https://doi.org/10.17221/128/2025-SWR

Balancing data quality in predictive geochemical mapping using machine learning: A Czech regional case study on topsoil nickel

Jan Skála, Daniel Žížala, Robert Minařík*

Research Institute for Soil and Water Conservation, Prague, Czech Republic

**Corresponding author: skala.jan@vumop.cz*

Citation: Skála J., Žížala D., Minařík R. (2026): Balancing data quality in predictive geochemical mapping using machine learning: A Czech regional case study on topsoil nickel. Soil & Water Res., 21: 89–97.

Abstract: Machine learning makes geochemical mapping highly adaptable, as its data-driven nature allows predictions to evolve with new information. In this study, topsoil nickel (Ni) data were compiled from various sources, each with different sampling times and analytical methods. To effectively use such imbalanced data into spatial modelling, it was necessary to test how the data uncertainty propagated through the final maps. A comprehensive benchmark of the quantile random forest algorithm was conducted to identify conditions under which the model performs optimally. Predictive maps of topsoil Ni at a 20-metre resolution were subsequently generated and compared using a multi-faceted evaluation strategy. This approach assessed how model adjustments – particularly those addressing the uncertainty introduced by the regression-based conversion of legacy measurements – affected the performance. Extensive benchmarking revealed that while out-of-sample validation showed only modest improvements (e.g., root mean square error (RMSE) reduced from 12.6 to 11.2 mg/kg) when modifying training data, covariates, or algorithm parameters, the resulting prediction grids differed substantially. The analysis also demonstrated that output variability across model scenarios occurred at different spatial scales: weighting approaches had localised effects, whereas high variability in the input data propagated more broadly across the region.

Keywords: data uncertainty; prediction maps; topsoil geochemistry

A key challenge in predictive soil mapping using machine learning (ML) is the scarcity of high-quality calibration data, driving growing efforts to recover and reuse legacy datasets (Richer de Forges et al. 2025). In the Czech Republic, a long-standing tradition of soil surveying provides valuable opportunities for integrating such legacy data into digital soil mapping (DSM) workflows (Borůvka et al. 2022). Over time, various institutions have compiled legacy datasets containing ancillary information relevant to the DSM, including geographic, geological, soil, and climatic variables. When combined with high-resolution digital elevation and remote sensing data, these datasets support the generation of accurate and interpretable digital soil maps (Wadoux et al. 2020). Machine learning has been used to map various soil properties in the Czech Republic (Borůvka et al. 2022; Žížala et al. 2022; Skála et al. 2025).

Predictive mapping of trace elements hinges on selecting covariates that can capture weak or non-linear relationships between environmental drivers and target elements (Kirkwood et al. 2016; Wilford

Supported by the Technology Agency of the Czech Republic (Project No. SS03010364) and the Ministry of Agriculture of the Czech Republic (Institutional support MZE-RO0223).

https://doi.org/10.17221/128/2025-SWR

et al. 2016). When conventional covariates fail to adequately represent the spatial variability – particularly across large and diverse regions – spatial proxies can serve as effective alternatives to enhance the ML performance (Hengl et al. 2018). However, geochemical calibration data are often sparse due to the cost and complexity of wet mineralisation. To address this limitation, legacy datasets from diverse surveys are often integrated, despite variations in the methodology and analytical rigour. This integration introduces heterogeneity in the data quality, which must be carefully addressed in predictive modelling (Malone & Searle 2021). Weighting methods are commonly employed to adjust models for varying degrees of reliability and measurement error in calibration data (van der Westhuizen et al. 2022; Richer-de-Forges et al. 2025). Nevertheless, few studies have systematically evaluated the performance of data-driven predictive algorithms under such heterogeneous data fusion conditions, particularly in the context of geochemical mapping. This study aims to assess how weighting influences the accuracy of predictive geochemical mapping using data-driven algorithms, specifically random forest. In the regional case study, a topsoil nickel (Ni) dataset was compiled by augmenting recent sampling data with legacy data to better represent the feature space of numerous covariates. However, the legacy data exhibited increased uncertainty due to longitudinal changes in the wet extraction techniques driven by technological and regulatory developments. Regression-based harmonisation steps aligned the datasets. Benchmarking assessed whether penalising regressed data could offset the variable reliability during calibration and whether spatial proxies enhance weighted model performance beyond environmental covariates. Nickel warrants detailed spatial modelling because its soil concentrations are controlled by highly variable geogenic sources and by diverse anthropogenic inputs. From an environmental perspective, Ni is also a regulated contaminant due to its potential phytotoxicity and human health risks, making reliable predictive mapping essential for both geochemical interpretation and risk assessment.

MATERIAL AND METHODS

Study area and topsoil data. This study investigates topsoil Ni concentrations across a geologically and environmentally diverse region of the Czechia, covering approximately 11 000 km^2 with 9 967 samples (Figure 1). The area includes metamorphic and

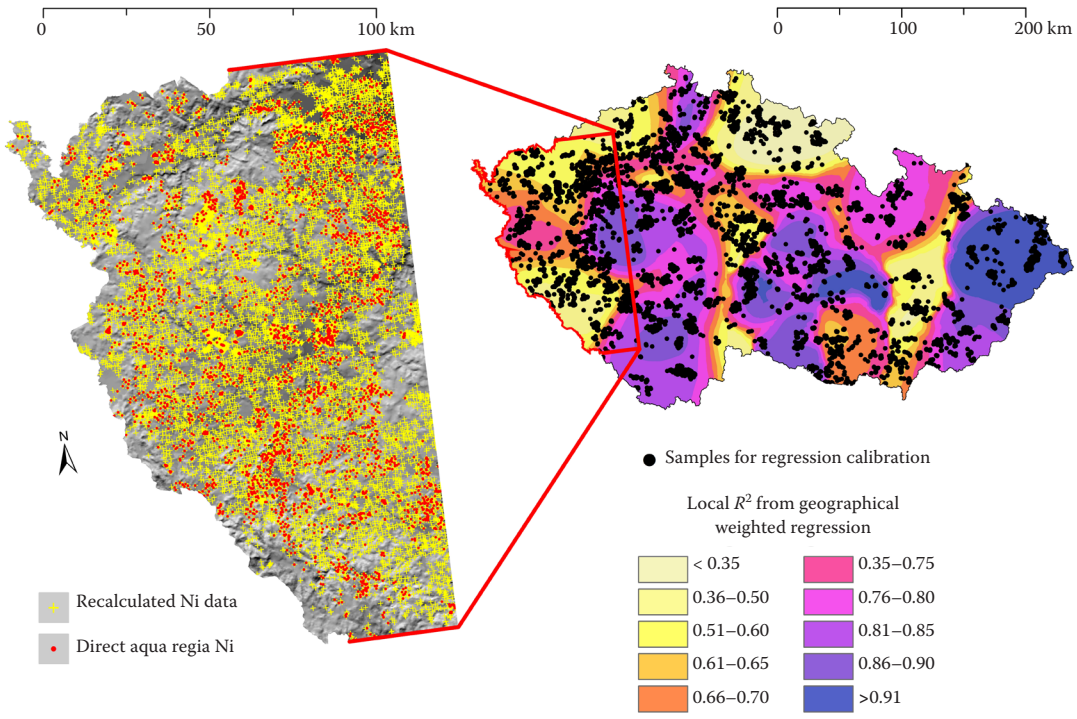

Figure 1. Spatial distribution of the training data, classified by direct analytical records and regression-based recalculations, with local reliability indicated by geographically weighted regression (GWR)-derived local R^2

https://doi.org/10.17221/128/2025-SWR

plutonic formations in mountainous zones (800 to 1 300 m elevation), historically linked to Variscan ore mineralisation, as well as sedimentary rocks and distinct lithologies such as the basaltic Doupov and gabbroid-rich Kdyně massifs. The region spans a broad anthropogenic gradient – from the mining-impacted Ore Mountains and the industrialised Pilsen agglomeration to low-impact zones near the Czech-Bavarian border.

The data for this study were sourced from nationwide soil monitoring programmes managed by the Czech Ministry of Agriculture, including the Registry of Contaminated Areas and the Monitoring of Contaminants in the Food Chain (Skála et al. 2025). While sampling protocols remained consistent, wet-mineralisation methods evolved, with legislative updates shifting practice from nitric-acid digestion to standardised aqua regia (Vácha et al. 2014). As a result, datasets now include results from multiple wet extraction techniques. A representative subset of 7 106 parallel measurements (Figure 1) enabled the integration of direct aqua regia concentrations and recalculated values into the spatial modelling workflow, harmonised using two distinct regression approaches:

(1) The first model used log-transformed Ni concentrations in a multiple linear regression with the pH, clay content, and organic carbon as covariates, fitted via ordinary least squares (OLS). The model demonstrated high reliability, achieving an R^2 of 0.90.

(2) The second harmonisation approach applied geographically weighted regression (GWR), which fits local linear models using a moving kernel (Brunsdon et al. 1998). We used an adaptive bi-square kernel in pseudo-stepwise mode via a GWmodel (Gollini et al. 2015). The GWR captured spatially varying relationships between the aqua regia Ni and predictors (nitric-acid Ni, pH, clay and organic carbon). The GWR improved the model fit, reduced the residual spatial autocorrelation, and provided spatial reliability estimates, which remained high across the study area (local R^2; Figure 1).

Environmental covariates. Environmental covariates were selected following the SCORPAN framework (McBratney et al. 2003), reflecting the soil-forming factors relevant to the surface geochemistry (Wilford et al. 2016). Relief was represented by eleven terrain derivatives from the 4th-generation Digital Terrain Model of the Czech Republic (DMR 4G, © ČÚZK) with a spatial resolution of 5 m per pixel. Climate data included long-term averages of temperature, precipitation, and evapotranspiration based on interpolated grids from the Czech Hydrometeorological Institute (CHMI), with a spatial resolution of 500 m per pixel. The soil and geology were captured using categorical covariates – soil types, parent rock, and texture classes from soil legacy maps (Žížala et al. 2022) and national geological maps (GEOČR500, Czech Geological Survey). To enhance the base model, additional covariates were incorporated to better capture the environmental variability:

(1) The lithology was supplemented with geophysical datasets from the Czech Geological Survey, including Bouguer gravity, airborne magnetic (ΔT), and gamma-radiometric total counts, provided at a 1 000 m resolution.

(2) The mining-related contamination was modelled using kernel density estimates of mine-waste deposits (© Czech Geological Survey), calculated with the spatstat R package directly at a 20 m resolution (Baddeley et al. 2015).

(3) The diffuse pollution was represented by long-term average deposition rates of Ni and PM_{10}, interpolated from the International Society for Knowledge Organization (ISKO) time-series data (https://www.chmi.cz/historicka-data/ovzdusi) at a 1 000 m resolution.

(4) The urban and traffic pressures were represented using night-time light imagery with native resolution ~500 m (Elvidge et al. 2021) a custom traffic-line density metric derived from vector traffic data at a 20 m resolution.

(5) The floodplain vulnerability was included using hydrological data from the DIBAVOD database (© VÚV T. G. Masaryka).

(6) The soil and landscape variability were characterised using a cloud-free bare soil composite from Sentinel-2 imagery with a spatial resolution of 20 m.

(7) The topographic complexity was addressed through multi-scale terrain metrics following Behrens et al. (2018), extending the base 20 m resolution with five octave levels.

(8) The land management effects were captured using a categorical grid of cropping intensity derived from the Land Parcel Identification System (LPIS) spatial data.

All the covariates were harmonised to a uniform spatial resolution of 20 × 20 m across the study area. Variables originally available at different resolutions were resampled using bilinear interpolation. Further details on covariate processing are provided in Skála et al. (2025) and Žížala et al. (2022).

https://doi.org/10.17221/128/2025-SWR

Generic workflow of model building. Random forest (RF), an ensemble-based ML method (Breiman 2001), is widely used in geochemical predictive mapping (Kirkwood et al. 2016; Hengl et al. 2018). In addition to its robustness to non-linear relationships, ability to model complex interactions, and insensitivity to outliers, RF is also highly suitable for our study because it is scalable and can naturally incorporate sample weights. In this study, RF models were trained in the quantile mode (QRF; Meinshausen 2006) using the caret package (Kuhn 2008) in R software. Model training was conducted under controlled conditions, with hyperparameter tuning limited to the mtry value – optimised via 20 iterations of random search (Bergstra & Bengio 2012). The number of trees was fixed at 500 to balance robustness and computational efficiency. The model accuracy was assessed using repeated 5-fold cross-validation, and the final model was fitted using the hyperparameter combination that minimised the root-mean-square error (RMSE), calculated as:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}} \quad (1)$$

where:

$\hat{y}_i$ – the prediction value for the target parameter;

y_i – the corresponding observed value;

n – the total number of validation points.

The final model performance was assessed using the coefficient of determination (R^2), calculated as the squared Pearson correlation between the observed and predicted values, and the RMSE (see Equation 1), which quantifies the average prediction error across the calibration and validation sets, which were defined conditionally based on the type of input data.

Model benchmarking to address input reliability and covariate variability. To address the RF's sensitivity to data artefacts, a multi-step benchmarking strategy assessed the model robustness across varying input reliability and covariate set-ups. Benchmarking involved three phases: (1) model calibration with 3 219 direct aqua regia samples; (2) training with all 9 967 topsoil Ni records, including the recalculated values; and (3) calibration using only converted data, reserving direct measurements for the validation. For mixed datasets, both weighted and unweighted QRF models were developed. Sample weights addressed any data balance, based on the regression reliability: the global R^2 from the OLS or local R^2 from the GWR. The direct aqua regia measurements received full weight (w_i = 1), while the recalculated values were weighted proportionally to their regression reliability (w_i = OLS global R^2 or w_i = GWR local R^2). Beyond the input reliability, the model performance depends on the explanatory strength of the covariates (Wadoux et al. 2020). To evaluate this, QRF variants were tested with: (*i*) baseline environmental covariates, and (*ii*) an extended set including spatial proxies, which may help compensate for incomplete feature spaces (Hengl et al. 2018). These proxies were generated using a modified buffer distance framework optimised for large datasets (Hengl et al. 2018; Žížala et al. 2020). Ni concentrations were split into 20 quantile classes, and the distances to the nearest point in each class were used as covariates.

A multi-way modelling strategy led to the calibration and evaluation of 17 models, each based on different input datasets and covariate configurations (Figure 2). Validation sets varied accordingly:

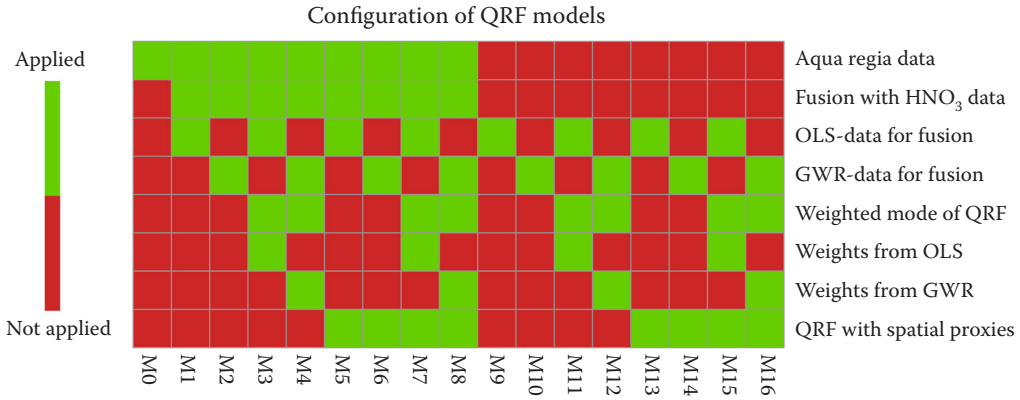

Figure 2. Overview of the trained models: input data composition, weighting strategies and covariate configurations
OLS – ordinary least squares; GWR – geographically weighted regression; QRF – quantile random forest

https://doi.org/10.17221/128/2025-SWR

Models 0–8, trained on non-inflated or mixed data, were validated using a 20% out-of-sample subset excluding converted data. Models 9–16, trained solely on converted data, were validated against the full set of direct aqua regia measurements.

Advanced cross-comparison of predictive surfaces. Predictive surfaces were generated at 20 m resolution using a tiled, parallelised approach. To compare model outputs, spatial concordance in continuous raster predictions were analysed by computing the mean absolute differences (MAD) for all the model pairs, summarising the average cell-by-cell deviation between raster layers x_i and y_i as:

$$\text{MAD}=\frac{\sum_{i=1}^{n}|x_i-y_i|}{n} \quad (2)$$

This metric quantified the concordance between) the model predictions across the spatial domain. To explore scale-dependent variation, comparisons were extended using the diffR package (Pontius & Santacruz 2023), aggregating raster values into windows that doubled in size at each step. To visualise local similarities and differences between selected model outputs, advanced raster algebra was applied within a moving window framework, using the relative width of the 90% prediction interval – calculated as $(Q_{95} - Q_5)/Q_{50}$ – where values were normalised to the median (Arrouays et al. 2014). This uncertainty metric served as the input for the focal raster correlation analysis using the terra package (Hijmans 2024), with Pearson correlations computed in a 19 × 19-pixel window to capture the spatial concordance while preserving sensitivity to the local variation.

RESULTS

Ranking the performance of model variants. Table 1 presents the performance metrics for all the model variants. The QRF models trained on direct analytical data or mixed datasets (Models 0–8) did not consistently outperform those based solely on the recalculated data (Models 9–16). While the recalculated-data models showed slightly higher RMSE, the differences were minimal. RF models are robust to noise, but sensitive to input data quality (Hengl et al. 2018). Our results suggest that augmenting training data with converted values had limited impact on the performance. Notably, QRF models using data converted via GWR slightly outperformed those using OLS in terms of the RMSE and R^2. Despite these gains, the R^2 values remained low across all the models, indicating that the covariates did not fully explain the spatial variability in the topsoil Ni concentrations. However, the spatial proxies consistently improved the performance modestly, with the most notable gains observed in spatially extended models calibrated using levelled data from multiple OLS regressions. The validation strategy also showed that the weighting input data in the QRF algorithm – using R^2 from regression levelling – had limited impact on the model performance, especially when spatial proxies were included. The top-performing model (Model 8) integrated direct analytical data with GWR-transformed inputs in a weighted QRF framework, enhanced by expanded spatial covariates.

Advanced comparison of predictive surfaces. Beyond global metrics, evaluating spatial patterns in predictive surfaces is crucial. Table 2 presents an adjusted matrix of the mean absolute differences between the output rasters, allowing a detailed comparison of the spatial agreement across the model variants. The lower triangle shows pairwise comparisons among the models calibrated with fused datasets (direct and converted data), while the diagonal entries compare the mixed-data models with their converted-data counterparts under identical configurations. The smallest mean differences – highlighted in bold italics – occurred between the weighted and unweighted models trained on the same data source (GWR or OLS) and using the same

Table 1. Performance metrics of the quantile random forest (QRF) models calibrated using various combinations of input data and covariates

Data fusion*			Converted data only*		
Model	R^2	RMSE	Model	R^2	RMSE
M0	0.21	11.8	–	–	–
M1	0.18	11.9	M9	0.15	12.6
M2	0.20	11.8	M10	0.21	12.1
M3	0.20	11.8	M11	0.15	12.6
M4	0.21	11.7	M12	0.22	12.1
M5	0.28	11.4	M13	0.24	12.1
M6	0.28	11.3	M14	0.28	11.8
M7	0.28	11.3	M15	0.24	12.1
M8	0.29	11.2	M16	0.28	11.7

RMSE – root mean square error; *the validation subsets differed between the two model groups

https://doi.org/10.17221/128/2025-SWR

Table 2. Cross-comparison of the mean absolute differences (MADs) between the Ni estimation grids from the model variants (Models 1–8); the diagonal entries show the MAD between the models trained solely on converted data (diagonal pair) vs. fused data under identical conditions (e.g., Model 1 vs. Model 9)

	M1	M2	M3	M4	M5	M6	M7	M8	Diagonal pair
M1	1.38								M9
M2	1.38	1.01							M10
M3	***0.45***	1.39	1.40						M11
M4	1.35	***0.45***	1.33	1.08					M12
M5	1.42	1.89	1.43	1.89	0.80				M13
M6	**1.95**	1.38	**1.95**	1.41	1.46	0.75			M14
M7	1.47	**1.92**	1.48	**1.91**	***0.39***	1.45	0.86		M15
M8	**1.95**	1.37	**1.95**	1.40	1.48	0.38	1.46	0.83	M16

Bold – highest MAD values; bold italics – smallest MAD values

covariate set (basic or spatially extended). The highest MAD values – bold in Table 2 –occurred when the models differed in multiple aspects, including the regression method, covariate set, and weighting.

The diagonal comparisons in Table 2 revealed moderate QRF sensitivity to bias from uneven data quality, particularly in the models relying solely on environmental covariates or OLS-based levelling. In contrast, the GWR-based models exhibited lower MAD values, indicating greater reliability of the conversion approach.

The multiscale analysis of selected outputs (Figure 3) showed that the input data differences affect the broad-scale patterns across the entire predictive surface. In contrast, changes in the covariate configurations produced more localised effects, and weighting had highly fine-scale impacts. These results suggest that the input data bias shapes the overall model structure, while the covariate enrichment and weighting offer limited corrective influence.

To further explore the spatial variability in prediction uncertainty, seven model variants with high divergence in multiscale comparisons (Figure 3) were analysed using focal correlation via raster algebra. The correlation rasters in Figure 4D–I show how the relative width of the 90% prediction interval

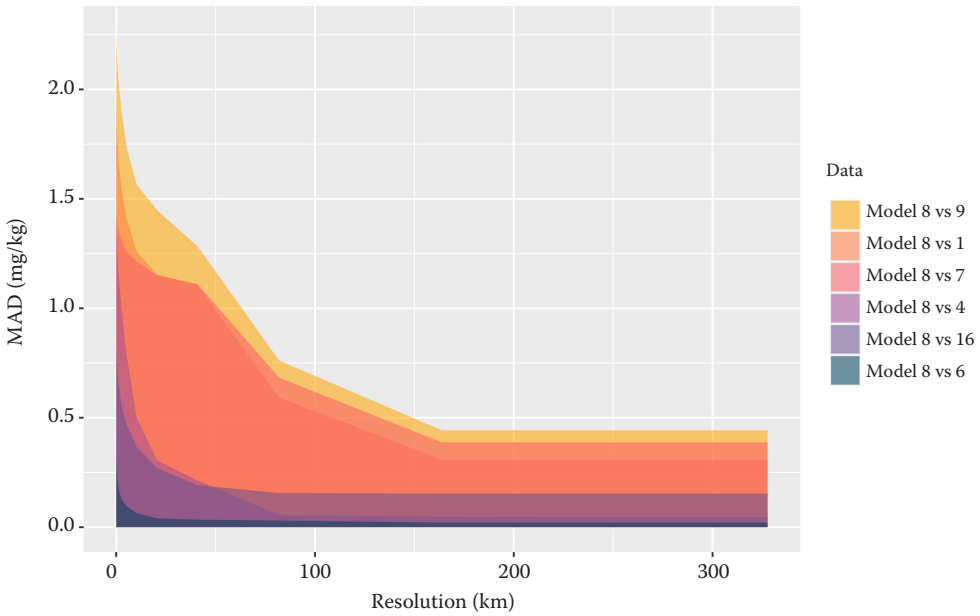

Figure 3. Mean absolute differences (MADs) between the selected grid pairs across multiple spatial scales, starting at the original 20 m resolution and doubling at each step

https://doi.org/10.17221/128/2025-SWR

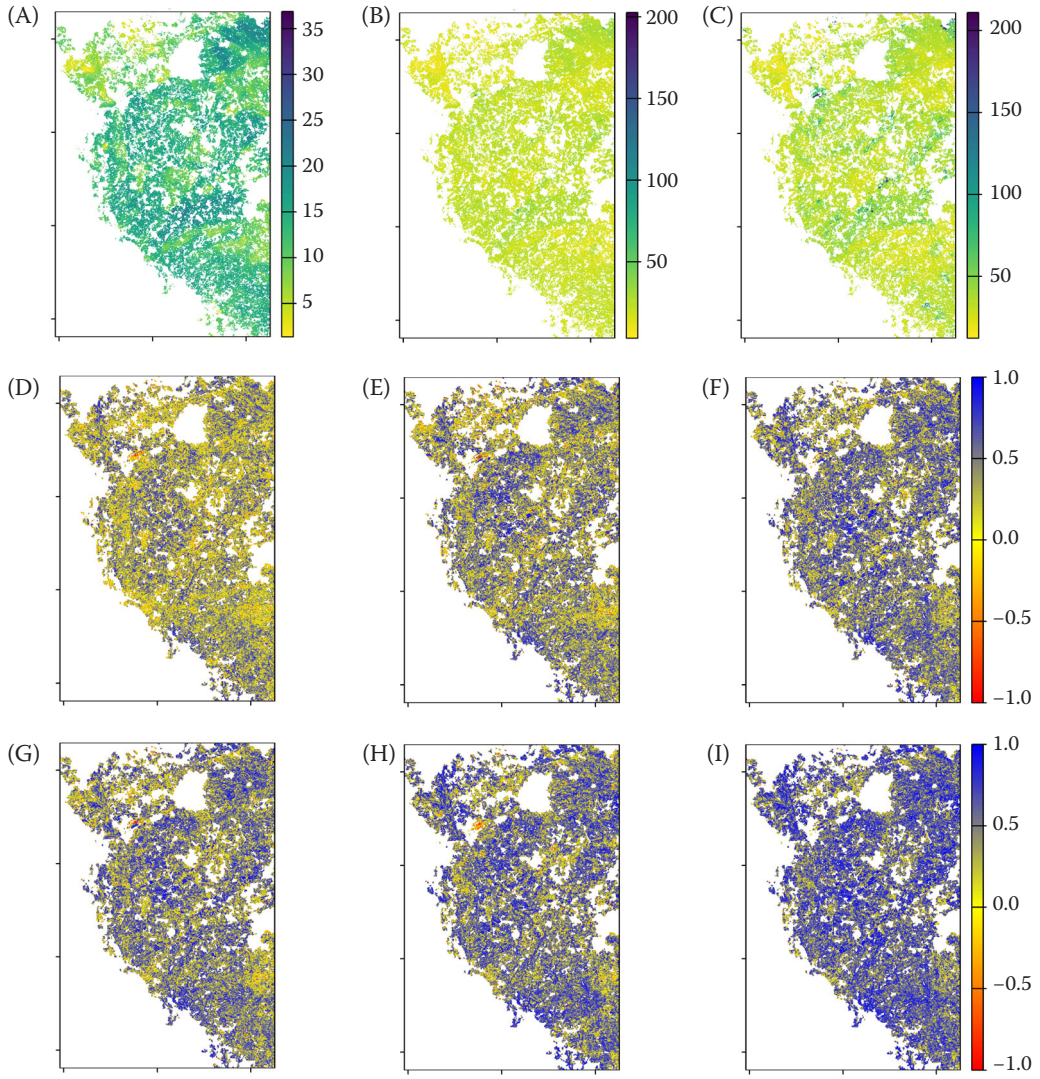

Figure 4. Predicted quantiles of the topsoil Ni content (mg/kg) from Model 8: Q_5 (A), Q_{50} (B), and Q_{95} (C); (D–I) show the focal correlation of the relative quantile spread from Model 8 compared to the other models: M9 (D), M1 (E), M7 (F), M4 (G), M16 (H), and M6 (I)

– expressed as the ratio of interval spread to the median prediction – varies spatially from the best-performing model (Model 8; Figure 4A–C). Figure 4 reveals several notable trends:

(1) The highest focal correlation was observed between the weighted and unweighted variants of the spatial QRF model using the GWR-converted data (Figure 4I).
(2) High concordance also occurred when comparing the weighted models using the same spatial QRF configuration, but different proportions of levelled data (Figure 4H).
(3) Moderate agreement was found between the models differing only in the covariate sets (Figure 4G).
(4) Lower concordance was noted when the models were based on data inflated using different regression techniques (GWR vs. OLS; Figure 4F).
(5) Marked divergence from the best-performing model was seen when multiple conditions were altered – specifically, when the competing model used different regression methods, modified covariates, and omitted weighting (Figure 4E and D).

DISCUSSION

The validation metrics (Table 1) show that weighting had little effect on the model performance, with only small differences across scenarios, consistent

https://doi.org/10.17221/128/2025-SWR

with Helfenstein et al. (2024). Weighting showed slightly greater influence when inflated data varied in reliability, particularly with GWR-based weights. Importantly, the uncertainty introduced through conversion did not significantly affect the prediction intervals (Figure 4H). The inclusion of spatial proxies modulated the influence of the input data. GWR-based weights had a stronger effect than the OLS-based ones, highlighting the need to assess the levelling method reliability as weighting alone appears insufficient to correct the data bias or artefacts. The proportion of converted data also plays a role. Richer de Forges et al. (2025) demonstrated that incorporating field estimates into laboratory measurements of the clay content, weighted by uncertainty, can significantly affect the QRF predictions. The gradual inclusion of field estimates in high-uncertainty areas improved the model accuracy and revealed risks of over-smoothing, underscoring the need to optimise both the data proportion and integration method. In this study, the proportion of levelled Ni data was relatively high and emerged as an influential factor affecting the prediction intervals (Figure 4F). These findings suggest that input data weighting should be considered in ML models when substantial variability in penalising weights is justified. In our case, weights were derived from the quality of regression fits used during data levelling. Other studies have explored alternative weighting strategies. Richer de Forges et al. (2025) and van der Westhuizen et al. (2022) employed iterative procedures based on residual variance from paired laboratory and field measurements. In contrast, Malone and Searle (2021) avoided weighting altogether, applying bootstrap resampling to assess the uncertainty propagation from the systematic conversion of field measurements. Their method could help produce the generation of uncertainty-sensitive outputs while mitigating the limitations associated with fixed weighting strategies.

In addition to these methodological considerations, it is important to recognise that the Ni concentrations in the study area originate from both geogenic and anthropogenic sources, each exhibiting distinct spatial signatures. Although explicit source apportionment between geogenic and anthropogenic inputs was beyond the scope of this study, the differing speciation of Ni from these sources typically leads to contrasting spatial patterns: broad lithologically controlled structures versus more localised anthropogenic enrichments. The multiscale comparisons are consistent with this distinction, indicating that coarse-scale variability is predominantly geogenic, whereas fine-scale deviations more likely reflect the anthropogenic influence. A similar analogy can be expected for other transition metals, particularly Co and Cr, which commonly exhibit a coarse-scale pattern of the topsoil content governed primarily by the composition of their host rocks.

CONCLUSION

The benchmarking results revealed a high degree of consistency between the directly measured and converted Ni data, highlighting the potential of data fusion to enhance the model performance. However, this also emphasises the need for the careful selection of regression models for data conversion and the implementation of uncertainty-aware strategies to capture data variability. The analysis showed that algorithmic parameterisation – particularly the use of penalising weights – had a largely localised effect. In contrast, modifying the ancillary data, such as the inclusion of spatial proxies, led to more pronounced regional differences. Concordance between the prediction grids declined further when multiple factors were altered simultaneously, including the input data, covariates, and weighting schemes. These findings provide practical guidance for integrating soil legacy data into predictive modelling workflows, especially in managing potential bias. The benchmarking served as a critical foundation for developing a nationwide model, the results of which are now accessible via the knowledge-based portal www.soilpass.vumop.cz.

Acknowledgements. We thank to all the staff of the Central Institute for Supervising and Testing in Agriculture for their many years of work in the field for Czech agrochemical testing, and the Ministry of Agriculture of the Czech Republic for allowing the use of the data. We are also grateful to the Czech Geological Survey for providing nationwide geophysical data and access to up-to-date data on closed and abandoned mining waste facilities. We also thank the Ministry of the Environment of the Czech Republic for authorising access to nationwide geophysical data.

REFERENCES

Arrouays D., McKenzie N., Hempel J., de Forges A.R., McBratney A.B. (2014): Global Soil Map: Basis of the Global Spatial Soil Information System. London, CRC Press.

https://doi.org/10.17221/128/2025-SWR

Baddeley A., Rubak E., Turner R. (2015): Spatial Point Patterns: Methodology and Applications with R. New York, Chapman & Hall/CRC.

Behrens T., Schmidt K., MacMillan R.A., Viscarra Rossel R.A. (2018): Multiscale contextual spatial modelling with the Gaussian scale space. Geoderma, 310: 128–137.

Bergstra J., Bengio Y. (2012): Random search for hyper-parameter optimization. Journal of Machine Learning Research, 13: 281–305.

Borůvka L., Vašát R., Šrámek V., Neudertová Hellebrandová K., Fadrhonsová V., Sáňka M., Pavlů L., Sáňka O., Vacek O., Němeček K., Nozari S., Oppong Sarkodie V.Y. (2022): Predictors for digital mapping of forest soil organic carbon stocks in different types of landscape. Soil and Water Research, 17: 69–79.

Breiman L. (2001): Random forests. Machine Learning, 45: 5–32.

Brunsdon Ch., Fotheringham S., Charlton M. (1998): Geographically weighted regression. Journal of the Royal Statistical Society: Series D (The Statistician), 47: 431–443.

Elvidge C.D., Zhizhin M., Ghosh T., Hsu F.C., Taneja J. (2021): Annual time series of global VIIRS nighttime lights derived from monthly averages: 2012 to 2019. Remote Sensing, 13: 922.

Gollini I., Lu B., Charlton M., Brunsdon C., Harris P. (2015): GWmodel: An R Package for exploring Spatial Heterogeneity using Geographically Weighted Models. Journal of Statistical Software, 63: 1–50.

Helfenstein A., Mulder V.L., Heuvelink G.B., Hack-ten Broeke M.J.D. (2024): Three-dimensional space and time mapping reveals soil organic matter decreases across anthropogenic landscapes in the Netherlands. Communications Earth & Environment, 5: 130.

Hengl T., Nussbaum M., Wright M.N., Heuvelink G.B.M., Graler B. (2018): Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. PeerJ, 6: e5518.

Kirkwood C., Cave M., Beamish D., Grebby S., Ferreira A. (2016): A machine learning approach to geochemical mapping. Journal of Geochemical Exploration, 167: 49–61.

Kuhn M. (2008): Building predictive models in R using the caret package. Journal of Statistical Software, 28: 1–38.

Malone B., Searle R. (2021): Updating the Australian digital soil texture mapping (Part 2): Spatial modelling of merged field and lab measurements. Soil Research, 59: 435–451.

McBratney A.B., Mendonça Santos M.L., Minasny B. (2003): On digital soil mapping. Geoderma, 117: 3–52.

Meinshausen N. (2006): Quantile regression forests. Journal of Machine Learning Research, 7: 983–999.

Pontius Jr., R.G., Santacruz A. (2023): diffeR: Metrics of Difference for Comparing Pairs of Maps or Pairs of Variables. Available at https://github.com/amsantac/diffeR

Richer-de-Forges A.C., Chen S., Heuvelink G.B.M., van der Westhuizen S., Orton T.G., Bourennane H., Arrouays D. (2025): Does digital soil mapping prediction performance of soil texture improve when adding uncertain field texture estimates? A study based on clay content. Geoderma, 456: 117277.

Skála J., Žížala D., Minařík R. (2025): Machine learning for predictive mapping of exceedance probabilities for potentially toxic elements in Czech farmland. Journal of Environmental Management, 380: 125035.

Vácha R., Sáňka M., Hauptman I., Zimová M., Čechmánková J. (2014): Assessment of limit values of risk elements and persistent organic pollutants in soil for Czech legislation. Plant, Soil and Environment, 60: 191–197.

van der Westhuizen S., Heuvelink G.B.M., Hofmeyr D.P., Poggio L. (2022): Measurement error-filtered machine learning in digital soil mapping. Spatial Statistics, 47: 100572.

Wadoux A.M.J.-C., Minasny B., McBratney A.B. (2020): Machine learning for digital soil mapping: Applications, challenges and suggested solutions. Earth-Science Reviews, 210: 103359.

Wilford J., de Caritat P., Bui E. (2016): Predictive geochemical mapping using environmental correlation. Applied Geochemistry, 66: 275–288.

Žížala D., Minařík R., Skála J., Beitlerová H., Juřicová A., Rojas J.R., Penížek V., Zádorová T. (2022): High-resolution agriculture soil property maps from digital soil mapping methods, Czech Republic. Catena, 212: 106024.

Received: October 27, 2025
Accepted: March 10, 2026
Published online: April 24, 2026