scientific reports

Check for updates

OPEN

High-Resolution prediction of soil pH in European temperate forests using Sentinel-2 and ancillary environmental data

Haidi Abdullah✉, Andrew K. Skidmore, Andjin Siegenthaler & Elnaz Neinavaz

Soil pH is a key indicator for understanding soil health status in forested ecosystems, yet high-resolution mapping of this variable, especially at a 30-m spatial resolution, remains limited. This study uses Sentinel-2 spectral data, in-situ soil pH measurements, topsoil physical properties from the Land Use/Cover Area Frame Survey (LUCAS) database, and elevation data to estimate soil pH across temperate forests in Europe using a Random Forest model. Despite challenges in signal penetration due to forest canopy cover, the model achieved high prediction accuracy ($R^2 = 0.62$) at 30 m resolution. Bulk density, available water capacity, and clay content were the most influential physical predictors, while Sentinel-2 bands, particularly SWIR (1.610 and 2.190 µm), NIR (0.842 µm), and red-edge (0.705 and 0.783 µm), captured key vegetation responses related to soil acidity. Spatial analysis showed higher model accuracy in central and southern Europe, with reduced performance in Scandinavia, likely due to more acidic soils and extreme seasonal variation. The model also revealed significant pH differences among forest types, with deciduous forests showing the highest values and coniferous the lowest. These findings demonstrate the potential of high-resolution remote sensing data for monitoring soil pH, supporting forest management, biodiversity conservation, and climate adaptation strategies.

Keywords Soil pH, Sentinel-2, Random forest, LUCAS topsoil database

Soil pH in forested areas is a crucial indicator of soil health, influencing tree vitality, nutrient availability, and soil micro-organism activities[1,2]. It plays a pivotal role by directly and indirectly affecting the availability and concentration of essential nutrients and minerals[3]. Increasing soil acidification, therefore, can be regarded as a harmful process in forest ecosystems, posing a risk to plant root systems and potentially leading to a decline in ecosystem productivity[4]. In general, the primary sources of soil acidification in forest ecosystems include timber harvesting activities, atmospheric deposition from industrial emissions, and the natural accumulation of organic biomass[5]. Despite the substantial reduction in industrial atmospheric deposition in recent decades, primarily due to regulations requiring flue gas treatments such as smokestack scrubbers[6,7], soil acidification remains a major issue due to intensive agricultural practices. For example, the widespread use of nitrogen-based fertilisers continues to be a key driver of soil acidification in many European regions, with ecological consequences particularly evident in forested areas[8].

European temperate forests are particularly vulnerable to environmental changes such as drought, insect infestations, and soil acidification, which influence soil composition and vegetation diversity[6,9]. These stressors highlight the importance of studying soil pH, as it plays a key role in nutrient availability, microbial activity, and ecosystem health. Traditional soil pH measurement methods, such as colourimetric and electrometric lab techniques, are known for their accuracy. However, they are costly, time-consuming and fail to capture the spatial heterogeneity of soil pH across large areas, as they do not fill the spatial gaps between field samples, leaving substantial areas unmeasured[10]. Proximal soil sensing, which uses electrical and electromagnetic sensors, offers a quicker, more practical alternative for field-scale pH estimation. While proximal sensing can capture some spatial variability, it remains limited in its coverage, making it suitable for field-scale studies rather than regional or continental assessments.

To overcome the limitations of traditional soil surveys, which do not spatially fill the gaps between field samples, various remote sensing approaches have been developed and utilised in recent years to estimate

Department of Natural Resources, Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, Hallenweg 8, 7522 NH Enschede, The Netherlands. ✉email: h.j.abdullah-1@utwente.nl

soil properties, including soil pH, at local and regional scales[11–14]. The ability to study soil pH using remote sensing data relies heavily on the spectral and spatial resolution offered by different sensors and platforms[11]. For example, optical remote sensing data from platforms such as Landsat and Sentinel-2 have been widely employed for soil pH estimation at local and small scales[15,16]. However, it is important to note that these studies have predominantly focused on agricultural farmland.

With recent advancements in remote sensing technologies, particularly the development of satellite hyperspectral imaging from sensors like DESIS, EnMap, and PRISMA, the application of remote sensing for soil pH estimation has expanded beyond agriculture systems to include other land cover types, such as forests. For instance, Abdullah, et al.[11] estimated soil pH using image spectroscopy data from DESIS satellite in two temperate forested areas in Europe, demonstrating the growing interest in applying remote sensing to non-agricultural ecosystems. Likewise, Vibhute and Kale[17] applied a Hyperion image of Earth Observing (EO)−1 spacecraft to map different soil types and estimate physicochemical soil properties, including soil pH. Additionally, recent studies have employed airborne hyperspectral data[18] and unmanned aerial vehicle (UAV) hyperspectral data to estimate soil properties, including soil pH, at very fine spatial scales[19].

To the best of our knowledge, no prior study has estimated soil pH in temperate forests at high resolution across a continental scale using optical remote sensing combined with environmental and topographic covariates. Our study addresses this gap by combining optical satellite data from Sentinel-2 with a large-scale dataset of in-situ soil measurements sourced from the European Soil Data Centre (ESDAC, 2018). This dataset, gathered through the Land Use/Cover Area Frame Survey (LUCAS), provides over 22,000 soil samples from European member states, which serve as invaluable training data for predicting soil pH at a broad, continental level. This integrative approach not only demonstrates the potential of merging satellite and ground-based data but also advances soil health assessments across diverse European landscapes.

Previous work conducted by Ballabio, et al.[20] generated topsoil property maps, including soil pH, across Europe at a spatial resolution of 500 m using the LUCAS dataset. Their approach incorporated remote sensing-derived indices, such as Normalised Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI) from MODIS products, but used these indices indirectly within a broader digital soil mapping framework. Specifically, soil chemical properties like pH were inferred through complex statistical techniques, including Gaussian Process Regression, supported by a wide range of harmonised environmental covariates. In this context, vegetation indices functioned as intermediate variables rather than direct predictors. Similarly, the International Soil Reference and Information Centre (ISRIC) developed global soil property maps at a 250 m resolution[21], relying on machine learning models trained on global soil observations and a suite of environmental covariates, including climate, terrain, and remote sensing-derived data.

In contrast, the current study presents a streamlined and scalable approach to estimating soil pH by integrating high-resolution Sentinel-2 surface reflectance data with environmental covariates such as topsoil physical properties and elevation data, all at a 30 m spatial resolution. Rather than relying on interpolated environmental layers or ancillary datasets such as climate or land cover, this approach uses spectral reflectance from forest canopies along with topsoil physical properties as the primary input. This approach builds on the widely recognised biochemical link between vegetation characteristics and soil properties, particularly soil pH. Recent research has reinforced this relationship. For instance, Abdullah, et al.[11], successfully estimated soil pH using DESIS spectral reflectance, while Skidmore, et al.[22] demonstrated that canopy-level spectral information can reliably predict microbial alpha diversity, which is closely associated with soil pH.

The novelty of this study lies not only in its direct use of spectral data but also in its global applicability and in highlighting both the strengths and limitations of this approach. By eliminating the need for extensive field campaigns and leveraging freely available satellite imagery, our method offers a cost-effective, easy-to-implement solution for environmental monitoring and soil management, especially in data-scarce regions. Moreover, by improving the spatial resolution of soil pH maps from 500 m to 250 m down to 30 m, the study enables detailed, local-scale assessments. This finer granularity is particularly valuable for forest management, biodiversity conservation, reforestation planning, and tracking the impacts of climate change on forest ecosystems. In areas where forest health depends on soil chemistry, our approach provides a vital tool for guiding sustainable landscape management.

Methods

An overview of the workflow for estimating soil pH in European temperate forests, using in-situ LUCAS topsoil data and Sentinel-2 imagery, is illustrated in Fig. 1 and further elaborated in the following sub-sections.

Study area

In this study, the European temperate forests have been selected as the study site for estimating and mapping soil pH. These forests span a broad geographical range, generally located between approximately 45° N and 70° N latitudes and 10° W and 30° E longitudes. The region is considered to encompass parts of Western, Central, and Northern Europe, including the United Kingdom, France, Germany, Poland, Austria, Sweden, and Finland. Data and definitions provided by Olson, et al.[23] on worldwide Terrestrial Ecoregions were applied to define the boundaries of the European temperate forests. Initially, the temperate climate zone was extracted and further refined by clipping it using the boundaries of EU countries (Fig. 2). The CORINE 2018 land use and land cover data (LULC)[24] were employed to pinpoint forested areas within Europe's identified temperate climate zone. Forested areas, including coniferous, deciduous, and mixed forests, were extracted within the temperate climate zone using the boundary identified by Olson, et al.[23] for further analysis.

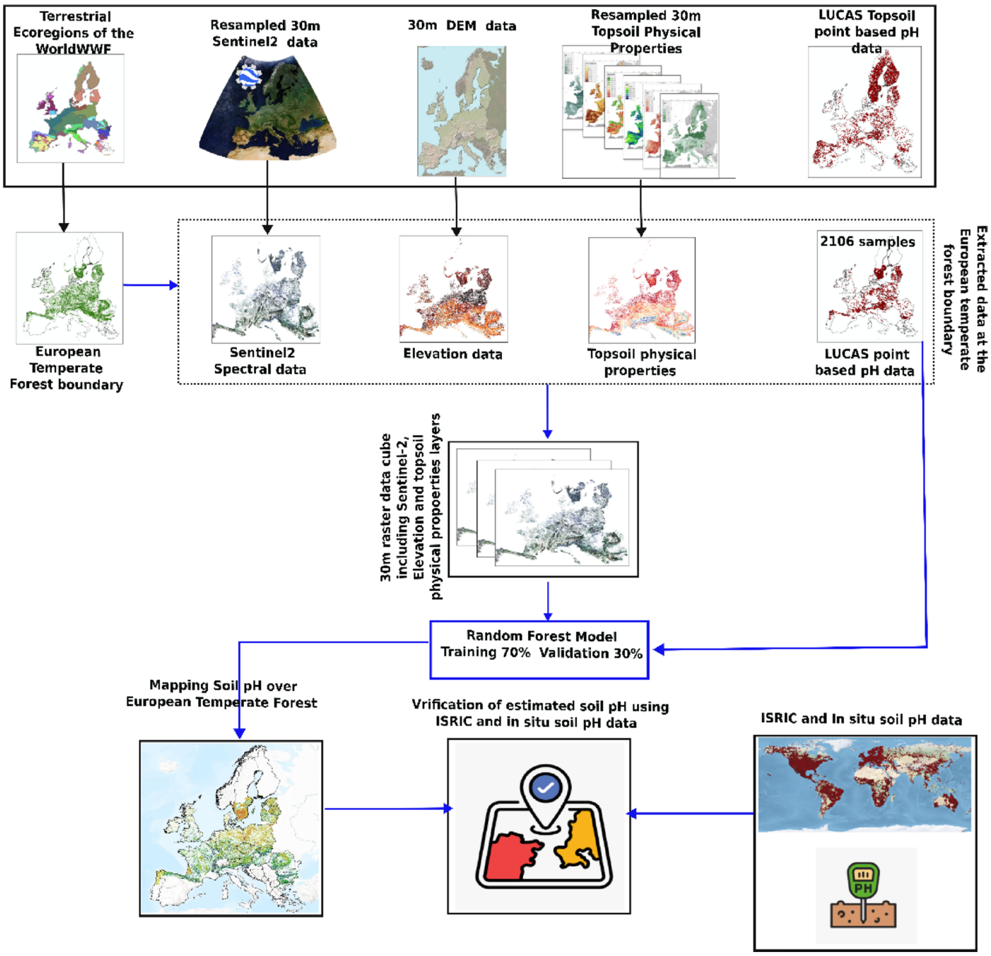

Fig. 1. Graphical workflow for estimating soil pH in European temperate forests using Sentinel-2 imagery, in-situ soil pH measurements, topsoil physical properties from the LUCAS database, and elevation data.

In situ soil pH data

Two different datasets were used in this study for modelling and validation. The in-situ soil pH dataset from the LUCAS project (used for modelling) and the BIOSPACE project (used for validation) were applied at European and local scales, respectively. The in-situ soil pH data were obtained from the EUROPEAN SOIL DATA CENTRE (ESDAC) 2018, developed as part of the LUCAS project. The LUCAS project started in 2009 to sample and analyse the main properties of topsoil, including soil pH, in 23 European Member States. This topsoil survey represents the first attempt to build a consistent spatial database of soil cover across Europe based on standard sampling and analytical procedures, with all soil samples being analysed in a single laboratory. In this study, the LUCAS topsoil dataset collected in 2018 was used. The database contains 18,984 geo-referenced soil samples distributed across 25 European countries. A standardised sampling procedure collected approximately 0.5 kg of topsoil (0–20 cm). The samples were then dispatched to a central laboratory for physical and chemical analyses. For more details on the LUCAS datasets and the methodology employed to generate the soil pH map, refer to Ballabio, et al.[20]. As the main aim of this study is to map soil pH in European temperate forests, the data from non-forested areas were excluded from the analysis. Initially, all soil samples collected from forested areas were extracted, which amounted to approximately 5,603 samples from the LUCAS dataset. Further, using the temperate forest boundaries defined earlier (Sect.“Study area”), the samples were refined, narrowing it down to around 2,106 samples, specifically from temperate forests.

At the local scale, the in-situ measurements obtained from the BIOSPACE project were collected from two ecologically distinct temperate forest sites: the Bavarian Forest National Park in Germany and the Veluwe region, which includes Veluwezoom and Hoge Veluwe National Parks in the Netherlands[22]. The Bavarian Forest National Park is a mountainous area located near the Czech border, characterized by high annual precipitation (1200–1800 mm), mixed forests dominated by Norway spruce (*Picea abies*) at higher elevations and European beech (*Fagus sylvatica*) in lower areas, and a long history of soil acidification due to industrial pollution[25].

In contrast, the Veluwe National Parks are situated in the lowlands of the Netherlands and comprise heathlands, dunes, and a mosaic of coniferous and deciduous stands, with species like beech, oak (*Quercus L*), birch (*Betula L.*), and Scotch pine (*Pinus sylvestris*). This region has a temperate climate with moderate rainfall (~ 850 mm annually) and is affected by nitrogen deposition from intensive agriculture, leading to soil acidification and

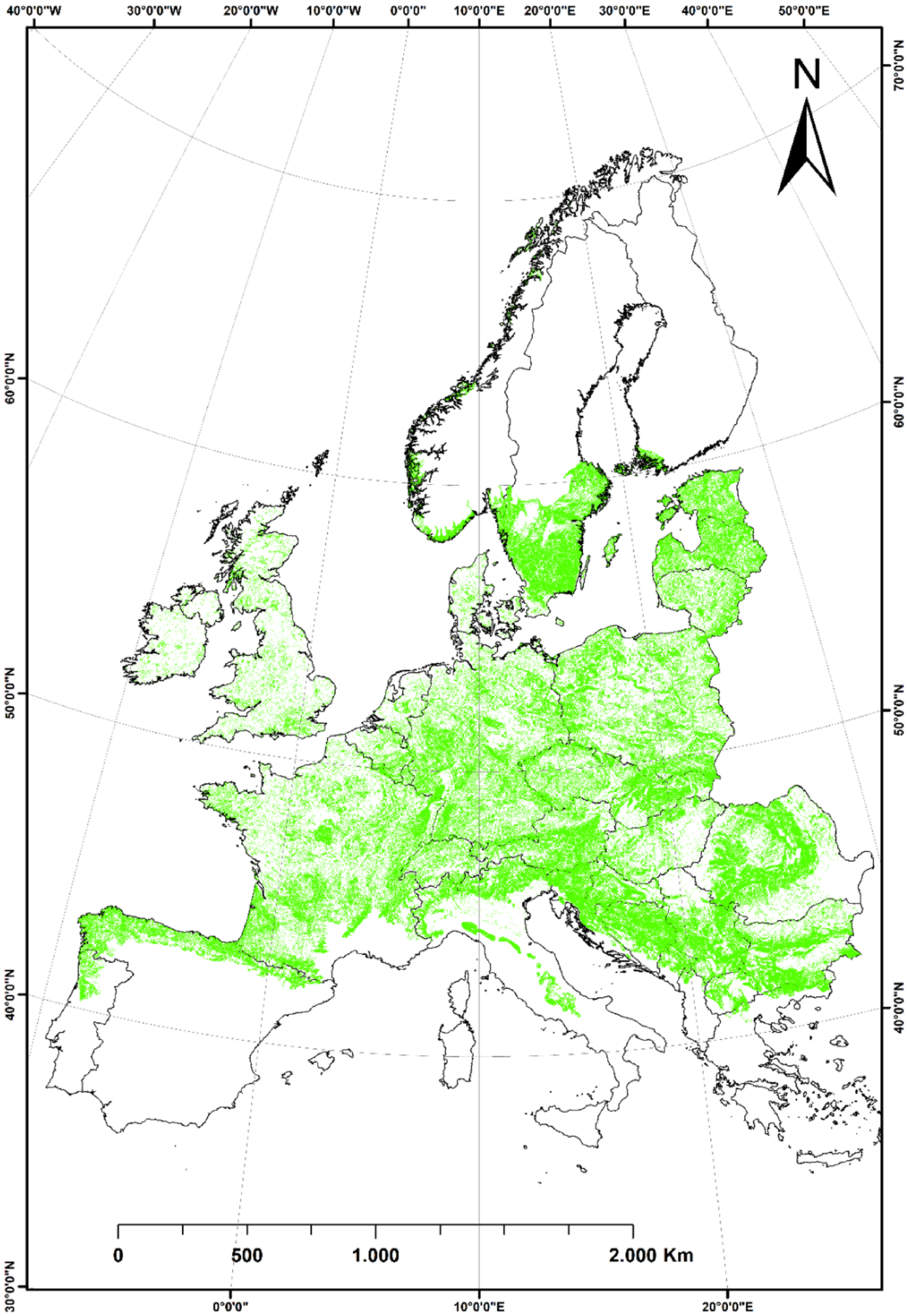

Fig. 2. Spatial extent of temperate forest regions across Europe (shown in green). The boundaries of the temperate forest biome are based on the global ecoregion classification by Olson, et al.[23].The map was created using ArcGIS Pro version 3.1.1.

extremely low pH levels in some areas (as low as 3.1)[6]. As mentioned, these local data are part of the BIOSPACE project and were collected during the summers of 2020 and 2021. The in-situ soil samples were taken from the top 0–10 cm of the mineral soil layer, following the careful removal of the litter layer. Further details on the field campaigns and soil sampling protocols are provided in Abdullah, et al.[11]. To facilitate comparison, soil pH values from 197 samples collected in the Bavarian Forest National Park ($n = 97$) and the Veluwe National Park ($n = 100$)—were evaluated against the soil pH estimates produced in this study. The coefficient of determination (R^2) was calculated to assess the degree of correlation between the two datasets.

Satellite and environmental covariates data

In this study, cloud-free, mosaicked Sentinel-2 satellite data and environmental covariates were used to estimate and map soil pH across European temperate forests (Table 1). Data processing and analysis were conducted on

Category	Covariate	Description
Sentinel-2 Spectral Bands	B2 (Blue)	Reflectance in blue wavelength
	B3 (Green)	Reflectance in green wavelength
	B4 (Red)	Reflectance in red wavelength
	B5, B6, B7 (Red-edge)	Reflectance in red-edge bands
	B8, B8 A (NIR)	Reflectance in near-infrared
	B11, B12 (SWIR)	Reflectance in shortwave infrared
Topsoil Properties	Clay content	% clay in topsoil
	Bulk density	Dry soil mass per volume
	Available water capacity	Water that can be stored in the soil
	Coarse fragments	The volume of rock fragments in the soil
	Sand content	% sand in topsoil
	Silt content	% silt in topsoil
Digital Elevation Model	Elevation	Height above sea level

Table 1. Overview of the covariates used in the random forest model for estimating soil pH in European temperate forests.

the Google Earth Engine (GEE) cloud computing platform, which enabled efficient handling of large image collections and the generation of the mosaicked Sentinel-2 dataset.

Sentinel-2 scenes with more than 10% cloud cover were excluded to ensure high image quality. The spectral bands (i.e., B2, B3, B4, B5, B6, B7, B8, B8 A, B11, and B12) were resampled to a 30 m spatial resolution using bilinear interpolation. These canopy reflectance values served as predictors in a Random Forest (RF) model. Since direct soil reflectance is often obscured by dense vegetation in forested areas, the spectral signals primarily represent the vegetation canopy. This canopy reflectance has been shown to carry indirect information about belowground soil conditions, including pH, as supported by Skidmore, et al.[22], who demonstrated that canopy reflectance derived from satellite image spectroscopy can serve as a proxy for underlying soil properties.

Only pixels classified as forests in the CORINE LULC (2018) dataset were used for model training to reduce mixed-pixel effects and improve signal quality. Furthermore, monthly median reflectance composites were generated from March to December 2018 to mitigate noise caused by seasonal variation, cloud contamination, and occasional understory exposure. These reflectance values were temporally aligned with the LUCAS soil data collection period.

To complement the reflectance data from Sentinel-2, a set of environmental covariates was selected based on their theoretical relevance to soil pH dynamics and their availability at high spatial resolution. These covariates included topsoil physical properties and elevation data. Elevation information was obtained from the Copernicus GLO-30 DEM, which offers a native spatial resolution of 30 m and aligns well with the Sentinel-2 reflectance data. Elevation was included due to its known influence on microclimate, runoff, erosion, and water accumulation factors that can significantly shape soil pH[26,27]. Moreover, elevation acts as a proxy for climatic gradients, as higher altitudes tend to exhibit lower temperatures, increased precipitation, and more acidic soils[28].

While climatic variables such as precipitation and temperature are widely recognised as important long-term drivers of soil acidification, primarily through processes like leaching and acid deposition, they were not included in this study for several reasons. First, although these variables are significant over extended timescales, their gradual effects are less relevant when modelling soil pH at a specific point in time[29]. In this study, we focused solely on data from the year 2018, limiting the relevance of long-term climatic trends. Second, including commonly available climate datasets (e.g., WorldClim) would compromise model parsimony, as their coarse spatial resolution is misaligned with the finer 30 m resolution of the Sentinel-2 imagery and soil property data used in this study. Third, our goal was to produce a high-resolution, spatially consistent soil pH map, which necessitated using covariates available at finer spatial scales.

As a result, the final set of covariates included clay, silt, sand content, bulk density, available water capacity (AWC), coarse fragments, and elevation. These soil properties were derived from Ballabio, et al.[20], originally available at 500 m resolution and subsequently resampled to 30 m using bilinear interpolation. Topsoil variables, particularly clay content, bulk density, and AWC, were prioritised for their established roles in influencing pH through buffering capacity, nutrient retention, and soil water dynamics. While relevant, Soil organic carbon (SOC) was excluded due to its high correlation with other variables (especially clay)[30,31], limited spatial coverage in forested regions, and negligible impact on model accuracy during preliminary evaluations. All datasets, including Sentinel-2 mosaics, environmental covariates, and elevation, were clipped to the temperate forest boundary defined in Sect. "Study area" to ensure ecological and geographic consistency in the modelling process.

Estimation and mapping of soil pH using random forests and remote sensing data

The in-situ soil pH measurements from the LUCAS dataset, integrated with Sentinel-2 canopy reflectance (i.e., bands B2–B12), topsoil physical properties, and digital elevation model (DEM) data, were used to estimate and map soil pH across temperate forested regions of Europe. An RF regression model was selected for its ability to handle high-dimensional data and capture complex, non-linear relationships between variables[32,33]. The selection of predictors was based on the premise that the spectral signatures of vegetation can indirectly represent underlying soil chemical properties, such as pH, through vegetation–soil interactions[11,22]. In-situ soil pH measurements from the LUCAS dataset served as the response variable (Y). The dataset was randomly partitioned into training (70%) and testing (30%) subsets, and model performance was assessed using a holdout cross-validation approach to evaluate generalisation capability.

To fine-tune model parameters, we performed cross-validation across models trained with varying numbers of decision trees (e.g., 100, 200, 300), ultimately selecting 300 trees based on validation performance, as this configuration yielded the lowest root mean square error (RMSE) and the highest R^2. Out-of-bag (OOB) samples were used to estimate feature importance, providing insights into the relative contribution of each predictor and allowing a comparison between spectral and environmental variables.

To improve model reliability and account for spatial and spectral variability within forested environments, we stratified the analysis by forest type (e.g., coniferous, deciduous, and mixed) based on CORINE LULC classifications. Recognising that satellite reflectance of vegetated surfaces can fluctuate due to seasonal, phenological, and atmospheric factors, we further assessed the model's temporal robustness by applying it to Sentinel-2 imagery from a different year (i.e., 2024). For this purpose, monthly median reflectance composites from March to December 2024 were generated to evaluate the model's sensitivity to interannual and seasonal spectral variations in canopy reflectance. In addition to internal validation, the model's predictive accuracy was externally validated using in situ data from the BIOSPACE dataset. Soil pH values predicted by the RF model (trained on the LUCAS dataset) were extracted for corresponding BIOSPACE locations and compared to observed values. Model performance was quantified using the R^2 and the normalised root mean square error (nRMSE).

Verifying and comparing estimated soil pH with existing soil pH data

To verify and compare the soil pH data estimated in this study, the existing soil pH data at the European scale, namely ISRIC, was applied. The ISRIC dataset was generated at standard soil depths of 0 cm, 5 cm, 15 cm, 30 cm, 60 cm, 100 cm, and 200 cm; for our analysis, soil data with a 15 cm depth was selected to be aligned with the *in-situ* soil measurements obtained from the LUCAS dataset. The data are provided in raster format with a spatial resolution of 250 m and were generated using an ensemble of machine learning methods applied to a large dataset of soil profiles and remote sensing data. For detailed information on the methodology used to estimate soil pH at various depths in the ISRIC dataset, please refer to Hengl, et al.[21]. To enable a comparison between the soil pH estimated in this study and the soil pH derived from the ISRIC dataset, the ISRIC soil pH data was resampled to a 30 m resolution to align it with the soil pH data estimated in this study.

Further, the percentage match between the ISRIC dataset at the European scale and retrieved pH using the RF approach was calculated. This process involved computing the absolute error between the predicted soil pH values and the ISRIC soil pH values at the corresponding depth (15 cm). The absolute error was determined on a pixel-by-pixel basis, using the difference between the two raster datasets. Next, the absolute error was normalized to obtain a percentage error by dividing it by the known range obtained from the original pH data and multiplying the result by 100. This step allowed us to quantify the degree of error relative to the total possible pH range. Finally, we calculated the percentage match by subtracting the percentage error from 100, resulting in a spatial representation of how closely the two datasets aligned. High percentage values indicated a strong agreement between the soil pH estimated in this study and the ISRIC soil pH data.

Analysing spatial variation in soil pH across forest types

To investigate the spatial variation of estimated soil pH across European temperate forests, we conducted a comparative analysis of soil pH among three forest types: coniferous, deciduous, and mixed forest stands. These forest types were identified based on CORINE LULC data and the temperate climate zone using the boundary defined by Olson, et al.[23]. The forest type classification was overlaid on the soil pH estimation data, allowing us to group and calculate the mean and standard deviation of soil pH for each forest type.

The purpose of this analysis was twofold. First, understanding the variation in soil pH across forest types is critical because soil pH plays a fundamental role in shaping forest ecosystem dynamics, including nutrient cycling, tree species distribution, and soil microbial activity. Differences in forest composition can lead to distinct patterns of soil acidity due to variations in litter input, root activity, and microbial processes[34]. Second, exploring this variation helps assess the applicability of remote sensing-based soil pH estimation models across diverse forested landscapes, ensuring that the models perform consistently under varying ecological conditions. To further examine this, we conducted a complementary analysis using the raw *in situ soil pH* data provided by the LUCAS dataset. This additional analysis aimed to test whether the variation in soil pH across different forest types observed in our model estimates aligns with the variation present in the measured in-situ values.

Results

Mapping soil pH in European temperate forests and model validation

The spatial distribution of soil pH across European temperate forests, estimated using an RF model with Sentinel-2 satellite data and the LUCAS topsoil database, shows significant regional variability (Fig. 3). Model predictions indicate soil pH values ranging from 3.1 to 7.4. Acidic soils (pH <5) are predominantly found in northern and eastern regions, including forested areas of Scandinavia, the Baltic states, and parts of Poland. In contrast, more alkaline soils (pH ≥ 7) are present in southern and central regions, particularly Spain, Italy, France, and Germany.

Areas in southern Sweden and parts of Finland exhibited relatively low soil pH values. Similar patterns of low pH values are also visible across the Baltic region. Meanwhile, more neutral to slightly alkaline soils are evident in central and western Europe, with distinct spatial gradients visible at national and regional levels.

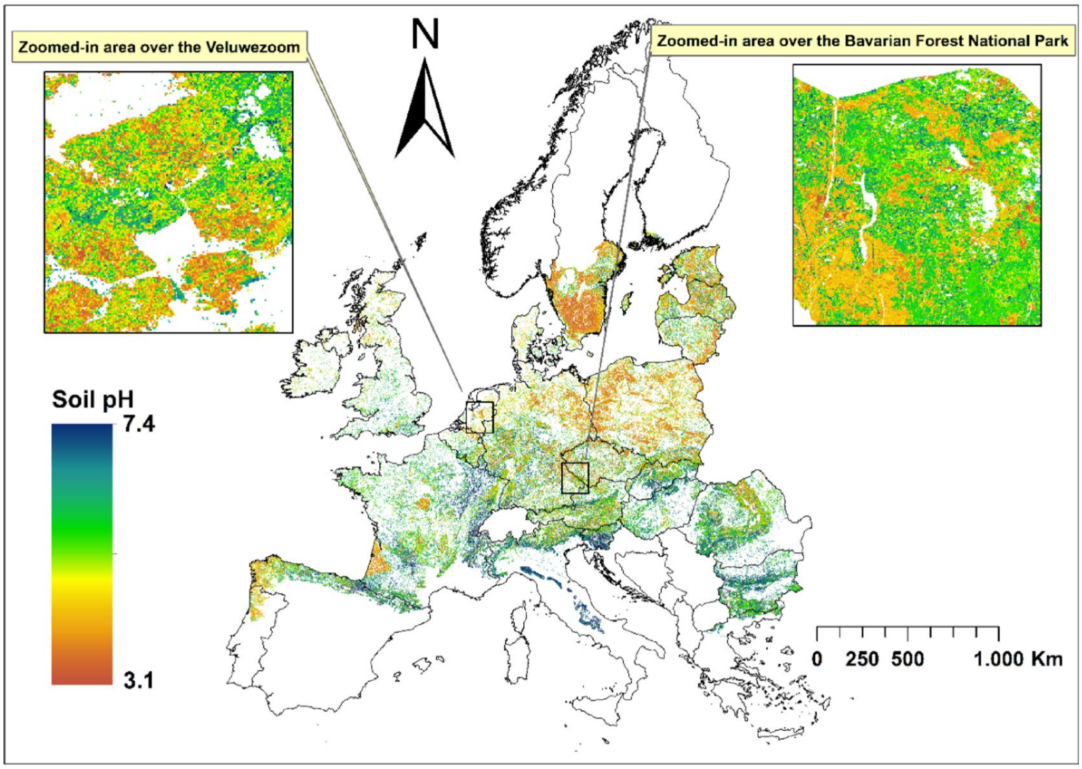

Fig. 3. Spatial distribution of estimated soil pH using Sentinel-2 data, environmental and topographic covariates, and the Random Forest model over the European Temperate forests for 2018. The soil pH model was developed in MATLAB R2024b, and the map visualisation was produced using ArcGIS Pro version 3.1.1.

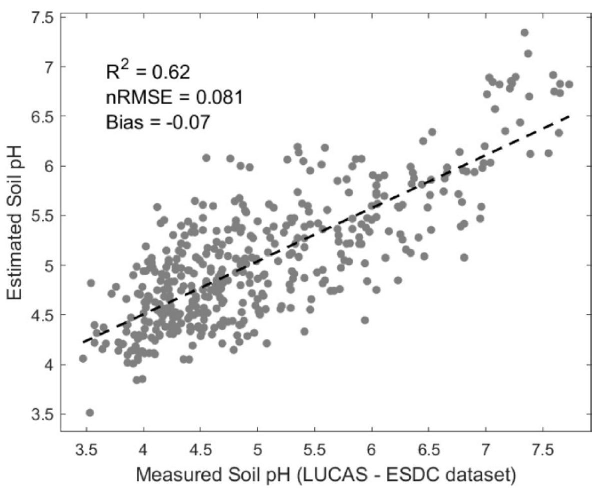

Fig. 4. Measured versus predicted soil pH derived from the random forest analysis using Sentinel-2 data.

As part of the temporal robustness assessment, a soil pH map was generated using Sentinel-2 data from 2024, and the same RF model framework was applied in 2018. The spatial distribution of estimated pH values in 2024 closely resembled the 2018 results, with similar regional patterns and gradients observed across the European temperate forests. Notably, acidic soils remained concentrated in northern and eastern regions, while more neutral to alkaline soils persisted in central and southern Europe. The predicted soil pH range in 2024 extended slightly from 3.4 to 7.8 compared to the 2018 range of 3.1 to 7.4, reflecting a marginal increase in variability but confirming stable model performance over time (see Supplementary Material, Figure S1).

The model's performance is shown in the scatter plot (Fig. 4), where predicted soil pH values are plotted against measured values from the LUCAS dataset. The model achieved an R^2 value of 0.62, indicating a moderate to strong correlation between the estimated and observed soil pH values. The nRMSE was 0.081, and the bias was − 0.07, reflecting a slight underestimation in predicted soil pH values.

On the other hand, model validation against independent in-situ pH measurements from the BIOSPACE project shows a moderate to strong positive correlation between the predicted and observed values at both test sites. As illustrated in Fig. 5, the model achieved an R^2 of 0.59 and an RMSE of 0.18 in the Bavarian Forest National Park, while in the Veluwe National Park, the R^2 was slightly lower at 0.49 but with a smaller RMSE of 0.16. Despite the lower correlation, these results indicate reliable predictive performance and slightly better accuracy at the Veluwe site.

To further evaluate the consistency of the independent in-situ pH measurements from the BIOSPACE project, we compared them with soil pH values obtained from existing large-scale soil databases, namely ISRIC and ESDAC/Ballabio (2019). In the Bavarian Forest region, the BIOSPACE data showed moderate agreement with both sources: ISRIC (R^2 = 0.38, RMSE = 0.36) and ESDAC (R^2 = 0.47, RMSE = 0.29). In the Veluwe region, the comparison revealed a weaker correlation with ISRIC (R^2 = 0.30, RMSE = 0.20), while ESDAC displayed a relatively stronger correlation (R^2 = 0.44, RMSE = 0.22) (see Supplementary Material, Figure S2).

A complementary comparison was conducted with the ISRIC soil pH data by calculating the percentage match between the estimated soil pH map and the ISRIC dataset. Figure 6 displays the spatial distribution of the percentage match across Europe, using a color gradient from red (0–10%) to dark blue (90–100%) to represent increasing levels of agreement. The highest agreement levels (dark blue, 90–100%) were predominantly observed in southern Europe, particularly in Spain, as well as along the Carpathian Mountains in Eastern Europe. In contrast, lower agreement (yellow to green, 30–50%) was evident in northern regions, especially Scandinavia. Central Europe exhibited moderate correspondence, with match percentages generally ranging from 40 to 70%.

To better understand the drivers of soil pH prediction in this study, feature importance was assessed using OOB estimates from the RF model. As shown in Fig. 7, both environmental covariates and Sentinel-2 spectral bands contributed to the model's performance. The selected features—highlighted by the red dashed line—were divided into two main groups. Among the soil's physical properties, AWC, bulk density, and clay content showed the highest importance scores, indicating their substantial role in enhancing prediction accuracy. Regarding the Sentinel-2 spectral data, key predictors included the SWIR bands (i.e., B11 at 1.61 μm and B12 at 2.19 μm), near-infrared (B8 at 0.84 μm), and red-edge bands (B5 at 0.705 μm and B7 at 0.783 μm).

Variation of soil pH across forest types

The spatial variation of estimated soil pH across European temperate forests was assessed by comparing values among different forest types: coniferous, deciduous, and mixed stands. Figure 8 presents the average soil pH values for these three forest types. The results indicate that coniferous forests have an average soil pH of approximately 5.0, while deciduous forests exhibit a higher average pH of around 6.0. Mixed forests show an average soil pH close to that of deciduous forests.

To validate these patterns, we compared them with raw in-situ measurements from the LUCAS dataset. The complementary analysis revealed a consistent trend: coniferous forest soils were generally more acidic than those in deciduous and mixed forests.

Discussion

This study investigates the potential of using freely available Sentinel-2 satellite data to estimate soil pH at a high spatial resolution (30 m) across European temperate forests, offering valuable insights for forest soil monitoring. By integrating Sentinel-2 reflectance with topsoil physical properties and elevation data, we demonstrated the feasibility of generating spatially continuous soil pH estimates over large forested areas. The RF model showed strong predictive performance in internal cross-validation using LUCAS data (R^2 = 0.62; nRMSE = 0.081; bias = − 0.07). External validation with BIOSPACE in-situ samples yielded slightly lower yet still reasonable accuracy (R^2 = 0.59 in the Bavarian Forest, and R^2 = 0.49 in the Veluwe region), indicating that local soil and vegetation

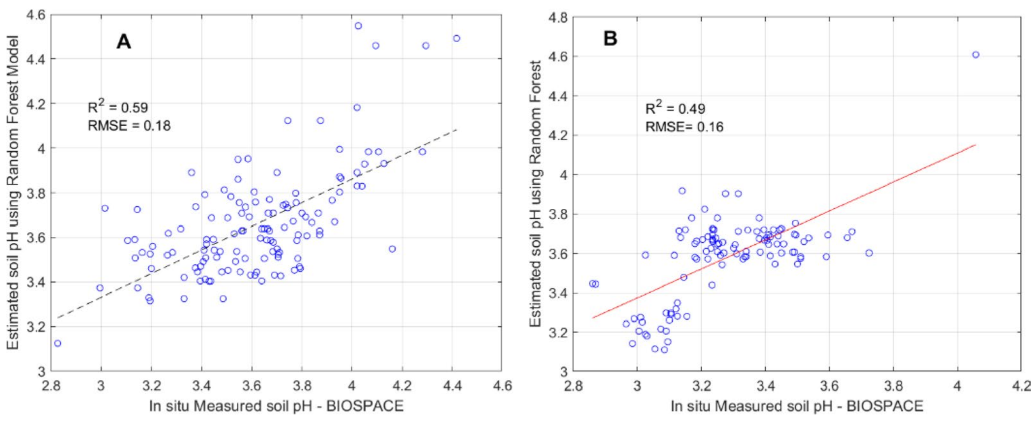

Fig. 5. Comparison of soil pH estimates from this study with independent *in-situ* measurements from the BIOSPACE project for the Bavarian Forest National Park **(A)** and the Veluwe National Park **(B)**. Soil pH was estimated using a Random Forest model based on Sentinel-2 surface reflectance and environmental covariates.

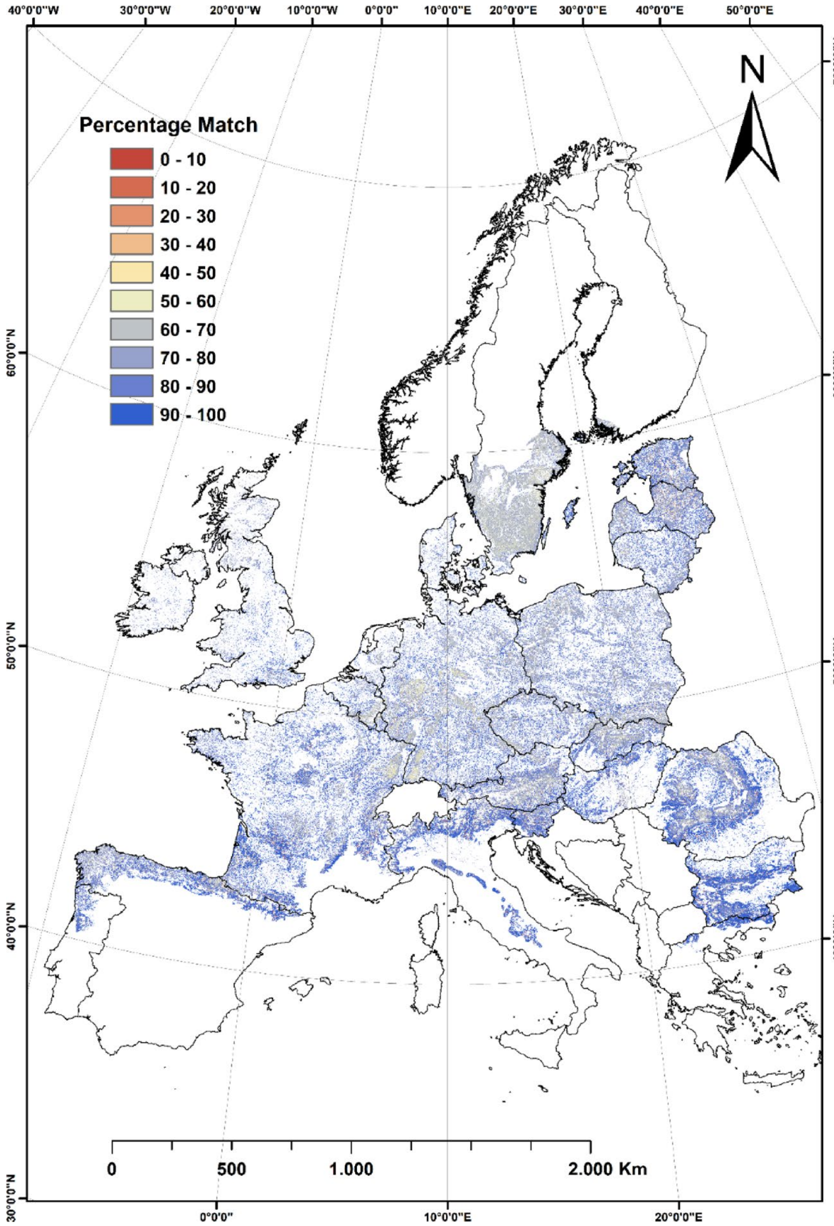

Fig. 6. Percentage match between the estimated soil pH in this study and the ISRIC dataset across Europe. The map was created using ArcGIS Pro version 3.1.1.

heterogeneity may affect model generalizability. Overall, the findings underscore the potential of Sentinel-2-based modelling to support soil pH monitoring in forest ecosystems where ground data are limited.

Spectral and environmental drivers of soil pH Estimation

Spectral reflectance data from Sentinel-2 played an important complementary role in the predictive modelling of soil pH, especially in relation to vegetation and soil interactions. Among the spectral bands, NIR, SWIR, and red-edge bands demonstrated notably higher importance relative to other bands in the Sentinel-2 imagery (Fig. 7). Although environmental covariates such as AWC and clay content had stronger overall importance in the model, these spectral bands (i.e., NIR, SWIR, and red-edge) contributed valuable information by capturing canopy-level responses that reflect underlying soil acidity[35].

Previous studies have shown that high-resolution hyperspectral data can effectively capture soil chemical properties through vegetation reflectance. For example, our earlier analysis using DESIS imagery over temperate forests in Germany and the Netherlands[11] identified the NIR and red-edge bands as significant predictors of soil pH, even though DESIS data lacks SWIR coverage. The agreement between the multispectral Sentinel-2 data and hyperspectral DESIS imagery underscores the importance of these specific wavelengths for inferring soil acidity through vegetation–soil interactions.

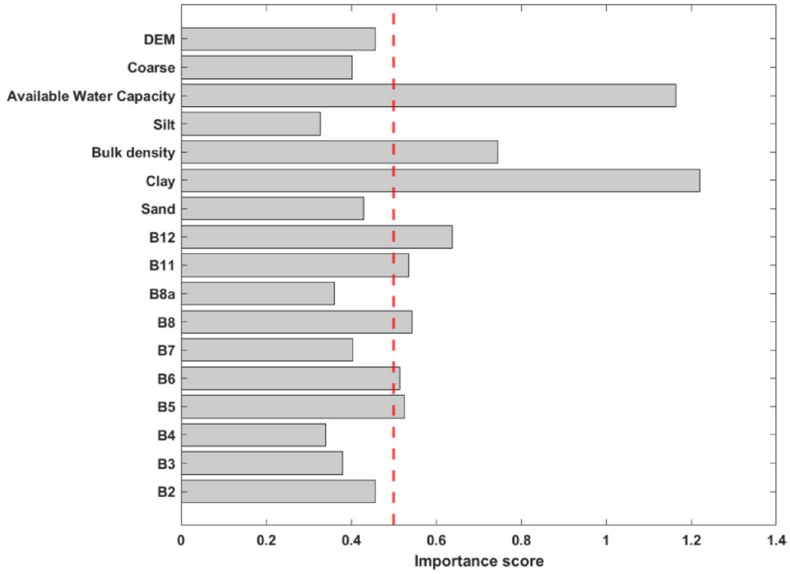

Fig. 7. Feature importance calculated using the random forest method. Features selected in this study are highlighted in dashed red line.

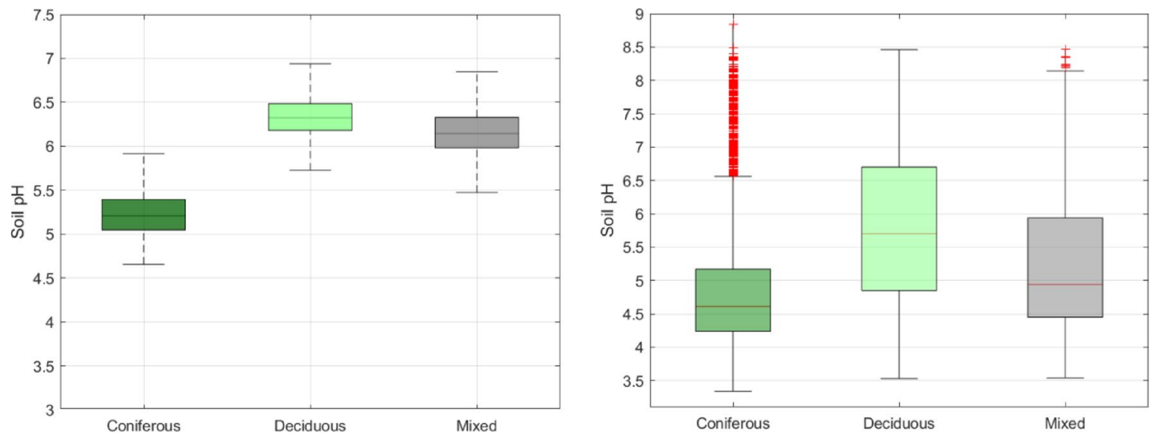

Fig. 8. Variation in soil pH across three forest types—coniferous, deciduous, and mixed—based on estimated pH using Sentinel-2 and environmental covariate data (left) and in-situ pH measurements from the LUCAS dataset (right).

While Sentinel-2 does not provide the spectral richness of hyperspectral platforms like DESIS, EnMAP, or PRISMA, its combination of strategically positioned red-edge and SWIR bands allows it to capture vegetation signals indicative of soil conditions effectively. This makes it a practical choice for large-scale applications, where its spatial resolution and revisit frequency offer advantages over more data-intensive hyperspectral platforms.

A well-established physiological basis supports the utility of these spectral bands. NIR reflectance is particularly sensitive to internal leaf structure, including the spongy mesophyll layer, which strongly influences reflectance in this wavelength region[36,37]. As soil pH regulates nutrient uptake, microbial activity, and root development, it has downstream effects on plant vigor and canopy condition[38,39]. For instance, acidic soils can inhibit nitrogen and phosphorus availability, limiting chlorophyll synthesis and resulting in detectable changes in canopy reflectance[40,41]. As a result, spectral variation in NIR domain can serve as an indirect proxy for edaphic conditions, especially in nutrient-stressed environments. On the other hand, the red-edge region is sensitive to chlorophyll concentration and plant stress, which further enhances the detection of soil acidity effects. Since red-edge reflectance shifts in response to changes in chlorophyll and nitrogen content, it captures subtle physiological responses associated with nutrient limitations and acid-induced stress[39,42]. Therefore, NIR spectral data offer a powerful means to assess soil conditions via plant functioning, particularly in forest ecosystems where direct soil observation is challenging.

SWIR reflectance, while less explored in earlier soil pH studies due to sensor limitations, emerged here as another important contributor. SWIR bands are highly responsive to leaf water content, which itself is influenced by soil moisture, an attribute that is modulated by pH, soil texture, and organic matter content[43,44]. Additionally, SWIR spectral responds to biochemical constituents such as lignin and cellulose, which are affected by plant species composition and stress levels[45]. These biochemical signals may indicates shifts in forest composition or stress responses associated with long-term acidification.

The role of topsoil physical properties

Beyond spectral reflectance, topsoil physical properties played a pivotal role in determining soil pH patterns across temperate forests (Fig. 7). The most influential variables in our model were clay content, bulk density, and AWC, each contributing uniquely to soil acidity regulation through their influence on hydrological, chemical, and biological soil processes.

Among these physical properties, clay content stood out as a critical factor due to its dual function in buffering pH fluctuations and maintaining soil moisture[46]. Soils rich in clay exhibit a higher cation exchange capacity, allowing them to retain and exchange base cations such as calcium, magnesium, and potassium key buffers against acidification[47,48]. This capacity stabilises soil pH under environmental stress and moderates the availability of toxic elements like aluminium, which become more soluble at low pH levels. Additionally, clay-rich soils hold more water, which supports microbial activity and root respiration, further influencing nutrient cycling and acid-base reactions[49].

Complementing the role of clay, bulk density emerged as another significant yet often overlooked contributor to pH regulation, primarily through its effect on soil porosity and aeration. Higher bulk density results in compacted soils with reduced oxygen diffusion, leading to anaerobic microsites and altered redox conditions. These processes affect the decomposition of organic matter and the release of hydrogen and other ions that can influence pH[50,51]. Dense soils also restrict root proliferation and microbial habitat quality, indirectly reducing buffering capacity and amplifying pH variability at the microscale.

Adding to this complex interplay, AWC directly influences the soil's ability to retain and supply water for plant and microbial use. In ecosystems subject to periodic drought or extreme precipitation, AWC becomes especially important. High AWC buffers the ecosystem against moisture stress, helping maintain microbial functioning and root nutrient uptake, both essential for stable pH regulation[52]. Conversely, in coarse-textured soils with low AWC, the leaching of base cations increases, promoting acidification and the accumulation of hydrogen ions. These processes are particularly pronounced under changing climatic conditions, making AWC a critical parameter for future soil health resilience.

Model validation and comparative performance

Our approach to soil pH mapping offers significant advancements compared to existing large-scale studies, particularly in terms of spatial resolution and the integration of spectral data. For example, Ballabio, et al.[20] developed a European-wide soil pH map at a 500 m resolution using the LUCAS dataset and vegetation indices such as NDVI and EVI derived from MODIS. While their methodology incorporated these indices as part of a broader digital soil mapping framework, our study stands apart by directly utilizing Sentinel-2 surface reflectance data at a much finer spatial resolution of 30 m. This higher resolution is crucial as it enables a more precise representation of soil pH variability across diverse forest ecosystems, enhancing its utility for forest management and environmental monitoring applications.

Building on similar methodologies, Poggio, et al.[53] also used remote sensing data alongside environmental covariates to model soil pH across European landscapes. While their study demonstrated the feasibility of using remote sensing data at a coarser resolution, our approach takes a step further by integrating Sentinel-2 spectral data with topsoil physical properties and elevation data. This combination allows for high-resolution, site-specific pH estimates, which improve the model's ability to capture localized soil variability a key factor often overlooked in coarser-scale studies.

Further validation of our model using 2024 Sentinel-2 imagery revealed its robustness, with predicted soil pH distributions closely aligning with those from 2018. Notably, the predicted pH range expanded slightly, from 3.4 to 7.8 in 2024 compared to 3.1 to 7.4 in 2018. This minor change suggests temporal consistency and highlights the model's potential for long-term monitoring. Unlike previous studies, such as those conducted by Ballabio, et al.[20] and Poggio, et al.[53], which did not explicitly consider multi-year validation, our model's stability over time further underscores its value for continuous soil health assessments. This is particularly important for forest management and environmental policy, where the ability to track soil health over multiple years is crucial, especially in areas vulnerable to acidification or climate-driven changes.

When comparing our results to other large-scale datasets, such as the ISRIC global soil property maps[21], our model outperformed these datasets, especially in regions with high ecological homogeneity. For instance, in the Veluwe region of the Netherlands, our model achieved a higher R^2 (0.59) compared to the ISRIC dataset (R^2 = 0.30). This enhanced performance is likely due to the incorporation of local soil data from the LUCAS project and the use of high-resolution Sentinel-2 data, which are better suited to capturing the fine-scale variability of soil pH in forested ecosystems.

These comparisons highlight the potential of high-resolution remote sensing data, such as Sentinel-2, for improving soil pH estimation across forested regions. Unlike the coarser-resolution datasets used in previous studies, our approach provides a more detailed and accurate understanding of soil conditions across Europe. This advancement has important implications for forest management and policy, particularly in efforts to mitigate soil acidification and support biodiversity conservation.

Regional and ecological variability

Regional comparisons with ISRIC soil pH data revealed clear spatial differences in model performance across Europe. Lower agreement in northern regions such as Scandinavia likely reflects a combination of inherent soil characteristics and environmental conditions. These include the widespread presence of naturally acidic podzolic soils, reduced biological activity due to long winters, and persistent snow cover, which may obscure or distort canopy reflectance and reduce model reliability[54,55]. Moreover, shorter growing seasons and pronounced seasonal variability in vegetation can lead to less stable spectral signals, thereby limiting the accuracy of pH predictions derived from remote sensing.

Conversely, higher model performance in central and southern Europe likely results from more stable climatic conditions, longer vegetation periods, and more homogeneous forest and soil types. These regions are dominated by Luvisols and Cambisols soils that tend to have higher base saturation and buffering capacity and feature relatively consistent canopy characteristics that enhance the quality of spectral observations[56,20]. The improved agreement here suggests that both ecological and climatic stability support more accurate remote-sensing-based modelling of soil properties.

Beyond regional contrasts, our analysis also revealed strong ecological patterns linked to forest type. Soil pH was consistently lower under coniferous forests, followed by mixed stands, and highest under deciduous forests. This trend aligns with well-documented differences in litter quality and decomposition rates among forest types. Coniferous species, such as spruce and pine, produce acidifying litter rich in lignin and resin, decomposing slowly and returning fewer base cations to the soil[57]. In contrast, deciduous species like beech and oak generate litter that decomposes more rapidly and supports greater microbial activity, leading to enhanced nutrient cycling and a less acidic soil environment. These ecological differences contribute to the spatial heterogeneity in soil pH observed across the study area and underscore the need to account for vegetation type when interpreting soil predictions based on canopy reflectance. Together with regional climatic and soil differences, forest composition plays a critical role in shaping the environmental controls of soil acidity.

Implications, applications, limitations, and future directions

The high-resolution (30 m) soil pH maps generated in this study perform strongly as a decision-support tool for adaptive forest management, environmental monitoring, and EU policy implementation. By resolving site-specific nutrient dynamics, species suitability, and acidification hotspots, which are features often obscured in coarser (250–500 m) datasets, these maps enable more precise reforestation planning, species selection, soil-liming treatments, and early warning of acidification risks. Leveraging freely available Sentinel-2 imagery, our approach provides a scalable, cost-effective framework for routine soil-health monitoring across regional to continental scales. This capability directly supports EU strategies such as the Biodiversity Strategy and Forest Europe, particularly in ecologically sensitive zones like Scandinavia, where acidification pressures are highest[58]. Although developed for temperate forests, the underlying random-forest framework used in this study is readily adaptable to other biomes through regional recalibration and the inclusion of additional covariates, such as soil organic carbon content, microbial-activity proxies, and terrain-derived moisture indices. For example, integrating high-resolution pH maps with biodiversity indicators, such as species-richness metrics, and microbial community datasets could further facilitate multi-dimensional assessments of forest health and resilience by revealing how soil chemistry interacts with biological diversity under varying environmental pressures[22].

However, despite these strengths, the approach is hampered by the current unavailability of (regional) high-resolution climate variables such as precipitation and temperature (e.g., the best-available global climate products are coarser than 250 m), rendering the integration of ancillary covariates with our 30 m Sentinel-2 composites impractical. The agreement between our model pH product and existing soil pH data declined at higher latitudes, particularly across Scandinavia and the Baltic states, where percentage-match values fall below 50%. This decrease stems from a combination of ecological and technical factors: (i) evergreen needle-leaf canopies prevalent in boreal zones exhibit muted spectral responses to pH-driven nutrient shifts[59,60]; (ii) thick organic layers and slow mineralisation rates decouple soil chemistry from canopy reflectance[61,62]; and (iii) persistent cloud cover, fog, or snow contamination during the short leaf-on window further reduces the chance of obtaining remote sensing images, especially large area composites as used in this study.

To address these spatial inconsistencies and enhance model generalizability, future research should incorporate the Sentinel-2 imagery used in this study with lidar-derived metrics of canopy height, leaf area index, and gap fraction from GEDI or ICESat-2 alongside C-band SAR backscatter from Sentinel-1 to better characterise sub-canopy moisture and ecosystem structural variables[63,64]. Embedding ecosystem structure and moisture variables within a stacked random-forest or deep-learning framework will train models on biome-specific relationships and decouple non-soil signals from true soil chemistry. At the same time, adopting phenology-aware compositing strategies, such as regionally tailored temporal windows or high-percentile indices that preserve spectral extremes, will prevent the "flattening" effects inherent to median composites[65,66]. Finally, as finer-scale climate layers and microbial community datasets become available, incorporating these complementary covariates will better constrain nutrient-availability contexts, strengthen the ecological interpretability of pH retrievals, and support evidence-based interventions to safeguard forest ecosystems under changing environmental conditions.

Conclusion

This study highlights the significant potential of using Sentinel-2 spectral data combined with soil physical properties and in-situ measurements to estimate soil pH in European temperate forests. One of the challenges posed by forest canopies is their potential to interfere with the retrieval of accurate spectral reflectance data from the forest floor, as canopy cover can obscure the underlying soil and vegetation. Although this limitation was not directly investigated in our study, it is an inherent factor in remote sensing applications over forested areas. Nevertheless, our findings demonstrate that accurate soil pH prediction is feasible, at a 30-m spatial resolution, suggesting that the combined approach can effectively mitigate some of these challenges. The NIR, SWIR, and red-edge spectral bands emerged as crucial predictors of soil pH, highlighting the indirect influence of soil chemistry on vegetation health and spectral reflectance. Our analysis revealed that forest type plays a critical role in soil pH variability, with coniferous forests exhibiting the lowest pH levels and deciduous forests the highest.

Moreover, the application of RF model in this study offers a scalable and cost-effective approach to soil pH monitoring across large forested areas, supporting more informed decision-making in forest management and environmental policy. The integration of high-resolution remote sensing data with in-situ measurements paves the way for more sustainable forest management practices, ensuring that forests remain resilient in the face of increasing environmental pressures, including nitrogen deposition and climate change.

Overall, the methods and results presented here have broad implications for biodiversity conservation, soil health assessment, and long-term forest ecosystem monitoring, particularly within the context of climate change and land-use policies.

Data availability

The Sentinel-2 imagery used in this study is freely available from the Copernicus Open Access Hub (https://scihub.copernicus.eu/). The topsoil physical property data (e.g., pH, bulk density, clay content) were obtained from the LUCAS Soil database, which is accessible via the European Soil Data Centre (ESDAC) (https://esdac.jrc.ec.europa.eu/). Additional reference data from the ISRIC SoilGrids database can be accessed at https://soilgrids.org/. All data used in this study are publicly available, and the specific processing scripts and model outputs can be provided by the corresponding author upon reasonable request.

Received: 22 January 2025; Accepted: 23 May 2025
Published online: 05 August 2025

References

1. Brady, N. The nature and properties of soils. (1984).
2. Rousk, J., Brookes, P. C. & Bååth, E. Contrasting soil pH effects on fungal and bacterial growth suggest functional redundancy in carbon mineralization. *Appl. Environ. Microbiol.* **75**, 1589–1596 (2009).
3. Sparks, D. L., Singh, B. & Siebecker, M. G. *Environmental Soil Chemistry* (Elsevier, 2022).
4. Arnold, G. *Soil Acidification and Imbalanced Nutrient Availability in Scots Pine Forest Soils in the Netherlands: Causes, Extent and Control* (Wageningen University and Research, 1993).
5. Kauppi, P., Kämäri, J., Posch, M., Kauppi, L. & Matzner, E. Acidification of forest soils: model development and application for analyzing impacts of acidic deposition in Europe. *Ecol. Model.* **33**, 231–253 (1986).
6. Skidmore, A., Abdullah, H., Torres-Rodriguez, A. & Neinavaz, E. Forest soils further acidify in core natura 2000 areas amongst unaware government policy. *Ecol. Ind.* **159**, 111621 (2024).
7. Basu, S. Chemical and biochemical processes for NOx control from combustion off-gases. *Chem. Eng. Commun.* **194**, 1374–1395 (2007).
8. Stevens, C. J., Dise, N. B., Mountford, J. O. & Gowing, D. J. Impact of nitrogen deposition on the species richness of grasslands. *Science* **303**, 1876–1879 (2004).
9. Seidl, R. et al. Forest disturbances under climate change. *Nat. Clim. Change.* **7**, 395–402 (2017).
10. Rossel, R. V., Adamchuk, V., Sudduth, K., McKenzie, N. & Lobsey, C. Proximal soil sensing: an effective approach for soil measurements in space and time. *Adv. Agron.* **113**, 243–291 (2011).
11. Abdullah, H. et al. Temperate forest soil pH accurately quantified with image spectroscopy. *Remote Sens. Applications: Soc. Environ.* **34**, 101161 (2024).
12. Ghazali, M. F., Wikantika, K., Harto, A. B. & Kondoh, A. Generating soil salinity, soil moisture, soil pH from satellite imagery and its analysis. *Inform. Process. Agric.* **7**, 294–306 (2020).
13. Charishma, D. et al. Estimation of top soil properties by Sentinel-2 imaging. *Geol. Ecol. Landscapes* **8**, 1–10 (2024).
14. Gogumalla, P. et al. Detecting soil pH from open-source remote sensing data: a case study of Angul and Balangir districts, Odisha state. *J. Indian Soc. Remote Sens.* **50**, 1275–1290 (2022).
15. Xia, C. & Zhang, Y. Comparison of the use of Landsat 8, Sentinel-2, and Gaofen-2 images for mapping soil pH in Dehui, Northeastern China. *Ecol. Inf.* **70**, 101705 (2022).
16. Davis, E., Wang, C. & Dow, K. Comparing Sentinel-2 MSI and Landsat 8 OLI in soil salinity detection: A case study of agricultural lands in coastal North Carolina. *Int. J. Remote Sens.* **40**, 6134–6153 (2019).
17. Vibhute, A. D. & Kale, K. V. Mapping several soil types using hyperspectral datasets and advanced machine learning methods. *Results Opt.* **12**, 100503 (2023).
18. Angelopoulou, T. et al. Evaluation of airborne hyspex and spaceborne PRISMA hyperspectral remote sensing data for soil organic matter and carbonates Estimation. *Remote Sens.* **15**, 1106 (2023).
19. Webb, H., Barnes, N., Powell, S. & Jones, C. Does drone remote sensing accurately estimate soil pH in a spring wheat field in Southwest Montana? *Precision Agric.* **22**, 1803–1815 (2021).
20. Ballabio, C. et al. Mapping LUCAS topsoil chemical properties at European scale using Gaussian process regression. *Geoderma* **355**, 113912 (2019).
21. Hengl, T. et al. SoilGrids250m: global gridded soil information based on machine learning. *PLoS One.* **12**, e0169748 (2017).
22. Skidmore, A. K. et al. eDNA biodiversity from space: predicting soil bacteria and fungi alpha diversity in forests using DESIS satellite remote sensing. *Int. J. Remote Sens.* 1–31. https://doi.org/10.1080/01431161.2025.2464958 (2025).
23. Olson, D. M. et al. Terrestrial ecoregions of the world: A new map of life on Earth: A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. *BioScience* **51**, 933–938 (2001).
24. Copernicus, L. & Monitoring, S. in *Copernicus Land Monitoring Service* (2018).
25. Heurich, M., Beudert, B., Rall, H. & Křenová, Z. National parks as model regions for interdisciplinary long-term ecological research: the Bavarian Forest and Šumavá National Parks underway to transboundary ecosystem research. In *Long-term ecological research: Between theory and application* (eds Müller, F. et al.) 327–344 (Springer, Dordrecht, 2010).
26. Karungi, J. et al. Elevation and cropping system as drivers of microclimate and abundance of soil macrofauna in coffee farmlands in mountainous ecologies. *Appl. Soil. Ecol.* **132**, 126–134 (2018).
27. Jenny, H. *Factors of Soil Formation: a System of Quantitative Pedology* (Courier Corporation, 1994).
28. Körner, C. The use of 'altitude'in ecological research. *Trends Ecol. Evol.* **22**, 569–574 (2007).
29. Chalchissa, F. B. & Kuris, B. K. Modeling the impacts of extreme climate scenarios on soil acidity (pH and exchangeable aluminum) in Abbay River Basin, Ethiopia. *Heliyon* 10 (2024).
30. Miles, N., van Antwerpen, R. & Ramburan, S. Soil organic matter under sugarcane: levels, composition and dynamics. *Proc. S. Afr. Sug. Technol. Assoc.* **89**, 161–169 (2016).
31. Zhong, Z. et al. Relationship between soil organic carbon stocks and clay content under different Climatic conditions in central China. *Forests* **9**, 598 (2018).
32. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).

33. Liaw, A. & Wiener, M. Classification and regression by randomForest. *R News* **2**(3), 18–22 (2002).
34. Heděnec, P. et al. Mycorrhizal association of common European tree species shapes biomass and metabolic activity of bacterial and fungal communities in soil. *Soil Biol. Biochem.* **149**, 107933 (2020).
35. Mulder, V., De Bruin, S., Schaepman, M. E. & Mayr, T. The use of remote sensing in soil and terrain mapping—A review. *Geoderma* **162**, 1–19 (2011).
36. Gausman, H. W. Leaf reflectance of near-infrared. *Photogrammetric Eng.* **40**, 183–191 (1974).
37. Slaton, M. R., Hunt, R. Jr, Smith, W. K. & E. & Estimating near-infrared leaf reflectance from leaf structural characteristics. *Am. J. Bot.* **88**, 278–284 (2001).
38. Alam, S. M., Naqvi, S. S. M. & Ansari, R. Impact of soil pH on nutrient uptake by crop plants. *Handb. Plant. Crop Stress.* **2**, 51–60 (1999).
39. Robson, A. *Soil Acidity and Plant Growth* (Elsevier, 2012).
40. Curran, P. J. Remote sensing of foliar chemistry. *Remote Sens. Environ.* **30**, 271–278 (1989).
41. Sparks, D. L. Environmental soil chemistry: an overview. *Environ. Soil. Chem.* **2**, 1–42 (2003).
42. Tyler, G. & Olsson, T. Plant uptake of major and minor mineral elements as influenced by soil acidity and liming. *Plant. Soil.* **230**, 307–321 (2001).
43. Ceccato, P., Flasse, S., Tarantola, S., Jacquemoud, S. & Grégoire, J. M. Detecting vegetation leaf water content using reflectance in the optical domain. *Remote Sens. Environ.* **77**, 22–33 (2001).
44. Zárate-Valdez, J. L., Zasoski, R. J. & Läuchli, A. E. Short-term effects of moisture content on soil solution pH and soil Eh. *Soil Sci.* **171**, 423–431 (2006).
45. Asner, G. P. Biophysical and biochemical sources of variability in canopy reflectance. *Remote Sens. Environ.* **64**, 234–253 (1998).
46. Weil, R. & Brady, N. The nature and properties of soils, 15th edn., edited by: Fox, D. *D.[Google Scholar]* (2016).
47. Phadungchewit, Y. The role of pH and soil buffer capacity in heavy metal retention in clay soils. PhD thesis (McGill University, Montreal, Quebec, Canada, 1990).
48. Gong, Y., Cao, Q. & Sun, Z. The effects of soil bulk density, clay content and temperature on soil water content measurement using time-domain reflectometry. *Hydrol. Process.* **17**, 3601–3614 (2003).
49. Dvořáčková, H., Dvořáček, J., Hueso González, P. & Vlček, V. Effect of different soil amendments on soil buffering capacity. *PloS One*. **17**, e0263456 (2022).
50. Brosinsky, A. et al. Analysis of spectral vegetation signal characteristics as a function of soil moisture conditions using hyperspectral remote sensing. *J. Indian Soc. Remote Sens.* **42**, 311–324 (2014).
51. Kooch, Y., Haghverdi, K., Nouraei, A. & Francaviglia, R. Soil properties are affected by vegetation types in a semi-arid mountain landscape. *Pedobiologia* **102**, 150932 (2024).
52. Koupai, J. A., Eslamian, S. S. & Kazemi, J. A. Enhancing the available water content in unsaturated soil zone using hydrogel, to improve plant growth indices. *Ecohydrol. Hydrobiol.* **8**, 67–75 (2008).
53. Poggio, L. et al. SoilGrids 2.0: producing soil information for the Globe with quantified Spatial uncertainty. *Soil* **7**, 217–240 (2021).
54. Akselsson, C., Hultberg, H., Karlsson, P. E., Karlsson, G. P. & Hellsten, S. Acidification trends in South Swedish forest soils 1986–2008—Slow recovery and high sensitivity to sea-salt episodes. *Sci. Total Environ.* **444**, 271–287 (2013).
55. Darmody, R. G., Thorn, C. E., Schlyter, P. & Dixon, J. C. Relationship of vegetation distribution to soil properties in Kärkevagge, Swedish Lapland. *Arct. Antarct. Alp. Res.* **36**, 21–32 (2004).
56. Šantrůčková, H., Cienciala, E., Kaňa, J. & Kopáček, J. The chemical composition of forest soils and their degree of acidity in central Europe. *Sci. Total Environ.* **687**, 96–103 (2019).
57. Chodak, M., Klimek, B. & Niklińska, M. Composition and activity of soil microbial communities in different types of temperate forests. *Biol. Fertil. Soils*. **52**, 1093–1104 (2016).
58. European Environment, A. *Monitoring of Surface Water Acidification* (European Environment Agency, Copenhagen, 1996).
59. Lhotáková, Z. et al. Foliage biophysical trait prediction from laboratory spectra in Norway Spruce is more affected by needle age than by site soil conditions. *Remote Sens.* **13**, 391 (2021).
60. Croft, H., Chen, J., Zhang, Y. & Simic, A. Modelling leaf chlorophyll content in broadleaf and needle leaf canopies from ground, CASI, Landsat TM 5 and MERIS reflectance data. *Remote Sens. Environ.* **133**, 128–140 (2013).
61. Näsholm, T. et al. Boreal forest plants take up organic nitrogen. *Nature* **392**, 914–916 (1998).
62. Markiet, V. & Mõttus, M. Estimation of boreal forest floor reflectance from airborne hyperspectral data of coniferous forests. *Remote Sens. Environ.* **249**, 112018 (2020).
63. Dubayah, R. et al. The global ecosystem dynamics investigation: High-resolution laser ranging of the Earth's forests and topography. *Sci. Remote Sens.* **1**, 100002. https://doi.org/10.1016/j.srs.2020.100002 (2020).
64. Udali, A., Lingua, E. & Persson, H. J. Assessing forest type and tree species classification using Sentinel-1 C-band SAR data in Southern Sweden. *Remote Sens.* **13**, 3237 (2021).
65. Lewińska, K. E., Frantz, D., Leser, U. & Hostert, P. Usable observations over Europe: evaluation of compositing windows for Landsat and Sentinel-2 time series. *Eur. J. Remote Sens.* **57**, 2372855 (2024).
66. Xie, S., Liu, L. & Yang, J. Time-series model-adjusted percentile features: improved percentile features for land-cover classification based on Landsat data. *Remote Sens.* **12**, 3091 (2020).

Acknowledgements

This research was funded through the BIOSPACE project, financially supported by the EuropeanResearch Council (ERC) within the framework of the European Union's Horizon 2020 research andinnovation programme, as indicated by grant agreement number 834709. The authors would like toexpress their sincere appreciation for the support received during the field data collection processfrom the Bavarian Forest National Park (BNFP), the Bavarian State Forest Enterprise (BayerischeStaatsforsten), and Neuburg Forest in Germany, and the Bavarian Forest Data Pool. We are gratefulfor the support of the Hoge Veluwe National Park, Veluwezoom National Park and Het Loo RoyalEstate in the Netherlands, and permission to access and conduct field work in the Veluwe area. We gratefully acknowledge that this research was initially supported by the BIOSPACE project, funded by the European Research Council under the European Union's Horizon 2020 programme (Grant Agreement No. 834709), and is now carried forward under the Horizon Europe project OBSGESSION (Grant Agreement No. 20004485), which provides our main funding. We also thank the Bavarian Forest National Park, the Bavarian State Forest Enterprise (Bayerische Staatsforsten), Neuburg Forest and the Bavarian Forest Data Pool in Germany, as well as Hoge Veluwe National Park, Veluwezoom National Park and Royal Estate Het Loo in the Netherlands, for their invaluable assistance with field data collection.

Author contributions

H.Abdullah: Conceptualization, Data curation, Formal analysis, Software, Visualization,, Investigation, Methodology, Validation, Writing – original draft, Writing – review& editing. A.K. Skidmore: Conceptualization, Investigation, Methodology, Validation, Writing – review & editing, Project administration. A. Siegenthaler: Writing – review & editing. E. Neinavaz: Writing – review & editing.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-03942-4.

Correspondence and requests for materials should be addressed to H.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.