



Assessing Groundwater Spring Potential Using Random Forest: A Comparison of High-Resolution Austrian and Pan-European Datasets

Master Thesis

for the attainment of the Master's degree "Master of Science",
abbreviated "MSc"

submitted within the University Master Program for Further Education

"Geographical Information Science & Systems – (UNIGIS MSc)"

at the Department of Geoinformatics - Z_GIS,
Faculty of Digital and Analytical Sciences,
University of Salzburg

submitted by

BSc Regina Folzwinkler

Supervisor:
Assoz. Prof. Dr. Dirk Tiede

Ramingstein, August 2026

Declaration

"I hereby declare that I have written this thesis on my own and that I have not used any sources or other tools and resources other than those indicated. All passages that have been taken literally or in substance from the cited sources are clearly identified as such.

I confirm that I have not used any illustrations from textbooks or other sources unless they are subject to a license permitting free use.

I further confirm that I have used artificial intelligence (AI)-based applications exclusively for those purposes explicitly permitted in the [guidelines on the use of generative AI in the UNIGIS program](#), in its current version. Any deviations from this policy must be explicitly approved in writing by the supervisor. I bear full responsibility for all content of this thesis.

This thesis has not previously been submitted in the same or a similar form as a Bachelor's/Master's/Diploma thesis/Dissertation/Master Thesis."

Regina Folzwinkler

Content

List of Tables	5
List of Figures	6
List of Abbreviations	8
Abstract	10
1 Introduction	11
1.1 Intention	11
1.2 Background	11
1.3 Research Aims and Purpose	13
1.4 Previous research	15
1.5 Hydrological Background	16
2 Methods	17
2.1 Study Area	17
2.1.1 Spring Points and Pseudo-absence Points	17
2.2 Geological Classification	18
2.3 Sentinel images from Google Earth Engine (GEE)	21
2.3.1 Sentinel - 2 images	21
2.3.2 Sentinel -1 images	22
2.3.3 Explorative wetness mask	23
2.4 Terrain Parameters	23
2.5 Hydrological Indices	29
2.6 Soil Data	33
2.7 Training Data Preparation	33
2.8 Random Forest Classification	37
2.8.1 Model Validation and Uncertainty Assessment	38
2.8.2 Design Sensitivity	38
2.8.3 Spatial Cross-Validation	39
2.9 Correlation Analysis of the Predictor Variables	39
2.10 Spring Potential Maps	40
2.11 Independent Evaluation of the Prediction Maps	41
2.11.1 Success Rate	41
2.11.2 Independently Mapped Waterlogged Areas	41
2.11.3 Qualitative Expert Review	42
2.12 Ablation Experiments	42
2.12.1 Removal of the Satellite Variables	42
2.12.2 Within-Area Variability of the Predictors	43
2.12.3 Local discriminative power of the predictors	43

3	Results	44
3.1	Model Performance Metrics	44
3.2	Sensitivity to the Sampling Design	44
3.3	Variable Importance	46
3.4	Sensitivity of the tree numbers	47
3.5	Correlation Analysis	48
3.6	Spring Potential Maps	51
3.7	Spatial cross-validation	56
3.8	Independent Evaluation of the Maps	57
3.8.1	Success rate	57
3.8.2	Independently mapped waterlogged areas	59
3.8.3	Qualitative expert review	59
3.9	Ablation experiments	60
3.9.1	Removing the satellite variables	60
3.9.2	Predictors without variation inside a project area	60
3.9.3	Removing the differences between the project areas	60
4	Discussion	63
5	Conclusion	71
	References	73
	Data References	75

List of Tables

Table 1 - Classification of lithological units into permeability classes.	20
Table 2 - Overview of the Austrian project areas, including area size [m ²] and the number of mapped springs per area.	35
Table 3 - Overview of the European project areas, including area size [m ²] and the number of mapped springs per area.	37
Table 4 - Comparison of the Austrian (AT) and European (EU) model. Top: final models (complete fixed-per-area pseudo-absence sets with class compensation; single runs), used for the spring potential maps. Bottom: Monte Carlo uncertainty analysis (100 realisation; balanced design without class compensation, pseudo-absences drawn from the fixed-per-area pool). ..	44
Table 5 - Model performance of the tested sampling designs and class ratios (single runs, Random Forest, 500 trees). Single runs are subject to random variation of up to ~0.1 MCC; reliable uncertainty estimates are given in table 4.1. The variants “fixed-per-area, full set, compensation” are the final model configuration used for the spring potential maps. The spatially uncontrolled design (EU MCC 0.56, AT 0.49) is only reported in the text.	45
Table 6 - Mean variable importance by predictor group (mean ± SD over 100 Monte Carlo realisations, balanced design, pseudo-absences drawn from the fixed-per-area pool). Topography: DEM and derived variables including the hydrological indices (slope, aspect, general curvature, plan curvature, profile curvature, TWI, SPI, STI, flow accumulation); Remote sensing: Sentinel-1 and Sentinel-2, Geology: 6 permeability classes. Only the top 20 of 22 predictor variables are considered. This is why the group values do not sum up to 100%.....	46
Table 7 - Model performance with different numbers of trees (final model configuration: fixed-per-area pseudo-absences with class compensation; mean ± SD over 10 repeated runs per setting).	47

List of Figures

Figure 1 Water retention planning in Southern France, part of the training areas. Plan: Oliver Krische, Holzer & Krische GmbH, (2022).	12
Figure 2 - Pond within a pond system of a water retention facility at Krameterhof. Photograph: Josef Andreas Holzer, 2012.	13
Figure 3 - Lithological units of the Austrian project areas reclassified into six permeability classes (L1–L6).	19
Figure 4 - Lithological units of the European project areas reclassified into six permeability classes (L1–L6).	20
Figure 5 - Selected Sentinel-2 false-color composite (B8–B4–B3) of one Austrian project area used for wetness assessment.	22
Figure 6 - Sentine -1 VH composite derived from median backscatter during February–March (red), April–May (green), and their difference (blue). The composite was used to visualize seasonal changes in radar backscatter. The displayed colours do not represent natural colours, but a false-colour composite used to visualize seasonal changes in radar backscatter.	23
Figure 7 - High-resolution (5 m) digital elevation model used for the Austrian model.	24
Figure 8 - Low-resolution (30m) Copernicus Digital Surface Model used for the European model.	25
Figure 9 - Slope (degree) derived from the Austrian 5 m DEM.	26
Figure 10 - Slope (degrees) derived from Copernicus 30m DSM.	26
Figure 11 - Aspect derived from the Austrian 5 m DEM.	27
Figure 12 - Aspect derived from Copernicus 30m DSM.	27
Figure 13 - Curvature derived from the Austrian 5 m DEM.	28
Figure 14 - Curvature derived from Copernicus 30m DSM.	28
Figure 15 - Flow accumulation derived from the high-resolution Austrian DEM (5 m).	30
Figure 16 - Flow accumulation derived from the Copernicus DSM (30 m) used for the European model.	30
Figure 17 - Topographic Wetness Index (TWI) derived from the Austrian 5 m DEM.	31
Figure 18 - Topographic Wetness Index (TWI) derived from the Copernicus 30m DSM.	31
Figure 19 - Stream Power Index (SPI) derived from the Austrian 5m (DEM).	32
Figure 20 - Stream Power Index (SPI) derived from the Copernicus 30m DSM.	32
Figure 21 - Spatial Distribution of Austrian project areas and their project IDs.	34
Figure 22 - Spatial Distribution of Project Areas in Europe and their Project IDs.	36
Figure 23 - Importance of single predictor variables, given as mean over 100 Monte Carlo realisations with the standard deviation. Only the top 20 of 22 predictor variables are considered. The mean predictor value is 4,5%.	47
Figure 24 - Sensitivity of both models to the number of trees. Each point is the mean of 10 independent runs; the error bars and the shaded band in the left panel give the standard deviation over these runs. The differences between 100, 250, 500 and 1000 trees are smaller than the scatter between repeated runs with the same setting, so the number of trees was not decisive for the result. All runs use the balanced training set and the random 70/30 split described in section 3.8.1.	48
Figure 25 - Single variable correlation with spring presence in both models.	48

Figure 26 - Pearson and Spearman correlation of each predictor with spring presence, for the Austrian model (left) and the European model (right). Grey points show the linear point-biserial correlation, blue points the rank correlation. Variables shown in orange.	49
Figure 27 - Correlation between predictor variables, calculated only at the verified spring locations. Only pairs with $ r \geq 0.70$ are shown. Every strong pair consists of two variables from the same family, and no strong correlation occurs between families. Correlated predictors share their importance in a random forest, which is why the variable importance is reported by group rather than for single variables.	50
Figure 28 - Prediction Map Austria trained with the full dataset.	52
Figure 29 - Prediction Map created with LOAO.	53
Figure 30 - European Prediction Map created with full dataset.	54
Figure 31 - Prediction Map created with LOAO.	55
Figure 32 - Performance of the Austrian and the European model under random five-fold cross-validation and under leave-one-area-out validation, and the same spatial validation with the six satellite variables removed. All values except the AUC are calculated at the fixed threshold of 0.5 on uncalibrated probabilities.	56
Figure 33- Success rate curves for the Austrian model (64 mapped springs) and the European model (91 mapped springs), logarithmic horizontal axis. The horizontal axis gives the share of a project area that is searched, ranked by predicted potential, and the vertical axis the share of mapped springs found within this share.	58
Figure 34 - Ablation experiment 2: share of project areas in which a predictor takes a single value, by predictor group and model. The number of variables per group is given in brackets. ..	60
Figure 35 - Ablation experiment 3: standardised difference between springs and non-springs of the same project area, after the differences between the areas have been removed by centring each predictor on the mean of its own area. Positive values mean that springs have higher values than the non-springs of their own area, negative values the opposite. In the European model the five soil variables are constant in every evaluable area and could not be evaluated. Variables marked with an asterisk could be evaluated in fewer areas than stated in the panel title, because they are constant elsewhere. Only project areas with at least two springs and two pseudo-absences were used.	61

List of Abbreviations

AT	Austria / Austrian model
AUC	Area Under the Curve
CI	Confidence Interval
CV	Coefficient of Variation
DEM	Digital Elevation Model
DSM	Digital Surface Model
EPSG	European Petroleum Survey Group (coordinate reference system codes)
ESDAC	European Soil Data Center
EDSB	European Soil Database
EU	Europe / European model
GEE	Google Earth Engine
GIS	Geographic Information System
GPS	Global Positioning System
LMT	Logistic Model Tree
LOAO	Leave-One-Area-Out
MC	Monte Carlo
MCC	Matthews Correlation Coefficient
NIR	Near-Infrared
NDVI	Normalized Difference Vegetation Index
NISAR	NASA-ISRO Synthetic Aperture Radar
RF	Random Forest
S1	Sentinel-1
S2	Sentinel-2
SAR	Synthetic Aperture Radar
SD	Standard Deviation
SPI	Stream Power Index

STI	Sediment Transport Index
TWI	Topographic Wetness Index
VH	Vertical transmit, horizontal receive (radar polarisation)

Abstract

Increasing frequencies of extreme weather events (Adler et al., 2017), such as droughts, require resilient water management strategies. Accurate knowledge of groundwater spring locations supports sustainable water resource management and previous studies have assessed spring potential with machine learning within single contiguous study areas (Rahmati et al., 2018; Chen et al., 2020; Sheikh et al., 2022). This study examines whether the approach can be transferred to small project areas that are widely distributed across Austria and Europe. Two Random Forest models were trained on field-verified springs: an Austrian model based on high-resolution national data and a European model based on pan-European data. Both models combined geological, soil and terrain variables with Sentinel-1 and Sentinel-2 data. The Austrian model performed better than the European model (MCC 0.47 versus 0.37), and the confidence intervals over 100 Monte Carlo realisations did not overlap. The placement of the pseudo-absence points was decisive. With a spatially uncontrolled design the order of the two models was reversed, which shows that background sampling can change the conclusions. A spatial validation, in which complete project areas were left out, showed that the ranking of the predictions is largely retained in unseen areas, while the absolute probabilities are not transferable without recalibration. Independently mapped waterlogged areas that were not used for training received higher predicted probabilities than the model area. The project areas are small and dispersed over several countries, which is a more heterogeneous setting than in the previous contiguous studies. Open pan-European data can support spring potential mapping within surveyed project areas, while high-resolution national data are more accurate. In practice the prediction maps can only be used with a recalibrated threshold.

1 Introduction

1.1 Intention

Over the past decades, rising global surface temperatures and increasing atmospheric water vapor content have been observed (Adler et al., 2017). While there is no significant trend in global mean precipitation, the frequency and intensity of extreme weather events, such as heavy rainfall and prolonged droughts, are increasing in many regions (Adler et al., 2017). These regional droughts can lead to increased fire hazard and frequency, and the size of wildfires also appears to be expanding (Moreira et al., 2011). It has been noted, that human interventions in the landscape, often driven by socio-economic interests, have resulted in simplified topographical structures and a decline in biodiversity (Moreira et al., 2011). Soil moisture is closely related to wildfire occurrence and behaviour due to its influence on vegetation and fuel moisture (Krueger et al., 2015). Maintaining adequate soil moisture can therefore play a significant role in preventing wildfires or at least limiting their size. Combined with increasingly frequent heavy rainfall events, agricultural and forestry landscapes are coming under growing pressure.

1.2 Background

Hydrological management practices like constructing artificial ponds, spring tapping, collecting surface water and dealing with accumulating infiltrating water can enhance the efficiency and sustainability of water availability in various ways.



Figure 1 Water retention planning in Southern France, part of the training areas. Plan: Oliver Krische, Holzer & Krische GmbH, (2022).

Artificial ponds constructed from local soil materials have the capacity to store excess water and maintain elevated soil moisture over extended periods due to the water-holding capacity of the surrounding soil. In water retention systems, multiple interconnected ponds (figure 1) can be distributed across the landscape and be additionally supplied by groundwater springs. A permanent groundwater supply is essential because ponds that rely solely on precipitation may dry out during prolonged droughts, whereas spring-fed ponds can maintain water availability and soil moisture even during dry periods. In addition, artificial ponds offer ecosystem services like increasing biodiversity in threatened areas, as demonstrated by experiments with dragonfly species (Santos et al., 2021). Their impact on local social-ecological systems, as well as their role in combined aquacultural food production, makes them a significant core element of traditional landscapes and regional identity, as exemplified by the golden humped tench of Poirino (Delpero and Volpato, 2022). In most project areas of this study groundwater springs

with a small discharge are mainly unused water resource. Without tapping groundwater springs with a small discharge, they just seep away again or are removed by drainage.



Figure 2 – Groundwater spring fed Pond within a pond system of a water retention facility at Krameterhof. Photograph: Josef Andreas Holzer, 2012.

Groundwater springs as well as seepage water springs play a significant role in providing water for such artificial ponds (figure 2) and drinking water for humans, but also for livestock and wild animals (Ranjan et al., 2021). Groundwater springs are typically characterized by a concentrated, steady outflow from deeper aquifers, often providing reliable and clean water year-round. Spring tapping is relatively cost-efficient and easy to maintain. In contrast, seepage springs emerge diffusely from shallow soil layers, usually have lower and more variable discharge, and are more susceptible to seasonal drying and contamination (Ranjan et al., 2021). With changing precipitation patterns, springs can become seasonal, and aquifers with low permeability take long to refill (Ranjan et al., 2021). Consequently, identifying potential groundwater spring locations is crucial for the planning of water retention measures and for reducing drought stress and wildfire risk.

1.3 Research Aims and Purpose

To assess hydrological conditions on agricultural and forestry land, monitoring the seasonal development of wet areas and identifying potential springs may provide valuable information. For this purpose, multispectral Sentinel-2 images as well as Sentinel-1 radar data can be used for annual observations. In particular, the seasonal variation of wet areas may provide significant hydrological insights. Primarily, such seasonal changes could offer clues about the potential discharge rates of specific springs.

Nowadays, the use of methods such as dowsing rods, despite being considered unscientific, remains common and can be quite costly. This study aims to compare two Random Forest models that provide an initial hydrological spring potential assessment of agricultural and forestry land based on satellite data, whereby one model relies on high-resolution national data and the other exclusively on European datasets, in order to assess the transferability of the approach at a continental scale. Although pan-European hydrogeological maps provide valuable information on hydrogeological units at continental scale (Duscher et al., 2015), they are not intended for groundwater spring potential mapping. Existing machine learning studies on groundwater spring potential have therefore focused on single and contiguous study areas (Rahmati et al., 2018; Arabameri et al., 2019; Chen et al., 2020; Sheikh et al., 2022). In contrast, this study develops and evaluates a model using numerous geographically separated small project areas distributed across Austria and Europe. Thereby the robustness is tested by a broader range of environmental conditions.

This leads me to the following questions:

1. How well can spring potential be modelled using European datasets compared with high-resolution Austrian datasets?
2. Which predictor variables contribute most to model performance?
3. To what extent does the quality and resolution of the geological data limit model performance?
4. Do the models assign increased spring potential to independently mapped waterlogged areas that were not used for model training?

1.4 Previous research

It has been shown that Sentinel-2 images in false colour composition offer high sensitivity for the identification of surface water (Li et al., 2022). Overall, this spectral combination allows for a clear differentiation between green vegetation, surface water, rocks, and shadow. However, it should be noted that the timing of image acquisition is crucial, as the state of vegetation can significantly affect spectral reflectance and, consequently, the detectability of wet areas.

Recent research has demonstrated that the combination of Sentinel-1 SAR and Sentinel-2 optical imagery provides significant advantages for multitemporal monitoring of surface water dynamics. Sentinel-1 enables observations independent of cloud cover, while Sentinel-2 supplies complementary information on vegetation and surface water. In combination with digital elevation models, these datasets improve the detection of surface water and reduce misclassifications caused by vegetation and terrain effects (Chen and Zhao, 2022).

Early GIS-based spring potential mapping studies focused on statistical and knowledge-driven approaches. Corsini et al. (2009) applied Weight of Evidence (WofE) and Radial Basis Function Neural Networks (RBFNN) to predict spring occurrence in the Northern Apennines, Italy. The study identified lithology, tectonic structures, and stratigraphic relationships as key controlling factors for spring emergence. In particular, the distance to faults and thrusts showed a strong influence on spring distribution (Corsini et al., 2009).

Groundwater potential mapping has been addressed using a variety of statistical, multicriteria decision analysis and machine learning approaches (Rahmati et al., 2018). In their study conducted in Iran, Rahmati et al. (2018) identified relative slope position, distance to faults, and lithology as the most influential predictor variables. The authors compared three models including Logistic Model Trees (LMT), Support Vector Machines and C4.5 decision trees. The LMT achieves a mean AUC-ROC value of 0.904, which indicates excellent predictive performance (Rahmati et al., 2018).

The potential of using Sentinel-2 imagery in combination with AI-based methods such as Random Forest for groundwater spring modelling has been demonstrated in a case study from China (Chen et al., 2020). The results obtained were highly convincing, achieving an accuracy (AUC) of 0.909.

Likewise, Sheikh et al. (2022) reported an AUC value of 0.94 for a Logistic Regression (LR) model applied in the Sindh Basin of the Kashmir Himalaya. Similar to Chen et al. (2020), the study

incorporated hydrogeological, topographic and lithological predictor variables together with NDVI. Distance to streams was found to be the most important predictor variable, followed by lithology (Sheikh et al., 2022).

1.5 Hydrological Background

Groundwater represents the subsurface component of the hydrological cycle and originates from the infiltration of precipitation into the soil and underlying geological formations. After percolating through permeable layers, groundwater may re-emerge at the land surface in the form of springs, which represent natural discharge points of groundwater flow systems (Hölting and Coldewey, 2013).

Springs are localized points of groundwater discharge where subsurface water emerges naturally at the land surface and subsequently contributes to surface runoff within the respective catchment area. The discharge of a spring is commonly expressed as spring yield and is usually measured in litres per second (l/s). Springs with continuous discharge throughout the year are referred to as perennial springs, whereas springs that flow only periodically are classified as intermittent springs (Hölting and Coldewey, 2013).

The occurrence of springs is closely related to the permeability of the underlying geological formations. However, wet areas are not necessarily caused by groundwater discharge and may also result from temporary surface water accumulation. During field surveys, measurements of electrical conductivity, pH, and water temperature are routinely compared with local surface runoff characteristics to help distinguish groundwater-fed springs from surface water accumulation.

The occurrence of springs is controlled by a complex interaction of geological, hydrological and topographical factors. As natural discharge points of groundwater systems, springs provide valuable information about subsurface flow processes and local groundwater conditions.

2 Methods

2.1 Study Area

The study areas were selected based on the availability of previously digitized spring locations obtained during on-site field surveys. These data were collected between 2021 and 2025 by an experienced geologist with expertise in hydrogeology and spring identification. Spring verification included hydrogeological assessment as well as measurements of electrical conductivity, pH, and water temperature to distinguish groundwater-fed springs from temporary surface-water accumulation. The original dataset distinguishes groundwater-related wet areas from temporary seepage zones caused by episodic surface-water accumulation. It also contains mapped fountains and drainage structures, which were excluded from this study. The study area comprises multiple project areas with and without verified spring locations. Only verified spring locations and pseudo-absence points were included in the analysis. Some uncertainty remains because, for several spring locations, only the spring chambers had been digitized, while the original groundwater emergence points were unknown. The study areas and spring points have been evaluated regarding their usage for the machine learning model. The samples are distributed across Europe (figure 22), whereas most of the projects are in Austria (figure 21). Fewer are in the Austrian near environment and only a small amount is in the western part of Europe. Unfortunately, the Swiss samples could not be integrated due to missing harmonized data. There are nearly no recorded samples in the eastern part of Europe. The study areas cover a broad range of climatic and environmental conditions. Nevertheless, all sites are located within the temperate climate zone.

2.1.1 Spring Points and Pseudo-absence Points

The identification of springs in the field is not always straightforward. In many cases, groundwater does not emerge from a clearly defined point source but rather appears as diffuse wet areas. Consequently, the visible wetness pattern does not necessarily correspond to the true groundwater emergence point (Hölting and Coldewey, 2013). In addition, the recorded spring locations are subject to a certain positional uncertainty due to GPS measurements under field conditions.

To avoid assigning negative samples in the immediate vicinity of verified spring locations, non-spring points (pseudo-absences) were sampled from the model prediction grid, which guarantees complete predictor information for every point. Several sampling designs were implemented and systematically compared (table 5). The initial, spatially uncontrolled design adopted from earlier project work, a design with a fixed number of pseudo-absences per project

area and an area-weighted design with a minimum number of points per area. In addition, balanced and imbalanced class ratios were tested with and without compensation for sparse categories. Based on this comparison, fixed-per-area design (EU: 6, AT: 10 points per area) with class compensation was adopted for the final models. In all designs, a minimum distance of 200m to any verified spring location was implemented. Pseudo-absences were restricted to the project areas themselves, whereas verified springs located in the immediate surroundings of a project area were retained and assigned to the nearest area (maximum distance 3 km). This restriction ensures that pseudo-absences are only placed where the field surveys allow the absence of unrecorded springs to be assumed with reasonable confidence.

The Random Forest model is built upon a combination of geological, hydrological, surface soil and elevation variables, following the feature set established in previous spring detection research (Chen et al., 2020). Sentinel-1 and Sentinel-2 satellite data were additionally incorporated to capture surface moisture and vegetation conditions of spring occurrence.

2.2 Geological Classification

For the Austrian model (figure 3), lithology data from the GEOFAST dataset provided by GeoSphere Austria at a scale of 1:50 000 was used. Although the GEOFAST dataset covers harmonized geological data for the whole of Austria, not all project areas are available as vector data. This resulted in the exclusion of some project areas containing verified spring points. Given the broad spatial distribution of sampling points across Austria and the resulting diversity of lithological units, the features were reclassified into six permeability classes.

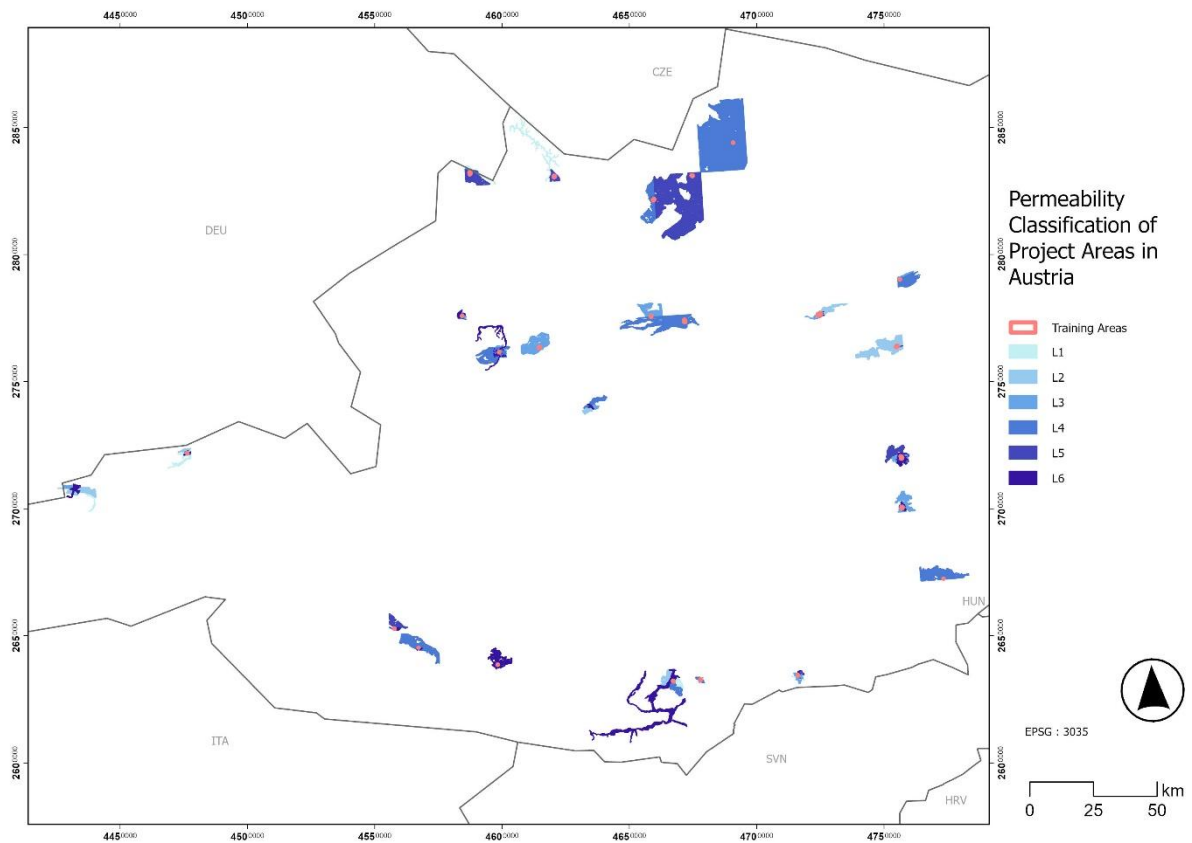


Figure 3 - Lithological units of the Austrian project areas reclassified into six permeability classes (L1–L6).

For the EU model (figure 4), lithology is based on the European Geological Data Infrastructure (EGDI) Pan-European Surface Geology dataset at a scale of 1:1 000 000. The data is provided as an INSPIRE-conformant WFS layer (GeologicUnitView) and harmonized across national geological surveys for European-scale applications.

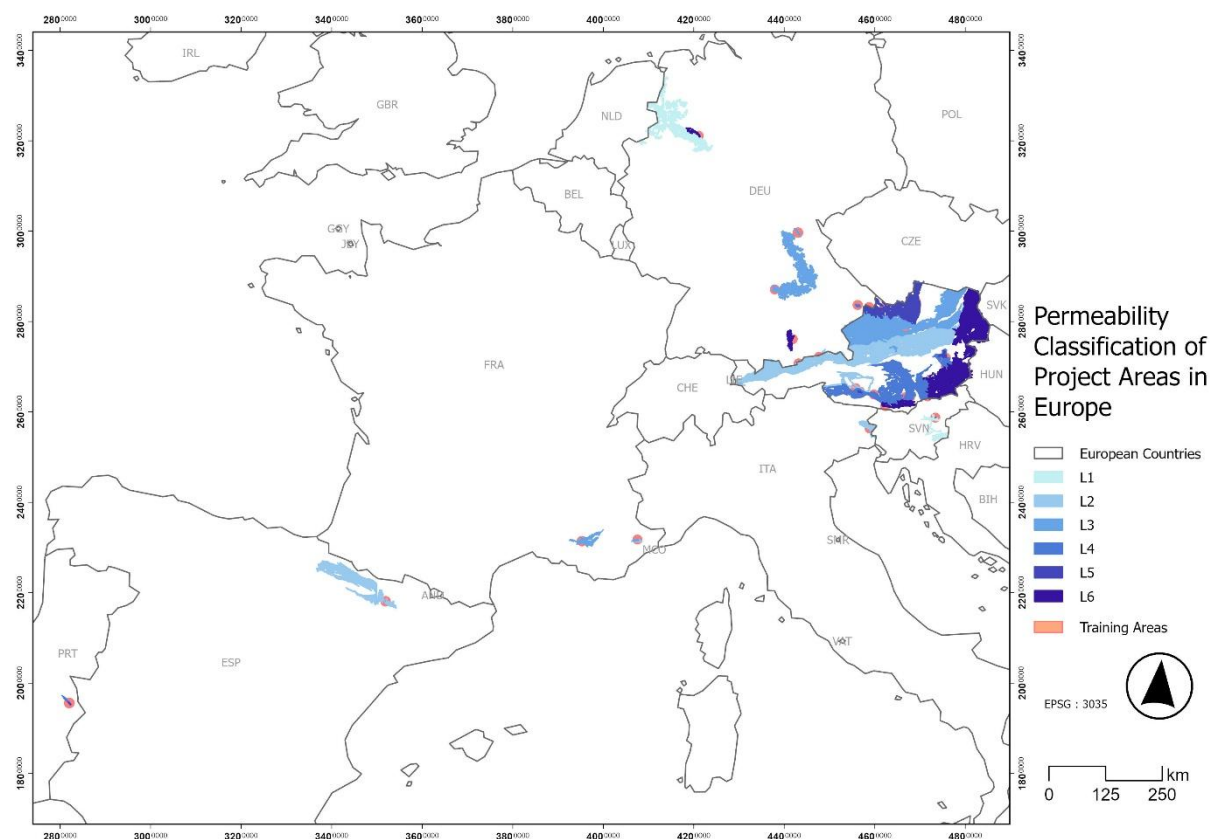


Figure 4 - Lithological units of the European project areas reclassified into six permeability classes (L1–L6).

The lithological units (Table 1) were manually assigned to permeability classes based on their hydrogeological characteristics, including rock type, grain size and degree of fracturing or karstification, following hydrogeological theory (Hölting and Coldewey, 2013).

Table 1 - Classification of lithological units into permeability classes.

Class	Permeability	Typical lithologies	Hydrological relevance
L1	Very high	Karstified limestone, dolomite, coarse gravel, alluvium, strongly fractured rock	Very high spring potential, fast infiltration, typical karst and fissure springs
L2	High	Sandstone, conglomerate, fractured volcanic rocks (basalt, tuff)	High spring potential, good groundwater storage and flow
L3	Moderate	Marl, loamy sediments, weathered metamorphic rocks	Moderate spring potential, locally favourable spring conditions
L4	Low	Claystone, shale, siltstone, weakly fractured crystalline rocks	Low spring potential, limited infiltration, surface runoff generally dominates

L5	Very low	Massive granite, gneiss, compact metamorphic basement	Very low spring potential, groundwater flow mainly limited to fractured zones
L6	Uncertain / heterogeneous	Glacial till, moraines, mixed or unknown lithology, anthropogenic deposits	Uncertain behaviour, context-dependent hydrology

2.3 Sentinel images from Google Earth Engine (GEE)

All Sentinel-1 and Sentinel-2 image processing, including image collection filtering, image selection and data export was managed with Google Earth Engine (GEE). GEE provides cloud-based access to the full Sentinel archive, enabling efficient processing of large image collections without local computational resources. All images were exported in EPSG:3035 at 10 m spatial resolution as GeoTIFF files for further analysis in ArcGIS Pro.

2.3.1 Sentinel - 2 images

The aim was to select the most appropriate image (figure 5) in which vegetation cover is low, typically at the beginning or end of the growing season. Waterlogged areas are extraordinary well detectable and therefore field surveys are typically conducted at this time of the year. While a visual inspection of the time series can make it obvious, which image is most suitable, an automated approach requires a more precise description of the relevant spectral characteristics. All Sentinel-2 images were reviewed in false colour composition, using bands 8, 4 and 3. By analysing the histograms of the red, green, and blue channels, it was found that the most suitable image is characterized by a comparatively low peak in the red histogram, combined with a high variance. To automatically select the optimal image, two conditions were defined: (1) the peak must be < 1100 and (2) the variance must exceed $> 20\,000$. If no image fulfilled both selection criteria, the image with the lowest red-band modal value was selected as a fallback. Only Sentinel-2 images with less than 20% cloud cover were considered for further analysis. The image should serve as the starting point for building the dataset used in the random forest workflow. This kind of selection only works within the selected project polygons, since the landscape is mostly composed of forestry and agricultural land and a low proportion

of buildings and other artificial surfaces. Attempts to use this selection model for the whole catchment area did fail due to different landscape composition.

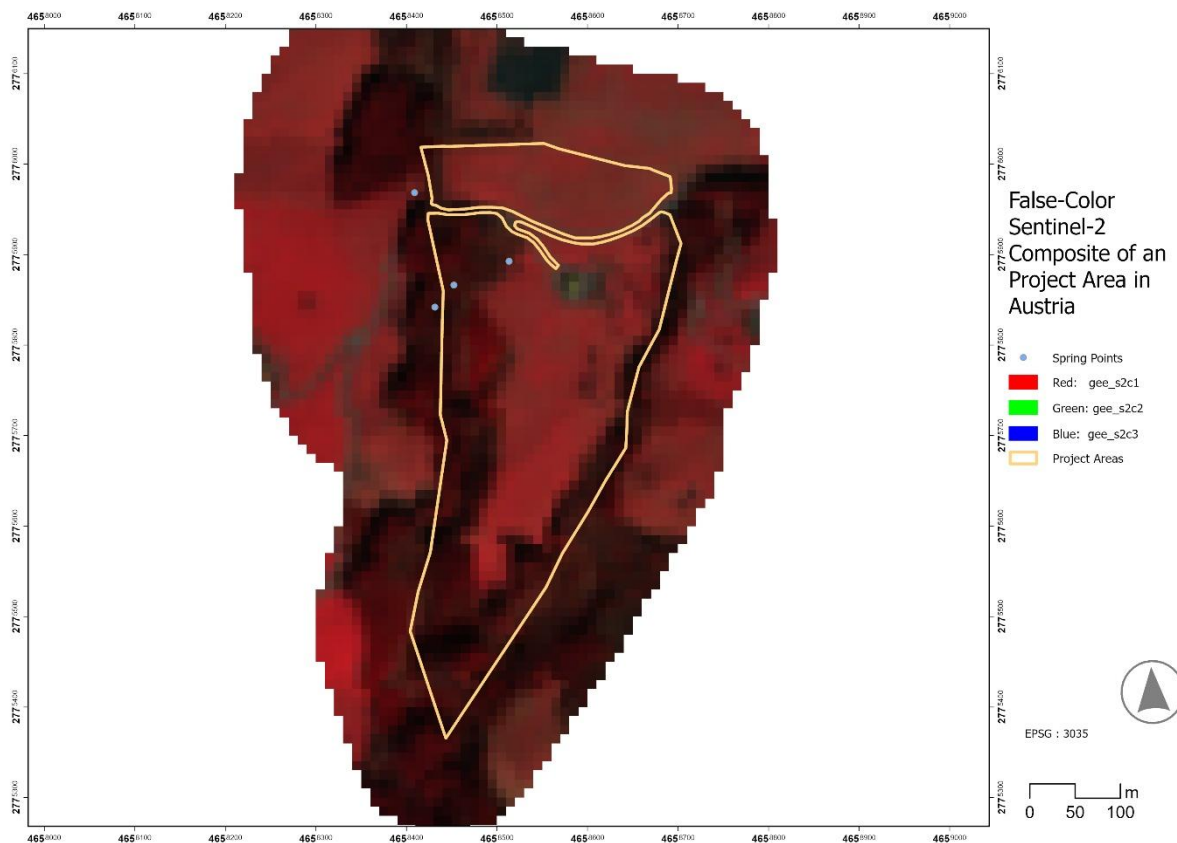


Figure 5 - Selected Sentinel-2 false-color composite (B8–B4–B3) of one Austrian project area used for wetness assessment.

2.3.2 Sentinel -1 images

Sentinel-1 SAR images (VH polarization, descending orbit) were acquired for the period February to end of May 2025 (figure 6). To characterize seasonal changes in surface moisture, median VH backscatter composites were calculated for two time periods: February–March and April–May. In addition, a temporal change layer (VH_change) was derived by subtracting the wet-season median from the later-season median. These three raster layers (VH_wet (Feb-March), VH_dry (April-May) and VH_change (VH_dry – VH_wet)) were extracted as predictor variables for the random forest model.

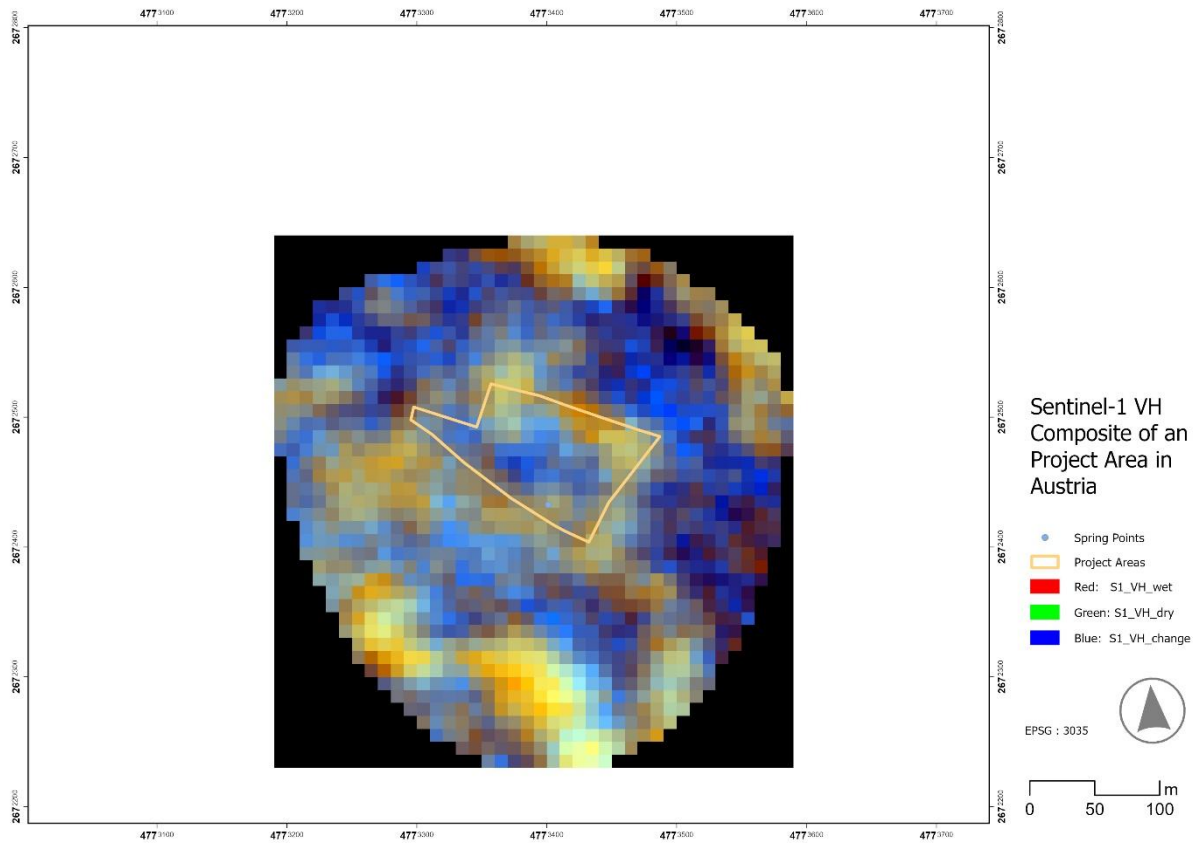


Figure 6 - Sentinel-1 VH composite derived from median backscatter during February–March (red), April–May (green), and their difference (blue). The composite was used to visualize seasonal changes in radar backscatter. The displayed colours do not represent natural colours, but a false-colour composite used to visualize seasonal changes in radar backscatter.

2.3.3 Explorative wetness mask

In addition to the predictor variables used for the Random Forest model, an exploratory mask was generated by combining the selected Sentinel-2 image with areas showing a decrease in Sentinel-1 VH backscatter. Pixels fulfilling both criteria were classified as potential wet areas. This step was performed solely for visual exploration and was not included in the random forest model. Nevertheless, visual comparison indicated that the explorative wetness mask covered the buffer zones of several verified spring locations.

2.4 Terrain Parameters

For the Austrian model, a Digital Elevation Model (DEM) with 5 m resolution was used both as a predictor variable and as the basis for calculating the hydrological indices. The EU model relies on the Copernicus Digital Surface Model (DSM) at 30 m resolution. To ensure, that the complete catchment area was included in the hydrological analysis, a 2.5km buffer was added around each project area before deriving the terrain parameters and hydrological indices. As is common practice for the calculation of hydrological indices, both elevation models were pre-

processed using the *Fill* tool in ArcGIS (figure 6 and 7) to remove surface artificial sinks and ensure continuous flow paths for the calculation.

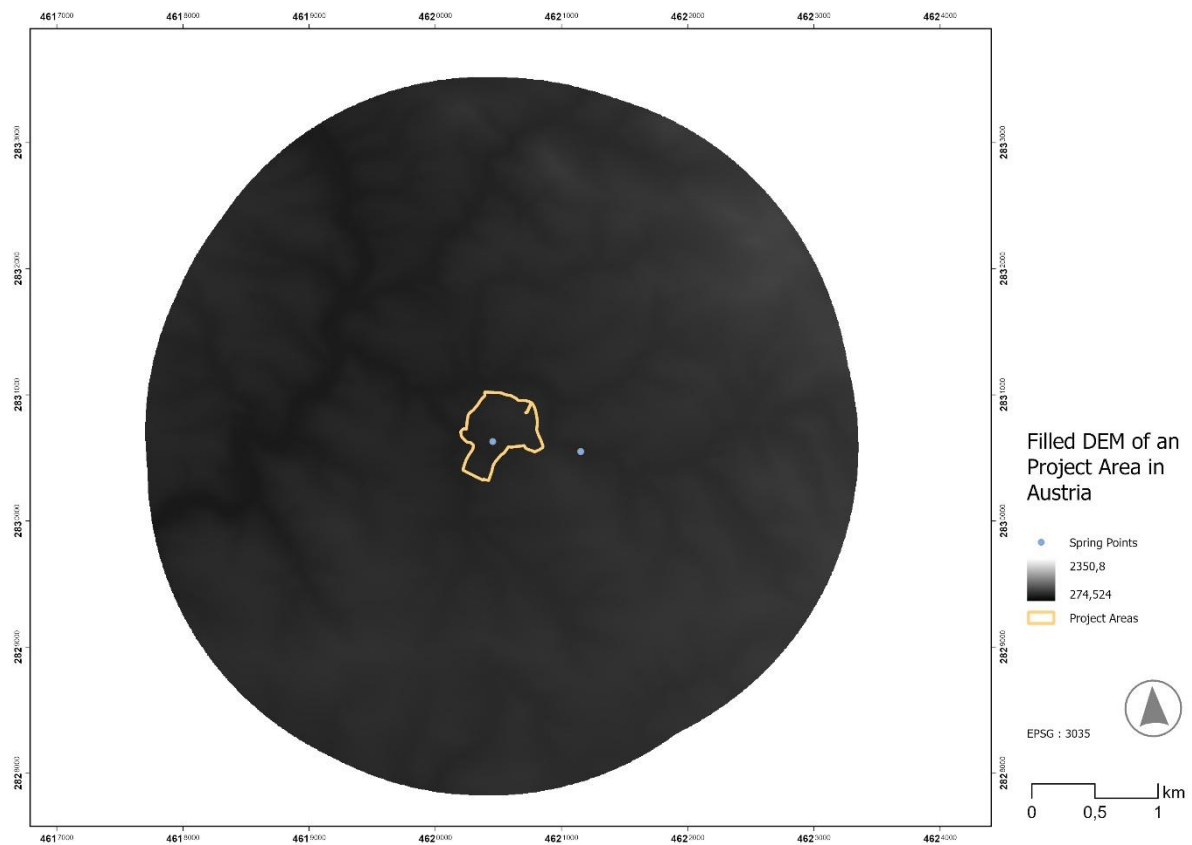


Figure 7 - High-resolution (5 m) digital elevation model used for the Austrian model.

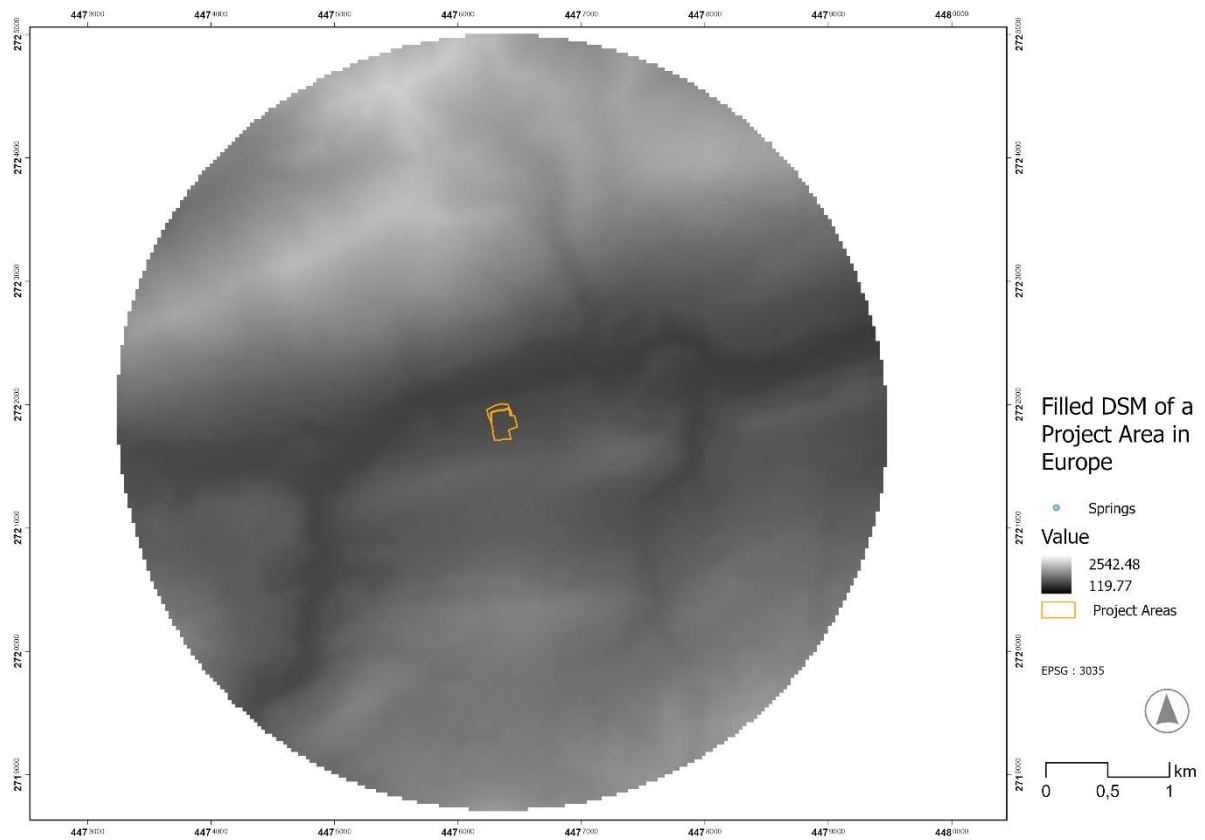


Figure 8 - Low-resolution (30m) Copernicus Digital Surface Model used for the European model.

From each respective elevation model, slope (figure 9 and 10), aspect (figure 11 and 12) and general (figure 13 and 14), plan and profile curvature were derived.

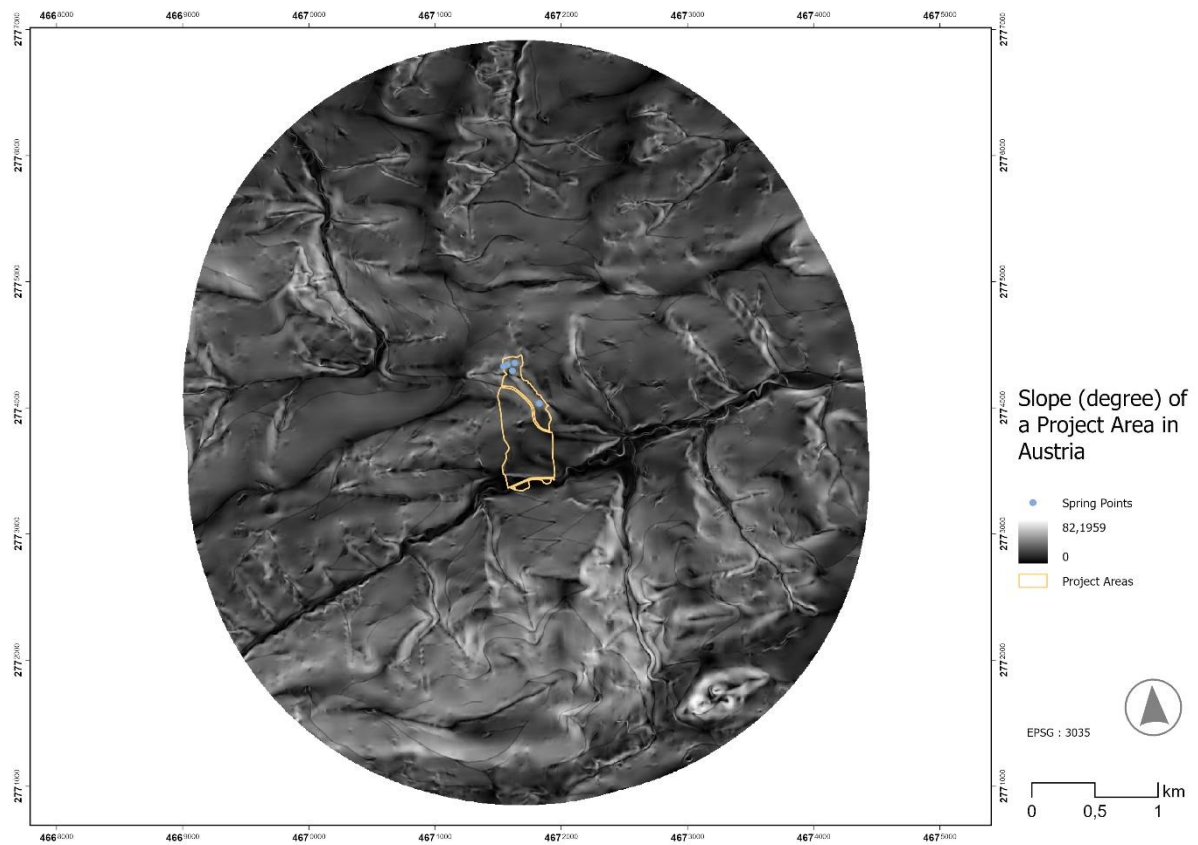


Figure 9 - Slope (degree) derived from the Austrian 5 m DEM.

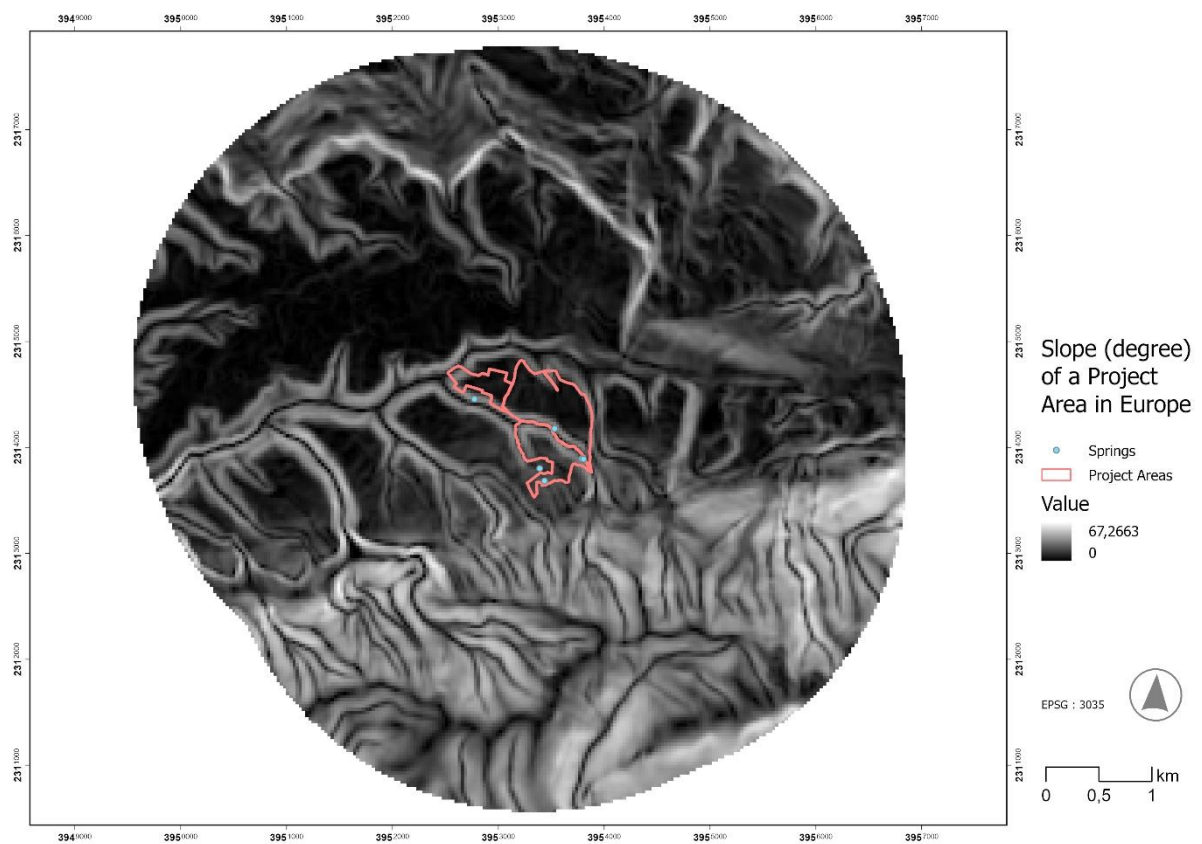


Figure 10 - Slope (degrees) derived from Copernicus 30m DSM.

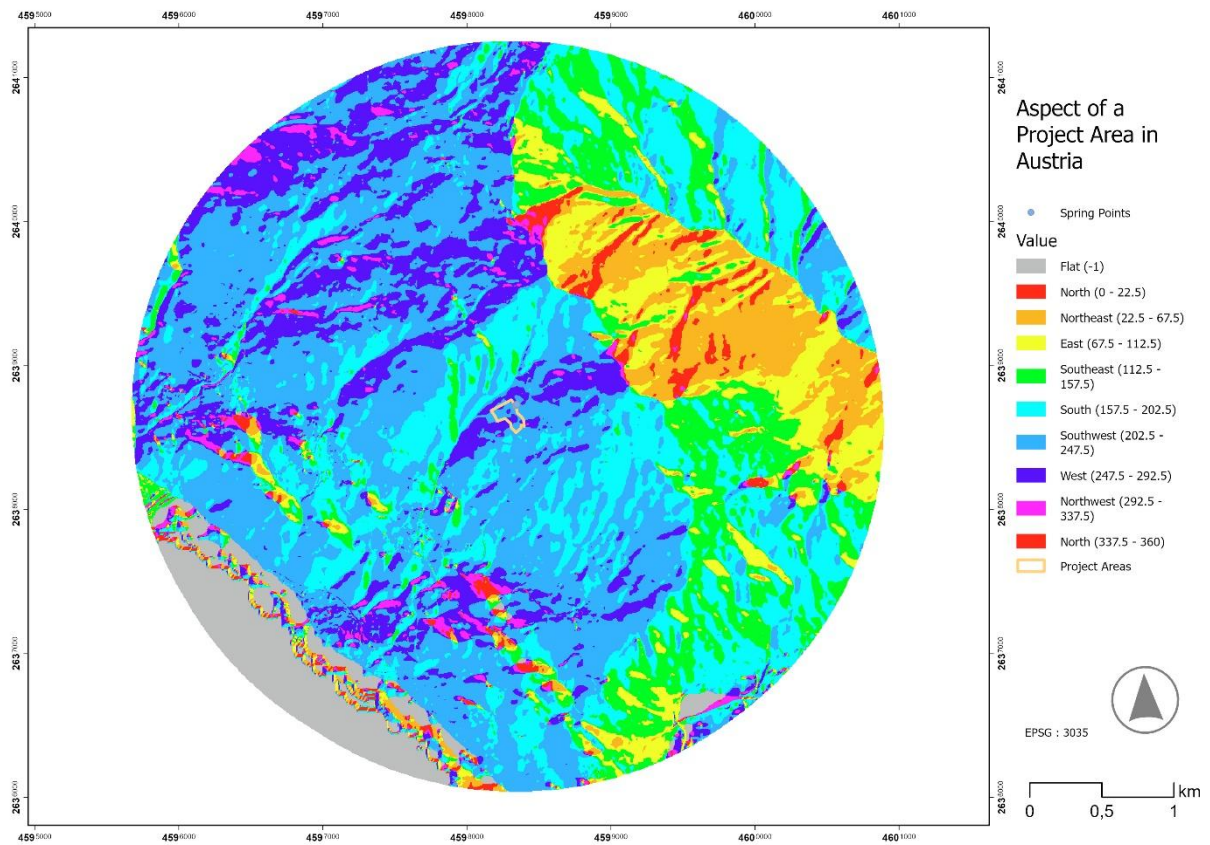


Figure 11 - Aspect derived from the Austrian 5 m DEM.

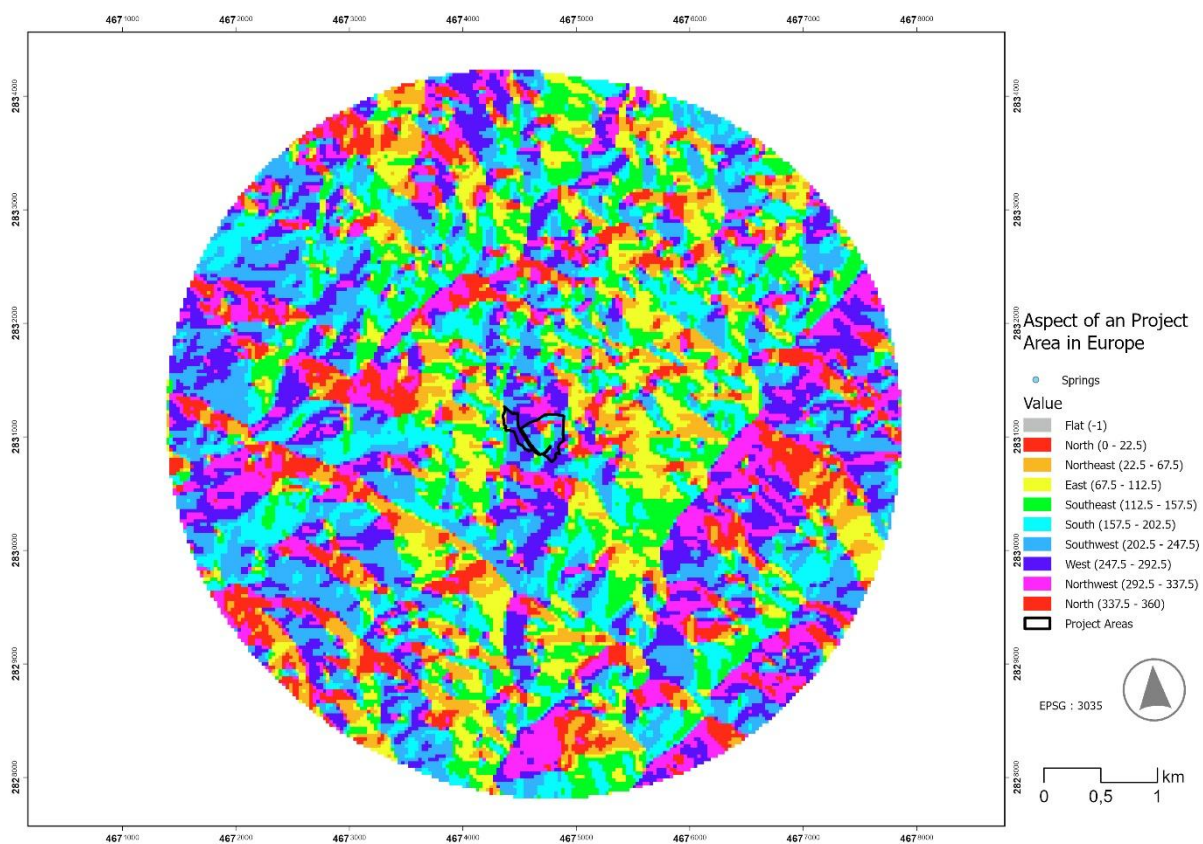


Figure 12 - Aspect derived from Copernicus 30m DSM.

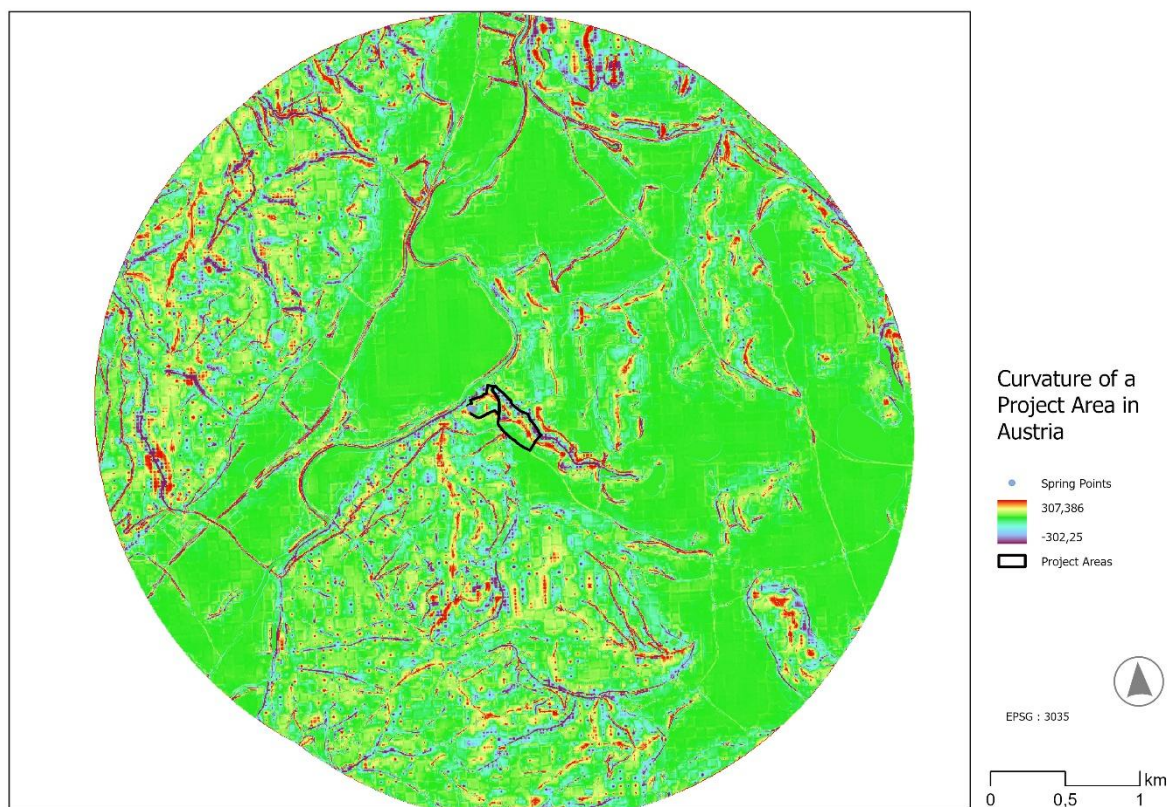


Figure 13 - Curvature derived from the Austrian 5 m DEM.

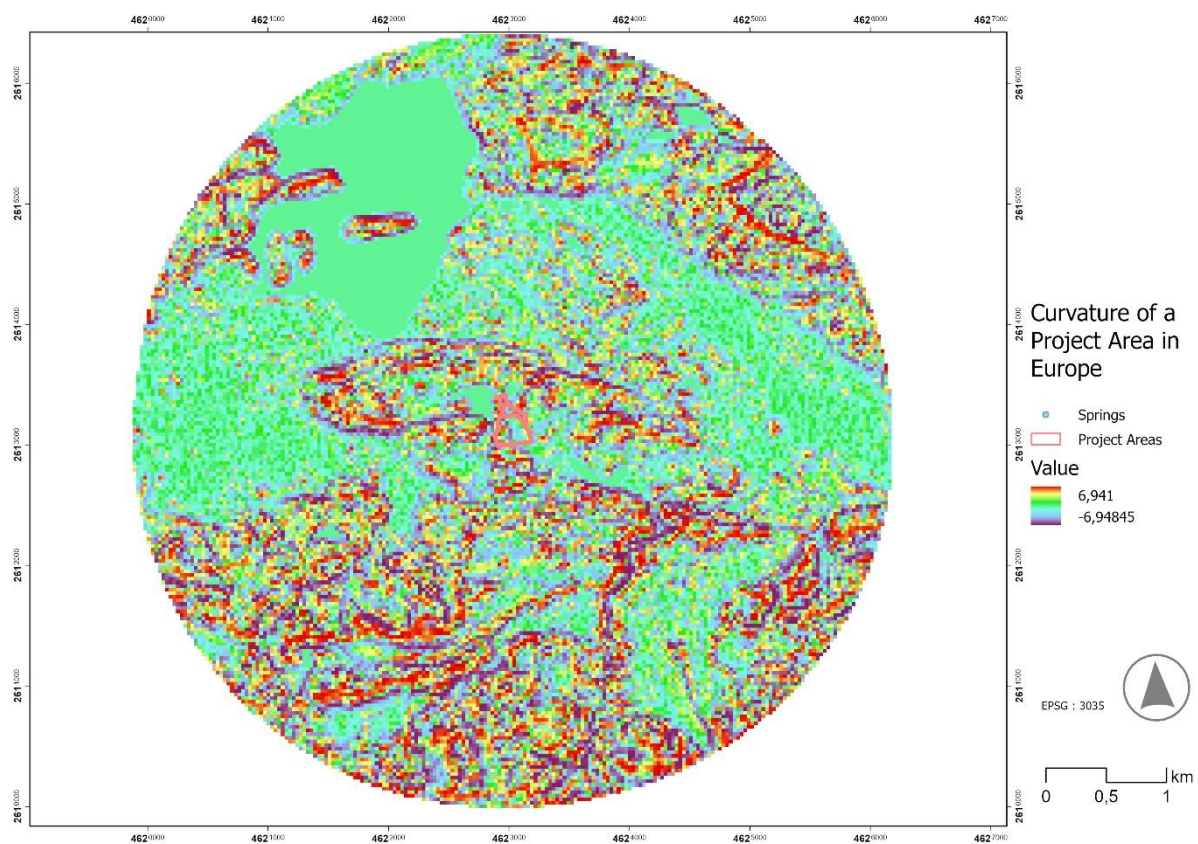


Figure 14 - Curvature derived from Copernicus 30m DSM.

2.5 Hydrological Indices

Based on the same elevation models, four hydrological indices were calculated: the Topographic Wetness Index (TWI) (figure 17 and 18), the Stream Power Index (SPI) (figure 19 and 20), the Sediment Transport Index (STI), and Flow Accumulation (figure 15 and 16).

All hydrological indices were calculated using the respective cell size of each elevation model, which is 5 m for the Austrian model and 30 m for the EU model, to ensure consistency between the spatial resolution of the input data and the derived indices.

The following indices were derived from each elevation model after sink filling and D8 flow direction calculation:

TWI:

$$TWI = \ln \left(\frac{(flow_{acc} + 1) \times cellsize}{\tan(\beta + 0.001)} \right)$$

SPI:

$$SPI = (flow_{acc} + 1) \times cellsize \times \tan(\beta + 0.001)$$

STI:

$$STI = \left(\frac{(flow_{acc} + 1) \times cellsize}{22.13} \right)^{0.6} \times \left(\frac{\sin(\beta + 0.001)}{0.0896} \right)^{1.3}$$

where *flow_{acc}* represents the upslope contributing area derived from the flow accumulation raster, *cellsize* is raster the cell size (m), and β is the slope angle (radians). A small constant (0.001 and +1) was added to avoid undefined mathematical operations, such as division by zero or the logarithm of zero. The implemented formulations are mathematically equivalent to the standard equations reported in the literature (Arabameri et al., 2019). Differences arise from the software implementation, which expresses the specific catchment area as flow accumulation

multiplied by the raster cell size.

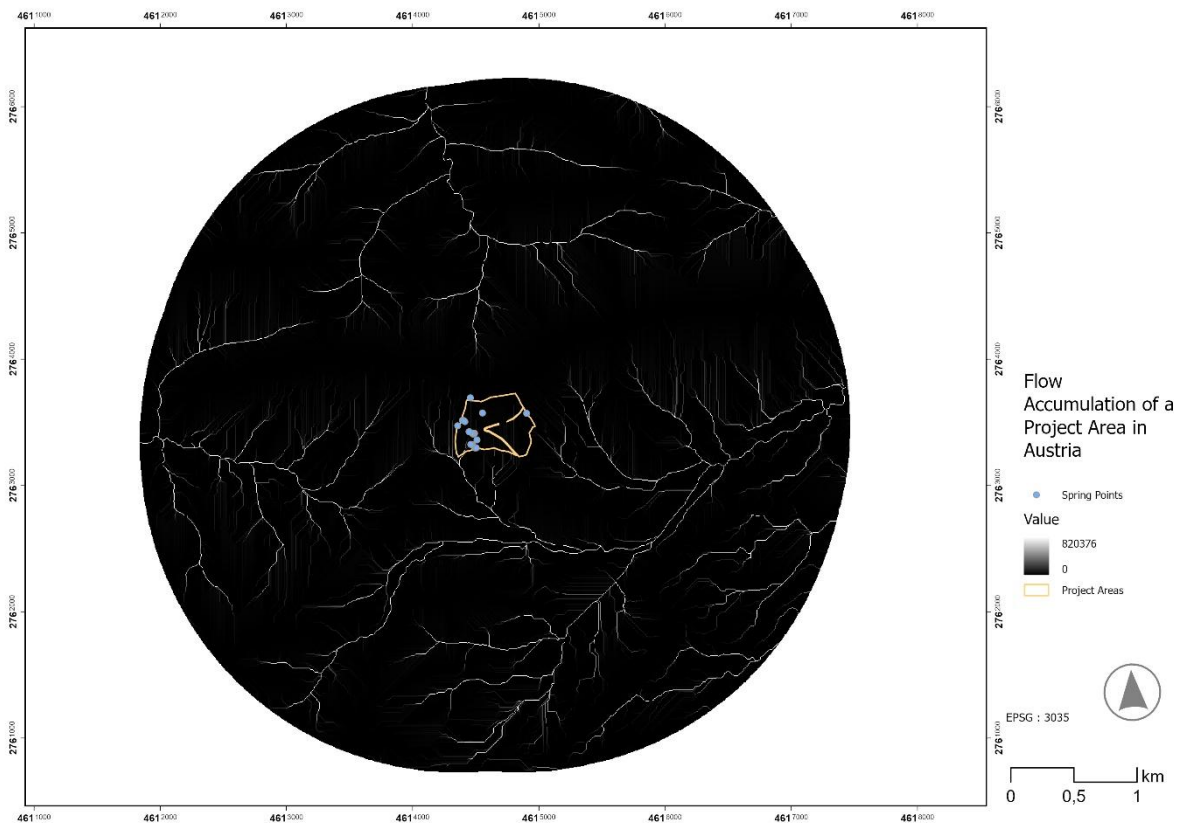


Figure 15 - Flow accumulation derived from the high-resolution Austrian DEM (5 m).

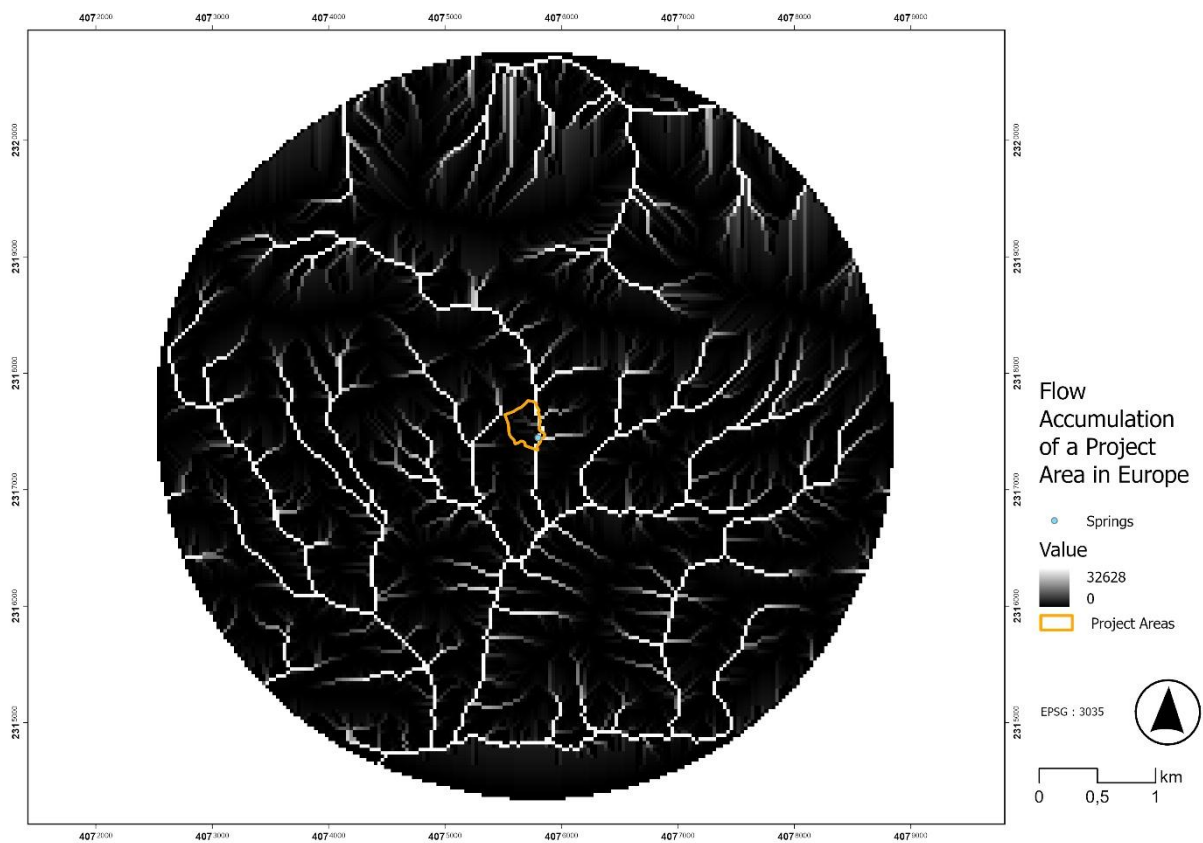


Figure 16 - Flow accumulation derived from the Copernicus DSM (30 m) used for the European model.

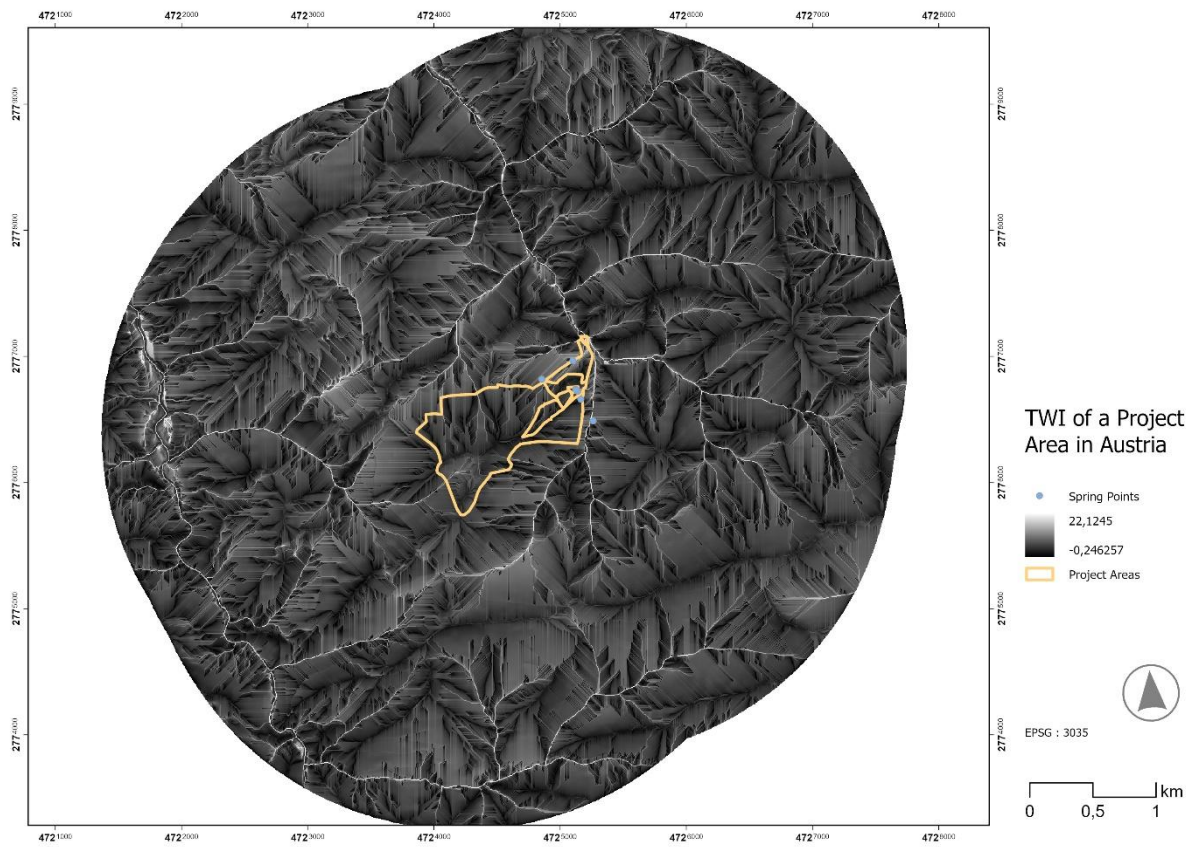


Figure 17 - Topographic Wetness Index (TWI) derived from the Austrian 5 m DEM.

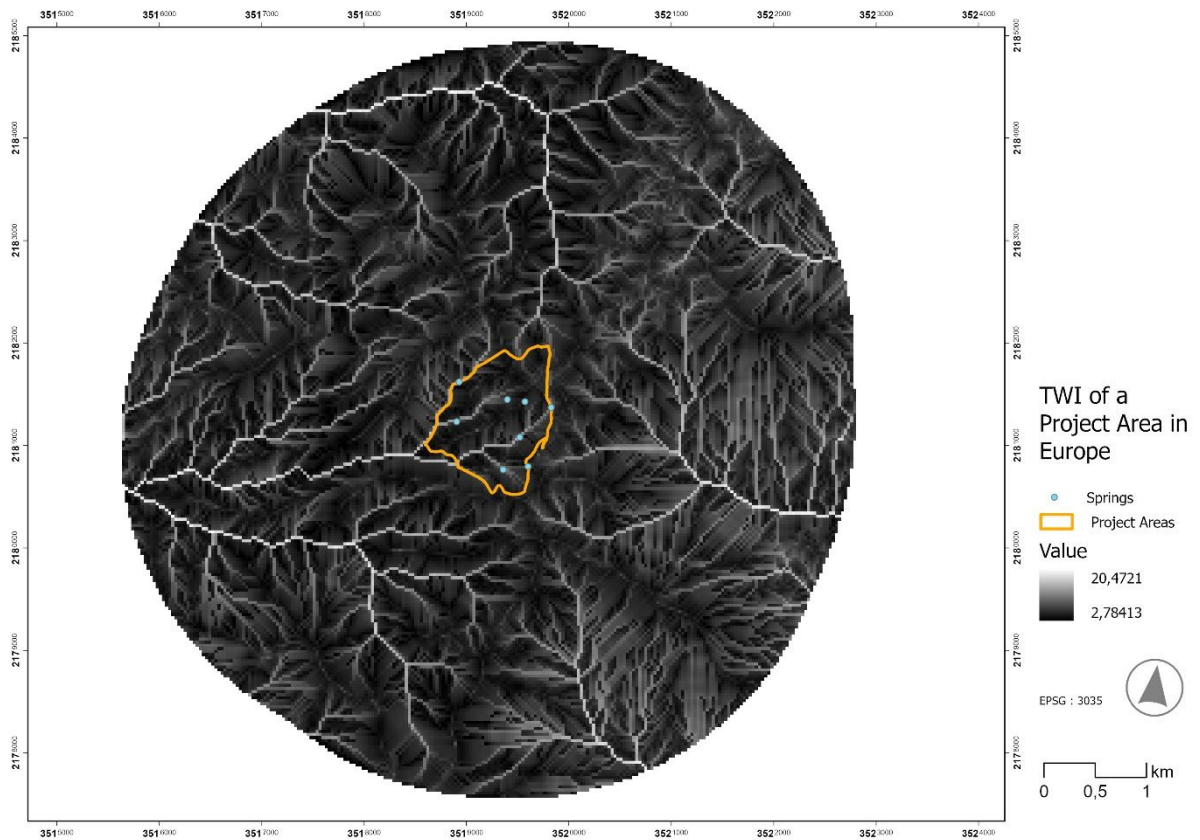


Figure 18 - Topographic Wetness Index (TWI) derived from the Copernicus 30m DSM.

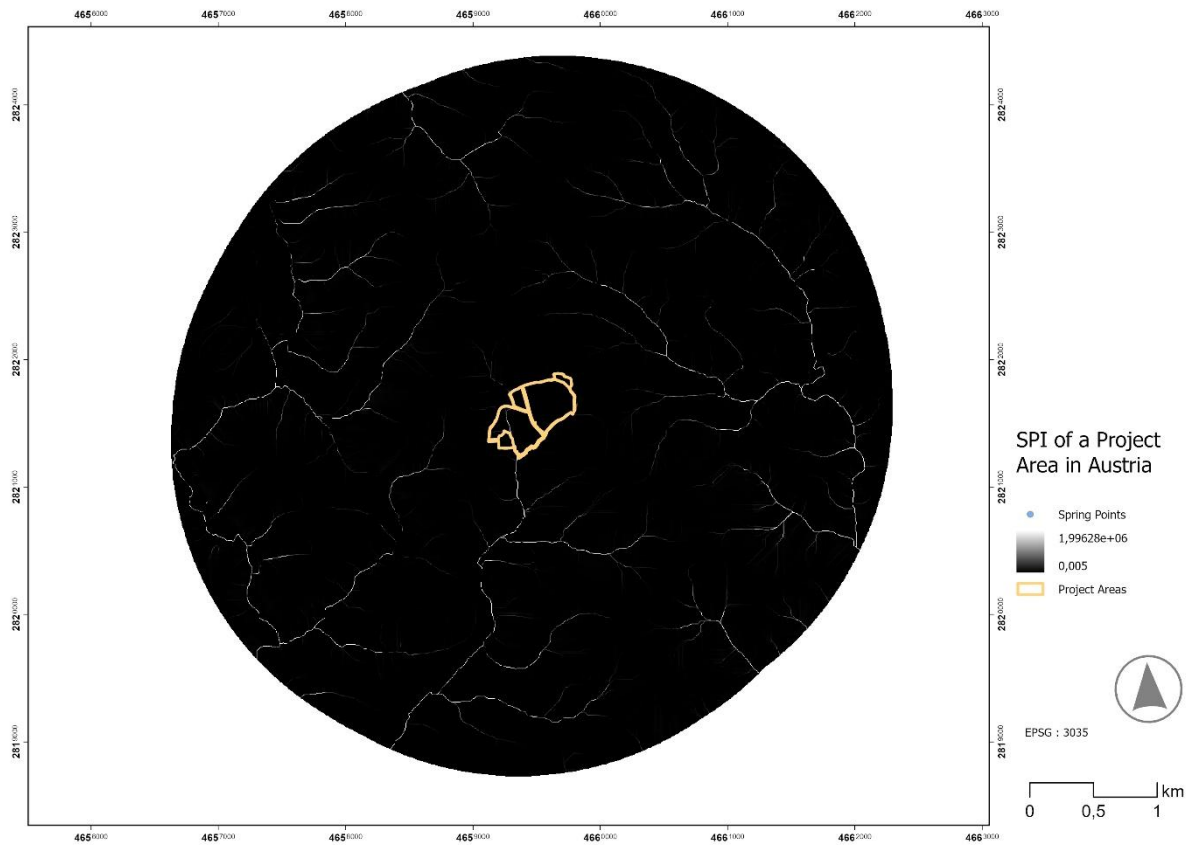


Figure 19 - Stream Power Index (SPI) derived from the Austrian 5m (DEM).

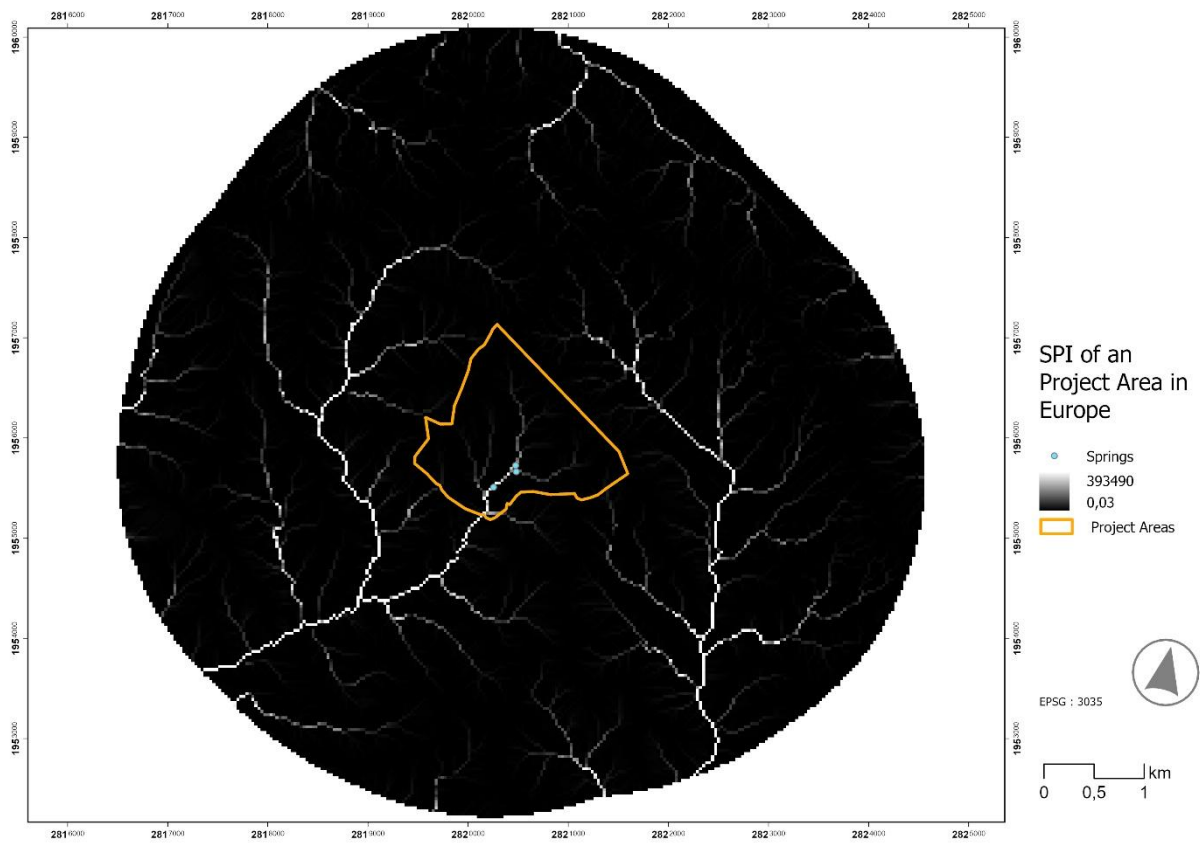


Figure 20 - Stream Power Index (SPI) derived from the Copernicus 30m DSM.

2.6 Soil Data

For the Austrian model, soil data was obtained from the Austrian Digital Soil Map (Österreichische Bodenkarte, bodenkarte.at) at 1 km resolution, which is freely available as a vector dataset. The polygons were converted to raster and the values at the sample locations were extracted with *Extract Multi Values to Points*. The following variables were extracted: sand, silt, and clay content (%), carbonate content (Kalkgehalt) and permeability (Durchlässigkeit, 10 classes).

For the EU model, soil data was obtained from the European Soil Database v2 (ESDB v2), distributed as a 1 km × 1 km raster library through the European Soil Data Centre (ESDAC), hosted by the European Commission's Joint Research Centre (Panagos et al., 2022). Although the dataset is publicly documented, access to the raster library is provided upon request through ESDAC. The following variables were extracted: parent material hydrogeological type (PMH), hydrogeological class (HG), depth to rock (DR), impermeable layer depth (IL), and soil texture class (TEXT).

The soil variables of the two models are not directly comparable. The Austrian dataset consists mainly of numerical variables, for example the sand, silt and clay content in percent. Some of the European variables are nominal classes instead. In the raster version of the ESDB these classes are stored as numbers, but the numbering follows the alphabetical order of the class codes and not the order of the property itself. For some variables, such as depth to rock, impermeable layer depth and texture class, the numbers still follow a meaningful order from low to high. For the parent material hydrogeological type and the hydrogeological class this is not the case. Two classes can share the same storage and permeability behaviour while receiving different numbers and the legend does not explain the coding. It was therefore not possible to recode all European variables into an ordinal scale or to find a direct equivalent for each Austrian variable. The European soil variables were used as numerical values in the model. This must be kept in mind when the results of the two models are compared.

2.7 Training Data Preparation

In general, the project areas are of small size. In the Austrian model the area ranges from 0.007 km² to 0.79 km² with a median of 0.10 km². Moreover, all 22 areas show a smaller size than 1 km², and 16 of them are smaller than 0.25 km². In Europe the areas range from 0.007 km² to 2.17 km² with a median of 0.10 km², and 31 of the 34 areas are smaller than 1 km². The total mapped area is 3.77 km² for Austria and 10.40 km² for Europe.

The number of springs per area is small as well. The median is two springs per area and the highest number in a single area is 17. In Austria 8 of the 22 project areas contain no spring at all, and in Europe 12 of the 34 areas. In these areas only pseudo-absence points were added to the training data. One Austrian area and five European areas contain a single spring only.

Verified spring locations and pseudo-absence points (see section 2.1.1) were merged into one training dataset per study region, comprising 64 springs and 220 pseudo-absences for Austria and 91 springs and 204 pseudo-absences for Europe. All predictor variable values were extracted to the training points using the *Extract Multi Values to Points* tool in ArcGIS Pro.

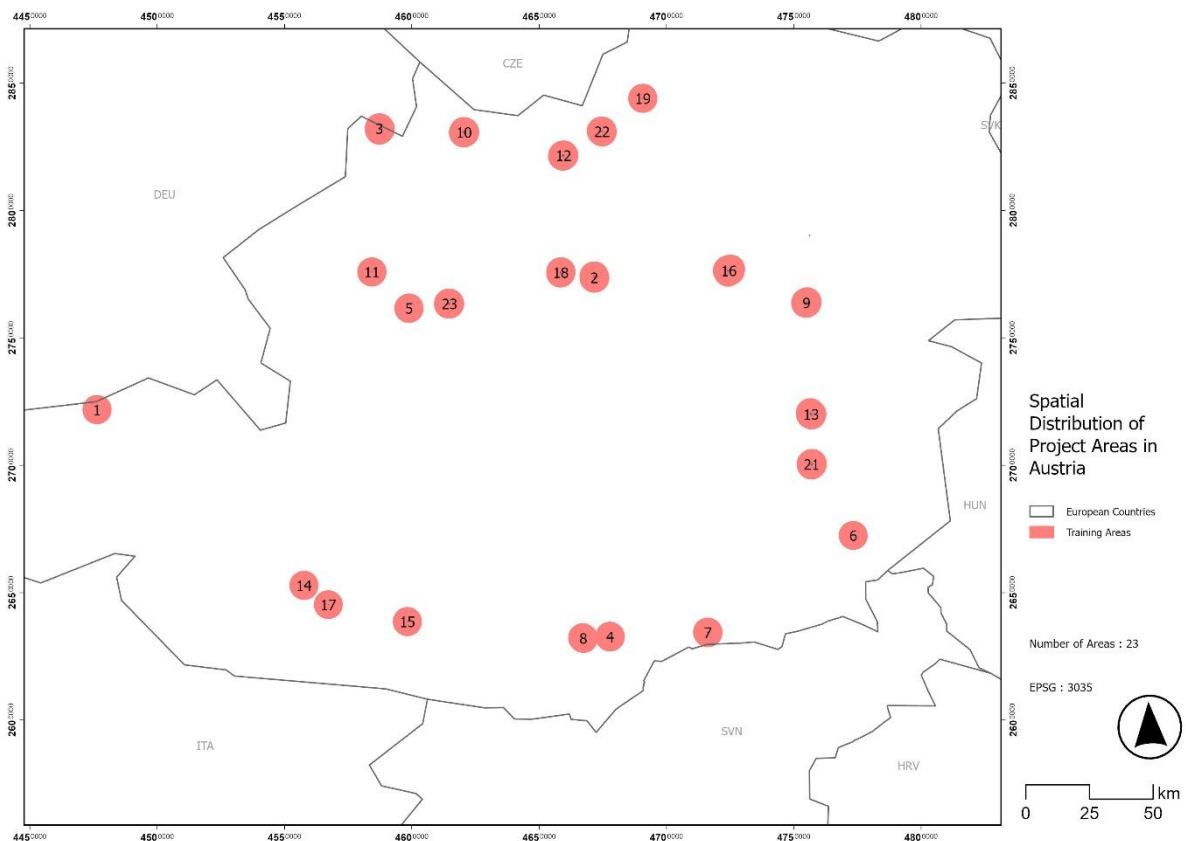


Figure 21 - Spatial Distribution of Austrian project areas and their project IDs.

Table 2 - Overview of the Austrian project areas, including area size [m²] and the number of mapped springs per area.

Project ID	Area [m2]	Spring Points
2	46883	2
4	314167	5
5	233455	0
6	88732	6
7	40153	0
9	11062	1
10	84096	0
11	84736	4
12	303221	3
14	272433	2
15	32945	0
17	345649	7
18	28520	0
22	24634	2
23	788256	7
24	6595	0
25	115310	4
26	31805	0
30	238223	0
31	279517	2
33	232224	17
35	167179	2

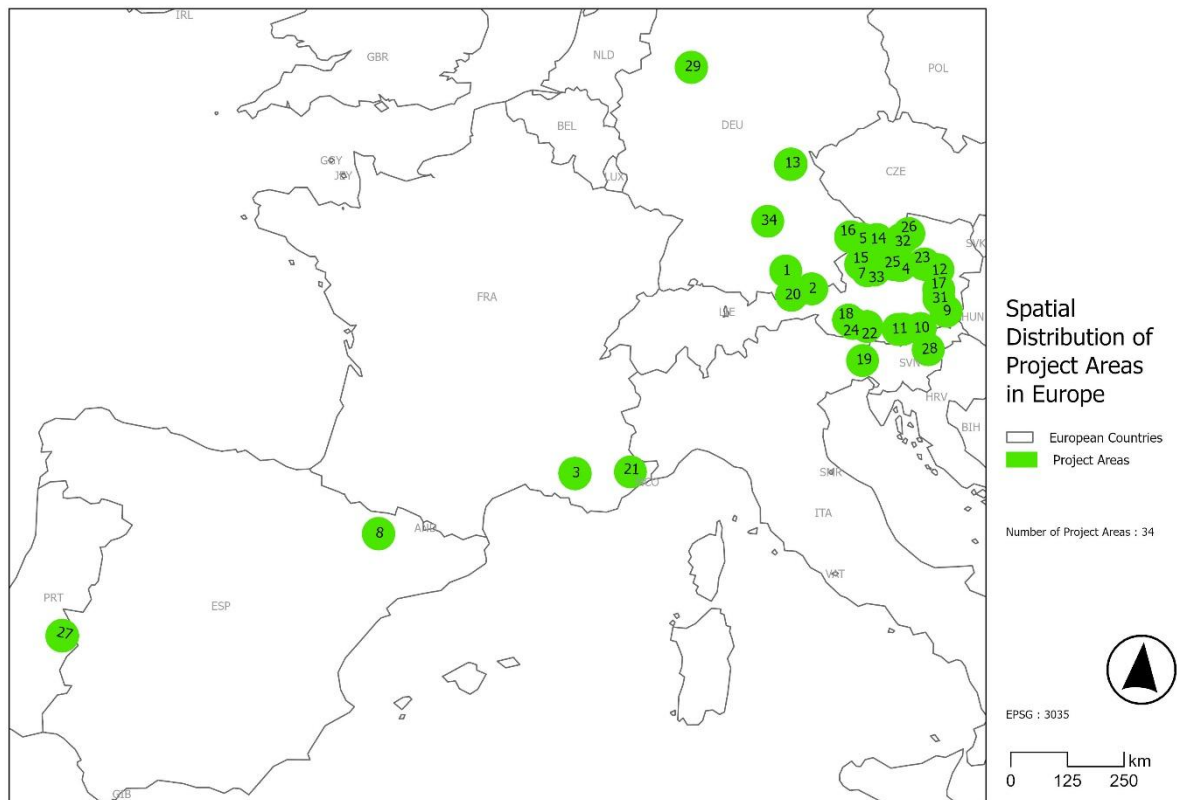


Figure 22 - Spatial Distribution of Project Areas in Europe and their Project IDs.

Table 3 - Overview of the European project areas, including area size [m²] and the number of mapped springs per area.

Project ID	Area [m ²]	Spring Points
1	25733	1
2	46883	2
3	745822	6
4	314167	5
5	233455	0
6	88732	4
7	40153	1
8	1130408	8
9	11062	1
10	84096	0
11	84736	3
12	303221	2
13	1568855	0
14	272433	2
15	32945	0
16	223343	0
17	345649	7
18	28520	0
19	37556	4
20	6764	2
21	88941	1
22	24634	2
23	788256	13
24	6595	0
25	115310	4
26	31805	0
27	2172572	3
28	46358	1
29	593398	0
30	238223	0
31	279517	2
32	132950	0
33	232224	17
34	22753	0

As shown in figure 21 and figure 22, the project areas are widely distributed throughout Austria and Europe, covering a range of geological and environmental conditions. The dataset also includes project areas where no springs were identified.

2.8 Random Forest Classification

The Random Forest algorithm was selected for spring potential mapping based on its established performance in a comparable study (Chen et al., 2020). Two separate models were trained and compared: the Austrian model (AT) with high-resolution national datasets and the

European model (EU) with available pan-European data. The comparison should show whether the approach can be transferred from national to continental scale.

Both models were implemented with the Forest-based Classification and Regression tool in ArcGIS Pro with identical hyperparameters: 500 trees, a minimum leaf size of 5, a maximum tree depth of 10, and 100 % of the training data available per tree. The target variable was binary categorical (1 = spring, 0 = non-spring), as in previous studies (Rahmati et al., 2018; Chen et al., 2020). Variable importance was evaluated with the measure provided by the tool. The spring potential maps were created from the predicted class probabilities, which were exported as raster surfaces. Model performance was assessed with a nested resampling design (section 2.8.1). In addition, the influence of the number of trees and of the sampling design was tested (section 2.8.2).

2.8.1 Model Validation and Uncertainty Assessment

The validation separates two sources of random variation. First, the training dataset was randomly split into 70 % training and 30 % validation data (Rahmati et al., 2018; Chen et al., 2020). This split was repeated 15 times, and the performance metrics (F1-score, Matthews Correlation Coefficient (MCC), sensitivity, and overall accuracy) were summarised as the median of the 15 splits. This step reduces the influence of the random split. Further, the whole procedure was repeated 100 times (Monte Carlo realisations). In each realisation, a new set of pseudo-absence points was drawn, while the verified spring locations stayed the same. This shows how sensitive the models are to the selection of the pseudo-absence points. The results are reported as the mean of the 100 realisations together with the standard deviation and a 95% confidence interval.

2.8.2 Design Sensitivity

The methodological comparisons were done step by step. Within these steps, the best configuration was kept and used for the next test. First the pseudo-absence sampling design (section 2.1.1), then the class ratio and the compensation for sparse categories, and finally the number of trees. The class ratio is the number of pseudo-absence points per spring. Ratios of 1:1, 1:2 and 1:3 were compared with the full set of pseudo-absence points, this corresponds to a ratio of 1:3.4 for Austria and 1:2.2 for Europe. Then each ratio was tested with and without the option to compensate for sparse categories. Because the configurations were optimised step by step and not in all combinations, interactions between the design choices were not evaluated.

The number of trees was varied between 100, 250, 500 and 1000, with 10 repetitions per setting, in order to check whether the fixed value of 500 trees influences the results.

2.8.3 Spatial Cross-Validation

The validation described in section 2.8.1 is based on random splits. Because the spring locations are strongly clustered within the project areas, a random split can place points of the same project area in the training and in the validation set at the same time. The performance is probably measured to some extent too optimistic in this way. To be able to measure how strong the effect is an additional block validation was carried out. Therefore, each project area was left out once and could so be predicted. The training was based on all the remaining samples. In further explanations this cross-validation is called leave-one-area-out (LOAO). Additionally, a random 5-fold cross-validation was calculated with the same model and the same parameters (500 trees, minimum leaf size 5, maximum depth 10, compensation for sparse categories). The only difference between the two designs is the fold structure, so the difference in performance can be attributed to the spatial separation of training and validation data. Both designs were evaluated with the MCC, Specificity, Sensitivity, F1-Score, Accuracy and with the area under the ROC curve (AUC). The AUC was added because it does not depend on a classification threshold. When there is a spatial shift, the uncalibrated probabilities are not comparable at the fixed threshold of 0.5. However the AUC still indicates whether the springs are ranked higher than the non-springs. In addition, the AUC was calculated separately for every excluded project area, to see whether the transfer fails in all areas or only in single ones.

2.9 Correlation Analysis of the Predictor Variables

The correlation analysis was carried out after the models had been trained. The predictor set, and in particular the remote sensing variables, was selected in an exploratory way (see sections 2.1.1 and 2.3). The correlations are therefore a plausibility check of this selection and of the modelling results. They were not used to select or remove variables, and no variable was excluded. All results of this section are used only in a descriptive way.

The connection between each continuous predictor and the samples was calculated with the point-biserial correlation (Pearson, Spring = 0/1). Lithology was excluded, since it is a categorical variable for which a Pearson coefficient cannot be interpreted in a meaningful way. A positive coefficient means that these variables take higher values at spring locations and a negative coefficient means lower values.

Most predictors are not normally distributed and this is why the results were checked for robustness. For every predictor, the Pearson coefficient was compared with the rank-based Spearman coefficient. In addition to that the skewness and the Shapiro-Wilk test for normality were calculated and the distributions were inspected as histograms. Predictors with a difference of $|\text{Pearson} - \text{Spearman}| \geq 0.10$ were reported separately. The reason for this was that the linear coefficient does not describe the association for this variable correctly. Collinearity between the predictors was measured with a Pearson correlation matrix and restricted to the verified spring locations only. Pairs with $|r| \geq 0.7$ were considered strongly correlated. All predictors were kept in the models, because in models like RF the predictive performance is robust against correlated input variables (Breiman, 2001). It has to be noted that the variable importance measure is not equally robust. Correlated predictors share their contribution and can show an incorrect tendency compared to uncorrelated predictors (Strobl et al., 2008). The importance values documented in section 3.3 therefore have to be read at the level of the predictor groups rather than as an exact ranking of single variables within a group.

It has to be kept in mind that the linear correlation is not the same as the variable importance reported in section 3.3. Random Forest detects non-linear and combined effects, for example the categorical lithology or the strongly skewed flow indices, which a single linear coefficient cannot show. A low correlation coefficient therefore does not mean that a variable is unimportant for the model.

2.10 Spring Potential Maps

For the final prediction, both models were trained on the complete training dataset with the compensation for sparse categories switched on. The trained models were subsequently applied to the prediction grid and the predicted class probability for spring presence was calculated for every point of this grid. The values were then converted into a raster surface. Moreover the prediction grid is based on the Sentinel point grid and has a resolution of 10 m in both regions. It covers the project areas with an additional buffer of approximately 100 - 500 m, depending on there was a digitized spring close to the project area boundary. This is the area where complete predictor information is available. The resulting values have to be read as a relative ranking of the spring potential and not as absolute probabilities of occurrence. Because the training data was balanced and compensated, while the real share of springs in the landscape is much smaller, the predicted probabilities are not calibrated to the true prevalence. For the interpretation, the raw values were therefore additionally converted with a prior correction to the real prevalence, in order to show which absolute probability a given raw value corresponds to. Additionally, the results of the LOAO validation were visualized and compared

to the predictive maps. The main difference is that in the LOAO maps the areas with their springs were not part of the training.

2.11 Independent Evaluation of the Prediction Maps

In addition to the model-based performance metrics, the prediction maps were assessed by further independent inspections. These comprised a quantitative success rate, a comparison with independently mapped waterlogged areas and a qualitative review by the project geologist who conducted the field surveys.

2.11.1 Success Rate

The success rate describes how strongly the predicted spring potential concentrates the verified spring locations within a small proportion of the study area. All cells of the prediction surface were ranked by their predicted probability in descending order. For increasing shares of the ranked area (1%, 2%, 5%, 10%, 20%, 30%, 40% and 50%), the number and share of verified spring locations falling within this area were determined. The result is a cumulative curve of captured springs plotted against the share of area. A model without any information would find as many springs as the share of area searched, so that 10 % of the area would contain 10 % of the springs. The distance of the curve above this expectation therefore indicates the gain compared with a random search.

Three variants of the curve were calculated. In the first variant, all project areas were included in the training data. In the second variant, each area was predicted by a model that had been trained without it (LOAO), and the cells of all areas were subsequently ranked within one pooled list. In the third variant, the cells of each area were ranked among themselves only, which corresponds to the survey of a single unknown area. All cells of the prediction grid were included in the ranking. A restriction to the project polygons was tested but not applied, since a part of the verified springs is located outside the polygons and such a restriction would retain only those springs that happen to lie sufficiently far inside.

2.11.2 Independently Mapped Waterlogged Areas

The original field dataset also contains waterlogged points, which were mapped as temporary seepage zones and were excluded from the training data. These points were used as an independent plausibility check and to answer research question 4. The predicted probability was extracted at each waterlogged point and compared with the distribution of the predicted probabilities across the entire model area. If the model responds to wet conditions in a consistent way, the waterlogged points should receive distinctly higher probabilities than the

area as a whole. The check was carried out separately for both regions, which were distinguished by the spatial position of the points.

However, it should be noted that this test does not represent a fully independent validation. The models include wetness-related predictors (Sentinel-1 and Sentinel-2), the waterlogged points were mapped by the same geologist as the spring locations, and spring potential and surface wetness are not equivalent. The result is therefore interpreted as an indication of plausibility rather than as a proof.

2.11.3 Qualitative Expert Review

In addition to the quantitative measures, the final spring potential maps were reviewed by the project geologist, who also carried out the field mapping of the spring locations. Several maps of different project areas were presented together with the project area boundaries and the verified spring locations. The review was conducted informally and focused on whether the predicted high-potential zones are hydrogeologically plausible and whether the maps are useful for planning future field surveys.

The assessment is based on a single expert and on a first visual impression. It is therefore reported as a qualitative statement without statistical significance.

2.12 Ablation Experiments

Three ablation experiments were carried out after the main modelling. In each experiment, a defined part of the information was removed and the resulting effect was measured. The first removes a group of predictors, the second identifies predictors that carry no information within a project area at all, and the third removes the differences between the project areas. None of these experiments was used to select or to tune the model.

2.12.1 Removal of the Satellite Variables

The Sentinel images were selected and scaled separately for each project area. In principle, this could create an artificial difference between the areas that is unrelated to spring occurrence. If this were the dominant effect of the satellite variables, a model without them should transfer better to an unknown area rather than worse.

To test this, the LOAO validation described in section 2.8.3 was repeated with the six satellite variables removed. The remaining 16 predictors comprise the terrain variables, the hydrological indices, the lithology and the soil variables. Model settings, training points and fold structure are identical to the full run, so that the set of predictors is the only difference. The two runs are compared using the AUC, because the AUC is threshold-independent and is therefore not affected by the shift in calibration described in section 3.7.

2.12.2 Within-Area Variability of the Predictors

The soil data have a cell size of 1 km, whereas the project areas have a median size of about 0.1 km². A predictor that takes only a single value within a project area cannot separate springs from non-springs in that area; it can only separate one area from another.

Consequently, a predictor that is constant within a project area is effectively already ablated there, since it cannot contribute to the separation within that area, whether or not it is included in the model. To quantify how frequently this occurs, the number of distinct values of every predictor was counted separately for each project area and compared with the number of points in that area. A predictor with a single value in an area is counted as constant there. The calculation is based on the training table only and requires no model runs.

2.12.3 Local discriminative power of the predictors

The variable importance and the correlations in sections 3.3 and 3.5 are calculated over all points together. In such a calculation most spring–absence pairs come from two different project areas, so the values describe mainly what separates the areas from each other. To measure what separates springs from non-springs inside the same area, the between-area part has to be removed. In this experiment the between-area component of every predictor is removed by centring the predictor on the mean of its own project area and dividing by the standard deviation of that area. What remains is only the variation inside an area. Predictors that are constant in an area drop out automatically, because their standard deviation is zero. The centred values were then pooled and the standardized difference between springs and non-springs was calculated, together with the share of project areas in which the difference points in the same direction as the pooled effect.

Because this measure is based on means, it is sensitive to skewed distributions in the same way as a linear correlation. Several predictors have a skewness between 5 and 13 (see section 3.5). The calculation was therefore repeated with ranks instead of values. For each area, the probability that a randomly chosen spring has a higher value than a randomly chosen non-spring of the same area. A value of 0.5 means no separation, and extreme values cannot distort it. Both measures are bivariate. A random forest also uses interactions between predictors, so a predictor with a small value can still contribute inside the model. A predictor that shows no local separation at all can contribute little inside an area.

3 Results

3.1 Model Performance Metrics

The final models were evaluated with 100 Monte Carlo realisations of the pseudo-absence sample (section 2.8.1). As you can see in table 4, the Austrian model reached an MCC of 0.465 ± 0.059 (95 % confidence interval [0.453; 0.476]) and a median accuracy of 0.73. The European model reached an MCC of 0.370 ± 0.057 (95 % confidence interval [0.359; 0.382]) and a median accuracy of 0.68. The confidence intervals do not overlap, so the Austrian model performed better than the European model. The variation between the realisations was moderate in both regions (coefficient of variation of the MCC: 12.6 % for AT, 15.4 % for EU).

Single model runs with identical data differed by up to 0.1 MCC. This confirms that single runs are not sufficient for a reliable comparison and that the repeated evaluation was necessary.

Table 4 - Comparison of the Austrian (AT) and European (EU) model. Top: final models (complete fixed-per-area pseudo-absence sets with class compensation; single runs), used for the spring potential maps. Bottom: Monte Carlo uncertainty analysis (100 realisation; balanced design without class compensation, pseudo-absences drawn from the fixed-per-area pool).

	AT	EU
Final model (fixed-per-area, class compensation; single run)		
MCC	0.48	0.39
F1 (spring class)	0.61	0.59
Sensitivity (spring class)	0.70	0.64
Median accuracy	0.79	0.72
Monte Carlo uncertainty analysis (balanced, no compensation; 100 realisations)		
MCC (mean $\pm$ SD)	0.465 ± 0.059	0.370 ± 0.057
95 % confidence interval (MCC)	[0.453; 0.476]	[0.359; 0.382]
Coefficient of variation (MCC)	12.6 %	15.4 %
Median accuracy	0.73	0.68

3.2 Sensitivity to the Sampling Design

The sampling design of the pseudo-absence points had a strong influence on the model results (table 5). With the first spatially uncontrolled design, the European model reached an MCC of 0.56 and the Austrian model 0.49. With the controlled fixed-per-area design, the order turned around: AT 0.47 and EU 0.37. The difference of about 0.19 MCC for the European model shows

the spatial bias of the uncontrolled design. In this design, 40 % of the pseudo-absences came from only two large project areas without springs, so the model could partly learn the region instead of the spring conditions. The area-weighted design reached slightly lower values than the fixed-per-area design (single runs: EU 0.34, AT 0.42).

The class ratio and the compensation for sparse categories also had a clear effect. Without compensation, the sensitivity for the spring class dropped strongly in the imbalanced designs (EU 1:2: 0.36; EU area-weighted: 0.24; AT 1:3: 0.32). At the same time, the overall accuracy stayed high (0.75–0.81), because the models mostly predicted the larger non-spring class. With compensation, the sensitivity stayed at 0.61–0.70 and the MCC stayed nearly the same (table 5).

Table 5 - Model performance of the tested sampling designs and class ratios (single runs, Random Forest, 500 trees). Single runs are subject to random variation of up to ~0.1 MCC. The variants “fixed-per-area, full set, compensation” are the final model configuration used for the spring potential maps. The spatially uncontrolled design (EU MCC 0.56, AT 0.49) is only reported in the text.

Variant	Springs + pseudo-absences	Compensation	MCC	F1 (spring)	Sensitivity (spring)	Median accuracy
EU fixed-per-area, balanced subsample	91 + (91 of 204)	No	0.38	0.62	0.65	0.704
EU fixed-per-area, full set (1:2), no compensation	91 + 204	No	0.32	0.45	0.36	0.750
EU fixed-per-area, full set (1:2), compensation	91 + 204	Yes	0.39	0.59	0.64	0.722
EU area-weighted, no compensation	91 + 284	No	0.31	0.35	0.24	0.804
EU area-weighted, compensation	91 + 284	Yes	0.34	0.52	0.61	0.728
AT fixed-per-area, balanced subsample	64 + (64 of 220)	No	0.37	0.70	0.70	0.684
AT fixed-per-area, full set (1:3), no compensation	64 + 220	No	0.45	0.47	0.32	0.812
AT fixed-per-area, full set (1:3), compensation	64 + 220	Yes	0.48	0.61	0.70	0.791
AT area-weighted, compensation	64 + 201	Yes	0.42	0.58	0.65	0.765

3.3 Variable Importance

The variable importance was averaged over the 100 Monte Carlo realisations (balanced design, pseudo-absences drawn from the fixed-per-area pool; section 2.8.1). In both regions, the topographic variables (including hydrological indices) were the most important group (EU: 48.8 ± 2.0 %; AT: 40.2 ± 3.8 %) with 10 variables. The remote sensing variables (Sentinel-1 and Sentinel-2) followed in second place (EU: 27.6 ± 1.5 %; AT: 23.9 ± 3.1 %) containing 6 variables. The soil variables were much more important in the Austrian model (24.2 ± 3.2 %) than in the European model (11.5 ± 1.4 %). Both datasets contained similar number of 5 variables. Geology was the smallest group (EU: 6.6 %; AT: 5.7 %). However, lithology was one of the strongest and most stable single variables in both regions (table 6). The variable importance consists of a different amount of variables, that is why the group importance is not directly comparable with each other.

The single variables (figure 23) show a robust core in both regions. This are lithology, the Sentinel-1 VH backscatter, profile curvature, and STI. In the Austrian model, the most important single variable was a radar channel (Sentinel-1 VH April–May, 6.4 %). In both regions, the Sentinel-1 variables were more important than the Sentinel-2 variables. The Sentinel-1 difference channel was useful in the European model (4.2 %), but it was almost not important in the Austrian model (1.4 %). TWI and flow accumulation were less essential and less stable in the Austrian model (about 3.2 %, $CV > 50$ %) than in the European model (4.5 – 4.7 %)

Table 6 - Mean variable importance by predictor group (mean $\pm$ SD over 100 Monte Carlo realisations, balanced design, pseudo-absences drawn from the fixed-per-area pool). Topography: DEM and derived variables including the hydrological indices (slope, aspect, general curvature, plan curvature, profile curvature, TWI, SPI, STI, flow accumulation); Remote sensing: Sentinel-1 and Sentinel-2, Geology: 6 permeability classes. Only the top 20 of 22 predictor variables are considered. This is why the group values do not sum up to 100%.

Group	EU model	AT model
Topography (10 variables)	48.8 ± 2.0 %	40.2 ± 3.8 %
Remote sensing (6 variables)	27.6 ± 1.5 %	23.9 ± 3.1 %
Soil (5 variables)	11.5 ± 1.4 %	24.2 ± 3.2 %
Geology (1 variable)	6.6 ± 0.7 %	5.7 ± 0.8 %

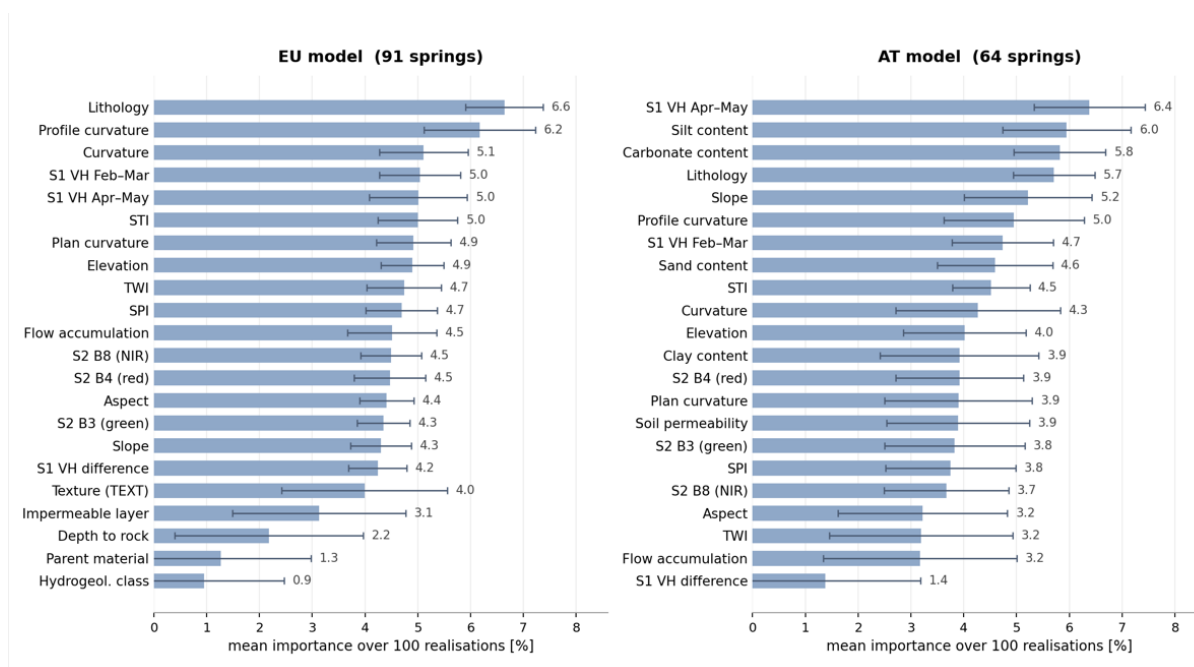


Figure 23 - Importance of single predictor variables, given as mean over 100 Monte Carlo realisations with the standard deviation. Only the top 20 of 22 predictor variables are considered. The mean predictor value is 4,5%.

3.4 Sensitivity of the tree numbers

The sensitivity of the model performance to the number of trees was tested with the final model configuration (100, 250, 500, and 1000 trees; 10 repeated runs each; table 7 and figure 24). The mean MCC varied only within a small range (EU: 0.325–0.359; AT: 0.441–0.479) and showed no trend towards more trees. The differences between the settings were smaller than the standard deviation of the repeated runs. In the European model, the standard deviation dropped from 0.060 with 100 trees to about 0.032 with 250 or more trees (figure 24).

Table 7 - Model performance with different numbers of trees (final model configuration: fixed-per-area pseudo-absences with class compensation; mean $\pm$ SD over 10 repeated runs per setting).

Trees	EU model			AT model		
	MCC (mean $\pm$ SD)	Accuracy	Sensitivity	MCC (mean $\pm$ SD)	Accuracy	Sensitivity
100	0.342 $\pm$ 0.060	0.693	0.645	0.479 $\pm$ 0.048	0.795	0.680
250	0.325 $\pm$ 0.033	0.687	0.629	0.447 $\pm$ 0.055	0.786	0.645
500	0.359 $\pm$ 0.032	0.710	0.621	0.473 $\pm$ 0.054	0.795	0.670
1000	0.355 $\pm$ 0.032	0.709	0.618	0.441 $\pm$ 0.041	0.785	0.640

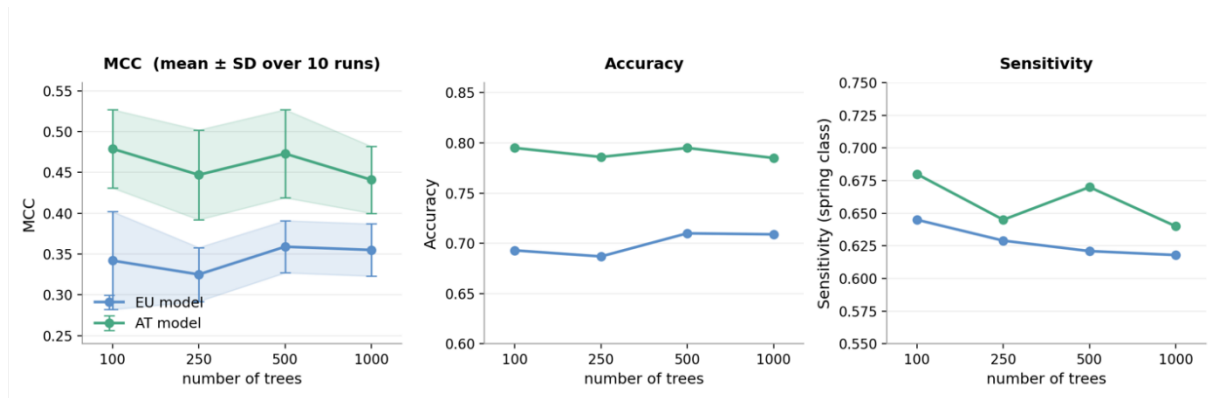


Figure 24 - Sensitivity of both models to the number of trees. Each point is the mean of 10 independent runs; the error bars and the shaded band in the left panel give the standard deviation over these runs. The differences between 100, 250, 500 and 1000 trees are smaller than the scatter between repeated runs with the same setting, so the number of trees was not decisive for the result. All runs use the balanced training set and the random 70/30 split described in section 2.8.1.

3.5 Correlation Analysis

The correlation between the single predictors and the spring occurrence is weak in both regions. The highest point-biserial coefficients are $r = 0.24$ for the Sentinel-1 VH backscatter of April–May in Austria and $r = -0.24$ for the profile curvature in Europe. All other predictors stay below $|r| = 0.25$ (figure 25).

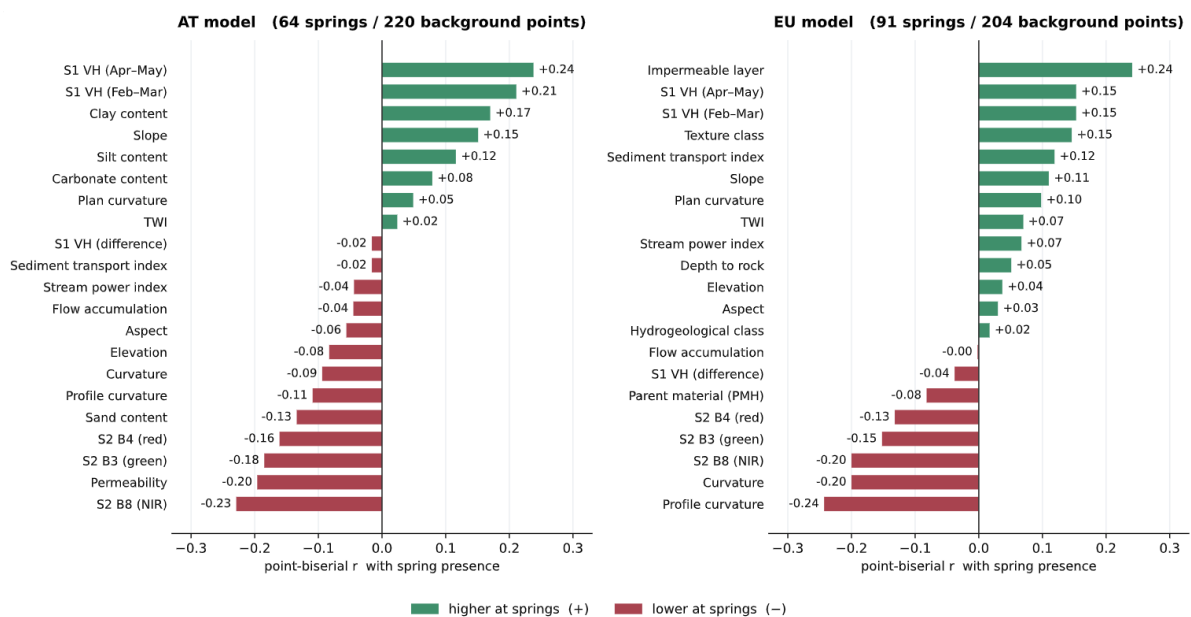


Figure 25 - Single variable correlation with spring presence in both models

The direction of the coefficients is consistent between the two regions. All three Sentinel-1 channels except the difference channel show positive values at spring locations (AT: 0.21 and 0.24; EU: 0.15 for both periods), while all three Sentinel-2 bands show negative values (AT: -0.16 to -0.23; EU: -0.13 to -0.20). Spring locations are therefore characterised by a higher radar backscatter and a lower optical reflectance than the surrounding points. The slope is slightly

positive in both models (AT: 0.15; EU: 0.11) and the general and profile curvature are negative (EU: -0.20 and -0.24), which corresponds to concave, converging terrain positions.

The soil variables behave differently in the two regions. In the Austrian model, the soil permeability ($r = -0.20$) and the clay content ($r = 0.17$) reach comparatively high values, while in the European model the depth of the impermeable layer is the strongest soil variable ($r = 0.24$). This matches the higher importance of the soil group in the Austrian model (section 2.3).

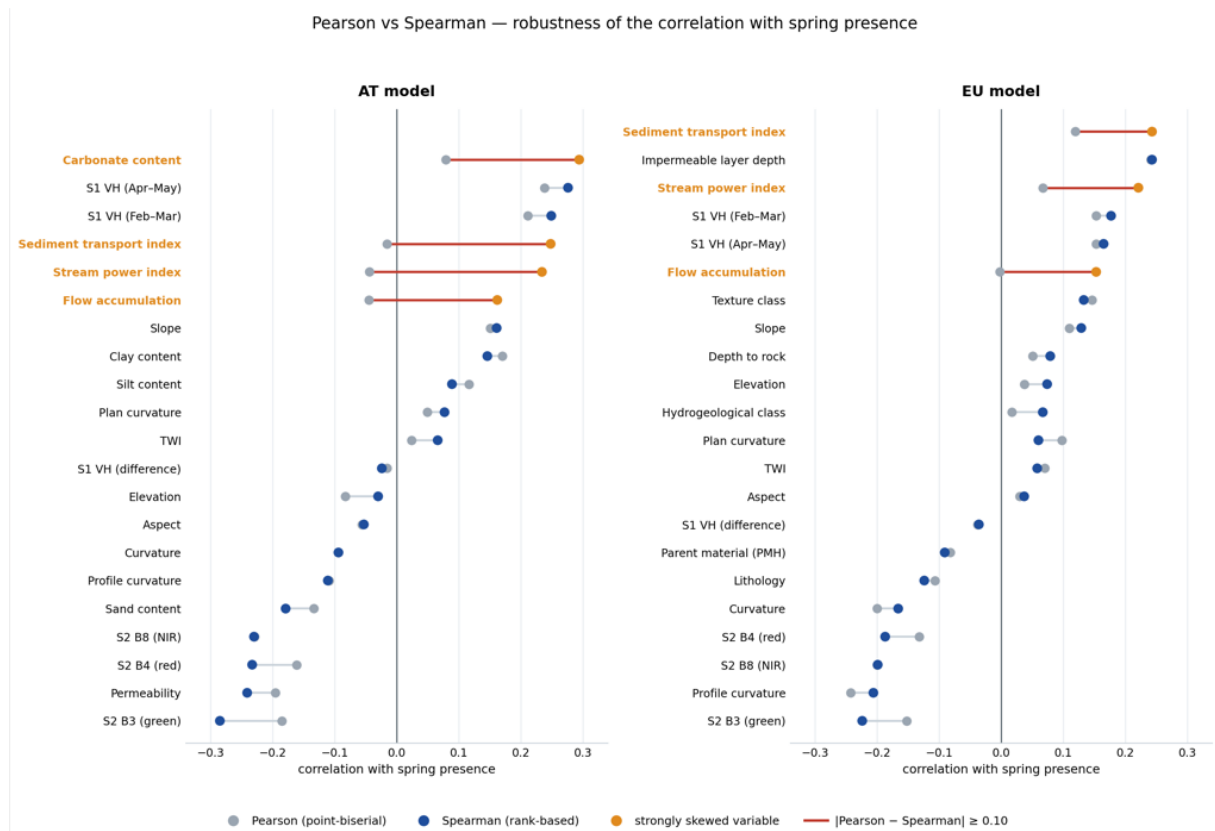


Figure 26 - Pearson and Spearman correlation of each predictor with spring presence, for the Austrian model (left) and the European model (right). Grey points show the linear point-biserial correlation, blue points the rank correlation. Variables shown in orange are strongly skewed variables.

Almost all predictors are not normally distributed (figure 26) (Shapiro-Wilk $p < 0.05$ for 21 of 22 variables in Austria and for 21 of 22 variables in Europe). For most predictors, the rank-based Spearman coefficient is nearly identical to the Pearson coefficient, so the analysis is robust against the deviation from normality. Only the strongly right-skewed variables differ clearly ($|\text{Pearson} - \text{Spearman}| \geq 0.10$). Flow accumulation, SPI and STI differ in both regions, and additionally the carbonate content in Austria. For these variables, the Pearson coefficient is flattened by extreme values and stays near zero, while Spearman shows a weak positive monotone association with the springs (AT: SPI 0.23, STI 0.25, flow accumulation 0.16; EU: SPI 0.22, STI 0.24, flow accumulation 0.15). The hydrological indices are therefore not unrelated to the spring occurrence, as the Pearson value alone would suggest.

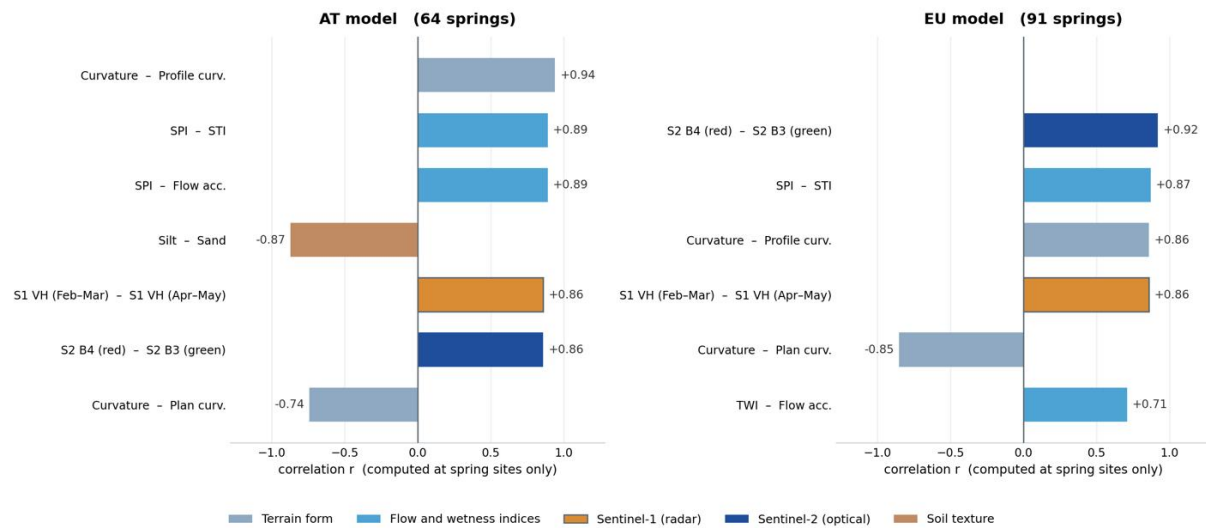


Figure 27 - Correlation between predictor variables, calculated only at the verified spring locations. Only pairs with $|r| \geq 0.70$ are shown. Every strong pair consists of two variables from the same family, and no strong correlation occurs between families. Correlated predictors share their importance in a random forest, which is why the variable importance is reported by group rather than for single variables.

Restricted to the verified spring locations, all strong correlations ($|r| \geq 0.7$) fall within the same variable family (Figure 27). The general and the profile curvature reach $r = 0.94$ in Austria and $r = 0.86$ in Europe, the general and the plan curvature $r = -0.74$ and $r = -0.85$. SPI and STI correlate with $r = 0.89$ (AT) and $r = 0.87$ (EU), SPI and flow accumulation with $r = 0.89$ in Austria and TWI and flow accumulation with $r = 0.71$ in Europe. The two Sentinel-1 VH periods correlate with $r = 0.86$ in both regions, and the Sentinel-2 red and green bands with $r = 0.86$ (AT) and $r = 0.92$ (EU). In the Austrian soil data, sand and silt correlate with $r = -0.87$.

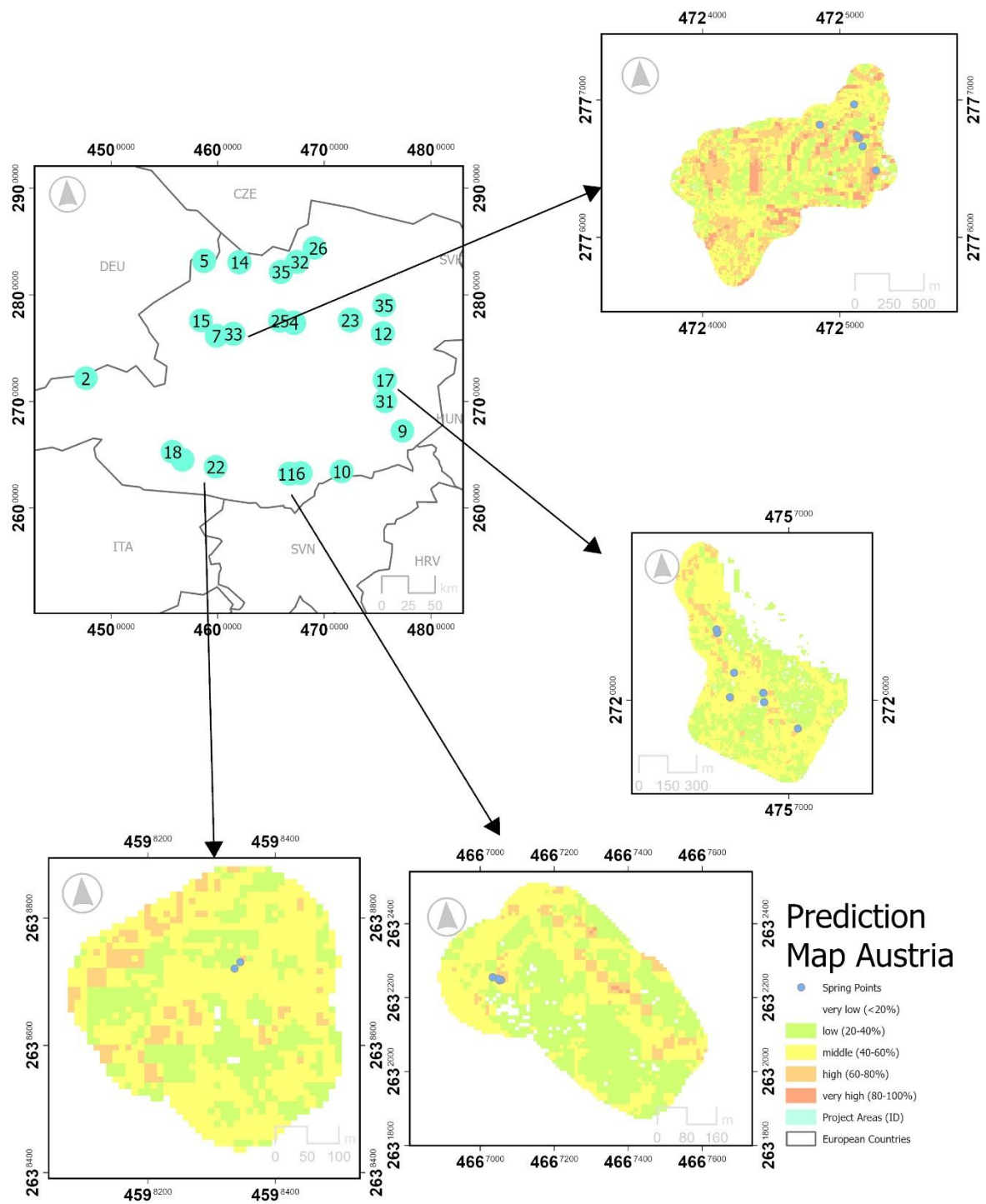
These values are not surprising, because the variables of one family are derived from the same source. The curvatures are second derivatives of the same surface, SPI, STI and TWI are all calculated from the flow accumulation, and the Sentinel channels come from the same sensor. No pair of variables from two different predictor groups exceeds $|r| = 0.7$. The highest values across the groups are found between the Sentinel-1 VH channels and the Sentinel-2 red and green bands (AT: -0.42 and -0.36 ; EU: -0.36 and -0.37), while the correlation between Sentinel-1 VH and the Sentinel-2 NIR band is close to zero (AT: -0.09 ; EU: -0.07).

Because the correlations are only linear and only pairwise, they cannot be compared directly with the variable importance of section 3.3. Lithology, for example, is the strongest single predictor in the model, but as a categorical variable it is not part of this analysis.

3.6 Spring Potential Maps

Figures 28 and 30 show the spring potential maps of the final models (predicted spring probability trained with full dataset, 10 m resolution). Whereas figure 29 and 31 show spring potential trained with LOAO, so the predicted area was not included in the training dataset. High predicted probabilities concentrate in topographic convergence zones such as hillslope hollows

and along drainage lines, whereas ridges show low values.



EPSG : 3035

Figure 28 - Prediction Map Austria trained with the full dataset.

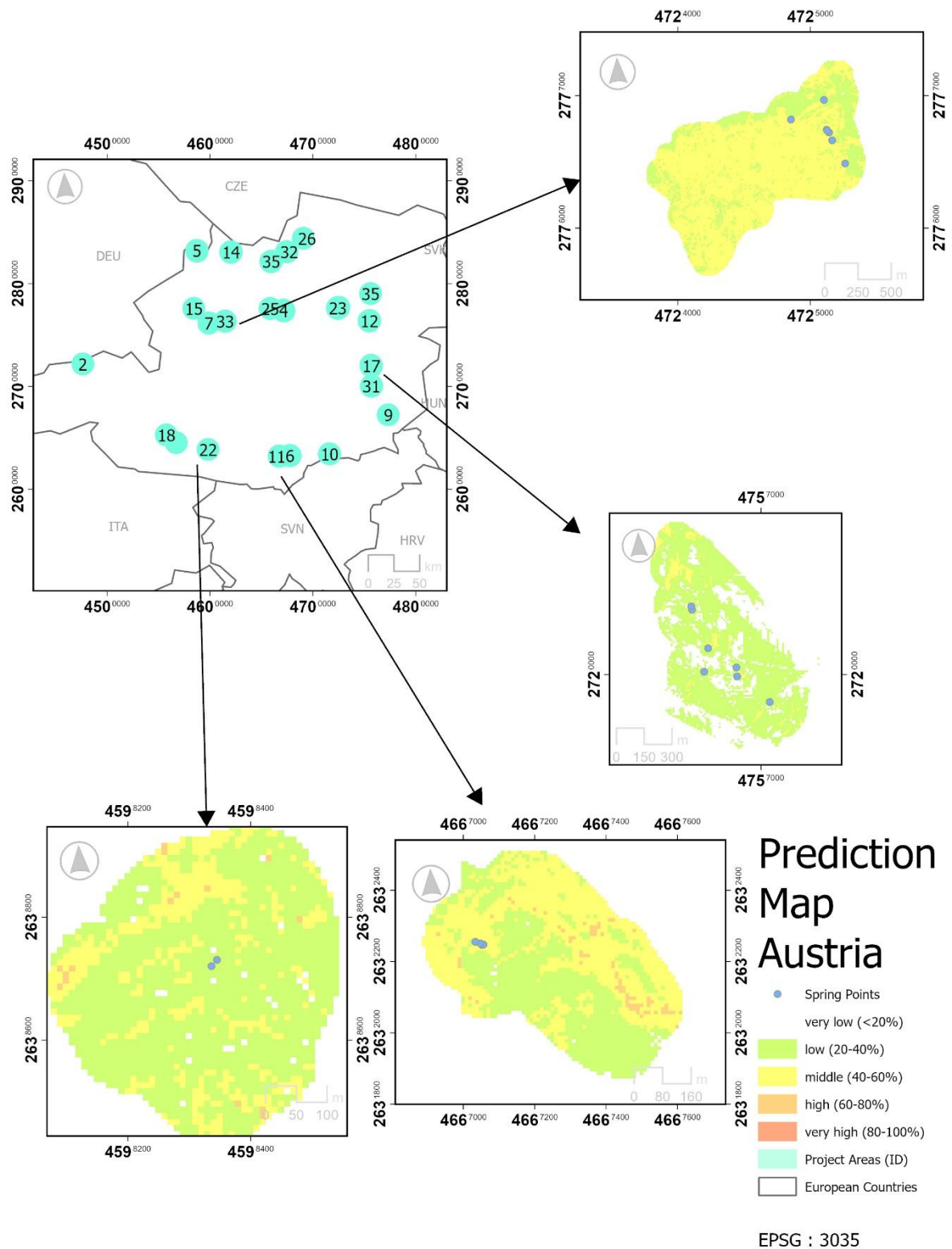


Figure 29 - Prediction Map created with LOAO.

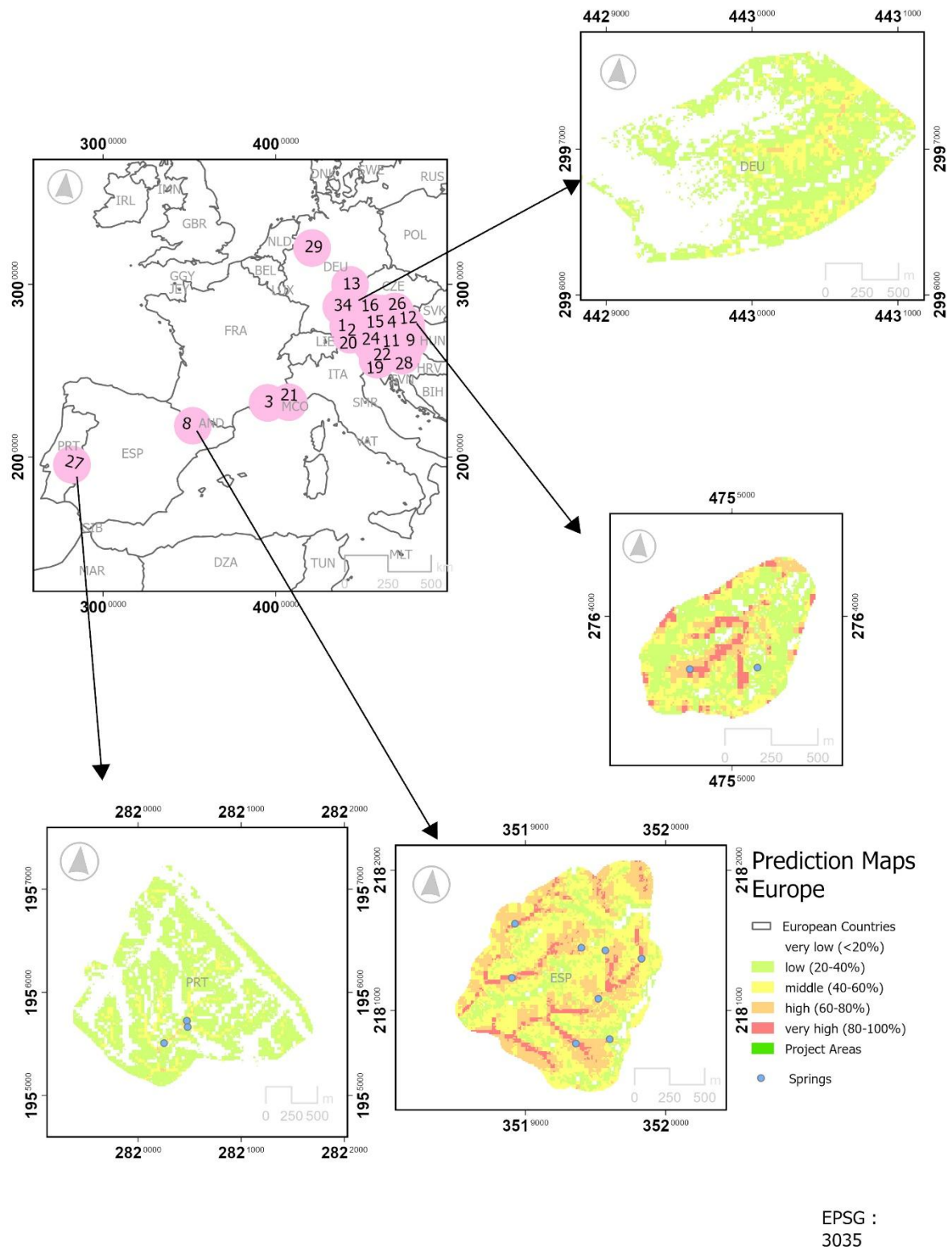


Figure 30 - European Prediction Map created with full dataset.

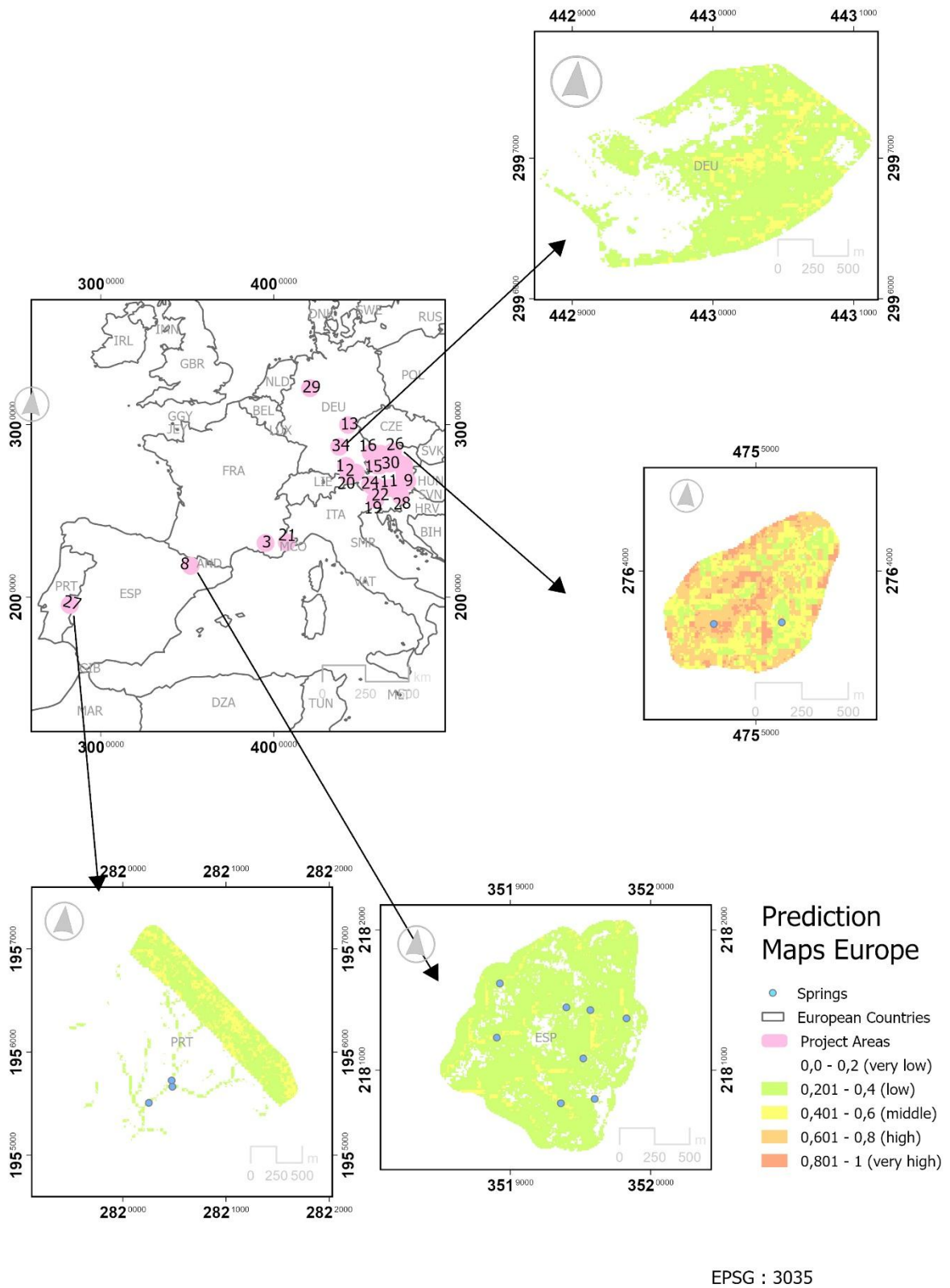


Figure 31 - Prediction Map created with LOAO.

The Prediction maps (LOAO) (figure 29 and 31) do not have an adapted threshold. This is why the spring probabilities seem to be massively underestimated. The Prediction maps (full dataset)

were additionally checked with an independent dataset: the mapped waterlogged areas from the field surveys, which were not used for model training. These points received clearly higher predicted probabilities than the model area on average (AT: mean 0.48 and 46.0 % above 0.5 at 654 points, versus 0.40 and 31.7 % over the whole area; EU: mean 0.45 and 38.0 % above 0.5 at 898 points, versus 0.35 and 22.9 %). The models therefore assign increased spring potential to independently mapped wet locations.

3.7 Spatial cross-validation

Figure 32 compares the two validation designs. In the random five-fold design the points of every project area are distributed over all folds, so the model always knows at least some points of the area it is tested on. In the leave-one-area-out design a whole project area is removed from the training data and then predicted. Model, settings and data are identical in both designs, only the way the folds are formed differs.

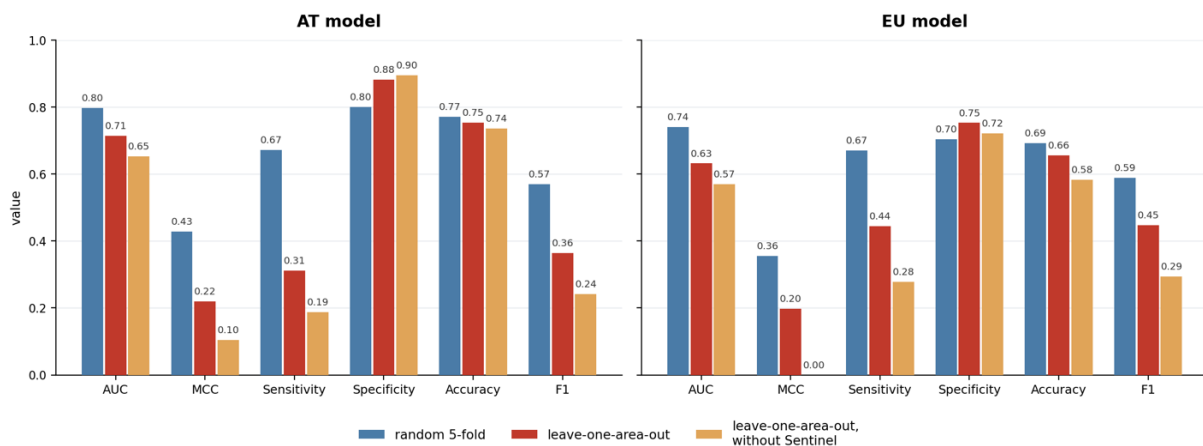


Figure 32 - Performance of the Austrian and the European model under random five-fold cross-validation and under leave-one-area-out validation, and the same spatial validation with the six satellite variables removed. All values except the AUC are calculated at the fixed threshold of 0.5 on uncalibrated probabilities.

The AUC drops from 0.797 to 0.715 for the Austrian model and from 0.741 to 0.632 for the European model. The loss is therefore 0.08 and 0.11 respectively. The ranking of the points is largely kept when the model is applied to an area it has never seen.

The threshold-based measures fall much more strongly. For the Austrian model the MCC drops from 0.43 to 0.22 and the sensitivity from 0.67 to 0.31, while the specificity rises from 0.80 to 0.88. If the model were simply worse in an unknown area, both sensitivity and specificity would fall. That one falls while the other rises shows something different: the predicted probabilities are compressed towards the middle, so fewer points reach the fixed threshold of 0.5. The model becomes undecided rather than wrong. The same pattern appears in the European model, where the sensitivity falls from 0.67 to 0.44 and the specificity rises from 0.70 to 0.75.

The comparison between the two models follows the resolution of the input data. The Austrian model, which uses national datasets, keeps a higher AUC under spatial validation (0.715) than the European model with pan-European datasets (0.632).

The third row of each block shows the same spatial validation with the six satellite variables removed. The AUC falls by a further 0.06 in both models, from 0.715 to 0.653 for Austria and from 0.632 to 0.570 for Europe. For the European model the MCC reaches zero, which means that the classification at the fixed threshold is no better than chance. This result is relevant for the interpretation of the satellite variables. The Sentinel images were selected and scaled per project area, which could in principle create an artificial separation between areas. If this were the main effect, removing the satellite variables would improve the transfer to an unseen area. The opposite happens in both models. Nevertheless, the small sample size has to be kept in mind and limits the statistical meaning.

A single AUC can only be calculated for project areas that contain springs and pseudo-absences. This applies to 14 of the 22 Austrian areas and to 21 of the 34 European areas. For the European model the median AUC per area is 0.667 with all variables and 0.542 without the satellite variables. The AUC is above 0.5 in 19 of 21 areas with all variables, but only in 12 of 21 areas without them. Areas with only one or two springs have a very uncertain area AUC, since a single point changes the value substantially. The per-area values are therefore reported for completeness, and the interpretation is based on the pooled values and the median.

In the European model one project area could not be predicted. It contains a lithology class that occurs nowhere else, and the tool cannot apply a trained model to a category it has not seen. This area is missing from the pooled values of the spatial validation, and one fold of the random five-fold design is missing for the same reason. This explains the different numbers of points between the designs.

3.8 Independent Evaluation of the Maps

3.8.1 Success rate

The success rate was calculated in three variants (figure 33), each of which addresses a different question. The first variant shows how well the fitted model reproduces the spring locations it was trained on, whereas the second and third variants show the behaviour of the models in a project area that had not been part of the training data. The first variant is therefore optimistic by construction, and the difference between it and the two remaining variants forms the essential aspect of the comparison.

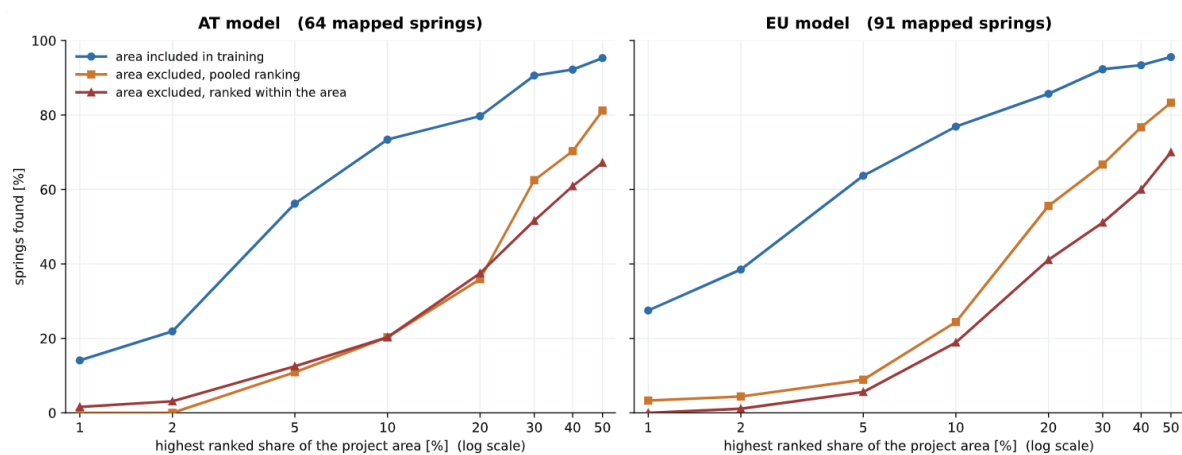


Figure 33- Success rate curves for the Austrian model (64 mapped springs) and the European model (91 mapped springs), logarithmic horizontal axis. The horizontal axis gives the share of a project area that is searched, ranked by predicted potential, and the vertical axis the share of mapped springs found within this share.

For the fitted model, the concentration of verified springs is strong. Within the 5% of the area with the highest predicted probability, the Austrian model captures 56.2% of the verified springs and the European model 63.7%; within the top 10% these values rise to 73.4% and 76.9%, and within the top 20% to 79.7% and 85.7%. In both models, the median spring is therefore located within the highest 5% of the ranked area. However, it should be noted that these values describe how tightly the models concentrate known spring locations and do not indicate the performance to be expected at unknown locations. When a project area is excluded from the training data and subsequently ranked among its own cells, the values decrease considerably. Within the top 10% of an unseen area, 20.3% of its springs are found by the Austrian and 18.9% by the European model, within the top 20% 37.5% and 41.1%, and within the top 30% 51.6% and 51.1%. In the smallest area shares, however, the two models behave differently. The European model does not exceed a random ranking below 5% of the area, whereas the Austrian model already reaches approximately 2.5 times the value of a random ranking at 5%. From 10% of the area onwards, both models correspond to a factor of approximately two compared with a random search, and the two curves remain close to each other over the remaining range, although the models were built from independent data sources. If the cells of all project areas are ranked in one pooled list instead, the curves lie higher, although the extent of this effect differs between the two models. In the European model, the pooled ranking is superior over almost the entire range, and the gap opens up at 20% of the area (55.6% against 41.1%). In the Austrian model, by contrast, the two leave-one-area-out variants are nearly identical up to 20% of the area; at that point the within-area ranking even lies marginally higher. The pooled ranking separates only from 30% onwards. The pooled ranking benefits from the fact that some project areas contain more springs than others, which provides no advantage when a single project

area is surveyed. Consequently, the difference between the two leave-one-area-out variants can be regarded as a measure of how much of the apparent model performance results from differences between the project areas rather than from the localisation of a spring within an individual area. In the European model this share is considerable, whereas in the Austrian model it becomes relevant only in the upper half of the ranked area.

3.8.2 Independently mapped waterlogged areas

The prediction maps were additionally checked against an independent dataset, namely the waterlogged areas mapped during the field surveys as points, which had not been used for model training. At these points, the predicted probabilities are higher than the average over the model area in both models. This test was made with the full dataset. In the Austrian model, the 654 waterlogged points reach a mean predicted probability of 0.48, and 46.0% of them exceed a probability of 0.5, compared with 0.40 and 31.7% over the whole model area. In the European model, the corresponding values are 0.45 and 38.0% at 898 points, compared with 0.35 and 22.9%. Both models therefore assign an increased spring potential to independently mapped wet locations, which answers research question 4.

3.8.3 Qualitative expert review

The prediction maps were reviewed by the project geologist, who had carried out the field mapping himself, and a first impression was recorded. According to his assessment, the predictions perform well on silicate bedrock. In addition to the spring locations themselves, waterlogged sites are indicated as well, and mass movements are clearly reflected in the maps. On carbonate bedrock, by contrast, the predictions perform poorly. In general, the more natural the landscape, the more reliably the springs are detected. Waterlogged areas caused by surface water or by drainage structures are detected. This underlines the results of 3.8.2. Although these areas are in reality larger than in the maps.

Three types of false positives were identified. Ditches and thalwegs are assigned a high potential even where they are dry. Small forest patches, both deciduous and coniferous, are likewise assigned a high potential. Fills and areas reshaped by excavation are incorrectly classified as wet.

The false positives constitute the clearest limitation of the maps and are taken up in the discussion. However, it should be noted that the assessment is based on a single expert and on a first visual impression of a small number of maps. It is therefore reported as a qualitative statement without statistical value.

3.9 Ablation experiments

3.9.1 Removing the satellite variables

Under leave-one-area-out validation, the AUC decreases from 0.715 to 0.653 in the Austrian model and from 0.632 to 0.570 in the European model, which corresponds to a loss of 0.062 in both models. In the European model, the MCC drops to zero, which means that the classification at the fixed threshold is no better than chance.

This result addresses the objection that the satellite variables could carry an artificial signal. The Sentinel images were selected and scaled separately for each project area and could therefore, in principle, separate the areas from each other without providing any information about spring occurrence. If this were the dominant effect, a model without these variables would transfer better to an unknown area. The opposite is observed, in both models and to the same extent.

3.9.2 Predictors without variation inside a project area

In the Austrian model, the five soil variables take a single value in 66% of the project areas and the lithology in 36% (figure 34). In the European model, this effect is considerably stronger: the five soil variables are constant in 94% of the 34 areas and the lithology in 85%. All terrain variables and all satellite variables vary from point to point in every area of both models.

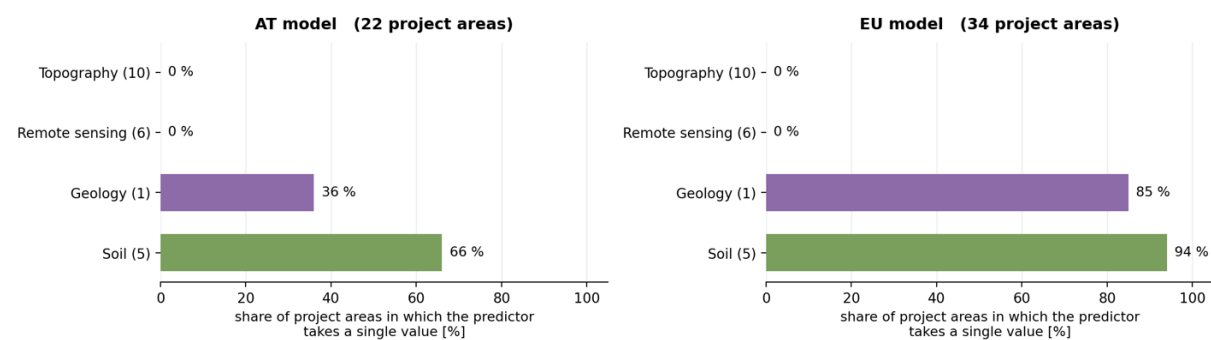


Figure 34 - Ablation experiment 2: share of project areas in which a predictor takes a single value, by predictor group and model. The number of variables per group is given in brackets.

The difference between the two models follows from the origin of the data. The Austrian lithology is derived from a national geological map, the European lithology from the pan-European dataset, and the European soil variables from a 1 km raster library. Given a median area size of 0.10 km², a single 1 km cell covers an entire project area.

3.9.3 Removing the differences between the project areas

After every predictor had been centred on the mean of its own project area, the mean standardised difference between springs and non-springs differs distinctly between the two models (figure 35). In the Austrian model it amounts to 0.51 for the satellite variables, 0.20 for the terrain variables, 0.17 for the lithology and 0.10 for the soil variables. In the European model,

the satellite variables and the terrain variables are on the same level at 0.30 each, whereas the lithology reaches only 0.04. The European soil variables could not be evaluated, since they are constant in every evaluable area, that is, in each of the 17 project areas that contain at least two springs and two pseudo-absences.

The same difference is apparent among the individual predictors. In the Austrian model, the five strongest single predictors are all satellite bands (S2 B8, S2 B3, S1 April–May, S1 February–March and S2 B4). In the European model, by contrast, three terrain variables are among the five strongest: S2 B8 (−0.47), profile curvature (−0.46), TWI (+0.45), S2 B3 (−0.43) and curvature (−0.43).

Consequently, the local separation within a project area is carried predominantly by the satellite variables in the Austrian model, whereas it is shared between the satellite and the terrain variables in the European model. The value of 0.04 for the European lithology means that this predictor contributes practically nothing to the separation within an area, which is consistent with its coarse resolution and with the high share of areas in which it is constant (section 3.9.2).

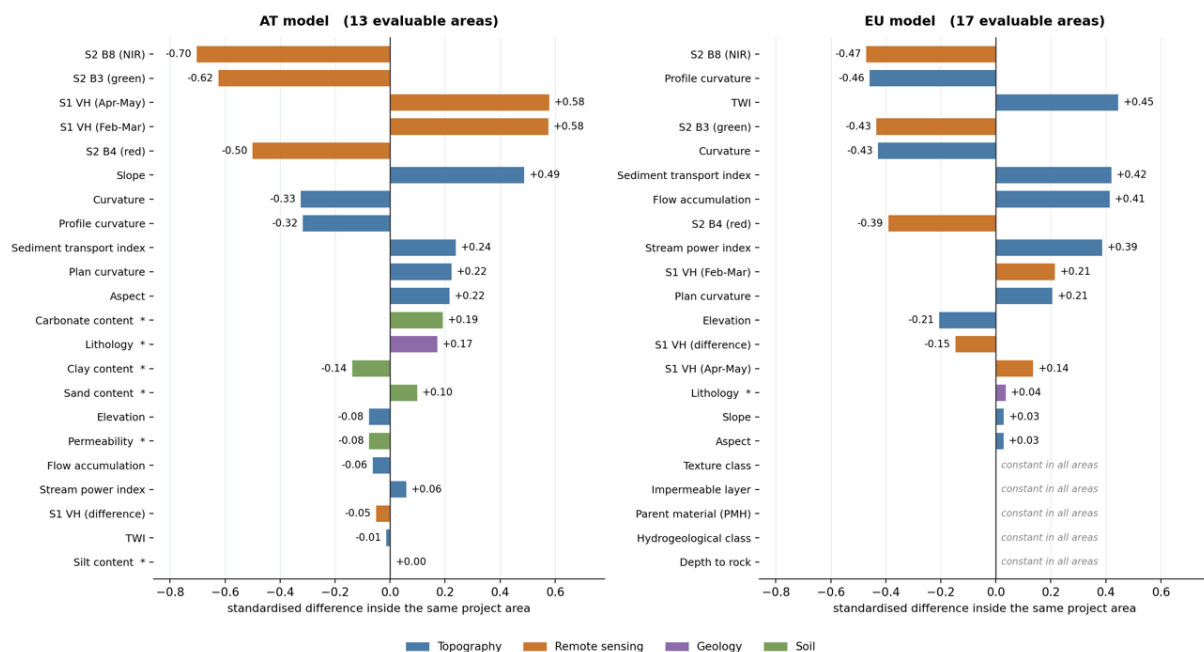


Figure 35 - Ablation experiment 3: standardised difference between springs and non-springs of the same project area, after the differences between the areas have been removed by centring each predictor on the mean of its own area. Positive values mean that springs have higher values than the non-springs of their own area, negative values the opposite. In the European model the five soil variables are constant in every evaluable area and could not be evaluated. Variables marked with an asterisk could be evaluated in fewer areas than stated in the panel title, because they are constant elsewhere. Only project areas with at least two springs and two pseudo-absences were used.

In the Austrian model, the group order differs from the pooled variable importance in section 3.3. The terrain variables remain in the upper range in both calculations, whereas the satellite variables move from third to first place and the soil variables from second to last. Both results are correct, since they measure different properties. The great majority of all spring-absence

pairs in the pooled calculation originate from two different project areas, so that the pooled values describe mainly what separates the areas from each other. Once this component has been removed, 16 of the 22 predictors still vary in every project area, whereas the remaining six vary only in part of the areas and were evaluated there.

The direction of the effects corresponds to the physical expectation. In both models, springs are darker in the near infrared band, which water absorbs most strongly. In the Austrian model, the cross-polarised radar backscatter is higher at springs, which indicates denser vegetation. Since the measure is based on means, the calculation was repeated with ranks; the ranking of the groups does not change.

4 Discussion

The Austrian model achieved higher performance metrics than the European model, which indicates that spring and non-spring locations could be distinguished more reliably with the more detailed national datasets. This difference is statistically robust, since the confidence intervals of the two models did not overlap over the 100 Monte Carlo realisations of the pseudo-absence sample. Although the performance metrics remained below those reported by Rahmati et al. (2018) and Chen et al. (2020), the present study relied on substantially smaller and more heterogeneous datasets and focused on the transferability of predictor variables between national and European data sources. Furthermore, predictor variables such as distance to faults and distance to streams, which were identified as important factors in previous studies (Corsini et al., 2009; Rahmati et al., 2018; Chen et al., 2020) were not included in the present models and might have improved their performance.

The two validation designs address two different questions, and the difference between them is the central result of this study. Under random five-fold validation, in which every project area is represented in the training data, both models reach useful values (AUC 0.797 for Austria and 0.741 for Europe). Under leave-one-area-out validation, in which a whole area is removed and subsequently predicted, the AUC falls to 0.715 and 0.632. A loss of 0.08 and 0.11 is moderate, and both models remain clearly above chance level. The threshold-based measures, however, decrease considerably more strongly. The MCC halves in both models, and the sensitivity drops from 0.67 to 0.31 and 0.44 respectively.

This difference is informative rather than contradictory. If a model were simply worse in an unknown area, sensitivity and specificity would decrease together. Instead, the specificity rises while the sensitivity falls (Austria 0.80 to 0.88, Europe 0.70 to 0.75). The predicted probabilities are compressed towards the middle, so that fewer points reach the fixed threshold of 0.5. The model therefore becomes undecided and it retains the ability to rank locations within the new area, even though the absolute probabilities are no longer comparable to those of the training areas. For practical application this means that a fixed probability threshold cannot be transferred to a new project area, whereas a relative ranking of the area's own cells can. The AUC is the only threshold-free measure reported here and therefore the only one that describes the transfer directly. The threshold-based measures should therefore be interpreted with caution whenever the maps are applied to an area that was not part of the training data.

The gap between the two designs also has a methodological consequence. Published spring potential models that are validated on randomly split points within the same areas report values

comparable to the random five-fold results of the present study. These values describe how well a model interpolates within known terrain and not how well it generalises to unknown terrain. Where the intended application is a new, unsurveyed area, this distinction is essential and a spatially blocked validation is the only design that indicates how the model will behave there. The comparison of the sampling designs showed how strongly the placement of the pseudo-absence points influences the apparent model quality. With the initial, spatially uncontrolled design, the European model seems to perform better than the Austrian model. Whereas with the controlled fixed-per-area design this order was reversed. Uncontrolled background sampling can therefore not only bias performance values but even reverse qualitative conclusions. This should be kept in mind when published studies are compared: very high performance values that were obtained with uncontrolled sampling and validation within the same area, such as the AUC of 0.94 reported by Sheikh et al. (2022), may partly reflect spatial autocorrelation rather than true predictive skill.

In addition, repeated model runs with identical data differed by up to 0.1 MCC. Single runs are therefore not a reliable basis for model comparison and central results of this study are reported as means over 100 realisations. The class imbalance likewise required attention. Without compensation for sparse categories, the sensitivity for the spring class dropped to 0.24–0.36, while the overall accuracy remained high because the models predicted mainly the majority class. With compensation, the sensitivity remained at 0.61–0.70 with a nearly unchanged MCC. Class compensation was therefore essential for a meaningful spring model.

The manual reclassification of lithological units into permeability classes represents a necessary simplification of complex geological conditions, as also demonstrated by Chen et al. (2020), who grouped lithological units despite using geological data at a scale of 1:10 000. Consequently, geological formations with transitional or mixed lithological characteristics cannot always be assigned unambiguously to a single permeability class. This limitation is particularly relevant for the Austrian dataset, which contains considerably more detailed geological information than the European dataset. In contrast, the European geological dataset is already highly generalised due to its scale. Accurate groundwater spring potential mapping benefits from detailed local geological information, since small-scale lithological differences may determine the location of spring outlets. The inclusion of structural geological variables such as distance to faults could have complemented the generalised lithological classification, as fault zones commonly increase bedrock permeability and provide preferential pathways for groundwater flow (Hölting and Coldewey, 2013). Nevertheless, lithology remained the strongest and most stable single predictor in both models, which is consistent with Chen et al. (2020), although the topographic variables formed the most important predictor group overall. In

addition, the relatively coarse spatial resolution of the geological datasets is likely to have reduced the spatial precision of the predictions. These datasets were nevertheless used because they provide consistent coverage across the entire study area. Previous studies often relied on considerably more detailed geological maps (Rahmati et al., 2018; Chen et al., 2020), whereas the present study used geological information at scales of 1:50 000 for the Austrian model and 1:1 000 000 for the European model. The coarser resolution may therefore have contributed to the reduced model performance.

The ablation experiments qualify how the importance of lithology should be read. Lithology was the strongest single predictor in the variable importance of both models, but once the differences between the project areas are removed, it separates springs from non-springs of the same area only weakly (+0.17 in the Austrian and +0.04 in the European model). The reason is that lithology takes a single value in 36% of the Austrian and in 85% of the European project areas, so that it cannot separate anything at all within these areas. Lithology therefore mainly indicates to the model which project area a point belongs to, and the project areas differ in their spring density. This is useful information for a model that covers several regions, but it is not the information required by a practitioner surveying a single site. The same applies even more strongly to the soil variables, which are constant in 66% of the Austrian and in 94% of the European areas and could not be evaluated at all in the European ablation. The importance values of the soil and geology variables should therefore not be read as evidence that these datasets locate spring outlets.

The Sentinel-1 variables used in this study were derived from seasonal VH backscatter and therefore reflect a combination of moisture-related and vegetation-related changes. The C-band wavelength of approximately 5.5 cm reaches only the uppermost centimetre of the soil, and only where the vegetation is low, for example on grassland or arable land (Tripathi and Tiwari, 2022; Rahmati et al., 2026). The project areas of this study are largely forested and are located in mountainous terrain, so that the backscatter describes the structure and water content of the vegetation layer rather than the moisture of the soil beneath it. Another recent study demonstrates the same limitation from the opposite direction (Massart et al., 2024). In this pan-European retrieval of surface soil moisture from Sentinel-1, pixels with dense vegetation have to be removed before the backscatter can be used at all. Dense vegetation is therefore not a disturbance of the signal but a boundary of the method (Massart et al., 2024). Consequently, the predictive value of the Sentinel-1 variables in this study should be read as an ecological response to persistent groundwater availability rather than as a direct detection of soil moisture. This interpretation is supported by the considerable variation in their importance

ranking across repeated model runs, which was larger than that of the topographic and geological predictors.

The high ranking of the Sentinel-2 variables should likewise be interpreted with caution. Since many verified spring locations are situated beneath forest canopy, the recorded spectral signal is unlikely to represent the spring outlet directly. Instead, Sentinel-2 probably captures vegetation characteristics, canopy conditions or land-cover patterns associated with persistent groundwater availability. Consequently, the predictive value of the Sentinel-2 variables may partly reflect indirect ecological responses rather than the direct detection of groundwater discharge. However, many trees at permanently wet sites are deciduous species, and at the acquisition times of the Sentinel data used here, namely late winter and early spring, they carry no leaf cover. The sensors can therefore partly capture the ground surface at exactly these sites, which may strengthen the moisture-related part of the signal. At the same time, the presence of deciduous stands can itself indicate persistent groundwater availability.

The exploratory combination of Sentinel-2 imagery with temporal Sentinel-1 VH information visually highlighted several verified spring locations, including sites located in forested areas where optical imagery alone is often limited by canopy cover. Although this observation was not evaluated quantitatively and was not part of the Random Forest analysis, it indicates that complementary information from optical and radar data may be useful for identifying groundwater-influenced environments. Future research should therefore investigate integrated multi-sensor approaches within time series and on additional bands, in particular the SWIR bands of Sentinel-2 and the L-band data expected from the upcoming NISAR and ROSE-L missions. Similarly, Chen et al. (2020) and Sheikh et al. (2022) included NDVI among the predictor variables of their models, which indicates that vegetation characteristics derived from multispectral imagery may provide useful information for groundwater spring potential mapping. In the present study, NDVI was not included in the analysis because a preliminary visual inspection did not identify a consistent association with verified groundwater spring locations. Moreover, NDVI received a considerably lower importance in these studies than the Sentinel variables reached in the present study, which supports the decision to use the spectral bands and the radar backscatter directly instead of a single vegetation index.

The ablation experiments show how strongly the optical signal dominates at the local scale. Once the differences between the project areas are removed, the near infrared band is the strongest single predictor in both models (-0.70 in the Austrian and -0.47 in the European model), and all three optical bands carry a negative sign (green -0.62 and -0.43 , red -0.50 and -0.39). Springs are therefore darker than the non-spring locations of their own area, in every

band and in both models. This is the most stable finding of the ablation analysis and one on which the two models, built from independent datasets, agree without qualification.

Low reflectance in these bands is produced by open water, by wet soil, by topographic and canopy shadow, and by dense vegetation. Only the first two are the target of this study. The qualitative expert review reported exactly the ambiguity that follows from this. Small forest patches and shaded ditches were assigned a high potential although they carry no water, whereas the correctly identified sites were mostly open and wet. The false positives are therefore a predictable consequence of the properties to which the models respond. The same property explains why the leaf-off acquisition dates support and limit the approach at the same time. Without leaf cover, the sensor can observe the ground at permanently wet sites, which strengthens the moisture-related part of the signal, but a dry shaded ditch is equally dark at the same time of year and cannot be distinguished on reflectance alone.

Taken together with the weak local contribution of the geological and soil variables, this leads to a clear statement about what the models do. They detect the surface expression of water and of features that resemble it, and not the hydrogeological setting that produces a spring. This is consistent with the field observation that a groundwater-fed spring is separated from temporary surface water by the physicochemical properties of the water, namely its electrical conductivity, its pH and its temperature, and not by any property visible from above. A predictor that captures the subsurface condition directly is missing from both models, and none of the datasets compared here provides such a predictor at the required spatial resolution.

In the present study, the definition of true absence locations was particularly challenging, because it was often not possible to verify with certainty that no additional undocumented springs existed within the project areas. Consequently, some pseudo-absence points may have represented unrecognised spring locations, which introduces an unavoidable source of uncertainty into the model training. The direct comparison of model performance between studies should likewise be interpreted with caution. Spring inventories are often incomplete, particularly with respect to small, intermittent or concealed spring outlets. Differences in inventory quality, sampling strategy and field verification procedures may therefore influence the reported model performance and limit the comparability of prediction metrics across studies.

Two further methodological limitations should be mentioned. First, the model configurations were optimised sequentially, so that interactions between design choices were not systematically evaluated. Second, although a leave-one-area-out validation was carried out in addition to the random splits, the number of project areas and the sampling size are still small, so that the transfer results rest on few independent units.

The different spatial resolutions of the elevation data (5 m and 30 m) influenced the derived topographic indices such as TWI, SPI and flow accumulation. The coarser 30 m resolution smooths local terrain features and small structures, which results in more generalised index values. In addition, the Austrian model was based on a Digital Elevation Model (DEM), whereas the European model relied on a Digital Surface Model (DSM). Since DSMs represent the elevation of the Earth's surface including vegetation and building structures, they are generally less suitable for hydrological analyses than a DEM. However, TWI and flow accumulation were less important and less stable in the Austrian model, although the Austrian elevation data have the finer resolution. At 5 m resolution, microtopography fragments the modelled flow paths, whereas the smoothed 30 m data produce more stable index values. Hydrological indices therefore appear to have an optimal scale, and a higher resolution is not automatically better. This is consistent with the observation that the influence of the elevation data resolution on the overall predictive performance is difficult to isolate. Previous studies reported strong predictive performance with elevation data of approximately 30 m resolution (Rahmati et al., 2018; Chen et al., 2020; Sheikh et al., 2022), which suggests that factors such as training data quality, predictor selection, geological setting and sampling strategy may have a greater influence on model performance than the elevation resolution alone.

The ablation experiments support this interpretation independently of the variable importance. Within a project area, the topographic wetness index, the stream power index and flow accumulation separate springs from non-springs clearly in the European model (+0.39 to +0.45), but not at all in the Austrian model (−0.06 to +0.06), whereas slope behaves in the opposite way (+0.49 in Austria and +0.03 in Europe). The two models therefore do not merely weight the same terrain signal differently, but they respond to different terrain properties.

Field investigations indicated that spring occurrence is influenced not only by the bedrock geology but also by the properties of the overlying soil, in particular its thickness and hydraulic conductivity. For this reason, soil-related predictor variables were included in the present study. However, both the Austrian and the European soil datasets have an effective spatial resolution of approximately 1 km, which is insufficient to represent the pronounced small-scale variability of soil conditions around individual spring outlets. Consequently, the predictive contribution of the soil variables may have been underestimated. Since both datasets share the same resolution, the considerably higher soil importance in the Austrian model (24.2% versus 11.5%) cannot be explained by resolution. It reflects the thematic quality of the variables: the Austrian data provide physical soil properties such as texture fractions, carbonate content and permeability, whereas several of the harmonised European soil variables carried almost no stable signal in the models. The Austrian eBOD dataset illustrates the potential of higher-

resolution soil information; its incomplete spatial coverage, however, prevented its consistent application across all Austrian project areas. Despite these limitations, the available datasets remain suitable for capturing broader regional differences in soil properties among the geographically dispersed project areas.

Two additional checks support the robustness of the models. First, the number of trees had no systematic effect on model performance (100 to 1 000 trees; the differences were smaller than the standard deviation of repeated runs), so that the 500 trees used in all models are not a critical setting. Second, the spring potential maps were compared with the mapped waterlogged areas from the field surveys, which had not been used for model training. These independently mapped wet locations received higher predicted probabilities than the model area on average in both regions (AT: 0.48 versus 0.40; EU: 0.45 versus 0.35). This supports the plausibility of the maps beyond the statistical validation.

Beyond the aggregated performance measures, the spring potential maps were evaluated in the way in which they would be used in practice, namely as a ranking of an area's own cells, produced by a model that has never seen this area. Of the springs in an unseen area, 20% (Austria) and 19% (Europe) fall within the highest ranked 10% of that area, 38% and 41% within the highest 20%, and 52% and 51% within the highest 30%. Compared with a random ranking of the same cells, this corresponds to a factor of approximately two across the whole range from 10% to 30% of the area. The two models, built from independent datasets, arrive at almost the same value.

The second evaluation addresses the distance from a spring to the nearest highly ranked cell, which is closer to the question asked by a field surveyor. Under leave-one-area-out prediction, the median distance from a spring to the nearest cell in the top 5% of its area is 22 m in the Austrian and 32 m in the European model, compared with 41 m and 67 m for a randomly placed point in the same area. At the top 10% level, the values are 14 m against 30 m and 20 m against 42 m. Again, both models reach a factor of approximately two.

A factor of two is a modest effect and clearly below what the same models achieve when the area is part of the training data, where 56% (Austria) and 64% (Europe) of the springs fall within the highest ranked 5%. The difference between these two sets of values is the price of transfer to an unsurveyed area, and it is considerable. At the same time, a factor of two is not without value in this application. Halving the area that has to be walked in order to find the same number of springs is a relevant saving in a field survey, particularly because the method requires no site visit beforehand and can be applied to any area covered by the input datasets. The maps should therefore be understood as a means of ordering field work and not as a substitute for it, and the highly ranked cells as candidates that still require verification on site. If verified spring

locations are available, the area can be included in the training data and the prediction can be extended to the surrounding terrain, so that the model operates under conditions it has already seen.

Longer radar wavelengths are a promising direction for further work, because they penetrate the vegetation layer that limits the C-band. This direction will become more accessible in the near future: Rahmati et al. (2026) name the upcoming NISAR and ROSE-L missions as open data sources for soil moisture monitoring, with ROSE-L planned as a European mission with full coverage. A repetition of this study with such data would allow a direct test of whether the limitation identified here can be overcome.

A second point concerns the way in which the radar data were used. The radar predictors of this study are two seasonal composites and their difference. In the within-area analysis, the difference of the two composites showed almost no local separation, although each composite on its own did. Averaging over a period retains the signal, whereas subtracting one period from the other removes it. A time series with several composites per year would instead describe how persistent the signal is at a location and would preserve this information.

Future studies may benefit from incorporating higher-resolution soil information together with additional hydrogeological predictor variables, such as distance to faults and distance to streams (Rahmati et al., 2018; Chen et al., 2020; Sheikh et al., 2022). In addition, preprocessing approaches such as Frequency Ratio weighting, as applied by Chen et al. (2020), may further improve model performance by enhancing the representation of relationships between predictor variables and spring occurrence. Further studies could also aim to distinguish waterlogged areas caused by temporary surface water from true groundwater spring outlets. Field investigations showed that electrical conductivity, pH and water temperature are the parameters used for this differentiation, and that the temperature is particularly informative, since groundwater-fed springs maintain relatively stable temperatures throughout the year. Targeted measurements of these parameters, or thermal remote sensing where feasible, could therefore complement the presented approach.

5 Conclusion

This study demonstrated that Random Forest models based on geological, topographic, soil and remote sensing variables can identify potential groundwater spring locations. The Austrian model achieved a higher predictive performance than the European model, which highlights the importance of high-quality regional datasets and at the same time demonstrates the general transferability of the methodological approach.

The topographic variables formed the most important predictor group in both regions, whereas lithology was the strongest and most stable single predictor. The Sentinel-1 and Sentinel-2 variables provide complementary information, most probably through indirect ecological responses associated with persistent groundwater availability. The generalisation of the geological data limits the models mainly within the project areas, since lithology is constant in a large share of them and can therefore separate the areas from each other but not the locations within an individual area. The difference in soil importance between the two regions reflects the thematic data quality rather than the spatial resolution. In addition, the models assigned an increased spring potential to independently mapped waterlogged areas that had not been used for model training, which supports the plausibility of the resulting maps.

The results of this study demonstrate that groundwater spring potential modelling is feasible using geographically dispersed local datasets. Before an application at a larger scale, however, the models should be improved with a larger number of training points and with remote sensing variables that address soil moisture more directly. This suggests that, given harmonised geological and environmental datasets, a European-scale groundwater spring potential map could be developed. Such a product could support the identification of groundwater-dependent areas and contribute to water resource management under the increasing drought conditions associated with climate change. However, such an approach would require harmonised geological datasets with sufficient spatial detail across Europe and a considerably larger number of sampling size and sampling areas. Moreover, an Austrian spring potential map could provide a useful basis for the designation of groundwater protection zones, since it indicates at an early stage which parts of an area are likely to contain groundwater discharge points and where field investigations should therefore be concentrated.

The combined spring inventories of the present study together with those published by Corsini et al. (2009), Rahmati et al. (2018), Chen et al. (2020), and Sheikh et al. (2022) may provide a valuable foundation for future deep learning approaches. If harmonized geological classifications and consistent predictor datasets were available, convolutional neural networks

could potentially learn complex spatial patterns associated with spring occurrence directly from the surrounding terrain and environmental context, rather than relying solely on predefined topographic indices. Such an approach would more closely resemble the visual interpretation commonly applied by hydrogeologists, who identify potential spring locations by evaluating the spatial configuration of landforms, geological boundaries, and surrounding terrain rather than considering individual predictor variables in isolation.

References

- Adler, R.F., Gu, G., Sapiano, M., Wang, J.-J., Huffman, G.J., 2017. Global Precipitation: Means, Variations and Trends During the Satellite Era (1979–2014). *Surv Geophys* 38 (4), 679–699. 10.1007/s10712-017-9416-4.
- Arabameri, A., Roy, J., Saha, S., Blaschke, T., Ghorbanzadeh, O., Tien Bui, D., 2019. Application of Probabilistic and Machine Learning Models for Groundwater Potentiality Mapping in Damghan Sedimentary Plain, Iran. *Remote Sensing* 11 (24), 3015. 10.3390/rs11243015.
- Breiman, L., 2001. Random Forests. *MACH LEARN* 45 (DOI: 10.1023/A:1010933404324), 32.
- Chen, W., Li, Y., Tsangaratos, P., Shahabi, H., Ilia, I., Xue, W., Bian, H., 2020. Groundwater Spring Potential Mapping Using Artificial Intelligence Approach Based on Kernel Logistic Regression, Random Forest, and Alternating Decision Tree Models. *Applied Sciences* 10 (2), 425. 10.3390/app10020425.
- Chen, Z., Zhao, S., 2022. Automatic monitoring of surface water dynamics using Sentinel-1 and Sentinel-2 data with Google Earth Engine. *International Journal of Applied Earth Observation and Geoinformation* 113, 103010. 10.1016/j.jag.2022.103010.
- Corsini, A., Cervi, F., Ronchetti, F., 2009. Weight of evidence and artificial neural networks for potential groundwater spring mapping: an application to the Mt. Modino area (Northern Apennines, Italy). *Geomorphology* 111 (1-2), 79–87. 10.1016/j.geomorph.2008.03.015.
- Delpero, A., Volpato, G., 2022. Integrated pond aquaculture and regional identity: ethnobiology of the golden humped tench of Poirino highlands, Northwest Italy. *Journal of ethnobiology and ethnomedicine* 18 (1), 31. 10.1186/s13002-022-00529-5.
- Duscher, K., Günther, A., Richts, A., Clos, P., Philipp, U., Struckmeier, W., 2015. The GIS layers of the “International Hydrogeological Map of Europe 1:1,500,000” in a vector format. *Hydrogeol J* 23 (8), 1867–1875. 10.1007/s10040-015-1296-4.
- Hölting, B., Coldewey, W.G., 2013. *Hydrogeologie: Einführung in die allgemeine und angewandte Hydrogeologie*, 8th ed. Springer Spektrum, Berlin, Heidelberg, 438 pp.
- Krueger, E.S., Ochsner, T.E., Engle, D.M., Carlson, J.D., Twidwell, D., Fuhlendorf, S.D., 2015. Soil Moisture Affects Growing-Season Wildfire Size in the Southern Great Plains. *Soil Science Society of America Journal* 79 (6), 1567–1576. 10.2136/sssaj2015.01.0041.
- Li, Z., Zhang, X., Xiao, P., 2022. Spectral index-driven FCN model training for water extraction from multispectral imagery. *ISPRS Journal of Photogrammetry & Remote Sensing* 192, 361. 10.1016/j.isprsjprs.2022.08.019.
- Massart, S., Vreugdenhil, M., Bauer-Marschallinger, B., Navacchi, C., Raml, B., Wagner, W., 2024. Mitigating the impact of dense vegetation on the Sentinel-1 surface soil moisture retrievals over Europe. *European Journal of Remote Sensing* 57 (1). 10.1080/22797254.2023.2300985.
- Moreira, F., Viedma, O., Arianoutsou, M., Curt, T., Koutsias, N., Rigolot, E., Barbati, A., Corona, P., Vaz, P., Xanthopoulos, G., Mouillot, F., Bilgili, E., 2011. Landscape--wildfire interactions in southern Europe: implications for landscape management. *Journal of environmental management* 92 (10), 2389–2402. 10.1016/j.jenvman.2011.06.028.
- Panagos, P., van Liedekerke, M., Borrelli, P., Köninger, J., Ballabio, C., Orgiazzi, A., Lugato, E., Liakos, L., Hervas, J., Jones, A., Montanarella, L., 2022. European Soil Data Centre 2.0: Soil data and knowledge in support of the EU policies. *European Journal of Soil Science* (73(6)).
- Rahmati, M., Balenzano, A., Bechtold, M., Brocca, L., Fluhrer, A., Jagdhuber, T., Karamvassilis, K., Mengen, D., Reichle, R.H., Kim, S., Taghizadeh-Mehrjardi, R., Walker, J., Zhu, L., Montzka, C., 2026. Soil moisture retrieval from Sentinel-1: Lessons learned after more than a decade in orbit. *Remote sensing of environment* 333, 115146. 10.1016/j.rse.2025.115146.

- Rahmati, O., Naghibi, S.A., Shahabi, H., Bui, D.T., Pradhan, B., Azareh, A., Rafiei-Sardooi, E., Samani, A.N., Melesse, A.M., 2018. Groundwater spring potential modelling: Comparing the capability and robustness of three different modeling approaches. *Journal of Hydrology* 565, 248–261. 10.1016/j.jhydrol.2018.08.027.
- Ranjan, P., Pandey, P.K., Pandey, V., Lepcha, P.T., 2021. Development of seepage spring for rural water security. *IOP Conf. Ser.: Mater. Sci. Eng.* 1070 (1), 12047. 10.1088/1757-899X/1070/1/012047.
- Santos, F., Nicasio, K., Silva, K., Martins, J., Périco, E., Dalzochio, M., Veras, D., Cajaiba, R.L., 2021. Can artificial ponds retain dragonfly (Insecta: Odonata) biodiversity? A preliminary study in the Brazilian Amazon. *Austral Entomology* 60, 707. 10.1111/aen.12574.
- Sheikh, H.A., Bhat, M.S., Alam, A., Ahsan, S., Shah, B., 2022. Assessing the groundwater spring potential of Sindh basin in the Kashmir Himalaya. *Arab J Geosci* 15 (23). 10.1007/s12517-022-10965-y.
- Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., Zeileis, A., 2008. Conditional variable importance for random forests. *BMC bioinformatics* 9, 307. 10.1186/1471-2105-9-307.
- Tripathi, A., Tiwari, R.K., 2022. Synergetic utilization of sentinel-1 SAR and sentinel-2 optical remote sensing data for surface soil moisture estimation for Rupnagar, Punjab, India. *Geocarto International* 37 (8), 2215–2236. 10.1080/10106049.2020.1815865.

Data References

Bundesamt für Eich- und Vermessungswesen (BEV), 2023. Digital Terrain Model (DTM) of Austria, 5 m resolution. Available at: <https://www.bev.gv.at> (accessed October 2025).

Bundesforschungs- und Ausbildungszentrum für Wald, Naturgefahren und Landschaft (BFW), 2023. eBOD – Digitale Bodenkarte Österreichs. Available at: <https://bodenkarte.at> (accessed 3 January 2026).

European Commission, Joint Research Centre (JRC), 2004. The European Soil Database distribution version 2.0. European Commission and the European Soil Bureau Network, EUR 19945 EN. Distributed through the European Soil Data Centre (ESDAC). Available at: <https://esdac.jrc.ec.europa.eu> (accessed 30 November 2025).

European Geological Data Infrastructure (EGDI), 2023. EGDI 1:1 Million Pan-European Surface Geology [Geospatial dataset]. EuroGeoSurveys / GEUS. Available at: <https://www.europe-geology.eu> (accessed 10 October 2025).

European Space Agency (ESA), 2021. Copernicus Global and European Digital Elevation Model (COP-DEM). Available at: <https://doi.org/10.5270/ESA-c5d3d65> (accessed October 2025).

European Space Agency (ESA), 2024. Copernicus Sentinel-1 data. Processed by ESA, accessed via Google Earth Engine (accessed May 2025).

European Space Agency (ESA), 2024. Copernicus Sentinel-2 MSI Level-2A data. Processed by ESA, accessed via Google Earth Engine (accessed November 2025).

GeoSphere Austria, 2025. Geological Units 1:50 000 (GeoFAST) [Geospatial dataset]. Available at: <https://maps.geosphere.at> (accessed 3 December 2025).