Soil type and content of macro-elements determine hotspots of Cu and Ni accumulation in soils of subarctic industrial barren : inference from a cascade machine learning

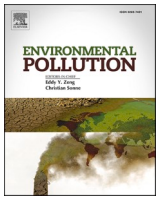
Environmental Pollution
Dvornikov, Yury; Slukovskaya, Marina; Gurinov, Artem; Vasenev, Viacheslav
https://doi.org/10.1016/j.envpol.2025.126457

Environmental Pollution 377 (2025) 126457

Contents lists available at ScienceDirect

Environmental Pollution

journal homepage: www.elsevier.com/locate/envpol



Soil type and content of macro-elements determine hotspots of Cu and Ni accumulation in soils of subarctic industrial barren: inference from a cascade machine learning☆

Yury Dvornikov [a,b,*], Marina Slukovskaya [a,c,d], Artem Gurinov [e,f], Viacheslav Vasenev [a,g]

[a] Smart Urban Nature laboratory, Peoples' Friendship University of Russia, Miklukho-Maklaya, 8/2, Moscow, 117198, Russia
[b] Laboratory of Carbon Monitoring in Terrestrial Ecosystems, Institute of Physicochemical and Biological Problems of Soil Science of the Russian Academy of Sciences, Institutskaya str., 2, Pushchino, 142290, Russia
[c] Tananaev Institute of Chemistry and Technology of Rare Elements and Mineral Raw Materials, Kola Science Centre, Russian Academy of Sciences, Apatity, Academic Campus 26a, 184209, Russia
[d] Laboratory of Nature-inspired Technologies and Environmental Safety of the Arctic region, Kola Science Centre, Russian Academy of Sciences, Apatity, Fersmana str. 14, 184209, Russia
[e] Laboratory of Geomorphology, Institute of Geography, Russian Academy of Sciences, Staromonetny Lane, 29/4, Moscow, 119017, Russia
[f] Faculty of Geography and Geoinformation Technologies, National Research University Higher School of Economics, Myasnitskaya str., 20, Moscow, 101000, Russia
[g] Soil Geography and Landscape Group, Wageningen University, Wageningen, 6700AA, Netherlands

ARTICLE INFO

Keywords:
Heavy metals
Cu
Ni
Digital soil mapping
Gradient boosting machines
Smelter impact
Aerial pollution

ABSTRACT

Aerial technogenic pollution from the activity of ferrous and non-ferrous metallurgy resulting in degradation of vulnerable natural ecosystems is a principal environmental problem in Russian Arctic. The industrial barren in the vicinity of Monchegorsk (Kola Peninsula) has been forming since 1950-s in the impact zone of the copper-nickel smelter. Soil heterogeneity, complete or partial degradation of vegetation, and rugged terrain intensified by soil erosion result in complex lateral spatial redistribution patterns of aerial depositions of Cu and Ni emitted by the smelter. In this research, we applied cascade machine learning (gradient boosting machines) to quantitatively describe these patterns. An extensive soil sampling campaign (n=506) across an area of 343 ha has revealed an extremely high levels of contamination (max bulk concentrations of Cu and Ni - 29.87 and 30.12 g/kg). We showed that soil types and the content of macro-elements (Ca and Fe) mapped based on the conventional set of predictors (topography, hydrology, landscape' spectral properties) explained spatial variability and especially hotspots of Cu and Ni contents with a higher accuracy compared to the models where interactions between macro-elements and heavy metals are not considered. This approach is a promising tool for mapping heavy metals' distribution in eroded, degraded, and highly polluted areas, which can be very useful to support land reclamation plans and allocate bioremediation measures.

1. Introduction

Aerial technogenic pollution and degradation of ecosystems (especially vulnerable Arctic and Subarctic environments) resulted from the activity of ferrous and non-ferrous metallurgy are an important ecological problems. According to Arctic Monitoring and Assessment Programme (AMAP) 2002 report, aerial emissions of heavy metals (HM) from the industry in Kola Peninsula largely contribute to overall HM pollution in the Arctic (AMAP, 2005). Copper-nickel (Cu-Ni) smelter in the vicinity of Monchegorsk town belonging to Kola mining and metallurgical company (MMC) (formerly «Severonikel») is among the leading global producers of nickel, copper, and cobalt and, therefore, one of the most significant sources of HM emissions in Northern Europe (Kashulina et al., 2014). The industrial barren (IB) around the Kola MMC today is considered one of the largest in the World (Kozlov and Zvereva, 2007). The influence of the smelter emissions on vegetation and soils of the IB as well as the effects of reclamation and remediation measures have been thoroughly investigated by several monitoring studies, survey campaigns, national and international projects summarized on the scheme (Fig A.1 in the Supplement).

☆ This paper has been recommended for acceptance by Dr Hefa Cheng.

* Corresponding author. Smart Urban Nature laboratory, Peoples' Friendship University of Russia, Miklukho-Maklaya, 8/2, Moscow, 117198, Russia.
E-mail address: dvornikov_yua@pfur.ru (Y. Dvornikov).

https://doi.org/10.1016/j.envpol.2025.126457
Received 31 December 2024; Received in revised form 22 April 2025; Accepted 12 May 2025
Available online 13 May 2025

It is likely that IB has been growing since 1950 because the production before this year was quite low, and later has increased twofold (Pozniakov, 1993), and IB patches could be visible over an area 2–3 km from the pollution source (to the North and South) according to prevailing meridional winds (Barcan and Kovnatsky, 1998). However, the extension of the impact area is uneven and is related to local topographic and landscape factors (Rees and Kapitsa, 1994). Nowadays, the impacted area continues to expand towards background landscape despite the decline of aerial emissions (Lyanguzova et al., 2016). This is associated with the low leaching rate of HM (100–200 years for Cu and 160–270 years Ni) from native soils (Barcan, 2002). The exact area of IB is debatable, but its patches can be traced up to 6–10 km from the pollution source (Kryuchkov, 1993).

For the first time, research on the direct impact of aerial pollution on soils and vegetation has begun after the establishment of a joint Russian-Norwegian pollution commission in 1988 (Norseth, 1994). Later, extremely high concentrations of Cu and Ni in soils of IB have been confirmed and various effects of aerial pollution were published (Evdokimova, 1995; Kozlov et al., 1993). Several attempts have already been made to investigate the spatial distribution and temporal dynamics of PTE (potentially toxic elements) in soils of the IB at sites located up to 75 km away from the smelter (Kashulina, 2018, 2017; Lyanguzova et al., 2016), and an exponential decrease in the contaminants' content with the distance from the pollution source was reported (Lyanguzova et al., 2016). Although the general patterns in spatial-temporal variability of soil pollution within IB were described, there are still gaps in quantitative assessments of local variations related to topography, soil types and properties. A detailed map based on proximal sensing and regression kriging was developed in 2020 (Dvornikov et al., 2022) for a pilot plot - two ha within the IB. Soil type and topographic wetness index (TWI) were the most important predictors of HM at 1.5 m resolution. This research aims to expand our 2020 attempts to a larger area of IB (~343 ha) by using the cascade machine learning approach. Machine learning (ML) methods for mapping HM and generally PTE in soils is largely applied in digital soil mapping (DSM) (Wadoux et al., 2020) since many processes in soils including HM accumulation and redistribution hold a certain non-linear character. Given the well-established guidelines for estimating the uncertainties of these methods, they appear to be promising if applied even to the relatively large areas. In highly polluted areas such as IB where the buffer capacity of soils is exceeded, the choice of optimal covariates for predicting the spatial distribution of HM in accordance with the traditional *scopran* principle (McBratney et al., 2003) is challenging (Yang et al., 2021). The spatial distribution of target variables is firstly influenced by the anthropogenic factors and can secondly be controlled by geomorphological processes: erosion, accumulation, and within-soil-processes. In these cases, a clear hypothesis on potential HM environmental drivers must be defined (Wadoux and McBratney, 2021) for avoiding the use of this method as a black box.

In this research, we attempted to predict two main contaminants in IB topsoil – bulk concentrations of Cu and Ni by applying ML with many predictors explaining the spatial variability of main landscape parameters (topography, vegetation, geology, geomorphology, and soil type/chemistry) at high (2 m) spatial resolution. We've applied a cascade ML, where content of macro-elements (Ca and Fe) in soils better predicted by conventional environmental drivers are further transferred to the set of predictors for explaining topsoil HM spatial variability that is less related to initial set of covariates. Our initial research hypotheses were as follows: 1) distance from the pollution source (anthropogenic factor) is not the leading factor controlling the spatial variability of HM in the topsoil; 2) soil pollution largely depends on soil type (organic matter content) and TWI (geomorphology) as shown in a previous study (Dvornikov et al., 2022); 3) HM concentration doesn't dependent on the geology and outcrops; 4) cascade ML with prediction of other elements (connectors) in soils (Ca and Fe) may provide a better accuracy of the final HM distribution maps if used as predictors.

2. Materials and methods

2.1. Study site description

The research area is located on the Kola Peninsula over the Polar Circle. The regional subarctic climate is considered cold with no dry season and cold summers - Dfc according to updated Köppen-Geiger classification (Beck et al., 2018). The native vegetation of the area around Kola MMC before the smelter establishment was dominated by northern taiga species: *Picea abies* and *Pinus sylvestris* (Manninen et al., 2015).

The study site (67.95 N, 32.82 E) is in the vicinity of Monchegorsk situated at 3.5 km to the west from the city and minimum at 967 m from the nearest pollution source within the impact zone of the smelter, max ~3.5 km (Fig. 1). The sampling area covers ~343 ha including both the smelter impact zone and area with partially preserved natural vegetation on the slopes of Travyanaya Varaka (hill); mapping area excluding lakes and anthropogenic objects is 329.5 ha. The studied industrial area is sparsely vegetated with the domination of *Salix* L. sp. and *Betula pubescens* Ehrh. trees (Koptsik et al., 2016). The site has a diverse meso-topography, including the top of the hill, hill slopes with rock outcrops, a few local depressions and surrounding moraine terrain - alternating hummocks, crests and hollows. The prevailing soil types are Skeletic Leptic Entic Podzol (Arenic, Toxic) (hilltop and slopes) (further – Podzol) and Dystric Rheic Hemic Histosol (Toxic) (local depressions) (further - Histosol), according to IUSS World Reference Base (IUSS Working Group WRB, 2015). Semi-natural soils of remediation sites are here identified as Skeletic Leptic Entic Podzol (Arenic, Organo-transportic, Toxic) (further here – Histosol/Podzol). Some patches of artificially deposited sands and gravel are distributed along anthropogenic objects. The previous research showed that the identified soil types differ in the capacity to accumulate the HM, and in the ratio of soluble and non-soluble forms of metals. In Podzol, Ni was mainly accumulated in non-soluble forms with the bioavailability of 1.5 %, while Cu was predominantly (60 %) available for plants (Slukovskaya et al., 2019). Metal bioavailability in Histosols was 19 % for Ni and 57 % for Cu (Slukovskaya et al., 2021).

2.2. Sampling design

Prior to the field work, we have defined a geospatial coordinates of sampling units (Fig. 1). Here, we define sampling unit as point in space within the study area where topsoil material was further collected, according to (de Gruijter et al., 2007). Two different strategies were selected for this experiment: 1) conditioned Latin Hypercube Sampling (cLHS) (Minasny and McBratney, 2006) to distribute sampling units in space for the model training (n=400); 2) spatially stratified random sampling (SSRS) to distribute the independent set of sampling units for the validation of final maps (map accuracy) (n=82). At each sampling unit, we have collected a composite of topsoil material in June 23–29, 2022. Each composite consisted of five aliquots bulked together (de Gruijter et al., 2007). Each aliquot represents a 100 cm^3 of soil material collected from the upper 5 cm of the soil profile by soil drill AM-7. A significant decline of HM concentrations below upper 5 cm of soil has been reported (Kashulina, 2017). We used a ~2×2 m envelope strategy to define the positions for collecting aliquots, where the geospatial coordinates of the sampling unit correspond to centre of the envelope. Since the planned mapping resolution was chosen to be 2 m (aligned to ArcticDEM resolution), samples were collected over the area corresponding to one pixel. The rationale of chosen sampling design is further discussed in Section A.1 in the Supplement.

During the fieldwork, location of 20 training sampling units has been shifted due to the hampered access to them: mean shift distance was 8.1 m, maximal - 22 m. Three training sampling units could not practically be accessed due to high water level in the vicinity of water bodies and were therefore excluded. In addition, we have randomly chosen 27

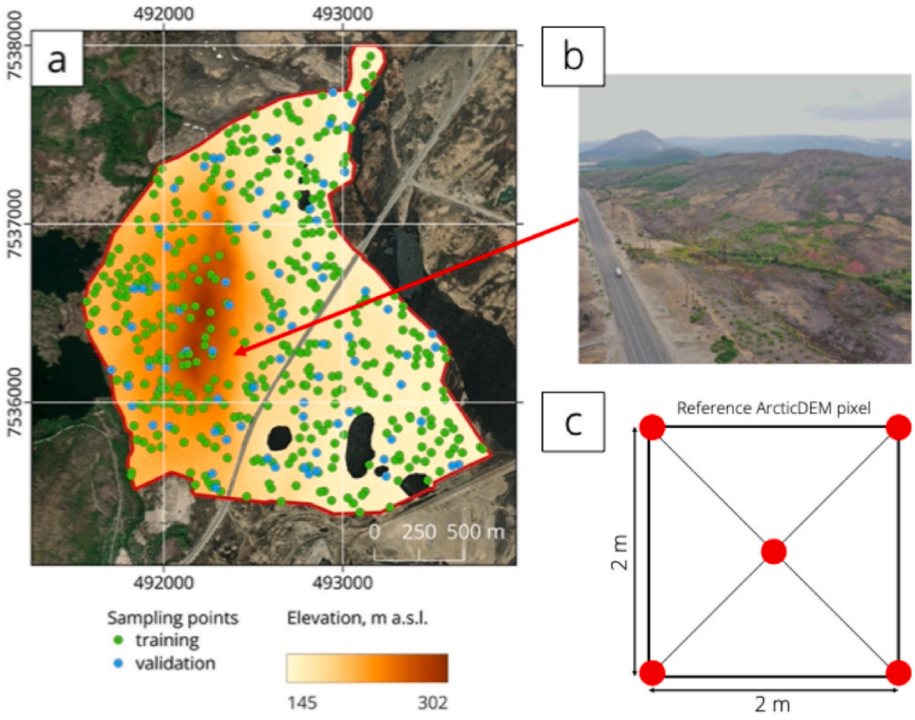


Fig. 1. Location of sampling points over the study area (red outlined): sampling units for model training were defined by cLHS, for validation – by SCSS (a), view on a Travyanaya Varaka Hill with patches of artificially planted vegetation (b); sampling unit scheme connected to ArcticDEM reference pixel (2 × 2 m): red dots indicate collected aliquots (c). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

sampling units for assessing short-distance variability of HM concentrations. These units were sampled twice (1–2 m away from the original sampling pixel location) by applying the same sampling strategy. In total, 506 samples were collected: 397 for model training, 82 for validation, and 27 for the assessment of short-distance variability.

2.2.1. Conditioned Latin Hypercube Sampling (cLHS)

We have used eleven raster ancillary data for cLHS: elevation, slopes, aspect, upslope contributing area (flow accumulation), flow direction, total curvature, TWI (Beven and Kirkby, 1979), topographic position index (TPI) and TPI-based landform classification (Weiss, 2001), geological formations, Euclidian distance from pollution sources, and land cover raster (Table A.1 in the Supplement). All topographic variables were derived from ArcticDEM at 2 m spatial resolution (Porter et al., 2018). Geological formations (intrusive, effusive) were manually digitized from 1:50,000 geological map of this region prepared in 1971, resulted from geological survey in this area undertaken in 1969–1972 (Zaytsevskiy, 1971). Distances from the pollution sources were calculated as raster data using *distanceFromPoints* function of the package *raster* in R environment (R Core Team, 2021). Land cover map was obtained by classifying ESA Copernicus Sentinel-2 composite image. Land cover map has also been used for masking anthropogenic objects and define areas with different prevailing soil types having the great impact of spatial HM variation (Dvornikov et al., 2022) (Table A.1). Vector data were rasterized and resampled to ArcticDEM resolution. We performed cLHS using *clhs* package (Roudier et al., 2012) available in R environment (R Core Team, 2021).

2.2.2. Spatially stratified random sampling (SSRS)

For sampling units assigned for the validation of the final map (test set), we have divided the entire study area into 41 hexagonal grid cells according to Uber H3 Grid using package *h3jsr* available in R environment. Within each hexagon (level=9), we have randomly selected two points (sampling units).

2.2.3. Laboratory analytics

All samples were delivered to certified Soil Ecological Laboratory of RUDN University (SEL RUDN, https://selab.rudn.ru/) where they were 1) dried at a room temperature, 2) partially grinded (peat samples), 3) sieved through 1 mm, 4) cleaned from plant remnants. All prepared samples were further filled and compacted into ~10 cm^3 plastic cuvettes and measured using Olympus© Vanta C portable X-ray fluorescence analyzer (pXRF) – a method that was tested for IB samples giving comparable results with ICP-MS measurements (Paltseva et al., 2022). All samples were measured three to four times (three measurements with ~1.5 min exposure, nine-twelve measurements per sample) with different placements of pXRF analyzer.

In addition, a set of samples (n=56) was measured again by pXRF to estimate an analytical error (Table A.2 in the Supplement). These samples were randomly chosen both from Histosol set (high organic matter concentrations) and Podzol set (wide range of concentrations). From the initial list of pXRF estimated elements, we have selected concentrations of Cu and Ni (main target variables) as well as Ca and Fe (auxiliary variables).

2.2.4. Pre-processing of measured concentrations

Pre-processing of laboratory analytics included subtraction of blank cuvette concentrations, estimations of internal pXRF error, variability of HM concentrations between measurements, variability between unit sub-samplings, and variability between sub-measurements (n=56) (analytical error) (Table A.2). We have re-measured one sample with standard deviation significantly exceeding 30 % of the mean value across three measurements and excluded the outlier. According to internal pXRF assessments, mean measurement error across all samples didn't exceed 1 %, maximal didn't exceed 6 %, 5 %, 2 %, and 1 % for Cu, Ni, Ca, and Fe respectively (Table A.2).

Mean analytical error was found to be 18.7, 14.8, 7.8, and 7.4 % for Cu, Ni, Ca, and Fe respectively (percentage of root-squared differences between two sets from the mean set's values) (n=56). It should be noted that highest differences (percentages) (>30 %) correspond to relatively

low concentrations (<1 g/kg for Cu and Ni). High concentrations are less sensitive in this case (Table A.2).

Measured concentrations were characterized by a high short-distance variability: up to 90 % for four elements (max – for Cu) (n=27). Mean root-squared difference from the mean of two values was found to be 29.6, 29.8, 11.1, 11.9 % for Cu, Ni, Ca, and Fe respectively (Table A.2).

We have also applied a correction coefficients to pXRF concentrations which is an accepted technique for obtaining accurate results (Peralta et al., 2020). A methodological work has recently been done for Olympus Vanta C analyser using the artificially polluted substrates with various concentrations of HM and organic matter (Romzaykina et al., 2024). Final concentrations of Cu, Ni, Ca, and Fe have been corrected by multiplying values with coefficients from these findings: 0.83, 0.88, 0.89, and 0.69.

2.2.5. Covariates

For model training, we added 23 more covariates based on ESA Copernicus Sentinel-2 satellite data. These variables included surface reflectance values in ten spectral bands (VIS-NIR-SWIR ranges) at native resolution of 10–20 m, Tasseled Cap Transformation (TCT) bands namely brightness, greenness, wetness, and ten spectral indices: widely used vegetation indices NDVI, SAVI, and EVI, modified normalized difference water index - MNDWI (Xu, 2006), bare soil index (BSI) (Diek et al., 2017), normalized burn ratio (NBR) (Key and Benson, 1999) since IB area is vulnerable to fires (Kozlov and Zvereva, 2007), clay-mineral ratio (CMR) (Alasta, 2011), plant senescence reflectance index (PSRI) (Merzlyak et al., 1999), red-edge chlorophyll index ($CI_{red\text{-}edge}$) (Gitelson et al., 2003), heavy metal stress sensitive index (HMSSI) (Zhang et al., 2018). TCT coefficients were derived from (Nedkov, 2017) and applied to Level-1C data (TOA). Other bands were calculated based on Level-2A (SR) data. All covariates were derived from the median July–August composites across 2021–2023 (six acquisitions in total – two per summer season of 2021, 2022, and 2023.

From the initial set of layers used in cLHS (n=11) we have excluded total curvature as having relatively stable values across study area and land cover map which was replaced by digital soil map prepared using our field survey descriptions and very-high resolution satellite mosaics. Flow accumulation raster was log-transformed. Additionally, we've added raster layer of geomorphons (Jasiewicz and Stepinski, 2013), Landsat-8/9 derived land surface temperature (LST) for late June 2023 as in (Ermida et al., 2020), snow depth estimation (similarly to (Meyer et al., 2022)), proxy of pollution distribution according to prevailing wind directions and exponential decrease of concentrations with distance from pollution source (Lyanguzova et al., 2016), and canopy height (Tolan et al., 2024). Thus, we have prepared 38 covariates compiled as a single multi-band raster file with spatial resolution of 2 m corresponding to ArcticDEM resolution. These covariates quantitatively describe the geology, geomorphology, soils, vegetation, hydrology, and anthropogenic factors potentially influencing the spatial HM distri in soils. For all predictors, equality of distributions between sampling points and the entire statistical population (SP) (raster values of the study area considered) was checked by T-test for numeric variables and Chi-Squared test for categorial variables.

2.2.6. Model tuning

Gradient boosting machines (GBM) is a widely applied ensemble algorithm based on decision tree but builds a numerous decision trees consecutively learning from the previous (Friedman, 2001). The algorithm doesn't require data normalization and performs well both with numerical and categorial variables. GBM requires several hyperparameters to be tuned: i) number of trees, ii) tree/interaction depth (number of nodes in each tree), iii) learning rate/shrinkage, iv) minimum observation in a node. We have tested a set of hyperparameters to reveal the best performing. To avoid overfitting, we used repeated (n=5) 10-fold cross-validation. We looped through interaction depths [3, 5, 10, 15], number of trees [500, 1000, 3000], learning rate [0.001, 0.01] and minimum observations in nodes [5, 10, 15]. Separate models were obtained for Ca, Fe, Cu, and for Ni with 38 predictors and 397 observations. Predicted values (Kempen et al., 2010) of Ca were used for Cu modelling, predicted values of Ca, Fe were used for Ni modelling (cascade ML). The selection of these macro-elements (Ca and Fe) relies on 1) their high concentrations in both dominating soil types (Ca in Podzols, and Fe in Histosols), and 2) their presence in aerial technogenic dusts and particles (Evdokimova, 1995; Slukovskii et al., 2020). Further, the high Ca content in Podzols over the study area may strongly differentiate these soils from Histosols with the highest organic matter content. In its turn, soil organic matter acts as an important agent in the processes of Cu accumulation due to the formation of stable complexes with humic substances (Zamulina et al., 2022). Therefore, we expect to capture this negative dependence between Ca and Cu. In contrast, Fe must have a connection with Ni as it acts as a siderophile (Vodyanitsky, 2008) on one hand, and might deposit from the technogenic iron sulphide particles (Slukovskii et al., 2020) on the other.

Model tuning, building, retrieval of relative influence have been performed using *caret* package (Kuhn, 2008) in R environment (R Core Team, 2021). We have assessed the uncertainty of all models by cross-validation RMSE ($RMSE_{cv}$). Both dependent variables (concentrations) and independent ones (predictors) didn't undergo any transformations prior to modelling. Workflow diagram of sampling, modelling, and validation is shown on Fig. A.2 (Supplement).

2.2.7. Mapping

Accuracy of maps has been evaluated through comparing predicted values and measured Ca, Fe, Cu and Ni bulk concentrations of a test set (n=82) using the widely used metrics (R^2_{test}, $RMSE_{test}$, and MAE_{test}) calculated using *MLmetrics* package in R environment. $RMSE_{cv}$ values obtained at the previous stage were considered for all predicted pixels as a metric of the uncertainty (Fig. A.2).

3. Results

3.1. Concentrations of Cu, Ni, Ca, and Fe

Bulk concentrations of Cu and Ni in a training set (*n=397*) were positively linearly correlated excluding high values (*R=0.72*) reaching max values of 29.87 and 30.12 g/kg (Fig. A.3 in the Supplement), their mean concentrations were 2.29 and 3.61 g/kg respectively while medians were 0.87 and 1.56 g/kg. Distributions are shown on Fig. A.4a in the Supplement. Concentration of Ca and Fe were less related to other elements with a more uniform distribution (Fig. A.4b in the Supplement). Their mean concentrations were 18.72 and 33.92 g/kg. We have observed a moderate inverse correlation between Ca and Cu (*R=-0.35*) and between Ca and Ni (*R=-0.29*), and a higher dependence between Fe and Ni (*R=0.53*) among the training set.

3.2. Distribution of predictor's values

Distributions of 36 predictors among 38 at sampling locations of training and training+test sets were statistically equal to the distribution of the SP (*p-value* varied from 0.059 to 0.983 for the test set only and from 0.117 to 1 for the training+test). For the training set, *aspect* and *flowdir* distributions differed from SP (*p-value* = 0.034–0.037, *t*-test). For the entire set of sampling points, *aspect* distribution differed from SP (*p-value* = 0.04, *t*-test).

3.3. Model outputs

GBM model for Ca was the most robust among the four target variables: Ca spatial distribution was in a strong relation with topographic, soil, and spectral properties of landscapes (surface reflectance in visible – near infrared – short-wave infrared ranges). With the best tuning

parameters [3, 5, 1000, 0.01] for the interaction depth, min observations in a node, number of trees, and shrinkage, the model has shown robust results at a cross-validation ($RMSE_{cv}$ = 4.85 g/kg) and validation with independent test set ($RMSE_{test}$ = 3.9 g/kg) (Table A.3 in the Supplement). The inclusion of a test set into training could further improve the model performance ($RMSE_{cv}$ = 4.63 g/kg), but not significantly.

The considered predictors (n=38) could explain spatial distribution of Fe but have shown the weakest relation to this target variable (Table A.3) among all four elements. The model tended to regress to the mean value underestimating highest values (Fe > 40 g/kg) and overestimating lowest values (Fe < 20 g/kg) due to (1) low spatial variability of Fe in the study area, and (2) inability of predictors considered to explain this spatial variability. However, the cross-validation metrics ($RMSE_{cv}$ = 11.27 g/kg) were close to metrics obtained by comparing with the independent test set ($RMSE_{test}$ = 8.52 g/kg) (Table A.3). The inclusion of a test set into training couldn't help to improve the model performance ($RMSE_{cv}$ = 10.69 g/kg) but slightly reduced the angle between the regression and 1:1 lines. Compared to Ca, the model for Fe needed a smaller learning rate (0.001) with a higher number of trees (3000) to reduce the total RMSE, but at the same time the larger number of observations in a node (n=15) (Table A.3) due to lower spatial variability.

Predicted Ca concentrations were included as an independent predictor to train the GBM model for Cu – one of two important target variables. More complex nature of relationships between Cu concentration in the topsoil and the environmental drivers has resulted in a high $RMSE_{cv}$ (2.34 g/kg) and $RMSE_{test}$ (2.37 g/kg) values (Table A.3) close to the mean value due to inability to reproduce mean concentrations of SP (highest uncertainty).

Although Ni had an even more complex spatial distribution compared to Cu as revealed for the initial model with predictors (data not shown), a stronger GBM model was obtained for Ni due to the inclusion of previously predicted values of Ca and Fe into the training. The inclusion of Cu into prediction (data not shown) has helped to gain higher quality metrics of the model ($RMSE_{cv}$ = 2.67 g/kg) but has resulted in weaker metrics on the test dataset ($RMSE_{test}$ = 4.0 g/kg). We have therefore decided to keep predicted Ca and Fe in the model for Ni ($RMSE_{cv}$ = 3.34 g/kg) characterized by not significant differences between cross-validation and test metrics. The highest model uncertainty can be observed for most observations (close to mean) whereas lowest and highest values are predicted relatively accurately.

3.4. Environmental drivers of topsoil macro-elements, Cu and Ni bulk concentrations

The influence of environmental drivers was considered for models created using the entire sample set (*n=479*). In general, fitted Ca, Fe, Cu and Ni concentrations were in a good agreement with observed concentrations: 80 %, 44 %, 83 % and 85 % of variance was explained (Table A.3). All models were characterised by a predominance of one predictor (see Table A.4 in the Supplement for details).

The distribution of Ca was mainly driven by soil type (belonging to Histosol units) (18.4 %) – lower concentrations, compared to other soil types. The increase in Ca concentrations coincided with the increase in the surface reflectance in green range (560 nm) between 5 % and 10 % (relative influence 4.9 %) and with the increase in the surface reflectance in near-infrared range (705 nm) but in a more complex way with peaks at 14 % and >32 % surface reflectance (relative influence 4.6 %) – impact of vegetation patches. Inverse relationship was observed with elevation – the highest concentrations at low levels (<170 m) (relative influence 5.6 %) and a complex relationship with TCT wetness (relative influence 5.0 %). It can therefore be concluded that Ca distribution was rather driven by natural factors (soil, vegetation, topography, and humidification regime).

Overall decrease in Fe bulk concentration in the topsoil coincided with the distance from the pollution source (from 1000 to 3500 m) (relative influence 13.8 %), with the increase in LST from 24 to 35 °C (relative influence 7 %), and increase of wind factor (>0.7) (relative influence 8.4 %). Similarly to Ca, surface reflectance in near infrared range impacted distribution of Fe (relative influence 11.5 %) – higher concentrations in wet vegetated patches. Inverse relationship was also observed for elevation (relative influence 4.5 %): higher concentrations at a lowest geomorphic position. It can be concluded that the Fe in topsoil was driven rather by non-natural factor (proximity and position relative to source of emissions). Relative influence of natural drivers – vegetation and landscape properties of Podzol-dominated areas (captured by LST), and topography (higher concentrations for lower geomorphic position of Histosol-dominated areas) has also been found.

Concentration of Cu was remarkably higher in Histosols (15.6 %). Inverse non-linear relationship between Cu and Ca concentrations was also captured by the model: a gradual Cu concentration decrease from 5 to 2 g/kg was observed at Ca concentration range from 10 to 24.5 g/kg (relative influence 9.3 %) – these concentrations are characteristic of eroded, sparsely vegetated Podzols (most of the study area). Other important predictors were from topography group: flow accumulation (log-transformed) and elevation (8.1 % and 6.9 %, respectively): these explained the spatial variability of Cu within Histosol-dominated areas (zones of accumulation). Surface reflectance in NIR2 (865 nm) range >28 % associated with vegetation patches (wet conditions) have shown a higher Cu concentrations compared to other Podzol-dominated patches (relative influence: 5.8 %). Further, an inverse relationship was observed with the distance from the pollution source (at values ranging from 1100 to 2200 m) (relative influence 5.4 %). At longer distances, no effect of this predictor could be seen. The Cu concentration was obviously driven by the organic matter content: all environmental drivers implicitly indicate this. Within eroded Podzols with degraded and eroded organic layer, macro-element Ca explain the spatial variability through the proportion of mineral component. Within Histosol-dominated areas, topographic patterns (accumulations) are also important for re-distribution of Cu in the topsoil.

Predicted Fe concentration has significantly helped to explain the variability of Ni concentration with a relative influence of 22.4 %: an almost linear gradual increase in Ni concentrations from 3 to 14 g/kg coincided with the increase of Fe bulk concentration from 30 to 55 g/kg. It was followed by soil type (Histosols – higher Ni concentrations compared to other soil types) – 14.3 %, and further by TWI – higher concentrations at non-slopes, Ca concentration - similar patterns to Cu, and elevation - also similar patterns to Cu – spatial distribution within Histosol-dominated areas) (4.8, 4.3, and 4.2 % respectively).

3.5. Final maps of Cu and Ni distribution and uncertainty

Map for Ca has demonstrated a uniform spatial distribution for the study area revealed by the model (Fig. 2a). Histosol-dominated landscapes and gully-bottoms on the Eastern slopes of Travyanaya Varaka Hill were characterized by lower concentrations. Vegetated stream valley patches at footslopes were characterized by concentrations above average (>30 g/kg). Average concentrations characterized most of the area covered by Podzols. Within Podzol patches the variation was driven by the vegetation, wetness regime and topographic patterns. Uncertainty was assessed by $RMSE_{cv}$ (4.63 g/kg) being robust at all concentration levels.

Patterns of high Fe concentration (>45 g/kg) are associated with peat distribution (predominantly Histosols) in the flat wet area in the South-East which is quantitatively described by LST (<27 °C) (Figs. 2b and 3). At the same time, the distance from the pollution source explained lower concentrations of similar patches in the North of the study area. Average Fe concentrations (~40 g/kg) were associated with small dense vegetation patterns. Uncertainty was assessed by $RMSE_{cv}$ (10.69 g/kg), but mostly this metric describes the uncertainty of lowest and highest values.

The highest Cu and Ni contents (Fig. 2c and d) were associated with

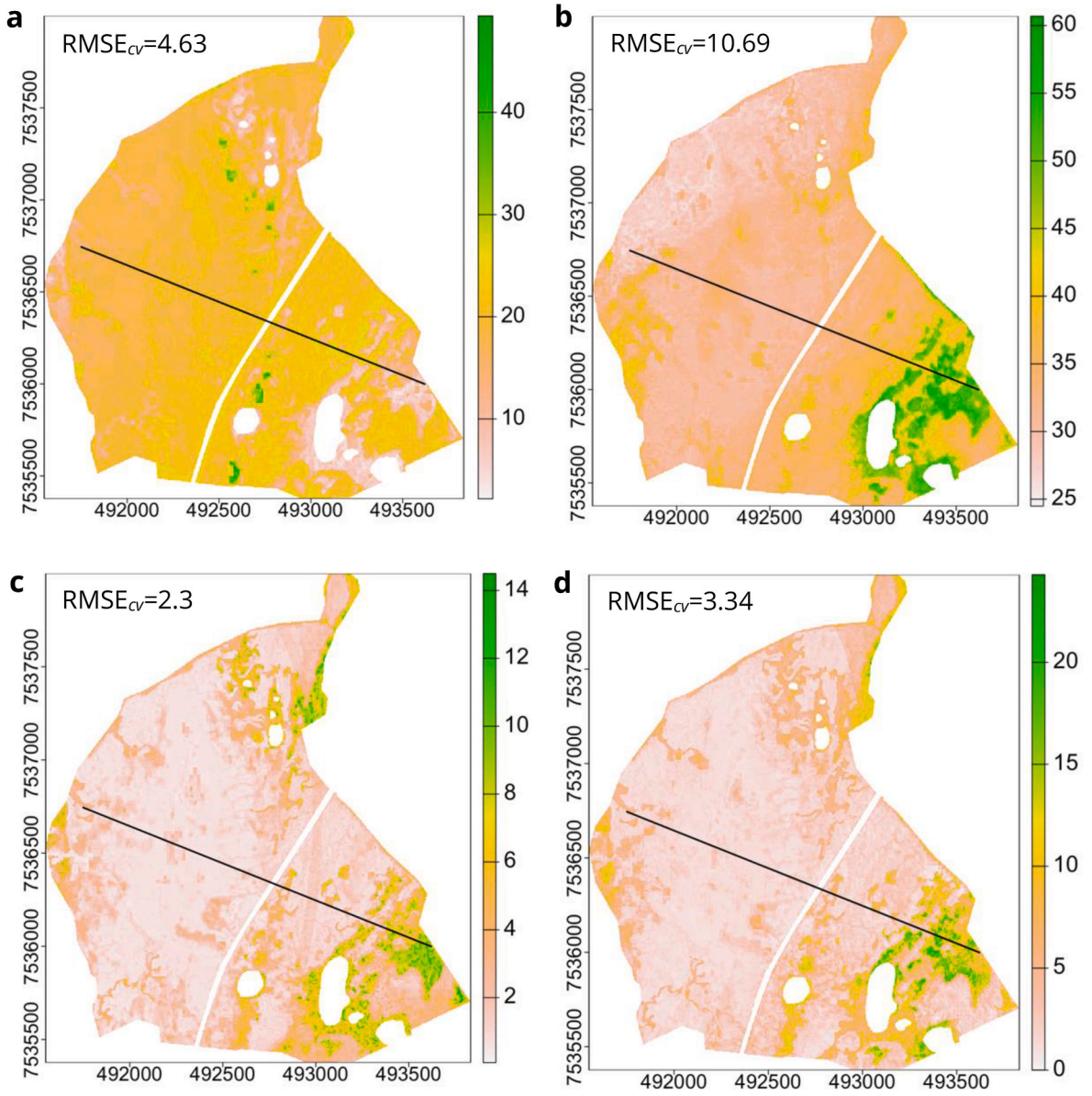


Fig. 2. Maps of target variables based on GBM models with a sample set (n=479): bulk concentrations of Ca (a), Fe (b), Cu (c), and Ni (d) given in g/kg. As solid shown is the profile line – from NW to SE (Fig. 3).

Histosols. Outside of these patches, the distribution of Cu was associated with Ca, while Ni – with Fe (Fig. 3). Models were unable to reproduce the highest concentrations of Cu and Ni, observed in the laboratory. The highest uncertainties are fair for the lowest concentrations, and it was estimated by RMSE$_{cv}$ for both elements (Table A.3).

4. Discussion

4.1. Soil contamination and environmental drivers of macro-elements, Cu and Ni concentrations in the topsoil of IB

4.1.1. Anthropogenic emissions and impact

Concentrations of Cu, Ni, and Fe in aerosols can reach 3.48, 33.1, and 52 g/kg (Opekunova et al., 2021) (Table A.5 in the Supplement). Although it is obvious, that the main source of elevated concentrations of these elements in soils is the deposition from the atmosphere, the relation of the topsoil concentrations with e.g. amount of precipitations has a more complex nature. Temporal monitoring of HM in the topsoil has shown a negative correlation between concentrations and precipitation amount of the year before sampling (Kashulina, 2018). Hyperbolic or exponential decline of HM concentrations in soils at a regional scale has been mentioned in many previous studies (Barcan and Kovnatsky, 1998; Evdokimova et al., 2011; Lyanguzova et al., 2016). If the local impact area of IB is considered, this effect of the distance from the pollution source seems to be more complex: many other factors like soil type and chemistry as well as vegetation and geomorphology do prevail in this case, as it was also suggested in (Kashulina, 2018).

We have found that only Fe has a distinctive relationship with non-natural re-distribution factors: distance and direction to pollution source, at least in the Northern direction. Native Al-Fe-humus Podzols are rich in Fe, but the situation is also hampered by the additional aerial technogenic deposition of Fe from the smelter in forms of pentlandite (Ni,Fe)9S8, pyrrotite Fe7S8(Nix), chalcopyrite CuFeS2 along with chalcosite Cu2S, covellite CuS, cuprite Cu2O, tenorite CuO, and metals Cu and Ni (Barcan, 2002). To a lower extent, the Fe concentrations are controlled by a natural factor (tree vegetation).

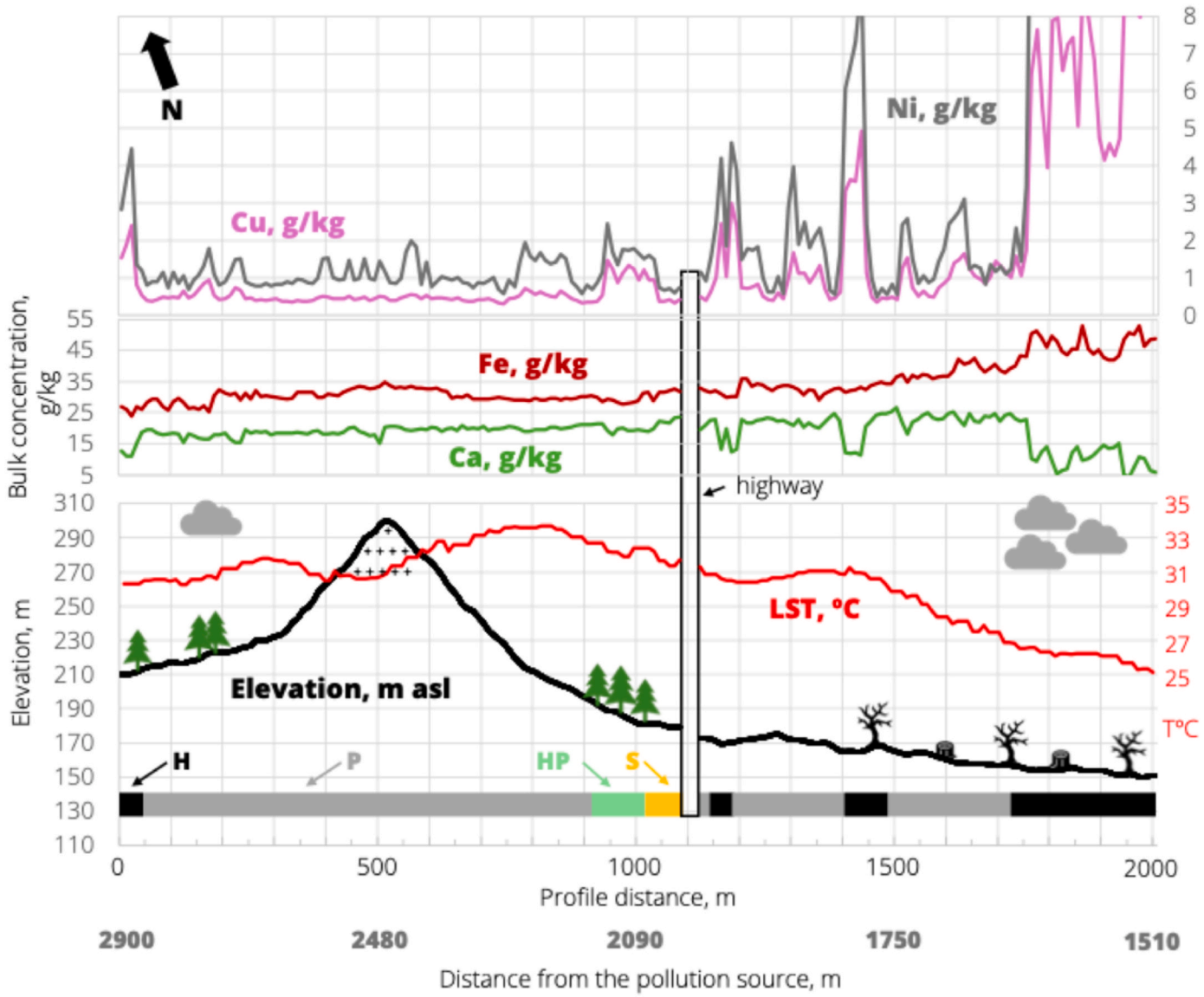


Fig. 3. Concentrations of pollutants with relation to other important environmental drivers along a representative two km profile within the impact zone of the smelter. Codes for soil types: H – Histosols, P – Podzols, HP – Histosol/Podzol, S – technogenic sand. Black crosses – parent rock outcrops. Profile line is shown on maps (Fig. 2).

4.1.2. Natural factors and re-distribution patterns

Maximal concentrations of Cu and Ni in the topsoil found in this research are comparable with those in Noril'sk ore (Boyd et al., 2009) and generally coincide with maximal concentrations found in other studies through ~40 years (Table A.5 in the Supplement). In general, bulk concentrations of Ni exceed concentrations of Cu both in Podzols and Histosols which seems to be the same even outside the impact area of the smelter (up to 47 km) (Lyanguzova and Barcan, 2019).

Bulk concentrations of Cu and Ni in the topsoil within IB are not associated with intrusive geological formations (Zaytsevskiy, 1971): the modelling results have therefore supported (median Cu and Ni sampled close to outcrops are 0.41 and 1.47 g/kg, respectively) our initial hypothesis stated in the beginning of this study. This was also partially confirmed in (Barcan, 2002).

In earlier publications devoted to soil pollution in the impact area it was argued that maximal proportion of pollutants belong to illuvial horizons of native soils (including the upper part of the subsoil) – 85–95 % (Barcan and Kovnatsky, 1998) and later - in the topsoil – organic horizon O or eroded illuvial horizon BFer (>80 % of acid-extractable forms of HM) (Elsukova et al., 2019; Kashulina, 2017; Lyanguzova et al., 2015). This is especially true for Cu content in organic-rich non-eroded Histosols since Cu is a stronger complexing agent and can be leached less effectively (Evdokimova et al., 2011). This pattern has been clearly captured by our GBM model for Cu. However, the reduction of organic matter content in the topsoil of Histosols may trigger the migration of a large portion of HM down the soil profile (Kashulina, 2017). In this study, we've found an inverse relation of Cu with Ca whose concentrations are considerably lower in Histosols, because 1) Histosol has a lower mineral proportion, and 2) Ca served as an alkaline geochemical barrier for metals deposited from atmospheric emissions (Slukovskaya et al., 2021). Within Histosol units, Cu and Ni are further re-distributed by topographic factors (flow accumulation and TWI) – concentrations of these elements are lower on slopes, as it was also shown in our pilot study (Dvornikov et al., 2022). Within Podzol-dominated landscapes Cu is strongly controlled by Ca which has a clear natural spatial variability. Forested areas on the western slope of Travyanaya Varaka Hill are characterized by lower Ca concentration compared to impacted eroded vegetation-free Podzol-dominated areas (Fig. 3) being absorbed by plant roots. Within them, no relations between Ca and Cu could be observed. At the same time, young vegetated

patches, often in a wet condition, previously described as «birch transitional community» or «birch woodlands» (Kozlov, 2001), are characterized by highest Ca concentration (>30 g/kg). These patches could confidently be captured by near-infrared band (>32 % surface reflectance) of Sentinel-2 composite. Here, the Ca distribution impact on Cu is high and was clearly captured by a GBM model. High topsoil concentrations of Cu within all vegetated patches can be related to developed A0 organic horizon of Podzols serving as a biogeochemical barrier (fixation) for migration of pollutants down the soil profile (Lyanguzova et al., 2015). Another (or additional) explanation of higher HM concentrations within those patches can be their higher load in spring due to thicker snow trapped by vegetation patches. The concentrations of Cu/Ni were in a good agreement with winter deposition amount of these elements (Kashulina, 2017) and the concentration of HM in the snow cover (Kashulina, 2018). However, we haven't found any relationships between estimated snow thickness and concentrations of Cu and Ni in our study area. In a vegetation-free Podzol-dominated landscapes, Cu and Ni might undergo further migration down the soil profile.

Spatial distribution of Ni (Fig. 2d) within IB has more complex patterns, than for Cu. Geochemically, Ni element is a siderophile (Vodyanitsky, 2008). In IB, the bulk content of Ni in the upper organic horizon exceeds the bulk content of Cu whereas acid-extractable forms of these metals demonstrate an opposite tendency (Lyanguzova et al., 2016). Lower ability of Ni to form complexes with humic and fulvic acids suggests a lower influence of soil type on the Ni spatial distribution captured by the GBM model. Spatial variation of Ni correlates with Fe to a larger extent explained by the distance to the pollution source and LST (from 24 to 35 °C) – these temperatures are recorded in Podzol-dominated landscapes since lower temperatures (21–24 °C) correspond to waterlogged depressions occupied by Histosols with incomparable high concentrations of Fe (Figs. 2b and 3). The relationships between Ni and Fe contents can be explained by the absorption of technogenic magnetite Fe3O4 particles within Podzols (Lyanguzova et al., 2016) at a surface layer (topsoil). Interestingly, none of the geomorphological predictors (TPI or geomorphons) had a significant relative influence in our models. The geomorphological impact was rather expressed by means of other variables: for degraded Podzols covering the larger proportion of the study area, the intense erosion (by wind and water) seems to promote the further redistribution of pollutants within the landscape.

4.1.3. Conceptual model of HM contamination within IB

In our previous pilot research, we have shown that in a small plot (two ha) within a highly degraded landscape, Histosol-dominated areas are characterized by maximal Cu and Ni concentrations. Within this soil type unit, the distribution of these pollutants is controlled by TWI (Dvornikov et al., 2022). This study has confirmed these patterns. However, the distribution of HM pollutants within Podzol-dominated landscapes, that predominate in this region, has a way more complex structure. Cu as largely controlled by Ca whose spatial variability has a clear natural background and indicates the organic matter content in soils. In these areas, we expect the migration of atmospherically polluted Cu down the soil profile into illuvial horizon as well as down the lower geomorphic levels. For this reason, no relations between anthropogenic controls (distance and direction from the pollution source) and Cu could be captured by GBM model. And that is not the case for Ni, that depends on Fe concentration to a large extent. In its turn, the concentration of Fe is controlled by a direct anthropogenic impact (deposition from aerial emissions). Scarce vegetated patches have the own impact: they are indicative for elevated Cu and Ni due to an additional load of organic matter as well as Fe from aerial deposition, and the variability of Ca doesn't explain the variability of Cu anymore here, except in the «birch woodlands».

We have also found a high spatial variability of HM pollutants at short distances (Table A.2 in the Supplement). This hampers the accurate estimations of HM at a high resolution, but the major patterns could still be captured (Fig. 2). The conceptual model presented in this paper, can further help to undertake a cost-efficient measure for the remediation of IB.

5. Conclusions

Two main conclusions can be drawn out of this research. First, it is a methodological conclusion. In highly polluted areas, prediction of HM concentrations must be done with the consideration of other macroelements in soils that, in its turn, are explained by conventionally used covariates related to topography, hydrology, and spectral properties. The ML-DSM approach in this case must first be applied to elements explained by natural (or anthropogenic, if considered) factors (in our case Ca and Fe). Further, predicted values of these elements must subsequently be used as new covariates for predicting aerially transported pollutants (HM, Cu and Ni in this case). Such cascade ML has helped to increase the accuracy of Cu and Ni maps and to reduce their uncertainty.

Second conclusion describes the within-landscape processes of HM deposition and redistribution. In this study, we were able to create a detailed picture of spatial distribution of main HM pollutants which is in a general agreement with patterns found in our earlier research (at some extent), but our GBM models could quantitatively describe the non-linear relationships between Cu and Ni concentration and other natural and anthropogenic drivers. These two pollutants had a different pattern of spatial distribution. Cu is rather controlled by natural factor (namely organic matter content). Ni is more strongly associated with anthropogenic emissions (through Fe). These complex quantitatively described HM spatial variability patterns have a clear practical application, as they allow to identify key areas of the territory that most require a bioremediation measures.

CRediT authorship contribution statement

Yury Dvornikov: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Marina Slukovskaya:** Writing – review & editing, Methodology, Investigation, Formal analysis, Data curation. **Artem Gurinov:** Writing – review & editing, Investigation, Formal analysis, Data curation. **Viacheslav Vasenev:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Soil survey and data analysis was supported by Russian Science Foundation Project N№19-77-30012. Database on the soil pollution is developing in accordance with the governmental assignment FSSF-2024-0023, and RUDN University Scientific Projects Grant System, project N№202414-2-000. We thank Victor Yakovenchuk and Andrey Vashkov (GI KSC RAS) for providing a legacy cartographic data, laboratory assistants Evgenia Avakyan (Saint-Petersburg Mining University), Polina Kosovskikh and Erlan Abdullaev (RUDN University) for soil samples' preparation and laboratory analytics, GIS-engineer Yana Egorova for preparing a digital geological map, Natalia Saltan (PABGI) for providing necessary literature data, and Gerard Heuvelink (WUR) for constrictive comments on sampling strategy, mapping, and quality assessment.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envpol.2025.126457.

Data availability

Data will be made available on request.

References

Alasta, A.F., 2011. Using Remote sensing data to identify iron deposits in central western Libya. In: International Conference on Emerging Trends in Computer and Image Processing. Bangkok, pp. 56–61.

AMAP, 2005. AMAP Assessment 2002: Heavy Metals in the Arctic. Oslo.

Barcan, V., 2002. Leaching of nickel and copper from soil contaminated by metallurgical dust. Environ. Int. 28, 63–68. https://doi.org/10.1016/S0160-4120(02)00005-3.

Barcan, V., Kovnatsky, E., 1998. Soil surface geochemical anomaly around the copper-nickel metallurgical smelter. Water Air Soil Pollut. 103, 197–218.

Beck, H.E., Zimmermann, N.E., McVicar, T.R., Vergopolan, N., Berg, A., Wood, E.F., 2018. Present and future Köppen-geiger climate classification maps at 1-km resolution. Sci. Data 5, 180214. https://doi.org/10.1038/sdata.2018.214.

Beven, K.J., Kirkby, M.J., 1979. A physically based, variable contributing area model of basin hydrology. Hydrol. Sci. Bull. 24, 43–69. https://doi.org/10.1080/02626667909491834.

Boyd, R., Barnes, S.J., De Caritat, P., Chekushin, V.A., Melezhik, V.A., Reimann, C., Zientek, M.L., 2009. Emissions from the copper-nickel industry on the Kola Peninsula and at Noril'sk, Russia. Atmos. Environ. 43, 1474–1480. https://doi.org/10.1016/j.atmosenv.2008.12.003.

de Gruijter, J., Brus, D., Bierkens, M., Knotters, M., 2007. Sampling for Natural Resource Monitoring. Springer, Berlin. https://doi.org/10.2136/sssaj2007.0018br.

Diek, S., Fornallaz, F., Schaepman, M.E., De Jong, R., 2017. Barest pixel composite for agricultural areas using landsat time series. Remote Sens. 9, 1245. https://doi.org/10.3390/rs9121245.

Dvornikov, Y.A., Slukovskaya, M.V., Yaroslavtsev, A.M., Meshalkina, J.L., Ryazanov, A., Sarzhanov, D.A., Vasenev, V.I., 2022. High-resolution mapping of soil pollution by Cu and Ni at a polar industrial barren using proximal and remote sensing. Land Degrad. Dev. 1–14. https://doi.org/10.1002/ldr.4261.

Elsukova, E.Y., Opekunova, M.G., Opekunov, A.Y., 2019. Technogenic transformation of heavy metal streams in soils in regions under influence of copper-nickel production. International Research Journal 12, 118–124. https://doi.org/10.23670/IRJ.2019.90.12.024.

Ermida, S.L., Soares, P., Mantas, V., Göttsche, F.M., Trigo, I.F., 2020. Google Earth engine open-source code for land surface temperature estimation from the landsat series. Remote Sens. 12, 1–21. https://doi.org/10.3390/RS12091471.

Evdokimova, G.A., 1995. Ecological-Microbiological Foundations of Soil Protection in the far North. KSC RAS, Apatity.

Evdokimova, G.A., Kalabin, G.V., Mozgova, N.P., 2011. Contents and toxicity of heavy metals in soils of the zone affected by aerial emissions from the severonikel Enterprise. Eurasian Soil Sci. 44, 237–244. https://doi.org/10.1134/S1064229311020037.

Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 29, 1189–1232. https://doi.org/10.1214/aos/1013203451.

Gitelson, A.A., Gritz, Y., Merzlyak, M.N., 2003. Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. J. Plant Physiol. 160, 271–282. https://doi.org/10.1078/0176-1617-00887.

IUSS Working Group WRB, 2015. World Reference Base for Soil Resources 2014, Update 2015. International Soil Classification System for Naming Soils and Creating Legends for Soil Maps, World Soil Resources Reports No. 106. FAO, Rome.

Jasiewicz, J., Stepinski, T.F., 2013. Geomorphons-a pattern recognition approach to classification and mapping of landforms. Geomorphology 182, 147–156. https://doi.org/10.1016/j.geomorph.2012.11.005.

Kashulina, G.M., 2018. Monitoring of soil contamination by heavy metals in the impact zone of copper-nickel smelter on the kola peninsula. Eurasian Soil Sci. 51, 467–478. https://doi.org/10.7868/s0032180x1804010x.

Kashulina, G.M., 2017. Extreme pollution of soils by emissions of the copper–nickel industrial complex in the Kola Peninsula. Eurasian Soil Sci. 50, 837–849. https://doi.org/10.7868/s0032180x17070036.

Kashulina, G.M., De Caritat, P., Reimann, C., 2014. Snow and rain chemistry around the "Severonikel" industrial complex, NW Russia: current status and retrospective analysis. Atmos. Environ. 89, 672–682. https://doi.org/10.1016/j.atmosenv.2014.03.008.

Kempen, B., Heuvelink, G.B.M., Brus, D.J., Stoorvogel, J.J., 2010. Pedometric mapping of soil organic matter using a soil map with quantified uncertainty. Eur. J. Soil Sci. 61, 333–347. https://doi.org/10.1111/j.1365-2389.2010.01232.x.

Key, C.H., Benson, N.C., 1999. The Normalized burn ratio (NBR): a landsat TM radiometric measure of burn severity. In: United States Geological Survey, Northern Rocky Mountain Science Center. Bozeman, MT.

Koptsik, G.N., Koptsik, S.V., Smirnova, I.E., 2016. Alternative technologies for remediation of technogenic barrens in the Kola Subarctic. Eurasian Soil Sci. 49, 1294–1309. https://doi.org/10.1134/S1064229316090088.

Kozlov, M.V., 2001. Snowpack changes around a nickel-copper smelter at Monchegorsk, northwestern Russia. Can. J. For. Res. 31, 1684–1690. https://doi.org/10.1139/cjfr-31-10-1684.

Kozlov, M.V., Haukoja, E., Yarmishko, V.T. (Eds.), 1993. Aerial Pollution in Kola Peninsula: Proceedings of International Workshop. University of Turku, St. Petersburg, Apatity, p. 417.

Kozlov, M.V., Zvereva, E.L., 2007. Industrial barrens: extreme habitats created by non-ferrous metallurgy. Rev. Environ. Sci. Biotechnol. 6, 231–259. https://doi.org/10.1007/s11157-006-9117-9.

Kryuchkov, V.V., 1993. Extreme anthropogenic loads and the Northern ecosystem condition. Ecol. Appl. 3, 622–630. https://doi.org/10.2307/1942095.

Kuhn, M., 2008. Building predictive models in R using the caret package. J. Stat. Software 28, 1–26. https://doi.org/10.18637/jss.v028.i05.

Lyanguzova, I.V., Barcan, V.S., 2019. Comparative analysis of the level of pollution of the organogenic horizon of al-fe-podzols and bog soils in the local impact zone of the copper-nickel plant. Principy èkologii 8, 57–68.

Lyanguzova, I.V., Goldvirt, D.K., Fadeeva, I.K., 2016. Spatiotemporal dynamics of the pollution of Al–Fe-humus podzols in the impact zone of a nonferrous metallurgical plant. Eurasian Soil Sci. 49, 1189–1203. https://doi.org/10.1134/S1064229316100094.

Lyanguzova, I.V., Goldvirt, D.K., Fadeeva, I.K., 2015. Transformation of polymetallic dust in the organic horizon of al-fe-humus podzol (field experiment). Eurasian Soil Sci. 48, 701–711. https://doi.org/10.1134/S1064229315050051.

Manninen, S., Zverev, V., Bergman, I., Kozlov, M., 2015. Consequences of long-term severe industrial pollution for above ground carbon and nitrogen pools in northern taiga forests at local and regional scales. Sci. Total Environ. 536, 616–624. https://doi.org/10.1016/j.scitotenv.2015.07.097.

McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma. https://doi.org/10.1016/S0016-7061(03)00223-4.

Merzlyak, M.N., Gitelson, A.A., Chivkunova, O.B., Rakitin, V.Y., 1999. Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening. Physiol Plant 106, 135–141. https://doi.org/10.1034/j.1399-3054.1999.106119.x.

Meyer, H., Kostrova, S.S., Meister, P., Lenz, M.M., Kuhn, G., Nazarova, L., Syrykh, L.S., Dvornikov, Y.A., 2022. Lacustrine diatom oxygen isotopes as palaeo precipitation proxy - holocene environmental and snowmelt variations recorded at Lake Bolshoye Shchuchye, Polar Urals, Russia. Quat. Sci. Rev. 290, 107620. https://doi.org/10.1016/j.quascirev.2022.107620.

Minasny, B., McBratney, A.B., 2006. A conditioned Latin hypercube method for sampling in the presence of ancillary information. Comput. Geosci. 32, 1378–1388. https://doi.org/10.1016/j.cageo.2005.12.009.

Nedkov, R., 2017. Orthogonal transformation of segmented images from the satellite Sentinel-2. Comptes Rendus de L'Academie Bulgare des Sciences 70, 687–692.

Norseth, T., 1994. Environmental pollution around nickel smelters in the Kola Peninsula (Russia). Sci. Total Environ. 148, 103–108.

Opekunova, M., Opekunov, A., Elsukova, E.Y., Kukushkin, S., Janson, S., 2021. Comparative analysis of methods for air pollution assessing in the Arctic mining area. Atmos. Pollut. Res. 12, 76–88. https://doi.org/10.1016/j.apr.2020.08.017.

Paltseva, A., Slukovskaya, M.V., Romzaykina, O.N., Sarzhanov, D.A., Drogobuzhskaya, S. V., Dvornikov, Y.A., Vasenev, V.I., 2022. Proximal sensing of soil pollution by heavy metals using a portable X-ray fluorescence analyzer in Subarctic industrial barren: limitations and perspectives. In: Rakshit, A., Ghosh, S., Vasenev, V.I., Pathak, H., Rajput, V.D. (Eds.), Soils in Urban Ecosystem. Springer Singapore, Singapore, pp. 243–254. https://doi.org/10.1007/978-981-16-8914-7_11.

Peralta, E., Pérez, G., Ojeda, G., Alcañiz, J.M., Valiente, M., López-Mesas, M., Sánchez-Martín, M.J., 2020. Heavy metal availability assessment using portable X-ray fluorescence and single extraction procedures on former vineyard polluted soils. Sci. Total Environ. 726. https://doi.org/10.1016/j.scitotenv.2020.138670.

Porter, C., Morin, P., Howat, I., Noh, M.-J., Bates, B., Peterman, K., Keesey, S., Schlenk, M., Gardiner, J., Tomko, K., Willis, M., Kelleher, C., Cloutier, M., Husby, E., Foga, S., Nakamura, H., Platson, M., Wethington Jr., M., Williamson, C., Bauer, G., Enos, J., Arnold, G., Kramer, W., Becker, P., Doshi, A., D'Souza, C., Cummens, P., Laurier, F., Bojesen, M., 2018. Arctic. https://doi.org/10.7910/DVN/OHHUKH [WWW Document].

Pozniakov, V.Y., 1993. The "Severonikel" smelter complex: history of development. In: Kozlov, M.V., Haukioja, E., Yarmishko, V.T. (Eds.), Aerial Pollution in Kola Peninsula: Proceedings of the International Workshop, pp. 16–19. St.-Petersburg, Apatity.

R Core Team, 2021. R: a Language and Environment for Statistical Computing.

Rees, W.G., Kapitsa, A.P., 1994. Industrial pollution in the Kol'skiy Poluostrov, Russia. Polar Rec. 30, 181–188.

Romzaykina, O.N., Slukovskaya, M.V., Paltseva, A.A., Losev, A.I., Korneykova, M.V., Vasenev, V.I., 2024. Rapid assessment of soil contamination by potentially toxic metals in the green spaces of Moscow megalopolis using the portable X-ray analyzer. J. Soils Sediments. https://doi.org/10.1007/s11368-024-03758-1.

Roudier, P., Beaudette, D.E., Hewitt, A.E., 2012. A conditioned Latin hypercube sampling algorithm incorporating operational constraints. In: Digital Soil Assessments and Beyond. Proceedings of the 5th Global Workshop on Digital Soil Mapping. Sydney, Australia.

Slukovskaya, M.V., Vasenev, V.I., Ivashchenko, K.V., Dolgikh, A.V., Novikov, A.I., Kremenetskaya, I.P., Ivanova, L.A., Gubin, S.V., 2021. Organic matter accumulation by alkaline-constructed soils in heavily metal-polluted area of Subarctic zone. J. Soils Sediments 21, 2071–2088. https://doi.org/10.1007/s11368-020-02666-4.

Slukovskaya, M.V., Vasenev, V.I., Ivashchenko, K.V., Morev, D., Drogobuzhskaya, S., Ivanova, L., Kremenetskaya, I.P., 2019. Technosols on mining wastes in the subarctic: efficiency of remediation under Cu-Ni atmospheric pollution. Int. Soil Water Conserv. Res. 7, 297–307. https://doi.org/10.1016/j.iswcr.2019.04.002.

Slukovskii, Z.I., Dauvalter, V.A., Denisov, D.B., Siroezhko, E.V., Cherepanov, A.A., 2020. Geochemistry features of sediments of small urban arctic Lake Komsomolskoye, Murmansk region. In: IOP Conference Series: Earth and Environmental Science. Institute of Physics Publishing. https://doi.org/10.1088/1755-1315/467/1/012004.

Tolan, J., Yang, H.I., Nosarzewski, B., Couairon, G., Vo, H.V., Brandt, J., Spore, J., Majumdar, S., Haziza, D., Vamaraju, J., Moutakanni, T., Bojanowski, P., Johns, T., White, B., Tiecke, T., Couprie, C., 2024. Very high resolution canopy height maps from RGB imagery using self-supervised vision transformer and convolutional decoder trained on aerial lidar. Remote Sens. Environ. 300. https://doi.org/10.1016/j.rse.2023.113888.

Vodyanitsky, Y.N., 2008. Affinity of heavy metals and metalloids for carriers in soils. Agrokhimia 87–94.

Wadoux, A.M.J.C., McBratney, A.B., 2021. Hypotheses, machine learning and soil mapping. Geoderma 383, 114725. https://doi.org/10.1016/j.geoderma.2020.114725.

Wadoux, A.M.J.C., Minasny, B., McBratney, A.B., 2020. Machine learning for digital soil mapping: applications, challenges and suggested solutions. Earth Sci. Rev. 210, 103359. https://doi.org/10.1016/j.earscirev.2020.103359.

Weiss, A.D., 2001. Topographic Positions and Landforms Analysis (poster). In: ESRI International User Conference, July 2001. ESRI, San Diego, CA.

Xu, H., 2006. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. Int. J. Rem. Sens. 27, 3025–3033. https://doi.org/10.1080/01431160600589179.

Yang, Y., Cui, Q., Jia, P., Liu, J., Bai, H., 2021. Estimating the heavy metal concentrations in topsoil in the daxigou mining area, China, using multispectral satellite imagery. Sci. Rep. 11, 1–9. https://doi.org/10.1038/s41598-021-91103-8.

Zamulina, I.V., Gorovtsov, A.V., Minkina, T.M., Mandzhieva, S.S., Burachevskaya, M.V., Bauer, T.V., 2022. Soil organic matter and biological activity under long-term contamination with copper. Environ. Geochem. Health 44, 387–398. https://doi.org/10.1007/s10653-021-01044-4.

Zaytsevskiy, A.B., 1971. Geological map of Monche-Chuna-Volchya tundras area. In: Report on Geological Survey (1:50,000) Scale in 1969-1972. Ministry of Geology RSFSR, Murmansk, p. 5.

Zhang, Z., Liu, M., Liu, X., Zhou, G., 2018. A new vegetation index based on multitemporal Sentinel-2 images for discriminating heavy metal stress levels in rice. Sensors 18, 2172. https://doi.org/10.3390/s18072172.