complete sediment budgets for watersheds before and after timber harvesting and other disturbances that affect sediment supply, transport, and storage.

Acknowledgments

The authors have no financial conflicts of interest. This research was partially supported by an appointment to the U. S. Forest Service Research Participation Program administered by the Oak Ridge Institute for Science and Education (ORISE) through an interagency agreement between the U. S. Department of Energy and the U.S. Department of Agriculture Forest Service. ORISE is managed by Oak Ridge Associated Universities (ORAU) under DOE contract number DE-SC0014664. All opinions expressed in this paper are the authors' and do not necessarily reflect the policies and views of USDA, USFS, DOE, or ORAU/ORISE. Funding was provided by the California Department of Forestry and Fire Protection (CAL FIRE) as part of their long-term commitment to research at Caspar Creek. We thank the editor Ellen Wohl, D. Nathan Bradley, and two anonymous reviewers for comments and suggestions that improved this manuscript. We thank those who helped collect field data and complete lab analysis for the Caspar Creek study and offer a special thanks to CAL FIRE colleagues Pete Cafferata, Drew Coe, Lynn Webb, and US Forest Service personnel responsible for collecting and processing field samples and data, especially Elizabeth Keppeler and Sue Hilton. We also thank Jayme Seehafer for managing the sediment processing laboratory and data at the Redwood Sciences Laboratory and Susan Marshall for access to the muffle furnace at Humboldt State University. All data used in this paper are archived and available from the USDA Forest Service (https://www.fs.usda.gov/rds/archive/) and Harvard Dataverse (https://www.dataverse.harvard.edu/).

References

Badoux, A., Andres, N., & Turowski, J. (2014). Damage costs due to bedload transport processes in Switzerland. *Natural Hazards and Earth System Sciences (NHESS)*, *14*(2), 279–294. https://doi.org/10.5194/nhess-14-279-2014

Bagnold, R. A. (1966). An approach to the sediment transport problem from general physics. *US Government Printing Office, US Geological Survey Professional Paper*, no 422-I.

Bailey, E. H., Irwin, W. P., & Jones, D. L. (1964). The Franciscan and related rocks and their significance in the geology of western California. *California Division of Mines and Geology Bulletin*, *183*, 1–177.

Bakke, P. D., Basdekas, P. O., Dawdy, D. R., & Klingeman, P. C. (2002). Calibrated Parker-Klingeman model for gravel transport. *Journal of Hydraulic Engineering*, *125*(6), 657–660. https://doi.org/10.1061/(asce)0733-9429(1999)125:6(657)

Barrière, J., Oth, A., Hostache, R., & Krein, A. (2015). Bed load transport monitoring using seismic observations in a low-gradient rural gravel bed stream. *Geophysical Research Letters*, *42*, 2294–2301. https://doi.org/10.1002/2015GL063630

Bray, E. N., & Dunne, T. (2017). Observations of bedload transport in a gravel bed river during high flow using fiber-optic DTS methods. In Earth Surface Processes and Landforms (Vol. 42, pp. 2184–2198). https://doi.org/10.1002/esp.4164

Brune, G. M. (1953). Trap efficiency of reservoirs. *Eos, American Geophysical Union*, *34*(3), 407–418.

Brunner, G. W. (2016). HEC-RAS river analysis system—Hydraulic reference manual, version 5.0. *Hydrologic Engineering Centre (HEC), U. S. Army Corps of Engineers*. https://doi.org/CPD-68

Bunte, K., Abt, S. R., Asce, F., Potyondy, J. P., & Ryan, S. E. (2004). Measurement of coarse gravel and cobble transport using portable bedload traps. *Journal of Hydraulic Engineering*, *130*(9), 879–893.

Bunte, K., & MacDonald, L. H. (1995). Detecting change in sediment load: Where and how is it possible. In *Effects of Scale on Interpretation and Management of Sediment and Water Quality, Proceedings of a Boulder Symposium* (pp. 253–260). IAHS Publ.

Burt, R. (Ed.). (2014). Kellogg soil survey laboratory methods manual. United States Department of Agriculture, Natural Resources Conservation Service, National Soil Survey Center.

Burtin, A., Cattin, R., Bollinger, L., Vergne, J., Steer, P., Robert, A., et al. (2011). Towards the hydrologic and bed load monitoring from high-frequency seismic noise in a braided river: The "torrent de St Pierre", French Alps. *Journal of Hydrology*, *408*(1–2), 43–53. https://doi.org/10.1016/j.jhydrol.2011.07.014

Cafferata, P. H., & Reid, L. M. (2013). Applications of long-term watershed research to forest management in California: 50 years of learning from the Caspar Creek experimental watersheds. State of California, The Natural Resources Agency, Department of Forestry and Fire Protection.

Cafferata, P. H., & Spittler, T. E. (1998). Logging impacts of the 1970's vs. the 1990's in the Caspar Creek Watershed. In: Ziemer, R.R. (Technical Coordinator), Proceedings of the Conference on Coastal Watersheds: The Caspar Creek Story, General Technical Report PSW GTR-168. Albany, CA: U.S. Department of Agriculture, Forest Service, Pacific Southwest Forest and Range Experiment Station, 103–115.

Chow, V. Te. (1959). Open Channel Hydraulics Open Channel Hydraulics. https://doi.org/10.1016/B978-0-7506-6857-6.X5000-0

Collins, B. D., Montgomery, D. R., Fetherston, K. L., & Abbe, T. B. (2012). The floodplain large-wood cycle hypothesis: A mechanism for the physical and biotic structuring of temperate forested alluvial valleys in the North Pacific coastal ecoregion. *Geomorphology*, *139-140*, 460–470. https://doi.org/10.1016/j.geomorph.2011.11.011

Davies, T., & McSaveney, M. (2011). Bedload sediment flux and flood risk management in New Zealand. *Journal of Hydrology. New Zealand*, *50*, 181–189.

Dendy, F. E., & Champion, W. A. (1978). Sediment deposition in U.S. reservoirs. *USDA Miscellaneous Publication No. 1362, USDA Agricultural Research Service, Oxford, MS*. 90 p.

Dietrich, W. E., Kirchner, J. W., Ikeda, H., & Iseya, F. (1989). Sediment supply and the development of the coarse surface layer in gravel-bedded rivers. *Nature*, *340*(6230), 215–217.

Duck, R. W., & McManus, J. (1994). A long-term estimate of bedload and suspended sediment yield derived from reservoir deposits. *Journal of Hydrology*, *159*(1–4), 365–373. https://doi.org/10.1016/0022-1694(94)90267-4

Evitt, W. R., & Pierce, S. T. (1975). Early tertiary ages from the coastal belt of the Franciscan complex, Northern California. *Geology*, *3*(8), 433–436. https://doi.org/10.1130/0091-7613(1975)3<433:ETAFTC>2.0.CO;2

Ferguson, R., Hoey, T., Wathen, S., & Werritty, A. (1996). Field evidence for rapid downstream fining of river gravels through selective transport. *Geology*, *24*(2), 179. https://doi.org/10.1130/0091-7613(1996)024<0179:FEFRDF>2.3.CO;2

Gomez, B. (2006). The potential rate of bed-load transport. *Proceedings of the National Academy of Sciences*, *103*(46), 17170–17173. https://doi.org/10.1073/pnas.0608487103

Grant, G. E., & Wolff, A. L. (1991). Long-term patterns of sediment transport after timber harvest, Western Cascade Mountains, Oregon, USA. *Sediment and Stream Water Quality in a Changing Environment: Trends and Explanation*, *203*, 31–40.

Hahm, W. J., Rempe, D. M., Dralle, D. N., Dawson, T. E., Lovill, S. M., Bryk, A. B., et al. (2019). Lithologically controlled subsurface critical zone thickness and water storage capacity determine regional plant community composition. *Water Resources Research*, *55*, 3028–3055. https://doi.org/10.1029/2018WR023760

Hassan, M. A., Robinson, S. V. J., Voepel, H., Lewis, J., & Lisle, T. E. (2014). Modeling temporal trends in bedload transport in gravel-bed streams using hierarchical mixed-effects models. *Geomorphology*, *219*, 260–269. https://doi.org/10.1016/j.geomorph.2014.05.019

Hean, D. S., & Nanson, G. C. (1987). Serious problems in using equations to estimate bedload yields for coastal rivers in NSW. *Australian Geographer*, *18*(2), 114–124. https://doi.org/10.1080/00049188708702933

Heinemarm, H. G. (1981). A new sediment trap efficiency curve for small reservoirs. *Journal of the American Water Resources Association*, *17*(5), 825–830. https://doi.org/10.1111/j.1752-1688.1981.tb01304.x

Henry, N. (1998). Overview of the Caspar Creek watershed study. Proceedings of the Conference on Coastal Watersheds: The Caspar Creek Story.

Hiraoka, M., Gomi, T., Oda, T., Egusa, T., & Uchiyama, Y. (2015). Responses of bed load yields from a forested headwater catchment in the eastern Tanzawa Mountains, Japan. *Hydrological Research Letters*, *9*(3), 41–46. https://doi.org/10.3178/hrl.9.41

Hsu, L., Finnegan, N. J., & Brodsky, E. E. (2011). A seismic signature of river bedload transport during storm events. *Geophysical Research Letters*, *38*, L13407. https://doi.org/10.1029/2011GL047759

Jackson, W. L., & Beschta, R. L. (1982). A model of two-phase bedload transport in an Oregon coast range stream. *Earth Surface Processes and Landforms*, *7*(6), 517–527. https://doi.org/10.1002/esp.3290070602

Johnson, S. L., & Rothacher, J. S. (2019). Annual bedload accumulation from sediment basin surveys in small gauged watersheds in the Andrews Experimental Forest, 1957 to present. Environmental data initiative (v11). https://doi.org/10.6073/pasta/73b09502b3e6b0b2622f53810afa093b.

Keppeler, E. T., & Brown, D. (1998). Subsurface drainage processes and management impacts. *In: Ziemer, R.R. (Technical Coordinator), Proceedings of the Conference on Coastal Watersheds: The Caspar Creek Story, General Technical Report PSW GTR-168. Albany, CA: U.S. Department of Agriculture, Forest Service, Pacific Southwest Forest and Range Experiment Station*, 25–34.

Klösch, M., & Habersack, H. (2018). Deriving formulas for an unsteady virtual velocity of bedload tracers. *Earth Surface Processes and Landforms*, *43*((7), 1529–1541. https://doi.org/10.1002/esp.4326

Kondolf, G. M., & Gordon, M. W. (1993). The sizes of salmonid spawning gravels. *Water Resources Research*, *29*(7), 2275–2285. https://doi.org/10.1029/93WR00402

Lamb, M. P., Dietrich, W. E., & Venditti, J. G. (2008). Is the critical Shields stress for incipient sediment motion dependent on channel-bed slope? *Journal of Geophysical Research*, *113*, F02008. https://doi.org/10.1029/2007JF000831

Lane, S. N., Richards, K. S., & Chandler, J. H. (1995). Morphological estimation of the time- integrated bed load transport rate. *Water Resources Research*, *31*(3), 761–772. https://doi.org/10.1029/94WR01726

Lane, S. N., Westaway, R. M., & Murray Hicks, D. (2003). Estimation of erosion and deposition volumes in a large, gravel-bed, braided river using synoptic remote sensing. *Earth Surface Processes and Landforms: The Journal of the British Geomorphological Research Group*, *28*(3), 249–271. https://doi.org/10.1002/esp.483

Langenheim, V. E., Jachens, R. C., Wentworth, C. M., & McLaughlin, R. J. (2013). Previously unrecognized regional structure of the Coastal Belt of the Franciscan Complex, Northern California, revealed by magnetic data. *Geosphere*, *9*(6), 1514–1529. https://doi.org/10.1130/GES00942.1

Leeks, G. J. L., & Marks, S. D. (1997). Dynamics of river sediments in forested headwater streams: Plynlimon. *Hydrology and Earth System Sciences Discussions*, *1*(3), 483–497.

Lewis, J. (1998). Evaluating the impacts of logging activities on erosion and suspended sediment transport in the Caspar Creek Watersheds. In R. R. Ziemer (Ed.), (Technical Coordinator)*Proceedings of the Conference on Coastal Watersheds: The Caspar Creek Story, General Technical Report PSW GTR-168*, (pp. 55–69). Albany, CA: U.S. Department of Agriculture, Forest Service, Pacific Southwest Forest and Range Experiment Station.

Lisle, T. E., & Lewis, J. (2009). Effects of sediment transport on survival of salmonid embryos in a natural stream: A simulation approach. *Canadian Journal of Fisheries and Aquatic Sciences*, *49*(11), 2337–2344. https://doi.org/10.1139/f92-257

Lisle, T. E., & Napolitano, M. B. (1998). Effects of recent logging on the main channel of North Fork Caspar Creek: *In: Ziemer, R.R. (Technical Coordinator), Proceedings of the Conference on Coastal Watersheds: The Caspar Creek Story, General Technical Report PSW GTR-168. Albany, CA: U.S. Department of Agriculture, Forest Service, Pacific Southwest Forest and Range Experiment Station*, 81–85.

Mackin, J. H. (1948). Concept of the graded river. *Geological Society of America Bulletin*, *59*(5), 463–512. https://doi.org/10.1130/0016-7606(1948)59[463:COTGR]2.0.CO;2

MATLAB (2018). *9.4.0.813654 (R2018a)*. Natick, Massachusetts: The MathWorks Inc.

May, C. L., & Lee, D. C. (2004). The relationships among in-channel sediment storage, pool depth, and summer survival of juvenile salmonids in Oregon coast range streams. *North American Journal of Fisheries Management*, *24*(3), 761–774.

Merz, J. E., & Chan, L. K. O. (2005). Effects of gravel augmentation on macroinvertebrate assemblages in a regulated California river. *River Research and Applications*, *21*(1), 61–74. https://doi.org/10.1002/rra.819

Montgomery, D. R., Buffington, J. M., Peterson, N. P., Schuett-Hames, D., & Quinn, T. P. (2011). Stream-bed scour, egg burial depths, and the influence of salmonid spawning on bed surface mobility and embryo survival. *Canadian Journal of Fisheries and Aquatic Sciences*, *53*(5), 1061–1070. https://doi.org/10.1139/f96-028

Mueller, E. R., Pitlick, J., & Nelson, J. M. (2005). Variation in the reference Shields stress for bed load transport in gravel-bed streams and rivers. *Water Resources Research*, *41*, W04006. https://doi.org/10.1029/2004WR003692

Muhs, D. R., Prentice, C., & Merrits, D. J. (2003). Marine terraces, sea level history and Quaternary tectonics of the San Andreas fault on the coast of California. *Quaternary Geology of the United States, INQUA 2003 Field Guide Volume*, (1976), 1–18.

Napolitano, M. B. (1996). Sediment transport and storage in North Fork Caspar Creek, Mendocino County, California: Water years 1980–1988. *M.S. Thesis*. 148 p.

Nash, E., & Sutcliffe, V. (1970). River flow forecasting through conceptual models Part I—A discussion of principles. *Journal of Hydrology*, *10*(3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

Parker, G. (1990). Surface-based bedload transport relation for gravel rivers. *Journal of Hydraulic Research*, *28*(4), 417–436. https://doi.org/10.1080/00221689009499058

Parker, G., & Klingeman, P. C. (1982). On why gravel bed streams are paved. *Water Resources Research*, *18*(5), 1409–1423. https://doi.org/10.1029/WR018i005p01409

Parker, G., Klingeman, P. C., & McLean, D. G. D. (1982). Bedload and size distribution in paved gravel-bed streams. *Journal of Hydraulic Engineering*, *108*(4), 544–571.

Phillips, C. B., & Jerolmack, D. J. (2014). Dynamics and mechanics of bed-load tracer particles. *Earth Surface Dynamics*, *2*(2), 513. https://doi.org/10.5194/esurf-2-513-2014

Rausch, D. L., & Schreiber, J. D. (1981). Sediment and nutrient trap efficiency of a small flood-detention reservoir. *Journal of Environmental Quality*, *10*(3), 288–293. https://doi.org/10.2134/jeq1981.00472425001000030007x

Reice, S. R., & Url, S. (1985). Experimental disturbance and the maintenance of species diversity in a stream community. *Oecologia*, *67*(1), 90–97. https://doi.org/10.1007/BF00378456

Reid, L. M., Dewey, N. J., Lisle, T. E., & Hilton, S. (2010). The incidence and role of gullies after logging in a coastal redwood forest. *Geomorphology*, *117*(1–2), 155–169. https://doi.org/10.1016/j.geomorph.2009.11.025

Reid, L. M., & Hilton, S. (1998). Buffering the buffer. In R. R. Ziemer (Ed.), Technical Coordinator*Proceedings of the Conference on Coastal Watersheds: The Caspar Creek Story, Gen. Tech. Rep. PSW GTR-168* (pp. 71–80). Albany, CA: U.S. Department of Agriculture, Forest Service, Pacific Southwest Forest and Range Experiment Station.

Reid, L. M., Keppeler, E., & Hilton, S. (2017). Post-landslide recovery patterns in a coast redwood forest. Gen. Tech. Rep. PSW-GTR-258. Albany, CA: US Department of Agriculture, Forest Service, Pacific Southwest Research Station: 135-147., 258, 135–147.

Reid, L. M., & Keppeler, E. T. (2012). Landslides after clearcut logging in a coast redwood forest. In R. B. Standiford, T. J. Weller, D. D. Piirto, & J. D. Stuart (Eds.), Tech. Coords*Proceedings of Coast Redwood Forests in a Changing California: A Symposium for Scientists and Managers. Gen. Tech. Rep. PSW-GTR-238*, (pp. 163–172). Albany, CA: U.S. Department of Agriculture, Forest Service, Pacific Southwest Forest and Range Experiment Station.

Renard, K. G., Nichols, M. H., Woolhiser, D. A., & Osborn, H. B. (2008). A brief background on the U.S. Department of Agriculture Agricultural Research Service walnut gulch experimental watershed. *Water Resources Research*, *44*, W05S02. https://doi.org/10.1029/2006WR005691

Rice, R. M., Tilley, F. B., & Datzman, P. A. (1979). A watershed's response to logging and roads: South Fork of Caspar Creek, 1967–1976. Res. Paper PSW-RP-146. Berkeley, CA: U.S. Department of Agriculture, Forest Service, Pacific Southwest Forest and Range Experiment Station.

Richardson, P. W., Seehafer, J. E., Keppeler, E. T., Sutherland, D. G., Wagenbrenner, J. W. (2020a). Caspar Creek Experimental Watersheds Phase 1 (1962–1985) data. Forest Service Research Data Archive. https://doi.org/10.2737/RDS-2020-0017

Richardson, P. W., Seehafer, J. E., Keppeler, E. T., Sutherland, D. G., Wagenbrenner, J. W. (2020b). Caspar Creek Experimental Watersheds Phase 2 (1985–2017) data. Forest Service Research Data Archive. https://doi.org/10.2737/RDS-2020-0018

Richardson, P. W. & Wagenbrenner, J. W. (2020). Supplemental bed load data for the North Fork of Caspar Creek, CA. Harvard Dataverse (v1). https://doi.org/10.7910/DVN/ZKYNQC

Rickenmann, D., Turowski, J. M., Fritschi, B., Klaiber, A., & Ludwig, A. (2012). Bedload transport measurements at the Erlenbach stream with geophones and automated basket samplers. *Earth Surface Processes and Landforms*, *37*(9), 1000–1011. https://doi.org/10.1002/esp.3225

Rinderer, M., Jenewein, S., Senfter, S., Rickenmann, D., Schöberl, F., Stötter, J., & Hegg, C. (2009). Runoff and bedload transport modelling for flood hazard assessment in small alpine catchments-the PROMAB GIS model. In *Sustainable Natural Hazard Management in Alpine Environments*, (pp. 69–101). Berlin, Heidelberg: Springer.

Roth, D. L., Finnegan, N. J., Brodsky, E. E., Rickenmann, D., Turowski, J. M., Badoux, A., & Gimbert, F. (2017). Bed load transport and boundary roughness changes as competing causes of hysteresis in the relationship between river discharge and seismic amplitude recorded near a steep mountain stream. *Journal of Geophysical Research: Earth Surface*, *122*, 1182–1200. https://doi.org/10.1002/2016JF004062

Ryan, S. E., Bishop, E. L., & Daniels, J. M. (2014). Influence of large wood on channel morphology and sediment storage in headwater mountain streams, Fraser Experimental Forest, Colorado. *Geomorphology*, *217*, 73–88. https://doi.org/10.1016/j.geomorph.2014.03.046

Schmidt, K. H., & Ergenzinger, P. (1992). Bedload entrainment, travel lengths, step lengths, rest periods—Studied with passive (iron, magnetic) and active (radio) tracer techniques. *Earth Surface Processes and Landforms*, *17*(2), 147–165. https://doi.org/10.1002/esp.3290170204

Selker, J., van de Giesen, N. C., Westhoff, M., Luxemburg, W., & Parlange, M. B. (2006). Fiber optics opens window on stream dynamics. *Geophysical Research Letters*, *33*, L24401. https://doi.org/10.1029/2006GL027979

Troendle, C. A., & Olsen, W. K. (1994). Potential effects of timber harvest and water management on streamflow dynamics and sediment transport. *Sustainable Ecological Systems: Implementing an Ecological Approach to Land Management. USDA Forest Service General Technical Report RM-247, Fort Collins, CO,* 34-41.

Turowski, J. M., Badoux, A., & Rickenmann, D. (2011). Start and end of bedload transport in gravel-bed streams. *Geophysical Research Letters*, *38*, L04401. https://doi.org/10.1029/2010GL046558

Verstraeten, G., & Poesen, J. (2001). Modelling the long-term sediment trap efficiency of small ponds. *Hydrological Processes*, *15*(14), 2797–2819. https://doi.org/10.1002/hyp.269

Wagenbrenner, J. W., & Robichaud, P. R. (2014). Post-fire bedload sediment delivery across spatial scales in the interior western United States. *Earth Surface Processes and Landforms*, *39*(7), 865–876. https://doi.org/10.1002/esp.3488

Wilcock, P. R. (1998). Two-fraction model of initial sediment motion in gravel-bed rivers. *Science*, *280*(5362), 410–412. https://doi.org/10.1126/science.280.5362.410

Wilcock, P. R. (2001). Toward a practical method for estimating sediment-transport rates in gravel-bed rivers. *Earth Surface Processes and Landforms*, *26*(13), 1395–1408. https://doi.org/10.1002/esp.301

Wilcock, P. R., & Crowe, J. C. (2003). Surface-based transport model for mixed-size sediment. *Journal of Hydraulic Engineering*, *129*(2), 120–128. https://doi.org/10.1061/(ASCE)0733-9429(2003)129:2(120)

Wilcock, P. R., & Kenworthy, S. T. (2002). A two-fraction model for the transport of sand/gravel mixtures. *Water Resources Research*, *38*(10), 1194. https://doi.org/10.1029/2001WR000684

Wilcock, P. R., Pitlick, J., & Cui, Y. (2009). Sediment transport primer: Estimating bed-material transport in gravel-bed rivers. *Gen. Tech. Rep. RMRS-GTR-226. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station*, 78.

Wohl, E., & Scott, D. N. (2017). Wood and sediment storage and dynamics in river corridors. *Earth Surface Processes and Landforms*, *42*(1), 5–23. https://doi.org/10.1002/esp.3909

Wood, P. J., & Armitage, P. D. (1997). Biological effects of fine sediment in the lotic environment. *Environmental Management*, *21*(2), 203–217. https://doi.org/10.1007/s002679900019

Yager, E. M., Kenworthy, M., & Monsalve, A. (2015). Taking the river inside: Fundamental advances from laboratory experiments in measuring and understanding bedload transport processes. *Geomorphology*, *244*, 21–32. https://doi.org/10.1016/J.GEOMORPH.2015.04.002

Yager, E. M., Venditti, J. G., Smith, H. J., & Schmeeckle, M. W. (2018). The trouble with shear stress. *Geomorphology*, *323*, 41–50. https://doi.org/10.1016/j.geomorph.2018.09.008

References From the Supporting Information

Duan, N. (1983). Smearing estimate: A nonparametric retransformation method. *Journal of the American Statistical Association*, *78*(383), 605–610.

Keppeler, E. T. (2012). Sediment production in a coastal watershed: Legacy, land use, recovery, and rehabilitation. In R. B. Standiford, T. J. Weller, D. D. Piirto, & J. D. Stuart (Eds.), (Technical Coordinators)*Proceedings of Coast Redwood Forests in a Changing California: A Symposium for Scientists and Managers. Gen. Tech. Rep. PSW-GTR-238* (pp. 69–77). Albany, CA: U.S. Department of Agriculture, Forest Service, Pacific Southwest Research Station.

Lewis, J. (1996). Turbidity-controlled suspended sediment sampling for runoff-event load estimation. *Water Resources Research*, *32*(7), 2299–2310.

Lewis, J., & Eads, R. (2009). Implementation guide for turbidity threshold sampling: Principles, procedures, and analysis. Gen. Tech. Rep. PSW-GTR-212. Albany, CA: US Department of Agriculture, Forest Service, Pacific Southwest Research Station. 86 p. https://doi.org/10.2737/PSW-GTR-212

Vaughan, A. A., Belmont, P., Hawkins, C. P., & Wilcock, P. (2017). Near-channel versus watershed controls on sediment rating curves. *Journal of Geophysical Research: Earth Surface*, *122*, 1901–1923. https://doi.org/10.1002/2016JF004180