10.1071/MF24082

Marine and Freshwater Research

Supplementary Material

Estimation of soil loss and sediment yield by using the modified RUSLE model in the Indus River basin, including the quantification of error and uncertainty in remote-sensing images

Muhammad Waseem Boota[A,B,C,D], *Shan-e-hyder Soomro*[E,*], *Haoming Xia*[A,B,C,D], *Yaochen Qin*[A,C,D], *Syed Shahid Azeem*[F], *Chaode Yan*[G], *Weiran Luo*[A], *Ayesha Yousaf*[H], *and Muhammad Azeem Boota*[I]

[A] College of Geography and Environmental Science, Henan University, Kaifeng, 475004, PR China. Email: engr.waseemboota@gmail.com; xiahm@vip.henu.edu.cn; qinyc@henu.edu.cn; luowr2012@henu.edu.cn

[B] Henan Key Laboratory of Earth System Observation and Modeling, Henan University, Kaifeng, 475004, PR China.

[C] Key Laboratory of Geospatial Technology for the Middle and Lower Yellow River Regions (Henan University), Ministry of Education, Kaifeng, 475004, PR China.

[D] Key Research Institute of Yellow River Civilization and Sustainable Development and Collaborative Innovation Center on Yellow River Civilization Jointly Built by Henan Province and Ministry of Education, Henan University, Kaifeng, 475004, PR China.

[E] College of Hydraulic and Environmental Engineering, China Three Gorges University, Yichang, 443002, PR China.

[F] MM Pakistan Pty Ltd, Lahore, Pakistan. Email: shahid.csaap@gmail.com

[G] School of Water Conservancy and Transportation, Zhengzhou University, Zhengzhou, 450001, PR China. Email: ycd@zzu.edu.cn

[H] College of Mechanical and Electrical Engineering, Henan University of Technology, Zhengzhou, 450052, PR China. Email: engr.ayeshawaseem@qq.com

[I] Barani Agricultural Research Institute, Chakwal, Pakistan. Email: azeemnazir633@gmail.com

[*]Correspondence to: Shan-e-hyder Soomro College of Hydraulic and Environmental Engineering, China Three Gorges University, Yichang, 443002, PR China Email: shanhydersoomro110@hotmail.com

Quantification of uncertainty using remote sensing images

In the present investigation, an extensive evaluation of errors and uncertainties was performed to ascertain the validity of the sediment yield estimations obtained from remote sensing data. Comprehending and quantifying errors and uncertainties are paramount in remote sensing research, primarily when the acquired data is employed to guide environmental management and policy formulation. Since remote sensing imagery is influenced by multiple sources of error, including sensor inaccuracies, atmospheric variables, and constraints related to spatial resolution, it is imperative to systematically assess these elements to ascertain the robustness of the study's conclusions. To quantify the errors and uncertainties inherent in the data derived from remote sensing methodologies, we utilised a variety of statistical and analytical approaches. The principal method entailed juxtaposing the river widths derived from remote sensing techniques with the *in situ* measurements collected during field surveys. This comparative analysis was conducted across various cross-sectional evaluations of the LIRB to evaluate the precision of the remote sensing data. The root mean square error (RMSE) was computed to assess the mean deviation between the empirically observed (*in situ*) and the forecasted (remote sensing derived) values. The RMSE indicates the standard deviation of the residuals, thereby elucidating the degree to which the remote sensing data aligns with the actual empirical measurements. This investigation determined the RMSE to be 115.4 m, a value within an acceptable threshold for extensive sediment yield research endeavours. Furthermore, the coefficient of determination (R^2) was utilised to assess the robustness of the correlation between the empirical and forecasted values. The derived value of R^2 was ~0.8665, which signifies a robust association between the remote-sensing datasets and the *in situ* observational metrics. This substantial correlation degree implies that the remote sensing data can be deemed dependable for estimating sediment yield within the LIRB (Peña-Arancibia *et al.* 2013).

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(P_i - O_i)^2} \qquad \text{(S1)}$$

where Pi is the predicted value for the ith observation in the dataset, O_i is the observed value for the ith observation in the dataset, and N is the sample size. To enhance the rigor of the analysis, we employed various statistical techniques to address non-detections (*NoD*s) and incomplete datasets, which frequently pose significant challenges in remote sensing research. Maximum Likelihood Estimation (*MLE*) (Pan *et al.* 2002) was employed to derive parameter estimates in *NoDs* by optimising the likelihood function (Eqn S2). This approach is especially advantageous when addressing datasets with a substantial fraction of absent or censored observations. In our analytical framework, *MLE* yielded the most favourable results when confronted with elevated proportions (19–39%) of *NoD*s (White *et al.* 2023).

$$L(\vartheta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{(X_i - \vartheta_o)^2}{2\vartheta_1}} \qquad \text{(S2)}$$

where L (ϑ) is the likelihood for a continuous variable, ϑ is a vector with two values, with detail description is provided in United States Environmental Protection Agency (2009). The Kaplan–Meier method addressed right-censored data (Tehrany *et al.* 2015), yielding robust distribution and central tendency estimates (Eqn S3). *KM* demonstrated optimal efficacy with lower *NoD* proportions (10–20%), reliably estimating the median across varying sample sizes.

$$\begin{aligned} pe_i &= pe_i + \frac{A_i}{A_i + B_i}(1 - pe_{i+1}) \\ pd_y &= (1 - pe_i) + \left(\frac{j}{A_i + 1}\right) \cdot (pe_i - pe_{i+1}) \end{aligned} \qquad \text{(S3)}$$

where pe_j denotes the proportion of the sample exceeding the *i*th *RL*, once the exceedance probabilities are computed, plotting positions for detections i.e. cumulative probabilities on a probability plot can be calculated with equation pd_y.

$$\sigma_{KM} = \sqrt{\sum_{i=1}^{m} (f(x_i) - \mu_{KM})^2 \cdot F_{KM}(x_{(i-1)})]} \qquad \text{(S4)}$$

where the final Kaplan–Meier estimate (*FKM*) for each *i = 1* (each distinct detected value) is given by a product of these conditional probabilities, and a detailed discussion is provided in United States Environmental Protection Agency (2009). The modelled migration rate is followed by log-normal distribution with means drawn for different empirical distributions. The deviance scale is directly related to the average migration rates and predicted (Donovan and Belmont 2019) by using Eqn S5 as:

$$\sigma = 0.25\mu + 1.09 \qquad \text{(S5)}$$

We computed the probability of significance for each migration rate using a relationship (R^2 = 0.88) between migration rate and probability of relevance, and this technique has 9 to 52% of *NoD* for every iteration.

Validation of error and uncertainty estimates

To substantiate the error and uncertainty assessments, we conducted a comparative analysis between the outcomes derived from remote sensing data and the independent field measurements acquired from various segments of the LIRB. This validation procedure entailed the meticulous cross-referencing of river widths derived from remote sensing techniques with empirical cross-sectional surveys executed on or near the identical dates of the satellite image acquisitions. The validation showed that despite the uncertainties in remote sensing data, the methods used to estimate and correct these errors effectively produced reliable sediment yield estimates. The systematic patterns in the proportion of retained measurements were further examined to ascertain whether variables such as the year of acquisition, image resolution, or the distance of channel migration significantly influenced the measurements' precision. The investigation demonstrated that the natural logarithm of the mean migratory distance constituted the most pivotal variable ($P < 0.001$, $R^2 = 0.8901$), suggesting that the remote sensing data exhibited a notable responsiveness to alterations in channel morphology over temporal scales. Understanding and addressing error and uncertainty in remote sensing images is critical for sediment yield studies, especially in large and dynamic river systems like the LIRB. Remote sensing constitutes a robust methodology for assessing environmental transformations across extensive spatial regions and protracted temporal spans. Nevertheless, without comprehensive error and uncertainty evaluations, the information obtained through remote sensing may result in erroneous inferences, jeopardising the efficacy of soil conservation and watershed management initiatives. In this research endeavor, the comprehensive examination of errors and uncertainties bolstered the validity of the sediment yield estimates derived from remote sensing techniques. It elucidated the constraints and possible origins of inaccuracies in the data. By systematically quantifying these uncertainties, one can enhance the comprehension of the confidence intervals associated with the results, facilitating more informed decision-making regarding land management strategies and erosion mitigation techniques. In summation, the comprehensive examination of error and uncertainty undertaken in this research was pivotal in substantiating the remote sensing data, affirming that the results are dependable and pertinent to practical environmental management. The methodologies employed for estimating error and uncertainty, in conjunction with thorough validation against empirical field data, establish a comprehensive framework for applying remote sensing in the analysis of sediment yield.

Sensitivity analysis

The RUSLE model was initially evaluated by employing its foundational conceptual variables pertinent to each discrete storm event. The resultant data from this evaluation was subsequently juxtaposed with the observed suspended sediment yield at the catchment's outlet. Secondly, an examination of the sensitivity of the conceptual factors was conducted to ascertain which individual factors or a synergistic combination of factors exhibited greater responsiveness to the model output. This analysis was performed employing a variance-based sensitivity methodology aimed at elucidating the parameters of the highest sensitivity. This methodology was favoured in comparison to alternative methodologies, including the one-at-a-time (OAT) approach, which is also referred to as local sensitivity analysis. This technique involves the systematic alteration of individual parameters (by a specified percentage) while maintaining the constancy of all other variables. This methodology was favoured in comparison to alternative methodologies, including the OAT approach, which is also referred to as local sensitivity analysis. This technique involves the systematic alteration of individual parameters (by a specified percentage) while maintaining the constancy of all other variables. Conceptual variables consisted of α and ß, which required enhancement by calibration. Physical variables included *KLSCP*, which was assessed utilising remote sensing and GIS. These parameters were regarded as invariant throughout the study period. The hydrological variables comprised runoff volume and peak runoff discharge, both of which exhibited variability in response to storm events and the time of concentration. Consequently, a sensitivity analysis was conducted on both the conceptual and hydrological parameters employing the Sobol methodology (Wang and Solomatine 2019).

Eicken (1993) introduced a highly effective technique for estimating variance-based sensitivity indices by applying Monte Carlo simulation. The Sobol index measures the sensitivity of output to input variables, with higher values indicating greater influence. This approach constitutes a comprehensive and model-agnostic form of sensitivity analysis predicated on the variance decomposition principle. Its objective is to ascertain the extent to which each parameter and its interactions contribute to the total unconditional variance of the model's output. Let us denote the model as a function.

$$Y = f(x_1, x_2, x_3, ..., x_k) \quad \text{(S6)}$$

where x_1, x_2, …, x_k represent distinct independent variables, and Y denotes the resultant output of the model. Sobol proposed partitioning the function f into components of ascending dimensionality:

$$f(x_1, x_2, x_3, ..., x_k) = f_o + \sum_i f_i + \sum_i \sum_{j>i} f_{ij} + ... + f_{1,2,3,...,k} \qquad \text{(S7)}$$

where each term is a function only of the factors in its index, e.g. f_i (x_i), f_i = f_i (x_i, x_j). The uniqueness of the decomposition is guaranteed under the condition that the input factors are independent and that the individual components in (S2) possess square integrability and exhibit a mean of zero across the specified domain of existence. The total unconditional variance is defined as:

$$V(y) = \sum_i V_i + \sum_i \sum_{j>i} V_{ij} + ... + V_{1,2,...,k} \qquad \text{(S8)}$$

where

$$V_i = V(E(Y \mid \chi_i) \qquad \text{(S9)}$$

$$V_{ij} = V(E(Y \mid \chi_i, \chi j)) - V_i - V_j \qquad \text{(S10)}$$

Equation S7 contains the k terms of the first order. The V_{ij} terms are the second-order terms that explain that part of the effect of x_i and x_j that is not described by the first-order terms. This way, individual factors' variance contributions and interactions with the total output variance can be determined. The variance contributions can then be computed as sensitivity indices:

$$S_i = \frac{V_i}{V} \qquad \text{(S11)}$$

$$S_{ij} = \frac{V_{ij}}{V} \qquad \text{(S12)}$$

$$S_n = S_i + \sum_{j \neq i} S_{ij} + ... + S_{1,2,...,k} \qquad \text{(S13)}$$

The first-order index, S_i, serves as an indicator of the contribution of the variance attributed to individual parameter xi within the overall variance of the model's output. The partial variance V_i, as delineated in Eqn S11, is derived from the variance of the conditional expectation articulated in Eqn S9. The term S_i is frequently referred to as the primary effect of x_i on the dependent variable Y. This can be conceptualised as the proportion of the variance in the model's output that would, on average, be eliminated if x_i were constrained to a specific value within its defined range. The influence on the variance of the model's output resulting from the interaction between factors x_i and x_j is represented by S_{ij}. *STi* encapsulates the primary effect of *xi* alongside all interactions with other factors extending up to the *k*th order.

References

Donovan M, Belmont P (2019) Timescale dependence in river channel migration measurements. *Earth Surface Processes and Landforms* **44**, 1530–1541.

Eicken H (1993) Automated image analysis of ice thin sections—instrumentation, methods and extraction of stereological and textural parameters. *Journal of Glaciology* **39**, 341–352.

Pan J-X, Fang K-T, Pan J-X, Fang K-T (2002) Maximum likelihood estimation. *Growth Curve Models and Statistical Diagnostics* 77–158.

Pan JX, Fang KT (2002) Maximum likelihood estimation. In 'Growth Curve Models and Statistical Diagnostics'. Springer Series in Statistics, pp. 77–158. (Springer: New York, NY, USA) https://doi.org/10.1007/978-0-387-21812-0_3

Peña-Arancibia JL, van Dijk AIJM, Renzullo LJ, Mulligan M (2013) Evaluation of precipitation estimation accuracy in reanalyses, satellite products, and an ensemble method for regions in Australia and South and East Asia. *Journal of Hydrometeorology* **14**, 1323–1333. doi:10.1175/JHM-D-12-0132.1

Tehrany MS, Pradhan B, Mansor S, Ahmad N (2015) Flood susceptibility assessment using GIS-based support vector machine model with different kernel types. *CATENA* **125**, 91–101. https://doi.org/10.1016/j.catena.2014.10.017

United States Environmental Protection Agency (2009) Statistical analysis of groundwater monitoring data at RCRA facilities: unified guidance. March 2009. EPA 530/R-09-007. (US EPA) Available at https://archive.epa.gov/epawaste/hazard/web/pdf/unified-guid.pdf

Wang A, Solomatine DP (2019) Practical experience of sensitivity analysis: Comparing six methods, on three hydrological models, with three performance criteria. *Water* **11**, 1062.

White E, Shephard MW, Cady-Pereira KE, Kharol SK, Ford S, Dammers E, Chow E, Thiessen N, Tobin D, Quinn G (2023) Accounting for non-detects: application to satellite ammonia observations. *Remote Sensing* **15**, 2610.