

# A Spatially Referenced Regression Model (SPARROW) for Suspended Sediment in Streams of the Conterminous U.S.

Gregory E. Schwarz, Richard A. Smith, Richard B. Alexander, and John R. Gray, U.S. Geological Survey, Reston, Virginia

Mailing address: U.S. Geological Survey, 413 National Center, 12201 Sunrise Valley Dr., Reston, VA 20192, Phone (703) 648-5718, Fax (703) 648-5295, Email [gschwarz@usgs.gov](mailto:gschwarz@usgs.gov).

## INTRODUCTION

Suspended sediment has long been recognized as an important contaminant affecting water resources. Besides its direct role in determining water clarity, bridge scour and reservoir storage, sediment serves as a vehicle for the transport of many binding contaminants, including nutrients, trace metals, semi-volatile organic compounds, and numerous pesticides (U.S. Environmental Protection Agency, 2000a). Recent efforts to address water-quality concerns through the Total Maximum Daily Load (TMDL) process have identified sediment as the single most prevalent cause of impairment in the Nation's streams and rivers (U.S. Environmental Protection Agency, 2000b). Moreover, sediment has been identified as a medium for the transport and sequestration of organic carbon, playing a potentially important role in understanding sources and sinks in the global carbon budget (Stallard, 1998).

A comprehensive understanding of sediment fate and transport is considered essential to the design and implementation of effective plans for sediment management (Osterkamp and others, 1998, U.S. General Accounting Office, 1990). An extensive literature addressing the problem of quantifying sediment transport has produced a number of methods for estimating its flux (see Cohn, 1995, and Robertson and Roerish, 1999, for useful surveys). The accuracy of these methods is compromised by uncertainty in the concentration measurements and by the highly episodic nature of sediment movement, particularly when the methods are applied to smaller basins. However, for annual or decadal flux estimates, the methods are generally reliable if calibrated with extended periods of data (Robertson and Roerish, 1999). A substantial literature also supports the Universal Soil Loss Equation (USLE) (Soil Conservation Service, 1983), an engineering method for estimating sheet and rill erosion, although the empirical credentials of the USLE have recently been questioned (Trimble and Crosson, 2000). Conversely, relatively little direct evidence is available concerning the fate of sediment. The common practice of quantifying sediment fate with a sediment delivery ratio, estimated from a simple empirical relation with upstream basin area, does not articulate the relative importance of individual storage sites within a basin (Wolman, 1977). Rates of sediment deposition in reservoirs and flood plains can be determined from empirical measurements, but only a limited number of sites have been monitored, and net rates of deposition or loss from other potential sinks and sources is largely unknown (Stallard, 1998). In particular, little is known about how much sediment loss from fields ultimately makes its way to stream channels, and how much sediment is subsequently stored in or lost from the streambed (Meade and Parker, 1985, Trimble and Crosson, 2000).

This paper reports on recent progress made to address empirically the question of sediment fate and transport on a national scale. The model presented here is based on the SPAtially Referenced

Regression On Watershed attributes (SPARROW) methodology, first used to estimate the distribution of nutrients in streams and rivers of the United States, and subsequently shown to describe land and stream processes affecting the delivery of nutrients (Smith and others, 1997, Alexander and others, 2000, Preston and Brakebill, 1999). The model makes use of numerous spatial datasets, available at the national level, to explain long-term sediment water-quality conditions in major streams and rivers throughout the United States. Sediment sources are identified using sediment erosion rates from the National Resources Inventory (NRI) (Natural Resources Conservation Service, 2000) and apportioned over the landscape according to 30-meter resolution land-use information from the National Land Cover Data set (NLCD) (U.S. Geological Survey, 2000a). More than 76,000 reservoirs from the National Inventory of Dams (NID) (U.S. Army Corps of Engineers, 1996) are identified as potential sediment sinks. Other, non-anthropogenic sources and sinks are identified using soil information from the State Soil Survey Geographic (STATSGO) data base (Schwarz and Alexander, 1995) and spatial coverages representing surficial rock type and vegetative cover. The SPARROW model empirically relates these diverse spatial datasets to estimates of long-term, mean annual sediment flux computed from concentration and flow measurements collected over the period 1985-95 from more than 400 monitoring stations maintained by the National Stream Quality Accounting Network (Alexander and others, 1998), the National Water Quality Assessment Program, and U.S. Geological Survey District offices (Turcios and Gray, in press). The calibrated model is used to estimate sediment flux for over 60,000 stream segments included in the River Reach File 1 (RF1) stream network (Alexander and others, 1999).

SPARROW uses statistical methods to calibrate a simple, structural model of riverine water quality, one that imposes mass balance in accounting for changes in contaminant flux. As applied here, the mass-balance approach facilitates the interpretation of model results in terms of physical processes affecting sediment transport, and makes possible the estimation of various rates of sediment generation and loss associated with stream channels and features of the landscape. The statistical approach provides a basis for assessing the error of these inferred rates and of the error in extrapolated estimates of sediment flux made for streams in the RF1 network.

An important implication of the holistic modeling approach adopted in this analysis is that estimates of sediment production and loss are based on, and therefore consistent with, measurements of in-stream flux. Other ancillary information, such as direct measurements of long-term sediment storage and release from reservoirs (Steffen, 1996), is incorporated into the analysis by specifying additional equations explaining these ancillary variables. The imposition of cross-equation constraints affords this information a statistically consistent weight in explaining in-stream sediment flux. Thus, the methodology described here represents a general framework for synthesizing a wide spectrum of available information relevant to the understanding of sediment fate and transport.

## METHODOLOGY

The SPARROW methodology (Smith and others, 1997) has been modified to incorporate greater spatial resolution. The primary spatial reference frame for the model continues to be the RF1 reach network: all point sources and landscape features are referenced to a particular RF1 reach. However, considerable internal structure has been added to each reach. Reach watersheds are

delineated using the 1-kilometer HYDRO 1K digital elevation model (DEM) (U.S. Geological Survey, 2000a), and explicit pathways are defined between landscape features and their adjacent RF1 streams. The delineation method uses a “burn-in” process whereby the RF1 reach is first digitized in the 1-kilometer grid and then the elevations of RF1 grid cells are artificially lowered to insure that simulated flow from surrounding cells moves into them. Flow directions based on the steepest descent determine the extent of the reach watershed and the undefined tributary flow paths leading from the landscape to the RF1 channel cells. To insure the accurate determination of in-stream travel time, RF1 stream pathways continue to be defined by the line work of RF1 channels rather than by the grid-cell representation.

A schematic of a typical reach watershed, illustrating its spatial structure and associated features, is given in figure 1. Flow directions, represented by the arrows crossing each adjacent grid cell, define the movement of water in undefined tributaries leading to the RF1 stream. The “burn-in” method insures that all flow paths intersect a reach cell at some point within the watershed, although inconsistencies between the RF1 reach and the DEM-defined stream channel may artificially lengthen “off-RF1” flow paths and shorten “on-RF1” paths (see figure 1 for an example). The length of the flow path provides a rough estimate of the distance sediment must travel in smaller tributaries before reaching the larger streams included in the RF1 network. Travel time in small streams versus large rivers has been shown to be an important factor affecting the in-stream delivery of nutrients (Alexander and others, 2000) and could be of similar importance for sediment.

The enhanced spatial structure afforded by the DEM facilitates the incorporation of spatially integrating features into the model. “Off-channel” reservoirs, located on the grid net according to their geographic coordinates provided by the NID, act as potential sinks for sediment emanating from cells with flow paths that intersect the reservoir grid cell. Similarly, “off-RF1” monitoring stations can be located on the grid and given a basin representation. Although these stations are not useful for calibrating the delivery process within RF1 channels, they offer a high-resolution view of other processes affecting the movement of sediment across the landscape.

Other important spatial features identified in the model include point sources, located relative to RF1 streams based on geographic coordinates (S. Rubin, Environmental Protection Agency, written commun., 1999), and land associated with uses that serve as likely sources or sinks for sediment. Point-source loadings of total suspended solids are determined by methods developed by the National Oceanic and Atmospheric Administration for the National Coastal Pollutant Discharge Inventory (National Oceanic and Atmospheric Administration, 1993). Land use is taken from the 21-class, 30-meter resolution NLCD, and summarized according to the number of 30-meter cells of a given land-use class that are mapped to a corresponding 1-kilometer cell. NLCD land use is used to refine the areal extent of the various sediment erosion rates associated with different land covers identified in the NRI.

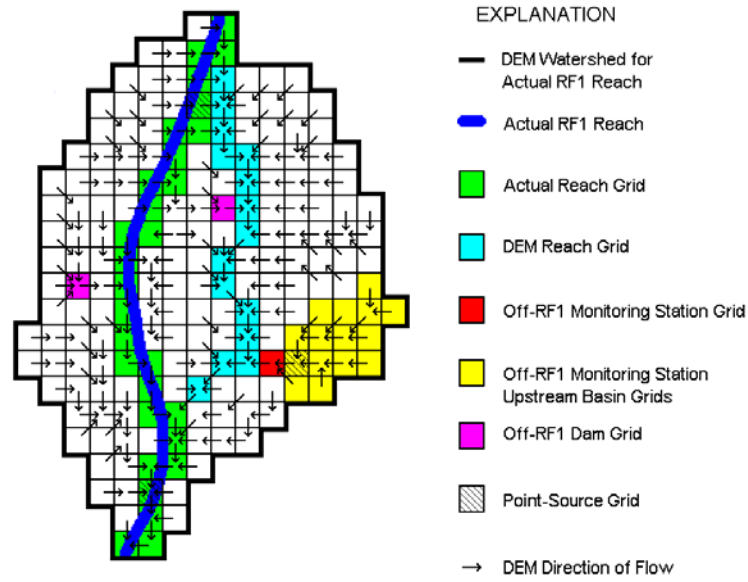


Figure 1. Schematic of a typical reach watershed illustrating the grid cell structure and identified attributes.

The mean annual suspended-sediment flux generated within and leaving reach watershed  $j$ , referred to as the incremental reach flux  $F_j$ , can be expressed as

$$(1) \quad F_j = \sum_{c=1}^{N_j} e^{-\delta \cdot \mathbf{d}_{c,j} + \alpha \cdot \mathbf{Z}_{c,j}} \boldsymbol{\beta} \mathbf{S}_{c,j},$$

where  $N_j$  is the number of 1-kilometer grid cells, indexed by  $c$ , in reach watershed  $j$ ,  $\mathbf{d}_{c,j}$  is a vector of factors describing the pathway from cell  $c$  to the outlet of reach  $j$ ,  $\boldsymbol{\delta}$  is a vector of coefficients associated with the pathway variables,  $\mathbf{Z}_{c,j}$  is a vector of landscape and climatic characteristics affecting the delivery of sediment within cell  $c$ ,  $\boldsymbol{\alpha}$  is a vector of coefficients associated with the  $\mathbf{Z}$  variables,  $\mathbf{S}_{c,j}$  is a vector of sediment sources, and  $\boldsymbol{\beta}$  is a vector of associated source coefficients.

The vector  $\mathbf{d}$  consists of (1) variables representing the landscape flow-path distance traversed to reach the RF1 stream, (2) the mean slope of the “off-RF1” flow path, the time of travel incurred along the RF1 stream, (3) variables affecting the retention of sediment in any reservoir located along the landscape or RF1 flow path, such as streamflow, reservoir age, and NID estimates of surface area or storage volume, and (5) other variables identifying possible sinks along the flow path, such as forested land or land classified by STATSGO as wetlands or alluvium. Variables included in the  $\mathbf{Z}$  vector include runoff, overland flow, slope, and indicators of soils or other factors affecting the movement of sediment off the field to channels. The source vector,  $\mathbf{S}$ , includes sediment erosion from the NRI and point-source loadings.

The 1-kilometer spatial detail used to determine  $F_j$ , corresponding to nearly 8 million grid cells for the more than 60,000 reaches in the conterminous U.S., places a heavy computational burden

on the iterative, non-linear least-squares, calibration method. To reduce the number of computations, the reach model is simplified by assuming the  $\mathbf{Z}$  variables take a single mean value  $\bar{\mathbf{Z}}_j$  for all cells in the reach and, for the  $\mathbf{d}$  variables, by substituting a second-order Taylor approximation about the reach-level mean  $\bar{\mathbf{d}}_j$ . The imposition of a common  $\bar{\mathbf{Z}}_j$  value for all cells in a reach is not restrictive given the spatial coarseness of existing information. The resulting approximation is

$$(2) \quad F_j \approx e^{-\delta' \bar{\mathbf{d}}_j + \alpha' \bar{\mathbf{Z}}_j} \sum_{c=1}^{N_j} \left\{ (1 - \delta' (\mathbf{d}_{c,j} - \bar{\mathbf{d}}_j)) \mathbf{S}'_{c,j} \boldsymbol{\beta} + (\boldsymbol{\beta} \otimes \boldsymbol{\delta})' \left( \mathbf{S}_{c,j} \otimes (\mathbf{d}_{c,j} - \bar{\mathbf{d}}_j) (\mathbf{d}_{c,j} - \bar{\mathbf{d}}_j)' \right) \boldsymbol{\delta} \right\}.$$

This approximation effectively converts the unit of observation in equation (1) from a 1-kilometer grid cell to a reach segment, replacing the non-linear terms dependent on individual cell values with non-linear and linear terms dependent on reach-level means, variances and covariances of the  $\mathbf{d}$  and  $\mathbf{S}$  variables.

To complete the model structure, individual reaches are combined to form a nested basin. Each nested basin  $i$  consists of the set  $J(i)$  of reaches upstream from monitoring station  $i$  and below any monitoring station located further upstream (if such stations exist) (see figure 2). The sediment load for nested basin  $i$ , denoted  $L_i$ , is equal to the sum of the incremental fluxes from the nested reach segments  $j \in J(i)$ , plus the monitored sediment discharged from the set  $U(i)$  of nested basins bounding the upper drainage of nested basin  $i$  (there may be more than one) and delivered to monitoring station  $i$ . The sediment load  $L_i$  is related to the upstream incremental fluxes,  $F_j$ , and monitored loads,  $L_u$ , according to a log-linear relation

$$(3) \quad \ln(L_i) = \ln \left( \sum_{j \in J(i)} e^{-\delta' \mathbf{d}_{j,i}} F_j + \sum_{u \in U(i)} e^{-\delta' \mathbf{d}_{u,i}} L_u \right) + \varepsilon_i,$$

where  $\mathbf{d}_{j,i}$  represents a vector consisting of the same variables in  $\mathbf{d}_{c,j}$ , but corresponding to the RF1-reach path extending from the downstream-end of reach  $j$  to the  $i^{\text{th}}$  monitoring station (accordingly,  $\mathbf{d}_{j,i}$  has values of 0 for all variables pertaining to “off-RF1” flow paths). In equation (3), an independent error term  $\varepsilon_i$  has been added to represent the combined effect of measurement and model error introduced at nested basin  $i$ .

Data on reservoir storage can be incorporated directly into the model by introducing an additional storage equation. Let  $\mathbf{d}^*$  and  $\boldsymbol{\delta}^*$  pertain to the subset of path variables and associated coefficients determining the rate sediment is stored in reservoirs, and define  $R_k$  as the annual amount of stored sediment measured at a reservoir on reach  $k$  (a similar analysis can be done for “off-RF1” reservoirs). The reservoir storage equation takes the form

$$(4) \quad \ln(R_k) = \ln \left( \sum_{j \in J(k)} \left( e^{\delta^{*'} \mathbf{d}_{j,k}^*} - 1 \right) F_j + \sum_{u \in U(k)} \left( e^{\delta^{*'} \mathbf{d}_{u,k}^*} - 1 \right) L_u \right) + w_k,$$

where  $w_k$  is a random error.

Joint estimation of equations (3) and (4), with the  $F_j$  and corresponding  $\alpha$ ,  $\beta$ , and  $\delta$  parameters defined by equation (2), is by non-linear three-stage least squares. To insure robust estimates and to facilitate the estimation of prediction error, the calibration of the model is repeated 200 times employing a bootstrap estimation algorithm (see Smith and others, 1997).

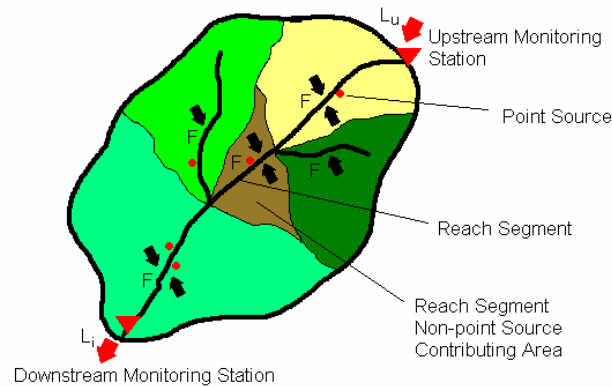


Figure 2. Schematic of a nested basin defined by upstream and downstream monitoring stations.

The flexible mathematical structure used in equations (1) - (3) is capable of accommodating a number of hypotheses concerning sediment fate and transport. Sites of sediment storage, identified in the model as a subset of the  $\mathbf{d}$  variables, can act as sediment sources or sinks, depending on the sign of corresponding  $\delta$  coefficients. A random coefficient form of the model allows storage sites to serve as sources in some regions and sinks in others. Such behavior can be inferred statistically by relating the prevalence of storage sites in nested basins to the magnitude of the squared residual  $\varepsilon$  in these basins (Godfrey, 1988). Non-point sources of sediment, such as soil erosion included under  $\mathbf{S}$ , are distinguished from sediment losses from storage (e.g., an alluvial plain) identified with  $\mathbf{d}$ , on the assumption that the former is a primary process due to weathering whereas the latter is a consequence of the accumulation of previously weathered material which is later released to streams under changing hydraulic conditions. Accordingly, the potential for storage loss in the model depends on the extent of accumulated upstream soil erosion due to weathering. The empirical validity of the USLE estimate of soil erosion can be evaluated through statistical hypothesis tests conducted on the relevant  $\beta$  coefficients. Alternative measures of soil erosion can also be empirically evaluated in the model by substituting variables serving as determinants of the USLE for the USLE erosion estimate.

The estimation of long-term suspended-sediment load at a monitoring station is based on the regression of the natural logarithm of instantaneous suspended-sediment concentration on current and lagged values of the natural logarithm of daily flow and other variables representing seasonal and trend effects. If the station has concentration data collected more frequently than a weekly basis, the regression model is modified to account for serial correlation. To be included in the analysis, a station must have at least 3 years of data between 1985 and 1995. Only data within the period 1985-95 are included in the regression.

Mean-annual suspended-sediment load is estimated by first simulating load for each day over the 1985-95 period and then averaging daily values on an annual basis. Simulated loads are obtained by taking the exponential of the sum of the predicted daily load given by the calibrated regression model with the time trend variable set to a base year of 1992 and a randomly selected residual from the regression model. For days having actual monitoring data, the daily load is computed by multiplying the measured instantaneous concentration by the daily flow. If a station has a data record with sufficient frequency to estimate a serial correlation parameter, the simulated daily load is based on the conditional prediction associated with past and future observed loads, plus a normally distributed random error having a correlation structure consistent with the conditional prediction and with the variance estimated by the regression model. The Monte Carlo process used to estimate simulated daily loads for the 1985-95 period is repeated 200 times, providing 200 values for estimating the mean and standard deviation of the average annual sediment load for a site.

### SUMMARY

The model described here is intended to empirically evaluate regional-scale processes affecting the long-term (i.e., decadal) transport of sediment in rivers. Additionally, the model will provide estimates of sediment mean annual flux for every reach included in the RF1 network. Error estimates for these process evaluations and stream predictions are determined using robust bootstrap methods. Future work will address the dynamic behavior of sediment flux associated with non-steady state streamflow conditions.

### REFERENCES CITED

- Alexander, R.B., Slack, J.R., Ludtke, A.S., Fitzgerald, K.K., and Schertz, T.L., 1998, Data from selected U.S. Geological Survey national stream water-quality monitoring networks, *Water Resources Research*, 34(9), pp. 2401-2405.
- Alexander, R.B., Brakebill, J.W., Brew, R.E., and Richard A. Smith, 1999, *ERF1—Enhanced River Reach File 1.2* (U.S. Geological Survey Open-File Report 99-457, Reston, Virginia).
- Alexander, R.B., Smith, R.A., and Schwarz, G.E., 2000, Effect of channel size on the delivery of nitrogen to the Gulf of Mexico, *Nature*, 403, pp. 758-761.
- Cohn, T. A., 1995, Recent advances in statistical methods for the estimation of sediment and nutrient transport in rivers, U.S. National Report to International Union of Geodesy and Geophysics, 1991-1994, *Review of Geophysics*, 33 Supplemental, available on the World Wide Web at URL <http://webserver.cr.usgs.gov/sediment/> accessed on 11/13/00.
- Godfrey, L.G., 1988, *Misspecification Tests in Econometrics*, Cambridge, 252 p.
- Meade, R. H., and Parker, R. S., 1985, Sediment in rivers of the United States, in *National Water Summary 1984*, U.S. Geological Survey Water-Supply Paper 2275, U.S. Government Printing Office, pp. 49-60.

- National Oceanic and Atmospheric Administration, 1993, *The National Coastal Pollutant Discharge Inventory, Point Source Methods Document*, Pollution Sources Characterization Branch, Strategic Environmental Assessments Division, Office of Ocean Resources Conservation and Assessment, National Ocean Service, 247 p.
- Natural Resources Conservation Service, 2000, *National Resources Inventory, 1992*, U.S. Department of Agriculture, data available on the World Wide Web at URL <http://www.nhq.nrcs.usda.gov/NRI/>, accessed on 11/15/00.
- Osterkamp, W. R., Heilman, P., and Lane, L. J., 1998, Economic considerations of a continental sediment monitoring program, *International Journal of Sediment Research*, 13(4), pp. 12-24.
- Preston, S.D. and Brakebill, J.W., 1999, Application of spatially referenced regression modeling for the evaluation of total nitrogen loading in the Chesapeake Bay watershed, *U.S. Geological Survey Water Resources Investigations Report 99-4054*, 12 p.
- Robertson, D. M. and Roerish, E. D., 1999, Influence of various water quality sampling strategies on load estimates for small streams, *Water Resources Research*, 35(12), pp. 3747-3759.
- Schwarz, G.E. and Alexander, R.B., 1995, *State Soil Geographic (STATSGO) Data Base for the Conterminous United States*, U.S. Geological Survey Open-File Report 95-449, Reston, Virginia.
- Smith, R.A., Schwarz, G.E., and Alexander, R.B., 1997, Regional interpretation of water-quality monitoring data, *Water Resources Research*, 33(12), pp. 2781-2798.
- Soil Conservation Service, 1983, *National Engineering Handbook*, Section 3 Sedimentation, U.S. Department of Agriculture.
- Stallard, R. F., 1998, Terrestrial sedimentation and the carbon cycle, *Global Biogeochemical Cycles*, 12(2), pp. 231-257.
- Steffen, L. J., 1996, A reservoir sedimentation survey information system – RESIS, in *Proceedings of the Sixth Federal Interagency Sedimentation Conference*, March 10-14, 1996, Las Vegas, NV: Sponsored by the Subcommittee on Sedimentation, Interagency Advisory Committee on Water Data, p. 29-37.
- Trimble, S. W. and Crosson, P., 2000, U.S. Soil Erosion Rates – Myth and Reality, *Science*, 289, pp. 248-250.
- Turcios, L.M., and Gray, J.R., in press, U.S. Geological Survey Sediment and Ancillary Data on the World Wide Web: Proceedings of the 7th Federal Interagency Sedimentation Conference, March 25-28, 2001, Reno, Nevada, 6 p.



- U.S. Army Corps of Engineers, 1996. National Inventory of Dams, data available on the World Wide Web at URL <http://crunch.tec.army.mil/> accessed 11/13/00.
- U.S. Environmental Protection Agency, 2000a, National water quality inventory, 1998 report to Congress, available on the World Wide Web at URL <http://www.epa.gov/305b/98report/>, accessed 9/15/00.
- U.S. Environmental Protection Agency, 2000b, The quality of our nation's water, 1998: EPA841-S-00-001, available on the World Wide Web at URL <http://www.epa.gov/305b/98report/98brochure.pdf>, accessed 9/27/00.
- U.S. General Accounting Office, 1990, Water pollution – Greater EPA leadership needed to reduce non-point source pollution, *U.S. General Accounting Office Report GAO/RCED-91-10*, 56 pp.
- U.S. Geological Survey, 2000a, Hydro 1k elevation derivative database, data available on the World Wide Web at URL <http://edcdaac.usgs.gov/gtopo30/hydro/namerica.html> accessed 11/13/00.
- U.S. Geological Survey, 2000b, National Land Cover Data, data available on the World Wide Web at URL <http://edcwww.cr.usgs.gov/programs/lccp/natlndcover.html>, accessed 11/13/00.
- Wolman, M. G., 1977, Changing Needs and Opportunities in the Sediment Field, *Water Resources Research*, 13(1), pp. 50-54.