Estimating suspended sediment concentrations in areas with limited hydrological data using a mixed-effects model

H. Andres Araujo,[1]* Andrew B. Cooper,[1] Marwan A. Hassan[2] and Jeremy Venditti[3]

[1] oo lo o d E mo l m m o F , B , B Gol , C d

[2] D m o o , o B Gol , o , B Gol , C d

[3] D m o o m o F , B , B Gol , C d

Abstract:

Sediment rating curves are commonly used to estimate the suspended sediment load in rivers and streams under the assumption of a constant relation between discharge (Q) and suspended sediment concentrations (SSC) over time. However, temporal variation in the sediment supply of a watershed results in shifts in this relation by increasing variability and by introducing nonlinearities in the form of hysteresis or a path-dependent relation. In this study, we used a mixed-effects linear model to estimate an average SSC–Q

A large portion of research in hydrology has investigated the relations between discharge and fine sediments, an area of research that continues expanding because predictions of fine sediment transport rate via hydraulically based functional relations are often more than two times different than rates calculated from direct sampling in the fi

Study section). Specific hydrometric stations can also provide an idea of the quality of data they produce. For example, the SEM annual load for Chilliwack River station at Vedder Crossing (station number 08MH001, discussed in the Case Study section) ranged between 16.2% and 43.0% from 1965 to 1976 (Environment Canada, 1992). Other sources of errors may arise from the data format itself. For example, data can also be irregularly spaced, leaving 'blanks' that can bias parameter estimation in the model. Additional errors and biases in the estimation may result from the data not conforming to the assumptions underlying the model being applied. These will be discussed in greater detail in the following sections.

d - m do lmo l o

Before fitting a mixed-effects model, it is advisable to perform an exploratory data analysis to help choose the grouping in the data set (e.g. seasons, months and weeks within the hydrologic cycle) and decide on the potential structure of random effects (e.g. which coefficients in Equation (1) should vary among groups).

In the case study, there are two levels of variation in the data: groups (same month in different years) and observations nested within groups (individual measurement of SSC and *Q* in each month measured in mg l^{-1} and $m^3 s^{-1}$, respectively). In this particular study, we used 12 mixed-effects models, one for each month across all years in the data set. The reason for formulating 12 mixed-effects models was to account for seasonal variability from month to month while acknowledging within-month variability from year to year. In each model corresponding to a month (e.g. mixed-effects model for June), observations between years (for that month) are independent, but observations within year (in that month) are correlated (Pinheiro and Bates, 2000). In this manner, we estimated an average set of coefficients for Equation (1) for each month (the fixed effects) but allow these to vary across years for that month (the random effects). The variability in the coefficients across years for the same month (i.e. June 1969 versus June 1970) may be due to hysteresis patterns, changes in the sediment sources and climatic factors (Figure 1). The software used to fit the mixed-effects models was the *nlme* library in R, version 2.10.2 (Pinheiro *et al.*, 2010).

Although there are several ways to formulate a linear mixed-effects model, we used different equations for the two levels of nested data (Singer, 1996; Snijders and Bosker, 1999; Lai and Helser, 2004; Zuur *et al.*, 2009). The first level describes the variability in SSC as a function of *Q* within each month, and the second level describes the variability in the SSC–*Q* relation among years in the same month (e.g. variability among the Januaries of the data set). The first level is represented by

$$= \beta_0 + \beta_1 \quad + \quad , \quad \sim (0, \sigma^2) \qquad (2)$$

where $= \log$ (SSC $+ 1$), $= \log Q(+ 1)$ for an addressing SSQ– to ensure normality of the residuals. The random error represents the within-month variance, and it is assumed to be independent and identically normally distributed with mean equal to zero and common variance σ^2. The second level is represented by

$$\beta_0 = \beta_0 + \quad_0 , \quad \beta_1 = \beta_1 + \quad_1$$

$$= \quad 0$$

should be used to represent the SSC–Q relation, several tests and measures of goodness of fit can be performed. For example, the likelihood ratio test can be used to compare the fit of nested models. In this case, the test will return a small value if the mixed-effects model is a good choice (Crawley, 2007). An alternative way to decide on what model to use is to compare the model scores of the Akaike information criterion (AIC), which we have included in this analysis, or the Bayesian Information Criterion (BIC). Each of these criteria examine the trade-offs between model complexity and improved goodness of fit for nonnested models (Burnham and Anderson, 1998) with lower values indicating greater parsimony.

CASE STUDY

The model was designed and tested using SSC and Q data from the Chilliwack River. The main reasons for

$m^3\ s^{-1}$, respectively. The largest winter rainfall event had an estimated discharge of 776 $m^3\ s^{-1}$, whereas the largest snowmelt event was 280 $m^3\ s^{-1}$ (Martin and Church, 1995). Despite the high flows of the winter, the highest monthly mean SSC was obtained for the snowmelt events in the spring (Figure 5). This outcome reflects the relative importance of the flood duration on the daily and monthly SSC. In the basin, the snowmelt events are of much longer duration than the rainfall counterpart. Of particular interest is the December 1975, which experienced the highest sediment yield in record (see Church *l*., 1989). Despite the large magnitude of sedimentation events on this month, the sediment concentration remained within the range of results obtained for the rest of the records, mixed-effect model, we were interested in the within

applied to a month (e.g. January) across all years with complete SSC and Q data could be used to extrapolate SSC in the Januaries where only Q data are present. Although we could have split the original data set into calibration and

SSC can minimize the need for intense monitoring and help managers and scientists determine the impact of suspended sediments on aquatic species. For example, in coastal British Columbia, the aquatic species contained in many coastal watersheds are susceptible to elevated SSC levels.

CONCLUSION

This study shows that mixed-effects models can help predict more accurate values of SSC than the standard rating curves. These models do so by directly modeling the intramonth and interyear variability in the SSC–Q relation. In the case study, the predictive power of both the mixed-effects model and the rating curve increased at higher values of SSC and Q, as their relation becomes stronger. For the low-flow summer months, the mixed-effects model proved to be a more valuable alternative than the sediment-rating curve, an important feature when estimating SSC in areas with limited hydrological data. This advantage can equip managers and scientist with better knowledge of suspended sediment and therefore aid addressing its effects on important physical, chemical and biological processes in rivers and streams.

ACKNOWLEDGEMENTS

The authors thank Ashley Page, Duncan Knowler, Erland MacIsaac, Athena Ogden, Carrie Holt and Ken Lertzman for their ideas or contributions to the manuscript and the valuable input of two anonymous reviewers. The Forest Investment Account–Forest Science Program from the Ministry of Forests and Range of British Columbia (FIA-FSP) Project Y092250 (Year 2) and the Natural Sciences and Engineering Research Council of Canada (NSERC) provided the funding for this project.

REFERENCES

Aitkin M, Longford N. 1986. Statistical modeling in school effectiveness studies.

Salant NL, Hassan MA, Alonso CV. 2008. Suspended sediment dynamics at high and low storm flows in two small watersheds.