Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Satellite-Based Estimation of Sediment Bulk Density: Role of Geomorphic Units in the Padma River, Bangladesh

Md Ataul Gani

gmdataul@bot.jnu.ac.bd

Jagannath University

Syed Asif Hasan

Jagannath University

Research Article

Keywords: Sediment bulk density, Geomorphic units (GUs), Sentinel-2, Random Forest (RF), Linear mixed models (LMMs), Large tropical river

Posted Date: May 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9664077/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Sediment bulk density (BD) is an important parameter influencing biomass productivity, root growth, water and nutrient uptake, and key biogeochemical processes such as denitrification, nitrogen fixation, and nitrification. The present study explores the potential of satellite data and river geomorphology to predict sediment BD. By analysing 210 sediment samples from the Padma River in Bangladesh, predictive models were developed using linear mixed models (LMMs) and machine learning (ML). The study shows the significant role of geomorphic units (GUs) in estimating BD, with GUs explaining 30.6% of BD variance compared to 12.9% for land use land cover types and 11.9% for mean sediment texture. Satellite-derived indices (NDVI, Sentinel-2 band 11 and SBL) showed strong correlations with field measurements (vegetation cover, soil moisture and % silt-clay), enabling accurate BD prediction without extensive sampling. Among machine learning models, Random Forests performed better than the others. LMM outperformed RF (R^2 = 0.75 vs 0.50), yielding results that capture a smoother, more constrained distribution and indicate more stable, interpretable predictions. Spatial prediction at 10 m resolution across 132.6 km^2 showed distinct patterns aligned with geomorphic understanding. Higher BD in vegetation islands and lower BD in the transverse, longitudinal and side bars. In addition, the spatial frequency distribution showed that RF's broader, multimodal distribution captures ecological complexity and extreme values more effectively, making it suitable for representing fine-scale heterogeneity. The present research presents a practical approach to large-scale sediment BD estimation, which has implications for understanding sediment dynamics, nutrient cycling, and river management in data-sparse regions.

1. Introduction

Sediment bulk density (BD) is a fundamental physical property that exerts critical control on numerous environmental processes in river systems (Bruland & Richardson, 2005). BD influences sediment porosity, water retention capacity, hydraulic conductivity, and resistance to erosion (Hillel, 2003). These properties, in turn, regulate biogeochemical processes including denitrification, nitrogen fixation, nitrification, and organic matter decomposition (Groffman et al., 2009; Morse et al., 2012). BD also serves as a key parameter for converting nutrient concentrations measured in mass units (mg kg^{-1}) to areal stocks (kg ha^{-1}), making it essential for carbon and nutrient budgeting (Poeplau et al., 2017).

Traditional methods for determining sediment BD are time-consuming, labour-intensive, and logistically challenging, particularly in large, dynamic river systems with limited accessibility (Mulder et al., 2011). Standard protocols require sample collection, lab work such as oven-drying, and precise volume measurements. All these procedures limit sampling frequency and spatial coverage (Blake & Hartge, 1986). This constraint has motivated research into alternative approaches to efficiently estimate BD over large areas using readily available data sources (Tranter et al., 2007; Jalabert et al., 2010).

Remote sensing offers a promising avenue for large-scale estimation of sediment properties (Ben-Dor et al., 2009). Multispectral imagery provides information on surface reflectance, vegetation cover, and

moisture content—all of which are related to sediment properties (Vaudour et al., 2019). The Normalised Difference Vegetation Index (NDVI) has been widely used as a proxy for vegetation cover and organic matter input, both of which influence BD through organic matter dilution and root penetration (Jobbágy & Jackson, 2000). Shortwave infrared (SWIR) bands are sensitive to soil moisture and clay content, providing additional predictive information (Summers et al., 2011; Castaldi et al., 2016). Recent advances in machine learning have enabled integration of these multiple data sources to predict soil properties with increasing accuracy (Hengl et al., 2017; Lamichhane et al., 2019).

Concurrently, geomorphological understanding has advanced to recognise that fluvial landforms—or geomorphic units (GUs)—integrate multiple environmental controls on sediment properties (Wheaton et al., 2015; Williams et al., 2020). GUs such as floodplains, bars, and channels represent distinct process-form assemblages shaped by specific combinations of flow regime, sediment supply, and vegetation dynamics (Buffington & Montgomery, 2013). Within each GU, sediments exhibit characteristic patterns of texture, organic matter content, and compaction that reflect their formative processes and post-depositional history (Nicholas et al., 2016). This geomorphic context may provide a powerful framework for understanding and predicting spatial variation in sediment properties.

The Padma River, a major distributary of the Ganges-Brahmaputra system in Bangladesh, provides an ideal setting for investigating these relationships. The river transports one of the highest sediment loads in the world (approximately 1.0-1.5 billion tons annually), creating a highly dynamic geomorphic landscape with extensive islands, sand bars, meandering channels, and expansive floodplains (Kuehl et al., 2020; Sarker et al., 2020; Gani et al., 2022). This geomorphic diversity, combined with pronounced seasonal hydrological variability, generates substantial spatial heterogeneity in sediment properties (Ophra et al., 2018). Understanding this heterogeneity is critical for predicting river behaviour, managing erosion and deposition, and assessing biogeochemical cycling in one of the world's most densely populated river corridors.

So far, no study has yet integrated geomorphic units with satellite-derived predictors to estimate BD in large tropical rivers. Therefore, this study addresses three interconnected research questions: (1) How do geomorphic units influence sediment bulk density (BD) in a large river system like the Padma River? (2) Can satellite-derived indices serve as effective proxies for field-measured predictors of BD? and (3) How well do linear mixed models and machine learning models perform to predict BD from satellite data with the integration of GUs? Therefore, the current research aims to develop a practical approach for large-scale sediment BD estimation that leverages freely available satellite data and geomorphic mapping, enabling potential applications in river management and biogeochemical assessment in data-sparse regions.

2. Materials and Methods

2.1 Study Area

The Padma River is the confluence of the Ganges-Brahmaputra river system, which flows approximately 120 km through Bangladesh before converging with the Meghna River (Gani et al., 2022). The study area encompasses a 50 km stretch of the Padma River corridor, spanning latitudes 23.20°N to 24.50°N and longitudes 89.50°E to 90.80°E (Fig. 1). The study area is characterised by a highly dynamic geomorphic landscapes including primary and secondary channels (C&S), vegetated islands (VI), side bars (SB), longitudinal bars (L), transverse bars (T), dry channels (ED), unvegetated banks (EK), and water depressions (WD). In the area, five primary land use/land cover (LULC) types were found according to the classification system (Gani et al., 2025), such as cropland (CL), natural vegetation (NV), land with water (LW), dry bare land (DBL), and water bodies (WB).

The climate of the study area is tropical humid, with a mean annual precipitation of 2,100 mm, approximately 80% of which occurs during the monsoon season (Gani et al., 2026). Temperature ranges from a mean minimum of 12°C in January to a mean maximum of 34°C in April. The river experiences pronounced seasonal hydrological variability, with discharge ranging from approximately 5,000 m³/s during the dry season (November-May) to over 90,000 m³/s during the monsoon season (June-October) (Gani et al., 2026).

2.2 Field Data Collection

Sediment cores were collected from different sites during the field campaigns in February 2020 (dry season) when the maximum sediment area was exposed. A stratified random sampling design was used to ensure representation across all identified GUs and LULC types (Webster & Oliver, 2007). In total, 210 samples were collected, 160 from the 0–5 cm layer and the remaining from the 5–10 cm depth using PVC cores of known volume (5 cm diameter × 5 cm height) for bulk density determination (Blake & Hartge, 1986). Geographic coordinates, vegetation cover (VC, %), and soil moisture (SM, %) were measured *in situ* using portable devices as described in Gani et al. (2025). Right after collection, all samples were placed in sealed polyethene bags and transported in a cold condition inside an icebox to the research laboratory of the Department of Botany, Jagannath University, Bangladesh, where they were refrigerated at 4°C until analysis (Schumacher, 2002).

2.3 Laboratory analysis

Sediment bulk density (BD) was determined using the core method (Grossman & Reinsch, 2002). Field-moist samples were weighed, oven-dried at 105°C for 48 hours until constant weight, and reweighed to determine dry mass. Bulk density (g cm⁻³) was calculated as:

$$BD = \frac{M_d}{V_c}$$

where M_d is the mass of oven-dried sediment (g) and V_c is the volume of the core (cm³) (Blake & Hartge, 1986).

For particle size distribution, samples were sieved to separate the sand fraction (0.075–4.75 mm), the sand-silt (0.075 – 0.053 mm) fraction, and the silt-clay (< 0.038 mm) fraction (ASTM International, 2017). Soil texture classes were assigned according to the US classification system (Soil Survey Staff, 2014). For subsequent analyses, mean soil texture (MST) values were derived from the relative proportions of sand, sand-silt, and silt-clay.

2.4 Satellite Data Acquisition and Processing

Two segments of Sentinel-2B images from the dry period (11 February 2020) were downloaded (https://scihub.copernicus.eu/). Images were selected based on the cloud cover < 10% and acquisition dates within ± 7 days of field sampling (Roy et al., 2014; Drusch et al., 2012). The image processing and analysis were performed using the semi-automatic classification plugin (SCP) tool in QGIS (Congedo, 2021). After collecting Sentinel-2 Level 1C products, an atmospheric correction was applied (Dark Object Subtraction, DOS1; Moran et al., 1992). Using Sentinel-2 bands, key indices, including the Normalised Difference Vegetation Index (NDVI) and the Soil Background Line (SBL), were calculated. These indices were used as proxies for environmental conditions that influence sediment BD.

2.5 Statistical Analysis and Modelling in R

All statistical analyses were performed in R version 4.3.1 (R Core Team, 2023) using the packages for data manipulation, visualisation, and modelling, including tidyverse (Wickham et al., 2019), ggplot2 (Wickham, 2016), nlme (Pinheiro & Bates, 2000), caret (Kuhn, 2008), and terra (Hijmans, 2024).

2.5.1 Preliminary Analysis of Spatial and Temporal Variability

First, the spatial and temporal variability of BD was explored using descriptive statistics and visualisation techniques. A paired t-test was performed to compare the measured BD at 0–5 cm and 5–10 cm depths. As there was no significant difference between the depths, data from the 0–5 cm depth were used for the analysis. Box plots were generated using ggplot2 to examine BD distribution across LULC types, soil texture classes, and GUs (Fig. 3). A Kruskal-Wallis rank test was applied to compare the BD among the LULC types, with a post hoc Dunn test with Bonferroni correction (R package “dunn.test”). Simple linear regression models were fitted to quantify the explanatory power of LULC, MST, and GUs individually (Table 2), using the lm function in base R (Chambers, 1992).

2.5.2 Linear Mixed Models (LMMs)

A series of linear mixed models was developed to predict sediment BD from both field-based and satellite-derived variables while accounting for the hierarchical structure of the data (Pinheiro & Bates, 2000). Model development followed a stepwise approach, progressively increasing complexity to identify the optimal random effects structure (Zuur et al., 2009). For each model type (field-based and satellite-based), four formulations were applied and compared as shown in Table 1.

Table 1
Description of different linear mixed models (LMM) used in the present study.

Data type symbol	Model type	Feature	Reference
F1, S1	Generalised least squares	No random effects, assuming independent errors	Fox & Weisberg, 2019
F2, S2	Random intercept model	Includes GUs as random intercepts to account for between-unit variability	Snijders & Bosker, 2012
F3, S3	Generalised least squares with variance structure	Accounting for heteroscedasticity across GUs	Pinheiro & Bates, 2000
F4, S4	Random intercept with variance structure	Combining random intercepts with GU-specific variance weights	Zuur et al., 2009

For field-based models (F1-F4), fixed effects included vegetation cover (VC, %), sediment moisture (SM, %), and silt-clay content (%; derived from particle-size analysis). For satellite-based models (S1-S4), fixed effects included NDVI, Sentinel-2 Band 11 (B11), and the soil background line (SBL). Models were fitted using Restricted Maximum Likelihood (REML) estimation via the nlme package (Pinheiro et al., 2023). Model comparison was conducted using Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and log-likelihood values, with lower AIC/BIC indicating better fit (Burnham & Anderson, 2002; Sakamoto et al., 1986). Likelihood ratio tests were employed to compare nested models, with significance assessed at $\alpha = 0.05$ (Lewis et al., 2011).

2.5.3 Machine Learning Models

To complement the LMM approach, five machine learning algorithms were used for comparison. Random Forest (RF), Extreme Gradient Boosting (XGB), Support Vector Machine (SVM), Neural Networks (NN) and generalised linear model (GLMNET). All models were trained using the same set of satellite-derived predictors (NDVI, B11 and SBL) and GUs as categorical predictors.

Table 2
Description of the machine learning models used in the present study.

Model name	Used R package	Feature	Reference
Random Forest (RF)	ranger	500 trees and mtry tuned via 5-fold cross-validation	Wright & Ziegler (2017); Brian (2005)
Extreme Gradient Boosting (XGB)	xgboost	Hyperparameters tuned via 5-fold cross-validation	Chen et al. (2023)
Support vector machine (SVM)	kernlab	A radial basis function kernel was used	Karatzoglou et al. (2004)
Neural network (NN)	nnet package	A single-hidden-layer feedforward was implemented	Venables & Ripley (2002)
Generalised linear models with elastic net regularisation (GLMNET)	glmnet package	The mixing parameter α was tuned via cross-validation.	Friedman et al. (2010)

2.5.4 Model Performance Evaluation Metrics

Model predictive performance was evaluated using three complementary metrics: the R-squared (R^2), Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) (Piñeiro et al., 2008; Alexander et al., 2015). The R^2 represents the proportion of variance in bulk density explained by the model, with higher values indicating better predictive accuracy. RMSE and MAE quantify prediction errors, with lower values indicating better model performance (Nagelkerke, 1991). For LMMs, conditional R^2 was reported, accounting for both fixed and random effects and ensuring a fair comparison with the standard R^2 of RF (Nakagawa & Schielzeth, 2013). Bar plots were generated to compare these metrics between the two modelling approaches visually.

Model prediction accuracy was assessed by plotting observed versus predicted values using ggplot2 visualisation functions. Prediction errors (calculated as observed minus predicted values) were analysed using kernel density estimation to compare the error distributions of LMM and RF models. Observations were sorted by actual bulk density values to facilitate pattern recognition. This visualisation enables assessment of how consistently each model captures the actual data patterns.

2.5.6 Spatial Analysis- Data Preparation and Model Implementation

Spatial Prediction and Mapping

All spatial layers (NDVI, B11, SBL and rasterised GUs) were stacked using the terra package (Hijmans, 2024). All rasters are aligned to a common resolution (10 m) and coordinate reference system (UTM

Zone 44N). Pixels corresponding to primary and secondary channels (C&S) were masked from the analysis due to the absence of training data (Meyer & Pebesma, 2021).

LMM predictions were generated using the predict () function from the nlme package (Pinheiro et al., 2023) with two levels: (i) for pixels where GU was present in training data: level = 1 predictions incorporating both fixed and random effects and (ii) for pixels where GU was not present in training data: level = 0 predictions using only fixed effects. The trained RF model was applied to the predictor stack using the predict () function with type = "response" (Wright & Ziegler, 2017). Predictions were generated for all unmasked pixels (n = 1,325,900).

Spatial prediction maps were visualised using ggplot2 with geom_spatraster () from the tidyterra package (Hernangómez, 2023). A continuous colour scale (viridis) was applied to represent BD values.

Spatial frequency distribution analysis

Predicted BD values from both models were extracted and analysed using the density () function in R (Venables & Ripley, 2002). Summary statistics (mean, standard deviation, coefficient of variation) were calculated, and density plots were generated using ggplot2.

3. Results

3.1 Spatial Variability of Sediment Bulk Density

Sediment bulk density (BD) exhibited noticeable spatial variability across the investigated land-use/land-cover (LULC) types, mean soil texture (MST) and geomorphic units (GUs) (Fig. 2). When grouped by LULC classes, land with water (LW) and dry bare land (DBL) generally showed higher BD values. In contrast, natural vegetation (NV) and cropland (CL) were associated with lower BD (Fig. 2a). Similarly, MST classes showed a clear gradient, with sand exhibiting higher BD and silt–clay-enriched sediments exhibiting lower BD (Fig. 2b). Across GUs, distinct differences were evident, indicating that geomorphic setting is a strong control on sediment compaction and packing density (Fig. 2c).

Simple regression analyses demonstrated that LULC, MST, and GUs each had a significant effect on BD (Table 3). The models used GUs as the explanatory variable, which accounted for the largest proportion of variance in BD (R^2 = 0.31), compared with models based on LULC (R^2 = 0.13) and MST (R^2 = 0.12).

Table 3
Regression models showing the impact of LULC, MST and GUs on BD estimation.

Model	Intercept	p value	R^2	RMSE
BD ~ LULC	1.696	7.514e-05	0.13	0.283
BD ~ MST	1.951	4.79e-05	0.13	0.284
BD ~ GUs	2.025	2.15e-10	0.31	0.255

3.2 Linear Mixed Models Performance

Comparison among the different linear mixed models (LMMs) showed substantial improvements in model fit when random effects and variance structures were included (Table 4). For both field-based and satellite-based datasets, the simplest models (F1 and S1) assuming independent errors performed poorly, with high AIC values (F1: AIC = 76.86; S1: AIC = 76.03) and low log-likelihood (F1: logLik = -33.43; S1: logLik = -33.01). Inclusion of GUs as random intercepts in both models (F2 and S2) dramatically improved model fit. This was evidenced by the substantial decrease in AIC (F2: AIC = 36.09; S2: AIC = 28.36) and increase in log-likelihood (F2: logLik = -12.04; S2: logLik = -8.18). Likelihood ratio tests confirmed that the random intercept models significantly outperformed their generalised least squares counterparts (F1 vs F2: L. Ratio = 42.76, $p < 0.0001$; S1 vs S2: L. Ratio = 49.67, $p < 0.0001$). The most complex models combining random intercepts with variance structures (F4 and S4) yielded the best overall performance. Model F4 showed the lowest AIC (28.33) and highest log-likelihood (-2.17) among field-based models, while model S4 showed the lowest AIC (24.97) and highest log-likelihood (-0.49) among satellite-based models. Likelihood ratio tests confirmed that these full models significantly outperformed the variance structure models (F3 vs F4: L. Ratio = 9.48, $p = 0.002$; S3 vs S4: L. Ratio = 19.64, $p < 0.0001$) (Table 4).

Table 4
Comparison among different LMMs. The best-fit models with the lowest AIC and Log-likelihood values are marked in bold.

Model code	Model type	Variable source	AIC	BIC	Log-likelihood	Test	L. Ratio	p-value
F1	Generalised least-square	Field	76.86	92.11	-33.43			
F2	Random intercept	Field	36.09	54.39	-12.04	F1 vs F2	42.76	< 0.0001
F3	Generalised least squares with variance structure	Field	35.81	69.39	-6.91	F2 vs F3	10.29	0.06
F4	Random intercept with variance structure	Field	**28.33**	**64.93**	**-2.17**	**F3 vs F4**	**9.48**	**0.002**
S1	Generalised least-square	Satellite	76.03	91.28	-33.01			
S2	Random intercept	Satellite	28.36	46.65	-8.18	S1 vs S2	49.67	< 0.0001
S3	Generalised least squares with variance structure	Satellite	42.61	76.16	-10.31	S2 vs S3	4.25	0.51
S4	Random intercept with variance structure	Satellite	**24.97**	**61.57**	**-0.49**	**S3 vs S4**	**19.64**	< 0.0001

3.3 Fixed and Random Effects in Best-Fitting LLM Model

In the best-performing field-based model (F4), fixed effects revealed significant negative relationships between BD and all three predictors (Table 5). Vegetation cover (VC) showed the strongest negative effect ($\beta = -0.237$, $p < 0.001$), indicating that a 1% increase in VC corresponds to a 0.00237 g cm^{-3} decrease in BD. In the best-performing models (F-4 and S-4), fixed effects showed significant negative relationships between BD and VC, silt–clay content, and SM (Table 4). For the satellite-based model, NDVI and Sentinel-2 band 11 also showed significant negative effects on BD, confirming that remotely sensed proxies of vegetation and moisture capture key controls on sediment BD.

Table 5
Results of LMMs (Models F-4 and S-4) showing fixed effects and random effects of different factors on predicting PDR. The significance of intercepts and slopes is indicated with ($p < 0.1$), * ($p < 0.05$) and *** ($p < 0.001$). The total number of observations was 160. (F = model with field-based data and S = model with satellite-based data).

Fixed effects	**Model F-4**		**Model S-4**	
Intercept	1.989	***	1.793	***
Slopes				
VC (%)	-0.237	***		
Silt-Clay (%)	-0.009	***		
SM (%)	-0.019			
NDVI			-0.397	*
SBL			-0.779	.
Sentinel-2 band 11			-0.681	.

The random effects associated with GUs indicated substantial between-unit variability, further emphasising the importance of geomorphic context in controlling BD patterns. Figure 3 presents the random effects (GU-specific intercept deviations) for the best-fitting models. The larger magnitude of random effects in Model S-4 (range: -0.2820 to + 0.3983) compared to Model F-4 (range: -0.1799 to + 0.2783) indicates that the satellite-based model relies more on GU-specific adjustments to achieve accurate predictions, compensating for the absence of direct field measurements.

3.4 LLM Model Diagnostics and Validation

Model diagnostics showed good agreement between observed and predicted BD values. The scatter plots showed that the models captured both the central tendency and a substantial portion of the observed BD variability (Fig. 4a). The figure also shows random scatter around zero with no funnel-shaped patterns, confirming that the variance structure successfully accounted for heteroscedasticity (Fig. 4b) with residuals approximately normally distributed (Fig. 4c). The approximately normal distribution (Shapiro-Wilk test: $W = 0.99$, $p = 0.34$) confirms that the normality assumption of the mixed model is satisfied. Together, these diagnostics validate the statistical assumptions underlying the LMM and confirm its suitability for predicting sediment BD.

3.5 Comparison Between Field-Based and Satellite-Based LLM Estimates

Comparison between field-based predictors (vegetation cover, soil moisture and silt-clay) and satellite-derived proxies (NDVI, band 11 and SBL) showed strong consistency in BD estimation (Fig. 5). Predicted values from both approaches closely followed the 1:1 line. This indicates that satellite-derived proxies can effectively substitute for field measurements in large-scale BD mapping. The correlation coefficient (R^2) of 0.82 indicates good agreement. Minor deviations at the extremes (BD < 1.3 g cm^{-3} and > 1.9 g cm^{-3}) suggested some saturation effects in spectral indices at very high or low vegetation/moisture conditions.

3.6 Machine learning model performance

Among the machine learning models, Random Forest (RF) demonstrated superior predictive performance across all metrics. The RF model achieved an R^2 of 0.50 (SD = 0.03), whereas XGB (R^2 = 0.49, SD = 0.03), SVM (R^2 = 0.45, SD = 0.04), Neural Network (R^2 = 0.49, SD = 0.04), and GLMNET (R^2 = 0.48, SD = 0.05) showed lower R^2 than RF. All ML models' R^2 values are lower than the LMM (Table 6).

Table 6
Comparison of different machine learning models' predictions based on satellite data.

Model	RMSE	MAE	R^2
RF	0.237	0.192	0.502
XGB	0.231	0.188	0.487
SVM	0.238	0.186	0.447
NN	0.228	0.184	0.495
GLMNET	0.241	0.201	0.480

3.7 Prediction accuracy among LMM and RF

LMM predictions cluster closer to the 1:1 line, indicating better accuracy, with a slight tendency to under-predict. RF predictions show greater deviation, consistently over-predicting bulk density relative to observed values (Fig. 6). RF overestimates bulk density here, suggesting possible bias or overfitting for this sample or condition.

Model performance comparison between LMM and RF showed that LMM achieved higher R^2 with higher RMSE and MAE, indicating superior predictive accuracy but higher error (Fig. 7). LMM had a smaller spread, no extreme outliers, and median errors closer to zero. On the contrary, RF showed either a tight median with many extreme positive outliers or a consistent positive bias with many extreme outliers

(Fig. 8a). Tracking predictions across ordered observations revealed that LMM more closely matched the actual BD values across the entire range. The LMM prediction line showed better alignment with observed values and narrower error ribbons. In contrast, RF predictions deviated slightly, particularly at higher BD values (Fig. 8b). This result indicated that LMM is better suited for capturing both global trends and local-scale variability in sediment BD.

3.8 Spatial Patterns of Predicted Bulk Density

3.8.1 Spatial Prediction and Mapping

Spatial prediction across the study area, based on approximately 1,325,900 pixels, showed clear differences in LMM and RF representations of the BD. The LMM-generated map showed smoother, more spatially coherent patterns, reflecting the model's tendency toward population-averaged estimates. In contrast, the RF map showed greater spatial heterogeneity, capturing fine-scale variability across different geomorphic settings (Fig. 9). Both maps successfully captured the major spatial patterns expected from geomorphic understanding. Higher BD values (red-orange, 1.8–2.1 g cm^{-3}) were observed in vegetated islands (VI), where finer sediments accumulate, and compaction occurs. Lower BD values (blue-green, 1.2–1.5 g cm^{-3}) were found in bar complexes (T, L and SB), reflecting coarse, loosely packed sandy sediments. Between these two, intermediate BD values were found in transitional areas such as dry channels (ED) and unvegetated banks (EK). The RF map, with sharper transitions at GU boundaries and more localised variation within GUs, better reflects the expected ecological reality of sediment properties in this dynamic river system (Fig. 9).

3.7.2 Spatial Frequency Distribution Analysis

The frequency distribution analysis reveals fundamental differences in the spatial predictions of RF and LMM. The LMM predictions followed approximately a normal distribution centred at ~ 1.38 g cm^{-3}. Thus, a relatively narrow range and reduced variance indicate conservative, population-centred estimates. The RF predictions exhibited a broader, multi-modal distribution with extended tails and a wider range (up to ~ 2.15 g cm^{-3}), capturing extreme values and potential sub-populations within the data. The multiple peaks in the RF distribution likely correspond to different geomorphic units, such as the primary peak (~ 1.4 g cm^{-3}), which represents sandy bars (T, L and SB), the secondary peak (~ 1.7 g cm^{-3}) corresponds to vegetated islands, and the tertiary peak (~ 1.9 g cm^{-3}) captures highly compacted sediments in ED and WD. Although the two distributions overlapped substantially (~ 87%), RF better represented ecological heterogeneity and local-scale extremes, whereas LMM provided more smoothed, statistically regularised estimates (Fig. 10).

4. Discussion

4.1 Geomorphic Controls on Sediment Bulk Density

The present research shows that geomorphic units (GUs) exert a dominant control on sediment bulk density (BD) in the Padma River system. GUs explained 30.6% of the variance in BD when considered alone and contributed substantially to predictive models combined with other environmental variables. The strong GU effect reflects the integration of multiple controlling factors within each geomorphic context. VI, for instance, experience regular deposition of fine sediments, prolonged saturation, and vegetation establishment. All the processes increase BD through compaction, organic matter incorporation, and reduced porosity (Nicholas et al., 2016; Walling et al., 2021). In contrast, active bars (T, L and SB) are characterised by frequent reworking, coarse sediment texture, and limited vegetation, resulting in consistently lower BD (Ashworth et al., 2015). This result supports the emerging understanding of geomorphology as an integrative framework for sediment properties and processes (Wheaton et al., 2015; Williams et al., 2020).

The better explanatory power of GUs compared to LULC (R^2 = 0.129) and MST (R^2 = 0.119) suggests that geomorphic complexity captures not only the current sediment state but also its depositional history and ongoing modification processes (Phillips, 2016). This phenomenon is critical in dynamic river systems, where sediments are frequently mobilised, transported, and redeposited. Thus, carrying the imprint of previous environments (Walling, 2013). The present results support the conceptualisation of GUs as "process-form assemblages" that integrate both the material properties and the formative processes of sedimentary environments (Buffington & Montgomery, 2013).

4.2 Remote Sensing Proxies for Sediment Properties

The strong agreement between field-based and satellite-based predictions (R^2 = 0.82; Fig. 6) demonstrates that remote sensing can effectively substitute for direct field measurements in large-scale BD assessment. This finding has important practical implications, as field sampling is time-consuming, labour-intensive, and logistically challenging in large river systems with limited accessibility (Mulder et al., 2011; Gani et al., 2025). The ability to produce a reliable BD map from Sentinel-2 data enables monitoring sediment properties across large spatial extents and temporal scales (Gholizadeh et al., 2018).

The successful application of Sentinel-2-derived indices (NDVI, Sentinel-2 band 11, and SBL) for BD prediction extends the growing literature on satellite-derived soil property indices (Vaudour et al., 2019; Angelopoulou et al., 2019). The significant negative relationship between NDVI and BD (β = -0.397, $p < 0.01$) suggests that vegetation contributes to the addition of organic matter, thereby reducing sediment BD by increasing porosity and lowering particle density (Jobbágy & Jackson, 2000). The strong effect of B11 (SWIR) on BD estimation (β = -0.779, $p < 0.001$) reflects the sensitivity of shortwave infrared reflectance to both soil moisture and organic matter content (Vaudour et al., 2019; Ben-Dor et al., 2009). SWIR bands have been successfully used to predict soil organic carbon, clay content, and moisture status in various environments (Summers et al., 2011; Castaldi et al., 2016). The present study confirms that SWIR data, combined with vegetation indices, provide a robust basis for estimating sediment BD without extensive field sampling.

4.3 Model Performance and Methodological Implications

The better performance of the Linear Mixed Model compared to Random Forest (R^2 = 0.75 vs 0.50) demonstrates that the LMM retains important advantages for certain applications. The interpretability of fixed and random effects provides insights into the magnitude and significance of individual predictors (Pinheiro & Bates, 2000). The ability to quantify uncertainty through standard errors and confidence intervals supports statistical inference (Gelman & Hill, 2007). For regional-scale assessments where population-level estimates are sufficient and statistical rigour is paramount, LMM remains a valuable approach.

However, the specific advantages of RF revealed in the present analysis—to capture non-linear relationships, improve the prediction of extreme values, and produce narrower error distributions—are particularly relevant for applications that require accurate local estimates (Hengl et al., 2017; Grimm et al., 2008). For example, carbon stock assessment requires accurate BD estimates at individual sampling points to convert concentration measurements to stocks (Poeplau et al., 2017). Similarly, habitat suitability modelling for benthic organisms may depend on identifying small areas of unusually high or low sediment density (Buffington & Montgomery, 2013). Thus, the ability to represent fine-scale spatial heterogeneity of both continuous and categorical predictors makes it well-suited for complex environmental systems where theoretical understanding is incomplete (Breiman, 2001; Strobl et al., 2009).

The selection between methods (either LMM or RF) should therefore be guided by the specific objectives of the study: RF for automatically handling interactions and fine-scale pattern detection, and LMM for statistically interpretable results and uncertainty quantification. In practice, a complementary approach using both methods may be optimal, leveraging the strengths of each for different aspects of the analysis (Kuhn & Johnson, 2013).

4.4 Implications for River Management and Biogeochemical Cycling

The spatial patterns of sediment bulk density revealed by the models have important implications for understanding biogeochemical processes in the Padma River system. Bulk density is a key control on sediment porosity, water retention, and gas exchange—all factors that influence microbial activity and nutrient cycling (Groffman et al., 2009). Areas of low BD, such as active bars (T, L, and SB), are characterised by high porosity and rapid water movement, potentially supporting aerobic processes such as nitrification (Bouwman et al., 2013). In contrast, high-BD in vegetation island sediments with finer texture and higher organic matter content may favour anaerobic processes including denitrification and methanogenesis (Groffman et al., 2009; Morse et al., 2012).

The geomorphic control on BD thus translates into spatial structuring of biogeochemical potential across the river corridor. This "biogeomorphic" perspective recognises that geomorphic processes create habitat templates that shape microbial communities and their activities (Corenblit et al., 2011;

Fisher et al., 2007). The maps of predicted BD (Fig. 10) provide a foundation for scaling up point measurements of biogeochemical processes to the river-corridor scale, addressing a fundamental challenge in aquatic biogeochemistry (Peterson et al., 2001; McClain et al., 2003).

For river management, the ability to map sediment properties at 10 m resolution across 130 + km² of river corridor offers unprecedented information for planning and decision-making. Areas of high BD may be targeted for restoration through vegetation establishment, thereby reducing compaction and improving habitat quality (Beechie et al., 2010). Conversely, areas of very low BD may be prioritised for bank stabilisation if erosion risk is high (Thorne, 1990). The temporal dimension—by applying our approach to multitemporal satellite imagery—could enable monitoring of changes in sediment properties in response to flow regulation, climate variability, or management interventions (Gholizadeh et al., 2018).

4.5 Limitations and Future Directions

The present study provides valuable insights into estimating BD using satellite data. However, several limitations can be acknowledged. Firstly, the training data (n = 160) is modest for machine learning applications. While cross-validation provides robust performance estimates, additional samplings would likely improve model accuracy and generalizability. Secondly, the study is limited to a single river system; application to other rivers would require local calibration or demonstration of model transferability. Finally, the study focused on surface sediments (0–10 cm); extending predictions to deeper layers would require additional sampling and potentially different remote sensing approaches.

The spatial resolution of the predictions (10 m) is adequate for reach-scale applications but may miss fine-scale variability important to certain processes. Future work can explore fusion with higher-resolution imagery (e.g., PlanetScope, 3 m) or UAV-based sensing to capture sub-pixel heterogeneity (Matese et al., 2015). Integration of additional remote sensing data—such as radar (Sentinel-1) for moisture estimation or LiDAR for topographic detail—can further improve predictions.

5. Conclusion

The present study shows that sediment bulk density in the Padma River can be reliably estimated from satellite data, thereby removing the need for intensive field sampling across large river systems. The present research led to three main findings. Firstly, geomorphic units exert stronger control on bulk density than either land cover or soil texture. Vegetation islands consistently showed higher bulk density due to fine sediment accumulation and compaction, whereas active bars had lower values because of coarse, frequently reworked sands. Secondly, satellite-derived indices—NDVI, Sentinel-2 band 11 (SWIR), and the SBL—served as effective proxies for vegetation cover, moisture, and organic matter. Predictions based solely on these indices/bands confirm that remote sensing can substitute for direct sampling over large areas. Thirdly, the Linear Mixed Model outperformed Random Forest, producing smoother, more interpretable spatial patterns, while Random Forest tended to overestimate extreme values. For studies where statistical inference and uncertainty quantification matter, LMM appears better suited than machine learning, at least for datasets of this size and structure.

The present research approach, combining geomorphic complexity, remote sensing proxies, and machine learning, is transferable to other data-sparse river systems and can be extended to predict additional sediment properties, such as organic carbon, nutrient content, and texture. As satellite data become increasingly abundant and accessible, such approaches will play an expanding role in environmental monitoring and management, bridging the gap between local measurements and regional assessments.

Declarations

Author Contribution

Conceptualisation: MAG; Data curation: MAG and SAH; Formal analysis: MAG; Methodology: MAG; Software: MAG; Visualisation: MAG and SAH; Writing : MAG and SAH.

References

1. Alexander, D. L. J., Tropsha, A., & Winkler, D. A. (2015). Beware of R^2: Simple, unambiguous assessment of the prediction accuracy of QSAR and QSPR models. *Journal of Chemical Information and Modeling*, 55(7), 1316–1322. 10.1021/acs.jcim.5b00206
2. Angelopoulou, T., Tziolas, N., Balafoutis, A., Zalidis, G., & Bochtis, D. (2019). Remote sensing techniques for soil organic carbon estimation: A review. *Remote Sensing*, 12(5), 842. https://doi.org/10.3390/rs11060676
3. Ashworth, P. J., Best, J. L., & Jones, M. A. (2015). The relationship between channel avulsion and sediment supply in braided rivers. *Geological Society of America Bulletin*, 127(3-4), 396–411. https://doi.org/10.1130/G19919.1
4. ASTM International. (2017). ASTM D6913: Standard test methods for particle-size distribution (gradation) of soils using sieve analysis. West Conshohocken, PA: ASTM International.
5. Beechie, T. J., Sear, D. A., Olden, J. D., Pess, G. R., Buffington, J. M., Moir, H., Roni, P., & Pollock, M. M. (2010). Process-based principles for restoring river ecosystems. *BioScience*, 60(3), 209–222.https://doi.org/10.1525/bio.2010.60.3.7
6. Ben-Dor, E., Chabrillat, S., Demattê, J. A. M., Taylor, G. R., Hill, J., Whiting, M. L., & Sommer, S. (2009). Using imaging spectroscopy to study soil properties. *Remote Sensing of Environment*, 113, S38–S55. https://doi.org/10.1016/j.rse.2008.09.019
7. Blake, G. R., & Hartge, K. H. (1986). Bulk density. In A. Klute (Ed.), *Methods of soil analysis: Part 1—Physical and mineralogical methods* (2nd ed., pp. 363–375). Madison, WI: American Society of Agronomy.
8. Bouwman, A. F., Beusen, A. H. W., Griffioen, J., Van Groenigen, J. W., Hefting, M. M., Oenema, O., Van Puijenbroek, P. J. T. M., Seitzinger, S., Slomp, C. P., & Stehfest, E. (2013). Global trends and

uncertainties in terrestrial denitrification and N_2O emissions. *Philosophical Transactions of the Royal Society B*, 368(1621), 20130112. https://doi.org/10.1098/rstb.2013.0112

9. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324
10. Bruland, G. L., & Richardson, C. J. (2005). Spatial variability of soil properties in created, restored, and paired natural wetlands. *Soil Science Society of America Journal*, 69(1), 273–284. https://doi.org/10.2136/sssaj2005.0273a
11. Buffington, J. M., & Montgomery, D. R. (2013). Geomorphic classification of rivers. In J. Shroder & E. Wohl (Eds.), *Treatise on geomorphology* (Vol. 9, pp. 730–767). San Diego: Academic Press.
12. Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). New York: Springer. http://dx.doi.org/10.1007/b97636
13. Castaldi, F., Palombo, A., Santini, F., Pascucci, S., Pignatti, S., & Casa, R. (2016). Evaluation of the potential of the current and forthcoming multispectral and hyperspectral imagers to estimate soil texture and organic carbon. *Remote Sensing of Environment*, 179, 54–65. https://doi.org/10.1016/j.rse.2016.03.025
14. Chambers, J. M. (1992). Linear models. In J. M. Chambers & T. J. Hastie (Eds.), *Statistical models in S* (pp. 95–144). Pacific Grove, CA: Wadsworth & Brooks.
15. Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., Zhou, T., Li, M., Xie, J., Lin, M., Geng, Y., & Li, Y. (2023). *xgboost: Extreme Gradient Boosting* (R package version 1.7.5.1). https://CRAN.R-project.org/package=xgboost
16. Congedo, L. (2021). Semi-Automatic Classification Plugin: A Python tool for the download and processing of remote sensing images in QGIS. *Journal of Open Source Software*, *6*(64), 3172. https://doi.org/10.21105/joss.03172
17. Corenblit, D., Baas, A. C. W., Bornette, G., Darrozes, J., Delmotte, S., Francis, R. A., Gurnell, A. M., Julien, F., Naiman, R. J., & Steiger, J. (2011). Feedbacks between geomorphology and biota controlling Earth surface processes and landforms: A review of foundation concepts and current understandings. *Earth-Science Reviews*, 106(3-4), 307–331.https://doi.org/10.1016/j.earscirev.2011.03.002
18. Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., & Bargellini, P. (2012). Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sensing of Environment*, 120, 25–36. https://doi.org/10.1016/j.rse.2011.11.026
19. Fisher, S. G., Heffernan, J. B., Sponseller, R. A., & Welter, J. R. (2007). Functional ecomorphology: Feedbacks between form and function in fluvial landscape ecosystems. *Geomorphology*, 89(1-2), 84–96. https://doi.org/10.1016/j.geomorph.2006.07.013
20. Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). Thousand Oaks, CA: Sage.

21. Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1), 1–22.
22. Gani, M. A., Gettel, G. M., van der Kwast, J., Irvine, K. A., & McClain, M. E. (2026). Nitrogen retention dynamics in a large floodplain river: A case study on the Padma River, Bangladesh. *Ecohydrology & Hydrobiology*, 100724.
23. https://doi.org/10.1016/j.ecohyd.2025.100724
24. Gani, M. A., van Dam, A. A., van der Kwast, J., McClain, M. E., Irvine, K. A., & Gettel, G. M. (2025). Estimation of potential denitrification and its spatiotemporal dynamics in seasonally inundated geomorphic units of a large tropical river using satellite data. *Science of the Total Environment*, 959, 178287. https://doi.org/10.1016/j.scitotenv.2024.178287
25. Gani, M. A., van der Kwast, J., McClain, M. E., Gettel, G. M., & Irvine, K. N. (2022). Classification of geomorphic units and their relevance for nutrient retention or export of a large lowland Padma River, Bangladesh: A NDVI-based approach. *Remote Sensing*, 14(6), 1481. https://doi.org/10.3390/rs14061481
26. Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.
27. Gholizadeh, A., Žižala, D., Saberioon, M., & Borůvka, L. (2018). Soil organic carbon and texture retrieving and mapping using proximal, airborne and Sentinel-2 spectral imaging. *Remote Sensing of Environment*, 218, 89–103. https://doi.org/10.1016/j.rse.2018.09.015
28. Grimm, R., Behrens, T., Märker, M., & Elsenbeer, H. (2008). Soil organic carbon concentrations and stocks on Barro Colorado Island—Digital soil mapping using Random Forests analysis. *Geoderma*, 146(1-2), 102–113.
29. Groffman, P. M., Butterbach-Bahl, K., Fulweiler, R. W., Gold, A. J., Morse, J. L., Stander, E. K., Tague, C., Tonitto, C., & Vidon, P. (2009). Challenges to incorporating spatially and temporally explicit phenomena (hotspots and hot moments) in denitrification models. *Biogeochemistry*, 93(1-2), 49–77. https://doi.org/10.1007/s10533-008-9277-5
30. Grossman, R. B., & Reinsch, T. G. (2002). Bulk density and linear extensibility. In J. H. Dane & G. C. Topp (Eds.), *Methods of soil analysis: Part 4—Physical methods* (pp. 201–228). Madison, WI: Soil Science Society of America.
31. Hengl, T., Mendes de Jesus, J., Heuvelink, G. B. M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I., Mantel, S., & Kempen, B. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLoS ONE*, 12(2), e0169748.https://doi.org/10.1371/journal.pone.0169748
32. Hernangómez, D. (2023). *tidyterra: Tidyverse methods and ggplot2 helpers for terra objects* (R package version 0.4.0). https://CRAN.R-project.org/package=tidyterra
33. Hijmans, R. J. (2024). *terra: Spatial data analysis* (R package version 1.7-46). https://CRAN.R-project.org/package=terra

34. Hillel, D. (2003). *Introduction to environmental soil physics*. San Diego: Academic Press.
35. Jalabert, S. S. M., Martin, M. P., Renaud, J. P., Boulonne, L., Jolivet, C., Montanarella, L., & Arrouays, D. (2010). Estimating forest soil bulk density using boosted regression modelling. *Soil Use and Management*, 26(4), 409–419. https://doi.org/10.1111/j.1475-2743.2010.00305.x
36. Jobbágy, E. G., & Jackson, R. B. (2000). The vertical distribution of soil organic carbon and its relation to climate and vegetation. *Ecological Applications*, 10(2), 423–436. https://doi.org/10.1890/1051-0761(2000)010[0423:TVDOSO]2.0.CO;2
37. Karatzoglou, A., Smola, A., Hornik, K., & Zeileis, A. (2004). kernlab—An S4 package for kernel methods in R. *Journal of Statistical Software*, 11(9), 1–20.
38. Kuehl, S. A., Allison, M. A., Goodbred, S. L., & Kudrass, H. R. (2020). The Ganges-Brahmaputra delta. In D. A. D. Evans & A. R. Muxworthy (Eds.), *Delta systems* (pp. 123–156). London: Geological Society of London.
39. Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, 28(5), 1–26.
40. Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. New York: Springer.
41. Lamichhane, S., Kumar, L., & Wilson, B. (2019). Digital soil mapping algorithms and covariates for soil organic carbon mapping and their implications: A review. *Geoderma*, 352, 395–413.
42. Lewis, F., Butler, A., & Gilbert, L. (2011). A unified approach to model selection using the likelihood ratio test. *Methods in Ecology and Evolution*, 2(2), 155–162.
43. Matese, A., Toscano, P., Di Gennaro, S. F., Genesio, L., Vaccari, F. P., Primicerio, J., Belli, C., Zaldei, A., Bianconi, R., & Gioli, B. (2015). Intercomparison of UAV, aircraft and satellite remote sensing platforms for precision viticulture. *Remote Sensing*, 7(3), 2971–2990.
44. McClain, M. E., Boyer, E. W., Dent, C. L., Gergel, S. E., Grimm, N. B., Groffman, P. M., Hart, S. C., Harvey, J. W., Johnston, C. A., Mayorga, E., McDowell, W. H., & Pinay, G. (2003). Biogeochemical hot spots and hot moments at the interface of terrestrial and aquatic ecosystems. *Ecosystems*, 6(4), 301–312.
45. Meyer, H., & Pebesma, E. (2021). Predicting into unknown space? Estimating the area of applicability of spatial prediction models. *Methods in Ecology and Evolution*, 12(9), 1620–1633.
46. Moran, M. S., Jackson, R. D., Slater, P. N., &Teillet, P. M. (1992). Evaluation of simplified procedures for retrieval of land surface reflectance factors from satellite sensor output. *Remote Sensing of Environment*, *41*(2–3), 169–184. https://doi.org/10.1016/0034-4257(92)90076-V
47. Morse, J. L., Ardón, M., & Bernhardt, E. S. (2012). Using environmental variables and soil processes to forecast denitrification potential and nitrous oxide fluxes in coastal plain wetlands. *Journal of Geophysical Research: Biogeosciences*, 117(G2), G02023. https://doi.org/10.1029/2011JG001923
48. Mulder, V. L., de Bruin, S., Schaepman, M. E., & Mayr, T. R. (2011). The use of remote sensing in soil and terrain mapping—A review. *Geoderma*, 162(1-2), 1–19.
49. Nagelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination. *Biometrika*, 78(3), 691–692.

50. Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalised linear mixed-effects models. *Methods in Ecology and Evolution*, 4(2), 133–142.
51. Nicholas, A. P., Ashworth, P. J., Sambrook Smith, G. H., & Sandbach, S. D. (2016). Numerical simulation of bar and island morphodynamics in anabranching megarivers. *Journal of Geophysical Research: Earth Surface*, 121(8), 1520–1541.
52. Ophra, S. J., Begum, S., Islam, R., & Islam, M. N. (2018). Assessment of bank erosion and channel shifting of Padma River in Bangladesh using RS and GIS techniques. *Spat. Inf. Res.* 26, 599–605. https://doi.org/10.1007/s41324-018-0202-2
53. Peterson, B. J., Wollheim, W. M., Mulholland, P. J., Webster, J. R., Meyer, J. L., Tank, J. L., Martí, E., Bowden, W. B., Valett, H. M., Hershey, A. E., McDowell, W. H., Dodds, W. K., Hamilton, S. K., Gregory, S., & Morrall, D. D. (2001). Control of nitrogen export from watersheds by headwater streams. *Science*, 292(5514), 86–90.
54. Phillips, J. D. (2016). Landforms as extended composite phenotypes. *Earth Surface Processes and Landforms*, 41(1), 16–26. https://doi.org/10.1002/esp.3764
55. Piñeiro, G., Perelman, S., Guerschman, J. P., & Paruelo, J. M. (2008). How to evaluate models: Observed vs predicted or predicted vs observed? *Ecological Modelling*, 216(3-4), 316–322.
56. Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. New York: Springer.
57. Pinheiro, J. C., Bates, D. M., & R Core Team. (2023). *nlme: Linear and nonlinear mixed effects models* (R package version 3.1-162). https://CRAN.R-project.org/package=nlme
58. Poeplau, C., Vos, C., & Don, A. (2017). Soil organic carbon stocks are systematically overestimated due to the misuse of the parameters bulk density and rock fragment content. *SOIL*, 3(1), 61–66. https://doi.org/10.5194/soil-3-61-2017
59. R Core Team. (2023). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/
60. Roy, D. P., Wulder, M. A., Loveland, T. R., Woodcock, C. E., Allen, R. G., Anderson, M. C., Helder, D., Irons, J. R., Johnson, D. M., Kennedy, R., Scambos, T. A., Schaaf, C. B., Schott, J. R., Sheng, Y., Vermote, E. F., Belward, A. S., Bindschadler, R., Cohen, W. B., Gao, F., & Zhu, Z. (2014). Landsat-8: Science and product vision for terrestrial global change research. *Remote Sensing of Environment*, 145, 154–172.
61. Sakamoto, Y., Ishiguro, M., & Kitagawa, G. (1986). *Akaike information criterion statistics*. Dordrecht: D. Reidel.
62. Sarker, M. H., Thorne, C. R., Aktar, M. N., & Ferdous, M. R. (2020). Morpho-dynamics of the Brahmaputra-Jamuna River, Bangladesh. *Geomorphology*, 358, 107119.
63. Schumacher, B. A. (2002). *Methods for the determination of total organic carbon (TOC) in soils and sediments*. Washington, D.C.: US Environmental Protection Agency.
64. Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modelling* (2nd ed.). London: Sage.

65. Soil Survey Staff. (2014). *Keys to soil taxonomy* (12th ed.). Washington, D.C.: USDA-NRCS.
66. Strobl, C., Malley, J., & Tutz, G. (2009). An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods*, 14(4), 323–348.
67. Summers, D., Lewis, M., Ostendorf, B., & Chittleborough, D. (2011). Visible near-infrared reflectance spectroscopy as a predictive indicator of soil properties. *Ecological Indicators*, 11(1), 123–131.
68. Thorne, C. R. (1990). Effects of vegetation on riverbank erosion and stability. In J. B. Thornes (Ed.), *Vegetation and erosion* (pp. 125–144). Chichester: Wiley.
69. Tranter, G., Minasny, B., McBratney, A. B., Murphy, B., McKenzie, N. J., Grundy, M., & Brough, D. (2007). Building and testing conceptual and empirical models for predicting soil bulk density. *Soil Use and Management*, 23(4), 437–443.
70. Vaudour, E., Gomez, C., Fouad, Y., & Lagacherie, P. (2019). Sentinel-2 image capacities to predict common topsoil properties of temperate and Mediterranean agroecosystems. *Remote Sensing of Environment*, 223, 21–33.
71. Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). New York: Springer.
72. Walling, D. E. (2013). The evolution of sediment source fingerprinting investigations in fluvial systems. *Journal of Soils and Sediments*, 13(10), 1658–1675.
73. Walling, D. E., Horowitz, A. J., & Owens, P. N. (2021). The role of sediment in the global carbon cycle. *Earth-Science Reviews*, 215, 103551.
74. Webster, R., & Oliver, M. A. (2007). *Geostatistics for environmental scientists* (2nd ed.). Chichester: Wiley.
75. Wheaton, J. M., Fryirs, K. A., Brierley, G., Bangen, S. G., Bouwes, N., & O'Brien, G. (2015). Geomorphic mapping and taxonomy of fluvial landforms. *Geomorphology*, 248, 273–295.
76. Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. New York: Springer.
77. Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., & Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686.
78. Williams, R. D., Bangen, S. G., Gillies, E., Kramer, N., Moir, H., & Wheaton, J. M. (2020). Let the river erode! Enabling lateral migration increases geomorphic unit diversity. *Science of the Total Environment*, 715, 136817.
79. Wright, M. N., & Ziegler, A. (2017). ranger: A fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software*, 77(1), 1–17.
80. Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., & Smith, G. M. (2009). *Mixed effects models and extensions in ecology with R*. New York: Springer. https://doi.org/10.1007/978-0-387-87458-6

Figures

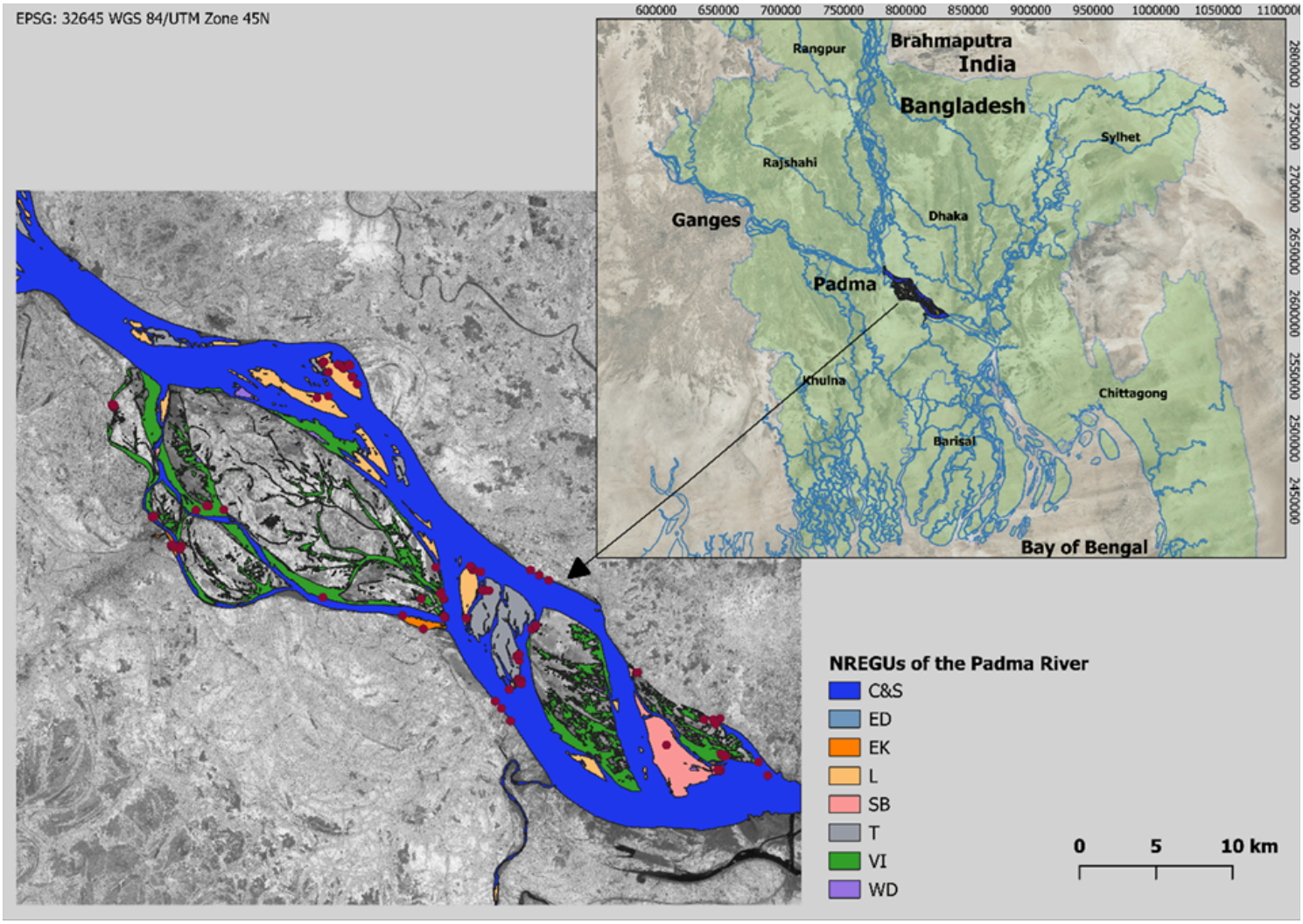


Figure 1

Map of the study area of the Padma River in Bangladesh showing sampling locations and geomorphic units (GUs) such as Secondary Channels (C&S), Vegetated Islands (VI), Sidebars (SB), Longitudinal Bars (L), Transverse Bars (T), Dry Channels (ED), Unvegetated Banks (EK), and Water Depressions (WD). (Collected from Gani et al., 2025).

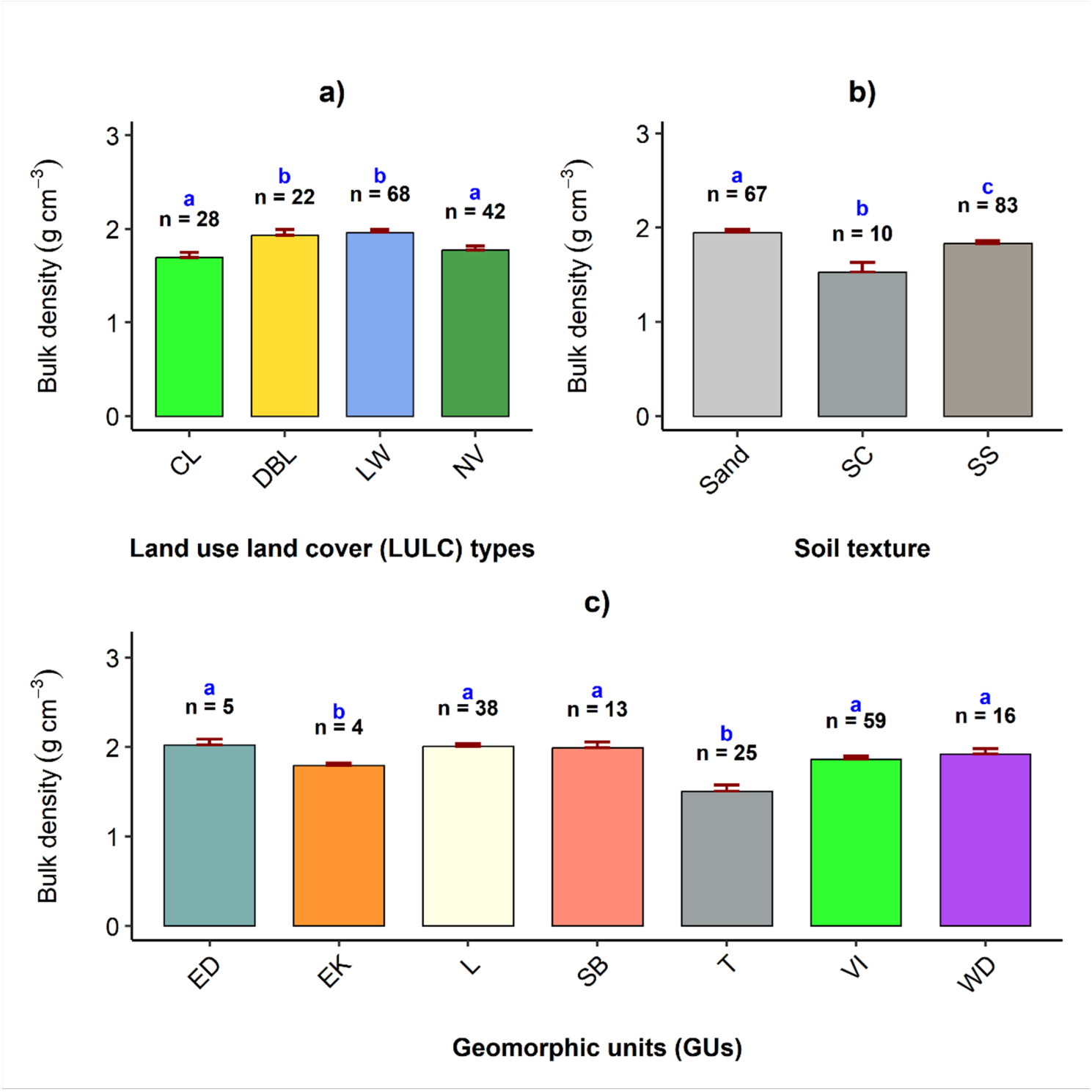


Figure 2

Variation of sediment bulk density (BD) across different categorical variables. a) BD by LULC types; b) BD by soil texture, and c) BD by GUs.

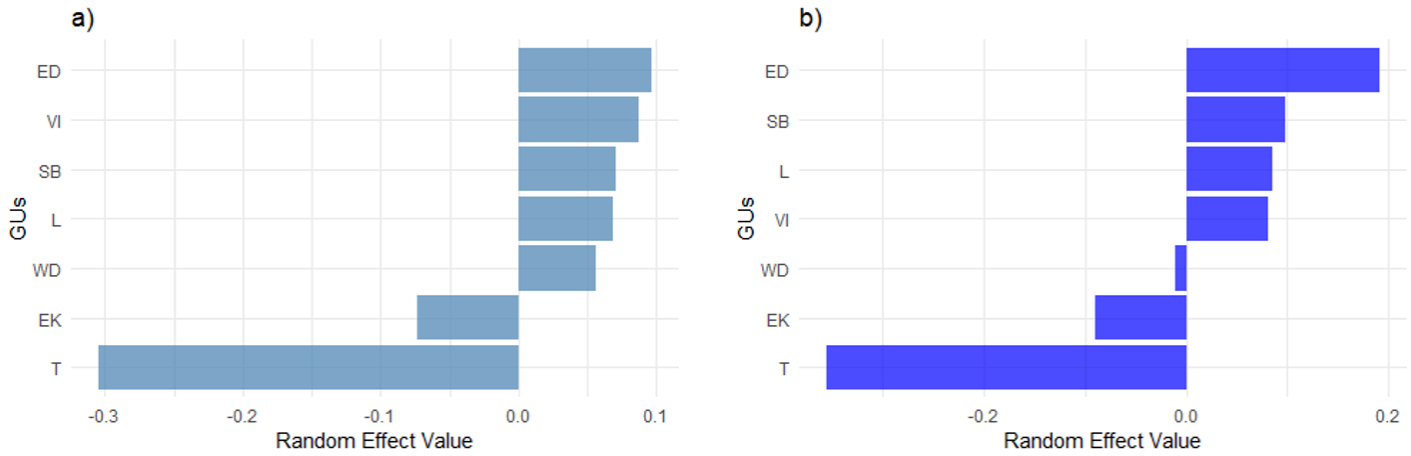


Figure 3

Random effects of different factors on predicting BD (Models F-4 and S-4). a) Random effect values of GUs in the case of Model F-4 (field-based and b) Random effect values of GUs in the case of Model S-4 (satellite-based). Positive values indicate GUs with higher-than-average BD (e.g., ED, SB, L and VI), while negative values indicate GUs with lower-than-average BD (e.g., EK and T).

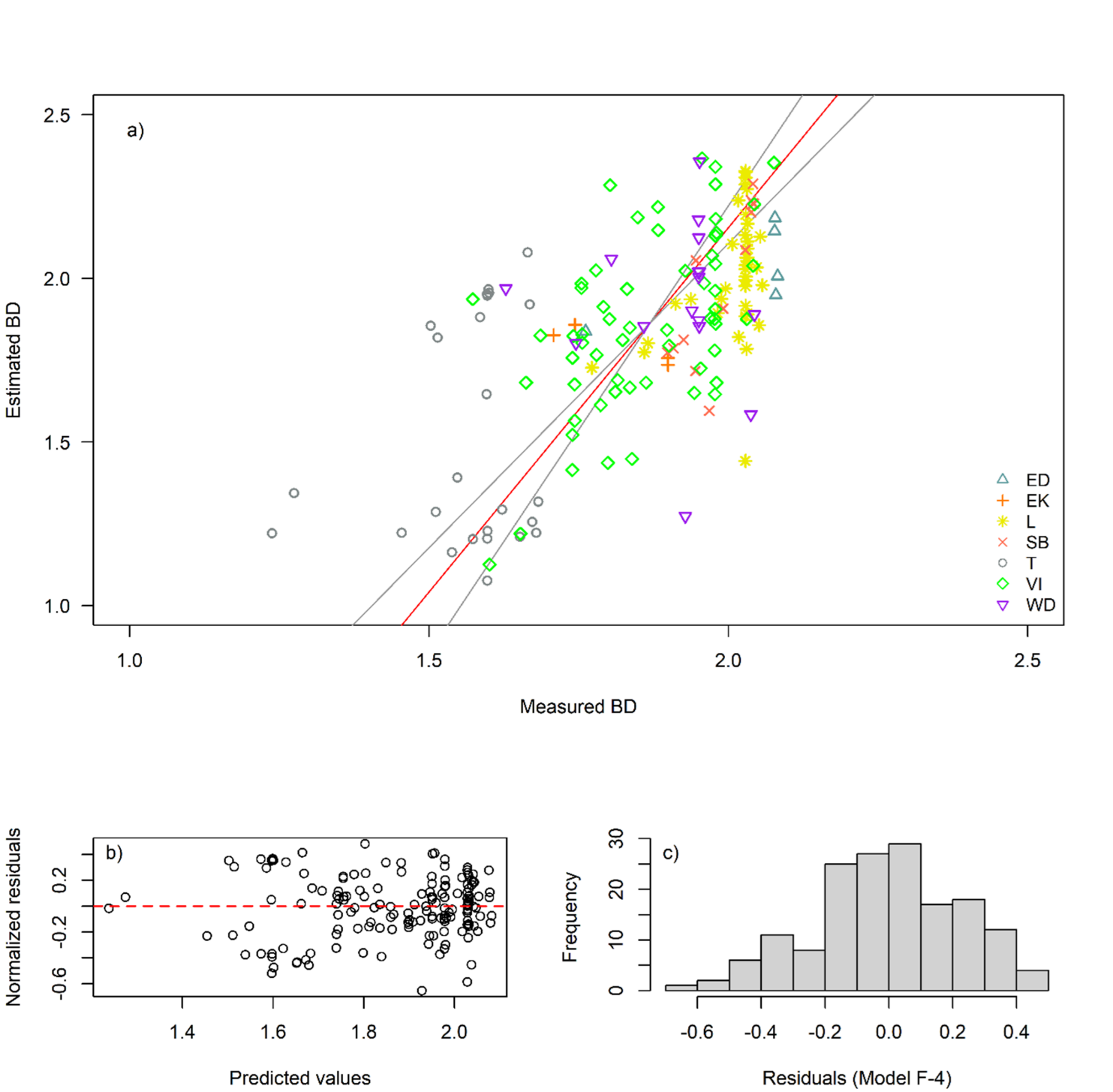


Figure 4

a) Scatter diagram of Model F-4 fit showing regressing BD from fixed (VC, SM and % Silt-Clay) and random effects (GUs); the diagram also shows the appropriate regression line (red line, standard major axis) with confidence intervals (2.5% and 97.5%) of intercepts and slopes. The close alignment of points around the 1:1 line, with no systematic deviation for any GU, indicates unbiased predictions across the BD range. b) Residuals plot showing model performance with no strong heteroscedastic patterns. c) Histogram of raw residuals showing an overlaid normal distribution. Secondary Channels (C&S),

Vegetated Islands (VI), Sidebars (SB), Longitudinal Bars (L), Transverse Bars (T), Dry Channels (ED), Unvegetated Banks (EK), and Water Depressions (WD).

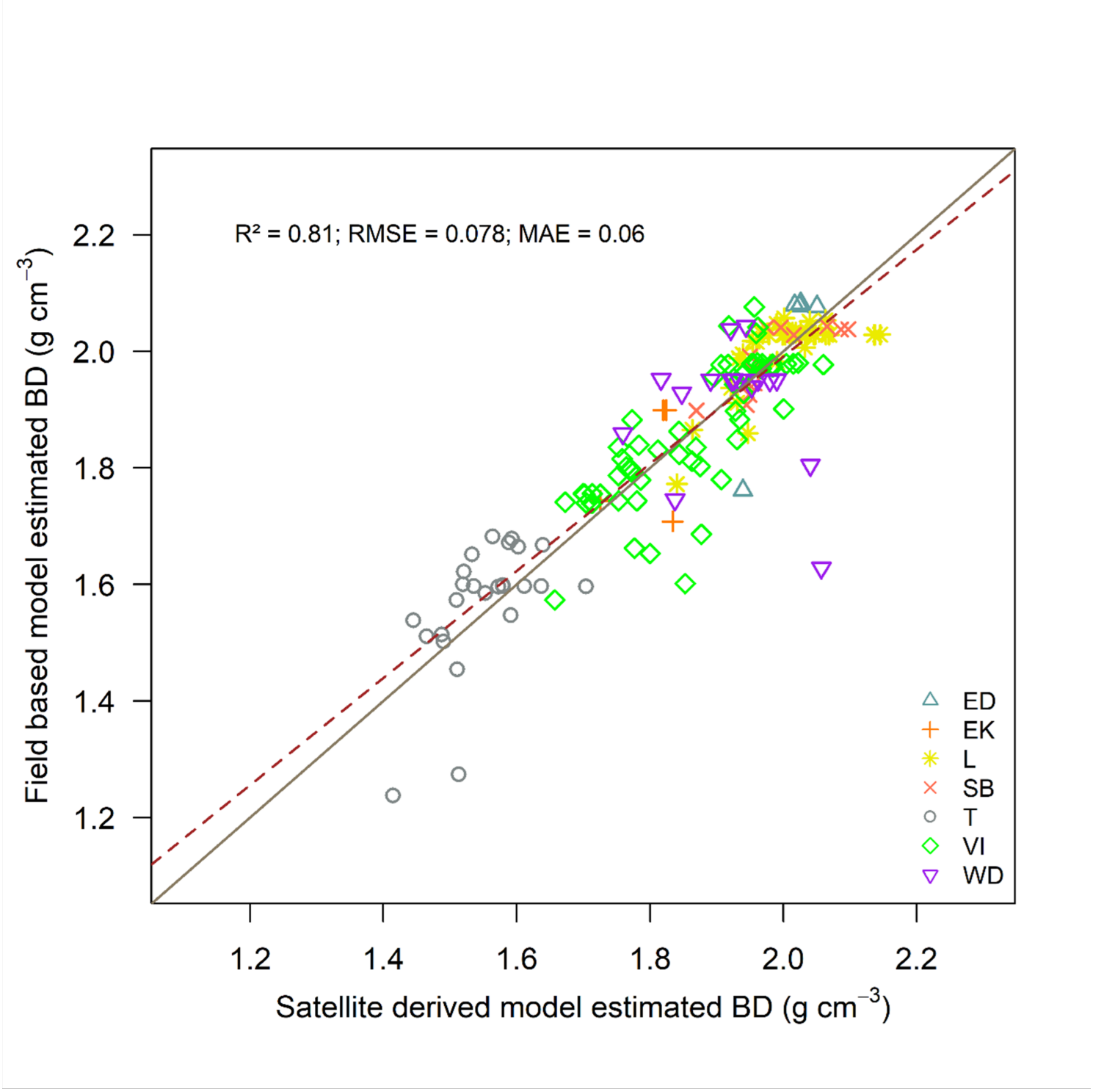


Figure 5

Field-based (vegetation cover, soil moisture and % Silt-Clay) vs satellite-based (NDVI, band 11 and SBL) estimation of BD using field-based and satellite-derived GUs. The dashed lines indicate the model-performed estimates, while the solid black line indicates the y = x line. The strong clustering of points around the 1:1 line across all GUs demonstrates that satellite-derived indices can effectively substitute

for field measurements in BD prediction. Each point represents a sample, with colours indicating different GUs: dry channel (ED = Cade blue), Unvegetated bank (EK = dark orange), Longitudinal bar (L = yellow), Side bar (SB = coral). Transverse bar (T=azure), Vegetation islands (VI=green), and Water depressions (WD=Purple).

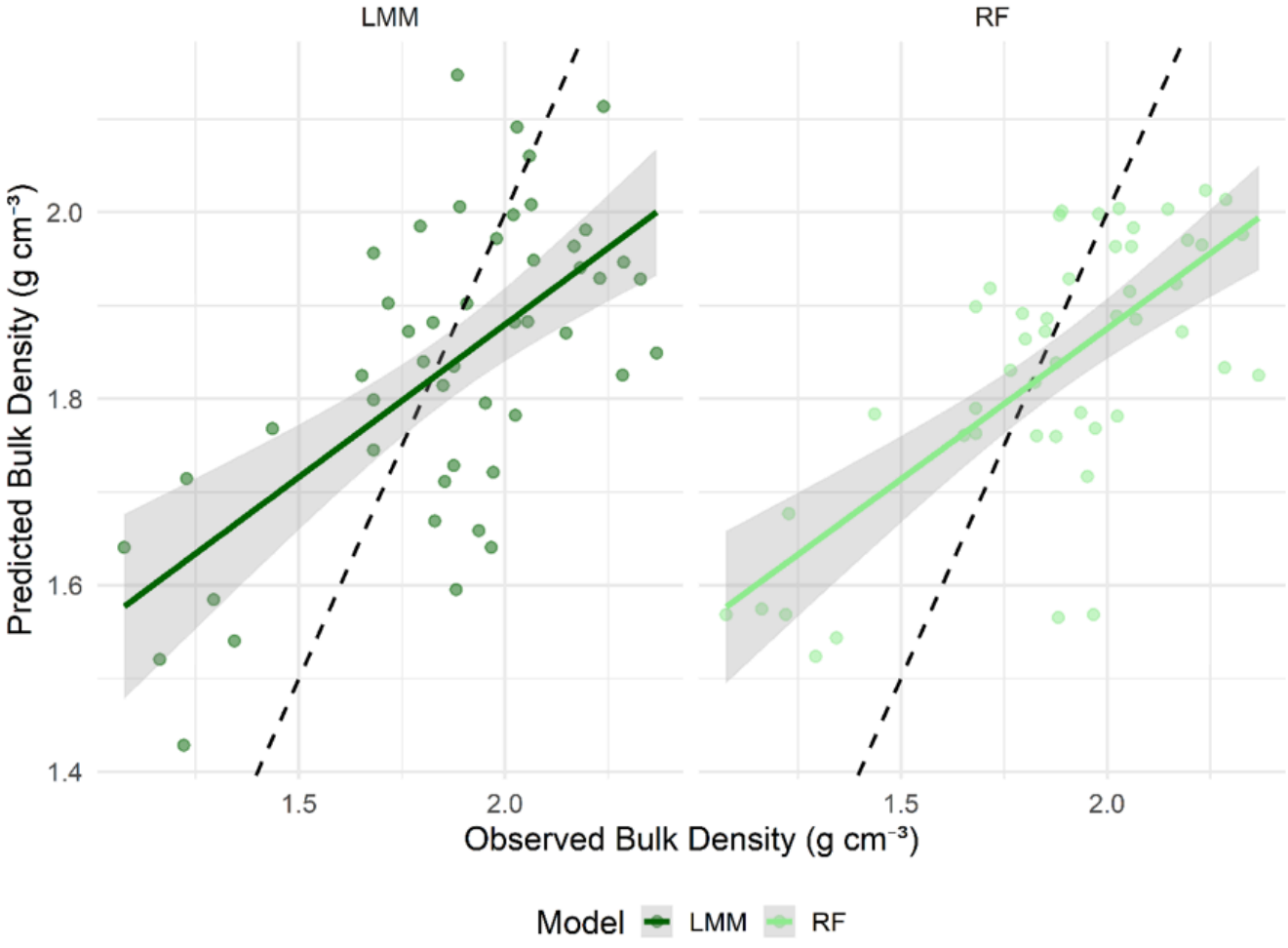


Figure 6

Observed vs predicted results of the LMM and RF models for estimating sediment BD.

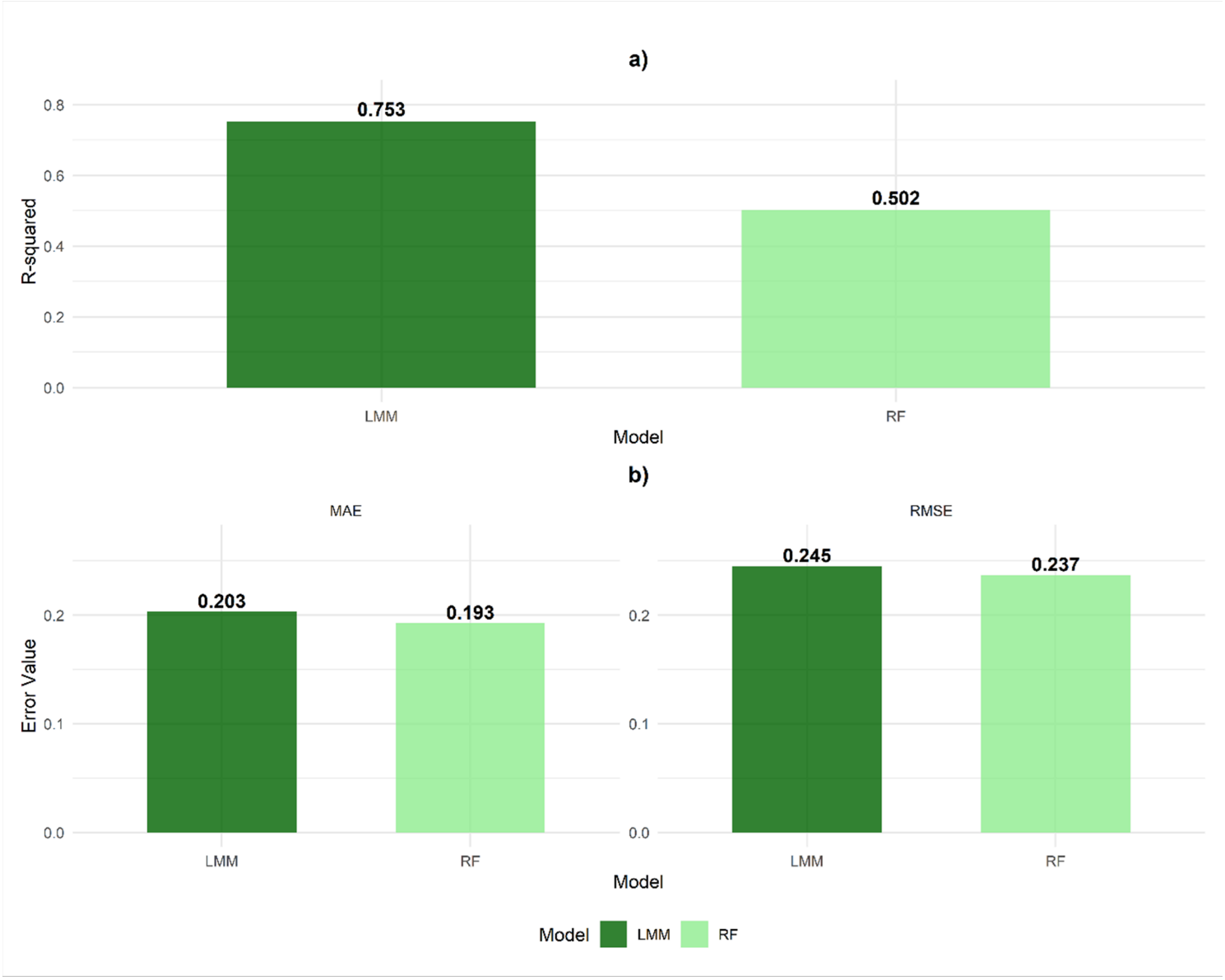


Figure 7

Model performance comparison between LMM and RF for estimating BD, showing prediction accuracy (R^2) and prediction errors (MAE and RMSE). Higher values of R^2 indicate better prediction accuracy, and Lower values of MAE and RMSE indicate better performance.

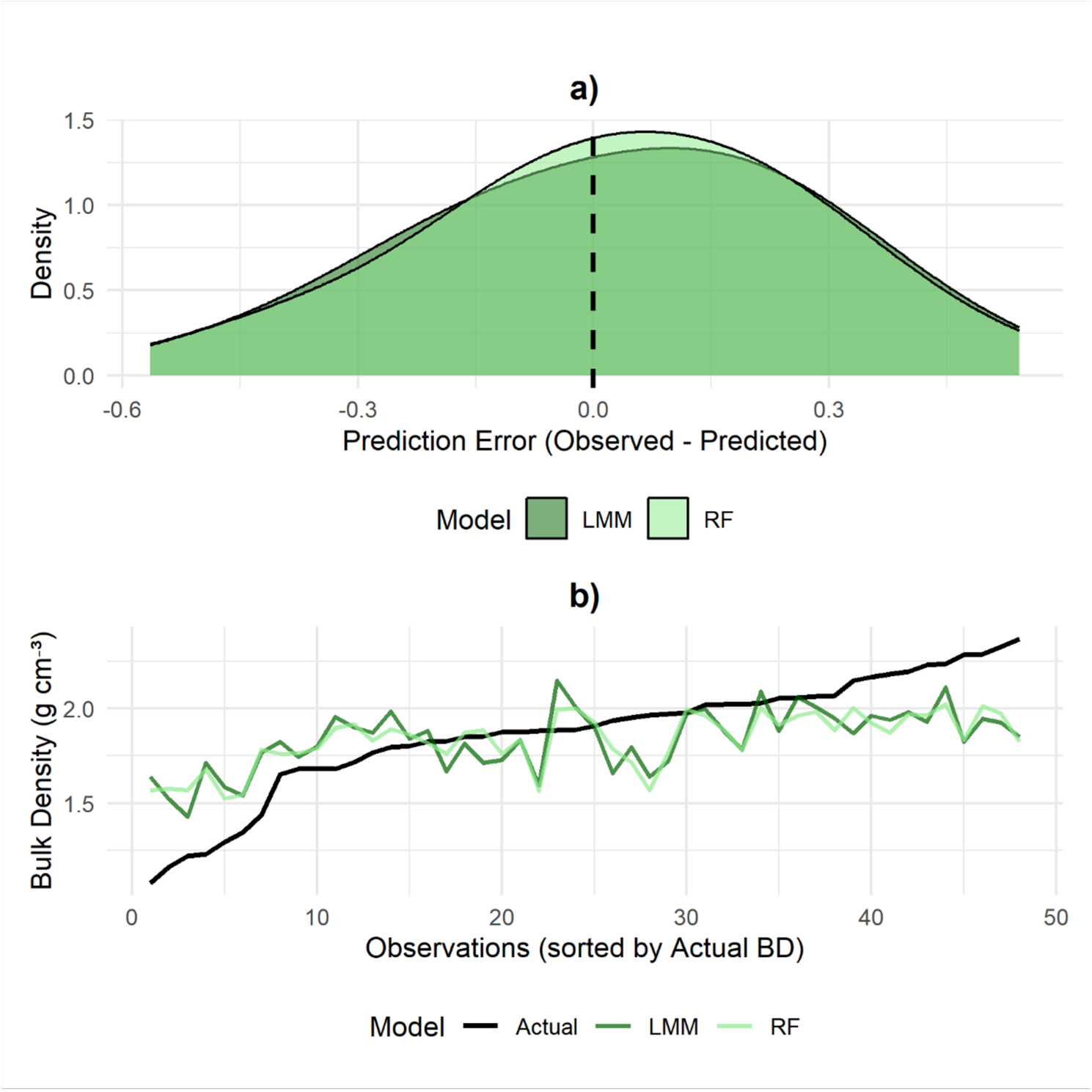


Figure 8

a) Comparison of prediction error distribution and b) Prediction tracking (sorted by actual BD) between LMM and RF for the estimation of sediment BD.

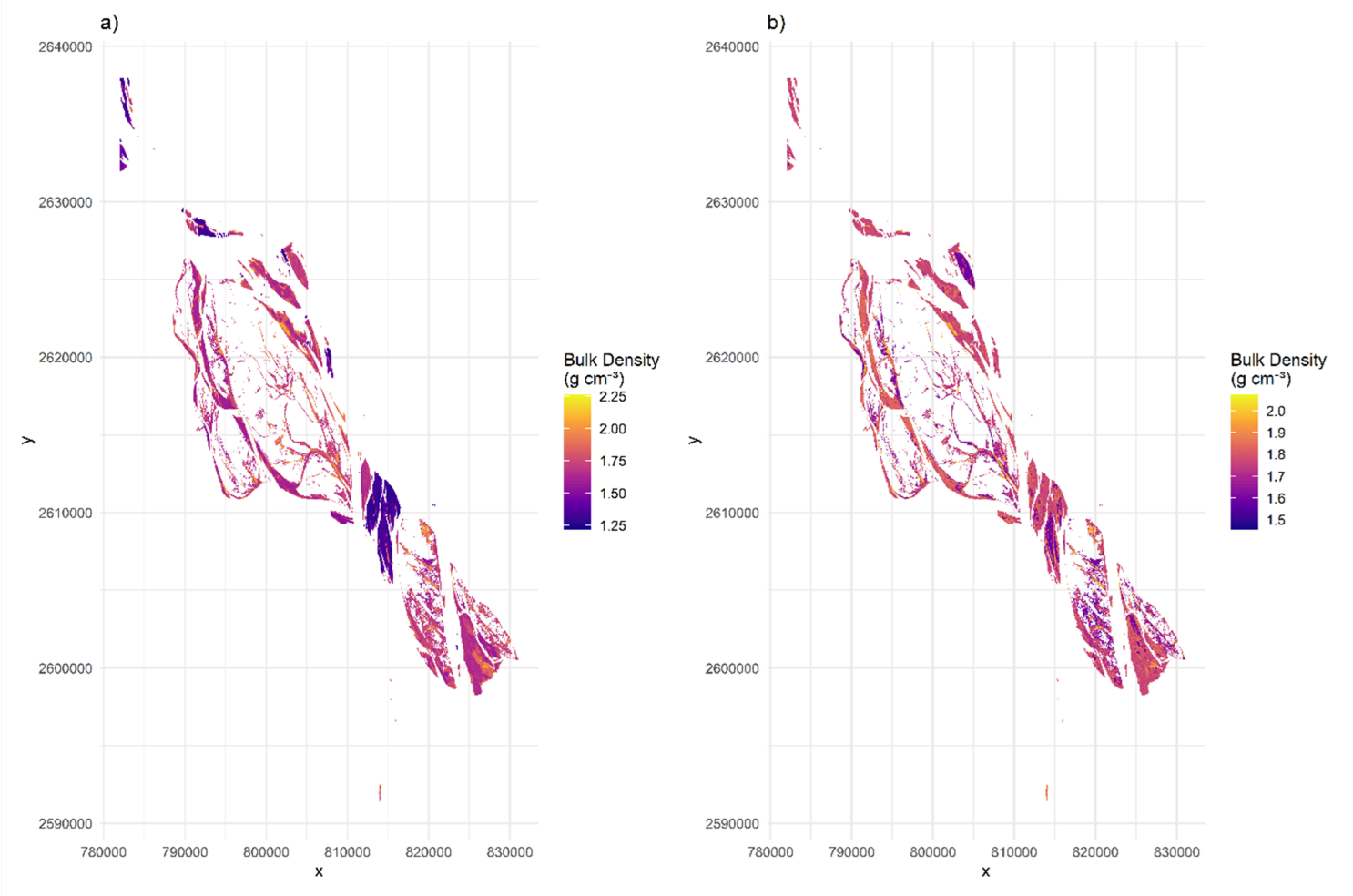


Figure 9

Application of the linear mixed model (LMM) and random forest (RF) for the estimation of sediment bulk density. a) LMM spatial prediction of sediment bulk density (g cm^{-3}); b) RF spatial prediction of sediment bulk density (g cm^{-3}).

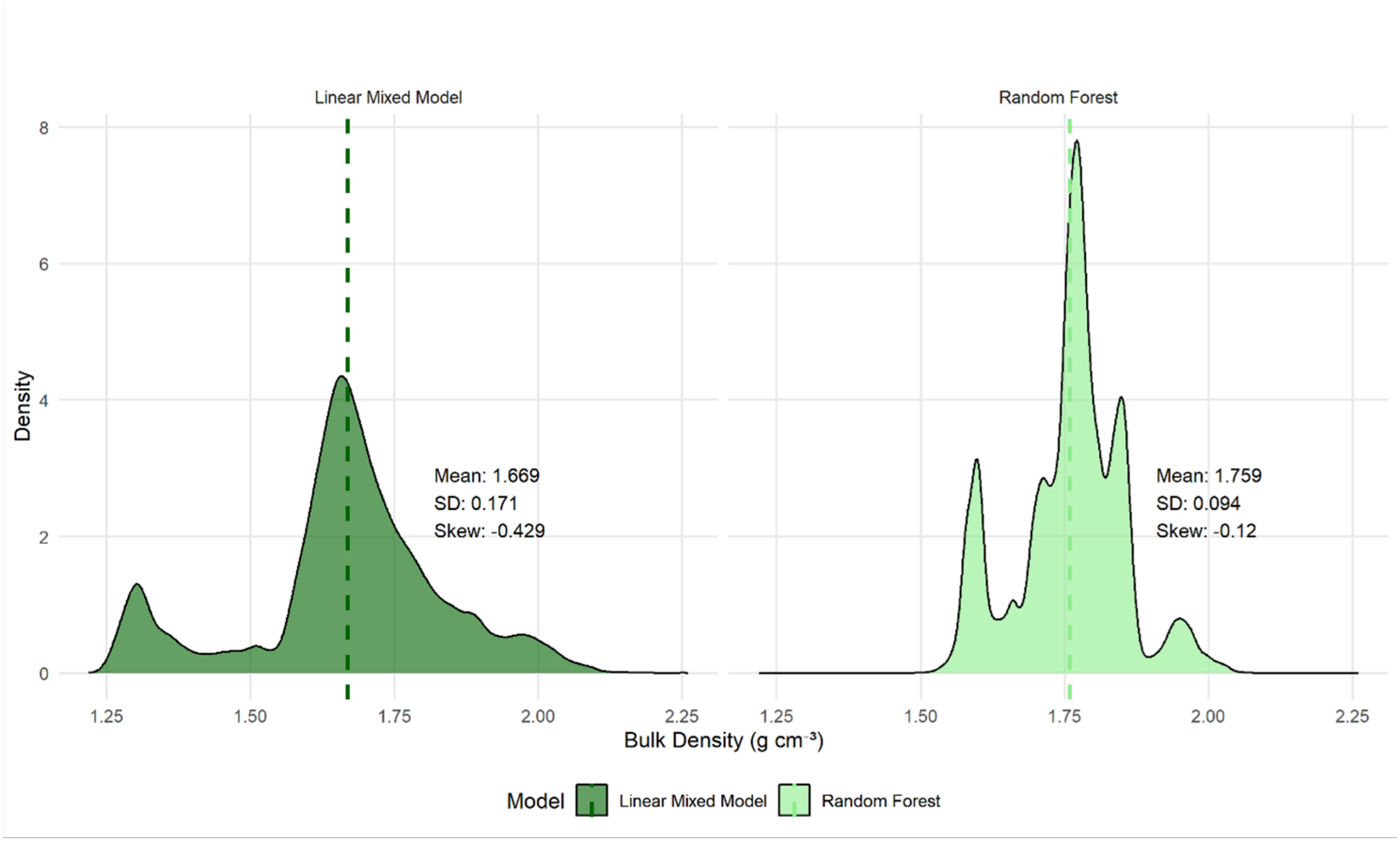


Figure 10

The frequency distribution of predicted BD values from an LMM and RF. The substantial overlap (87%) indicates both models capture the core spatial patterns, but RF better represents ecological heterogeneity and local-scale extremes.