EVALUATION OF GSSHA FOR SIMULATING SEDIMENT CONCENTRATIONS IN STEEP HAWAIIAN WATERSHEDS

A THESIS SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAIʻI AT MĀNOA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

IN

CIVIL ENGINEERING

MAY 2016

By

Jeffrey R. Nielson

Thesis Committee:

Oceana P. Francis, Chairperson
Roger W. Babcock, Jr.
Sayed M. Bateni

ACKNOWLEDGMENTS

I would like to acknowledge and thank the people who have helped me on my path to completing this thesis. First, I would like to thank my advisor Oceana Francis, for bringing me out to the University of Hawaii and taking me on as a Research Assistant, and the other members of my committee, Roger Babcock Jr. and Sayed Bateni, for working with me on several substantive engineering projects, including the developing a retrofit design for the outlet structure of Honokowai dam in west Maui, which led to the research questions addressed herein.

Second, I would like to thank Steve Turnbull for providing guidance on the use of GSSHA, while I was getting started, and Chuck Downer for visiting the University of Hawaii, while here on Oahu for other business, to chat about GSSHA.

I would also like to thank all the undergraduate and graduate students who I have worked with, especially Kim Falinski, for collaborating with me on projects.

Third, I would like to thank several organizations, including the State of Hawaii Department of Land and Natural Resources, Maui County, the National Fish and Wildlife Foundation, and the Kahoolawe Island Reserve Commission, for granting access to lands, and providing funding for my research.

Lastly, I would like to thank my wife, Amy Nielson, for supporting, encouraging, and inspiring me at home.

ABSTRACT

Physics-based hydrological models, like GSSHA, potentially have an accuracy advantage over empirical or semi-empirical models, when simulating processes in steep Hawaiian watersheds, which can be very dissimilar to lands (or field plot experiments) where the empirical relationships were developed. GSSHA, therefore, could potentially be a valuable tool for hydrology projects in Hawaii, but is seldom used in Hawaii. In this study, GSSHA was evaluated and used to predict streamflow and sediment concentrations in the upper region of Halawa watershed on leeward Oahu—a watershed with an average overland basin slope of 0.65—during single-event storms. Single-event simulations were the focus here, because of the importance of single-event storms to Hawaii's sediment issues (other studies claim that a large portion of annual sediment flux occurs during low-frequency, high-magnitude storms). Streamflow and sediment concentration data collected at 15-minute intervals were used to calibrate and validate the simulations. Several case studies were developed to test findings from past studies, including the finding that GSSHA tends to over predict sediment for low-frequency, high-magnitude storms, when calibrated with only high-frequency, low-magnitude storms (as stated in the GSSHA manual). It was found that GSSHA becomes less accurate in sediment concentration predictions for low-frequency storms (>2-year recurrence interval) when calibrated with only high-frequency storms (<2-year recurrence interval). It is shown that the degradation in accuracy was likely a result of over prediction in streamflow. It was also found that GSSHA made more accurate predictions of streamflow when the validation storm events were similar to calibration events in rainfall duration, accumulation, and peak intensity, indicating that one set of calibration parameters may not be sufficient for all storm events. Distinctively, however, it is shown that one

set of calibration parameters may be sufficient for all storm events for predicting sediment concentrations. GSSHA predicted sediment load with an 8.8% error for a validation storm six times larger (based on streamflow) than calibration. Additionally, in response to studies that have shown that many sediment transport equations over predict sediment flux in steep terrain by orders of magnitude, two sediment transport equations—Engelund-Hansen and Kilinc-Richardson—were tested and compared to determine which equation is better for steep terrain. It was found that the Engelund-Hansen equation consistently outperformed the Kilinc-Richardson equation, indicating that the Engelund-Hansen equation is a better choice for use in steep Hawaiian watersheds.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Predicting sediment concentration or discharge from watersheds is often fundamental in research pertaining to land-use change impacts, watershed management strategies, downstream water quality, health of aquatic ecosystems, and, *inter alia*, landscape evolution (*e.g.*, Huang 2015; Krysanova 2014; U.S. Geological Survey Field Center 2007; Lenat 1994; Stark 2001; Fabricius 2005 and references therein). The findings frequently support substantive decisions on, *e.g.*, the development of lands, selection of best management practices, preservation of natural resources, allocation of government funds, and environmental policy making (*e.g.*, Xie 2015; U.S. Geological Survey Field Center 2007; Dillaha 1988). Such studies (referred to here as "watershed sediment studies"), therefore, are of unambiguous importance.

Watershed sediment studies typically rely on models to make predictions about sediment concentrations and loadings. Commonly used models, referred to almost exclusively, as here, by their acronyms, include SWAT, HSPF, AGNPS, WEPP, MIKE-SHE, KINEROS, and GSSHA (for a quantitative comparison see Kalin 2003). Each of these models can be used to predict sediment discharge from watersheds, but each make predictions in different ways (Kalin 2003). For example, some of the models are empirical, while others are semi-empirical, and still others are physically-based, relying on conservation of mass or momentum laws, instead of experimental relationships. Because of these differences, there is not one model that is best for all applications. For this reason, and given the importance of models in watershed sediment studies in support of decision making, models are continually being created, evaluated, reviewed, and improved through research.

The Gridded Surface Subsurface Hydrologic Analysis (GSSHA) is a newer model from the list above, developed by the U.S. Army Corps of Engineers, with the capability of simulating sediment concentration, along with other watershed processes. GSSHA has many features that make it appealing for watershed sediment studies, including a basis in physics (conservation of mass equations are used in place of empirical relationships), process-based simulation, spatial and temporal distribution of parameters, and the choice of several different sediment transport equations (namely, Kilinc-Richardson and Engelund-Hansen) (Downer 2006). These features theoretically give GSSHA an accuracy advantage over empirical or semi-empirical models (which describes most of the other models listed above) in simulating processes of watersheds that are dissimilar from the watersheds (or field plot experiments) where the empirical relationships were developed (Ogden 2001, referring to GSSHA's predecessor CASC2D).

Hawaii is one such place, where watersheds can be very dissimilar from where empirical models are developed. Hawaii's watersheds are unique in topography, especially average steepness or slope, average watershed area (many small and intermediate watersheds), and spatial and temporal rainfall distribution (*e.g.*, Sahoo 2006, Giambelluca 2013). None of the empirical models listed above were developed for the specific application of modeling Hawaiian watersheds (Kalin 2003). Researchers could, therefore, potentially have an accuracy advantage over empirical models in using GSSHA to model Hawaiian watersheds. Moreover, there are a significant number of watershed sediment studies in Hawaii, pertaining to the management of coral reefs alone (U.S. Geological Survey 2015a), showing a need for the development and validation of models for Hawaii. Further, GSSHA can be used to simulate single-events, which are highly important to Hawaii's sediment issues, as shown by studies that claim that a large percentage of the annual watershed sediment load occurs during single-event, low-frequency,

high-magnitude storms (Sahoo 2006; Sustainable Resources Group 2012; Markus 2006). For these reasons, GSSHA is evaluated here, with particular emphasis on GSSHA's ability to predict sediment concentrations for single-events, in an intermediate-sized (5 mi^2), steep, Hawaiian watershed.

Additional motivations for this evaluation of GSSHA are (1) calls, from a body of literature related to GSSHA (summarized below), for more testing of GSSHA in predicting watershed flows and sediment (Downer 2003; Downer 2004; Shurtz 2009; McCarthy 2012; McCarthy 2013; Thompson 2014), including a cautionary statement regarding inaccurate sediment simulation in the GSSHA manual (Downer 2006), and (2) studies that show that *most* sediment transport equations tend to over predict sediment discharge by orders of magnitude (Yager 2012 and references therein), which begs the questions, (i) whether the equations in GSSHA over predict sediment in steep terrain, and (ii) which equation (Engelund-Hansen or Kilinc-Richardson) is more accurate in steep terrain for use in Hawaii.

1.1 GSSHA

GSSHA is a derivative of the Cascade Two-Dimensional Model (CASC2D), a two-dimensional, physically based, distributed parameter hydrological model (Downer 2006). Notably, in contrast to CASC2D, each process simulated by GSSHA has an independent time-step and update time, which increases computational efficiency, by enabling simultaneous simulations of processes with different time-steps (Downer 2006). GSSHA is capable of modeling hydrologic, hydraulic, and groundwater systems, for long-term studies, and single events, which were the focus here. Since its introduction in 2006 (version 1.43), GSSHA has been gaining traction as a useful and reliable hydrologic model. As evidence, in 2013 it was added to the Federal Emergency

Management Agency's list of models that meet minimum requirements for the National Flood Insurance Program (Federal Emergency Management Agency 2015).

Additionally, GSSHA has a growing body of peer-reviewed published applications, beginning several years before the introduction of the current version, which prove GSSHA's usefulness in different situations, and point to areas where evaluation is still needed (Downer 2003; Downer 2004; Sharif 2004 (referring to the pre-GSSHA CASC2D model); Sharif 2006; Swain 2013; Saha 2015). For example, in Downer (2003), GSSHA was evaluated for predicting runoff and soil moisture at watershed scale. GSSHA was calibrated and verified with only watershed discharge measurements, and then was used to estimate soil moistures. *Downer* found that GSSHA accurately predicted soil moistures during growing season, but that GSSHA under predicted outside of growing season, for unknown reasons, and calls for more research in this area.

Similarly, in Downer (2004), GSSHA was evaluated for predicting streamflow. 23 events were simulated (12 for calibration, 11 for validation) in succession for a single watershed. *Downer* found that GSSHA accurately predicted streamflow with 53% volume error and 15% peak error in calibration, and 17% volume error and 31% peak error in validation. Here, again, *Downer* calls for more application evaluations at different sites.

Also, GSSHA's predecessor CASC2D was used to facilitate the evaluation of errors in the estimation of radar rainfall data (Sharif 2004), and GSSHA was used to show the usefulness of radar rainfall data as part of a *nowcasting* system to improve flood warning and forecasting (Sharif 2006). In the latter study, percent errors of 17% in peak, and 25% in volume, and 9% in magnitude of the hydrograph's rising limb, were reported. (These were single-event simulations.)

Significantly larger errors were mentioned with simulations involving storm extrapolation, but no percent error calculations or other figures were given.

As a final example, Saha (2015) evaluated GSSHA in predicting monthly, seasonal, and annual streamflow, under climate-change rainfall scenarios. *Saha* ran continuous-event simulations and calibrated GSSHA with streamflow data from 2006 to 2010 and validated from 2010 to 2011. *Saha* did not report any percent error calculations for model performance, but did report a coefficient of determination or R2=0.62 and a Nash-Sutcliffe efficiency or NSE=0.59 for calibration, and R2=0.65 and NSE=0.61 for validation, and concluded satisfactory performance.

GSSHA has also been the focus in a number of graduate-level theses, which have evaluated GSSHA for predicting sediment discharge, snowmelt-induced streamflow, streamflow prediction compared to HEC-HMS, for long-term continuous- and single-events (Shurtz 2009; McCarthy 2012; McCarthy 2013; Thompson 2014).

For example, in Shurtz (2009), GSSHA was evaluated against HEC-HMS, a model with a longer-history of use and successful applications. GSSHA and HEC-HMS were calibrated and validated by simulating single-event storms from several watersheds. GSSHA validation percent error was reported to range from 2.2% to 33.2% for peak flow, and <0.1% to 67.3% for flow volume, across all simulations (but also reports over predictions in one validation of over 100%, and one validation failed to produce any runoff, which indicates a 100% under prediction error of flow volume and peak). Also, *Shurtz* adjusted initial moisture values to improve the fit of validations. *Shurtz*, having found that GSSHA performed well in one of three validations, concluded that GSSHA is unreliable when calibrated with a single event and calls for more research into single-event calibrations and how GSSHA can be used for single-event simulations.

McCarthy (2012) is one of the only reports available that is an evaluation of GSSHA's ability to simulate sediment discharge from watersheds. *McCarthy* used GSSHA to predict sediment discharge in a watershed in the Dominican Republic. GSSHA was calibrated for sediment using USLE calculations, not observations. The Kilinc-Richardson (KR) equation and Slope and Unit Discharge (SUD) method were compared. Approximately a one-year period was simulated. The results showed a large difference between the KR and SUD predictions, with the SUD predicting 5491.2% more sediment than KR. One key finding in *McCarthy* was that the dimensionless Erodibility Coefficient is a very sensitive soil erosion parameter that is useful for sediment calibration. *McCarthy* also advocated calibrating with long-term events and allowing a model *ramp up* period before simulations.

In McCarthy (2013), GSSHA was evaluated while predicting streamflow in long-term simulations including snowmelt. Large errors, however, were reported and no conclusions were made, aside from calls for more software and model development and evaluation.

Finally, in Thompson (2014), GSSHA was evaluated for simulating flows in snowmelt-dominated watersheds, as a continuation of McCarthy (2013). Again, GSSHA was used to simulate snowmelt processes for continuous long-term simulations. *Thompson* found that it was difficult to match peak flow, peak timing, and flow volume during calibration, and proposed a method of achieving a better calibration for flows with snow-melt processes.

These publications and theses have helped to define both the usefulness and limitations of GSSHA. They have shown how GSSHA can be used successfully in several specific applications, where GSSHA could potentially be improved, and where future research is needed. They, however, mostly focus on GSSHA's ability to predict runoff, soil moistures, streamflow, and streamflow processes, but not sediment. The one study above that is related to sediment

shows that more testing is needed before GSSHA can be relied upon to predict sediment discharge (McCarthy 2012).

Moreover, GSSHA developers point directly to potential weaknesses in GSSHA for sediment predictions. In the GSSHA manual, developers state that GSSHA has been observed to over predict sediment discharge from lower-frequency, high-magnitude storms when calibrated with high-frequency, low-magnitude storms (Downer 2006). This observation is based on a finding from a study on the CASC2D model by Ogden (2001), and was attributed to a lack of particle detachment limits, errors in hydrologic predictions, and errors in initial soil moistures. No quantification of over prediction, guidance on the proper use of GSSHA (if any), or anything more specific is given by the manual (Downer 2006).

Downer (2010), however, tested an updated version of GSSHA, with a new sediment transport formulation, and found that GSSHA is not only capable of predicting sediment from lower-frequency storms, when calibrated with only higher-frequency storms, but is able to simulate *storms of all sizes* (satisfactory accuracy implied). *Downer* tested a storm that was four times larger than calibration storms, found a maximum percent error in sediment volume of -47%, and concluded that there was no deterioration of accuracy (error range for sediment volume in calibration was -73% to 220%). These findings are encouraging, but no other GSSHA studies related to sediment prediction were found that corroborate Downer (2010).

Thus, based on this literature review, plenty of room was left for the evaluation here, which was necessarily about streamflow, but was focused on sediment transport, and attempted to validate GSSHA as a model capable of accurately predicting sediment concentrations in steep Hawaiian watersheds on the single-event scale.

1.1.1 GSSHA Sediment Processes

GSSHA uses the following governing mass balance equation for overland sediment routing:

$$\frac{\partial(q * c)}{\partial x} + \frac{\partial(q * c)}{\partial y} + \frac{\partial(h * c)}{\partial t} = E(x.y.t) + q(x, y, t)$$

where c is the sediment concentration, h is water depth, q is the unit discharge, E is sediment from sources (the sum of particles detached by rainfall and runoff), and Q is lateral sediment flux into channels.

GSSHA calculates sediment particle detachment by rainfall and runoff. GSSHA calculates a detachment capacity rate, D_i, for sediment particles detached by raindrops as a function of rainfall intensity, I, with an empirical exponential coefficient, β, and several factors that are commonly used in empirical models. This process is similar to Gabet (2003). D_i can be written as:

$$D_i = K_I C_W C_G C_i I^{\beta}$$

where K_I is a soil erodibility factor (for detachment by rainfall), C_W is a water depth factor, C_G is a cover factor, and C_i is a crop management factor.

GSSHA also calculates a detachment capacity rate, D_r, for sediment particles detached by runoff in relation to flow shear stress, τ, the sediment load, G, and the runoff sediment transport capacity, T. D_r can be written as:

$$D_r = a(\tau - \tau_{critical})^b \left(1 - \frac{G}{T}\right)$$

where a and b are empirical values.

1.1.2 GSSHA Sediment Transport Equations

GSSHA allows selection between several sediment transport equations. The equations used in this evaluation are the Kilinc-Richardson (KR) and Engelund-Hansen (EH).

The KR equation can be used to calculate sediment discharge from overland flow with respect to unit flow and soil and land cover properties. The KR equation for sediment discharge (q_s), as it appears in GSSHA, is as follows:

$$q_s = 25.5 q^{2.035} S^{1.664} \frac{K_d}{0.15}$$

where, q is the unit discharge of overland flow, S is the surface friction slope (*e.g.*, from Manning's Equation), and K_d is the dimensionless *Erodibility Coefficient*, composed of three empirical factors, similar to those of USLE, namely the Erodibility Factor (K), Cropping Management Factor (C), and the Conservation Practice Factor (P). This equation is similar to the

modified KR equation developed by Julien (1995), with the exception of the combining of the three empirical factors into K_d (Julien 1995).

The EH equation can also be used to calculate sediment discharge in overland flow, but quite differently than the KR equation. The EH equation calculates the volumetric sediment discharge with respect to particle size, as follows:

$$G_i = KF_i \frac{0.05BV^2h^{3/2}S^{3/2}}{(sg-1)^2 D_i\sqrt{g}}$$

where, *i* indicates the *i*th size fraction, G_i is the volumetric rate of sediment transport, *K* is a calibration coefficient, F_i is the proportion of material to total load, *B* is the flow width, *h* is the flow depth, *V* is the mean flow velocity, *S* is the slope of the water surface, *sg* is the specific gravity, D_i is the mean particle size, and *g* is gravitational acceleration (Engelund 1967).

GSSHA offers several other sediment transport methods, namely Slope and Unit Discharge (SUD), Unit Stream Power, Effective Stream Power, and Shear Velocity (collectively called the "shear stress methods"). These shear stress methods, however, were omitted from this study based on their accuracy in preliminary testing. Each of these methods over predicted sediment concentrations by orders of magnitude above the preliminary tests of EH and KR. This finding is similar to McCarthy 2012, which found that SUD drastically over predicted sediment in a case study of a watershed in the Dominican Republic.

Calibration is also achieved differently between EH and KR, and the shear stress methods (McCarthy 2012). The EH and KR are highly sensitive to the erodibility coefficient, *K*, which is composed of empirical parameters, making it a good choice to manipulate during calibration,

because these values are normally obtained from a table and not from site measurements. The shear stress methods do not depend on the erodibility coefficient, but are more sensitive to particle size and fraction.

1.2 OBJECTIVES

1. Develop case studies to evaluate GSSHA's accuracy in predicting sediment concentrations with an emphasis on single-event applications in steep Hawaiian watersheds.
    a. Test whether GSSHA over predicts sediment concentrations for lower-frequency storms when calibrated with higher-frequency storms, in support of the findings of Ogden (2001) and as stated in the GSSHA Manual (Downer 2006), or whether GSSHA accurately predicts sediment concentrations in these circumstances, in corroboration of Downer (2010).
    b. Compare the Engelund-Hansen and Kilinc-Richardson sediment transport equations and determine which equation is more accurate for steep Hawaiian watersheds.
2. Based on the outcomes from the case studies in Objective 1, determine whether GSSHA can be relied upon to predict watershed sediment concentration in steep Hawaiian watersheds during single-events.
3. Identify potential improvements for GSSHA, in general, and for use in Hawaii specifically.

CHAPTER 2

METHODS

2.1 GSSHA MODEL DEVELOPMENT

A GSSHA model was developed for upper Halawa watershed. The development process is described below.

2.1.1 Watershed Modeling System (WMS)

Aquaveo's WMS was used to develop and calibrate a GSSHA model of upper Halawa. WMS is a graphical user interface for many popular hydrologic models, including GSSHA. WMS features GIS tools, which were used to import topography, soil, and land cover data, and compute flow directions and accumulations. WMS also features an automated calibration module, which was used to inversely model watershed parameters, including stream channel roughness, overland roughness, initial soil moisture, and hydraulic conductivity, based on observed rainfall and streamflow. Most of the images in this section, below (e.g., GSSHA grid, index maps), were generated using WMS, including the following figure of upper Halawa watershed (a forested, mountainous, leeward Oahu watershed) delineated with the elevation layer and Landsat imagery.

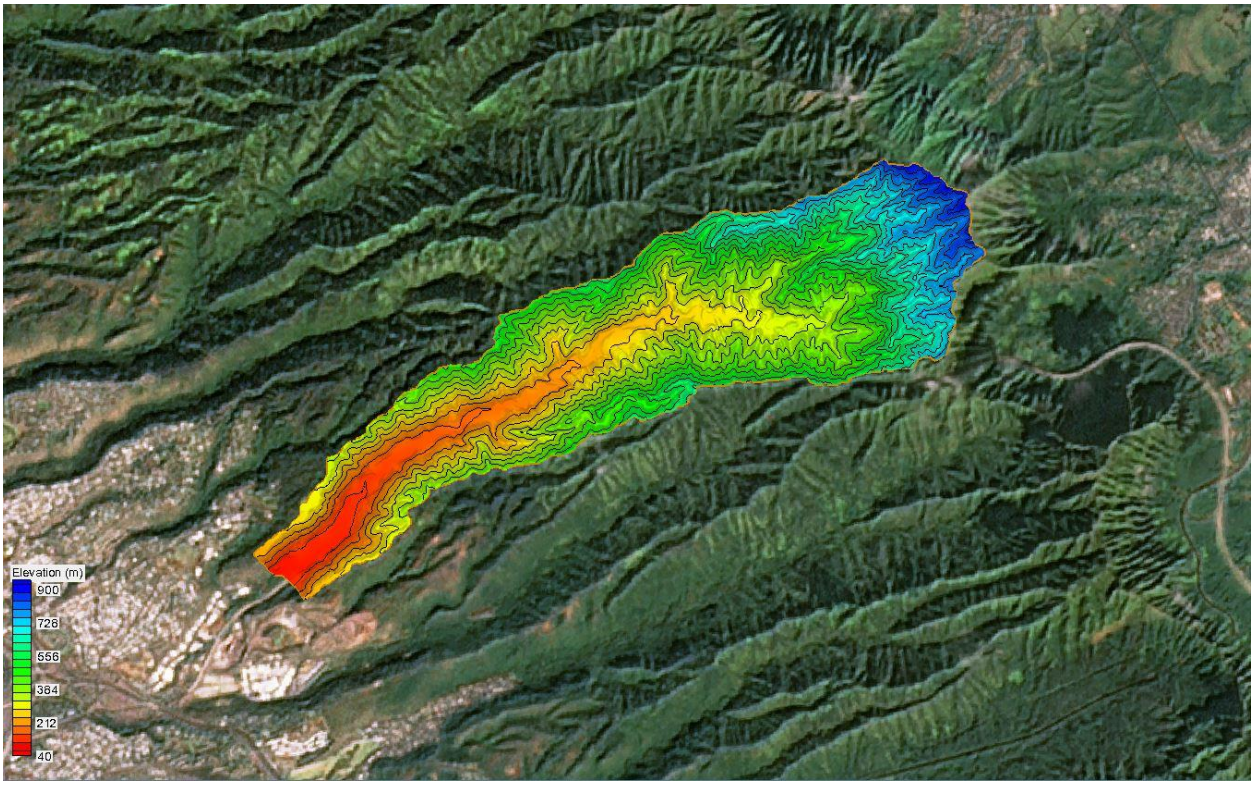

Figure 1: Upper Halawa watershed delineated with WMS.

2.1.2 GSSHA Grid

Using the WMS interface, a GSSHA grid was created from a 10-meter digital elevation map (DEM), to represent the upper Halawa Watershed. The DEM was obtained from the Hawaii Statewide GIS Program (http://planning.hawaii.gov/gis/download-gis-data/). A 10-meter cell size was specified for the GSSHA grid, because it is the minimum cell size, and should maximize the spatial resolution, and, therefore, model accuracy. A larger cell size could be used to reduce simulation time, which may be beneficial in some studies—especially long, continuous-event studies. Here, however, where only single-event storms were simulated, simulation time was not an issue. As a note, researchers have found the maximum cell size to be approximately 150 meters, with respect to accuracy in sediment transport modeling (Rojas 2008, referring to the

CASC2D model). Thus, a 10-meter cell size was determined to be the best choice for this application.

2.1.3 Stream Network

A stream network was developed for the GSSHA grid. Stream segments were created using the United States Department of Agriculture's Topographic Parameterization Program (TOPAZ) and transferred to the GSSHA grid. TOPAZ (included in WMS) computes flow accumulation and direction from elevation data. A minimum flow accumulation threshold of 15 hectares was found, through iteration, to best represent Halawa's stream network. After creation, each stream segment was classified with a feature type of *trapezoidal channel*, so that dimensions and parameters could be assigned. Initial estimates for channel roughness values and approximate stream dimensions were assigned, including channel depth, width, and side slope. Lastly, to enhance the functionality of the stream network, stream segments were reordered, smoothed, eliminating any *digital dams*, or regions of the grid that unrealistically stop the flow of runoff and sediment, and vertices were redistributed. The stream network shown in the figure below is the product of TOPAZ.

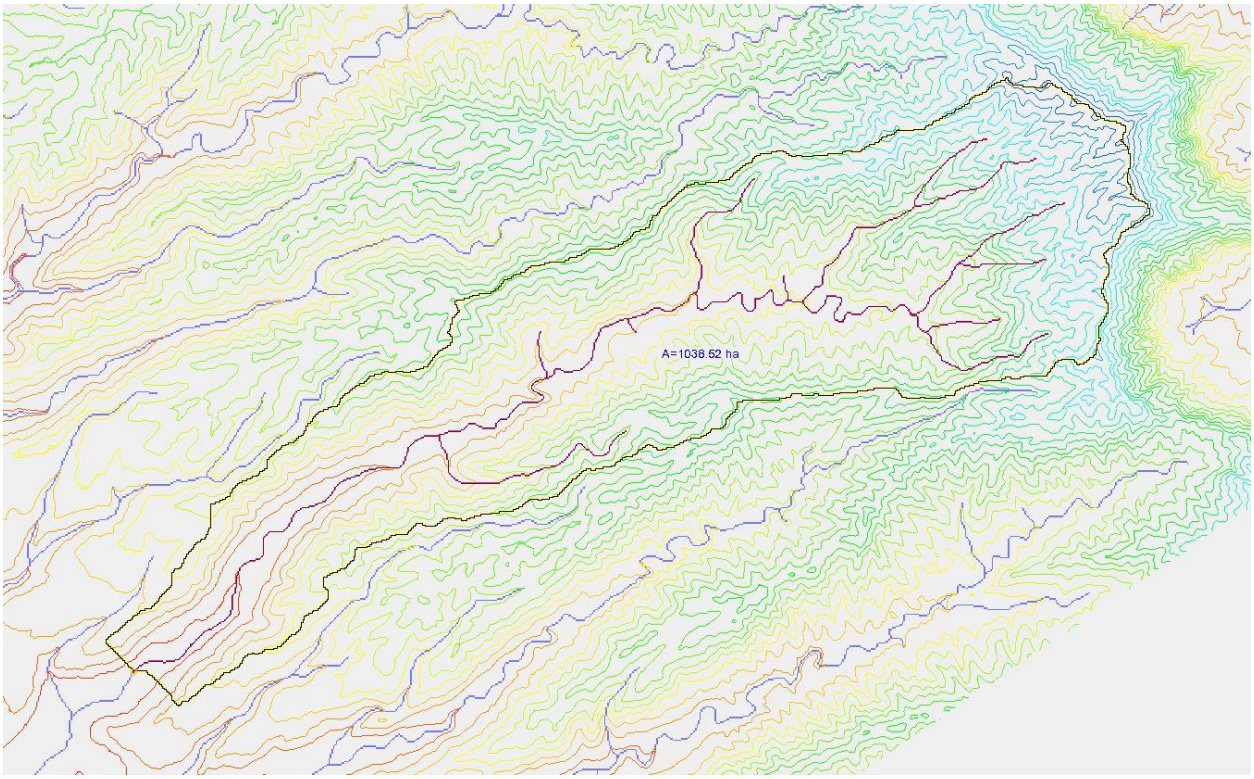

Figure 2: TOPAZ computed stream network.

2.1.4 Coverages and Index Maps

GSSHA *coverages* and *index maps* were derived. Coverages are feature layers that store parameter data by polygonal region and can be used to generate index maps. Index maps are grids, which correspond to the GSSHA grid, and are used to store parameter values cell-by-cell. The index maps facilitate spatial distribution of parameters and cell-by-cell calculations of watershed processes.

First, coverages for land cover and soil type were created by importing land use and soil type *shapefiles*, then by *mapping* necessary attributes to the respective layers. A land cover shapefile was created from a NOAA 2011 C-CAP image, using ArcGIS software. Soil type shapefiles were obtained from the Natural Resources Conservation Service, Web Soil Survey (Soil Survey Staff 2015). For the land use coverage, land cover code and name were mapped

from the land cover shapefile. Similarly, for the soil type coverage, texture, SCS soil type, hydraulic conductivity, initial moisture, field capacity, and wilting point values were mapped from the soil type shapefile.

Second, by spatial overlay, the shapefile-derived coverages were used to generate grid-based index maps. Index maps were generated for land cover and soil type, individually. A third index map was created with a combination of the two coverages, because certain watershed processes require both land cover and soil type data, *e.g.*, infiltration and soil erosion. The index maps were used to define parameter values for land cover and soil type at each cell of the GSSHA grid. The land cover index map is shown below.

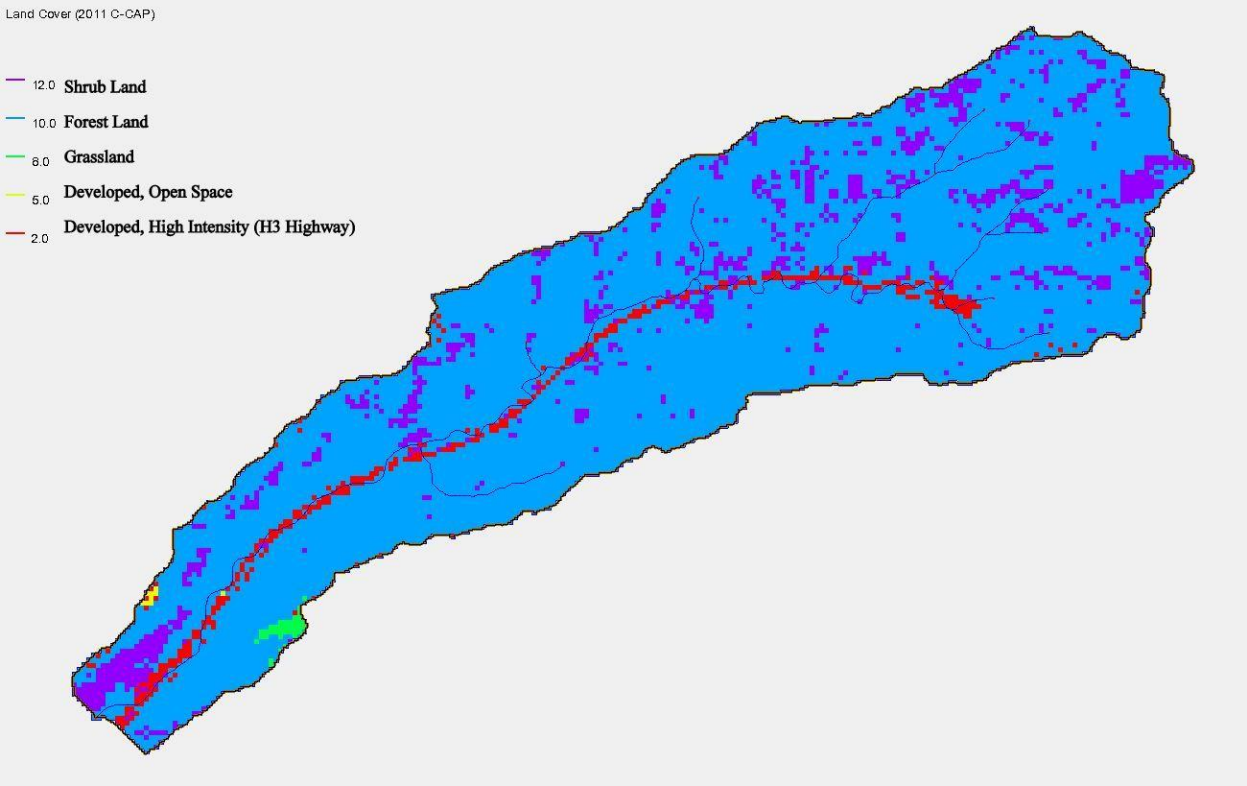

Figure 3: Land cover index map of upper Halawa.

2.1.5 Watershed Process Selection and Settings

Select watershed process modules were activated through the WMS Job Control menu. The modules activated, include overland flow (Alternating Direction Explicit) with Retention Depth, Infiltration (Green and Ampt with soil moisture redistribution), Evapotranspiration (Penman Method), Channel Routing (diffusive wave), and Soil Erosion (using Kilinc-Richardson and Engelund-Hansen sediment transport equations for each simulation). The total time for each simulation was 4320 minutes, to cover the duration of the storms simulated (described below in the *Storm Selection* section), and with a time step of 30 seconds. The following figure shows the WMS Job Control menu for GSSHA, where modules can be activated.

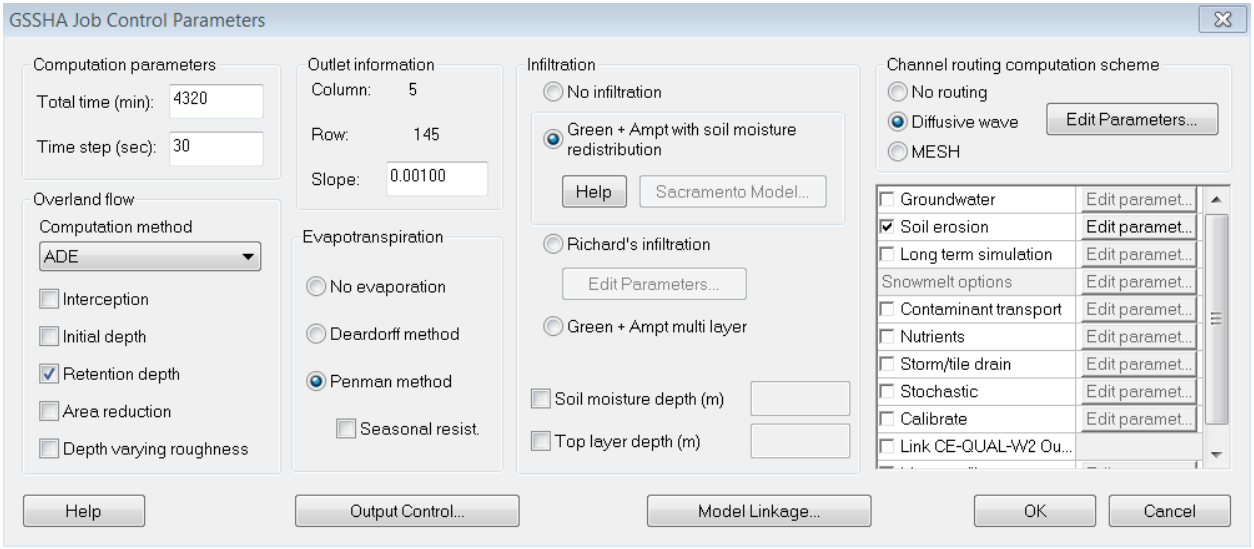

Figure 4: GSSHA Job Control menu.

2.1.6 Data Sources

As mentioned above, the elevation data was obtained from the Hawaii Statewide GIS Program. Also, the land cover and soil type data was obtained from a 2011 NOAA C-CAP image and the Web Soil Survey, respectively. These sources supplied data including, field capacity, wilting

point, and starting points for initial moisture and hydraulic conductivity. GSSHA, however, requires additional data to facilitate the selected processes, including overland roughness, capillary head, porosity, and coefficient for soil detachment by rainfall, to name a few. The table below is a list of the data requirements, type, and sources, separated by process. For data type, *specific* indicates that the data was obtained from the Halawa watershed, and *estimate* indicates that the data is an approximate value from a similar study area, a recommended value, or a default value.

Table 1: GSSHA Parameters and Data Sources

Overland Flow (ADE method)		
Parameter	Data Type	Data Source
Roughness (n) (for each land cover type)	Estimate	Engman (1986)

Infiltration (Green and Ampt w/ soil moisture redistribution)		
Parameter	Data Type	Data Source
Hydraulic conductivity (cm/hr)	Specific	Soil Survey Staff (2015)
Capillary head (cm)	Estimate	Rawls and Brakensiek (1989)
Porosity (m^3/m^3)	Estimate	Rawls and Brakensiek (1989)
Pore distribution index (cm/cm)	Estimate	Rawls and Brakensiek (1989)
Residual Saturation (m^3/m^3)	Estimate	Rawls and Brakensiek (1989)
Field Capacity (m^3/m^3)	Specific	Soil Survey Staff (2015)
Wilting point (m^3/m^3)	Specific	Soil Survey Staff (2015)

Channel Routing/Channel Erosion (Diffusive Wave)		
Parameter	Data Type	Data Source
Sediment Porosity	Estimate	WMS Default
Water Temp (C)	Estimate	WMS Default
Sand Size (mm)	Estimate	WMS Default

Soil Erosion (Kilinc-Richardson and Engelund-Hansen method)		
Parameter	Data Type	Data Source
Coeff for detachment by rainfall (1/J)	Estimate	Wicks and Bathurst (1996)
Rill erodibility coeff (s/m)	Estimate	WEPP Model Manual
Rill erodibity exponent (dimensionless)	Estimate	WEPP Model Manual
Critical rill detachment (Pa)	Estimate	GSSHAwiki Editors
Erodibility coeff (dimensionless)*	Estimate	(see below)
%Gravel	Specific	Particle Size Analysis
%Sand	Specific	Particle Size Analysis
%Silt	Specific	Particle Size Analysis
%Clay	Specific	Particle Size Analysis
*the erodibility coefficient is the product of the following three parameters		
Erodibility Factor (K)	Estimate	Parveen (2012)
Cropping Management Factor (C)	Estimate	Tirkey (2013)
Conservation Practice Factor (P)	Estimate	Tirkey (2013)

Initial Moisture		
Parameter	Data Type	Data Source
All soil types (silty clay, loam, bedrock, etc.)	Specific	Soil Survey Staff (2015)

The sources (whether specific or estimates) for some parameters listed, including initial moisture, hydraulic conductivity, and overland roughness were used for initial values, but these parameters were ultimately determined through calibration, which is described below. Parameters not listed, including channel roughness, were also estimated during calibration.

Rainfall, streamflow discharge, and sediment concentration data used for calibration and testing in this study were obtained from the United States Geological Survey's *Water Data for the Nation* service (U.S. Geological Survey 2015b). These data can be found below in the *Storm Data* section.

2.1.7 Automated Calibration

WMS offers an automated calibration module for GSSHA, which was used to calibrate the Halawa model. Several calibration methods can be selected (*e.g.*, Shuffled Complex Evolution, Multistart, Multilevel Single Linkage). Here, the default, Levenberg-Marquardt (LM)/Secant LM (SLM) was used, and proved to be adequate, according to percent error checks of hydrograph peak and volume, and additional statistical evaluation, which is described in *Results*. Moreover, Aqauveo recommends the LM/SLM method for *routine* use (Aquaveo 2014a). Parameters, including stream channel roughness, overland roughness, initial moisture, and hydraulic conductivity were defined as calibration parameters (flagged with negative values in the *mapping tables*), as recommended by Aquaveo 2014b, because of their sensitivity—small changes in these parameters can significantly affect the shape and size of the hydrograph. These parameters were allowed to vary during calibration across a specified range. Incremental 15-minute rainfall and corresponding streamflow discharge data were used as observed data. WMS automatically adjusted parameters with the LM/SLM method, until finding an optimum solution that matches a

computed hydrograph with the observed discharge data (specific values can be viewed in Appendix A).

2.1.8 Sediment Calibration

After the Halawa model was calibrated for observed streamflow discharge, it was calibrated for sediment concentration. This calibration was completed manually, by iteration. The Soil Erosion parameters were changed as needed after each iteration to force a match between computed and observed values. The model was calibrated for both sediment transport equations, KR and EH, independently. GSSHA does feature four other sediment transport methods, including Slope and Unit Discharge, but these shear force methods were omitted from this study based on poor performance in initial testing preliminary to this study.

2.2 MODEL ACCURACY EVALUATION

As recommended by the American Society of Civil Engineers (ASCE) for the evaluation of the results of single-event hydrologic modeling, *percent error* in peak streamflow, peak timing, and flow volume was used to evaluate GSSHA accuracy in prediction of streamflow (ASCE 1993; see also Green 1986). These criteria have been determined to be appropriate measures of accuracy in single-event simulation, because peak streamflow, time to peak, and flow volume are often the target outputs of single-event studies, and can be, therefore, more relevant measures of accuracy for single-events than other methods that evaluate the overall goodness-of-fit of the dataset, *e.g.*, the coefficient of determination or the Nash-Sutcliffe efficiency (Green 1986; Moriasi 2007).

For the evaluation of sediment concentration simulations, percent error in peak concentration, peak timing, were used. Percent error between observed and computed area under the concentration time-series curve, which yields non-intuitive units of mg-min/l, was not used. Instead, sediment concentrations were converted to sediment loads in tons (English or US tons, where 1 ton = 2000 lb.) for percent error comparisons, using the following equation from Woods (2001):

$$Q_s = Q * C * k$$

where, Q_s is sediment discharge in English tons/day, Q is streamflow in cfs, C is sediment concentration in mg/l, and k is a conversion factor (k=0.0027).

The coefficient of determination (R^2) and the Nash-Sutcliffe efficiency (NSE) were used to examine overall goodness-to-fit for further evaluation. These methods were used, because drawing conclusions from percent error calculations alone would have been difficult, due to a lack of sediment studies using GSSHA, and a lack of published thresholds using percent error as an indicator whether model performance is satisfactory. There are, however, studies that provide guidance on thresholds of satisfactory performance for R^2 and NSE. Thus, these methods were used to support conclusions on model accuracy. Moreover, the field of systematic quantification of watershed simulations is continually developing (*e.g.*, Moriasi 2007), so using R^2 and NSE here for single-event simulations should give researchers another data point for review, since these methods are used less frequently for single-event evaluations.

The R^2, below, describes the degree of collinearity between the observed and simulated data.

$$R^2 = \left[\frac{\sum_{i=1}^{n} (X_i - X_{avg})(Y_i - Y_{avg})}{\sqrt{\sum_{i=1}^{n} (X_i - X_{avg})^2 \sum_{i=1}^{n} (Y_i - Y_{avg})^2}} \right]^2$$

where, n is the number of observations, X_i and Y_i are the ith simulated and observed values, respectively; and X_{avg} and Y_{avg} are the averages of the simulated and observed values, respectively. R^2 ranges from 0 to 1, 1 indicating the best fit, and is used here for comparison between simulated events.

The NSE, below, describes how well the observed versus simulated data fit the 1:1 line.

$$NSE = 1 - \left[\frac{\sum_{i=1}^{n} (Y_i - X_i)^2}{\sum_{i=1}^{n} (Y_i - Y_{avg})^2} \right]$$

where, n is the number of observations, Y_i is the ith observation, X_i is the ith simulated value, and Y_{avg} is the average of the observed values. There are varying opinions on the correct use of NSE to assess model accuracy. A literature review in *Moriasi* shows that out of 92 samples, NSE values in calibration ranged from -0.23 to 0.95 with an average of 0.89 for daily streamflow prediction (Moriasi 2007). NSE values in validation of daily streamflow (128 samples) ranged from -1.81 to 0.89 with an average of 0.67. For daily sediment in calibration from only 2 samples, NSE ranged from -2.50 to 0.11 with an average of 1.20. NSE values in sediment

validation ranged (2 samples) ranged from -3.51 to 0.23 with an average of -1.64. Generally, *Moriasi* suggests that an NSE of 0.50 or greater is one indicator that the model's streamflow predictions are satisfactory, but gave no guidance for using NSE specifically for sediment (Moriasi 2007). Thus, here, NSE of 0.50 or greater was used as one indicator of satisfactory model prediction.

All percent error calculations were used synergistically with R^2 and NSE for comparisons between simulations and results from other studies, to draw conclusions on whether GSSHA's accuracy was satisfactory.

2.3 CASE STUDY DEVELOPMENT

Two case studies were developed to evaluate GSSHA and to satisfy the objectives herein. The development of the case studies is described below.

2.3.1 Watershed Selection

Upper Halawa watershed—a leeward Hawaiian watershed—was selected as the test watershed for case studies, based on (1) data availability and (2) terrain characteristics. The Halawa watershed is shown below.

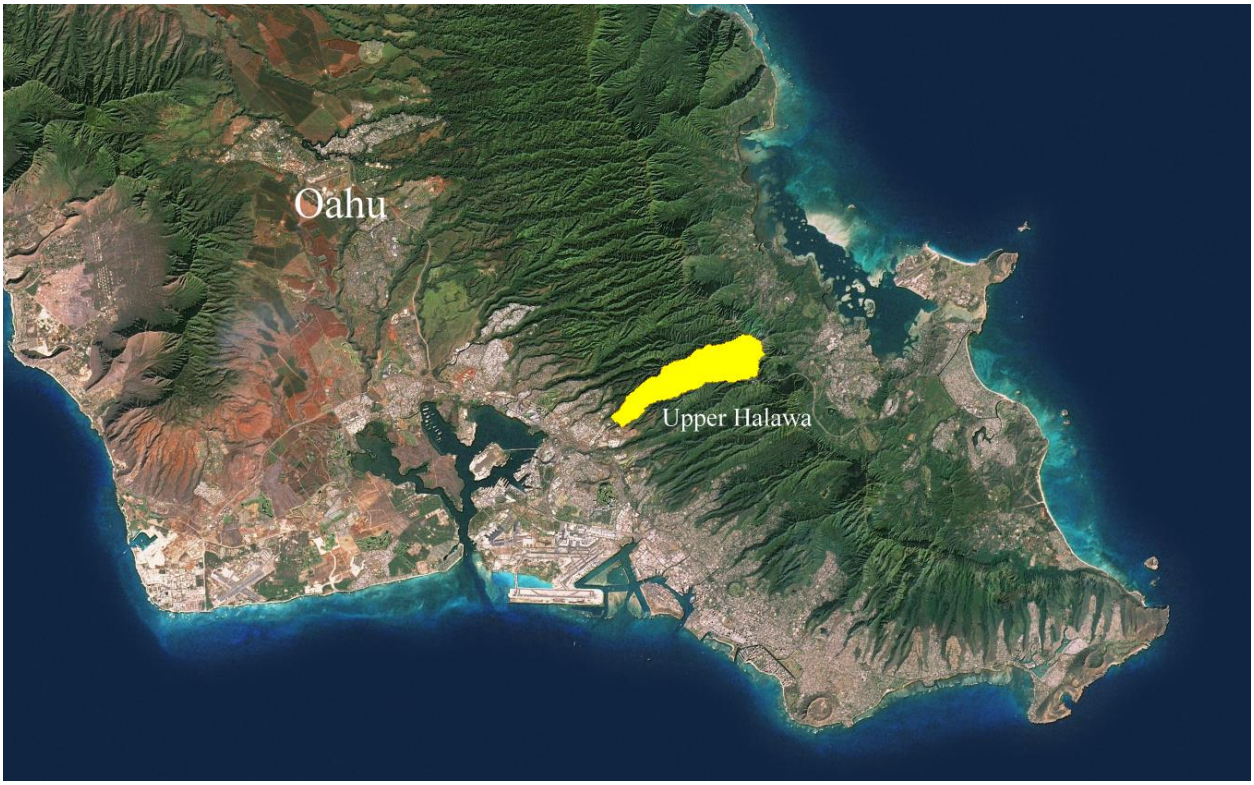

Figure 5: Location of Halawa watershed.

Streamflow and sediment discharge data were needed from a steep Hawaiian watershed, for model calibration and testing. The USGS provides these data from the monitoring of several steep Hawaiian watersheds, which could have been used for this study. The Halawa watershed, however, is unique among the monitored watersheds, because the USGS has divided Halawa into upper and lower portions for monitoring purposes. The lower potion contains some developed and flatter land. The upper portion is almost entirely steep mountainous terrain, making it an ideal laboratory for steep terrain testing. The USGS has corresponding rainfall, streamflow, and sediment data for upper Halawa from 10/01/2011 to the present day (U.S. Geological Survey 2015b).

Upper Halawa watershed has an area of approximately 4.01 square miles, a length of 6.35 miles, with an average slope along the stream of 0.079 (from outlet to upstream boundary), and an average overland basin slope of 0.65 (computed using WMS). Upper Halawa is composed of 86% evergreen forest, with the remainder primarily shrub land (11%), and less than 10% grassland and developed land, including the H3 highway (NOAA 2011 C-CAP). The soils types in upper Halawa are mainly silty clay and bedrock (Soil Survey Staff 2015). Upper Halawa gets approximately 132.2 inches (3358.9 mm) of rainfall accumulation per year (Giambelluca 2013).

2.3.2 Flood-Flow Frequency Analysis of Upper Halawa Watershed

A flood-flow frequency analysis was performed of upper Halawa watershed, and an equation was produced, to classify peak discharges from single-event storms by recurrence interval. This classification was needed to identify and distinguish between high- and low-frequency storms.

As recommended by the U.S. Water Advisory Committee on Water Data (U.S. Water Resources Council 1982), the Log-Pearson Type 3 (LP3) approach was selected and applied.

USGS peak annual discharge data from the past 31 years provided a basis for the analysis (U.S. Geological Survey 2015b). These data are shown below.

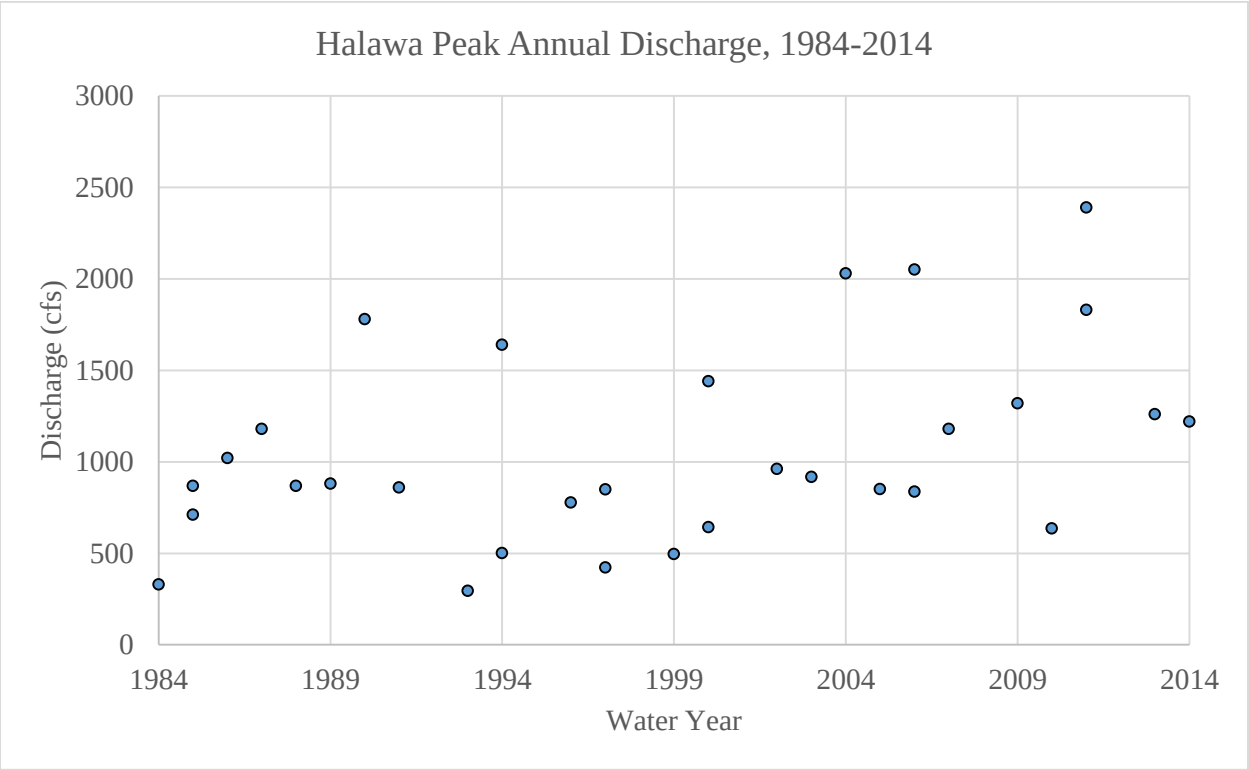

Figure 6: Halawa peak annual discharge data.

The following equation is the product of the LP3 analysis.

$$y = 464.56 * \ln(x) + 694.11$$

where, *y* is peak discharge (cfs) and *x* is the recurrence interval (years). This equation was used to classify the single-event storms in this study by recurrence interval.

The flood-flow frequency curve for 2- to 100-year storms, and the coefficient of determination for the equation (R^2), are displayed in the figure below.

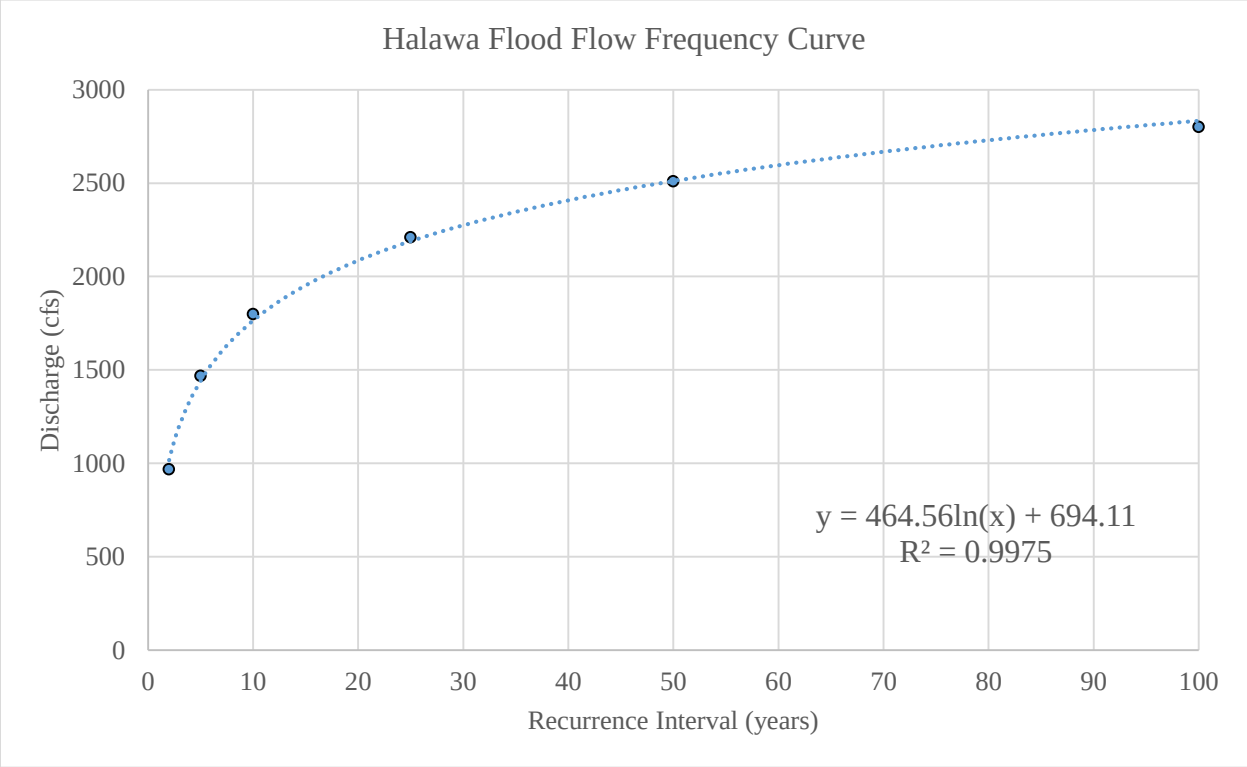

Figure 7: Halawa flood flow frequency curve.

The full LP3 analysis procedure can be found in the Appendix.

2.3.3 Storm Event Selection

The storms used for calibration and testing were selected based on (1) the time period of USGS streamflow, rainfall, and sediment data availability (10/01/2011-Present), (2) the time period in which USGS has approved data for publication (10/01/2011-09/14/2014, at the time of this report), (3) the time period in which USGS has consistent 15-minute rainfall data (01/01/2013-09/14/2014), and (4) the magnitude of streamflow with respect to recurrence interval. Following this selection criteria, the storms listed in the following table were identified (U.S. Geological Survey 2015b).

Table 2: Storm data for case studies.

Storm	Date	Recurrence Interval (years)	Peak Streamflow (cfs)	Total Rainfall Accumulation (mm)	Peak 15-Min Rainfall (mm)
Storm 1	01/27/2014-01/29/2014	<2-year	234	111.0	8.4
Storm 2	01/02/2013-01/04/2013	<2-year	120	141.5	4.1
Storm 3	04/06/2014-04/08/2014	<2-year	245	107.7	6.1
Storm 4	05/10/2013-05/12/2013	<2-year	215	202.2	6.9
Storm 5	05/27/2013-05/29/2013	~3-year	1260	529.1	26.4
Storm 6	07/19/2014-07/21/2014	~5 year	1400	240.3	18.3

Here, a *high-frequency storm* is defined as a storm of 2-year recurrence interval or below, and a *low-frequency storm* is a storm with a recurrence interval greater than 2 years. For example, based on these definitions, storms 1-4 are classified as high-frequency and storms 5 and 6 are classified as low-frequency.

A set duration of 72 hours was chosen for all events for simplicity in modeling, and because a three-day duration was determined adequate to capture the entire storm events, based on the streamflow data going from near zero, to peak, and back to zero within that 72-hour time frame. If a 24- or 48-hour interval was used, some rainfall data connected to the corresponding single-event floods, would have been omitted.

2.3.4 Storm Data

Rain data was obtained from the USGS (U.S. Geological Survey 2015b). The following six figures show 15-minute incremental rainfall and total accumulation for storms 1-6.

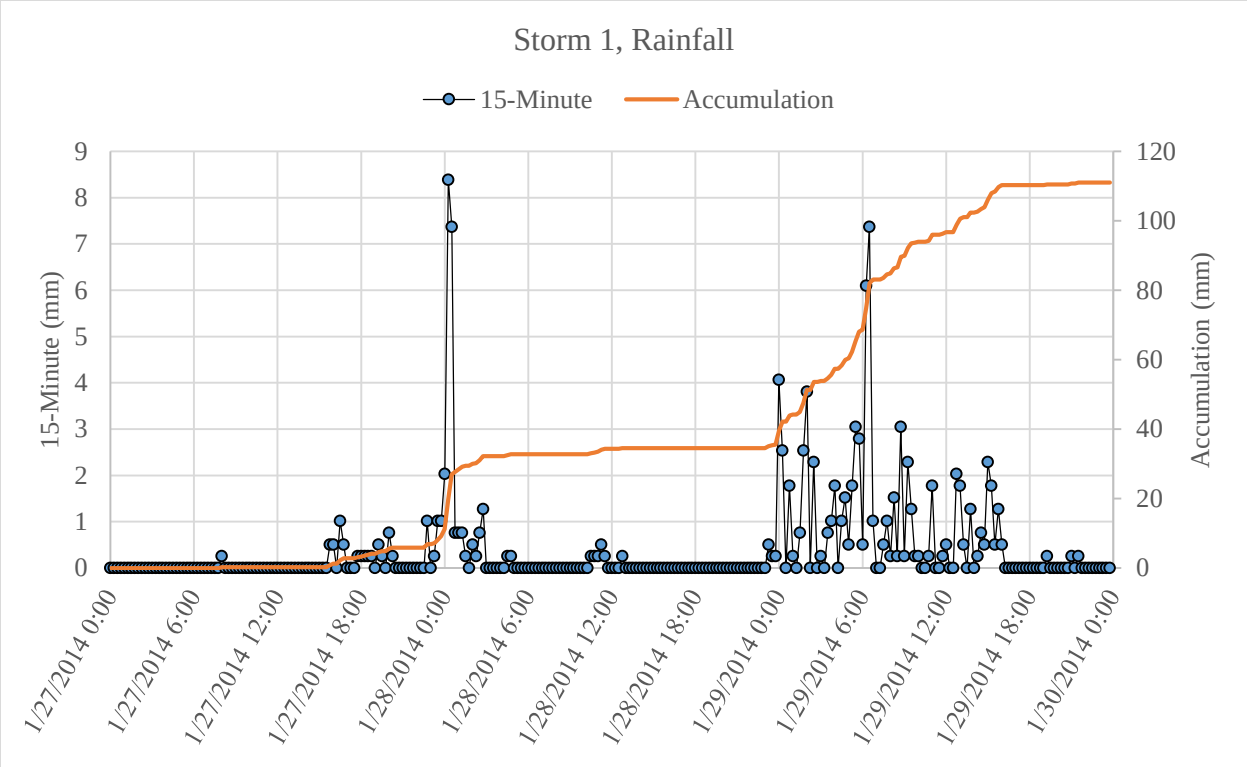

Figure 8: Storm 1 rainfall.

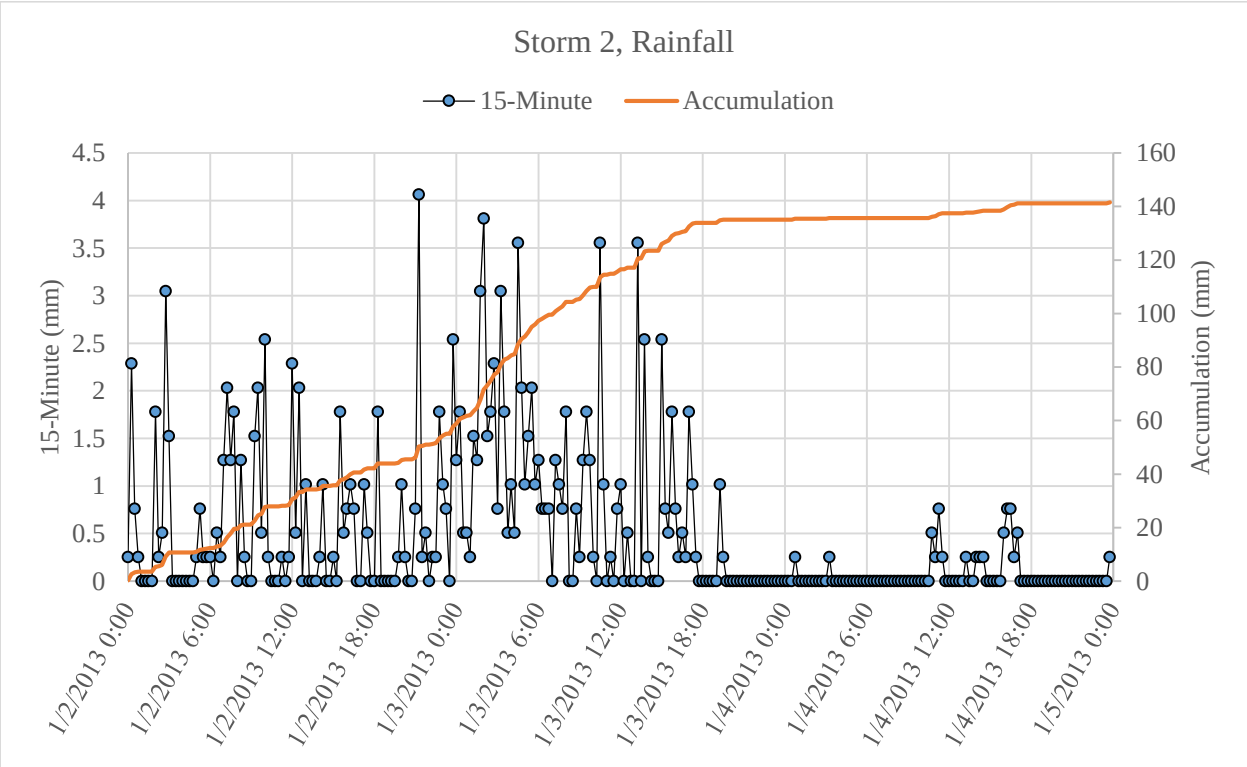

Figure 9: Storm 2 rainfall.

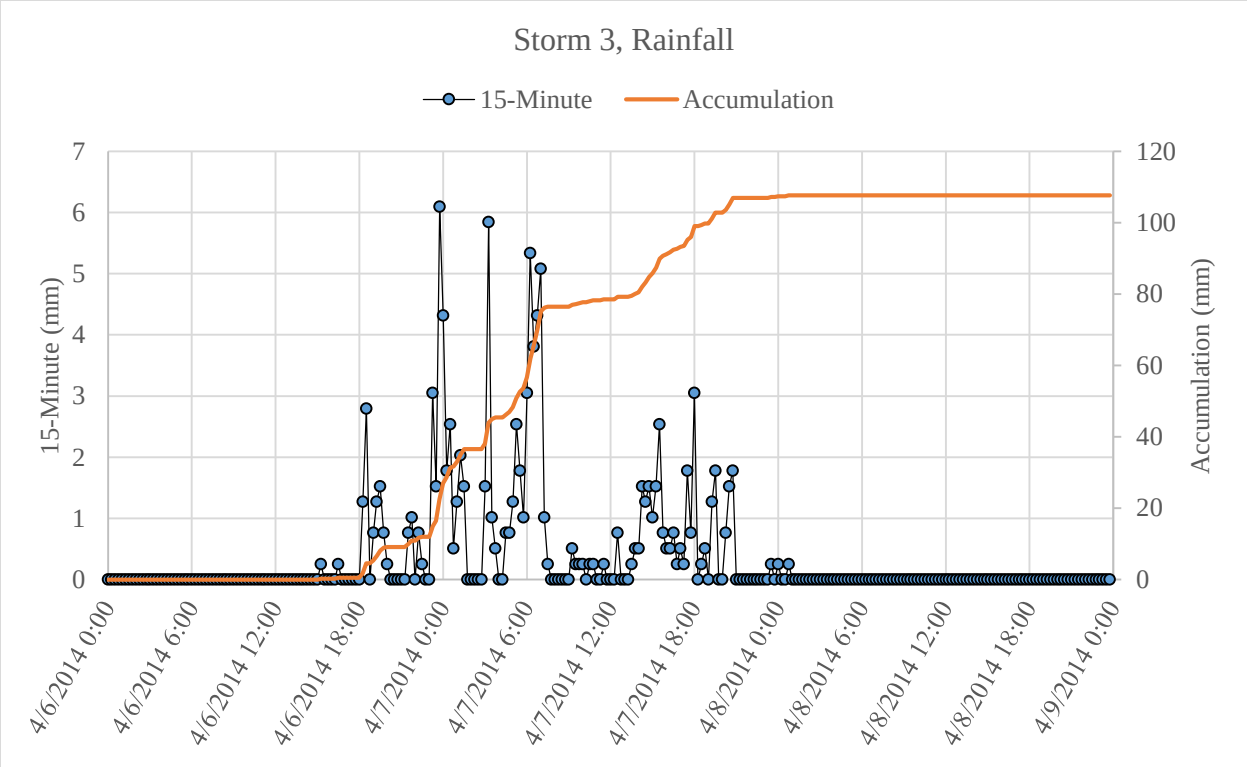

Figure 10: Storm 3 rainfall.

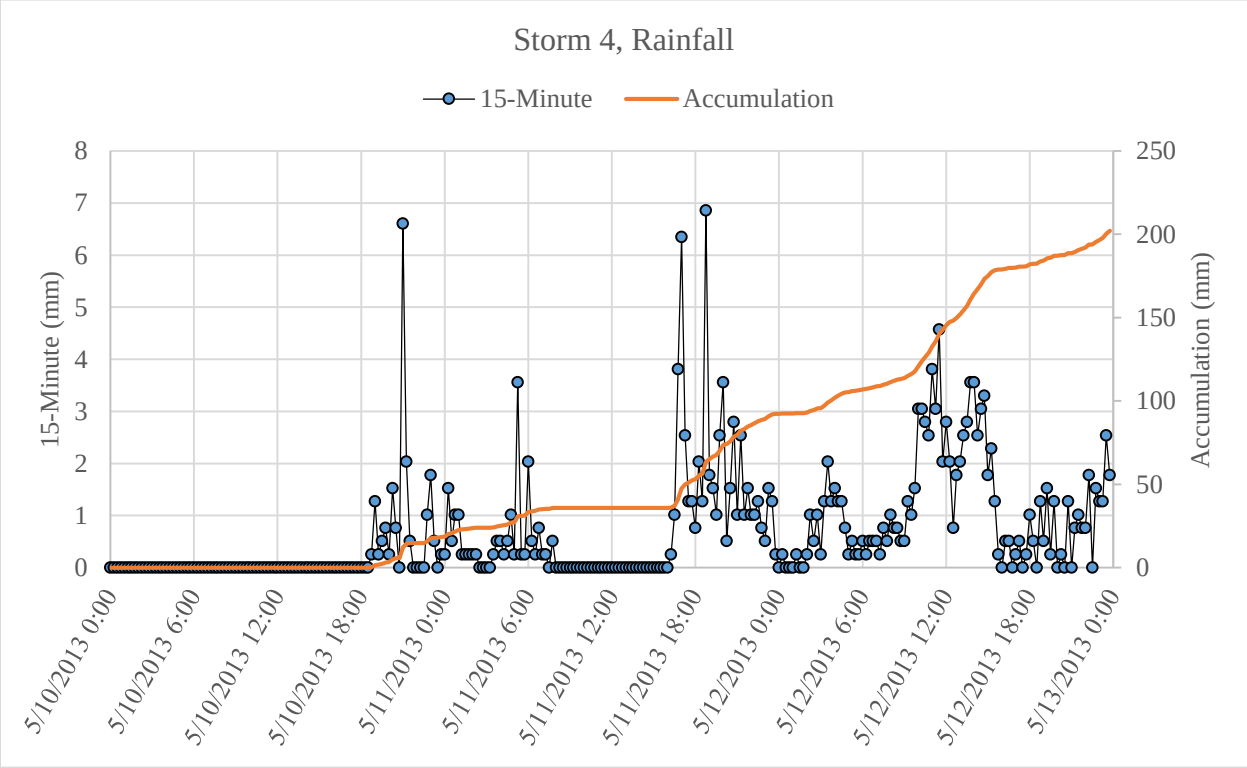

Figure 11: Storm 4 rainfall.

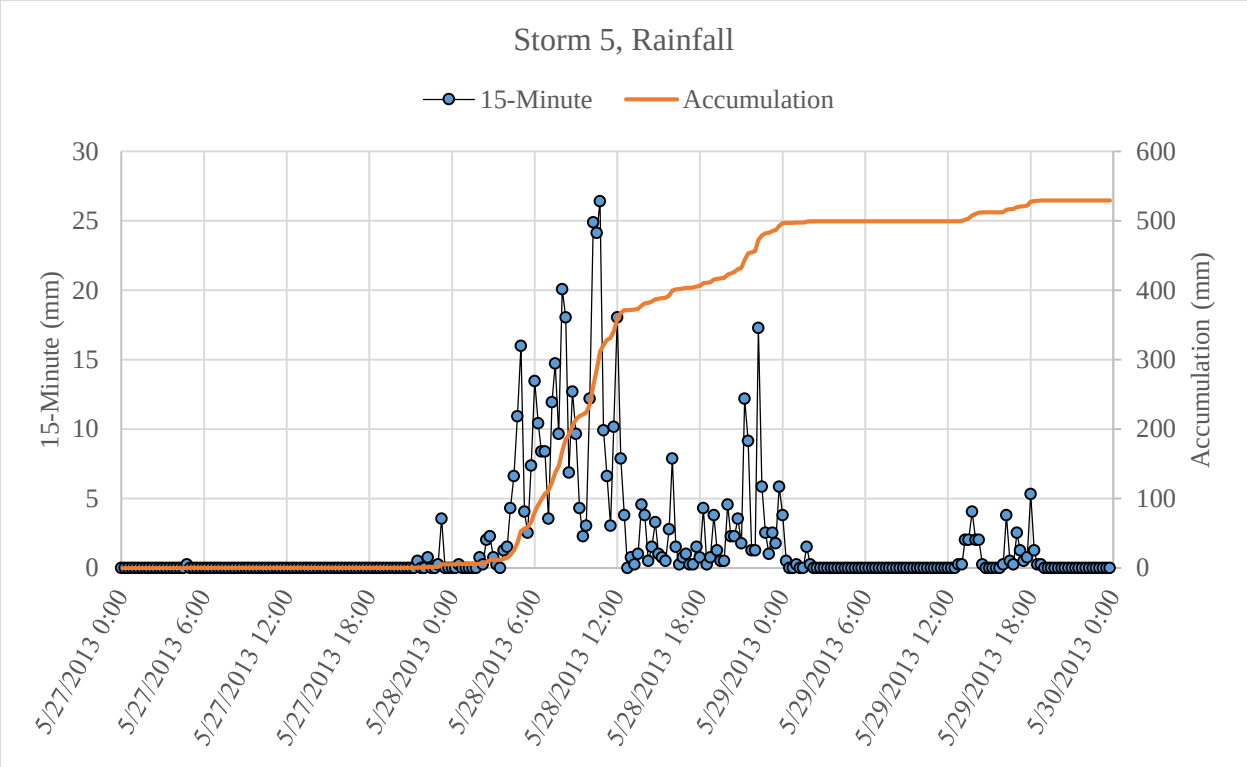

Figure 12: Storm 5 rainfall.

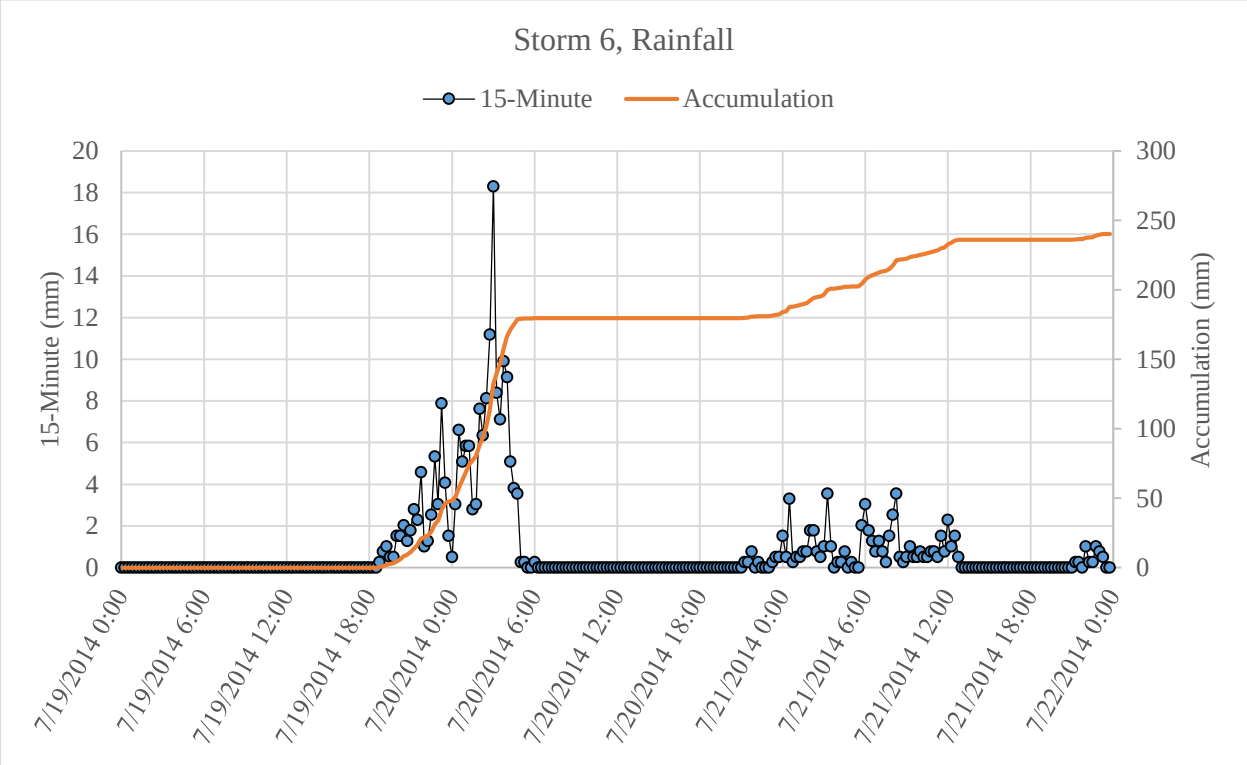

Figure 13: Storm 6 rainfall.

2.3.5 Description of Case Studies

Case 1: High-frequency storm calibration followed by simulation of several other high-frequency storms.

Case 1 entails using storm 1 for calibration and storms 2-4 for testing. The motivation for Case 1 is to establish a baseline for GSSHA accuracy in streamflow and sediment concentration prediction for single-event storms in steep Hawaiian watersheds. GSSHA will be calibrated with a high-frequency storm for Halawa (<2-year recurrence interval), with respect to streamflow and sediment transport with both Engelund-Hansen and Kilinc Richardson, and then tested with several storms of similar recurrence interval, but a range of rainfall characteristics. For example, in relation to Storm 1, which has a high peak 15-min rainfall and a low accumulation, Storm 2 has a higher accumulation but lower peak; Storm 3 has approximately the same accumulation but a lower peak; and Storm 4 has a much higher accumulation but a lower peak. The baseline established from the simulations of Case 1 will be used as a reference for Case 2, which is described below.

Case 2: High-frequency storm calibration followed by simulation of lower-frequency storms.

Case 2 entails using the calibration from Case 1, a high-frequency storm, to simulate low-frequency storms 5 and 6, which are the two largest storms from water years 2013 and 2014, respectively. The motivation for Case 2 is to determine whether GSSHA over predicts sediment concentrations for lower-frequency storms when calibrated with higher-frequency storms (Objective 1a), as observed by Ogden (2001) and as stated in the GSSHA manual (Downer

2006), although the opposite was observed in Downer (2010). Case 2 should allow verification and quantification of any over prediction and lead to guidance on the use of GSSHA in such circumstances.

The results from Case 2 will be used together with Case 1 to determine (1) whether the Engelund-Hansen or the Kilinc-Richardson equation is better suited for predicting sediment concentration in steep Hawaiian watersheds (Objective 1b), (2) to determine whether GSSHA can be relied upon for single-event watershed sediment studies in Hawaii (Objective 2), (3) to identify potential improvements for GSSHA (Objective 3), and, thus, to solve all objectives simultaneously.

Case 3: Run any additional simulations determined necessary to assess the accuracy of GSSHA in prediction sediment concentration for single-events.

The purpose of Case 3 was to run additional simulations that may be determined necessary based on the findings of cases 1 and 2. Based on findings in Case 2, it was determined that additional simulations were needed to assess the reasons that GSSHA may over predict sediment.

CHAPTER 3

RESULTS AND DISCUSSION

3.1 CASE 1

Case 1: High-frequency storm calibration followed by simulation of several other high-frequency storms

3.1.1 Storm 1 Streamflow Calibration

Storm 1 was the high-frequency storm used to calibrate GSSHA for streamflow (automated calibration) and sediment discharge (manual calibration). The results for streamflow calibration are displayed below.

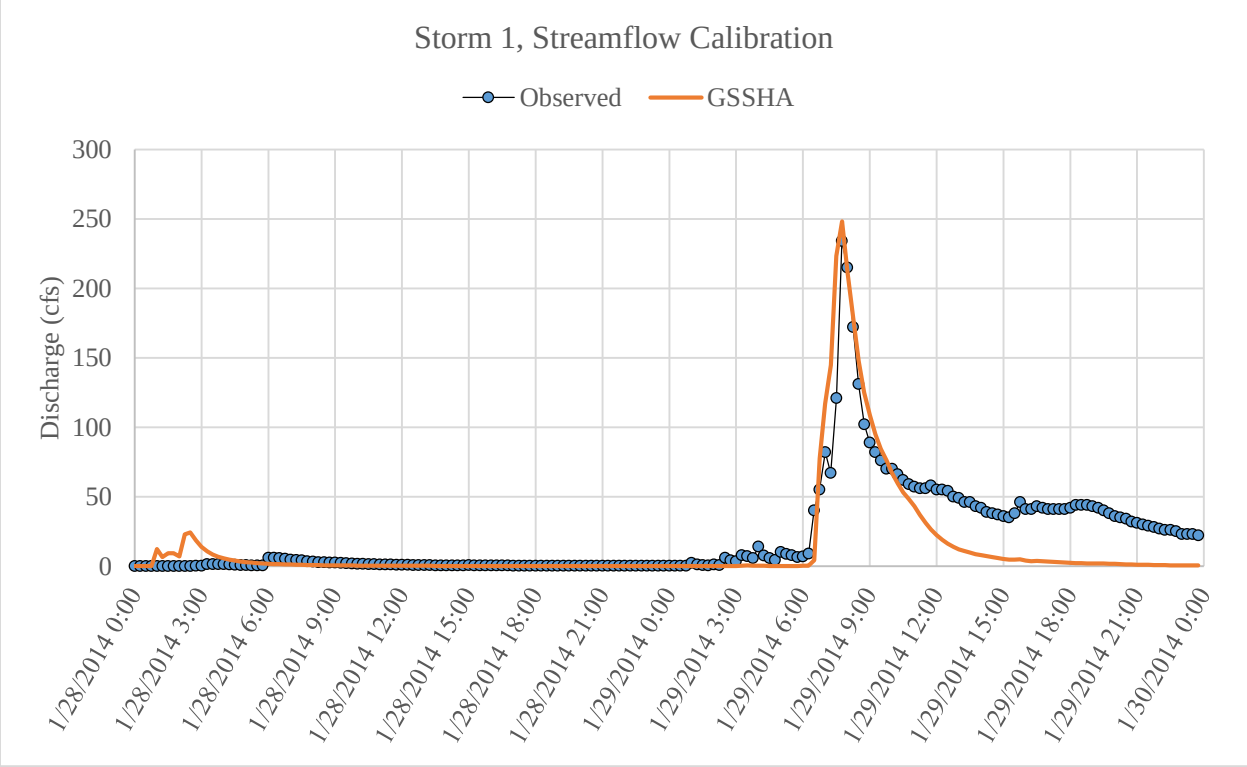

Figure 14: Storm 1 streamflow calibration.

The evaluation criteria were applied to the observed and computed hydrographs. The difference in peak is 14 cfs (234 cfs observed, 248 cfs computed), resulting in a 5.9% error. The difference in time to peak is 0 minutes (at 3345 minutes) for a 0.0% error. The difference in volume is 1286636 ft^3 (3650688 ft^3 observed, 2364052 ft^3 computed) for a -35.2% error, which is similar to the percent error in volume from other GSSHA calibration studies (Shurtz 2009; Downer 2004, 53% average percent error in flow volume during calibration).

The majority of the error in flow volume comes from the difference between the falling limbs of the hydrographs. This was also the case in Shurtz (2009), and could be explained by a lack of watershed storage. Here, for example, the error in the falling limb could be a result of small depressions in topography, which exist in the watershed, but are smoothed and cannot be represented accurately by a 10-meter-resolution grid. Such depressions would likely impound small quantities of runoff and generate more flow during the falling limb. To support this idea, changing the overland roughness to likely unrealistic values (near n=1.0) to slow runoff, in an attempt to compensate for lost watershed storage, has the effect of increasing volume in the falling limb (this also, however, causes error to increase in the rising limb and peak).

Additionally, between the two hydrograph curves, R^2=0.75, which is used as a reference point for the R^2 values in the tests that follow, and NSE=0.68, showing a good fit (0.50 or greater is considered satisfactory).

To match the observed values, and capture the dynamic response of the watershed, automated calibration selected very low initial soil moisture values (e.g., 86% of the watershed has initial moisture of 5%). These values are likely low if compared to actual values, considering this storm occurred during the wet season. Attempts were made to manually increase soil moistures, and increase hydraulic conductivity (because these operations have opposite effects

on runoff production), to compensate and hold the computed flow peak and volume near observed. These attempts, however, led to an increase in the false peak predicted by GSSHA early on in the event (notice the false peak occurring at approximately 2:00 on 1/28/2014 in Figure X above). A reason why one may want to have higher soil moistures in this case, is that it could create more soil moisture adjustability when simulating storms from the wet versus the dry season.

In summary, because the observed and computed hydrographs have a close visual fit, the percent error calculations are comparable to other studies (*e.g.*, Downer 2004), and NSE is above 0.5, this calibration was considered satisfactory.

3.1.2 Storm 1 Sediment Calibration

The following figure shows the sediment concentration calibration results.

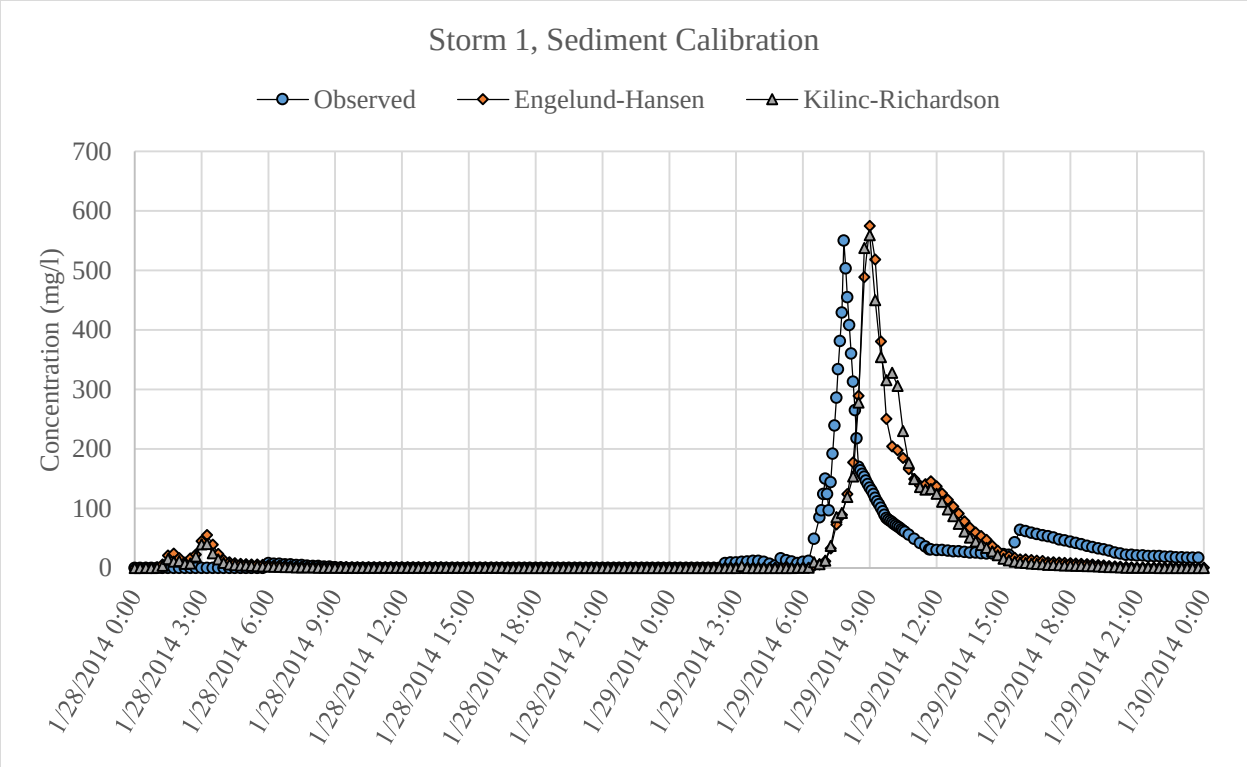

Figure 15: Storm 1 sediment calibration.

The evaluation criteria were applied to the observed and computed sediment concentration curves. The difference in peak between the observed and Engelund-Hansen (EH) is 25 mg/l (550 mg/l observed, 575 mg/l computed), resulting in a 4.5% error. The difference in peak between the observed and Kilinc-Richardson (KR) is 9 mg/l (550 mg/l observed, 559 mg/l computed), resulting in a 1.6% error. The difference in time to peak is 75 minutes (3345 minutes observed, 3420 minutes computed) for a 2.2% error.

Instead of comparing area under the curve of the sediment concentration time-series, which yields non-intuitive units of mg-min/l, total sediment loads were calculated in tons (English or US tons, where 1 ton = 2000 lb.). The difference in sediment load in tons between observed and EH is 0.5 tons (13.2 tons observed, 12.7 tons computed) for a -3.7% error. The

difference in sediment load in tons between observed and KR is 0.2 tons (13.2 tons observed, 13.0 tons computed) for a -1.5% error.

This calibration was done mostly by adjusting the Erodibility Coefficient. Here, as in McCarthy (2012), the dimensionless Erodibility Coefficient, which is a product of three empirical factors, was found to be, by far, the most sensitive parameter. Sensitivity testing showed that it is so sensitive that changes—even large changes—to the other soil erosion parameters (detachment coefficients and rill values) had little effect of sediment concentrations, once the model was calibrated with the necessary and very small Erodibility Coefficient value. The Erodibility Coefficient value used with KR for the largest percentage of watershed area is 0.000002, which does not leave much room for additional adjustability on the low-end of WMS (lowest value accepted is 0.000001).

The additional evaluation criteria, R^2 and NSE, if viewed alone, would indicate that this calibration is unsatisfactory (using NSE=0.5 or greater indicates satisfactory performance). Here, R^2=0.26 and NSE=-0.52 for EH, and R^2=0.25 and NSE=-0.61 for KR. This outcome for R^2 and NSE, however, was expected, because of the trailing peaks of the computed sediment concentration curves. The computed sediment concentration peaks (for both KR and EH) trail the computed hydrograph peak by 75 minutes, whereas the observed peaks occur within the same 15 minute interval. Since, R^2 and NSE evaluate the goodness of fit between the curves, the trailing peak in sediment concentration curve has an adverse effect on these evaluation methods for sediment concentration.

Following *ASCE*, which recommends synchronizing the time to peak for comparisons, time to peak between the computed hydrograph and sediment concentration curves were synchronized (ASCE 1993; see also Glysson 1987 and Green 1986). This process of eliminating

the error in timing does not change anything for the percent error calculations of concentration peak or sediment load comparisons, and it does allow for a better comparison of goodness of fit with R^2 and NSE. The figure below shows the time-adjusted curves for sediment concentration of Storm 1.

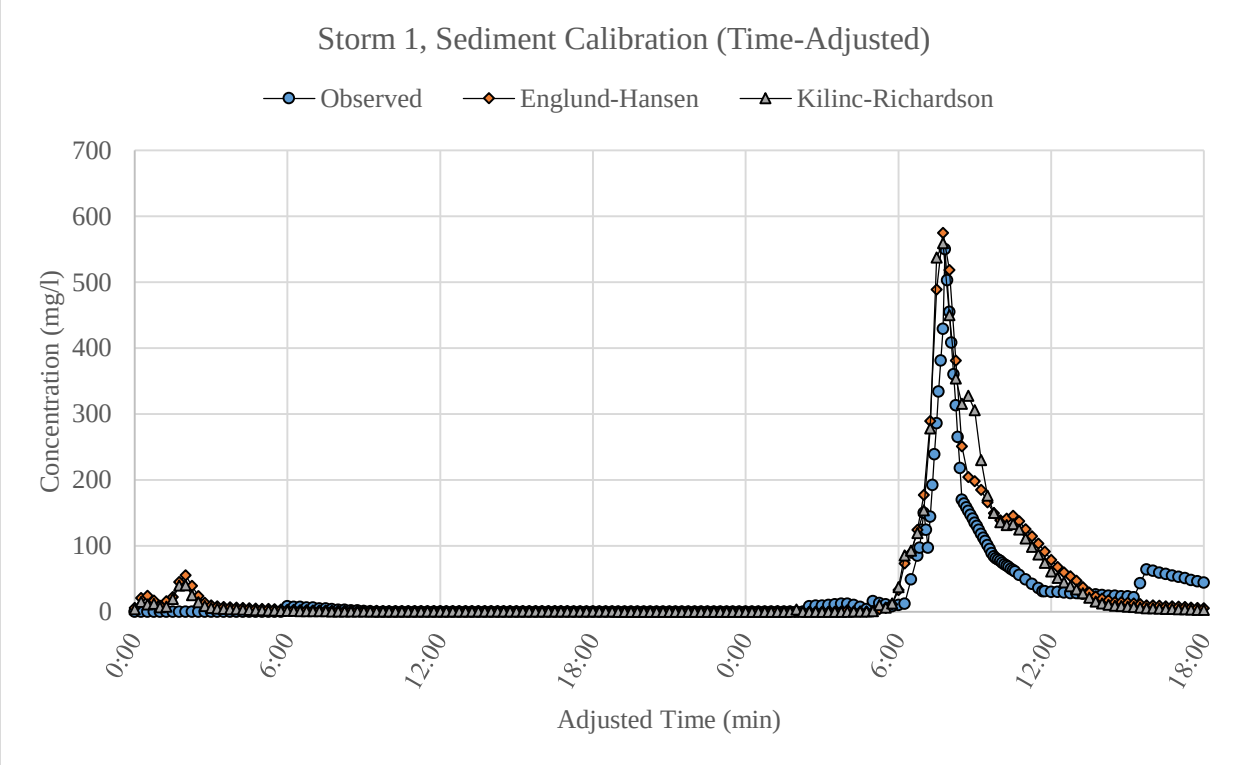

Figure 16: Storm 1 time-adjusted sediment calibration.

The additional evaluation criteria were applied to the time-adjusted curves for comparison, resulting in R2=0.85 and NSE=0.82 for EH, and R2=0.86 and NSE=0.83 for KR, which shows similar goodness of fit between EH and KR in calibration. These R2 and NSE values, viewed in context with the percent error calculations, which show relatively low errors compared to other studies (Shurtz 2009; Downer 2004), indicate a satisfactory sediment calibration.

It is shown below that GSSHA consistently predicts trailing peaks for sediment concentration for the other storms. These trailing peaks are at odds with observations. Observations show that the peaks of streamflow and sediment concentration occur within the same 15 minute interval. Attempts were made during calibration to correct this issue by changing several parameters, including specific gravity of sediment particles, and particle size fractions, but no satisfactory results were achieved. Changing these quantities had large effects of concentration magnitude, but small effect on the time to peak, within a realistic range of values. The large changes in magnitude interfered with calibration and negated any benefit from a shift in peak. Thus, for model performance evaluation, the trailing peaks were handled, as above, by adjusting the time to peak. Results, however, are given below for both before and after time adjustments were made.

3.1.3 Storm 2 Streamflow Validation

Storm 2 is the first high-frequency storm (<2-year recurrence interval) used to test the Storm 1 calibration of GSSHA. Although similar to Storm 1 in frequency, Storm 2 has a lower 15-minute peak rainfall value, by 51%, a 27% higher rainfall accumulation, and a longer sustained rainfall duration (within the overall 72-hour duration simulations).

The streamflow validation hydrograph is shown below.

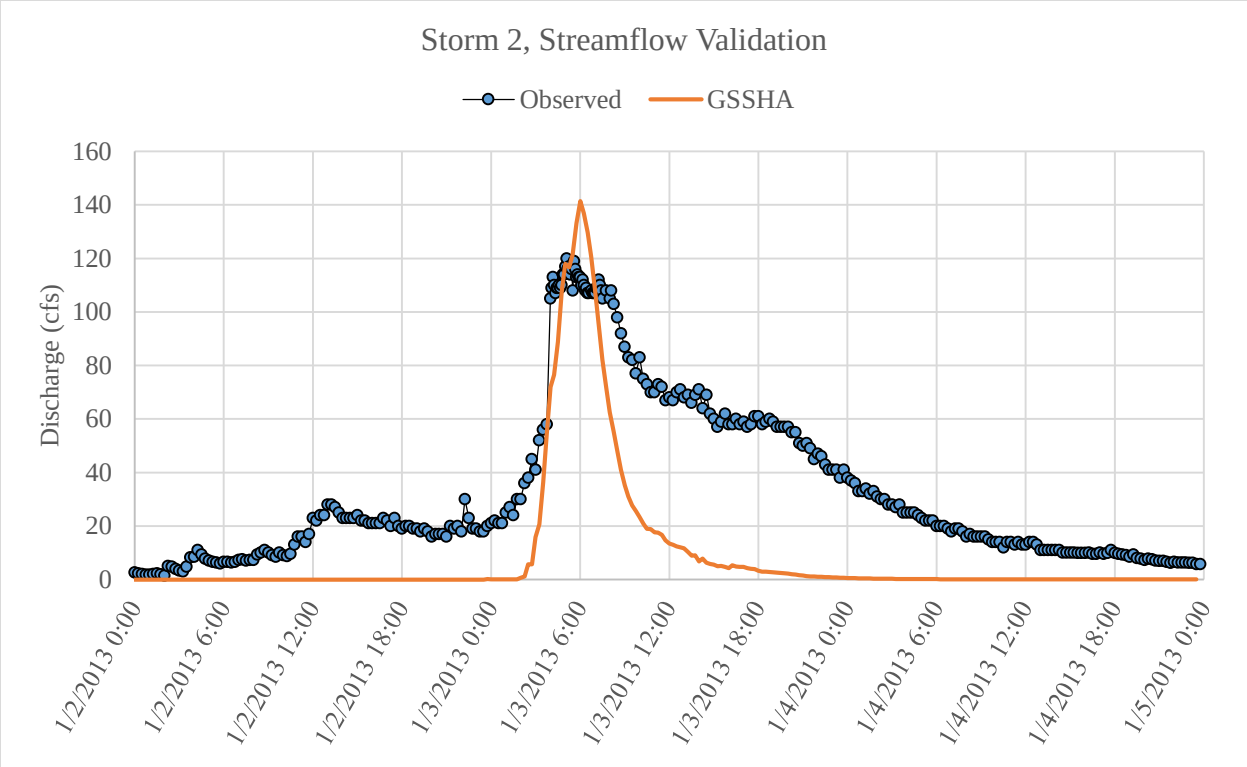

Figure 17: Storm 2 streamflow validation.

The evaluation criteria were applied to the observed and computed hydrographs. The difference in peak is 21 cfs (120 cfs observed, 141 cfs computed), resulting in a 17.5% error. The difference in time to peak is 55 minutes (observed at 1745 minutes, computed at 1800 minutes) for a 3.2% error. The observed curve of Storm 2, however, has a longer duration peak, and it appears that the peak of the computed curve occurs approximately in the center of the observed peak, showing a good fit. The difference in volume is 6008363 ft^3 (8313480 ft^3 observed, 2305117 ft^3 computed) for a -72.3% error, which is similar to *Shurtz* (67% error for single-event simulation), but is high compared *Downer* (17% on average), although *Downer* was simulating continuous events (Shurtz 2009; Downer 2004).

Additionally, between the two hydrograph curves, R^2=0.61, which could support a conclusion that GSSHA has predicted a relatively good hydrograph shape, peak, and timing, but

NSE=-0.03, which shows unsatisfactory overall goodness-of-fit accuracy, being below NSE=0.5 threshold.

GSSHA over predicted the peak and under predicted the volume for the streamflow of Storm 2. These errors may indicate that the set of calibration parameters may not be suitable for different types of storms. For example, the volume error in the rising limb may indicate that initial moisture levels need to be increased, which would cause more initial runoff. The peak error may be remedied by higher hydraulic conductivities. The volume error in the falling limb, similar to calibration, could be a result of a loss of watershed storage in grid resolution, or, again, that soil moisture levels need to be increased for this storm.

The theory that, for single-event modeling, GSSHA may not be able to use one set of calibration parameters to accurately simulate storms with different rainfall characteristics, is similar to the finding in Cunderlik (2009) that one calibration of HEC-HMS cannot be used for storms with different rainfall characteristics. In *Cunderlik* storms with different rainfall characteristics (autumn storms with long duration lower intensity rainfall versus summer storms featuring the converse), needed different calibrations for accurate predictions using HEC-HMS. Here, although Storm 1 and Storm 2 are from the same season, both occurring in the month of January, they are different in the same ways that the two storms were different in Cunderlik—Storm 2 being longer duration and lower intensity, and Storm 1 being the converse—and it appears that GSSHA may need a different set of calibration parameters to accurately simulate a hydrograph for Storm 2.

Shurtz (2009) also points to differences in storm characteristics to explain hydrograph prediction errors from GSSHA and HEC-HMS.

3.1.4 Storm 2 Sediment Validation

The following figure shows the sediment concentration validation results.

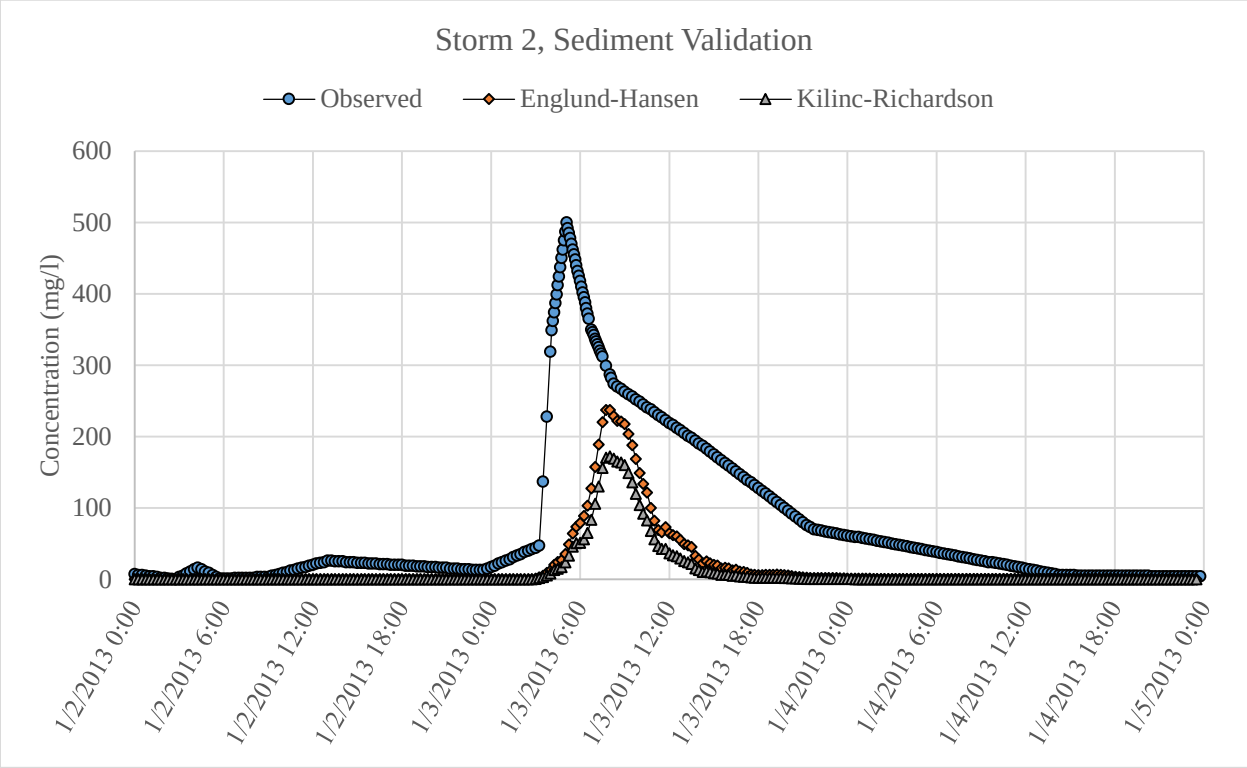

Figure 18: Storm 2 sediment validation.

The evaluation criteria were applied to the observed and computed sediment concentration curves from the Storm 2 simulations. The difference in peak between the observed and EH is 263 mg/l (500 mg/l observed, 237 mg/l computed), resulting in an error of -52.6%. The difference in peak between the observed and KR is 328 mg/l (500 mg/l observed, 172 mg/l computed), resulting in a -65.6% error. The difference in time to peak is 180 minutes (1740 minutes observed, 1920 minutes computed) for a 10.3% error.

Sediment concentration was converted into tons. The difference in sediment load in tons between observed and EH is 35.7 tons (42.7 tons observed, 7.0 tons computed) for a -83.6%

error. The difference in sediment load in tons between observed and KR is 37.9 tons (42.7 tons observed, 4.8 tons computed) for a -88.8% error. These errors are approximately double the error found in Downer (2010) during validation (-47%), but not far outside the range of error in calibration observed in Downer (2010), -73% to 220% (for calibration), and well within the maximum absolute error of that range.

The R^2 and NSE, between the observed and computed curves is R^2=0.49 and NSE=0.14 for EH and R^2=0.45 and NSE=-0.02 for KR. Based on these values and the results of the percent error calculations, both over predicted, goodness of fit was poor, but EH outperformed the KR in the test of Storm 2.

Again, peak time was adjusted so that the observed and computed peaks match, to compare the goodness of fit, independent of timing errors. The figure below displays the time-adjusted sediment concentration curves.

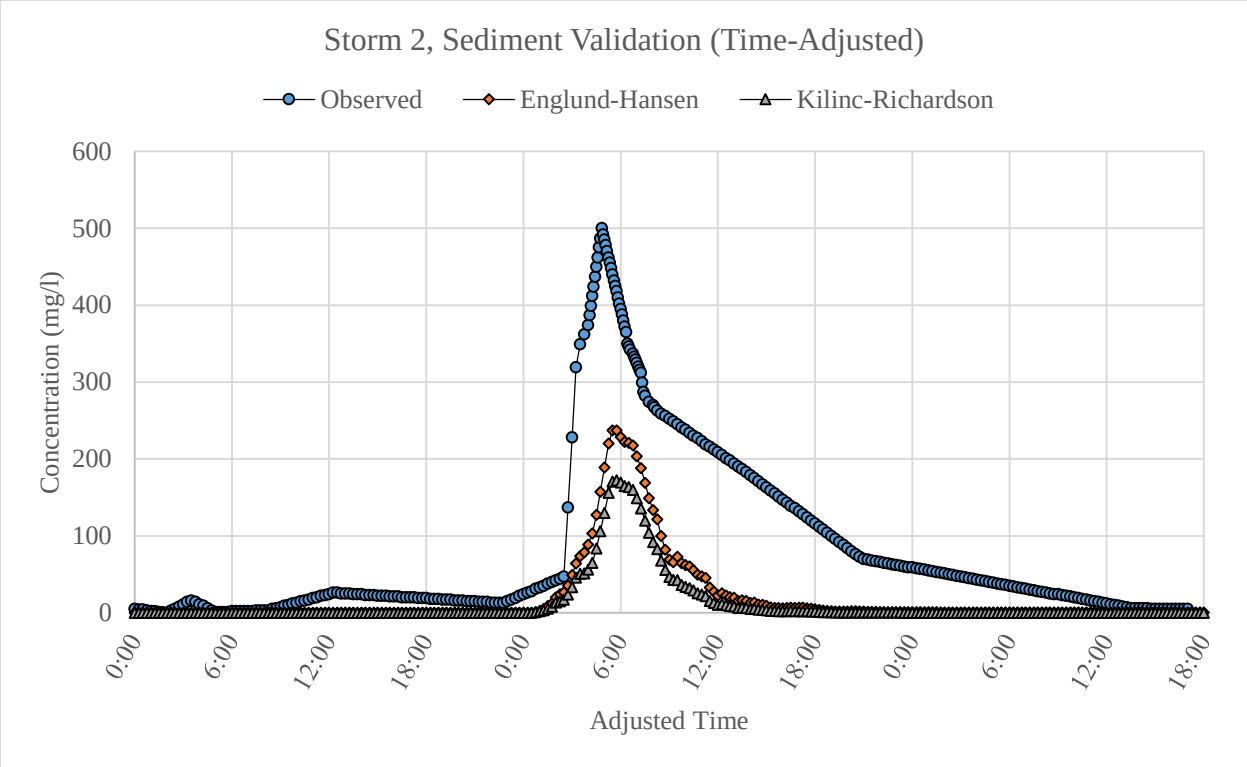

Figure 19: Storm 2 time-adjusted sediment validation.

The evaluation criteria were applied to the time-adjusted curves, resulting in R2=0.61 and NSE=-0.06 for EH, and R2=0.55 and NSE=-0.23 for KR.

3.1.5 Storm 3 Streamflow Validation

Storm 3 is the second high-frequency storm (<2-year recurrence interval) used to test the Storm 1 calibration of GSSHA. Although similar to Storm 1 in frequency, Storm 3 has a lower 15-minute peak rainfall value, by 27.4%, but nearly the same rainfall accumulation—a difference of -2.9%. Storm 3 is also different from Storm 1 in rainfall distribution, but not in the same way as Storm 2 is different than Storm 1. The rainfall of Storm 3 occurs in one pulse or burst, whereas the rainfall in Storm 1 occurs in two separate bursts, but both have approximately the same rainfall duration (see *Storm Data* in Chapter 2 herein for plots of rainfall (Figures 8-13)).

The streamflow validation hydrograph is shown below.

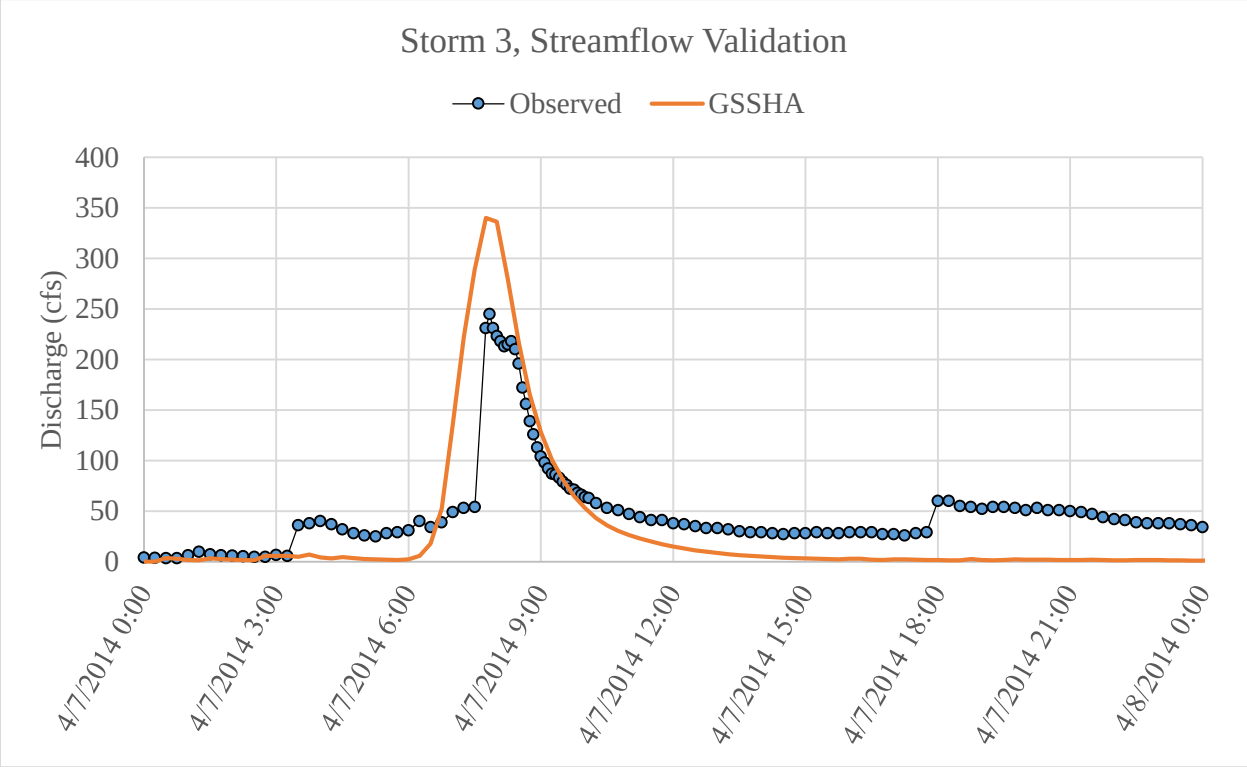

Figure 20: Storm 3 streamflow validation.

The evaluation criteria were applied to the observed and computed hydrographs from the test of Storm 3. The difference in peak is 95 cfs (245 cfs observed, 340 cfs computed), resulting in a 38.7% error. The difference in time to peak is 0 minutes (observed at 1905 minutes, computed at 1905 minutes) for a 0.0% error. The difference in volume is 2515782 ft^3 (5162055 ft^3 observed, 2646273 ft^3 computed) for a -48.7% error, which is less than *Shurtz* (67% error for single-event simulation), but is high compared *Downer* (17% on average), although *Downer* was simulating continuous events (Shurtz 2009; Downer 2004).

GSSHA showed better accuracy in predicting streamflow for Storm 3, compared to the first test of Storm 2. Again, however, a large part of the difference in flow volume occurs

between the falling limbs of the hydrographs. This could be viewed as further support for the idea that a 10-meter-resolution grid—while small compared to other models, and the same size as most of the best GIS data files for watershed modeling—may be too large to capture important topographic details of steep Hawaiian watersheds that translate into watershed storage.

Also, Storm 3's rainfall characteristics—accumulation, peak intensity, and duration—were a closer match to Storm 1's than Storm 2's. This could be viewed as further support for the theory that, like findings for HEC-HMS by Cunderlik (2009), GSSHA cannot accurately predict hydrographs from storms with different rainfall characteristics, using one set of calibrated parameters. For optimum accuracy in single-event simulations, simulated storms may need to have similar rainfall distribution characteristics as storms used for calibration.

Additionally, between the two hydrograph curves, R^2=0.65, which could support a conclusion that GSSHA has, again, predicted a relatively good hydrograph shape, peak, and timing. However, an NSE=0.16, although a better fit than the test of storm 2, shows unsatisfactory overall accuracy for streamflow prediction, using the threshold of NSE=0.5.

3.1.6 Storm 3 Sediment Validation

The following figure shows the sediment concentration validation results. For better time series shape comparison, between observed and computed, the computed values were assigned to the secondary axis.

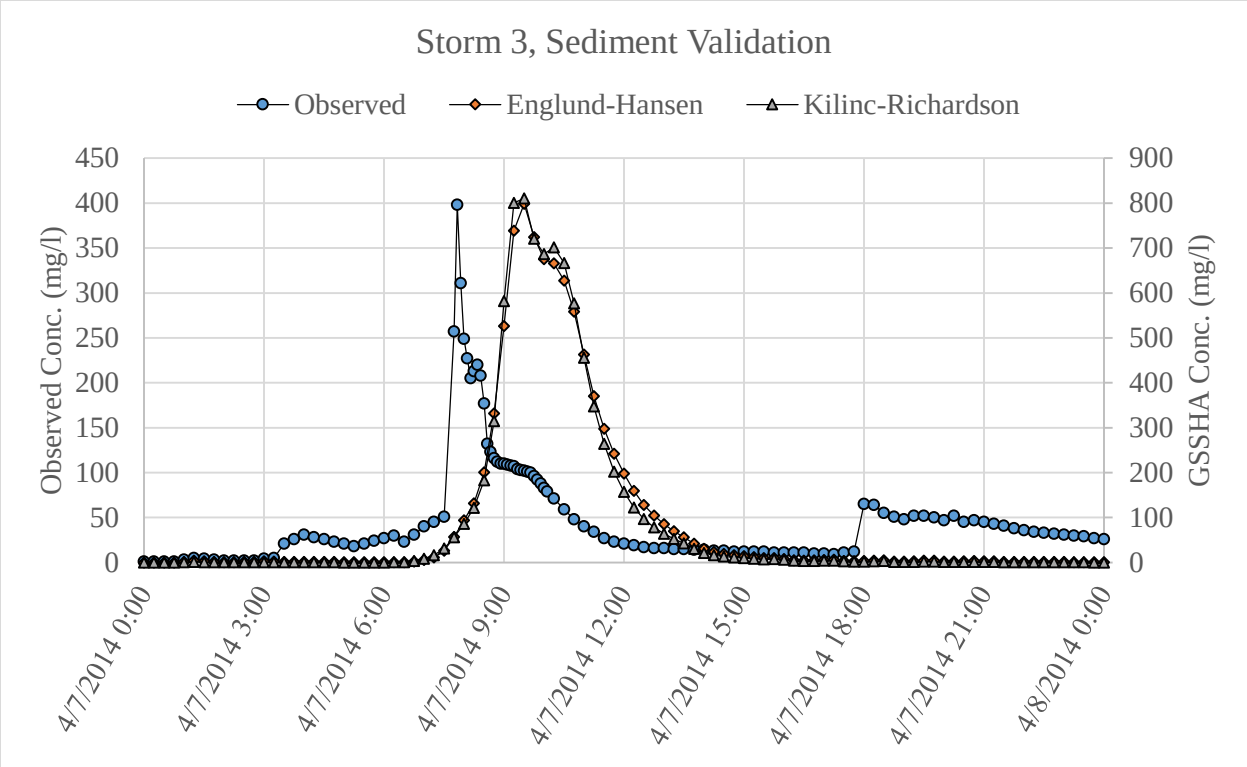

Figure 21: Storm 3 sediment validation.

The evaluation criteria were applied to the observed and computed sediment concentration curves from the test of Storm 3. The difference in peak between the observed and EH is 400 mg/l (398 mg/l observed, 798 mg/l computed), resulting in an error of 100.5%. The difference in peak between the observed and KR is 412 mg/l (398 mg/l observed, 810 mg/l computed), resulting in a 103.5% error. The difference in time to peak is 90 minutes (1920 minutes observed, 2010 minutes computed) for a 4.7% error.

Sediment concentrations were converted to tons. The difference in sediment load in tons between observed and EH is 5.6 tons (11.1 tons observed, 16.7 tons computed) for a 50.5% error. The difference in sediment load in tons between observed and KR is 5.8 tons (11.1 tons observed, 16.9 tons computed) for a 52.3% error. These errors are nearly equal in magnitude to the validation error observed in Downer (2010), -47%, which was considered satisfactory.

The R^2 and NSE, between the observed and computed curves is R^2=0.23 and NSE=-10.73 for EH and R^2=0.23 and NSE=-11.33 for KR. Based on these values the EH narrowly outperformed the KR in the test of Storm 3. Also, although both over predicted concentration peak by nearly 100%, both satisfactorily predicted sediment load, when compared to Downer (2010).

Peak time was adjusted so that the observed and computed peaks match, to compare the goodness of fit, independent of timing errors. The figure below displays the time-adjusted sediment concentration curves.

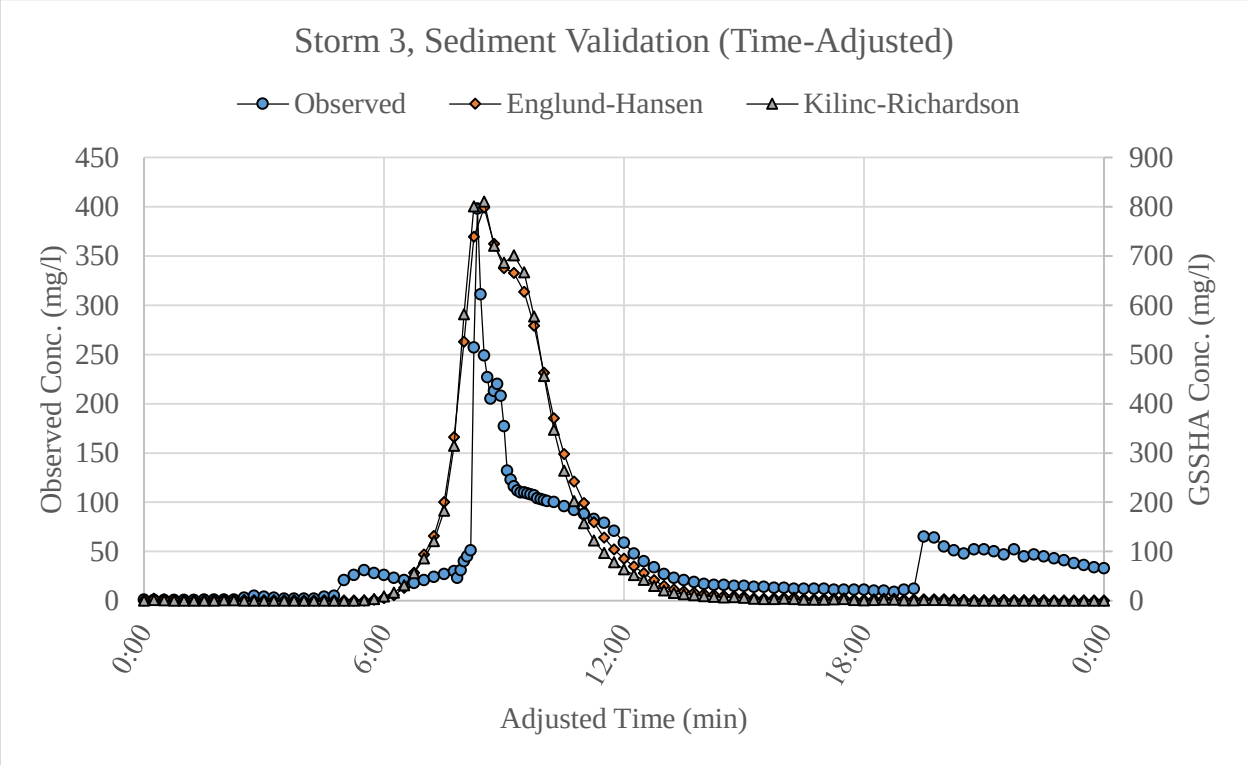

Figure 22: Storm 3 time-adjusted sediment validation.

The evaluation criteria were applied to the time-adjusted curves, resulting in R^2=0.78 and NSE=-1.31 for EH, and R^2=0.70 and NSE=-1.68 for KR.

The relatively high R^2 values could be misleading. An R^2 of 0.78 is nearly as high as the R^2 from the calibration, which would indicate a good fit. When comparing percent errors, however, and by observation, one can tell that GSSHA's accuracy in predicting sediment concentration for Storm 3 is not as good as the calibration storm, Storm 1. This discrepancy is due to insensitivity to proportionality of R^2. This insensitivity has also been noticed by other researchers (Legates 1999; Green 1986). The higher R^2 values are a result of the similar shape between the observed to computed curves, and are not indicative of the differences in magnitude. For example, as shown above, the observed and computed curves look relatively similar, if one neglects noticing that the computed curves are assigned to the secondary axis. This is an example of why R^2 should generally not be relied upon alone as a goodness of fit indicator in single-event studies. Here, R^2 seems to show how well GSSHA predicts the shape of hydrograph and sediment curves, but does not seem to be indicative of overall model performance.

3.1.7 Storm 4 Streamflow Validation

Storm 4 is the third and last high-frequency storm (<2-year recurrence interval) used to test the Storm 1 calibration of GSSHA. Storms 4 and 1 were observed to cause similar peak streamflow discharges of 215 cfs and 234 cfs, respectively. Storm 4 also has the closest 15-minute peak rainfall intensity to Storm 1, lower by 17.9%, and a similar rainfall distribution with two bursts, but approaches double the rainfall accumulation, at 82.2% higher. Further, Storm 4 is different from Storm 1, 2, and 3, in that Storm 4 occurs at the beginning of what is typically considered *dry season* in Hawaii (May through September), whereas 1, 2, and 3, occur in *wet season* (October through April). Thus, in this study, Storm 4 is the first test of GSSHA's performance of simulating a dry season storm after being calibrated by a wet season storm.

The streamflow validation hydrograph is shown below. The computed hydrograph was assigned to the secondary axis, due to large over prediction of streamflow.

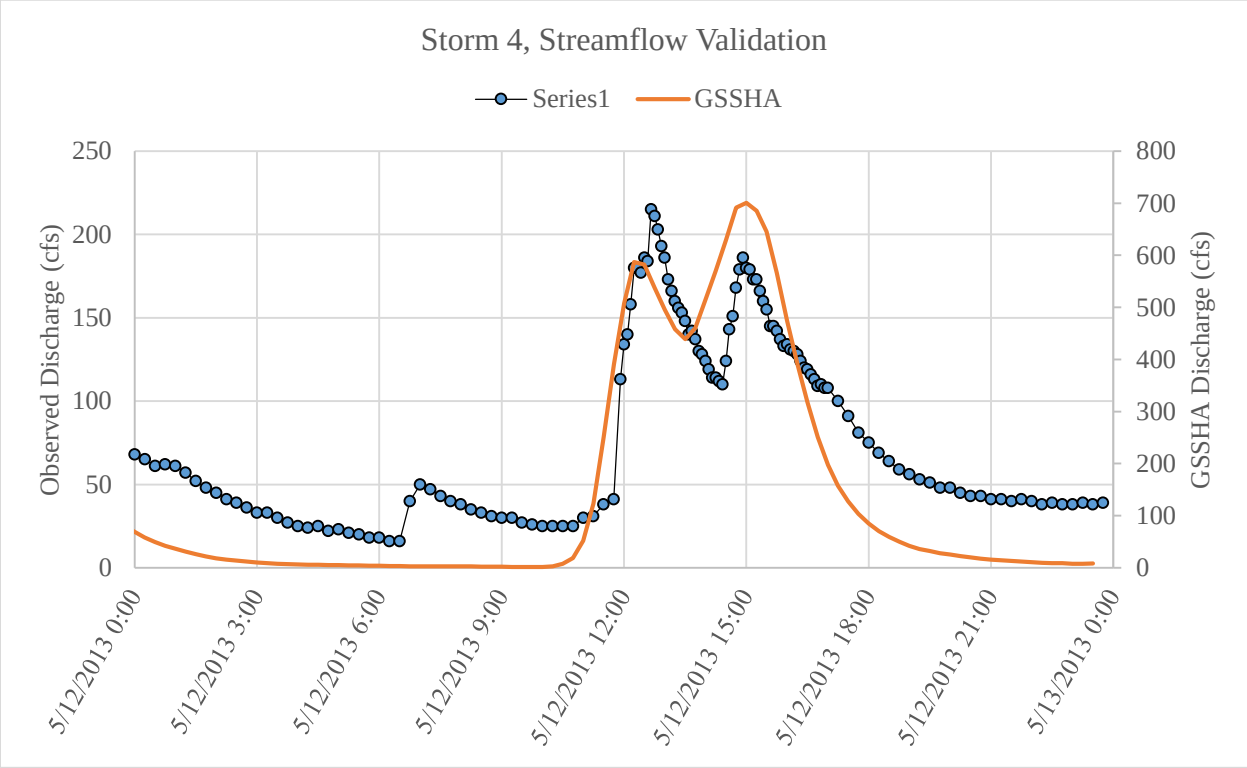

Figure 23: Storm 4 streamflow validation.

The evaluation criteria were applied to the observed and computed hydrographs from the test of Storm 4. The difference in peak is 486 cfs (215 cfs observed, 701 cfs computed), resulting in a 226.0% error. The difference in time to peak is 150 minutes (observed at 3630 minutes, computed at 3780 minutes) for a 4.1% error. There is, however, a bimodal peak in both the observed and computed hydrographs. The difference in time to peak, based on a comparison to the second of the bimodal peaks is 0 minutes (at 3780 minutes) for a 0.0% error. The difference in volume is 7604772 ft^3 (6720380 ft^3 observed, 14325152 ft^3 computed) for a 113.2% error,

which is high compared to *Shurtz* (67% error for single-event simulation), and *Downer* (17% on average), although *Downer* was simulating continuous events (Shurtz 2009; Downer 2004).

Additionally, R^2=0.74 and NSE=-6.26. Again, R^2 shows a similar shape between observed and computed hydrographs, but is insensitive to proportional differences, and NSE indicates a low degree of goodness of fit.

Here, comparing percent error calculations, R^2, and NSE, GSSHA displays poor water balance accuracy, as shown by large proportional differences in hydrographs, but good response to system dynamics, as shown by impressive similarity in hydrograph shape, even, notably, capturing the bimodal peak.

As shown above, there was a large proportional error between the observed and computed hydrographs (226% in peak and 113% in flow volume). These differences could be due to different watershed properties between wet and dry season. GSSHA was calibrated with a storm from wet season and tested with a storm from dry season. One could intuitively expect the calibration set to be different between wet and dry seasons, as found in Cunderlik (2009) with HEC-HMS. For example, soil moisture would likely be higher, at least on average, during the wet season. Applying this assumption here, the calibration set for Storm 1 would likely have higher soil moisture values than a calibration set for Storm 4. Initial soil moistures from Storm 1, therefore, would be unrealistically high for Storm 4, causing artificially high streamflow, which was seen in the computed hydrograph for Storm 4.

It is unlikely, however, that the errors in the Storm 4 hydrograph are due to this issue, because the initial soil moisture values in the calibration set for Storm one are lower than expected for a wet season storm (5% soil moisture over 86% of the watershed). The errors are more likely a result of the differences in rainfall distribution characteristics between storms 1 and

4, which would likely lead to different sets of calibration parameters—particularly hydraulic conductivities.

3.1.8 Storm 4 Sediment Validation

The following figure shows the sediment concentration validation results. Due to a large over prediction by GSSHA, the computed values were assigned to the secondary axis.

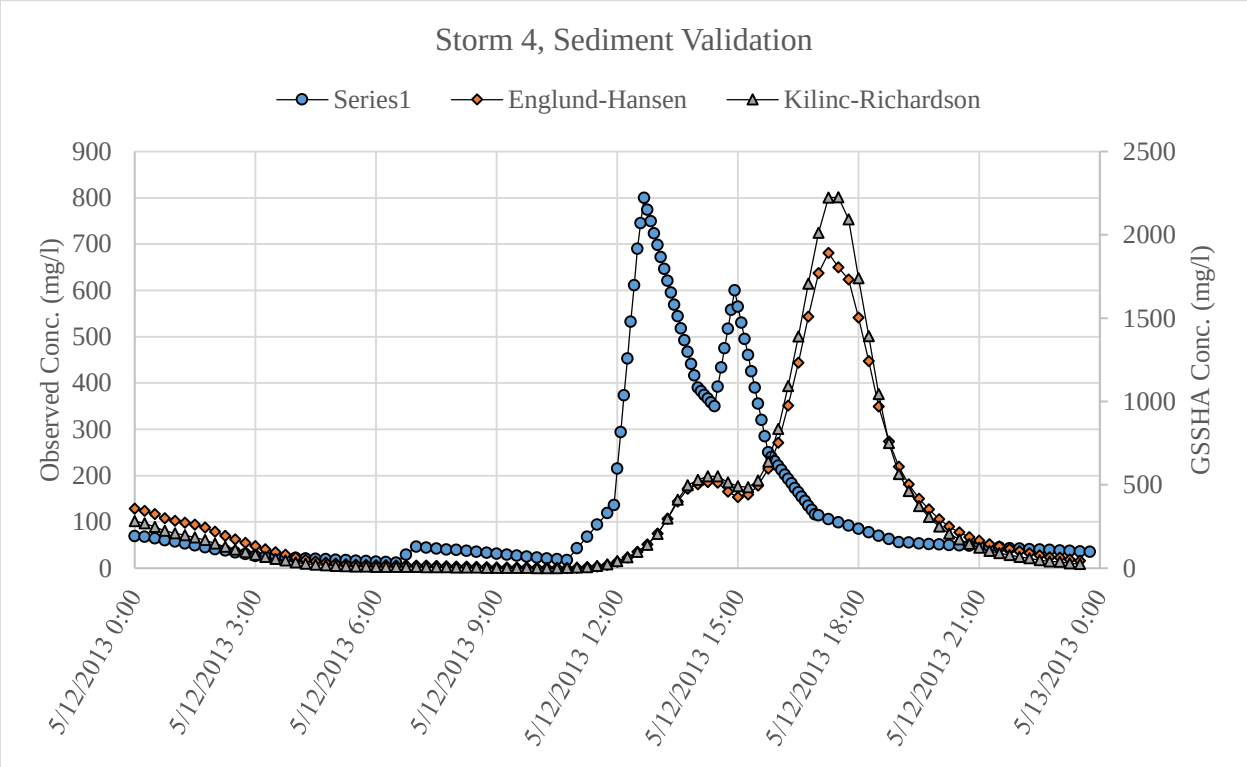

Figure 24: Storm 4 sediment validation.

The evaluation criteria were applied to the observed and computed sediment concentration curves from the test of Storm 4. The difference in peak between the observed and EH is 1092 mg/l (800 mg/l observed, 1892 mg/l computed), resulting in an error of 136.5%. The difference in peak between the observed and KR is 1426 mg/l (800 mg/l observed, 2226 mg/l

computed), resulting in a 178.3% error. The difference in time to peak is 300 minutes (3630 minutes observed, 3930 minutes computed) for an 8.3% error.

Sediment concentration was converted to tons and compared. The difference in sediment load in tons between observed and EH is 147.5 tons (44.5 tons observed, 192.0 tons computed) for a 331.5% error. The difference in sediment load in tons between observed and KR is 161.5 tons (44.5 tons observed, 206.0 tons computed) for a 362.9% error. These errors exceed the maximum error encountered in Downer (2010), which was 220%.

A comparison between the observed and computed curves yields R^2=0.12 and NSE=-5.03 for EH and R^2=0.11 and NSE=-6.77 for KR. Based on these values and the results of the percent error calculations, EH outperforms KR, but both over predict sediment.

Again, peak time was adjusted so that the observed and computed peaks match, to compare the goodness of fit, independent of timing errors. The figure below displays the time-adjusted sediment concentration curves.

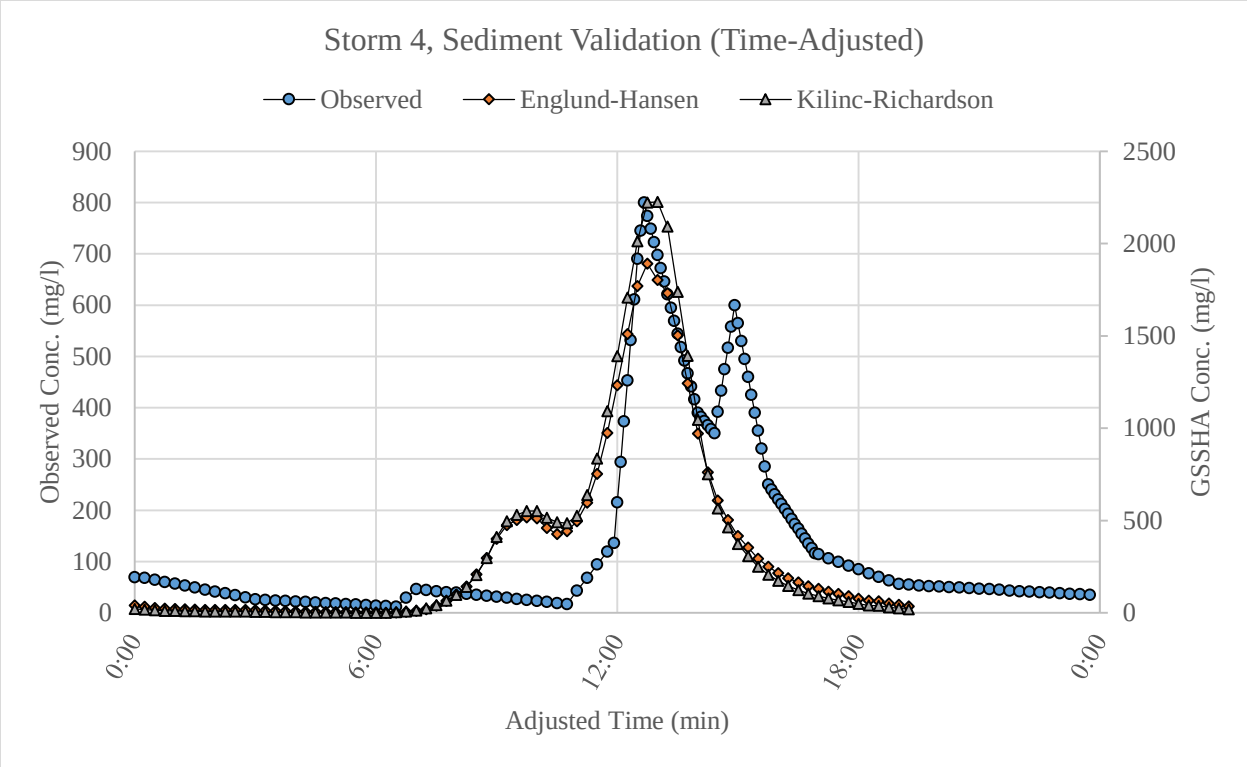

Figure 25: Storm 4 time-adjusted sediment validation.

The evaluation criteria were applied to the time-adjusted curves, resulting in R^2=0.53 and NSE=-0.22 for EH, and R^2=0.51 and NSE=-0.70 for KR.

3.1.9 Case 1 Summary

Case 1 accomplished the first part in testing whether GSSHA over predicts sediment discharge for lower-frequency storms when calibrated with higher-frequency storms, by establishing a baseline of accuracy for predicting storms of similar frequency. GSSHA was calibrated with Storm 1 and then used to predict streamflow and sediment concentration for Storms 2-4, which are of approximately the same frequency (<2-year events). Also, each storm was simulated using the EH and the KR transport equations to determine which equation is more accurate. The absolute mean error in sediment load for the validation simulations was 155% for EH and 168%

for KR (mean absolute error was mush lower without storm 4, at 67% and 71% respectively). Thus, Case 1 provided a baseline for comparison against the simulation of larger storms, and showed that EH outperforms KR for sediment prediction. The figure below shows the relationship between sediment load prediction error and peak discharge by storm, for storms 1-4 and compared to results of Downer (2010).

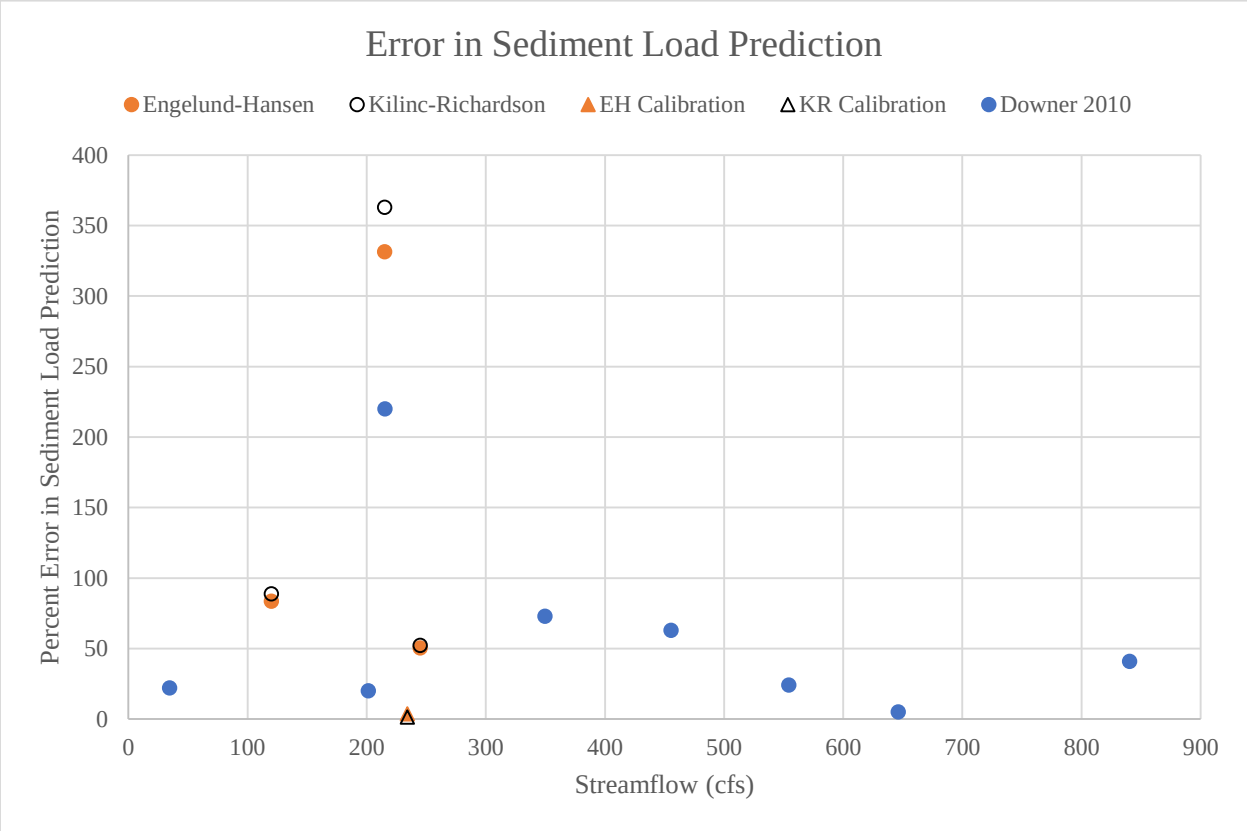

Figure 26: Error in sediment load prediction for storms 1-4.

The EH equation outperformed KR by a difference of 13% in absolute mean error. The two equations performed similarly for Storms 2 and 3, but EH outperformed KR by a difference of 31%, indicating that EH may be a better choice for single-event simulations in steep Hawaiian watersheds.

Case 1 also highlighted a few potential issues for using GSSHA for single-event simulations. First, GSSHA is more accurate when simulating storms with similar rainfall distribution characteristics as the calibration storm. For example, Storm 3 simulations had the lowest percent errors, and the rainfall distribution characteristics of Storm 3 were the most similar to Storm 1. This finding is similar to Cunderlik (2009), which found that HEC-HMS is more accurate when simulating storms with similar rainfall distribution characteristics. Second, GSSHA consistently predicted lagging sediment concentration peaks. The sediment concentration peaks for each storm, 1-4, trailed their respective hydrograph peaks by as much as 150 minutes, whereas, the observed concentration peaks fell within the same 15-minute interval as the observed hydrograph peaks.

Future studies could focus on other methods of calibrating GSSHA for single-events, to reduce sensitivity to rainfall distribution characteristics between storms (also using synthetic distributions, *e.g.*, Type 1 for Hawaii). Other methods were tried here, including calibration using multiple events (using several storms to train the model before testing), but those methods did not yield better overall performance for single-event simulations. A *ramp up* period, where the validation event is inserted into a series of calibrated storms, was also attempted, as recommended by McCarthy (2012), but, again, did not lead to better overall performance for single-event simulations.

3.2 CASE 2

Case 2: High-Frequency storm calibration followed by simulation of lower-frequency storms

3.2.1 Storm 5 Streamflow Validation

Storm 5 is the first low-frequency storm (>2-year recurrence interval) used to test the Storm 1 calibration of GSSHA. Storms 5 and 1 are highly dissimilar. Storm 5 produced approximately five times the streamflow as Storm 1 (compare with Downer (2010), which tested storms 2-4 times larger than calibration). Storm 5 is the record high for rainfall accumulation in 24-hours in the history of the USGS monitoring of Halawa (U.S. Geological Survey 2015b). Storm 1 also occurs during the dry season. Compared to Storm 1, Storm 5 has a 214.3% higher 15-minute peak rainfall intensity, and 376.7% higher rainfall accumulation. Thus, Storm 5 is an extreme test for GSSHA.

The streamflow validation hydrograph is shown below. The computed hydrograph was assigned to the secondary axis, due to large over prediction of streamflow.

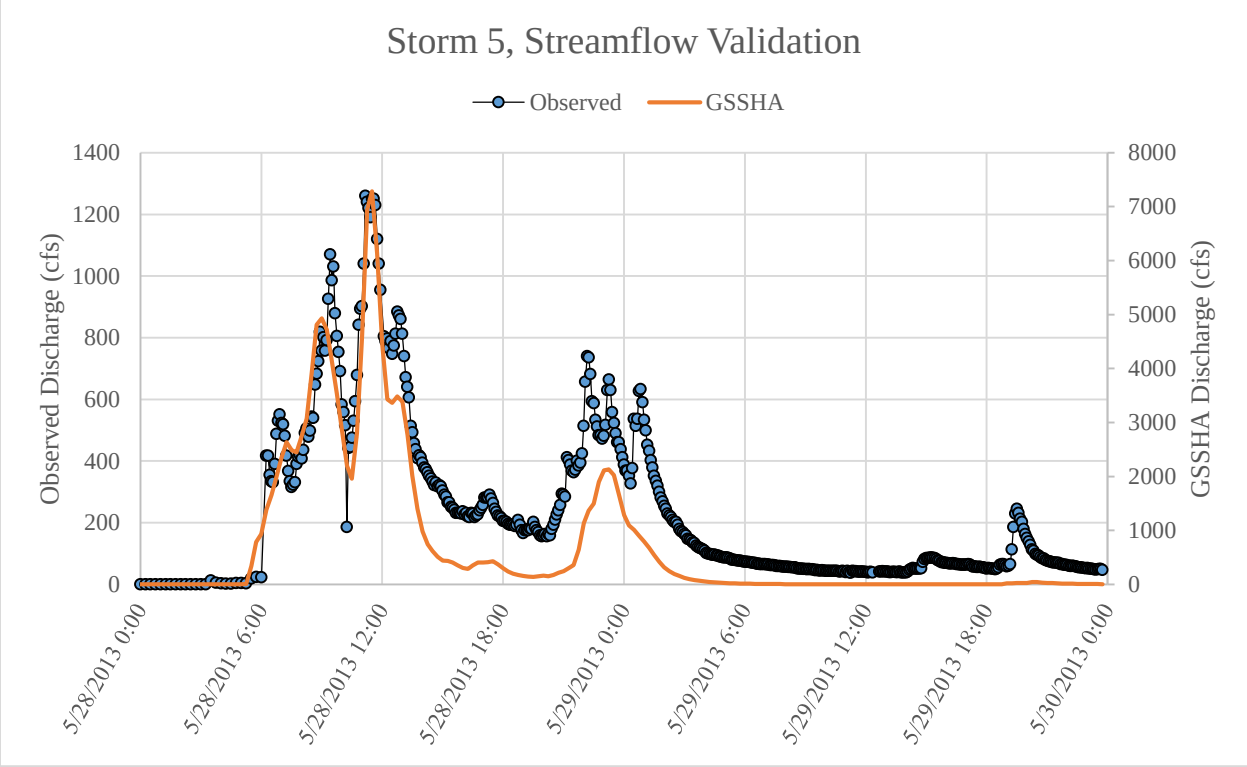

Figure 27: Storm 5 streamflow validation.

The evaluation criteria were applied to the observed and computed hydrographs for the test of Storm 5. The difference in peak is 6021 cfs (1260 cfs observed, 7281 cfs computed), resulting in a 477.9% error. The difference in time to peak is 15 minutes (observed at 2115 minutes, computed at 2130 minutes) for a 0.7% error. The difference in volume is 88389397 ft^3 (39282608 ft^3 observed, 127672005 ft^3 computed) for a 225.0% error, which is high compared to *Shurtz* (67% error for single-event simulation), and *Downer* (17% on average), although *Downer* was simulating continuous events (Shurtz 2009; Downer 2004).

Additionally, R^2=0.84 and NSE=-17.37. Again, R^2 shows a similar shape between observed and computed hydrographs, but is insensitive to proportional differences, and NSE indicates a low degree of goodness of fit.

Comparing percent error calculations, R^2, and NSE, GSSHA displays poor water balance accuracy, but good response to system dynamics, as shown by the impressive hydrograph shape prediction. GSSHA, notably, captured the bimodal peak. A large part of the water balance accuracy could be caused by the wet season calibration of this dry season storm.

3.2.2 Storm 5 Sediment Validation

The following figure shows the sediment concentration validation results.

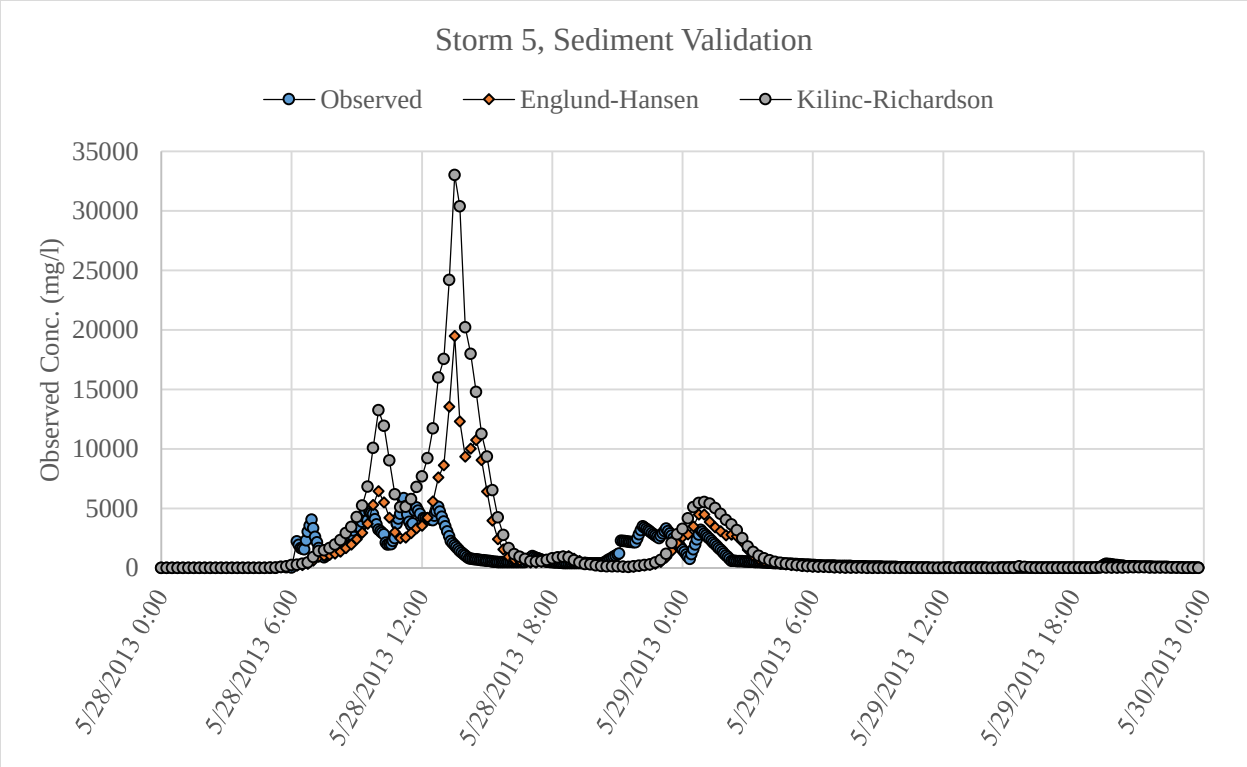

Figure 28: Storm 5 sediment validation.

The evaluation criteria were applied to the observed and computed sediment concentration curves from the test of Storm 5. The difference in peak between the observed and EH is 13643 mg/l (5860 mg/l observed, 19503 mg/l computed), resulting in an error of 232.8%. The difference in peak between the observed and KR is 27154 mg/l (5860 mg/l observed, 33014 mg/l computed), resulting in a 463.4% error. The difference in time to peak is 135 minutes (2115 minutes observed, 2250 minutes computed).

Sediment concentration was converted to tons and compared. The difference in sediment load in tons between observed and EH is 8312 tons (2865 tons observed, 11177 tons computed) for a 290.1% error. The difference in sediment load in tons between observed and KR is 18389 tons (2865 tons observed, 21254 tons computed) for a 641.8% error. These errors are much higher than the baseline mean absolute error in Case 1 of 155% EH and 168% for KR. These

errors are also much higher than the highest validation error in Downer (2010), which was an absolute maximum of 47%. This storm, however, is five times as big as the calibration storm, whereas the larger storms in *Downer* were 2-4 times larger than calibration.

A comparison between the observed and computed concentration curves yields R^2=0.22 and NSE=-2.38 for EH and R^2=0.23 and NSE=-11.47 for KR. Based on NSE and the results of the percent error calculations, EH outperforms KR by a large margin, but both EH and KR over predict sediment.

Again, peak time was adjusted so that the observed and computed peaks match, to compare the goodness of fit, independent of timing errors. The figure below displays the time-adjusted sediment concentration curves.

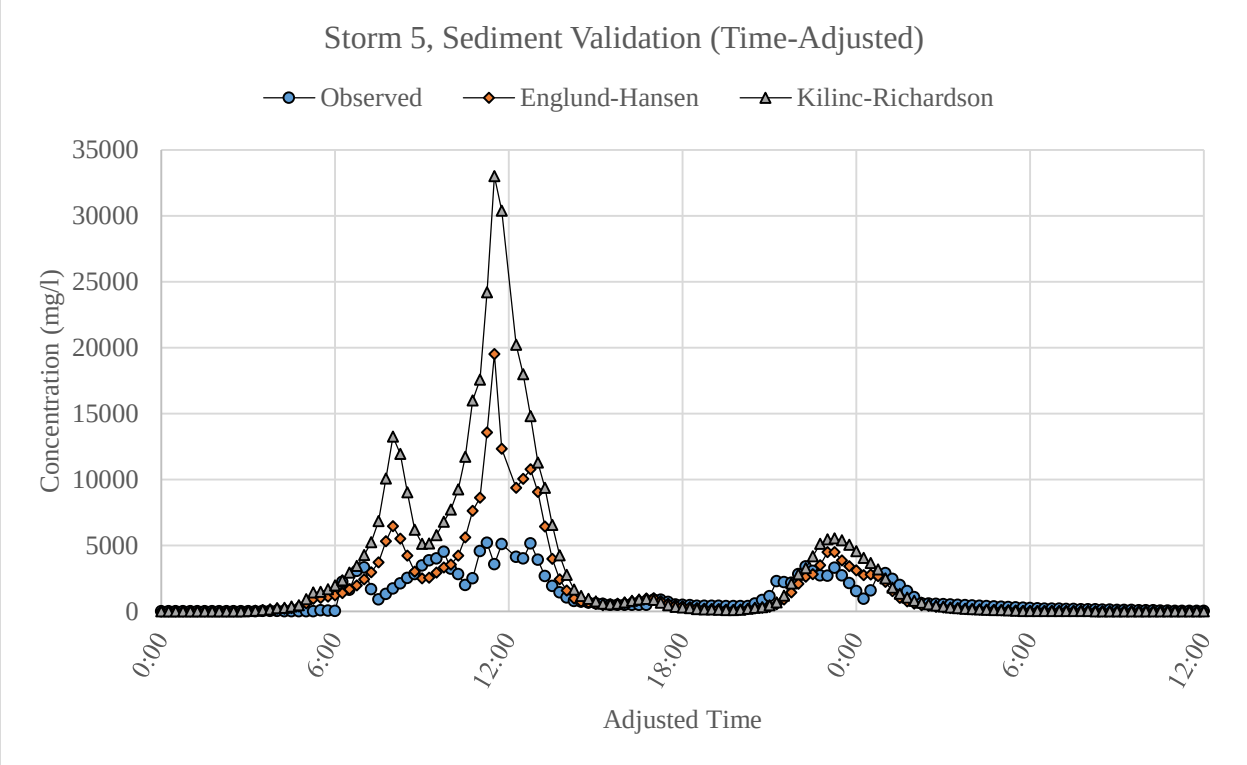

Figure 29: Storm 5 time-adjusted sediment validation.

The evaluation criteria were applied to the time-adjusted curves, resulting in R^2=0.67 and NSE=-0.93 for EH, and R^2=0.63 and NSE=-9.11 for KR.

3.2.3 Storm 6 Streamflow Validation

Storm 6 is the second and last low-frequency storm (>2-year recurrence interval) used to test the Storm 1 calibration of GSSHA. Storm 6 produced approximately six times the streamflow than Storm 1 (compare with Downer (2010), which tested storms 2-4 times larger than calibration). Also in contrast to Storm 1, Storm 6 has a 117.9% higher 15-minute peak rainfall intensity, 116.5% higher rainfall accumulation, and Storm 6 occurred during dry season. Storm 6, however, has a similar rainfall distribution to Storm 1, with two bursts of rainfall, for approximately the same duration.

The streamflow validation hydrograph is shown below. The computed hydrograph was assigned to the secondary axis, due to large over prediction of streamflow.

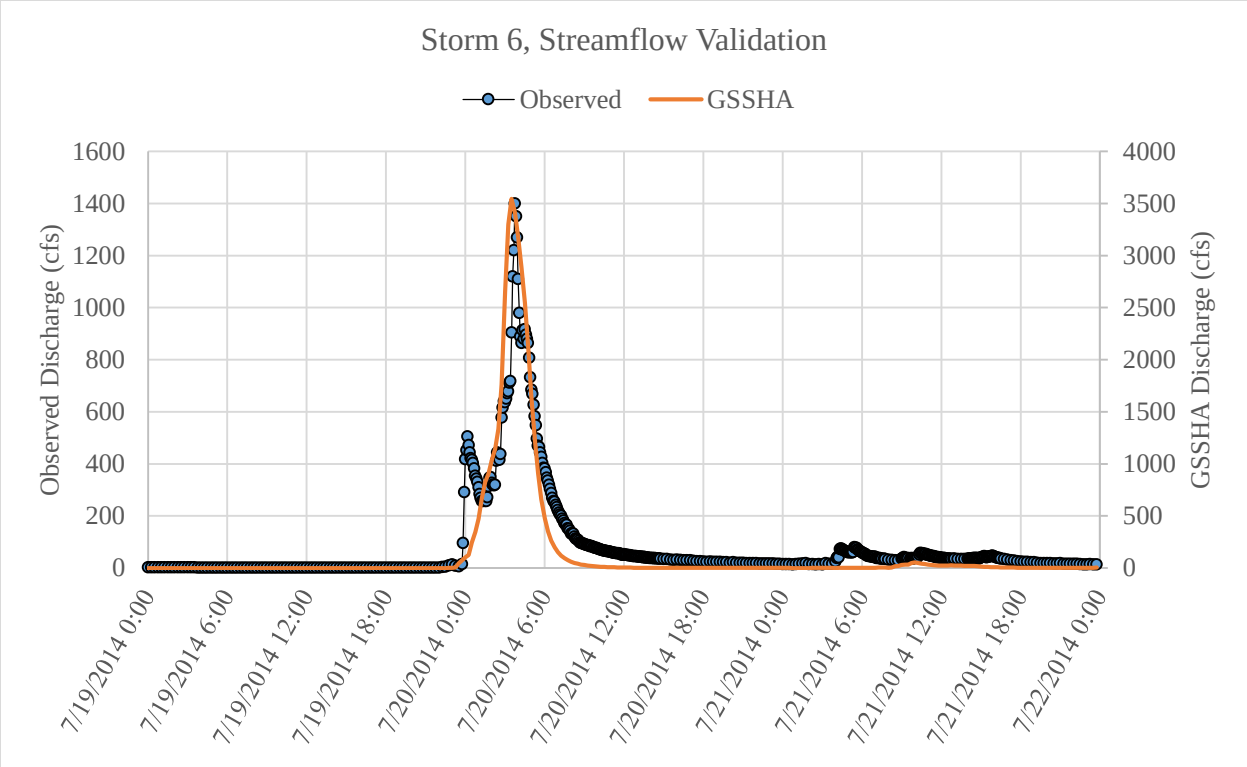

Figure 30: Storm 6 streamflow validation.

The evaluation criteria were applied to the observed and computed hydrographs for the test of Storm 5. The difference in peak is 2143 cfs (1400 cfs observed, 3543 cfs computed), resulting in a 153.1% error. The difference in time to peak is 15 minutes (observed at 1665 minutes, computed at 1650 minutes). The difference in volume is 16507799 ft^3 (19644480 ft^3 observed, 36152279 ft^3 computed) for an 84.0% error, which is high compared to *Shurtz* (67% error for single-event simulation), and *Downer* (17% on average), although *Downer* was simulating continuous events (Shurtz 2009; Downer 2004).

Additionally, R^2=0.86 and NSE=-3.37. Again, R^2 shows a similar shape between observed and computed hydrographs, but is insensitive to proportional differences, and NSE indicates a low degree of goodness of fit.

Comparing percent error calculations, R^2, and NSE, GSSHA displays poor water balance accuracy, but, again, good response to system dynamics, as shown by the impressive hydrograph shape prediction. A large part of the water balance accuracy could be caused by the wet season calibration of this dry season storm, as suspected for the Storm 4 and 6 tests.

3.2.4 Storm 6 Sediment Validation

The following figure shows the sediment concentration validation results.

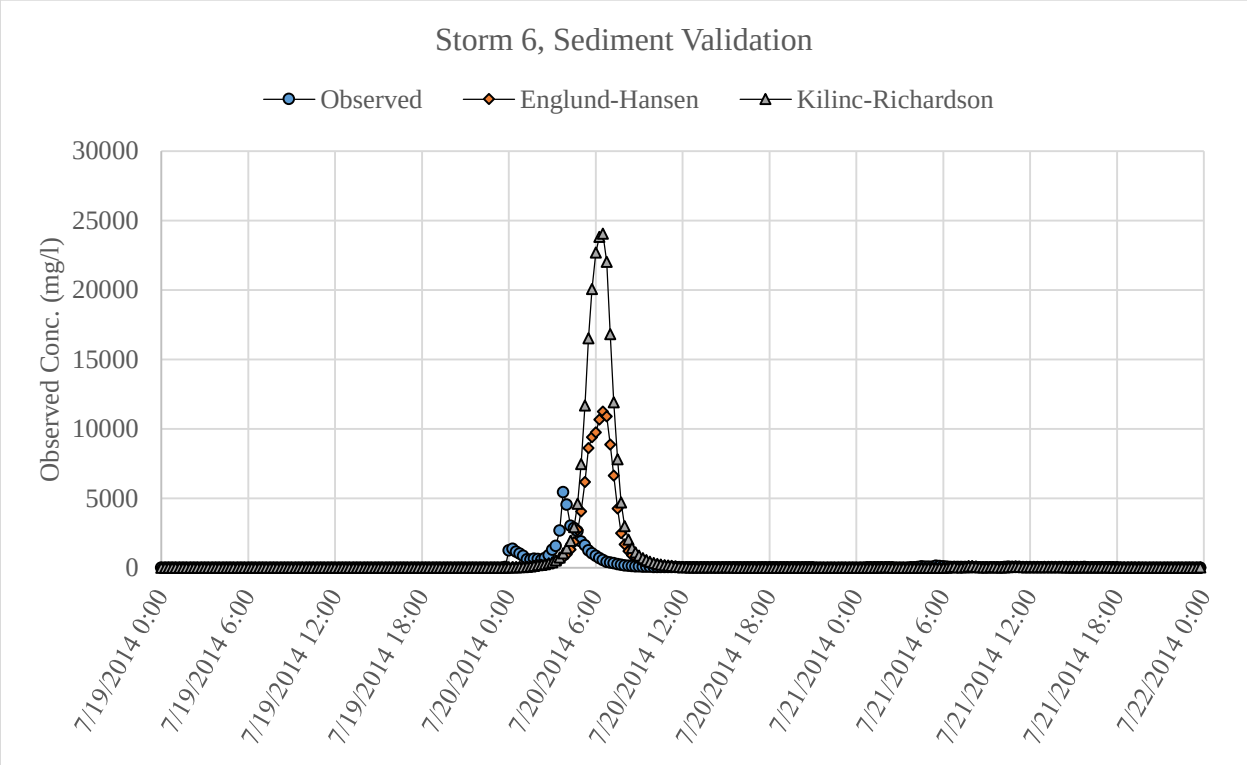

Figure 31: Storm 6 sediment validation.

The evaluation criteria were applied to the observed and computed sediment concentration curves from the test of Storm 6. The difference in peak between the observed and EH is 5807 mg/l (5430 mg/l observed, 11237 mg/l computed), resulting in an error of 106.9%.

The difference in peak between the observed and KR is 18612 mg/l (5430 mg/l observed, 24042 mg/l computed), resulting in a 343.7% error. The difference in time to peak is 135 minutes (1665 minutes observed, 1830 minutes computed).

Sediment concentration was converted to tons and compared. The difference in sediment load in tons between observed and EH is 4975 tons (904 tons observed, 5879 tons computed) for a 550.3% error. The difference in sediment load in tons between observed and KR is 10637 tons (904 tons observed, 11541 tons computed) for an 1176.7% error. These errors are much higher than the baseline mean absolute error in Case 1 of 155% EH and 168% for KR. These errors are also much higher than the highest validation error in Downer (2010), which was an absolute maximum of 47%. This storm, however, is six times as big as the calibration storm, whereas the larger storms in *Downer* were 2-4 times larger than calibration.

A comparison between the observed and computed curves yields R^2=0.09 and NSE=-3.08 for EH and R^2=0.07 and NSE=-32.73 for KR. Based on NSE and the results of the percent error calculations, EH outperforms KR by a large margin, but both EH and KR over predict sediment.

Below, peak time was adjusted so that the observed and computed peaks match, to compare the goodness of fit, independent of timing errors. The figure below displays the time-adjusted sediment concentration curves.

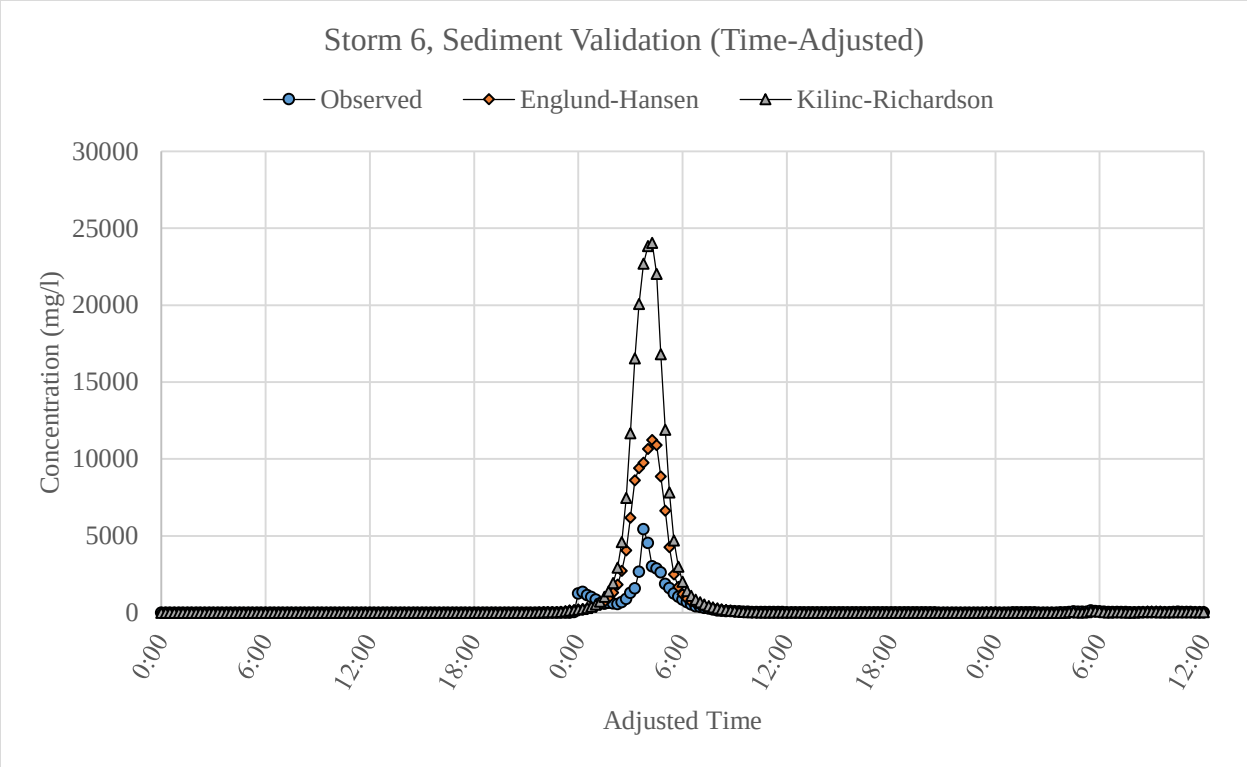

Figure 32: Storm 6 time-adjusted sediment validation.

The evaluation criteria were applied to the time-adjusted curves, resulting in R^2=0.83 and NSE=-3.08 for EH, and R^2=0.85 and NSE=-25.03 for KR.

3.2.5 Case 2 Summary

Case 2 tested whether GSSHA over predicts sediment discharge for lower-frequency storms when calibrated with higher-frequency storms. GSSHA was calibrated with Storm 1 and then used to predict streamflow and sediment concentration for Storms 5 and 6, which are five and six times larger than Storm 1, respectively. Also, each storm was simulated using the EH and the KR transport equations to determine which equation is more accurate. The baseline, absolute mean error established in Case1 for sediment load for the validation simulations was 155% for EH and 168% for KR (mean absolute error was mush lower without storm 4, at 67% and 71%

respectively). GSSHA over predicted sediment for Storm 5 by for EH and for KR. GSSHA over predicted sediment for Storm 6 by for EH and for KR. These percent errors are above the baseline set in Case 1 and also above the validation errors from Downer (2010) (-47%). Thus, Case 2 has shown that GSSHA over predicted sediment for higher-magnitude storms, when calibrated with lower-magnitude storms, for single-event simulations in steep Hawaiian watersheds. Although, the EH simulation for Storm 5 had a lower percent error in sediment load than the smaller Storm 4 (290% versus 332%). Case 2 has also shown support for the conclusion of Case1 that EH outperforms KR for sediment prediction. The figure below shows the relationship between sediment load prediction error and peak discharge by storm.

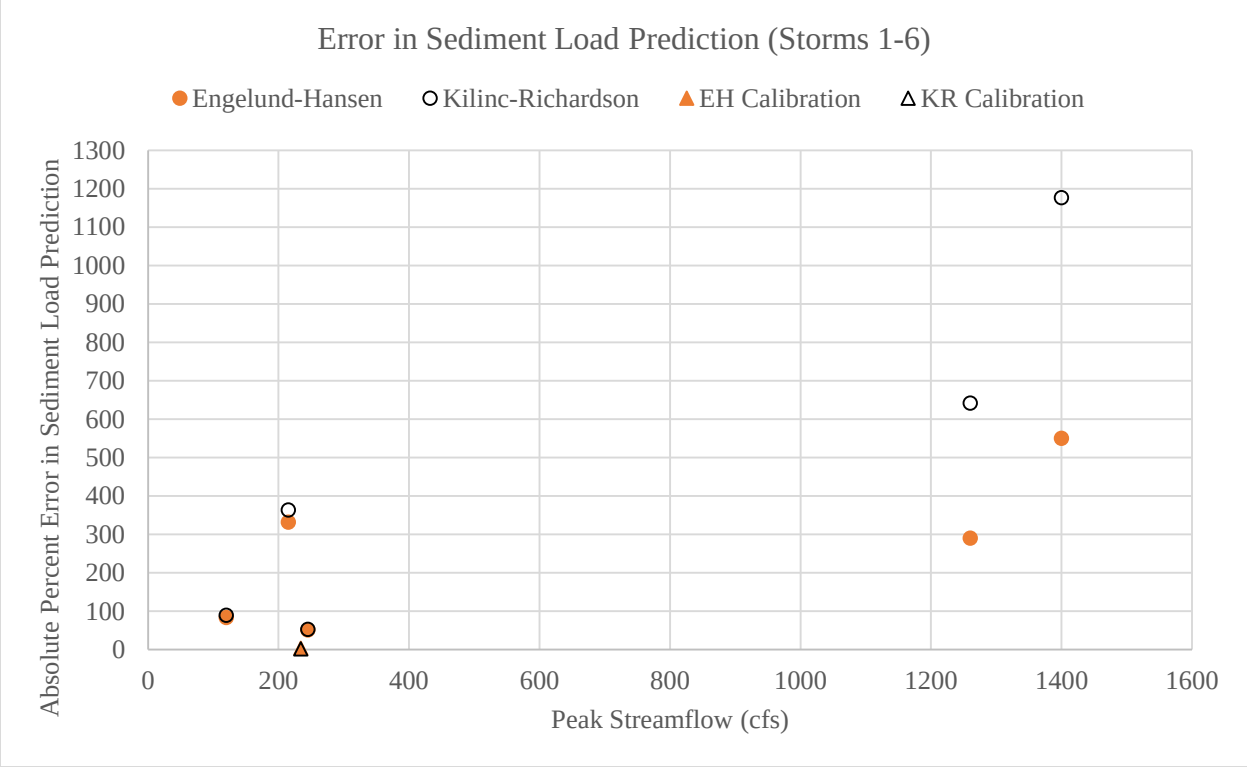

Figure 33: Error in sediment load prediction for storms 1-6.

The EH equation outperformed KR by a difference of 489% in absolute mean error, indicating that EH is likely a better choice for single-event simulations in steep Hawaiian watersheds. The EH simulation of Storm 5, notably, outperformed the smaller Storm 4, which is the only simulation that is at odds with the conclusion that GSSHA over predicts sediment for lower-frequency storms when calibrated with higher-frequency storms. The following figure shows the relationship between predicted sediment loads and peak streamflow by event.

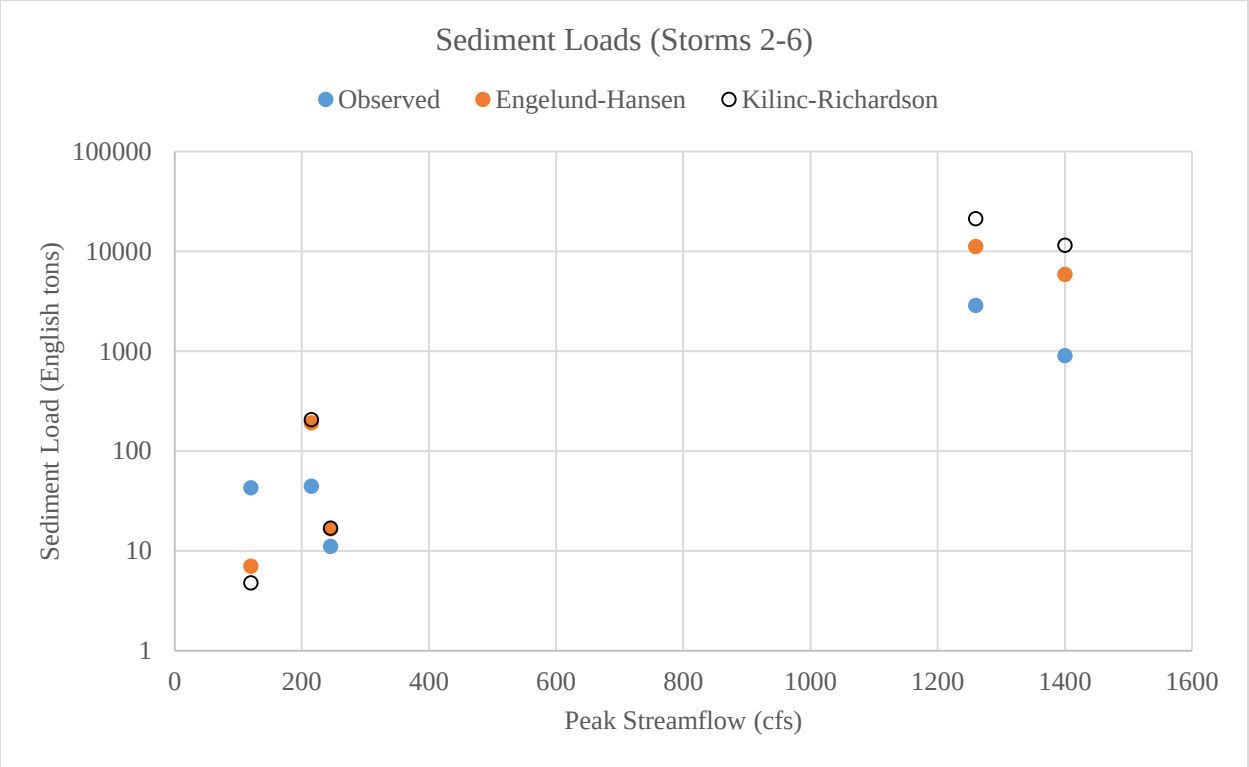

Figure 34: Sediment load predictions for storms 2-6.

The leading hypothesis for source of sediment load prediction error in the simulations of the lower-frequency storms is the over prediction of streamflow. This is similar to what was suspected by Ogden (2001). *Ogden* was the study that originally observed the over prediction, and attributed it to streamflow over prediction, as well as erosion limits, and errors in initial

moistures (Ogden 2001; Downer 2010). Case 3 was developed to test whether the over prediction in sediment load is caused by an over prediction of streamflow. The figure below shows the relationship between streamflow volume percent error and sediment load percent error.

Figure 35: Streamflow volume error to sediment load error.

The correlation coefficient between streamflow volume error and sediment load error for storms 1-6 is approximately 0.44 for both EH and KR, but 0.97 for storm 1-4.

3.3 CASE 3

Case 3: Run any additional simulations determined necessary to assess the accuracy of GSSHA in prediction sediment concentration for single-events.

The purpose of Case 3 was to test the hypothesis that the sediment prediction errors from cases 1 and 2 are the result of over prediction in streamflow. GSSHA was calibrated for predicting the streamflow of Storm 6, but the soil erosion calibration was left the same from Storm 1 calibration.

3.3.1 Storm 6, Streamflow Calibration

The following figure show the calibrated streamflow for Storm 6.

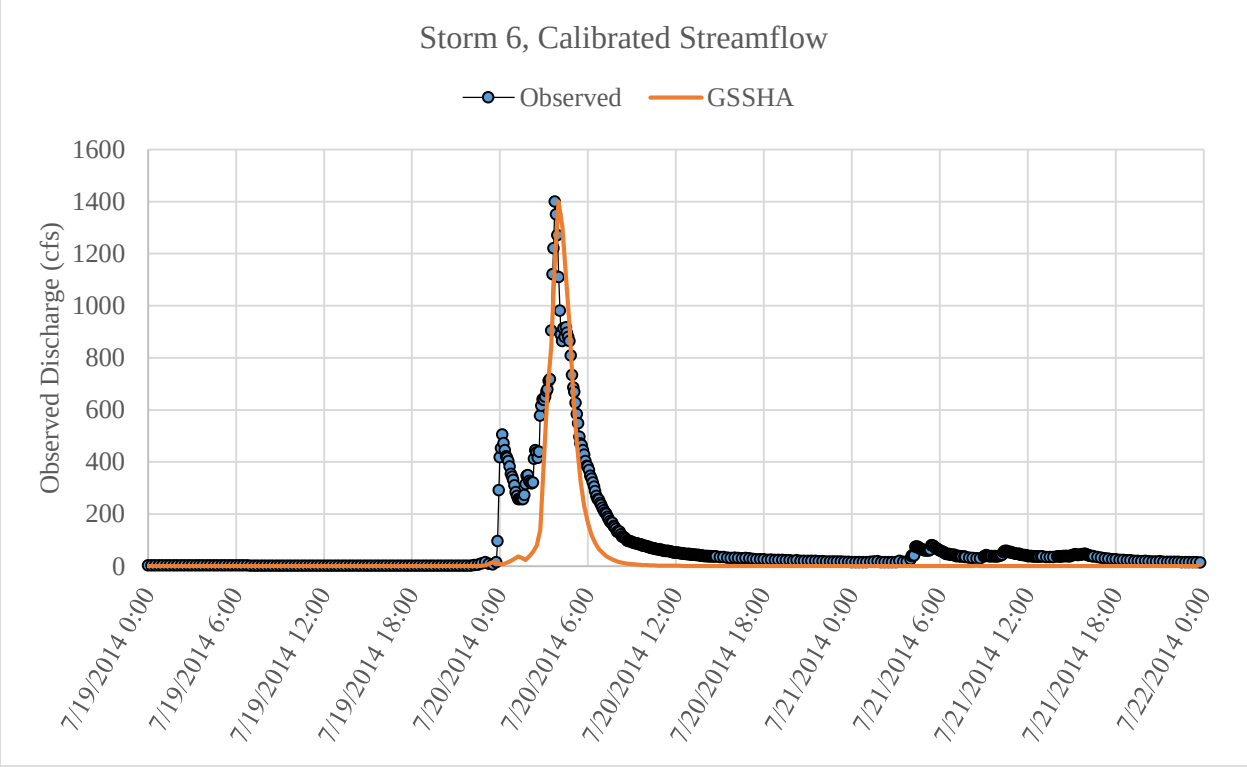

Figure 36: Storm 6 calibrated streamflow.

The evaluation criteria were applied to the observed and computed hydrographs. The difference in peak is 0 cfs (1400 cfs observed, 1400 cfs computed), resulting in a 0.0% error. The difference in time to peak is 0 minutes (at 1680 minutes) for a 0.0% error. The difference in volume is 9979701 ft^3 (19644480 ft^3 observed, 9664779 ft^3 computed) for a 50.8% error, which is similar to the percent error in volume from other GSSHA calibration studies (Shurtz 2009; Downer 2004, 53% average percent error in flow volume during calibration).

Additionally, between the two hydrograph curves, R^2=0.80, which is used as a reference point for the R^2 values in the tests that follow, and NSE=0.74, showing good fit (0.50 or greater is considered satisfactory).

3.3.2 Storm 6 (Calibrated Streamflow), Sediment Validation

The following figure shows the sediment validation with calibrated streamflow.

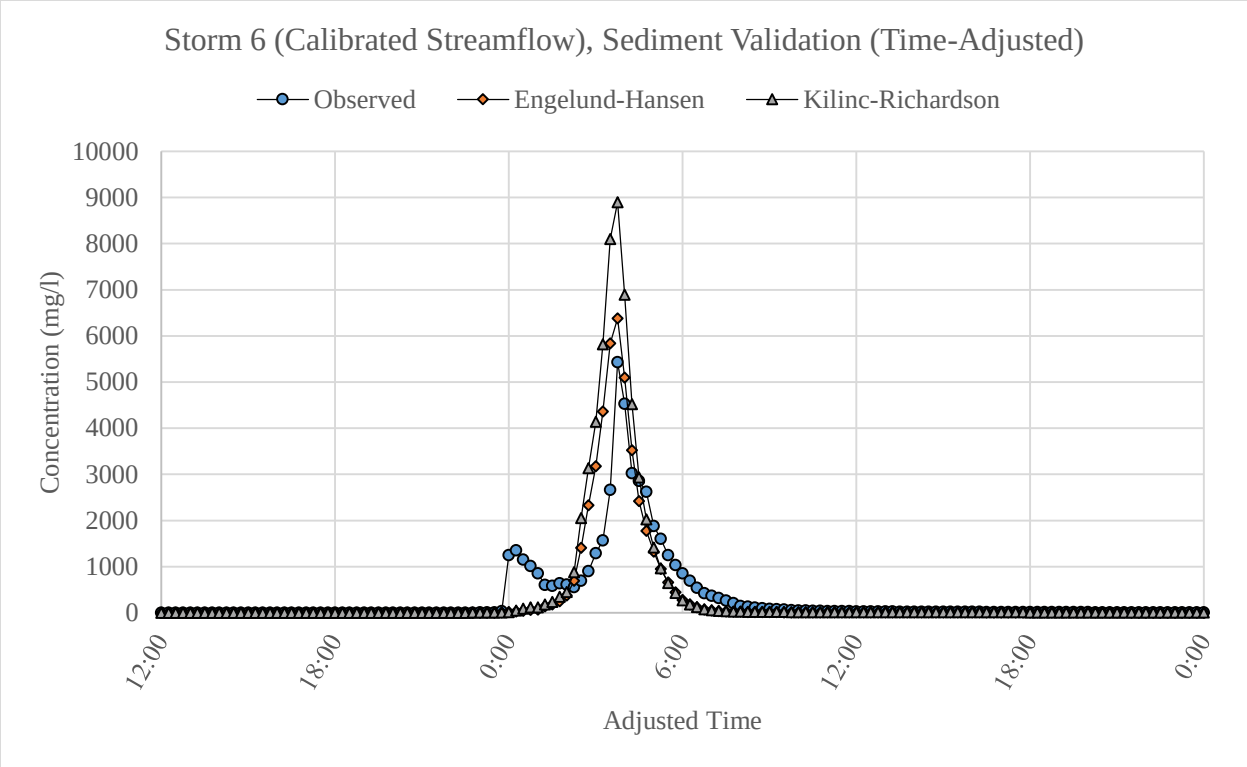

Figure 37: Storm 6 time-adjusted sediment validation with calibrated streamflow.

The evaluation criteria were applied to the observed and computed sediment concentration curves from the new test of Storm 6. The difference in peak between the observed and EH is 952 mg/l (5430 mg/l observed, 6382 mg/l computed), resulting in an error of 17.5%. The difference in peak between the observed and KR is 3469 mg/l (5430 mg/l observed, 8899 mg/l computed), resulting in a 63.9% error. The difference in time to peak is 135 minutes (2115 minutes observed, 2250 minutes computed), but time was adjusted to align the peaks.

Sediment concentration was converted to tons and compared. The difference in sediment load in tons between observed and EH is 80 tons (904 tons observed, 984 tons computed) for an 8.8% error. The difference in sediment load in tons between observed and KR is 392 tons (904 tons observed, 1296 tons computed) for a 43.4% error. These errors show a significant

improvement in prediction of sediment. They are much lower than the baselines from Case 1 and lower than the validation errors found by Downer (2010).

Thus, reducing the over prediction of streamflow by calibrating for streamflow, but maintaining the soil erosion calibration set from Storm 1 calibration, reduced the error in sediment prediction from 550.3% to 8.8% for EH and 1176.7% to 43.4% for KR. This shows that the source of the majority of error from the simulations of Storm 6 in Case 2 could be over prediction of streamflow.

A comparison between the observed and computed concentration curves yields R^2=0.78 and NSE=0.63 for EH and R^2=0.76 and NSE=0.06 for KR. These values show that EH outperforms KR, and indicate satisfactory performance from EH.

These results imply that, although GSSHA over predicts sediment for lower-frequency storms, when calibrated with higher-frequency storms, and GSSHA is sensitive to rainfall distribution characteristics, likely needing multiple calibration sets to accommodate multiple types of storms, GSSHA may be able to accurately predict sediment concentrations and discharge for any storm, with one set of soil erosion calibration parameters, using the Engelund-Hansen sediment transport equation, so long as the streamflow predictions are accurate.

CHAPTER 4

CONCLUSION

The purpose of this study was to evaluate GSSHA's accuracy and potential usefulness in predicting sediment concentrations in steep Hawaiian watersheds. GSSHA, being physically-based, has the potential to outperform other commonly used empirical or semi-empirical models for studies in Hawaii, where watersheds can be highly different from the watersheds and laboratory field plots on which the empirical models are based. GSSHA, therefore, has the potential to be a valuable tool for sediment studies in Hawaii, which often have findings that support decisions that have economic, environmental, and conservation impacts. There is a paucity of research, however, testing GSSHA's accuracy in predicting sediment concentrations, and no known research validating its use for sediment concentration of discharge prediction in Hawaii.

The first objective of the study was to develop case studies to test whether GSSHA over predicts sediment concentrations for lower-frequency storms when calibrated with higher-frequency storms, in support of the findings of Ogden (2001) and as cautioned by the GSSHA Manual (Downer 2006), or whether GSSHA accurately predicts sediment concentrations in these circumstances, in corroboration of Downer (2010). The first objective also called for a comparison between two sediment transport equations used by GSSHA—the Engelund-Hansen (EH) and Kilinc-Richardson (KR)—to determine which is more accurate in steep terrain.

In response to the first objective, three case studies were developed. The case studies were all set in the upper Halawa watershed (average overland basin slope of 0.65) in leeward Oahu, to maintain a focus on the use of GSSHA in steep Hawaiian watersheds. Rainfall,

streamflow, and sediment concentration data were gathered from the USGS for six different storm events, which formed the basis of the case studies. Additionally, a flood-flow frequency analysis was performed to classify the storm data by recurrence interval, based on peak discharge, and to identify low- and high-frequency storms.

Case 1 was designed to (1) test GSSHA with high-frequency storms after being calibrated with a high-frequency storm, to set a baseline for prediction accuracy in sediment load before simulating low-frequency storms, and (2) begin comparison of the performance between the EH and KR equations. GSSHA was calibrated with Storm 1 and then used to predict streamflow and sediment concentration for Storms 2-4, which are of approximately the same frequency (<2-year events). Also, each storm was simulated using the EH and the KR transport equations to determine which equation is more accurate.

Case 2 was designed to (1) test whether GSSHA over predicts sediment discharge for lower-frequency storms when calibrated with higher-frequency storms, compared to the baseline set in Case 1, and (2) continue the comparison of the performance of the EH and KR equations. GSSHA was calibrated with Storm 1 and then used to predict streamflow and sediment concentration for Storms 5 and 6, which are five and six times larger than Storm 1, respectively. Also, each storm was simulated using the EH and the KR transport equations to determine which equation is more accurate.

The table below displays GSSHA's accuracy in predicting sediment loads during the simulations of Case 1 and Case 2.

Table 3: Sediment load prediction errors by storm.

Storm	Streamflow (cfs)	EH Sediment Load Error % (tons/tons)	KR Sediment Load Error % (tons/tons)
1 (Calibration)	234	3.7	1.5
2	120	83.6	88.8
3	245	50.5	52.3
4	215	331.5	362.9
5	1260	290.1	641.8
6	1400	550.3	1176.7

The absolute mean errors in sediment load for the high-frequency validation simulations 2-4 were 155% for EH and 168% for KR, which constituted the baseline. The absolute errors for the low-frequency validation simulations 5-6 were much higher than the baseline, as shown above, which indicates that GSSHA did over predict sediment for low-frequency storms, after being calibrated with a high-frequency storm. This finding is similar to Ogden (2001), and the cautionary statement in the GSSHA manual (Downer 2006), and could be viewed as at odds with Downer (2010).

In comparison to Downer (2010), the errors in sediment prediction are higher. *Downer* found a maximum sediment error of -47% (KR was the only equation used) for a storm that was four times larger than the calibration storms. Here, the maximum sediment error was 550% for EH and 1177% for KR for a storm that was six times larger than the calibration storm. These large differences in errors could be due to the larger proportional difference between calibration and validation storms, or a difference in calibration procedures. First, Downer used a lower-frequency storm that was four times larger than calibration, and here a storm six times larger was

tested. The larger difference in proportionality could be required to cause GSSHA to over predict as seen in Ogden and as cautioned in the GSSHA manual (Downer 2006). Future studies could test this idea and further quantify storm magnitude differences between calibration and validation and the sediment errors that result.

Second, *Downer* used several storms of similar frequency to calibrate GSSHA before testing with larger storms, whereas here GSSHA was calibrated with one storm before being tested with similar and larger storms. Calibrating with multiple storms (making a conglomerate event that is longer than the three-day storm used here) could give a parameter set that is more of an average of actual parameter values and possibly lead to better accuracy on average or in certain cases. If so, this could explain the difference between the findings here and in Downer (2010).

Also, these results are more similar to Downer (2010) when viewing only the high-frequency storms. Storm 4 had much higher errors than 2 and 3. Ignoring 4, the absolute mean error in sediment load for the validation simulations 2-3 were 67% for EH and 71% for KR. These values for 2-3 are within the range of error found in Downer (2010) (error range for sediment volume in calibration was -73% to 220%).

Additionally, it was observed from the sediment load errors from cases 1 and 2 that GSSHA is more accurate when the validation storms are similar in rainfall distribution characteristics. The more similar the rainfall distribution characteristics between calibration and validation storms, the more accurate GSSHA predicted streamflow and sediment. For example, Storm 3 had the most similar rainfall distribution characteristics to Storm 1 out of all of the high-frequency storms, having nearly the same total rainfall accumulation and rainfall duration. Storm three also had low hydrograph errors and the lowest error sediment load error.

The table above (Table 3) also shows that the EH outperforms KR in sediment load prediction. What is not apparent in the table is that the EH is not just outperforming the KR by predicting less sediment, but by consistently predicting sediment closer to observed values. EH predicted more accurate and lower sediment loads for storms 3, 4, 5, and 6, and more accurate and higher sediment loads for storm 2. Exact reasons for better performance are unknown and beyond the scope of this study. It may be beneficial, however, for future researchers to look into possible reasons to explain the better performance of the EH method.

Reasons for over prediction in general were considered. Ogden (2001) originally observed the over prediction, which led to the caution in the GSSHA manual (Downer 2006), and attributed it to streamflow over prediction, as well as erosion limits, and errors in initial moistures (Downer 2010). Streamflow volume over prediction was observed here to have a strong correlation to sediment load error (0.97 correlation coefficient) for storms 1-4.

Case 3 was designed to test whether the over prediction in sediment load is caused by an over prediction of streamflow. GSSHA was calibrated for the streamflow of Storm 6 and used to test sediment load prediction using the same soil erosion calibration obtain from the calibration of Storm 1. The results from Case 3 are combined with cases 1 and 2 in the table below.

Table 4: Sediment load prediction errors by storm with calibrated Storm 6.

Storm	Streamflow (cfs)	EH Sediment Load Error % (tons/tons)	KR Sediment Load Error % (tons/tons)
1 (Calibrated)	234	3.7	1.5
2	120	83.6	88.8
3	245	50.5	52.3
4	215	331.5	362.9
5	1260	290.1	641.8
6	1400	550.3	1176.7
6 (Calibrated)	1400	8.8	43.4

The percent errors in sediment load were drastically reduced with streamflow, as shown in the table above, by calibrating for streamflow, but maintaining the soil erosion calibration set from Storm 1 calibration. This shows that the source of the majority of error from the simulations of Storm 6 in Case 2 could be over prediction of streamflow.

A comparison between the observed and computed concentration curves yields R^2=0.78 and NSE=0.63 for EH and R^2=0.76 and NSE=0.06 for KR. These values show that EH outperforms KR, and indicate satisfactory performance from EH.

These results imply that, although GSSHA over predicts sediment for lower-frequency storms, when calibrated with higher-frequency storms, and may require multiple sets of calibration parameters for predicting streamflow, GSSHA may be able to accurately predict sediment concentrations and discharge for any storm, with one set of soil erosion calibration parameters, using the Engelund-Hansen sediment transport equation, so long as the streamflow predictions are accurate.

The second and third objectives of this study were to determine whether GSSHA can be relied upon to predict watershed sediment concentration in steep Hawaiian watersheds during single-events, based on the findings of the case studies, and make recommendations of potential GSSHA improvements.

In response to these objective the conclusions from cases 1, 2, and 3, are summarized below, and are supplemented by guidance on how GSSHA can be successfully used in single-event studies in Hawaii, and recommendations of potential improvements.

4.1 SUMMARY OF MAIN CONCLUSIONS

- GSSHA made more accurate predictions of sediment load in validation when validation storm events were similar in rainfall distribution characteristics (rainfall duration, accumulation, peak intensity) to calibration events. This is similar to the findings in Cunderlik (2009), which found the same to be true of HEC-HMS, and Shurtz (2009). The implication is that GSSHA can be sensitive to rainfall distribution characteristics for single-event simulations, and that one set of calibration parameters may not be sufficient to simulate multiple storms with different characteristics. Thus, caution should be used in simulating multiple single-event storms with different rainfall distribution characteristics from a single-event calibration. Also, using multiple storms for calibration may give better performance *on average*, as shown in Downer (2010).
- GSSHA was observed to over predict sediment loads for lower-frequency storms (>2-year recurrence interval) after being calibrated for higher-frequency storms (<2-year recurrence interval), similar to the findings in Ogden (2001), and as cautioned by the GSSHA manual (Downer 2006), but at odds with Downer (2010). Thus, caution should be taken when simulating storms five and six times larger than storms used for calibration. Storms two to four times larger may be okay, based on the findings of Downer (2010). More research should be done to further quantify storm magnitude differences between calibration and validation and the relationship between those differences and prediction error.
- GSSHA accurately predicted sediment loads in validation from a storm six times greater than the calibration storm, using the same set of soil erosion calibration

parameters. So, although it was found that one set of calibration parameters may not be sufficient to simulate accurate streamflow from multiple storms with different rainfall distribution characteristics, one set of soil erosion calibration parameters may be sufficient to simulate accurate sediment loads from multiple storms of different magnitude and rainfall distribution characteristics. Thus, so long as the streamflow prediction is accurate, GSSHA may be able to accurately predict sediment load for storms of all sizes.

- The Engelund-Hansen equation outperformed the Kilinc-Richardson equation in predicting sediment loads in upper Halawa watershed for every validation simulation. Thus, the Engelund-Hansen equation is likely a better choice for sediment load prediction in steep Hawaiian watersheds. Reasons for the difference in performance remain unknown. Some sediment transport equations have been found to be inaccurate in steep terrain, because they do not account for the immobility of larger grains (Yager 2012). Perhaps the Engelund-Hansen can account for the effects of these larger grains, because it calculates sediment discharge compartmentally by size, unlike the Kilinc-Richardson, and is, therefore, more accurate. More research should be done to determine why the Engelund-Hansen equation outperformed the Kilinc-Richardson equation.

REFERENCES

ASCE Task Committee. (1993). "Criteria for evaluation of watershed models." *Irrigation and Drainage*, 119(3), 429–442.

Aquaveo, LLC (2014a). "GSSHA—Calibration—Computer-Based Calibration of GSSHA Models." *WMS 10.0 Tutorials*. Retrieved from http://www.aquaveo.com/software/wms-learning-tutorials. Accessed [12/05/2015].

Aquaveo, LLC (2014b). "GSSHA—Calibration—Manual Calibration of GSSHA Models." *WMS 10.0 Tutorials*. Retrieved from http://www.aquaveo.com/software/wms-learning-tutorials. Accessed [12/05/2015].

Clark, G. M., and Woods, P. F. (2001). "Transport of Suspended and Bedload Sediment at Eight Stations in the Coeur d'Alene River Basin, Idaho." *U.S. Geological Survey, Books and Open-File Reports*, 00–472.

Cunderlik, J. M. (2009). "Calibration, verification, and sensitivity analysis of the HEC-HMS hydrologic model." *Assessment of Water Resources Risk and Vulnerability to Changing Climatic Conditions*, (August), 1–10.

Dillaha, T. A., Sherrard, J. H., Lee, D., Mostaghimi, S., and Shanholtz, V. O. (1988). "Evaluation of Vegetative Filter Strips as a Best Management Practice for Feed Lots." *Water Pollution Control Federation*, 60(7), 1231–1238.

Downer, C. W., and Ogden, F. L. (2003). "Prediction of runoff and soil moistures at the watershed scale: Effects of model complexity and parameter assignment." *Water Resources Research*, 39(3), 1045.

Downer, C. W., and Ogden, F. L. (2004). "GSSHA: Model to Simulate Diverse Stream Flow Producing Processes." *Journal of Hydrologic Engineering*, 9(3), 161–174.

Downer, C. W., and Ogden, F. L. (2006). "Gridded Surface Subsurface Hydrologic Analysis (GSSHA): User's Manual." Retrieved from http://wmsdocs.aquaveo.com/gssha_manual_erdc.pdf. Accessed [05/05/2015].

Downer, C., Ogden, F., Pradhan, N., Liu, S., and Byrd, A. (2010). "Improved Soil Erosion and Sediment Transport in GSSHA." *Erdc Tn-Swwrp-10-3*.

Engelund, F., and Hansen, E. (1967). "A monograph on sediment transport in alluvial streams." 65. Teknisk Forlag, Copenhagen, Denmark.

Fabricius, K. E. (2005). "Effects of terrestrial runoff on the ecology of corals and coral reefs: review and synthesis." *Marine Pollution Bulletin*, 50(2), 125–146.

Federal Emergency Management Agency (2015). “Hydrologic Models Meeting the Minimum Requirement of National Flood Insurance Program.” Retrieved from http://www.fema.gov/hydrologic-models-meeting-minimum-requirement-national-flood-insurance-program. Accessed [12/05/2015].

Gabet, E. J., and Dunne, T. (2003). “A stochastic sediment delivery model for a steep Mediterranean landscape.” *Water Resources Research*, 39(9), ESG 2–1–ESG 2–12.

Gassman, P. W., Reyes, M. R., Green, C. H., and Arnold, J. G. (2007). “The soil and water assessment tool: Historical development, applications, and future research directions.” *Transactions of the Asabe*, 50(4), 1211–1250.

Glysson, G. D. (1987). “Sediment-transport Curves.” *U.S. Geological Survey, Books and Open-File Reports*, Open-File Report 87-218.

Green, I. R. A., and Stephenson, D. (1986). “Criteria for comparison of single event models.” *Hydrological Sciences Journal*, 31(3), 395–411.

Giambelluca, T.W., Q. Chen, A.G. Frazier, J.P. Price, Y.-L. Chen, P.-S. Chu, J.K. Eischeid, and D.M. Delparte (2013). “Online Rainfall Atlas of Hawai‘i.” *Bull. Amer. Meteor. Soc.* 94, 313-316.

Huang, T., and Lo, K. (2015). “Effects of Land Use Change on Sediment and Water Yields in Yang Ming Shan National Park, Taiwan.” *Environments*, 2(1), 32–42.

Julien, P. Y. (1995). “Erosion and Sedimentation.” Cambridge University Press, Cambridge, New York.

Kalin, L., & Hantush, M. M. (2003). “Evaluation of sediment transport models and comparative application of two watershed models.” (EPA/600/R-03/139). Retrieved from https://clu-in.org/download/contaminantfocus/sediments/evaluating-sed-transport-models-Ada-600r03139_open.pdf. Accessed [05/05/2015].

Krause, P., and Boyle, D. P. (2005). “Advances in Geosciences Comparison of different efficiency criteria for hydrological model assessment.” *Advances In Geosciences*, 5(89), 89–97.

Krysanova, V., and Srinivasan, R. (2014). “Assessment of climate and land use change impacts with SWAT.” *Regional Environmental Change*, 15(3), 431–434.

Legates, D. R., and McCabe, G. J. (1999). “Evaluating the use of ‘goodness-of-fit’ measures in hydrologic and hydroclimatic model validation.” *Water Resources Research*, 35(1), 233–241.

Lenat, D. R., and Crawford, J. K. (1994). “Effects of land use on water quality and aquatic biota of three North Carolina Piedmont streams.” *Hydrobiologia*, 294(3), 185–199.

Markus, M., and Demissie, M. (2006). “Predictability of Annual Sediment Loads Based on Flood Events.” *Journal of Hydrologic Engineering*, 11(4), 354–361.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Binger, R. L., Harmel, R. D., and Veith, T. L. (2007). “Model evaluation guidelines for systematic quantification of accuracy in watershed simulations.” *Transactions of the ASABE*, 50(3), 885–900.

Ogden, F., & Heilig, A. (2001). “Two-dimensional watershed-scale erosion modeling with CASC2D.” In *Landscape erosion and evolution modeling* (pp. 277-320). Springer US.

Rojas, R., Velleux, M., Julien, P. Y., and Johnson, B. E. (2008). “Grid Scale Effects on Watershed Soil Erosion Models.” *Journal of Hydrologic Engineering*, 13(9), 793–802.

Saha, G. (2015). “Climate Change Induced Precipitation Effects on Water Resources in the Peace Region of British Columbia, Canada.” *Climate*, 3(2), 264–282.

Sahoo, G. B., Ray, C., and Carlo, E. H. De. (2006). “Calibration and validation of a physically distributed hydrological model, MIKE SHE, to predict streamflow at high frequency in a flashy mountainous Hawaii stream.” *Journal of Hydrology*, 327, 94–109.

Sharif, H. O., Ogden, F. L., Krajewski, W. F., and Xue, M. (2004). “Statistical analysis of radar rainfall error propagation.” *Journal of Hydrometeorology*, 5, 199–212.

Sharif, H. O., Yates, D., Roberts, R., and Mueller, C. (2006). “The Use of an Automated Nowcasting System to Forecast Flash Floods in an Urban Watershed.” *Journal of Hydrometeorology*, 7(1), 190–202.

Shurtz, K. M. (2009). “Automated Calibration of the GSSHA Watershed Model: A Look at Accuracy and Viability for Routine Hydrologic Modeling.” Brigham Young University.

Soil Survey Staff (2015), Natural Resources Conservation Service, U.S. Department of Agriculture. *Web Soil Survey*. Retrieved from http://websoilsurvey.nrcs.usda.gov/. Accessed [05/05/2015].

Stark, C. P., and Stark, G. J. (2001). “A channelization model of landscape evolution.” *American Journal of Science*, 301, 486–512.

Sustainable Resources Group Intn’l, I. (2012). *Wahikuli-Honokōwai Watershed Management Plan Volume 2 : Strategies and Implementation*. Report.

Swain, N., Christensen, S., Latu, K., Jones, N., Nelson, E., & Williams, G. (2013). "A geospatial relational data model for ingesting GSSHA computational models: A step toward two-dimensional hydrologic modeling in the cloud." *World Environ. Water Resour. Congr.*, 201, 2716-2725.

Thompson, B. G. (2014). "Modeling a Snowmelt-Dominated Watershed in Northern Utah Using GSSHA." Brigham Young University.

U.S. Geological Survey (2015a). Pacific Coastal and Marine Science Center. Sediment Studies, Pacific Coral Reefs Website. Retrieved from http://coralreefs.wr.usgs.gov/sediment.html. Accessed [12/05/2015].

U.S. Geological Survey (2015b). National Water Information System data available on the World Wide Web (Water Data for the Nation). Retrieved from http://waterdata.usgs.gov/nwis/. Accessed [05/05/2015].

U.S. Geological Survey Field Center (2007). *Sediment Transport Modeling Review Workshop, Review Panel Report*. Retrieved from https://www.usbr.gov/uc/rm/amp/twg/mtgs/07jun25/Attach_03a.pdf. Accessed [05/05/2015].

U.S. Water Resources Council. (1982). "Guidelines for Determining Flood Flow Frequency." *Bulletin 17B: Reston, Virginia, Hydrology Subcommittee,* Office of Water Data Coordination, U.S. Geological Survey, 182.

Xie, H., Chen, L., and Shen, Z. (2015). "Assessment of Agricultural Best Management Practices Using Models: Current Issues and Future Perspectives." *Water*, 7(3), 1088–1108.

Yager, E. M., Dietrich, W. E., Kirchner, J. W., and Mcardell, B. W. (2012). "Prediction of sediment transport in step-pool channels." *Water Resources Research*, 48, 1–20.

APPENDIX

FLOOD-FLOW FREQUENCY ANALYSIS OF UPPER HALAWA WATERSHED

The Log Pearson Type 3 (LP3) treatment as described in "Guidelines for Determining Flood Flow Frequency" by the Interagency Advisory Committee on Water Data ("Bulletin 17B") was used for this analysis (U.S. Water Resources Council 1982). All equations below are from Bulletin 17B. The following equation was used to calculate the base 10 logarithms for upper Halawa streamflows (Q) at the selected recurrence intervals (the LP3 distribution).

$$Log(Q) = \bar{X} + K * S$$

where K is a factor based on a given recurrence interval and a weighted skew coefficient (G_W), $\bar{X}$ is the mean of the base 10 logarithms of the sample data, and S is the standard deviation of the same.

S and $\bar{X}$ are defined by the following equations.

$$\bar{X} = \frac{\sum X}{N}$$

$$S = \sqrt{\frac{\sum (X - \bar{X})^2}{N - 1}}$$

where X is a base 10 logarithm of the sample data, and N is the number of annual discharge records.

K values for selected recurrence intervals were interpolated from a table in Appendix 3 of Bulletin 17B, after the weighted skew coefficient, G_W, was determined. G_W is a weighted combination of a station skew coefficient (G), and a generalized skew coefficient ($\overline{G}$), calculated to reduce the sensitivity in the analysis towards extreme events. For that purpose, Bulletin 17B recommends using G_W, instead of G alone, especially when the analysis consists of fewer records, *e.g.*, the suggested 10-year minimum. Below is the equation used to calculate G_W.

$$G_W = \frac{MSE_{\bar{G}}(G) + MSE_G(\bar{G})}{MSE_{\bar{G}} + MSE_G}$$

where G is the station skew coefficient, calculated from the sample data, $\overline{G}$ is the generalized skew coefficient, obtained from a skew-isoline map of annual maximum streamflow logarithms ("Plate I" of Bulletin 17B), MSE_G is the mean-square error of the station skew, and $MSE_{\overline{G}}$ is the mean-square error of the generalized skew. Here, because the generalized skew was determined from Plate I, $MSE_{\overline{G}} = 0.302$ was used, as directed by Bulletin 17B.

G and MSE_G, are defined by the equations below.

$$G = \frac{N * \sum(X - \bar{X})^3}{(N-1)(N-2) * S^3}$$

$$MSE_G = 10^{(A - B(Log_{10}(\frac{N}{10}))}$$

where A and B are values dependent upon the absolute value of G. Here, because the absolute value of G is less than 0.90, the following equations were used:

$$A = -0.3 + 0.08|G|$$

$$B = 0.92 - 0.26|G|$$

Figure 38 is Plate I from Bulletin 17B, used to obtain $\overline{G}$.

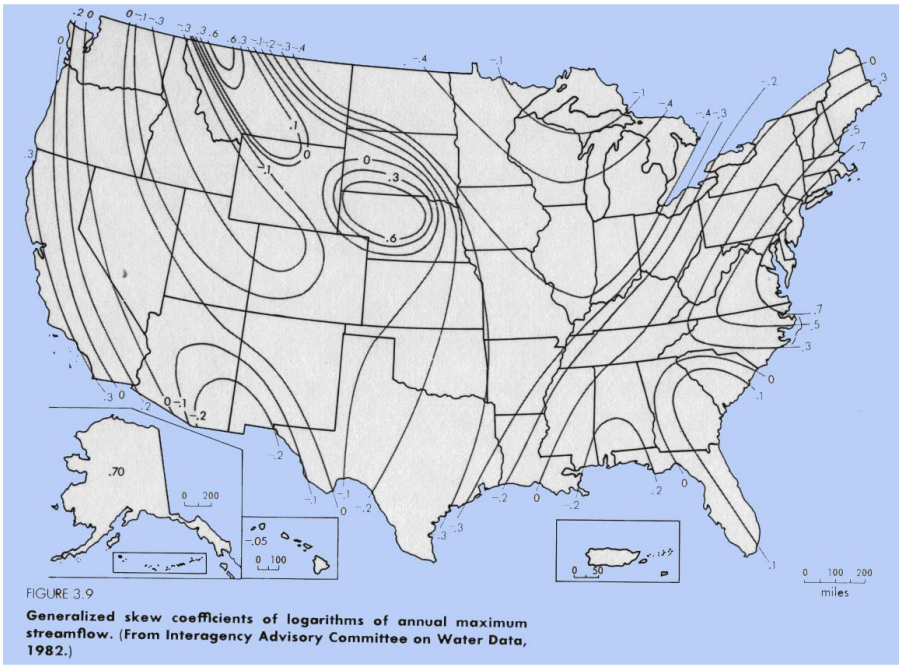

Figure 38: Generalized Skew Coefficients of Logarithms of Annual Maximum Streamflow

Before the statistical treatment was completed, outlier equations were used to determine whether any high or low values in the sample data should be excluded from the set. The

equations produce high and low thresholds. If data from the sample is higher or lower than the respective thresholds, then those data may depart too significantly from the trend, and may need to be removed from the analysis. The outlier equations are displayed below:

$$X_H = \bar{X} + K_N * S$$

$$X_L = \bar{X} - K_N * S$$

where X_H is the high outlier threshold, X_L is the low outlier threshold, KN is a factor based on N, the number of sample data, and can be obtained from a table in Appendix 4 of Bulletin 17B. S, again, is the standard deviation of the base 10 logarithms of the sample data.

The high outlier threshold was evaluated first, then the low outlier threshold, because the station skew is higher than 0.4—as specified in Bulletin 17B. No data, however, was removed as a result of these outlier tests.