Stuff removed from Data Collection section

### Data Collection

Data on precipitation (P), water discharge (Q), suspended sediment concentration (SSC) and turbidity (T) were collected during three field campaigns: January-March, 2012, February-July 2013, and January-March 2014, and several intervening periods of unattended monitoring by instruments with data loggers. Field sampling campaigns were scheduled to coincide with the period of most frequent storms in the November-May wet season, though large storms were sampled throughout the year.

#### Precipitation

Precipitation (P) was measured at three locations in Faga'alu watershed using Rainwise RAINEW tipping-bucket rain gages (RG1 and RG2) and a Vantage Pro Weather Station (Wx) (Figure 1). Data at RG2 was only recorded January-March, 2012, to determine a relationship between elevation and precipitation in the LOWER subwatershed. The total event precipitation (Psum) and event Erosivity Index (EI30) were calculated using data from RG1, with data gaps filled by 15 min interval precipitation data from Wx. While previous data suggest that precipitation increases with elevation, here we do not calculate watershed-mean precipitation, and instead use precipitation depth at RG1 to indicate the depth of rainfall during a storm event. Most sheetwash and rill erosion, which depends on rainfall intensity and EI30, occurred at the quarry, near the location of RG1. Rainfall data from RG1 is therefore most representative of rainfall at the quarry.

#### Water Discharge

Stream gaging sites were chosen to take advantage of an existing control structure (FG1) and a stabilized stream cross section (FG3)(Duvert et al, 2010). At FG1 and FG3, Q was calculated from 15 minute interval stream stage measurements, using a stage-Q rating curve calibrated to manual Q measurements made under baseflow and stormflow conditions (Figures 3 and 4). Stream stage was measured with non-vented pressure transducers (PT) (Solinst Levelogger or Onset HOBO Water Level Logger) installed in stilling wells at FG1 and FG3. Barometric pressure data collected at Wx were used to calculate stage from the pressure data recorded by the PT. Data gaps in barometric pressure from Wx were filled by data from stations at Pago Pago Harbor (NSTP6) and NOAA Climate Observatory at Tula (TULA) (Figure 1). Priority was given to the station closest to the watershed with valid barometric pressure data. Barometric data were highly correlated and the data source made little (<1cm) difference in the resulting water level. Q was measured in the field by the area-velocity method (AV) using a Marsh-McBirney flowmeter to measure flow velocity and channel surveys measure cross-sectional area (Harrelson et al., 1994; Turnipseed and Sauer, 2010).

AV-Q measurements could not be made at high stages at FG1 and FG3 for safety reasons, so stage-Q relationships were constructed to estimate a continuous record of Q. At FG3, the channel is rectangular with stabilized rip-rap on the banks and bed (Appendix Figure A1.1). Recorded stage varied from 4 to 147 cm. AV-Q measurements (n= 14) were made from 30 to 1,558.0 L/sec, covering a range of stages from 6 to 39 cm. The highest recorded stage was much higher than the highest stage with measured Q so the rating could not be extrapolated by a power law. Stream conditions at FG3 fit the assumption for Manning's equation, so the stage-Q rating at FG3 was created using Manning's equation, calibrating Manning's n (0.067) to the Q measurements (Figure 3).

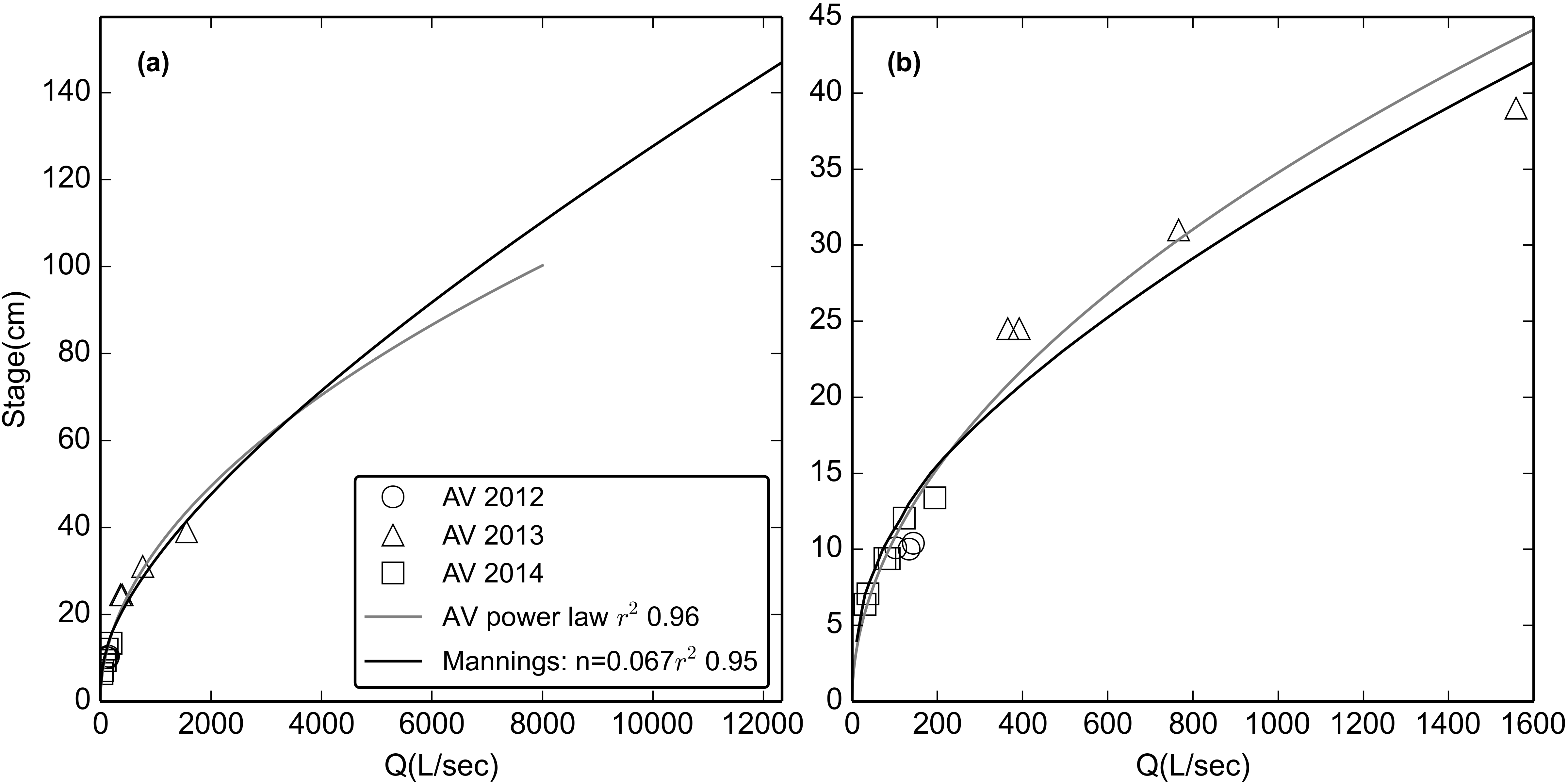


Figure 3. Stage-Discharge relationships for stream gaging site at FG3 for (a) the full range of observed stage and (b) the range of stages with AV measurements of Q. RMSE was 93 L/sec, or 32% of observed Q.

At FG1, the flow control structure is a masonry ogee spillway crest of a defunct stream capture. The structure is a rectangular channel 43 cm deep that transitions abruptly to gently sloping banks, causing an abrupt change in the stage-Q relationship (Appendix Figure A1.2). At FG1, recorded stage height ranged from 4 to 120 cm, while area-velocity Q measurements (n= 22) covered stages from 6 to 17 cm. Since the highest recorded stage (120 cm) was higher than the highest stage with measured Q (17 cm), and there was a distinct change in channel geometry above 43 cm the rating could not be extrapolated by a power law. The flow structure did not meet the assumptions for using Manning's equation to predict flow so the HEC-RAS model was used (Brunner 2010). The surveyed geometry of the upstream channel and flow structure at FG1 were input to HEC-RAS, and the HEC-RAS model was calibrated to the Q measurements (Figure 4). While a power function fit Q measurements better than HEC-RAS for low flow, HEC-RAS fit better for Q above the storm threshold used in analyses of SSY (Figure 4).

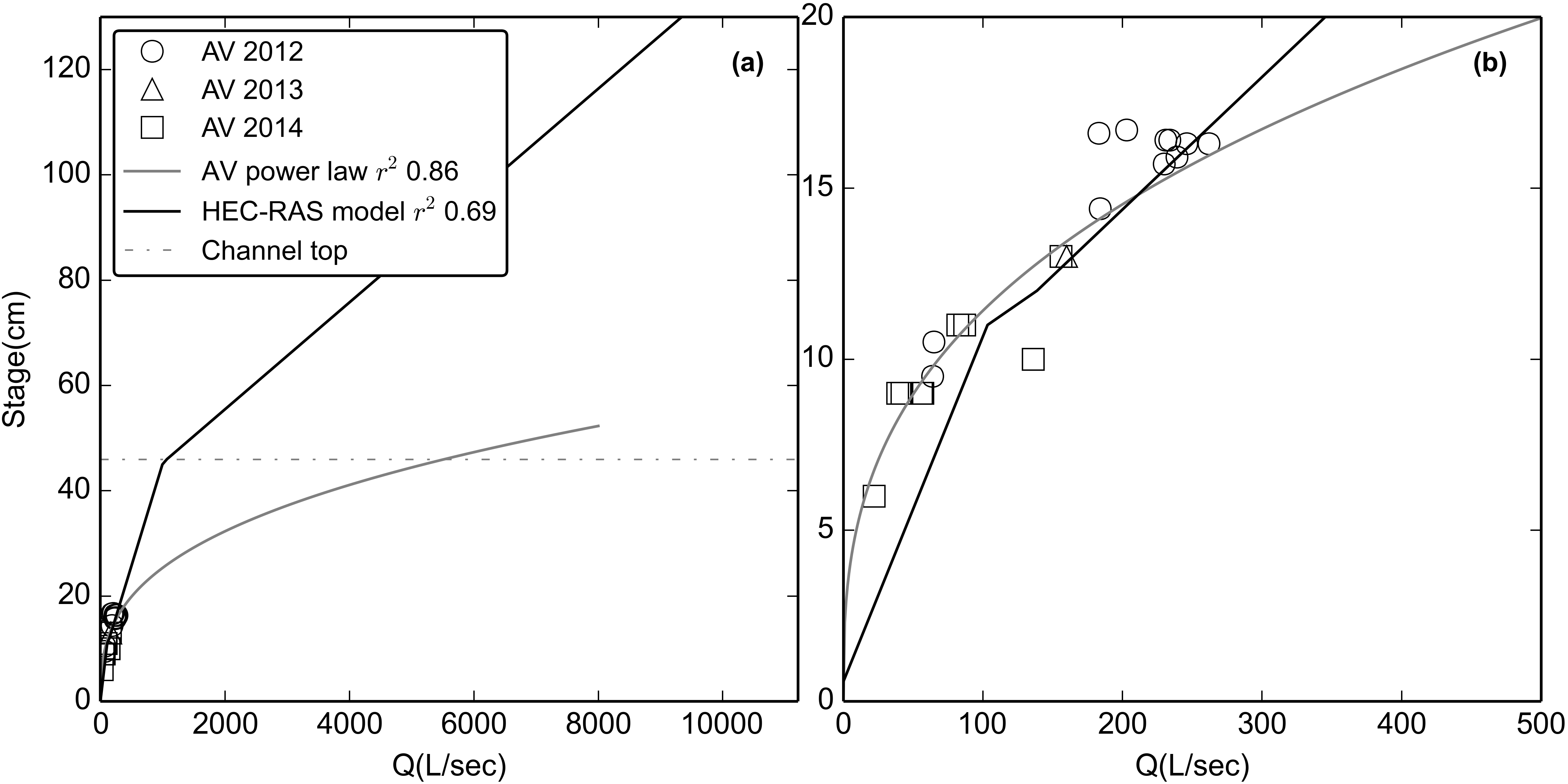


Figure 4. Stage-Discharge relationships for stream gaging site at FG1 for (a) the full range of observed stage and (b) the range of stages with AV measurements of Q. RMSE was 31 L/sec, or 22% of observed Q. "Channel Top" refers to the point where the rectangular channel transitions to a sloped bank and cross-sectional area increases much more rapidly with stage. A power-law relationship is also displayed to illustrate the potential error that could result if inappropriate methods are used.

Water discharge at FG2 was calculated as the product of the specific water discharge from FG1 (m³/0.9 km²) and the watershed area draining to FG2 (1.17 km²). This assumes that specific water discharge from the subwatershed above FG2 is similar to above FG1. Discharge may be higher from the quarry surface, which represents 5.7% of the LOWER\_QUARRY subwatershed, so Q, and thus SSY from the quarry are a conservative, lower bound estimate. The quarry surface is continually being disturbed, sometimes with large pits excavated and refilled in the course of weeks, as well as intentional water control structures implemented over time. Given the changes in the contributing area of the quarry, estimates of water yield from the quarry were uncertain, so we assumed a uniform specific discharge for the whole LOWER\_QUARRY subwatershed.

#### Continuous Suspended Sediment Concentration

Continuous SSC at 15 minute intervals was estimated from 1) linear interpolation of SSC measured from water samples, and 2) 15 min interval turbidity data (T) and a T-SSC relationship calibrated to stream water samples collected over a range of Q and SSC.

Stream water samples were collected by grab sampling with 500 mL HDPE bottles at FG1, FG2, and FG3. At FG2, water samples were also collected at 30 min intervals during storm events by an ISCO 3700 Autosampler triggered by a stage height sensor. Samples were analyzed for SSC on-island using gravimetric methods (Gray, 2014; Gray et al., 2000). Water samples were vacuum filtered on pre-weighed 47mm diameter, 0.7 um Millipore AP40 glass fiber filters, oven dried at 100 C for one hour, cooled and weighed to determine SSC (mg/L). From January 6, 2012, to October 1, 2014, 637 water samples were collected and analyzed for SSC: FG1 (n=59), FG2 (n=90 grab samples, n=198 from the Autosampler), and FG3 (n=159).

##### Interpolated grab samples

Interpolation of SSC values from grab samples could only be performed if at least three stream water samples were collected during a storm event (Nearing et al., 2007), and if they adequately captured the SSC dynamics of the storm event. SSC was assumed to be zero at the beginning and end of each storm if no grab sample data was available for those times (Lewis et al., 2001).

##### Turbidity-SSC relationships

Turbidity (T) was measured at FG1 and FG3 using three types of turbidimeters: 1) Greenspan TS3000 (TS), 2) YSI 600OMS with 6136 turbidity probe (YSI), and 3) Campbell Scientific OBS500 (OBS). All turbidimeters were permanently installed in protective PVC housings near the streambed where the turbidity probe would be submerged at all flow conditions, with the turbidity probe oriented downstream. Despite regular maintenance, debris fouling during storm and baseflows was common and caused data loss during several storm events (Lewis et al., 2001). Storm events with incomplete or invalid T data were not used in the analysis. A three-point calibration was performed on the YSI turbidimeter with YSI turbidity standards (0, 126, and 1000 NTU) at the beginning of each field season and approximately every 3-6 months during data collection. Turbidity measured with 0, 126, and 1000 NTU standards differed by less than 10% (4-8%) during each recalibration. The OBS requires calibration every two years, so recalibration was not needed during the study period. All turbidimeters were cleaned following storms to ensure proper operation.

At FG3, a YSI turbidimeter recorded T (NTU) at 5 min intervals from January 30, 2012, to February 20, 2012, and at 15 min intervals from February 27, 2012 to May 23, 2012, when it was damaged during a large storm. The YSI turbidimeter was replaced with an OBS, which recorded Backscatter (BS) and Sidescatter (SS) at 5 min intervals from March 7, 2013, to July 15, 2014 (OBSa), and was resampled to 15 min intervals. No data was recorded from August 2013-January 2014 when the wiper clogged with sediment. A new OBS was installed at FG3 from January, 2014, to August, 2014 (OBSb). To correct for some periods of high noise observed in the BS and SS data recorded by the OBSa in 2013, the OBSb installed in 2014 was programmed to make a burst of 100 BS and SS measurements at 15 min intervals, and record Median, Mean, STD, Min, and Max. All BS and SS parameters were analyzed to determine which showed the best relationship with SSC. Mean SS showed the highest r2 and is a physically comparable measurement to NTU measured by the YSI and TS (Anderson, 2005).

At FG1, the TS turbidimeter recorded T (NTU) at 5 min intervals from January 2012 until it was vandalized and destroyed in July 2012. The YSI turbidimeter, previously deployed at FG3 in 2012, was repaired and redeployed at FG1 and recorded T (NTU) at 5 min intervals from June 2013 to October 2013, and January 2014 to August 2014. T data was resampled to 15 min intervals to compare with SSC samples for the T-SSC relationship, and to correspond to Q for calculating SSY.

The T-SSC relationship can be unique to each region, stream, instrument or even each storm event (Lewis et al., 2001), and can be influenced by water color, dissolved solids and organic matter, temperature, and the shape, size, and composition of sediment. However, T has proved to be a robust surrogate measure of SSC in streams (Gippel, 1995), and is most accurate when a unique T-SSC relationship is developed for each instrument separately, using in situ grab samples under storm conditions (Lewis, 1996). A unique T-SSC relationship was developed for each turbidimeter, at each location, using 15 min interval T data and SSC samples from storm periods only (Figure 5). A "synthetic" T-SSC relationship was also developed by placing the turbidimeter in a black tub with water, and sampling T and SSC as sediment was added (Appendix 4, Figure 1), but results were not comparable to T-SSC relationships developed under actual storm conditions and were not used in further analyses.

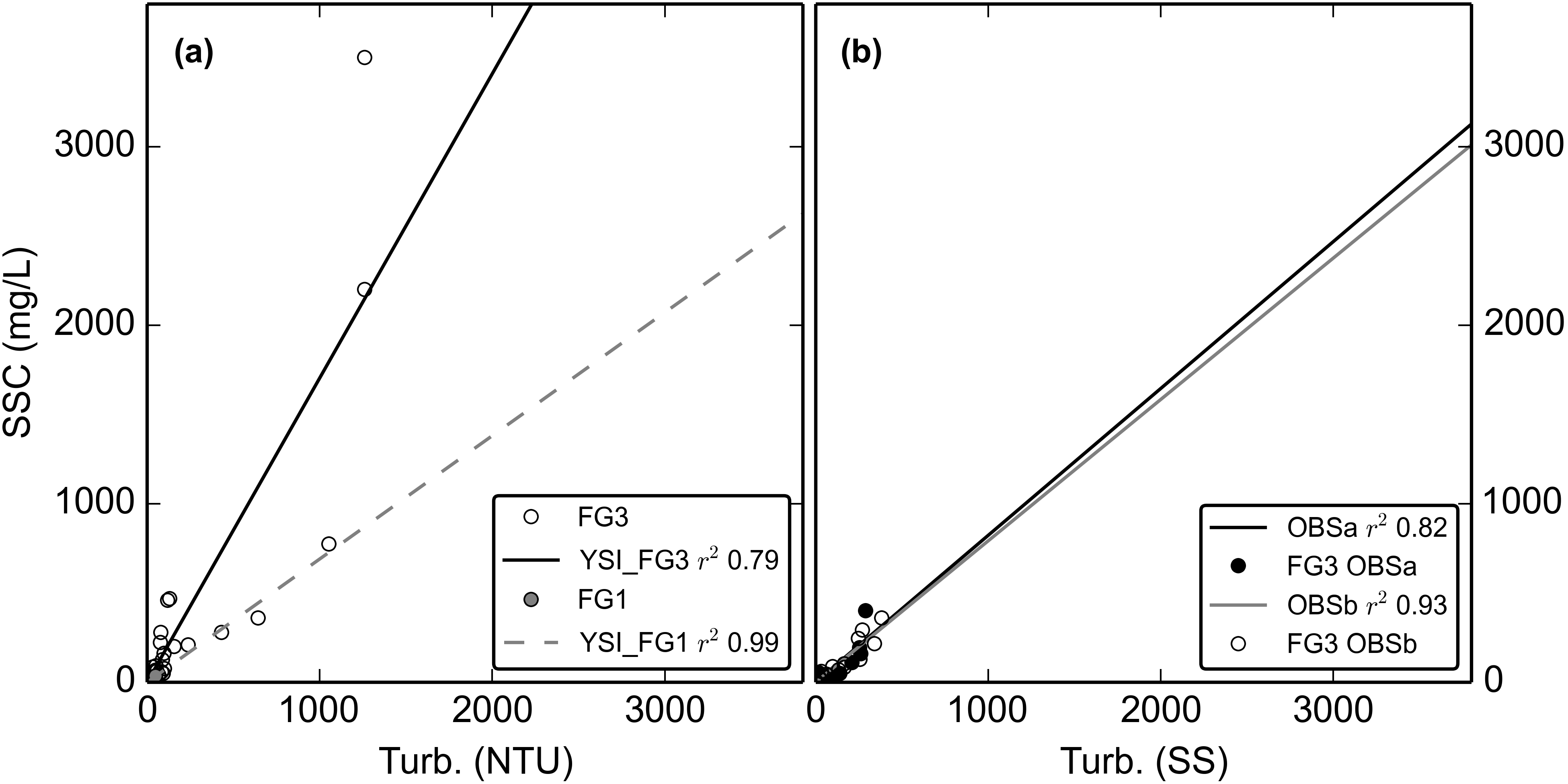


Figure 5. Turbidity-Suspended Sediment Concentration relationships for a) the YSI turbidimeter deployed at FG3 (02/27/2012-05/23/2012) and the same YSI turbidimeter deployed at FG1 (06/13/2013-12/31/2014). b) OBS500 turbidimeter deployed at FG3 (03/11/2013-07/11/2013) and c) OBS500 turbidimeter deployed at FG3 (01/31/2014-03/04/2014).

The T-SSC relationships varied among sampling sites and sensors but all showed acceptable r2 values (0.79-0.99). Lower scatter was achieved by using grab samples collected during stormflows only. For the TS (not shown) and YSI deployed at FG1, the r2 values were high (0.58, 0.99) but the ranges of T and SSC values used to develop the relationships were considered too small (0-16 NTU) compared to the maximum observed during the deployment period (1,077 NTU) to develop a robust relationship for higher T values. Instead, the T-SSC relationship developed for the YSI turbidimeter installed at FG3 (Figure 5a) was used to calculate SSC from T data collected by the TS and the YSI at FG1. For the YSI turbidimeter, more scatter was observed in the T-SSC relationship at FG3 than at FG1 (Figure 5a), which could be attributed to the higher number and wider range of values sampled, and to temporal variability in sediment characteristics. The OBSa and OBSb turbidimeters had high r2 values (0.82, 0.93) and compared well between the two periods of deployment (Figure 5b).

### Cumulative Probable Error (PE)

Uncertainty in SSYEV estimates arises from both measurement and model errors, including models of stage-discharge (stage-Q) and turbidity-suspended sediment concentration (T-SSC) (Harmel et al., 2006). The Root Mean Square Error (RMSE) method estimates the "most probable value" of the cumulative or combined error by propagating the error from each measurement and modeling procedure to the final SSYEV calculation (Topping, 1972). The resulting cumulative probable error (PE) is the square root of the sum of the squares of the maximum values of the separate errors:

|  |  |  |
| --- | --- | --- |
|  |  | Equation 6 |
| where PE is the cumulative probable error for individual measured values (±%), EQmeas is uncertainty in Q measurements (±%), ESSCmeas is uncertainty in SSC measurements (± %), EQmod is uncertainty in Q modeled by the Stage-Q relationship (RMSE, as ±% of the mean observed Q), ESSCmod is uncertainty in SSC modeled by the T-SSC relationship (RMSE, as ± % of the mean observed SSC)(Harmel et al., 2009). | | |

EQmeas and ESSCmeas were estimated using lookup tables (LUT) from the DUET-H/WQ software tool (Harmel et al., 2009). The effect of uncertain SSYEV estimates may complicate conclusions about contributions from subwatersheds, anthropogenic impacts, and SSYEV-Storm Metric relationships. This is common in sediment yield studies where successful models estimate SSY with ±50-100% accuracy (Duvert et al., 2012) but the difference in SSY from undisturbed and disturbed areas was expected to be much larger than the cumulative uncertainty. PE was calculated for SSYEV from the UPPER and TOTAL watersheds, but not calculated for SSYEV from the LOWER subwatershed since it was calculated as the difference of SSYUPPER and SSYTOTAL.

## Results

### Field Data Collection

#### Precipitation

Annual precipitation measured at RG1, with gaps filled with data from Wx, was 3,502 mm, 3,529 mm, and 3,709 mm in 2012, 2013, and 2014, respectively, which are approximately 94% of long-term precipitation (=3,800 mm) from PRISM data (Craig, 2009). No difference in measured P was found between RG1 and Wx, or between RG1 and RG2, so P was assumed to be homogenous over the watershed for all analyses. Rain gauges could only be placed as high as ~300 m (RG2), though the highest point in the watershed is ~600 m. Long-term rain gage records show a strong precipitation gradient with increasing elevation, with average precipitation of 3,000-4,000 mm on the lowlands, increasing to more than 6,350 mm at high elevations (>400 m.a.s.l.) (Craig, 2009; Dames & Moore, 1981; Wong, 1996). Precipitation data measured at higher elevations would be useful to determine a more robust orographic precipitation relationship. For this analysis, however, the absolute values of total precipitation in each subwatershed are not as important since precipitation and the erosivity index are only used as predictive storm metrics.

#### Water Discharge

Discharge at both FG1 and FG3 was characterized by periods of low but perennial baseflow, punctuated by short, flashy hydrograph peaks (FG1: max 8,356 L/sec, FG3: max 13,071 L/sec) (Figure 6). Though Q data was unavailable for some periods, storm events appeared to be generally smaller but more frequent in the October-April wet season. Storm events during the May-September dry season were less frequent but larger. The largest event in the three year study was observed in August 2014.

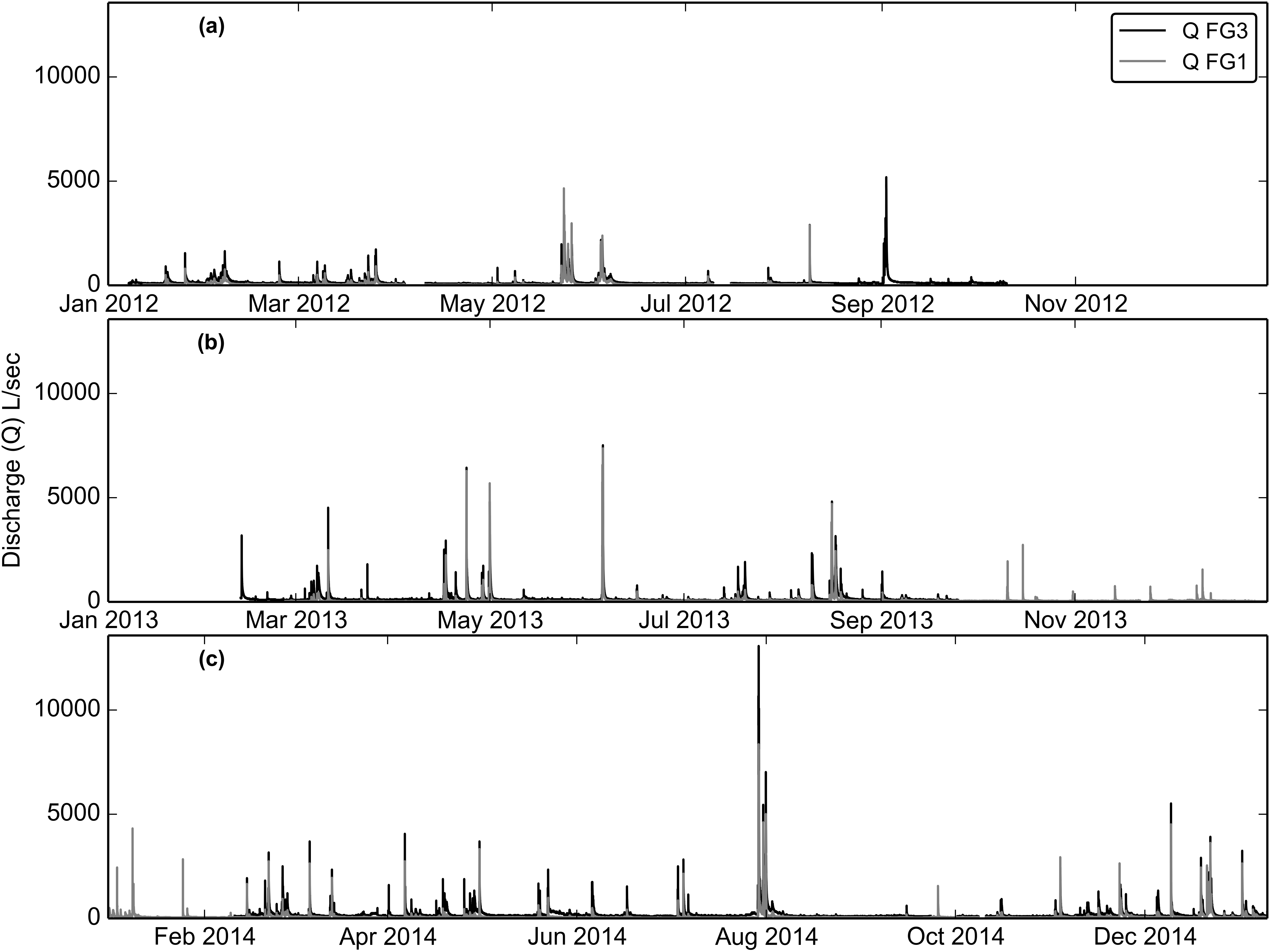


Figure 6. Time series of water discharge (Q), calculated from measured stage and the stage-discharge rating curves in a) 2012 b) 2013 and c) 2014.

#### Storm Events

Using the storm definition criteria, 210 events were identified from Q data at FG1 and FG3 between January, 2012, to July 2014; 169 events had simultaneous Q data at FG1 and FG3 (Appendix 3, Table 1). SSC data from T or interpolated grab samples were recorded during 112 events at FG1, and 74 events at FG3. Of those storms, 42 events had data for P, Q, and SSC at both FG1 and FG3 to calculate SSY from the LOWER subwatershed. SSY data from interpolated grab samples were collected at FG2 for 8 storms to calculate SSY from the LOWER\_QUARRY and LOWER\_VILLAGE subwatersheds separately. Storm event durations ranged from 1 hours to 2 days, with mean duration of 13 hours.

Most storm events showed a typical pattern, where a short period of intense rainfall caused a rapid increase in SSC downstream of the quarry (FG2) while SSC remained low at the undisturbed forest site (FG1)(Figure 7). The highest SSC was typically observed at FG2, with slightly lower and later peak SSC observed at FG3. SSC downstream of the undisturbed forest (FG1) typically increased more slowly, remained lower, and peaked later than the disturbed sites downstream of the quarry (FG2) and the village (FG3). Though peak SSC was highest at FG2, the highest SSY was measured at FG3 due to the addition of storm runoff from the larger watershed draining to FG3.

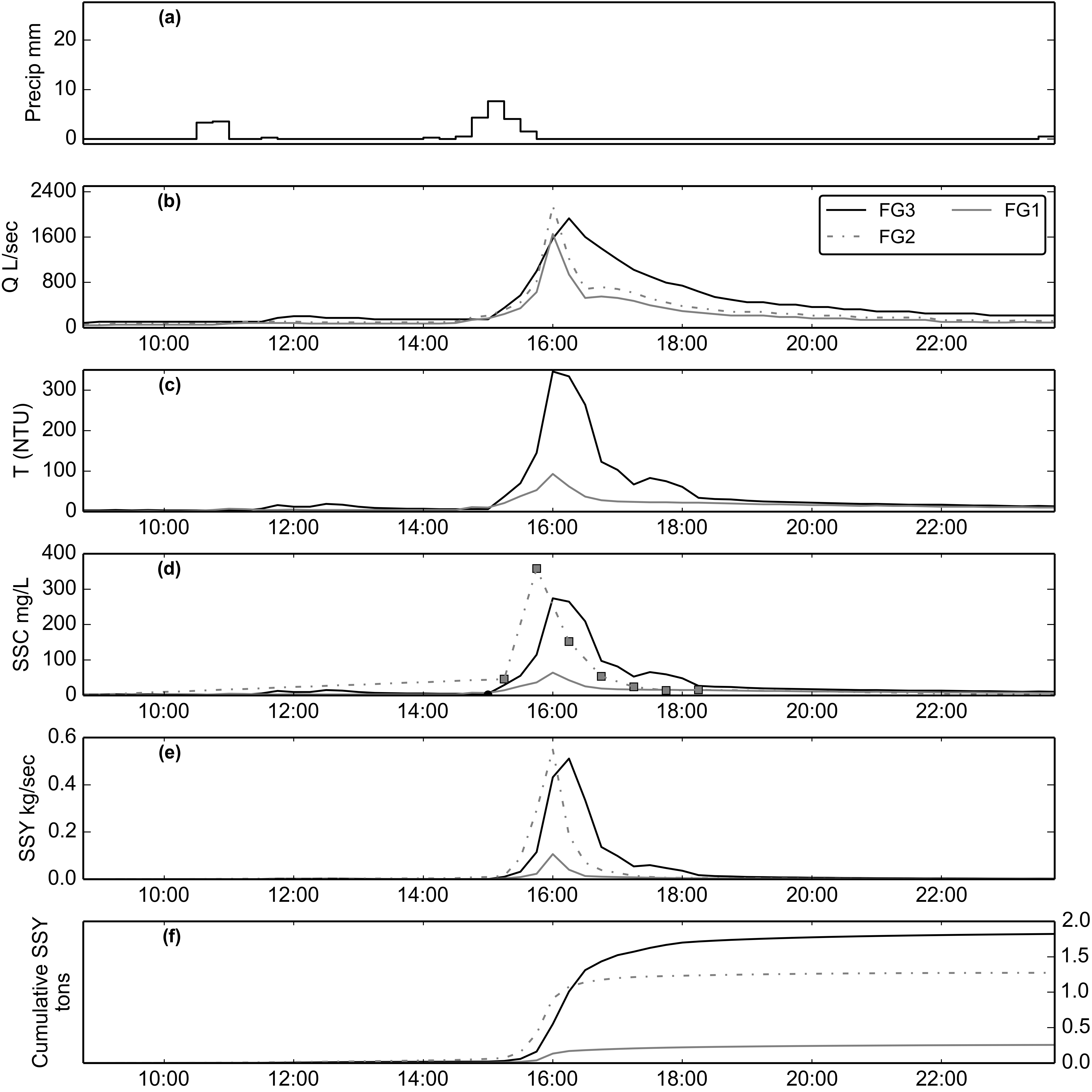


Figure 7. Example of storm event (02/14/2014). SSY at FG1 and FG3 calculated from SSC modeled from T, and SSY at FG2 from SSC samples collected by the Autosampler.

#### Suspended Sediment Concentration

Mean (μ) and maximum SSC of water samples, collected during non-stormflow and stormflow periods by grab and autosampler, were lowest at FG1 (μ=28 mg/L, max=500 mg/L), highest at FG2 (μ=337 mg/L, max=12,600 mg/L), and in between at FG3 (μ=148 mg/L, max=3,500 mg/L). At FG1, 0% of grab samples (n=14) were collected during non-stormflow (μ=8 mg/L (Figure 8a); 0% of grab samples (n=45) were collected during stormflow, μ= 35 mg/L (Figure 8b). At FG2, 0% of grab samples (n=21) were collected during non-stormflow (μ= 105 mg/L; 0% of grab samples (n=69) were collected during stormflow, μ= 409 mg/L. At FG3, 0% of samples (n=39) were collected during non-stormflow (μ= 52 mg/L; 0% of samples (n=120) were collected during stormflow, μ= 179 mg/L. This pattern of SSC values suggests that little sediment is contributed from the forest upstream of FG1, followed by a large input of sediment between FG1 and FG2, and then SSC is diluted by addition of stormflow with lower SSC between FG2 and FG3.

Probability plots of the SSC data collected at FG1, FG2 and FG3 showed they were highly non-normal, so non-parametric tests for statistical significance were applied. The Kruskall-Wallis test showed SSC samples from all three locations were significantly different for non-stormflow (p<0.000) and stormflow (p<0.000). The pair-wise Mann-Whitney test showed SSC samples were significantly different between FG1 and FG2 (non-stormflow, p=0.000; stormflow, p=0.000), but were not significantly different between FG2 and FG3 (non-stormflow, p=0.011; stormflow, p=0.576).

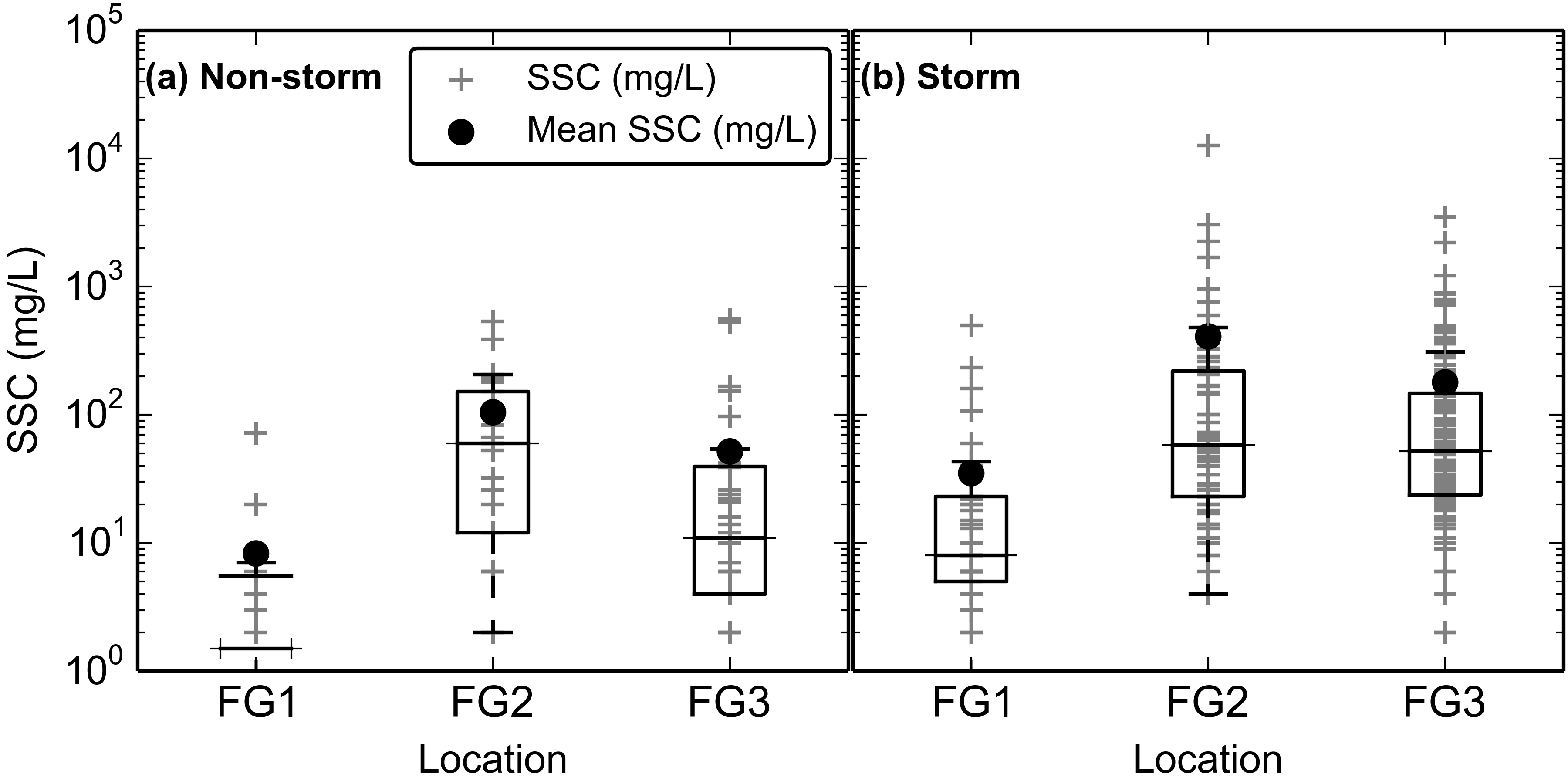


Figure 8. Boxplots of Suspended Sediment Concentration (SSC) from grab samples only (no Autosampler) at FG1, FG2, and FG3 during (a) non-stormflow and (b) stormflow.

SSC varied by several orders of magnitude for a given Q at FG1, FG2, and FG3 due to significant hysteresis observed during storm periods (Figure 9). At FG1, interannual variability of SSC during stormflow was assumed to be caused by randomly occurring landslides or mobilization of sediment stored in the watershed during large storm events. The maximum SSC sampled downstream of the undisturbed forest, at FG1 (500 mg/L), was sampled on 04/23/2013 at high discharge (QFG1= 3,724 L/sec) (Figure 9a). Anecdotal and field observations reported higher than normal SSC upstream of the quarry during the 2013 field season, possibly due to landsliding from previous large storms (G. Poysky, pers. comm.).

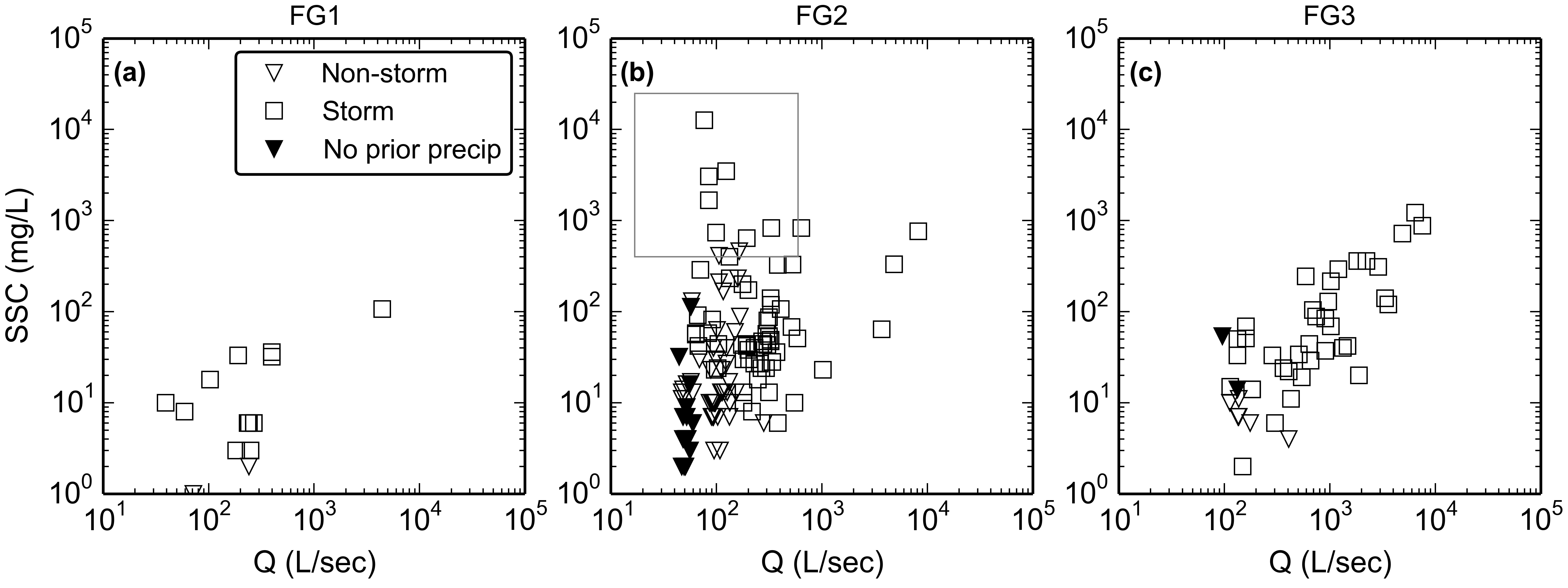


Figure 9. Water Discharge vs Suspended Sediment Concentration at a) FG1, b) FG2, and c) FG3 during non-stormflow and stormflow periods. The box in b) highlights the samples with high SSC during low flows, solid symbols indicate SSC samples where precipitation during the preceding 24 hours was 0 mm.

At FG2 and FG3, additional variability in the Q-SSC relationship was due to the changing sediment availability associated with quarrying operations and construction in the village. The high SSC values observed downstream of the quarry (FG2) during low Q were caused by two mechanisms: 1) precipitation events that did not result in stormflow, but generated runoff from the quarry with high SSC and 2) washing fine sediment into the stream during rock crushing operations at the quarry.

The maximum SSC sampled at FG2 (12,600 mg/L) and FG3 (3,500 mg/L) were sampled during the same rainfall event (03/05/2012), but during low Q (QFG3=287 L/sec)(Figure 9b-c). During this event, brief but intense precipitation caused high sediment runoff from the quarry, but did not increase Q above the defined storm threshold. SSC was diluted further downstream of the quarry at FG3 by the addition of runoff with lower SSC from the village.

Given the close proximity of the quarry to the stream, SSC downstream of the quarry can be highly influenced by mining activity like rock extraction, crushing, and/or hauling operations. During 2012, a common practice for removing fine sediment from crushed aggregate was to rinse it with water pumped from the stream. In the absence of retention structures the fine sediment was then discharged directly into the stream, causing high SSC during baseflow periods with no precipitation in the preceding 24 hours (solid symbols, Figure 9b-c). Riverine discharge of fine sediment rinsed from aggregate was discontinued in 2013, corresponding with low SSC during low Q in 2013 (Figure 9b-c). In 2013 and 2014, waste sediment was piled on-site and severe erosion of these changing stockpiles caused high SSC during storm events.

#### Cumulative Probable Error (PE)

The measurement error (RMSE) was 8.5 % for Q at FG1 and FG3 from the DUET-H/WQ LUT (Harmel et al., 2006), which included error in the area-velocity measurements (6%), continuous Q measurement in a natural channel (6%), pressure transducer error (0.1%), and streambed condition (firm, stable bed=0%). The model errors (RMSE) were 32% for the stage-Q rating curve using Manning's equation at FG3, and 22% using HEC-RAS at FG1.

The measurement errors (RMSE) for SSC measurements from the DUET/WQ LUT, were 16.3% for sample collection, which included error from interpolating over a 30 min interval (5%), sampling during stormflows (3%), and 3.9% for sample analysis which included measuring SSC by filtration (3.9%). The model errors (RMSE) of the T-SSC relationships were 16.0% (4 mg/L) for the YSI and TS at FG1, 113.0% (348 mg/L) for the YSI at FG3, and 46.0% (48 mg/L) for the OBS at FG3. Cumulative Probable Error (RMSE %) for SSY estimates at FG1 and FG3 were calculated from the measurement errors for Q (8.5%) and SSC grab samples (16.3%), and the model errors of the respective stage-Q and T-SSC relationships for that location. Cumulative Probable Errors (PE) in SSYEV were 28-49% (μ=43%) at FG1 and 36-118% (μ=94%) at FG3.