Fitting sediment rating curves using regression analysis: a case study of Russian Arctic rivers

NIKITA I. TANANAEV

Igarka Geocryology Lab, Bldg. 8A, 1st District, Igarka, Krasnoyarsk Krai, 663200, Russia
nikita.tananaev@gmail.com

Abstract Published suspended sediment data for Arctic rivers is scarce. Suspended sediment rating curves for three medium to large rivers of the Russian Arctic were obtained using various curve-fitting techniques. Due to the biased sampling strategy, the raw datasets do not exhibit log-normal distribution, which restricts the applicability of a log-transformed linear fit. Non-linear (power) model coefficients were estimated using the Levenberg-Marquardt, Nelder-Mead and Hooke-Jeeves algorithms, all of which generally showed close agreement. A non-linear power model employing the Levenberg-Marquardt parameter evaluation algorithm was identified as an optimal statistical solution of the problem. Long-term annual suspended sediment loads estimated using the non-linear power model are, in general, consistent with previously published results.

Key words suspended sediment load; suspended sediment rating curves; Russian Arctic; regression analysis

INTRODUCTION

The sediment rating curve is widely employed as an empirical technique for relating suspended sediment concentrations C (g m^{-3}) with water discharge Q (m^3 s^{-1}) (Colby, 1956). As such, it introduces a causative linkage between two variables, one (Q) treated as an independent predictor (Glysson, 1987). A plethora of previous publications on rating curves expresses this relationship using a power function of the form:

$$C = aQ^b \tag{1}$$

where a and b are the rating coefficient and exponent, respectively (Syvitski *et al.*, 2000). Alternative formulations of the sediment rating curve equation include the use of a power function with a constant (Asselman, 2000) and a simple linear fit (Mount & Abrahart, 2011). Each of these alternatives has its benefits and limitations.

One potential use of the sediment rating curve, as argued by Fenn *et al.* (1985) is for exploring the internal features of coincident sediment and discharge datasets rather than for obtaining a plausible predictive model. Sediment rating curve equations, nonetheless, are used widely in producing suspended load estimates for periods when only water discharge data are available. Fitting a rating curve in this context, is therefore a regression problem, and the plausibility of the solution relies on the performance of the statistical methods involved (Cox *et al.*, 2008). Implicit heterogeneity in the observational datasets poses a challenge for effective rating curve fitting. Depending on the time frame of monitoring, various intra-annual variations in sediment delivery and transport processes can be present in the datasets, including hysteresis and seasonality (Asselman, 2000). Fitting procedures based on discharge classes (Jansson, 1996), seasonal rating curve equations (Khanchoul & Jansson, 2008) and dataset separation by rising and falling limb stages of the hydrograph (Aquino *et al.*, 2009) enhance the performance of rating curves in these circumstances, if such distinctions can be clearly drawn on the basis of the data collected.

The sediment rating curve approach, *sensu stricto*, is only applicable to the analysis of measured values of discharge and concentration, which are frequently treated as daily averages. When continuous sampling at short intervals is performed, time series analysis or artificial neural networks with consideration of autocorrelation may produce the best result (Mount & Abrahart, 2011).

The quality of the rating curve model depends as much on the fitting method, as it does on the datasets collected as well as the nature of the key geomorphic processes responsible for the sediment delivery to the study streams. The principal assumptions of the rating curve concept hold true when runoff is the major driver of sediment source activation, e.g. during snowmelt and rainstorm events. Arctic watersheds underlain by permafrost provide an example of an

environment where geomorphic activity is frequently driven by heat (cryogenic processes). The purpose of the study was therefore to test the applicability of various sediment rating curve fitting techniques to the observational datasets for three medium to large rivers in the Russian Arctic. We assessed the efficiency of the resulting models and their general applicability to datasets where the observations are scarce and reflect the joint action of fluvial and cryogenic processes in generating suspended sediment fluxes.

STUDY SITES AND DATASETS

Gauging sites

This study employed datasets for discharge and suspended sediment concentration from three gauging stations on medium to large rivers of the Russian Arctic: the Anabar River at Saskylakh, the Lena River at Tabaga and the Indigirka River at Vorontsovo (Table 1, Fig. 1).

Table 1 Summary information for the study watersheds and gauging stations.

Gauge	φ (°N)	λ (°E)	A (km^2)	Q (m^3 s^{-1})
Anabar R. at Saskylakh	71.97	114.08	78 800	419
Lena R. at Tabaga	61.83	129.60	897 000	7 140
Indigirka R. at Vorontsovo	69.57	147.53	305 000	1 560

φ, gauge latitude; λ, gauge longitude; A, basin area; Q, mean annual discharge.
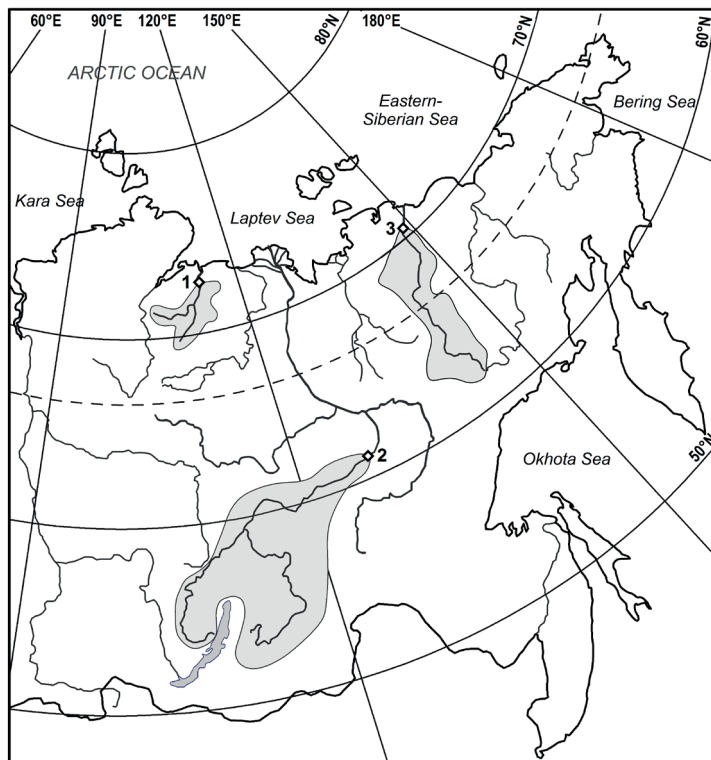

Fig. 1 Locations of the study gauging stations, showing their watersheds: the Anabar River at Saskylakh (1), the Lena River at Tabaga (2), and the Indigirka River at Vorontsovo (3).

Datasets

Table 2 summarizes the datasets used for the study sites.

Table 2 Summary statistics for the study datasets.

Gauge	T (year)	n	Q_T (m^3 s^{-1})	Q_d (m^3 s^{-1})	SSC_d (g m^{-3})
Anabar River at Saskylakh	7	47	469	2 240	39
Lena River at Tabaga	18	194	7 110	15 000	34
Indigirka River at Vorontsovo	17	185	1 590	4 330	213

T, length of the dataset; n, number of observations; Q_T, mean annual discharge for the years included in the dataset; Q_d and SSC_d, mean discharge and suspended sediment concentration for the dataset, respectively.

METHODS

Notwithstanding the opinion that the fitting and use of sediment rating curves is well-documented and standardized (Mount & Abrahart, 2011), there are still ongoing debates on the appropriateness and accuracy of various curve fitting procedures.

This study was not designed to test all existing fitting techniques, but rather to examine the most common procedures:
(a) linear regression on untransformed values (linear fit);
(b) linear regression on log-transformed values (power fit);
(c) non-linear regression (power fit).
The log-transformed power model requires bias correction to account for the inequality of the means of the initial and log-transformed data. A bias correction factor CF was applied to the log-linear power models, as described by Ferguson (1986):

$$CF = \exp(2.65s^2), \tag{2}$$

$$s^2 = \sum_{i=1}^{n} (\log C_i - \log \hat{C}_i)/(n-2) \tag{3}$$

where C_i and $\hat{C}_i$ are observed and predicted values, respectively, and n is the number of observations.

Non-linear regression fitting is regarded as an optimization problem, so the potential solutions can be numerous, depending on the chosen variety of the loss function optimization algorithms. In this study, the performance of the Levenberg-Marquardt, the Simplex (Nelder-Mead) and the Hooke-Jeeves algorithms was tested and compared. The former algorithm was developed for use innon-linear least squares solutions, while the latter two were designed for wider applications in non-linear optimization.

RESULTS

Initial data inspection

Selection of the most accurate fitting technique, as well as assessment of the applicability of any particular regression model, starts with data inspection, which is best performed graphically (Fig. 2). Only the sediment concentration data for the Lena River pass the Kolmogorov-Smirnov test for log-normality. In general, the suspended sediment distributions tend to be more skewed towards the left; the lower left parts of the scatter plots are overpopulated, though the degree of scatter remains low (Fig. 3).

Both histograms (Fig. 2) and scatter plots (Fig. 3) suggested that the overall quality of rating curves could be relatively poor. Threshold behaviour is characteristic of both the Lena and Indigirka River datasets, as the scatter increases significantly when discharges of 15 000 m^3 s^{-1} and 25 000 m^3 s^{-1}, respectively, are exceeded. For the Lena River this threshold value corresponds well with the effective discharge estimate of 16 000 m^3 s^{-1}, responsible for intense bank erosion (Tananaev, 2013). Above this threshold, the variability in suspended sediment concentration is ascribed to the introduction of significant amounts of wash load, originating from both the surrounding river basin and the eroded channel bank material.
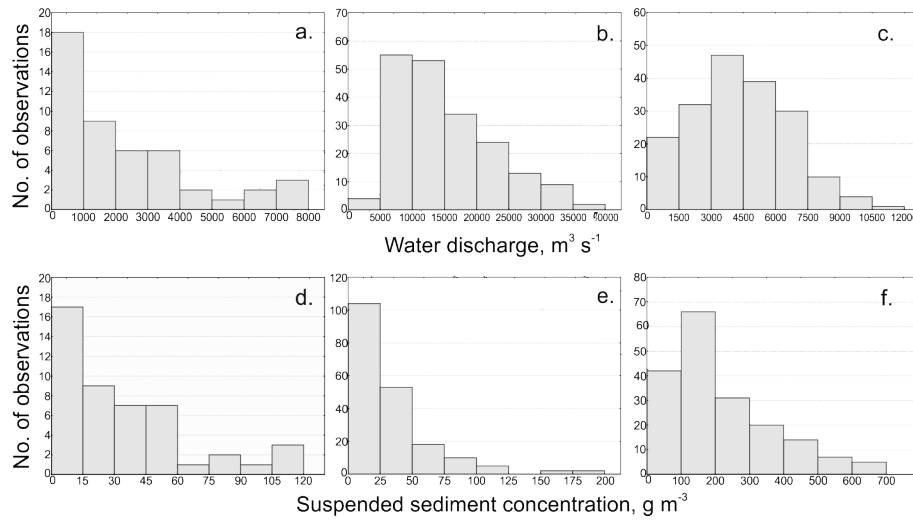
Fig. 2 Frequency distributions of water discharge and suspended sediment concentration for the Anabar (a, d), Lena (b, e) and Indigirka (c, f) rivers, respectively.
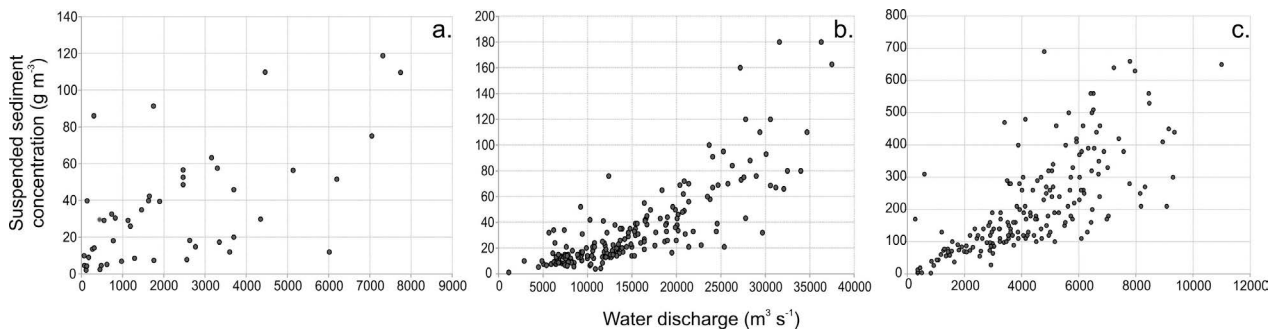

Fig. 3 Scatter plots of suspended sediment concentration *versus* water discharge for the Anabar River at Saskylakh (a), the Lena River at Tabaga (b), and the Indigirka River at Vorontsovo (c).

Table 3 Sediment rating curve equations for the Russian Arctic study rivers.

	Anabar at Saskylakh	Lena at Tabaga	Indigirka at Vorontsovo
Linear fit	$C = 0.01235\,Q$	$C = 0.00249\,Q$	$C = 0.0494\,Q$
Non-linear power fit:			
L-M	*$C = 0.685Q^{0.522}$*	$C = 0.0000028\,Q^{1.68}$	$C = 0.0304\,Q^{1.056}$
S	*$C = 0.683Q^{0.522}$*	$C = 0.0000028\,Q^{1.68}$	$C = 0.0304\,Q^{1.056}$
H-J	*$C = 0.684Q^{0.522}$*	$C = 0.000105\,Q^{1.32}$	$C = 0.0304\,Q^{1.056}$
Non-linear power fit with constant:			
L-M	*$C = 0.000015\,Q^{1.717} + 21.9$*	$C = 1.05\ 10^{-8}Q^{2.211} + 9.6$	*$C = 0.0134Q^{1.141} + 16.8$*
S	*$C = 0.000015\,Q^{1.720} + 21.9$*	$C = 1.05\ 10^{-8}Q^{2.211} + 9.6$	*$C = 0.0134Q^{1.142} + 16.8$*
H-J	*$C = 0.000307\,Q^{1.377} + 19.5$*	$C = 0.00023\,Q^{1.256} - 7.7$	*$C = 0.0141Q^{1.137} + 16.2$*
Log-linear power fit:			
OLS	*$C' = 0.857Q^{0.459}$*	$C' = 0.000118\,Q^{1.288}$	$C' = 0.0249\,Q^{1.067}$
	$CF = 1.448$	$CF = 1.117$	$CF = 1.168$
	$C = 1.241Q^{0.459}$	$C = 0.000132\,Q^{1.288}$	$C = 0.0291\,Q^{1.067}$
PLS	*$C' = 7.079\,Q^{0.172}$*	$C' = 0.021\,Q^{0.744}$	$C' = 1.014\,Q^{0.618}$
	$CF = 1.558$	$CF = 1.166$	$CF = 1.229$
	$C = 11.032Q^{0.172}$	$C = 0.0246\,Q^{0.744}$	$C = 1.246\,Q^{0.618}$

C', suspended sediment concentration, uncorrected for back-transformation bias; statistically insignificant parameters (*p*-value exceeds 0.05) are given in italics.

Regression models

Sediment rating curves were fitted for the study datasets, and the resulting equations are presented in Table 3. Linear fit was obtained using ordinary least squares (OLS). Non-linear models were built both with and without an additive constant using the Levenberg-Marquardt (L-M), Simplex(S) and Hooke-Jeeves (H-J) algorithms. Log-linear power fit used both OLS and partial least squares (PLS). The reason for considering PLS as an alternative to OLS is that the former implicitly covers the 'error-in-variables' issue in the measured data, as all observations are collected with indefinite instrumental errors.

DISCUSSION

Regression model efficiency

The performance of regression models is frequently judged visually (Fig. 4). Visual judgement, however, should not be substituted for the statistical assessment of model efficiency. The Nash-Sutcliffe criterion was employed for this purpose (Table 4). The most efficient models were those employing non-linear power fit with a constant, owing to the fact that an additional parameter is involved (Asselman, 2000), while the simple linear and log-linear PLS models yielded the poorest efficiencies.
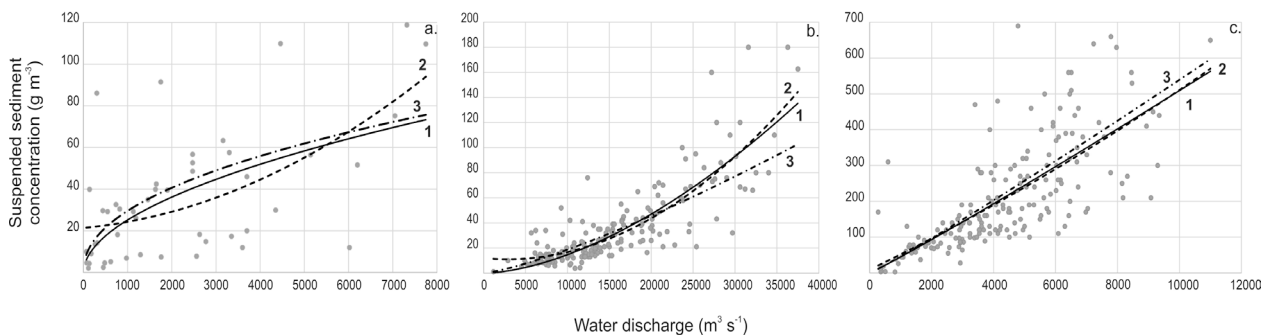

Fig. 4 Sediment rating curves for the Anabar River at (a) Saskylakh, (b) the Lena River at Tabaga and (c) the Indigirka River at Vorontsovo, obtained using non-linear power fit (1), non-linear power fit with additive constant (2) and log-linear fit (3).

Table 4 Nash-Sutcliffe efficiencies of the rating curves.

	Anabar at Saskylakh	Lena at Tabaga	Indigirka at Vorontsovo
Linear fit	0.23	0.60	**0.53**
Non-linear power fit:			
L-M	0.32	0.70	**0.53**
S	0.32	0.70	**0.53**
H-J	0.32	0.68	**0.53**
Non-linear power fit with constant:			
L-M	**0.36**	**0.71**	**0.53**
S	**0.36**	**0.71**	**0.53**
H-J	**0.36**	0.68	**0.53**
Log-linear power fit:			
OLS	0.30	0.66	0.52
PLS	0.20	0.44	0.45

Suspended sediment load estimates

The regression model accuracy cannot be judged directly, as high-resolution empirical observations are absent for the study period. The rating curve equations can, however, be used to produce estimates of the annual suspended sediment loads of the Russian Arctic study rivers (Table 5).

Table 5 Long-term seasonal and annual suspended sediment loads of the Russian Arctic study rivers.

	Anabar at Saskylakh	Lena at Tabaga	Indigirka at Vorontsovo
Spring flood			
Q (m^3 s^{-1})	2 440	22 100	5 020
t (days)	48	74	54
W_R (Mt)	0.41	7.87	5.76
Summer floods			
Q (m^3 s^{-1})	965	16 100	5 390
t (days)	18	20	19
W_R (Mt)	0.04	0.91	2.35
Summer low-flow			
Q (m^3 s^{-1})	94.2	5 980	1 170
t (days)	56	72	57
W_R (Mt)	0.003	0.23	0.30
Annual sediment load (Mt)	0.45	9.00	8.41

Q, mean seasonal discharge; t, season duration; W_R, seasonal suspended sediment load.

In general, the suspended sediment load estimates derived herein (Table 5) are broadly consistent with previously published results. For the Anabar River at Saskylakh, our estimate is close to the 0.4 Mt estimate from the papers of Gordeev *et al.* (1996) and Holmes *et al.* (2002). For the Lena River at Tabaga, our estimate exceeds the value of 7.7 Mt reported by Hasholt *et al.* (2005), and for the Indigirka R. at Vorontsovo, our estimate is lower than the estimates of 12.9 Mt (Gordeev *et al.*, 1996), 12.0 Mt (Hasholt *et al.*, 2005) and 11.1 Mt (Holmes *et al.*, 2002) reported previously.

CONCLUSION

Based on the visual inspection of the suspended sediment rating curves and the Nash-Sutcliffe criterion, a non-linear power model employing the Levenberg-Marquardt parameter evaluation algorithm was identified as an optimal statistical solution of the problem. Long-term annual suspended sediment loads for the study rivers estimated using the non-linear power model are, in general, consistent with those reported previously.

REFERENCES

Aquino, S., Latrubesse, E. & Bayer, M. (2009) Assessment of wash load transport in the Araguaia River (Aruanã gauge station), central Brasil. *Latin American Journal of Sedimentology and Basin Analysis* 16(2), 119–128.

Asselman, N.E.M. (2000) Fitting and interpretation of sediment rating curves. *J. Hydrol.* 234, 228–248.

Colby, B.R. (1956) Relationship of sediment discharge to streamflow. *US Geol. Survey Open File Report 56-27.*

Cox, N.J., *et al.* (2008) Fitting concentration and load rating curves with generalized linear models. *Earth Surf. Processes Landf.* 33, 25–39.

Fenn, C.R., Gurnell, A.M. & Beecroft, I.R. (1985) An evaluation of the use of suspended sediment rating curves for the prediction of suspended sediment concentration in a proglacial stream. *GeografiskaAnnaler. Series A* 67, 71–82.

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

Glysson, G.D. (1987) Sediment-transport curves. *US Geol. Survey Open File Report 87-218.*

Gordeev, V.V., *et al.* (1996) A reassessment of the Eurasian river input of water, sediment, major elements, and nutrients to the Arctic Ocean. *American Journal of Science* 296, 664–691.

Hasholt, B., *et al.* (2005) Sediment transport to the Arctic Ocean and adjoining cold seas. *15th Int. Northern Research Basins Symp. and Workshop, Luleå to Kvikkjokk*, 41–67.

Holmes, R.M., *et al.* (2002) A circumpolar perspective on fluvial sediment flux to the Arctic Ocean. *Global Biogeochemical Cycles* 16, 1098, doi: 10.1029/2001GB001849.

Jansson, M.B. (1996) Estimating a sediment rating curve of the Reventazón River at Palomo using logged mean loads with discharge classes. *J. Hydrol.* 183, 227–241.

Khanchoul, K. &Jansson, M.B. (2008) Sediment rating curves developed on stage and seasonal means in discharge classes for the Mellahwadi, Algeria. *GeografiskaAnnaler, Series A* 90, 227–236.

Mount, N.J. & Abrahart R.J. (2011) Load or concentration, logged or unlogged? Addressing ten years of uncertainty in neural network suspended sediment prediction. *Hydrol. Processes* 25, 3144–3157.

Syvitski, J.P.M., *et al.* (2000) Estimating fluvial sediment transport: The rating parameters. *Water Resour. Res.* 36, 2747–2760.

Tananaev, N.I. (2013) Hydrological and geocryological controls on fluvial activity of rivers in cold environments. In: *Cold and Mountain Region Hydrologic Systems under Climate Change* (Gelfan, A. *et al.*, eds). IAHS Publ. 360, 161–167.