Universität Zürich[UZH]

Geographisches Institut

Factors Predicting the Occurrence of Ecologically Valuable Springs: A Random Forest Modelling Approach

GEO 511 Master's Thesis

Author: Nicholas Roman von Holzen, 19-052-471

Supervised by: Dr. Ilja van Meerveld, Dr. Maria Staudinger

Faculty representative: Prof. Dr. Jan Seibert

26.08.2025

Abstract

Spring habitats are among the most threatened aquatic ecosystems in Switzerland. Effective protection requires assessing their ecological value (e.g., structural and faunistic condition), yet the existing FOEN method is resource-intensive and expensive. A modelling approach could simplify this assessment, thereby facilitating protection efforts. Previous research has largely focused on modelling spring occurrence and its predictors, rather than ecological value. This highlights the need for a new approach to modelling the ecological quality of springs. The aim was to determine whether structural parameters from the FOEN method can predict faunistic quality, and whether landscape and topographic variables can predict the structural and faunistic quality of springs.

An exploratory data analysis was performed on more than 750 Swiss springs structurally and faunistically assessed using the FOEN method. Additional potential predictor variables were generated in QGIS from publicly available landscape and topographic datasets. Various Random Forest regression and classification models were implemented in R. Faunistic results were modelled using respective structural data, while both structural and faunistic results were modelled using landscape and topographic variables. Key predictors were identified through permutation-based variable importance measures.

When modelling faunistic results using structure data, regression models showed similar performance, with R^2 values never exceeding 0.25, indicating only weak predictive tendencies. Classification models reached a maximum accuracy of 0.43 with lower balanced accuracy, reflecting poor prediction of the lowest fauna classes. Therefore, accurate prediction of the faunistic condition from the structural condition was not possible. Modelling structural and faunistic results with landscape and topographic data yielded better results for structural conditions, with the best regression model achieving an R^2 of 0.25, while the faunistic modelling did not exceed 0.15. Classification accuracy was substantially higher for structural modelling than for faunistic modelling, though balanced accuracy was similar. Overall, accurate prediction of ecologically valuable spring occurrence from landscape and topographic variables was not feasible. Across datasets, key predictor variables included Elevation, Number of Structures, Land Use, and Lithology, among others.

This study demonstrates the application of Random Forest models with field-derived data to assess spring habitats. The findings indicate that successful modelling of the faunistic quality requires additional structural parameters not currently included in the FOEN method, highlighting the need for a methodological expansion. Limitations include imbalanced class distributions dominated by highly rated springs. Future research should identify critical structural predictors that define the ecological value of springs and explore alternative modelling approaches, such as hybrid models or neural networks, to improve predictive performance.

Acknowledgements

This work would not have been possible without the support and contribution of several people, for which I would like to express my gratitude:

For the inspiring meetings, the shared ideas and tips, and the guidance in general:
Dr. Ilja van Meerveld & Dr. Maria Staudinger

For the valuable professional exchange, suggestions, interest, and organisational support throughout the entire duration of my thesis:
Dr. Daniel Küry & Pascal Stucki

For the provision of data:
Maxime Chèvre from the national data and information centre on Swiss fauna *info fauna*, as well as the owners of the spring data from various cantonal authorities.

To everyone who supported me in any way during the process of this thesis:
Jahel Hämmerli, Jonas Laube, Elena Santi, friends and family

I also acknowledge that I used several different OpenAI models as a support in the process of writing the R codes and the QGis Python codes. I also used DeepL, OpenAI models and the built-in Overleaf language tool Writefull to spell-check and improve the formulation of texts. However, this work and the results presented here are not the direct result of artificial intelligence tools.

Contents

List of Figures

List of Tables

1. Introduction

1.1 Definition and Characteristics of Springs

Springs are discrete points of groundwater discharge at the surface of the Earth (BAFU 2021, Kamp 1995) and often occur where an impermeable layer intersects the surface (Stucki et al. 2015). Springs as groundwater emergence sites are locally limited and must at least temporarily lead to discharge (DIN4049-3 as cited in Küry et al. (2019)). Springs are the only locations in the landscape where groundwater flows without mixing with surface water (Cantonati et al. 2022). The point where groundwater emerges at the surface and the areas influenced by this water form the spring habitat. Thus, springs are groundwater-dependent ecotones that connect surface and underground water bodies (BAFU 2021, Cantonati & Ortler 1998).

Springs are characterised by three different zones: the groundwater reservoir, the eucrenon (i.e., the spring area) and the hypocrenon (i.e., spring brooks running down from the outlet) (Illies & Botosaneanu 1963 as cited in Fernández-Martínez et al. (2024)). The eucrenon stretches from the groundwater exit point to the beginning of the spring stream. Since the spring stream has similar characteristics, it is also considered part of the spring habitat. Endorheic springs do not form a longer spring brook but rather seep away after a relatively short distance (Küry et al. 2019, BAFU 2021).

Spring ecosystems are usually small in size and differ significantly from streams, rivers, ponds and lakes in terms of water quality, substrate composition, adjacent terrain, bedrock geology, vegetation, and climatic conditions. Due to the interplay of hydrogeology, groundwater, soils, climatology, hydrology, water chemistry, biology, and possible human impacts, spring ecosystems can be considered complex systems (Fernández-Martínez et al. 2024, Cantonati et al. 2022).

If several outlets of the same type are found at one location, this is referred to as a spring system. If a spring consists of multiple types that create a common outflow, it is called a spring complex (Stucki et al. 2015). In addition, wandering springs are defined as those whose outlet migrates upward or downward, depending on the groundwater supply (Imesch & Küry 2024). Furthermore, a spring outlet can be artificial due to human alterations (Küry et al. 2019).

1.1.1 Spring Classification

Springs can be classified in different ways. The first spring classification system distinguished three main types according to hydromorphological criteria (Steinmann et al. 1915 & Thienemann 1926 as cited in Fernández-Martínez et al. (2024)): (1) rheocrenes or flowing springs with water feeding small streams; (2) limnocrenes or pool springs with lentic habitats; and (3) helocrenes or seepages with shallow or soggy wet zones. A clear distinction of these three main spring types is not always possible, leading to multiple mixed forms. The original typology by Steinmann et al., 1915 & Thienemann 1926 is still applied in Switzerland today, although in some parts of Switzerland it is divided more finely on a regional basis (Zollhöfer 1997 as cited in Küry et al. (2019)). In other countries, some extensions have been introduced, such as distinguishing springs according to geology, hydrology, water chemistry, water temperature, ecology, and human use (Fernández-Martínez et al. 2024).

Another approach is to subdivide springs according to their subterranean water pathways. This allows for differentiation between joint springs, karst springs, and loose rock springs (Hölting & Coldeway 2013 as cited in Küry et al. (2019)). Springs can also be classified according to the vegetation that occurs, which in turn depends on the geological substrate. Thus, warm-loving spring vegetation can be distinguished (Adiantion), spring habitats with hard, calcareous water (Cratoneurion), those with soft, low-calcium water (Cardamino-Montion), floodplain springs (Giessen), and the type known as "surplus-irrigated surfaces, springs without vegetation" (Delarze et al. 2016). For springs in the Alps, there is a supplementary typology also based on phytosociological surveys (Geissler 1976).

1.1.2 Spring Hydrology

Hydrologically, springs are characterised by stable temperatures, more or less constant cold-water discharge, and low oxygen saturation. The temperature of a spring is usually the same as that of groundwater and is roughly equivalent to the average annual temperature (or recharge temperature) of the respective location. Springs are therefore cool in summer and warm in winter (Stucki et al. 2015). The chemistry of spring water is determined by aquifer lithology, climate characteristics, vegetation, and soil properties. The concentration of ions in the water depends on the solubility of bedrock minerals and the residence time in the aquifer. With higher water temperatures, reactivity and weathering increase, influencing spring water chemistry. Precipitation amounts also affect water chemistry, although accelerated chemical reactions and dilution ultimately balance each other out with increased precipitation (Fernández-Martínez et al. 2024).

The vegetation and soil properties in the catchment area influence spring water chemistry by affecting infiltration, organic matter enrichment, and buffering (Fernández-Martínez et al. 2024). Spring water is often low in oxygen because, as groundwater, it remains underground for extended periods (Stucki et al. 2015). The climate, hydrogeology, and lithology of the aquifers determine whether a spring has permanent or intermittent flow (Cantonati et al. 2020a). Depending on temporal behaviour, springs that only flow intermittently are referred to as «Hungerquellen» (Stucki et al. 2015). The discharge of permanent springs varies significantly in different lithologies, for instance with larger drainage flows in karst systems compared to granite systems (Fernández-Martínez et al. 2024).

1.1.3 Flora and Fauna of Springs

The spring habitat is an ecotone at the interface between groundwater and surface water (Cantonati & Ortler 1998). The springs contain closely interconnected communities of groundwater, surface waters, and moist zones (the hygropetric zone) (BAFU 2021). The springs offer favourable conditions for highly specialised species due to their low temporal variability of environmental conditions and their particularly structure rich habitat (Peterka et al. 2023).

Organisms living in the actual source of spring water are called crenobionts. These highly specialised spring dwelling organisms have adapted to spring habitats to survive and complete their life cycles (Fernández-Martínez et al. 2024). A large number of crenobiont species are national priority species (BAFU 2021). Species associated with a spring during certain phases of their life cycles are referred to as crenophiles. Crenoxens may also routinely benefit from functions provided by springs but are not dependent on them. Stygobionts are classified as inhabitants of the groundwater ecosystem, while stygophiles mainly inhabit the spring sink. Rheophiles inhabit the spring downstream waters. During drought or flood periods, this water continuum from aquifer to stream may be interrupted (Fernández-Martínez et al. 2024). Typical spring inhabitants, such as certain species of stoneflies, caddisflies, and mayflies, are highly specialised and therefore sensitive to disturbances of their habitats (Stucki et al. 2015).

Aquatic spring species are cold-water specialists and gradually vanish downstream as water temperatures increase. Among these are glacial relicts that have survived only in Scandinavia and the Alps since the last ice age. Depending on the type of spring, different groups or species dominate. In karst springs, typical spring species regularly occur alongside groundwater species. In hillside fens and spring brooks, one may find damselfly larvae, various stonefly genera, and caddisfly larvae. In calcareaous sinter springs, amphipods often dominate, as well as stoneflies and sometimes fire salamander larvae. In seepage springs, one may find, among others, the caddisfly larva *Crunoecia irrorata* (Stucki et al. 2015).

The spring and its springbrook habitat are called the krenal, and its community is called the krenon (Stucki et al. 2015). The biodiversity of spring ecosystems is very high due to the convergence of species of various origins and is strongly related to the hydrogeological characteristics of the aquifer (Cantonati et al. 2020b). Spring habitats with similar environmental conditions often have very different species compositions, particularly in the mountains (Cantonati et al. 2006). However, the species richness of springs that occasionally dry out is always lower. They are occupied by specialist species that can withstand desiccation or bridge it with resistant stages (Stucki et al. 2015).

Springs can act as refuges for biodiversity during drought periods when the fluvial network is dry (Fernández-Martínez et al. 2024), or they can serve as refuges for relict species (Cantonati et al. 2012, Taxböck et al. 2017). Even minor disturbances can result in complete loss of habitat. Spring habitats that have developed over millennia can scarcely be restored once lost. Approximately 500 species in Europe rely on spring habitats (LfU 2023).

1.2 Ecological Importance and Decline of Springs

Spring habitats are considered hotspots of species diversity. Many animal species are highly adapted to spring habitats and occur only there. The springs therefore contain a high proportion of threatened species (Imesch & Küry 2024, BAFU 2021). Especially in the Alps, springs make a significant contribution to regional biodiversity (Seiler et al. 2021).

1.2.1 Decline of Spring Habitats in Switzerland

In Switzerland, the density of springs in the water-poor Jura was less than two springs per square kilometre in the near-natural state. In contrast, there were about 20 springs per square kilometre in the water-rich Central Plateau. In some parts of the Alps, this figure was even higher (BAFU 2021). However, spring habitats have declined dramatically in the past 200 years (Küry et al. 2019). Already in 1880, more than half of the springs in the Central Plateau were tapped, and the spring density there halved between 1880 and 1997 (Zollhöfer 1997 as cited in Lubini-Ferlin et al. (2014)). The decrease was most pronounced in settlement areas, followed by intensively used agricultural land at lower elevations (Küry et al. 2019, BAFU 2021). Seiler et al. (2021) assumed that by 1999, 95% of the springs in the Central Plateau had been tapped or otherwise impaired. Alluvial springs in floodplains and endorheic springs have also declined significantly (Stucki et al. 2015).

Spring habitats are among the most threatened aquatic ecosystems and should be considered worthy of protection, requiring preservation and, where necessary, enhancement (Cantonati et al. 2006, BAFU 2021). Today, most near-natural springs are located in forests, but they are often small and have only a modest discharge (Küry 2009, Stucki et al. 2015, Imesch & Küry 2024). Most of the untapped springs in open areas lie in grazing pastures (BAFU 2021). Springs in open landscapes can still be found frequently in the hills, as well as in the higher elevations of the Alps (BAFU 2021, Imesch & Küry 2024). Springs in open landscapes, especially pool springs and marshy or seepage springs, are currently valued highest in the Central Plateau.

1.2.2 Explanations for the Decline of Spring Habitats

The reasons for the decline in spring habitats are varied and, in some cases, multilayered. For some springs, one factor may be a sufficient explanation, while for others, it is the interplay of different factors that contribute to impairment or disappearance. An important factor in the decline of springs is the use of springs for drinking water, especially near settlements, but also in forests (BAFU 2021, LfU 2023). Even today, many areas of Switzerland and particularly the Alps see new springs being tapped for the supply of drinking water (Stucki et al. 2015). Currently, around 40% of Switzerland's drinking water comes from springs (Lubini-Ferlin et al. 2014, BAFU 2021, Stucki et al. 2015). When a spring tap is decommissioned, the structures are often not removed but left to decay (LfU 2023, Stucki et al. 2015).

The intensification of agriculture since the late nineteenth century has also contributed significantly to the loss of spring habitats. On the one hand, this can be explained by the intensification of the use of meadowlands and the drainage of seepage springs and flowing springs, as well as the drainage of entire wetlands and meadow landscapes (BAFU 2021, Lubini-Ferlin et al. 2014, Stucki et al. 2015, LfU 2023). On the other hand, runoff from agricultural land containing nutrients and pesticides into spring areas plays a major role in the destruction of spring habitats (BAFU 2021, Fernández-Martínez et al. 2024, LfU 2023). In overfertilised springs, plants grow exceptionally strongly and quickly cover the water surface, thereby restricting competitors in their development and preventing insects from reaching the water surface (Stucki et al. 2015). In addition, damage to springs in open landscapes from trampling is widespread. Livestock on pastures use springs as a watering place, significantly affecting spring habitats through the physical impact of their hooves and the addition of excrements as fertiliser (BAFU 2021, Stucki et al. 2015, Imesch & Küry 2024). Intensive recreational use can also locally damage spring habitats through trampling (BAFU 2021).

During the construction of infrastructure such as roads, particularly in forests, springs have been and still are being tapped, diverted, or channelled (BAFU 2021, Stucki et al. 2015). The same is the case for springs in open areas that were destroyed by expansion of settlement areas and developments (BAFU 2021). Various types of forest management have also contributed to the decline in spring habitats. Springs are very sensitive to the deposition of mown material or woody debris in wild dumps (Stucki et al. 2015). Moreover, spruce trees planted in locations where they do not naturally belong affect forest springs due to their dense crowns that provide shade and acidification caused by needle litter (LfU 2023). The list of factors impacting spring habitats also includes the tapping of springs for energy generation through small power plants, for cooling, irrigation, snowmaking systems, Kneipp pools, livestock watering, etc. (BAFU 2021, Lubini-Ferlin et al. 2014). Because spring organisms are so closely bound to low and constant temperatures, climate change poses a major overarching threat to spring communities (BAFU 2021, Stucki et al. 2015).

1.2.3 Effects of the Loss of Spring Habitats

Because of habitat loss, spring specialists are particularly under threat (BAFU 2021). In Switzerland alone, among molluscs, mayflies, stoneflies, caddisflies, and dragonflies, 96 species with a strong dependence on spring habitats have been found, 40% of which are classified as endangered, while many others are considered potentially endangered (BAFU 2021, Stucki et al. 2015). Not only is numerical reduction in habitats having a negative impact; the fragmentation of the remaining spring habitats also exerts a negative influence (Lubini-Ferlin et al. 2014). If the density of springs in a landscape area decreases, habitat connectivity is lost and species isolation increases (Imesch & Küry 2024, Lubini-Ferlin et al. 2014). Moreover, the role of spring habitats as refuges for Ice Age relicts is jeopardised by the loss of springs (Imesch & Küry 2024).

1.3 Protection of Spring Habitats in Switzerland

1.3.1 Legal Basis in Switzerland

Springs are considered bodies of water under the Federal Water Protection Act, and the protection provisions within this legislation apply to springs and spring habitats, regardless of whether the spring is public or private. In terms of water quality, as with other bodies of water, it is prohibited to contaminate springs, and the general duty of care applies here as well (Imesch & Küry 2024). With regard to the quantity of water, the springs are indirectly protected through provisions that ensure adequate residual flow (Lubini-Ferlin et al. 2014). However, there is no formal protective status for spring habitats in Switzerland. As a result, habitat protection is provided through the intervention regime of the Nature and Cultural Heritage Act, but only on the condition that the spring and its surroundings constitute a habitat worthy of protection (BAFU 2021). Various criteria are used to determine whether a habitat is worthy of protection: on the one hand, the presence of bankside areas or bogs; on the other, the occurrence of certain types of habitat, which are almost without exception assigned to nationally priority habitats (Küry et al. 2019). Another criterion for being worthy of protection is the presence of Red-List species, or nationally priority species (BAFU 2021).

The Nature and Cultural Heritage Act additionally contains a general clause enabling the designation of other sites that are particularly worthy of protection. However, to satisfy this clause, certain requirements regarding size and ecological quality must be met. Thus, an ecological assessment of the spring habitat is necessary for each individual case to assess whether it is worthy of protection (Lubini-Ferlin et al. 2014).

1.3.2 FOEN Method for Spring Habitat Assessment

The Federal Office for the Environment (FOEN) mandated the development of a method for a standardised survey and assessment of spring habitats. This was carried out in the work by Lubini-Ferlin et al. (2014). The assessment is based on structure (impairments, vegetation, and structural diversity) and the ecological dependence of the fauna on the springs. Following the procedure in the "Modulstufenkonzept", five status classes are distinguished. The springs in classes 1 and 2 of the structural assessment are considered significant spring habitats and are selected for faunistic assessment (Küry et al. 2019). A significant spring habitat is characterised by the following properties: type-specific expression of spring structures and vegetation, absence of major anthropogenic influences, characteristic biodiversity for the respective type of spring outlet, a spring-specific composition of the invertebrate fauna with a high proportion of spring-dependent species, and the presence of Red-List Species and Priority Species (Küry et al. 2019).

The FOEN field worksheets used to evaluate the springs in this thesis dataset are available in Appendices *A1 Spring Protocol Structure* and *A2 Spring Protocol Fauna*. These two complementary methods for the faunistic-ecological assessment of springs were originally developed in Germany, then tested in Switzerland, and slightly adapted for high-alpine areas. Both methods are intended to be carried out in parallel, as structurally impaired springs can also be inhabited by a remnant biocenosis (Lubini-Ferlin et al. 2014). The structural assessment consists of mapping habitat structures and impairments, based on the procedure by Schindler 2004 as cited in Küry et al. (2019). With relatively little effort, the structural survey presented in Lubini-Ferlin et al. (2014) provides a fairly precise picture of the state of each spring by evaluating its visible conditions. The header data at the top of the structural assessment protocol have no influence on the assessment of the spring. Assessment Part A on the left side of the protocol deals with impairments and yields Value A (impairment). Value A corresponds to the highest value of all observed parameters. Assessment part B on the right side addresses vegetation-use-structure and yields Value B. If a high number of substrates, number of flow types, and number of structures is found, a bonus (b) of 0.4 is automatically subtracted from the final assessment value. The overall result of the structural assessment is consequently derived from:

$$\text{Final Result} = \frac{A+B}{2} - b$$

The lower the resulting value, the better the structural condition of the surveyed spring. Based on this value, the spring is placed into one of five structure categories: *natural*, *conditionally natural*, *moderately impaired*, *damaged*, and *strongly damaged*.

The faunistic assessment presented in Lubini-Ferlin et al. (2014) consists of a faunistic sampling procedure following Schindler (2004). The animals are collected using a net or a sieve, sieved out and transferred to a white lab tray. Sampling is stopped only when no new taxon has been found for approximately 10 minutes. After identifying the animals in the laboratory, preferably at species level, the assessment procedure by Fischer (1996) is applied. It is based on the spring dependence of an organism, i.e., the stenotypy of the taxon in relation to the habitat and its frequency. The more strongly a species depends on springs, the higher its ecological score (ÖWZ). The final result of the faunistic survey is calculated by multiplying the ÖWZ of a taxon found by its abundance level, depending on the total number of taxa found:

$$\text{Ecological Value Sum (ÖWS)} = \frac{\sum(\text{ÖWZ} \cdot \text{Abundance})}{\text{Number of Taxa}}$$

The higher the resulting value, the better the faunistic condition of the surveyed spring. Based on this value, the spring is then placed in one of five faunistic assessment categories: *spring-typical*, *conditionally spring-typical*, *spring-compliant*, *non-spring*, and *very non-spring*. Unlike the structural assessment that can be done by anyone who is motivated and has basic knowledge of natural habitats, the faunistic assessment, especially the identification of the animals, can only be done accurately by experts.

1.3.3 Implementation of Spring Habitat Protection

Küry et al. (2019) explain how the significance of nature conservation (with respect to the Nature and Heritage Protection Act) can be determined from the surveys conducted according to the FOEN method: First, the number of status classes for the structural assessment is reduced from five to three categories A, B, and C according to the result of the main criterion of the structural assessment, which is the final result of the structural assessment. The faunistic assessment is likewise reduced to three categories A, B, and C, although here the result of the faunistic assessment is not the only main criterion. The number of taxa found, the number of species with an ecological score greater than 8, and the number of Red-List species are also included as main criteria for classification into categories A, B, and C. After this initial categorisation based on the main criteria, it is checked whether additional criteria are met that allow an upgrade to the next higher category. Secondary criteria for the structure categories are a large spring area ($\geq$ 100m²), a large discharge ($\geq$ 100l/s), a high degree of connectivity (spring complex/system with at least five spring outlets), the spring as a visually significant element in the landscape (scenic, historical or archaeological importance) and a diverse structure (bonus points in the overall structural assessment (yes/no)). If at least three of these secondary criteria are met, the spring is upgraded to the next higher structure category. Secondary criteria for the fauna categories are taxa with national priority (at least one species with priority 1 or $\geq$ 2 species with priority 2), as well as the number of endemic species ($\geq$ 2 endemic species). After a possible upgrade, the category according to the Nature and Cultural Heritage Act (national, regional, or local) for the spring habitat is determined using a simple matrix of categories A, B, or C for structure and fauna.

Based on spring assessment data derived from the FOEN method, it is therefore possible to assess which spring habitats need special protection, and the competent authorities can use these data as a scientific basis for decisions and protective measures relating to spring habitats. Currently, there is no national inventory of protected spring habitats. Hence, cantons and municipalities are responsible for the protection and enhancement measures for spring habitats. Even a spring habitat of national importance does not automatically receive protection; instead, it must be actively protected, like habitats of regional or local importance. Protection is enforced using the legal instruments and procedures specified by the relevant canton or municipality (Küry et al. 2019).

1.4 Structural Characteristics Influencing Spring Fauna

Various studies have shown that spring biodiversity is strongly dependent on spring types, since different spring types support different communities (Cantonati et al. 2012). An often underestimated factor for the occurrence of certain species is connectivity (the presence of a spring brook) (Ilmonen et al. 2009). Surrounding vegetation can also have a major influence on fauna composition, particularly in forest springs. For example, acidification is an important factor in the composition of taxa in springs and is of particular relevance in conifer forest springs. Some taxa that are under-represented in deciduous forest springs become quantitatively more relevant in conifer forest springs. Species richness and the total number of individuals are higher in deciduous forest springs than in conifer forest springs (Reiss & Chifflard 2018). On a medium- to large-scale perspective, factors influencing the occurrence of species, potential productivity, and hospitability of the spring include the location, altitude, and aspect of the spring, the geomorphology and variation in the soil and geology, and the regional climate, as well as the microclimate (Cantonati et al. 2020c, Kubíková et al. 2012).

1.4.1 Hydrological Factors

An important hydrological factor that influences spring fauna is flow velocity or flow rate (Cantonati et al. 2012, Kubíková et al. 2012). Bonettini & Cantonati (1996) and Fumetti & Nagel (2012) have shown that macroinvertebrate diversity is highest in springs with low flow and lowest in springs with high flow variability. Associated with spring discharge is the presence of moisture gradients in spring habitats. Staudacher & Füreder (2007) and Gerecke et al. (2011) have shown that in helocrenes, large changes in fauna composition can occur if a gradient is present from aquatic to semi-terrestrial moist areas. Other important factors for the fauna in springs include the temperature and depth of the water. Diurnal water temperature variability depends on sunlight, which warms the water, and on water depth. Annual temperature variability also has a major influence on the composition of the species community (Pešić et al. 2019).

Geochemistry (mineral richness, pH, and conductivity) regulates the composition and diversity of the biotic assemblage in springs (Peterka et al. 2023, Cantonati et al. 2012). Kubíková et al. (2012) found that the benthic assemblage is only influenced by water chemistry in areas with diverse geological bedrock or significant anthropogenic influence. Because springs are fed by groundwater, not only are temperatures stable, but the physical and chemical characteristics are also fairly stable (Cantonati et al. 2020c).

1.4.2 Influence of Substrate Characteristics and Microhabitats

Substrate type was found to be one of the main discriminating characteristics affecting invertebrate density in springs (Dumnicka et al. 2007). Species composition also depends on the substrate composition (Cantonati et al. 2012). Kubíková et al. (2012) demonstrated that substrate heterogeneity may be more important than substrate type. Substrate heterogeneity positively affects the occurrence of numerous species and may even override the influence of other environmental characteristics (Lindegaard et al. 1998 as cited in Kubíková et al. (2012)).

Spring ecosystems are characterised by a mosaic of highly heterogeneous microhabitats on a very small scale, each defined by size, water dynamics, and type of substrate (Fernández-Martínez et al. 2024, LfU 2023, Stucki et al. 2015). Laterally from the spring outlet, following the moisture gradient, there are various subhabitats, clearly indicated by characteristic plants, such as mosses or marsh plants. The boundaries between these sub-habitats are blurry and can shift depending on precipitation and spring discharge (Stucki et al. 2015). These strong ecological gradients and the spatial heterogeneity of microhabitat structures in small areas lead to high species richness (Seiler et al. 2021, Blattner et al. 2022). In spring ecosystems, multiple microhabitats can occur in close proximity and can contain distinct assemblages of organisms that may or may not interact with those in other microhabitats. Spring microhabitats include cave mouths, channels, whitewater flow, terraces, pools, spray zones, hyporheic habitats, peripheral riparian zones, wet or dry bedrock walls, and others. Each spring is a unique mosaic of microhabitats, each of which contributes to the biodiversity of the ecosystem (Stevens et al. 2021). Weigand (1998) found that the number of invertebrate taxa was significantly higher in springs with a high diversity of suitable microhabitats, and Staudacher & Füreder (2007) further specified that species diversity was primarily related to habitat complexity.

1.4.3 Influence of Stability and Disturbances

Ecosystem persistence is an evolutionarily important characteristic of springs (Nekola 1999, Cartwright et al. 2020). Long-term persistent springs function as paleorefugia on ecological and evolutionary timescales (Stevens et al. 2021). The flatworm *Crenobia alpina*, a classic glacial relict, lives in headwaters because it cannot tolerate warm temperatures, whereas the caddisfly *Rhyacophila laevis*, a tertiary relict, inhabits springs because it cannot tolerate low winter temperatures (Cantonati et al. 2012). However, recently emerging or developed springs serve as neorefugia, supporting the recent arrival of colonists (Stevens et al. 2021). Gerecke et al. (2009) also found that the high diversity of emerging insects reflects low long-term habitat stability, and vice versa.

Stability is not only important on evolutionary timescales, but can also matter for individual taxa over shorter periods. Some invertebrates are indicators of flow permanence in springs, particularly taxa with limited recolonisation capacity and sensitivity to drought (Scarsbrook et al. 2007, Gerecke et al. 2009). Temporary springs, on the contrary, are characterised by a high proportion of drought-resistant taxa or taxa with strong dispersal and recolonisation capacities (Cantonati et al. 2012). Disturbances caused by human activities, such as forest management and drainage, can affect bryophyte and macroinvertebrate taxa, as shown by Ilmonen et al. (2012). Local flooding events also significantly influence crenobiotic taxa in springs (Pešić et al. 2019).

1.5 Modelling the Occurrence of Springs

1.5.1 Models Used in the Past

A wide range of methods and techniques have been proposed and tested to study the potential of groundwater, and therefore also the occurrence of springs at a given point in the landscape. These methods include numerical/statistical methods, GIS-based methods, machine learning methods, as well as hybrid and ensemble methods.

Numerical methods, which employ mathematical equations, can provide precise predictions for groundwater assessment but require extensive and accurate input data (Sashikkumar & Colins 2017). If the necessary hydrogeological, groundwater, hydraulic conductivity, and precipitation data are limited or unreliable, this can impede the precision and reliability of numerical methods of groundwater potential (Nhu et al. 2023). Bivariate and multivariate statistical methods have shown promising outcomes in certain scenarios (Jaafarzadeh et al. 2021), but these approaches cannot account for complex nonlinear relationships and interdependencies between variables in groundwater systems (Nhu et al. 2023).

Various types of GIS-based models have been used for groundwater modelling, including frequency ratio, weights of evidence modelling, Dempster-Shafer theory, multi-criteria decision analysis, logistic regression, evidential belief function, certainty factor, and index of entropy (Rahmati et al. 2018).

In recent years, machine learning techniques have begun to replace conventional prediction methods due to their powerful capabilities for groundwater modelling (Tao et al. 2022). Machine learning methods can capture complex relationships, handle large and diverse datasets, adapt to changing conditions, integrate multiple geospatial data, and quantify uncertainties (Nhu et al. 2023). These advantages contribute to a more accurate and comprehensive modelling, enabling a better understanding and management of groundwater resources (Zaresefat & Derakhshani 2023). Among the many machine learning methods are classification and regression tree (CART), artificial neural network (ANN), naive Bayes (NB), k-nearest neighbour (KNN), boosted regression tree, mixture discriminant analysis, maximum entropy, generalised additive model, and aquifer sustainability factor (Chen et al. 2020, Naghibi et al. 2020, Nhu et al. 2023, Al-Fugara et al. 2020, Rahmati et al. 2018). Further machine learning tools such as multivariate adaptive regression spline (MARS), boosted regression tree (BRT), support vector machine (SVM), mixture discriminant analysis (MDA), and random forest (RF) offer key benefits. They facilitate the inclusion of numerous predictive variables and are robust in dealing with missing values (Al-Fugara et al. 2020). These approaches can also capture latent interactions among predictors (Friedman & Meulman 2003). A potential drawback of machine learning methods is their susceptibility to data overfitting, thereby producing unstable regression coefficients (Guisan & Thuiller 2005).

Very recently, new hybrid and ensemble methods have been introduced in which different models are integrated into a single model. In most cases, hybrid models show improved performance and produce more accurate groundwater potential models (Chen et al. 2020). The random forest modelling approach used in this work produced very good results in predicting the distribution of springs in Iran based on hydrological, geological and physiographical data (Naghibi et al. 2016, Naghibi & Pourghasemi 2015, Zabihi et al. 2016). In various studies, random forest models have shown better performance compared to other single machine learning tools (Naghibi et al. 2017, Knoll et al. 2019, Kumar & Pati 2022, Chen et al. 2020, Kumari et al. 2023).

1.5.2 Variables Predicting the Occurrence of Springs

Various explanatory variables have been shown to be useful in predicting the occurrence of springs. Slope degree (and consequently slope position and slope length) has a direct influence on runoff, erosion, and sediment transport, with gentle slopes generating less runoff, which implies that more precipitation is infiltrated or evaporated (Naghibi et al. 2020). Slope aspect is an important variable, as each aspect receives a certain amount of sunshine that determines snow melt rates, biomass production, vegetation cover, and thus the soil condition for infiltration (Naghibi et al. 2020, Zapata-Rios et al. 2016). Elevation indirectly influences spring occurrence because the development of drainage systems is significantly related to altitude, and generally at higher elevations, steeper slopes lead to lower infiltration rates (Rahmati et al. 2018). Terrain roughness promotes the presence of springs, as water discharge can occur when the topography intersects the water table (Fernández-Martínez et al. 2024). Consequently, convergence and plan curvature can also influence the occurrence of springs. Negative convergence refers to concavities (valleys), while positive convergence indicates convex features (ridges), which have a clear impact on water availability (Conoscenti et al. 2015). Plan curvature shows curvature along the direction opposite to the highest slope degree, also affecting water availability (Ayalew et al. 2004). Faults produce underground conduits that allow groundwater to flow, making distance to faults and fault density important variables for the occurrence of springs (Fernández-Martínez et al. 2024). Other variables that affect water availability and the occurrence of springs include stream density, topographic wetness, profile curvature, NDVI, soil, lithology, land use, distance to rivers, and distance to streets (Chen et al. 2018, Naghibi et al. 2016, Naghibi & Pourghasemi 2015, Rahmati et al. 2018, Zabihi et al. 2016). In statistical models, these explanatory variables serve as independent variables, whereas the occurrence of springs represents the dependent variable.

1.6 This Thesis: Motivations, Objectives, and Research Questions

Springs are small but important habitats for many highly specialised animal species. The decline of these biodiversity hotspots has led many of these species to be threatened. In addition to the loss of habitats, the increasing isolation of the remaining habitats exacerbates the situation. The spring habitats most at risk, and thus in greatest need of protection, are often located in places where conflicts of use, for example with agriculture, are most significant. However, all springs in good condition must be protected, also in forests or mountain areas where near-natural springs are still more common, since springs continue to be tapped even today. Due to climate change with altered frequency and intensity of precipitation, this situation is unlikely to improve. Moreover, alpine springs are already refuges for relict species, and in order to maintain this function and offer new refuges for species displaced by climate change, as many intact habitats as possible are required. Spring habitats are among the most endangered ecosystems, so springs that remain in a natural state must be protected at all costs.

However, for a long time, the Confederation and the cantons had only a very limited overview of the occurrence and condition of spring habitats (Küry et al. 2019). Following the Action Plan of the Swiss Biodiversity Strategy (BAFU 2017), the FOEN launched the 2019 pilot project "Tracing the Value of Water – Spring Habitat". This project focusses on surveying, protecting, and enhancing spring habitats. In the context of the pilot project, the Confederation, together with the cantons, seeks to improve knowledge about these threatened and little-noticed spring habitats. One of the tasks is to compile an inventory of spring habitats in Switzerland (BAFU 2021). Küry et al. (2019) explain that the cantons are responsible for ensuring that the number and location of near-natural spring habitats worthy of protection and the composition of their communities are known. They must ensure that spring habitats are identified, recorded, and assessed in a transparent way. Because assessing the fauna using the FOEN method requires a great deal of time and resources, it is not possible to inventory and evaluate the fauna at every potential site. Küry et al. (2019) clarify that a multistage procedure is therefore needed, which varies in each canton but essentially involves the following steps: First, the location of springs or potential spring sites must be identified. Next, a field survey is conducted to record and assess the structural characteristics of the site. After a selection process, the fauna and possibly the vegetation of the most significant spring habitats are investigated. Finally, springs that meet the protection criteria are placed under protection, employing suitable measures.

The aim of this thesis is to provide an additional tool for this spring inventory process. The Random Forest model will be used to help select which springs should undergo more detailed analyses. The spring inventory process should be simplified and made more efficient to obtain better results faster. Since the FOEN method is directly decisive whether a spring habitat will be protected, it is beneficial to apply methods or models that build on the FOEN approach already during the inventory phase. Moreover, since an ecological assessment is always required in Switzerland to determine whether a spring is worthy of protection, simplifying this process would make it easier and faster to place spring habitats under protection.

The research questions in this thesis pursue two objectives. An objective is to determine whether the structural parameters in the dataset can be used to predict the faunistic quality of spring habitats. The second objective is to investigate whether broader variables, such as those introduced in subsection 1.5.2, can also be used to predict the ecological value of springs. Hence, this thesis aims, on the one hand, to model the ecological value of spring habitats at known points in the landscape. On the other hand, it seeks to explore whether it is possible to make predictions about where in the landscape ecologically valuable springs are more likely to occur. From a scientific perspective, the goal of this work is to expand or adapt existing spring modelling approaches for the probability of occurrence to a modelling of the ecological value of springs. Answering the first research question should clarify whether it is possible to predict the faunistic status of a spring based on its structural assessment following the FOEN method. The second research question seeks to determine whether additional variables, as discussed in subsection 1.5.2, allow statements about the location of ecologically valuable springs. The third research question aims to identify the most important parameters of the models for research questions 1 and 2. Based on these objectives, the following research questions were formulated:

- Can the faunistic condition of a spring be predicted on the basis of its surveyed structural condition?
- Can the occurrence of an ecologically valuable spring be predicted using landscape & topographic variables?
- Which variables of the structural condition, along with the landscape & topographic variables, are the most important predictors?

2. Data & Methods

2.1 Swiss Springs Dataset

2.1.1 Description of the Dataset

The spring dataset used in this thesis contains data from spring surveys conducted using the FOEN method. For every spring listed, the dataset contains all the information collected in the field according to the *Spring Protocol Structure* and the *Spring Protocol Fauna*. Data from surveys conducted with this standardised FOEN method are collected in the MIDAT Sources information system. The excerpt provided from the MIDAT data portal originally included 787 springs (as of 26.8.2024). The springs were assessed as part of various projects, initiated by different clients, and carried out by different contractors. The data is owned by the respective cantons in which the springs are located. The cantons were informed about the use of the data in consultation with the national data and information centre on Swiss fauna 'info fauna'. The first entry in the dataset is dated 25.4.2002, and the most recent spring was assessed on 31.7.2023. During that span of more than 20 years, a total of 51 different evaluators assessed 787 springs in the field. As shown in Figure 1a, only a few springs were evaluated using the FOEN method in the early years, and there are no entries at all from 2008 and 2009. More than half of the springs were surveyed between 2019 and 2023.

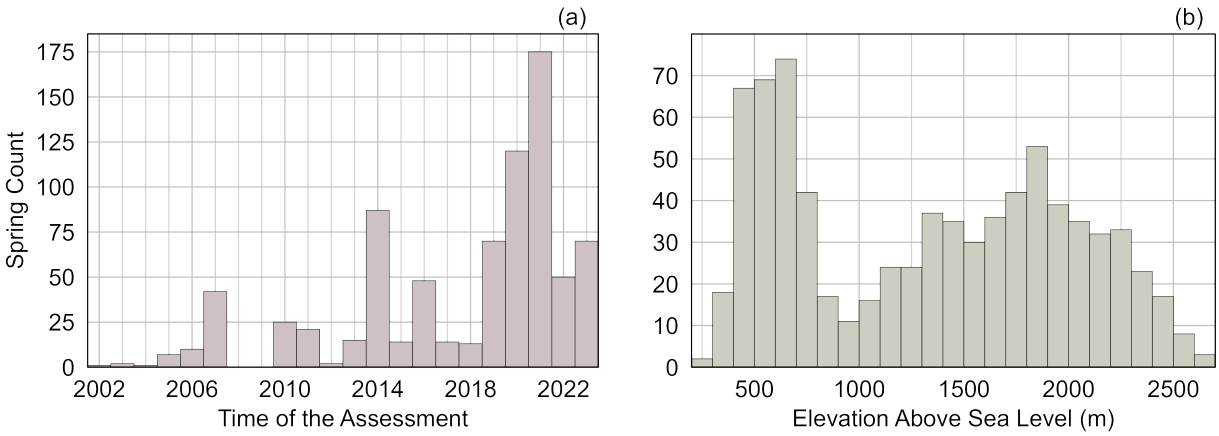


Figure 1: *Histogram plots of the assessed springs per calendar year (a) and of the springs per 100m elevation bins (b). Note that the ranges of the y-axes are different in the plots.*

The investigated springs are distributed throughout Switzerland, although certain regions or localities may show a higher number of entries depending on the project framework under which they were surveyed. There are only a few cantons with no entries at all, namely Appenzell Ausserrhoden, Geneva, Nidwalden, and Zug. As shown in Figure 1b, about 40% of the investigated springs lie below 1000 m.a.s.l., while the majority can be found in mountainous areas above 1000 m.a.s.l. The highest surveyed springs are located above 2700 m.a.s.l. Figure 2 illustrates the distribution of the surveyed springs from an overall Swiss perspective.

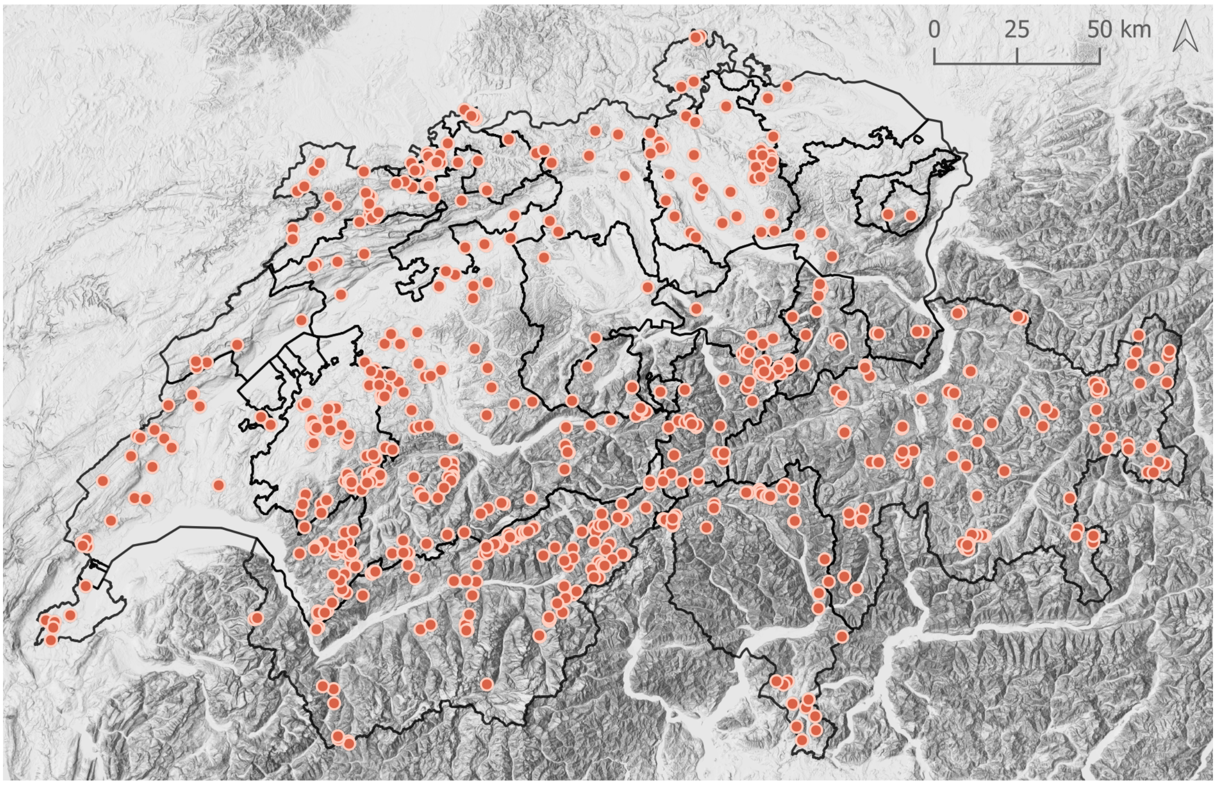


Figure 2: *Map of Switzerland showing the location of the assessed springs in the dataset. Note that not all springs are visible due to overlaying datapoints (map background: swisstopo).*

2.1.2 Dataset Processing

In the provided dataset, certain adjustments and additions were needed to enable further processing. German column headings were translated into English as faithfully as possible to present the results in this thesis. The linguistic standardisation of entries that were a mix of French and German, as well as indexing the springs with clear unique numbers, ensured the continued usability of the data. Certain variables were recalculated because the existing ones occasionally contained errors (Number of infrastructures and Number of taxa found). For categorical data, empty fields were replaced with 'none', as blank fields can cause problems in statistical analyses. It was confirmed that no empty values appear where none are expected. Missing ÖWZ values were added to the dataset using a recent list (provided by D. Küry), and some entries (Limnephilidae and Goeridae) were removed from the dataset because there is no ÖWZ classification at the family level for these.

Some variables/columns were removed from the original dataset for further processing. This includes the columns Spring discharge (time), Distance to spring discharge, Spring destroyed, and No outflow because all or almost all entries in these columns were identical or the fields were empty. The following columns were removed from the dataset as each had only very few entries, making statistical analysis impossible: three types of capture (150 captures in total), two types of relocation (80 relocations in total), three types of water damming (< 50 water dammings in total), six types of protection (< 100 protections in total), and five types of deposits (54 deposits in total). The columns Artificial substrates (19) and Filamentous algae (51) were also removed from the dataset due to the small number of values. The columns Revitalisation object (assessment) & Personal overall impression (as comparison for evaluation) were also removed because there were many empty fields and the existing entries were very subjective, making them unsuitable for statistical analysis. Certain springs had to be excluded from the dataset because the data contained errors or important variables such as Connectivity or Result structural assessment had no entries.

In the following columns, the data in the original dataset were presented as comma-separated lists: Infrastructure, Vegetation/Usage, Flow Diversity, and Special Structures. In order to analyse these data, these columns were converted into binary matrices and appended to the spring dataset. After data processing, 752 of the 787 springs remained in the dataset. In terms of structural assessment, the dataset contains 479 *natural*, 159 *conditionally natural*, 81 *moderately impaired*, 29 *damaged*, and 4 *strongly damaged* springs. Regarding the faunistic assessment, there are 199 *spring-typical*, 234 *conditionally spring-typical*, 239 *spring-compliant*, 73 *non-spring*, and 7 *very non-spring* springs in the dataset.

2.1.3 Dataset Analysis Methods

Before developing a Random Forest model, the dataset was analysed using various statistical methods. This exploratory statistical analysis served both to gain a better understanding of the dataset and to prepare for the Random Forest modelling. Identifying variables that are strongly correlated is important for creating a Random Forest model. Furthermore, the correlations found in the statistical analysis can help to better understand the results of the Random Forest model and place its conclusions in context.

Kruskal-Wallis-Tests

For all quantitative variables in the dataset, Kruskal-Wallis tests were carried out. Based on the results of the spring assessment (Result structural assessment and Result faunistic assessment), each spring was placed in one of five structure classes and one of five fauna classes. In the Kruskal-Wallis test, the test variable was defined as the dependent variable and the structure or fauna classification was defined as the independent variable. Using the p-value, it was then determined if there was a significant difference in the values of a variable between the different categories of structures or fauna, applying a significance level of 0.95. For the variables Water Temperature [℃], Discharge [l/s], Conductivity, Distance to next spring [m], and Number of outflows, empty fields had to be replaced with 'NA' before the Kruskal-Wallis test could be performed. The Kruskal-Wallis test enables to draw conclusions about whether the values of quantitative variables differ significantly among the various structure and fauna categories.

IQR-Median Heatmap
After assigning the springs to the five structure and fauna classes, the IQR and median value were calculated for all data points within these classes. To make IQR values comparable between different variables, a relative IQR (r_IQR) was calculated by dividing the IQR values of each category by the overall median value of the corresponding variable. The relative IQR values are presented in a heatmap, and the median values of each variable per category are shown numerically. Since some relative IQR values are extremely high due to the very small number of data points in certain variable-category combinations, these r_IQR values are coloured separately to maintain the comparability of the remaining data. Ultimately, this results in two heatmaps (one for the structure categories and one for the fauna categories), from which information on the variation of all quantitative variables within the dataset can be derived.

Boxplots
For selected quantitative variables, boxplots are created. The results of the Kruskal-Wallis tests were used as the criterion for determining which boxplots should be analysed in more detail. Where there are statistically significant differences between the fauna or structure categories, it makes sense to break down the point distribution in a boxplot more precisely. The data points for a given variable are distributed across the individual categories, with the structure categories on the left and the fauna categories on the right. Data points are explicitly displayed in the box plot of each category to illustrate how many data points fall into each structure or fauna category and to show the distribution of data points within each category. For certain variables, the y-axes had to be adjusted for the visibility of the remaining data, which means that a few extreme outliers are no longer visible in the plots.

Correlation Matrix
For all quantitative variables in the dataset, Spearman correlations were calculated. This makes it possible to determine the strength and direction of a monotonic relationship between two variables. Unlike Pearson's correlation, Spearman's correlation does not assume a normal distribution or a linear relationship of the data, which makes it robust against outliers or asymmetrically distributed data, as is the case in the present dataset. Since not all variables have data for every spring, only observations without missing values are taken into account when calculating correlations. As a result, the sample size may vary from one correlation to another. The correlations were calculated not only for the individual test variables, but also for the results of the spring assessments, that is, Value A: Impairment (highest value), Value B: Vegetation-use-structure, Result structural assessment, and Result faunistic assessment. This allows correlations between individual variables and the results of the spring assessments to be quickly identified.

Scatterplot Matrix
For 16 of the 17 quantitative variables in the dataset, scatterplots were generated, illustrating the relationships with Result structural assessment and Result faunistic assessment. The variable Spring area was not included because its values strongly correlate with those of Spring

size and therefore would not yield any new insights. In certain variables, the axes had to be adjusted for the visibility of the remaining data, meaning that a few extreme outliers are no longer visible in the plots. Scatterplots help identify patterns in distributions, clusters, or trends that may suggest potential correlations between two variables. Moreover, their straightforward graphical representation makes it easy to quickly derive connections which can then be pursued and verified in hypothesis formation or during the creation of the Random Forest model.

t-SNE

Categorical variables in the dataset were analysed using dimensionality-reducing visualisations (t-SNE). Such categorical variables are only found in the structural assessment part of the dataset. The t-SNE analysis was divided into one section for the Value A data and another for the Value B data (see structural assessment sheet). The categorical variables of the Value A data are divided into two types: the first type is two-level columns where there are only two possible answers, “yes” or “no” (this includes the columns Capture, Relocation, Water damming, Sole shoring, Deposits, Infrastructure, as well as the six different types of infrastructure). For each of these columns, a dummy variable is generated that takes the value 1 for “yes” and 0 for “no.” The second type of categorical variables consists of variables with several possible answer levels (this includes the columns Water extraction, Artificial drop, Trampling damage, and Discharges, each with 3 to 5 different levels). For each possible variable-level combination, a separate dummy column is generated, which can take values of 0 or 1. The Value B data also contain these two types of categorical variables. The two-level columns include the 16 vegetation / use types with 4-5 levels, the 12 different special structures, and the eight different types of flow diversity. Multi-level columns of the Value B data are the variable Summer shading (with 5 different levels) and Water-land interlocking (with 3 different levels). The dummy columns of both data types (two-level columns & multi-level columns) were merged, and a minimal jitter, i.e. a very small random deviation, had to be added to the entries 0 and 1. The jitter distinguishes identical values and prevents the creation of duplicates, to which the t-SNE analysis is very sensitive. Without the minimal jitter in the dummy dataset, only about a quarter of all springs would be included in the analysis.

Subsequently, the R package for t-SNE performs a non-linear, two-dimensional embedding of the dummy-coded data. The settings (perplexity = 30, eta = 200, max_iter = 5000, theta = 0.1) were selected through repeated tests and produce the most meaningful graphical results among the possible configurations. The results of the t-SNE analysis are then displayed in scatterplots. The data points are coloured according to the faunistic and structural assessments, revealing potential clusters in these evaluations as well. In general, t-SNE analyses allow complex categorical information to be represented in a two-dimensional space, visually highlighting potential clusters, patterns, or relationships among variables. However, it is important to note that the t-SNE axes in the plots do not permit a definitive interpretation because they do not share comparable coordinates. In fact, small distances on a t-SNE plot can be equated with similar data. However, it cannot be discerned how truly similar or dissimilar data points from different clusters are from t-SNE plots due to the absence of definitive coordinates. Moreover, because t-SNE is a stochastic procedure, each run produces slightly different results, limiting reproducibility.

Taxa Histograms

Faunistic surveys of the species found and their abundances are analysed using the Shannon Index and the Simpson Index. Both indices are widely used to quantify biodiversity in an ecosystem, such as a spring. From the TAXA columns in the dataset, a species matrix is constructed that lists the number of individuals for each taxon in each spring. Then these two indices are calculated from that matrix. The Shannon Index is calculated as follows:

$$H' = -\sum(p_k \cdot \ln(p_k))$$

Here, p_k is the proportion of individuals of species k in the total number of individuals (that is, $p_k = \frac{n_k}{N}$), where n_k is the number of individuals of species k and N is the total number of individuals in the sample). The Shannon Index measures diversity by considering both species richness and the evenness of individuals across those species. Its values range from 0 to 3, where higher values reflect a greater number of species and a more balanced distribution, indicating greater diversity.

The Simpson Index is calculated as follows:

$$D = \sum p_k^2$$

where p_k is again the relative proportion of species k. The Simpson Index represents the probability that two randomly selected individuals in a sample belong to the same species. Its values range from 0 to 1, where higher values indicate lower diversity, as fewer species are more dominant. While the Shannon Index is particularly sensitive to rare species, the Simpson Index places greater emphasis on the dominance of the most common species.

2.2 Landscape & Topographic Variables

2.2.1 Selection and Compilation of Variables

Answering the second research question requires data that are not collected using the FOEN method. Firstly, possible modelling variables from those presented in Subsection 1.5.2 were identified that have been used in previous studies to model the occurrence of springs. Secondly, all other freely available data collected by Swisstopo were analysed to assess their relevance. The criteria for meaningful variables were simple availability, relatively straightforward calculation, and potential significance for the ecological value of a spring. Table 1 lists all landscape & topographic variables that are included in the Random Forest modelling in this thesis. The variables Slope degree, Slope aspect, Slope length, Plan curvature, Profile curvature, Terrain roughness index (TRI), Topographic wetness index (TWI), and Convergence index were all calculated using DEM data. DEM data (SwissALTI3D with a resolution of 2m) were publicly available and downloaded within a radius of 1km around each spring. A resolution of 0.5m was also available, but was not used due to calculation time and the fact that the required analyses were sufficiently accurate at 2m. The variables Soil type, Soil depth, Water permeability, Water storage capacity, Waterlogging, and Soil skeleton content are all part of the soil suitability map of Switzerland and can be derived from it. The map, published in 1980, provides an overview of soil conditions from a large-scale perspective at a scale of 1:200'000. One of the geological variables is Lithology, which stands for the main lithological group extracted from the Lithology500 dataset. The lithological map of Switzerland which includes the Lithology500 dataset, provides an overview of the subsurface classified according to lithological-petrographic criteria at a scale of 1:500'000. Another geological variable is Distance to faults, which could be derived from the GeoCover dataset. The 2D geological models in the GeoCover dataset depict the near-surface area of the geological subsurface. The predominant Land use in the spring location was obtained from Switzerland's Land Use Statistics collected by the Federal Statistical Office. Other proximity variables include Distance to streets, Distance to lakes, and Distance to rivers calculated from the Swiss TLM dataset, a vector dataset of the large-scale topographic landscape model of Switzerland.

Various variables that have been successfully used to model the occurrence of springs were not selected as landscape & topographic variables. Stream Density and Fault Density were not selected as the closest distances to these features already serve as variables. The stream power index, sediment transport index, and NDVI influence the occurrence of springs, but their influence on ecological value appears to be very minor, which is why these variables were not selected. Solar radiation was not calculated, despite its potential importance, as it can only be calculated for a specific day, which offers no added value. Relative slope position and Drainage Density could not be calculated using the GIS application, but similar variables (Slope length and TWI) were calculated.

Table 1: *Overview and short explanation of landscape & topographic variables derived from various geospatial data sources.*

Data Source	Variable	Explanation
SwissALTI3D (DEM)	Slope Degree	Steepness of the terrain
	Slope Aspect	Direction the slope faces
	Slope Length	Flow length to the next local depression or flow path
	Plan Curvature	Curvature perpendicular to slope direction
	Profile Curvature	Curvature along slope direction
	Terrain Roughness Index TRI	Unevenness or roughness of the Earth's surface
	Topographic Wetness Index TWI	Tendency of a site to accumulate water
	Convergence Index	Degree to which water is concentrated or dispersed, based on topography
Soil Suitability Map of Switzerland	Soil Type	19 soil types as defined in the Soil Suitability Map
	Soil Depth	Depth of soil that can be rooted by plants
	Water Permeability	Permeability of the least permeable horizon in the top 50 cm
	Water Storage Capacity	Water retained in the soil by tension forces
	Waterlogging	Presence of excess water (slope or groundwater) in the soil
	Soil Skeleton Content	Mineral soil components larger than 2 mm
Lithological Map of Switzerland	Lithology	Subsurface classified into 25 types based on lithological-petrographic criteria
GeoCover	Distance To Faults	Shortest distance to a fault line or thrust fault
Switzerland's Land Use Statistics	Land Use	17 land use types as defined in Switzerland's Land Use Statistics
SwissTLM3D (Topographical landscape Model)	Distance To Streets	Shortest distance to a road or path
	Distance To Lakes	Shortest distance to a standing waterbody
	Distance To Rivers	Shortest distance to a flowing waterbody

2.2.2 Variable Processing and GIS Workflow

The raw DEM data, covering a radius of one kilometre around each spring, were loaded into QGIS in TIF format. Digital elevation models often contain artefacts or depressions that are not real, resulting in apparent local minima or plateaus where runoff can be blocked or is undefined. Using a GRASS algorithm, hydrologically consistent DEMs were created from the TIF files, ensuring that water could theoretically run off everywhere (see Appendix A3 for the QGIS Python code). All subsequent work steps are based exclusively on these filled DEMs to avoid calculation errors.

When calculating the DEM variables, which are explained in the following, calculations were often carried out at different scales. In a preliminary Baseline RF the modelling results of the variable replicates were then compared, and the most important replicate was selected as the final modelling variable. In most cases, replicates were calculated on three different scales. For a DEM resolution of 2 metres, a window size of 3 means that the calculation was performed on a 3x3 pixel window (i.e. 6x6 metres). With window = 13 or window = 25, the surrounding area is significantly larger, which means that the calculated variable is smoothed and interpreted on a larger scale.

Slope Degree (see Appendix A4 and A6 for the QGIS Python code):
The slope was calculated as an angle based on the local height differences in the x and y directions. The values are statistically averaged slopes over the selected scales based on a convex quadratic surface adjustment within the selected windows: 3x3 (36m^2), 13x13 (676m^2) and 25x25 (2'500m^2).

Slope Aspect (see Appendix A4 and A6 for the QGIS Python code):
The slope direction was added to the Slope degree calculations in the 3x3, 13x13 and 25x25 windows.

Slope Length (see Appendix A7 for the QGIS Python code):
In the 2000 m^2, 40'000 m^2 and 320'000 m^2 windows around the springs, the length in metres was calculated up to the next local low point or discharge path. This makes it possible to describe how far water flows across the terrain until it enters a channel or body of water.

Plan Curvature (see Appendix A5 and A6 for the QGIS Python code):
In the selected 7x7, 13x13 and 25x25 windows, a parabolic surface was adapted. Based on the height values of this terrain analysis, the curvature can be calculated in different directions. For the Plan curvature, the curvature was extracted perpendicular to the fall line, i.e. along the contour lines.

Profile Curvature (see Appendix A5 and A6 for the QGIS Python code):
The curvature along the drop line, in the direction of the maximum slope, was extracted from previous calculations.

Terrain Roughness Index (TRI) (see Appendix A7 for the QGIS Python code):
The standard deviation of the elevation values in the selected square windows around the spring grid cell (3x3, 13x13, and 25x25) corresponds to the TRI.

Topographic Wetness Index (TWI) (see Appendix A8 for the QGIS Python code):
The TWI was calculated in the selected windows (2000 m^2, 40,000 m^2 and 320,000 m^2) around the springs by combining information on the specific catchment area (i.e. how much upstream area is receiving water) and the slope gradient.

Convergence Index (see Appendix A9 for the QGIS Python code):
As a full flow path modelling approach could not be used due to technical limitations, a proxy for hydrological convergence was calculated. This proxy calculates how strongly the terrain collects or scatters water based solely on topography in the 3x3, 13x13, and 25x25 windows.

For the variables Slope degree and Slope aspect, the decision on which replica variable to use was based on a visual assessment on a map (map.geo.admin). Approximately 20 springs were consulted to select replicas that corresponded the most closely to reality, which in each case corresponded to the 3x3 window replicas. For TWI and Slope length replicates, the values were the same at all scales, so the smallest window was selected as the final variable. The other DEM variable replicates (Plan curvature, Profile curvature, TWI and Convergence index) were tested in Random Forest models where Result faunistic assessment was predicted. The replica of the variable with the highest predictive power was selected as the final variable. The 7x7 replicate was selected for Plan curvature, the 13x13 replicate for Convergence index, and the 25x25 replicate for Profile curvature and TRI.

All variables derived from the Soil Suitability Map of Switzerland and the Lithological Map of Switzerland (i.e. Soil type, Soil depth, Water permeability, Water storage capacity, Waterlogging, Soil skeleton content, and Lithology) could be assigned to the springs via a simple spatial join because they are polygon features. To calculate the distance variables (Distance to faults, Distance to streets, Distance to lakes, and Distance to rivers), the polyline features closest to the springs were identified, and the distance between the spring and feature was calculated. The predominant Land use identified at a spring location corresponds to the land use point feature (100 m grid) closest to the spring location.

2.3 Random Forest Modelling

Random Forests consist of an ensemble of classification or regression trees, each constructed using bootstrap samples of the training data and random selection of features in each split. The final prediction is obtained by aggregating the predictions of all trees in the ensemble (Svetnik et al. 2003). The results of the spring investigations according to the FOEN method (Result faunistic assessment and Result structural assessment) correspond to numerical values that can be categorised into five fauna or structure classes (see subsection 1.3.2). In the following explanations of the Random Forest modelling methods, two different types of models are used. If the aim is to model membership of a fauna or structure class, a Classification Random Forest is used where each tree casts a unique vote for the most popular class (Breiman 2001). If structural or faunistic values are to be modelled, Regression Random Forests are used, where tree predictors take on numerical values as opposed to class labels (Breiman 2001).

2.3.1 Datasets Used for Random Forest Modelling

For the use of the Swiss Springs dataset in Random Forest models, the adjustments explained in subsection 2.1.2 were made first. Some additional changes were made to the dataset before the Random Forest application. The original four columns describing the Type of capture of the springs were converted into a simple yes/no column, as only very few springs are captured at all, and even fewer springs apply to the subcategories. The three columns of Relocation, the four columns of Water damming, the seven columns of Sole shoring, and the six columns of Deposits were turned into yes/no columns for the same reason.

No further adjustments had to be made to the dataset for the landscape & topographic variables prior to Random Forest implementation, as the generation process, as described in subsection 2.2.2, was designed for direct Random Forest use.

2.3.2 Random Forest Model Variants

After the creation of an initial Baseline Random Forest model, several hyperparameter tuning approaches were applied to explore potential improvements for future models. The following analyses were carried out: First, it was tested whether increasing the number of trees improves model performance. In addition, the optimal number of variables to consider at each split and the optimal node size (minimum number of observations in each leaf node) were calculated. The effect of specifying maxnodes (maximum number of leaf nodes) in the model was also examined. Furthermore, cross-validation was checked to see whether it would improve model performance. Another analysis investigated whether removing the variables with IncMSE ≤ 0 would lead to improvements. In addition, it was tested whether transforming or standardising the data would have a positive effect on the results. None of these analyses led to any improvement, as the approach and packages used in R automatically perform a type of hyperparameter tuning, thereby automatically checking and optimising these parameters. For this reason, hyperparameter tuning was omitted in all subsequent Random Forest modelling steps.

Subsequently, various approaches within the Random Forest spectrum were tested to improve the model performance. Some approaches could be used for the Regression Random Forest models as well as for the Classification Random Forest models. Some approaches have been tested specifically for one or the other variant.

For the following approaches, a model was generated for both the Classification and Regression models, and the results were subsequently compared with the results of the Baseline model:

Regularized random forest (RRF) is a tree ensemble method that adds a regularisation penalty when selecting new features for node splits. At each split, a new feature is only chosen if it provides substantially more predictive information (measured by information gain) than features already used, controlled by a penalty coefficient. Compared to Baseline RF, RRF typically yields more compact and less redundant feature sets, improving efficiency without significant loss of predictive power (Deng & Runger 2012).

Conditional inference forest (CIF) is a method that builds multiple decision trees where variable selection and split point selection are separated, using statistical tests to identify associations between predictors and the target variable. This approach avoids the selection bias of standard Random Forests, which tend to favour variables with many categories or continuous values due to the use of Gini impurity. Compared to Baseline RF, CIF provides more reliable variable importance and unbiased selection by employing permutation tests and separating variable selection from the splitting process (Das et al. 2009).

Quantile regression forest (QRF) is an extension of the standard Random Forest that estimates the conditional distribution of the target variable for each observation, rather than only the mean prediction. This allows for a quantification of the prediction uncertainty. Compared to standard Random Forests, QRF provides robust confidence intervals for predictions by modelling the uncertainty associated with each estimate (Yang 2025).

Cost-sensitive random forest (CSRF) incorporates misclassification costs into the training by assigning higher weights to errors in minority classes. Each tree is evaluated and weighted according to its ability to minimise cost, rather than simple accuracy, and the final prediction is based on a weighted aggregation of tree output. Compared to a standard Random Forest, CSRF directly addresses imbalanced data and unequal class error costs, improving minority class detection by penalising misclassifications (Devi et al. 2019).

Extremely randomized trees (ERT) is an ensemble learning method that builds multiple unpruned decision trees, where the attribute and the split point at each node are selected completely at random. Unlike standard Random Forests, ERT does not use bootstrap samples and applies a higher degree of randomisation, in order to reduce variance more strongly while improving computational efficiency. This extreme randomisation leads to more robust models while typically maintaining predictive accuracy (Geurts et al. 2006).

For the following approach, a code was created for each Regression Random Forest model and the results were subsequently compared with the results of the Baseline RF model:

Stratified sampling (SSRF) (or Regressand Stratification) refers to the partitioning of regression data into cross-validation folds such that the distribution of the target variable is preserved across both training and test sets. By creating folds with similar output distributions, using stratification into multiple strata, this approach minimises the target shift introduced by standard k-fold cross-validation, leading to more reliable and less biased performance estimates (Sáez & Romero-Béjar 2022).

For the following approaches, a code was created for each Classification Random Forest model and the results were then compared with the results of the Baseline RF model:

Class weighting (CWRF) in Random Forests assigns different weights to classes during model training. This approach modifies the standard majority voting by introducing class-specific weights for each tree, thereby improving the classifier's ability to correctly identify minority classes in imbalanced datasets. In contrast to Baseline RF, Class weighting directly addresses class imbalance without altering the data distribution, leading to more balanced predictive performance (Zhu et al. 2018).

Dynamic weighting (DWRF) refers to assigning performance-based weights to individual decision trees during ensemble aggregation, rather than using fixed weights. These weights are dynamically determined so that trees contributing more to prediction quality have greater influence on the final outcome. Compared to standard Random Forests Dynamic Weighting leverages diversity and individual strengths of trees, leading to improved predictive performance, especially on imbalanced datasets (Shahhosseini & Hu 2020).

Threshold adjustment (TARF) using ROC Analysis (Receiver Operating Characteristic) is an approach for evaluating and improving the performance of classifiers by analysing how the true positive rate and the false positive rate change across all possible decision thresholds. Instead of relying on a fixed threshold, ROC analysis enables the identification of an optimal cut-off that can maximise metrics such as sensitivity and specificity, making model evaluation and threshold selection more robust (Carrington et al. 2023).

Weighted tree voting (WTV) in Random Forests assigns a specific weight to each tree in the ensemble, based on its individual out-of-bag (OOB) classification accuracy. During prediction, the final class label is determined by aggregating the class probabilities from all trees, each multiplied by its corresponding weight, rather than using simple majority voting. This adaptive weighting leads to more robust and accurate predictions, as trees with better generalisation performance have a stronger influence on the outcome (Paul et al. 2018).

Up-/downsampling (UDRF) (or Oversampling) is a sampling technique used to address class imbalance by increasing the number of minority class instances in the training data until both classes are equally represented. This approach aims to improve the model's ability to correctly classify minority class samples and to reduce bias toward the majority class. The ensemble structure improves noise robustness, and results in a more reliable classification performance on unbalanced datasets (More & Rana 2017).

2.3.3 Modelling Faunistic Condition Based on Spring Structure Data

The Random Forest approach to answer the first research question attempts to model the values of Result faunistic assessment or the fauna class affiliation using the extensive data of the structural assessments within the FOEN method. As Random Forest models are very susceptible to gaps in the data set, two different variants of the model had to be tested. In one variant, the variables Temperature, Discharge, Conductivity, Distance to next spring, and Number of outflows were removed from the model as they showed data gaps for some springs. The advantage of this variant is that all 752 springs are still included in the model. In the other variant, only Conductivity was excluded as a variable, meaning that only 449 springs would still be included in the model. Furthermore, it was tested whether the binary coding of the vegetation/usage data has an influence on the model quality. In one variant, binary coding with more than 50 columns was used; in another variant, a simplification was made using only one column for each of the 16 vegetation types, which indicates whether this vegetation type occurs anywhere in the spring area or not. These four dataset variants were subsequently tested in simple Baseline RF models for predicting Result faunistic assessment or fauna class affiliation. It was found that there are no major performance differences between the different dataset variants. For the sake of completeness, but also for clarity, it was decided to use the dataset variant with all variables (except Conductivity), but with only 449 springs and with only 16 vegetation/usage columns. All subsequent modelling steps for the modelling with spring structure data were performed with this dataset variant.

With the selected dataset, a Baseline Regression Random Forest and a Baseline Classification Random Forest model were each implemented, along with one model for the six Regression Random Forest variants and the ten Classification Random Forest variants (see subsection 2.3.2). In every R script, the corresponding classification or regression metrics are calculated (for further details on performance metrics see Appendix A10). A Baseline Regression Random Forest code is shown in Appendix A11. In the first part, the Random Forest model is created and trained, and in the second part the comparison metrics are calculated. Appendix A12 shows a Baseline Classification Random Forest code. All other model variants are structured in the same style with model-specific adjustments in the first part of the code.

2.3.4 Modelling Ecological Value Based on Landscape & Topographic Variables

The Random Forest approach to answer the second research question attempts to model the ecological value of springs (Result faunistic assessment & Result structural assessment or the respective class affiliation) using landscape & topographic variables (see section 2.2). Here, too, a classical Baseline Regression Random Forest and a Baseline Classification Random Forest model were implemented, along with one model for the six Regression Random Forest variants and the ten Classification Random Forest variants. In every Random Forest model, the corresponding classification or regression metrics are calculated.

2.3.5 Modelling Ecological Value Based on Mixed Spring Structure and Landscape & Topographic Variables

To determine whether model performance improves when using a combination of the two datasets previously used separately, both datasets were merged and used together to model Result faunistic assessment or the respective class affiliation. Here, too, a classical Baseline Regression Random Forest and a Classification Random Forest model were implemented, along with one model for the six Regression Random Forest variants and the ten Classification Random Forest variants. In every script, the corresponding metrics are calculated.

In total, this results in eight model groups: four regression model groups and four classification model groups. These include one group for modelling the faunistic assessment based on structural data, two groups for modelling the ecological value based on landscape & topographic variables, and one group for modelling the faunistic assessment using the mixed data. Within each model group, modelling is conducted using seven Regression Random Forest variants and eleven Classification Random Forest variants, respectively. This amounts to a total of 72 Random Forest models that were trained and evaluated in the context of this thesis.

2.3.6 Most Important Modelling Variables

Before identifying the most important modelling variables, it was necessary to determine the best Random Forest variant within each model group, that is, the variant with the best model performance metrics. Based on this information, the ten most important predictor variables for the selected Random Forest variant were then identified for the specific modelling case (for example modelling Result faunistic assessment using spring structure data). This was achieved by permuting all predictor variables in the model. For each predictor variable, a copy of the test dataset was created in which only that variable was randomly shuffled, i.e., permuted. Predictions were then generated using this modified dataset, and this process was repeated for each predictor variable. Comparing the model performance metrics for these permuted datasets with those from the original dataset provided information on how much model performance deteriorates without a variable, thereby indicating the importance of each individual variable. To assess whether the ten most important predictor variables of the selected model variant are also important in all other model variants, permutation of these ten selected variables was also applied to the other model variants. Ideally, the other model variants confirm the key variables identified by the best model. The calculation of the permutation weights of the predictor variables can be seen in the third part of the codes in Appendix A11 and Appendix A12.

3. Results

3.1 Dataset Analysis

The results of the dataset analysis, whose methods are explained in subsection 2.1.3, are presented in the following three subsections. All quantitative variables from the header section of the structural assessment sheet, from the Value A and Value B sections of the structural assessment sheet, and from the faunistic assessment sheet were analysed together, and their results are shown in subsection 3.1.1. The categorical variables from the Value A and Value B sections of the structural assessment sheet are introduced in subsection 3.1.2. Lastly, the results of the analysis of the faunistic assessments of the springs are presented in subsection 3.1.3.

3.1.1 Numerical Data

The results of the Kruskal-Wallis test for the numerical data variables in Table 2 indicate which variables show significantly different values among the structure and fauna categories. For a large proportion of the variables, there are in fact significant differences in values for both classifications, meaning for both the different structure classes and the fauna classes. This is the case for Elevation, Spring area, Number of flow types, Number of structures, Number of infrastructures, Number of red list species, Number of priority species, Number of species found, Number of outflows, and Water temperature. Some variables show significant differences for one classification but not for the other; this applies to Spring size, Number of substrates, and Conductivity. For four variables, neither classification reveals significant differences among the various categories. For the variables Spring stream length, Discharge, and Distance to next spring, this is likely due in large part to the fact that these are very difficult to estimate accurately with the available methods in the field. In addition, it is probable that not all 51 observers employed exactly the same strategy when evaluating these variables, which has resulted in a wide spread of the data and consequently no significant differences. Number of endemic species also does not show significant differences for either classification. This is largely because more than 700 of the 752 springs did not contain endemic species. Table 2 shows in the column “Data available” which variables provide data for all springs and which variables were recorded and added to the dataset only for some of the springs.

Table 2: *Results of the Kruskal-Wallis test determining statistically significant differences of datapoints among the five structure categories and among the five fauna categories ($\alpha = 0.05$). Two separate Kruskal-Wallis tests were carried out per variable, one for structure categories and one for fauna categories. P-values in green correspond to $p < 0.05$ (statistically significant differences among the categories), p-values in red correspond to $p \geq 0.05$ (no statistically significant differences among the categories).*

	Data available	**Structure Classes**	**Fauna Classes**
Elevation [m.a.s.l]	752 / 752	< 0.005	< 0.005
Spring Size [m^2]	752 / 752	< 0.005	0.16
Spring Area [m^2]	752 / 752	< 0.005	0.01
Spring Stream Length [m]	752 / 752	0.76	0.11
Number of Substrates	752 / 752	< 0.005	0.54
Number of Flow Types	752 / 752	< 0.005	< 0.005
Number of Structures	752 / 752	< 0.005	< 0.005
Number of Infrastructures	752 / 752	< 0.005	0.01
Number of Red List Species	752 / 752	< 0.005	< 0.005
Number of Priority Species	752 / 752	< 0.005	< 0.005
Number of Endemic Species	752 / 752	0.26	0.08
Number of Species Found	752 / 752	0.02	< 0.005
Water Temperature [°C]	665 / 752	< 0.005	< 0.005
Discharge [l/s]	736 / 752	0.79	0.37
Conductivity [µS/cm]	372 / 752	0.38	< 0.005
Distance to Next Spring [m]	555 / 752	0.38	0.45
Number of Outflows	607 / 752	< 0.005	< 0.005

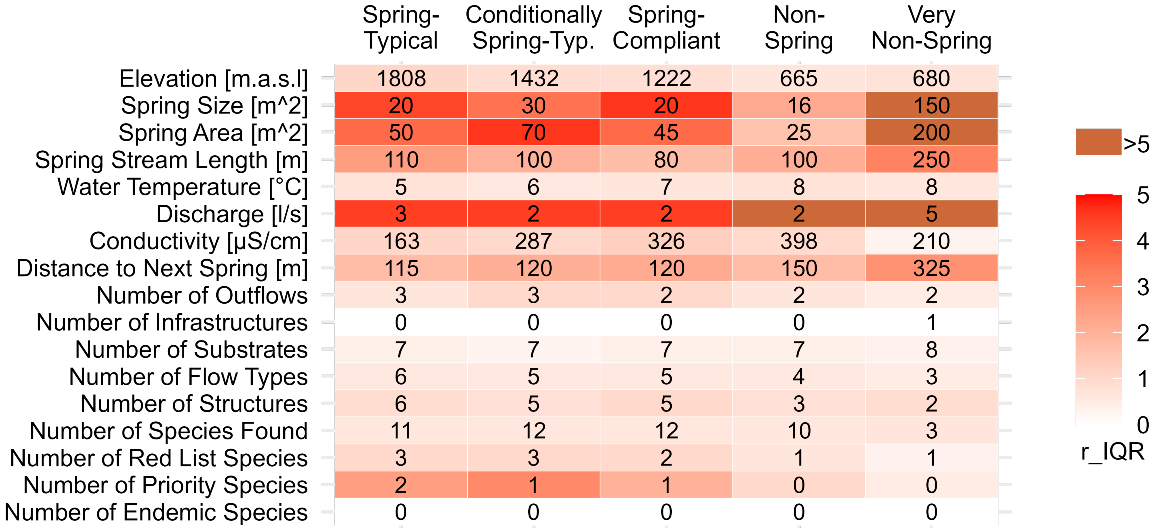
In Figure 3, the heatmap of the relative IQR values is shown for all numerical variables per structure category. IQR values allow statements about the variation of a variable's data points within a category. Dividing by the median of the corresponding variable ensures comparable scale sizes of the r_IQR values between different variables. High relative IQR values indicate extensive variation of the data points within a category and compared to the median of the variable, while low relative IQR values reflect low variation of the data points. Many variables show relatively low relative IQR values that do not change significantly between the different categories of structures. These variables include Elevation, Water temperature, Conductivity, Number of outflows, Number of infrastructures, Number of substrates, Number of flow types, Number of structures, Number of species found, Number of red list species, and Number of endemic species. Although the median of some of these variables changes considerably in the different categories (Elevation, Water temperature, Conductivity, Number of outflows & Number of structures), the variation in the IQR does not appear to be influenced by these alterations in the data structure. Other variables such as Spring size, Spring area, Spring stream length, Discharge, and Distance to next spring have relatively high relative IQR values, sometimes with large differences between individual categories. For example, in the highest structure category *natural*, there is a very high variation in Spring size and Spring area, whereas in the lowest structure classes *damaged* and *strongly damaged*, these variables show little variation. The median values of these two variables also differ substantially across the categories, in contrast to Discharge, whose median values remain relatively constant in the various categories, even though the relative IQR values reveal varying degrees of data variation. It is important to note that the category *damaged* contains only 30 springs and *strongly damaged* only 4. Depending on the variable, this number may be even smaller due to gaps in the data.

	Natural	Conditionally Natural	Moderately Impaired	Damaged	Strongly Damaged
Elevation [m.a.s.l]	1592	1280	688	597	488
Spring Size [m^2]	25	20	10	11	5
Spring Area [m^2]	60	50	19	16	6
Spring Stream Length [m]	100	100	90	100	150
Water Temperature [°C]	6	6	8	9	11
Discharge [l/s]	2	2	2	2	1
Conductivity [µS/cm]	262	266	248	433	609
Distance to Next Spring [m]	110	135	150	100	800
Number of Outflows	3	2	2	1	1
Number of Infrastructures	0	1	0	1	1
Number of Substrates	7	7	6	6	5
Number of Flow Types	6	5	4	3	3
Number of Structures	6	5	3	2	2
Number of Species Found	11	11	12	13	7
Number of Red List Species	3	2	2	2	2
Number of Priority Species	1	1	1	0	1
Number of Endemic Species	0	0	0	0	0

r_IQR: 0, 1, 2, 3, 4, 5, >5

Figure 3: *Structural assessment categories with median values of the individual variables and the relationship between the IQR and the median value of a variable highlighted in color according to the color scheme on the right* ($r_{\mathrm{IQR}} = \frac{IQR_{\text{per category}}}{Median_{\text{overall}}}$).

In Figure 4, the heatmap of the relative IQR values is shown for all numerical variables per fauna category. The heatmap is remarkably similar to that for the structure classes. Variables with low relative IQR values and minimal changes in categories are the same as in the structure classes. In contrast, the median values of Elevation, Conductivity, and Number of outflows show fewer marked differences among the fauna categories. The variables with high relative IQR values are likewise the same, with the only distinction that Distance to next spring tends to display lower values and less variability across the categories. Unlike in the structure categories, Spring size, Spring area, and Discharge show comparatively high relative IQR values in every fauna category, indicating substantial variation in these variables in all fauna categories. It should also be noted that the lowest-rated fauna category (*very non-spring*) contains only 10 springs, and this number can be even smaller for certain variables due to data gaps.



	Spring-Typical	Conditionally Spring-Typ.	Spring-Compliant	Non-Spring	Very Non-Spring
Elevation [m.a.s.l]	1808	1432	1222	665	680
Spring Size [m^2]	20	30	20	16	150
Spring Area [m^2]	50	70	45	25	200
Spring Stream Length [m]	110	100	80	100	250
Water Temperature [°C]	5	6	7	8	8
Discharge [l/s]	3	2	2	2	5
Conductivity [µS/cm]	163	287	326	398	210
Distance to Next Spring [m]	115	120	120	150	325
Number of Outflows	3	3	2	2	2
Number of Infrastructures	0	0	0	0	1
Number of Substrates	7	7	7	7	8
Number of Flow Types	6	5	5	4	3
Number of Structures	6	5	5	3	2
Number of Species Found	11	12	12	10	3
Number of Red List Species	3	3	2	1	1
Number of Priority Species	2	1	1	0	0
Number of Endemic Species	0	0	0	0	0

Figure 4: *Faunistic assessment categories with median values of the individual variables and the relationship between the IQR and the median value of a variable highlighted in color according to the color scheme on the right* ($r_{IQR} = \frac{IQR_{\text{per category}}}{Median_{\text{overall}}}$).

Figure 5 shows boxplots for four different variables, broken down by the structure categories on the left and the fauna categories on the right. In the boxplots for Elevation, one can clearly see that the median values increase with better-rated structure and fauna categories. Basically, the higher the spring is located, the better it is rated overall. However, the boxplots for each category also reveal a large spread in the data, even though the median values suggest a clear trend. The distribution of the data points appears very similar between the structural and faunistic assessments, with the biggest difference being the median value in the yellow categories, which differs by nearly 500 metres. In the plots for Number of substrates, no such strong trend is evident. In the structure categories, the two highest rated categories have the highest median value, but the same cannot be observed in the fauna categories. Here too, there is substantial data dispersion in each category. For Number of flow types, there is again a fairly clearer trend, with higher values found in higher-rated structure and fauna categories. Hence, the more flow types that can be identified in a spring, the better its overall rating. A nearly identical pattern emerges for Number of structures, where the higher the number of different structures, the better the springs are rated. Unlike Number of flow types, however, the values for Number of structures can reach higher levels, yet the median values remain almost the same in the two variables across nearly all categories.

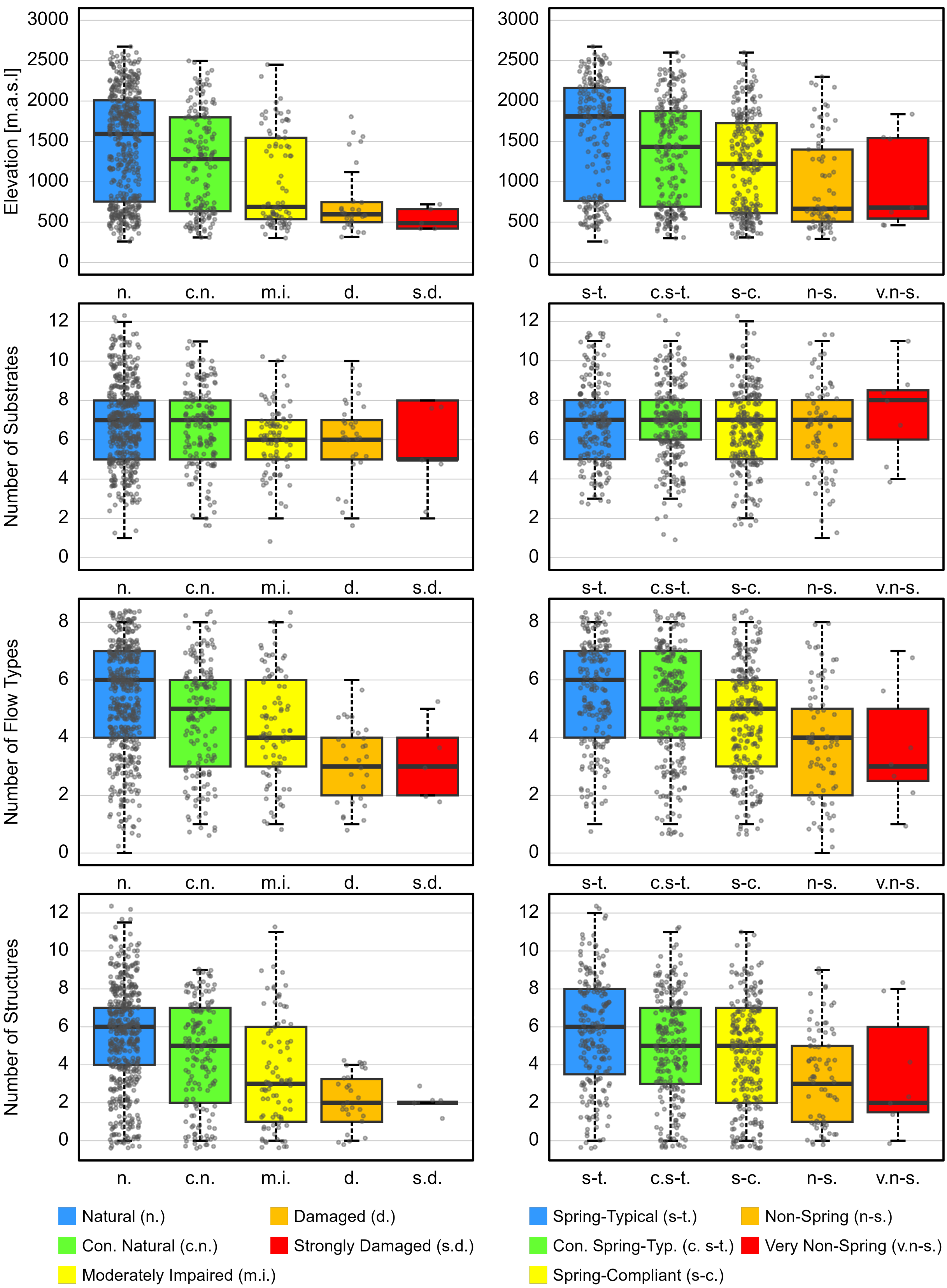


Figure 5: *Boxplots of four spring variables grouped by structure (left) and fauna (right) classes. Gray dots show individual data points, randomly jittered in the x-direction and slightly in the y-direction to improve visibility.*

Figure 6 also presents boxplots for four different variables. In the Number of outflows plots, numerous outliers stand out, some of which can reach very high values. Apart from these outliers, the median values for the Number of outflows likewise show the anticipated increase in better-rated springs, in both the structural assessment and the faunistic assessment. For Number of species found, fairly surprisingly, no trend emerges showing more species in springs with higher ratings. The median values in the two top categories are in fact lower than those of the two green and yellow categories. Thus, the number of springs found does not appear to be a determining factor for the quality of a spring. Regarding the Number of red list species, a clear trend towards higher values in better-rated categories can be observed, particularly in the faunistic assessment. A similar but less pronounced trend appears in the structural assessments as well. However, as with all boxplots, there is considerable data dispersion within each category. In the case of the Number of priority species, the fauna categories show a trend similar to that of Number of red list species, with higher numbers in better-rated springs. However, the median values in all 10 boxplots for this variable are very low, with outliers reaching values of up to 10.

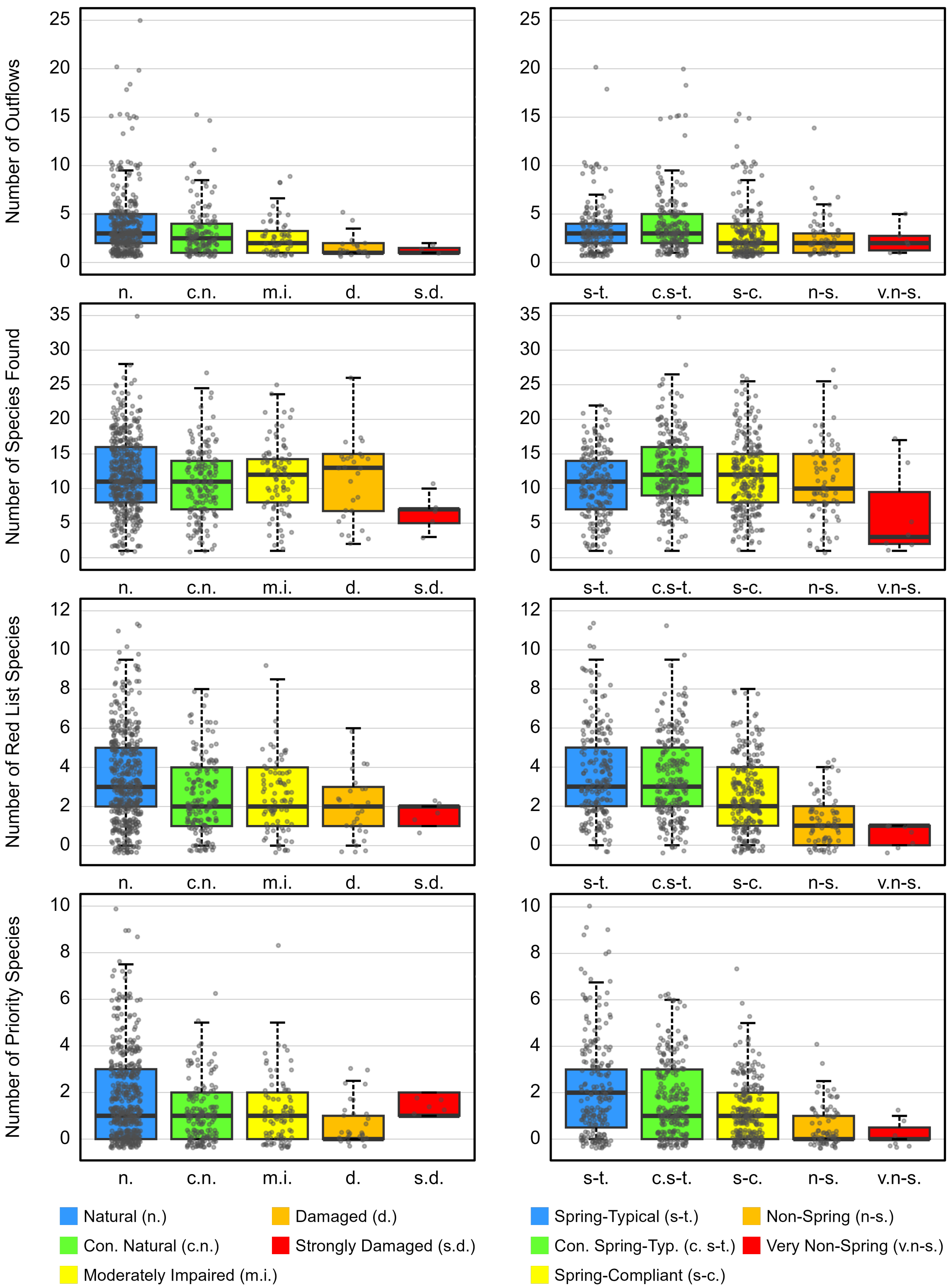


Figure 6: *Boxplots of four spring variables grouped by structure (left) and fauna (right) classes. Gray dots show individual data points, randomly jittered in the x-direction and slightly in the y-direction to improve visibility. Note that the y-axes of the Number of outflows plots are truncated to highlight the main data range, resulting in three outliers falling outside the plotted area.*

In Figure 7, a Spearman correlation matrix is shown, clearly representing the strength and direction of the monotonic relationship between all pairs of numerical variables and the results of the structural and faunistic assessments. The highest correlation value among all variable combinations occurs between Number of red list species and Number of priority species, with a Spearman correlation of 0.81. This is not surprising, as Red List Species are often also Priority Species and vice versa. A similarly high correlation value (0.80) is found between Spring area and Spring size. This likewise makes a great deal of sense, since on the one hand it is often very difficult to distinguish these two measures in the field, and on the other hand a larger Spring size automatically implies a larger Spring area. The highest negative correlation between two variables is that between Elevation and Water temperature (-0.76). With increasing elevation,

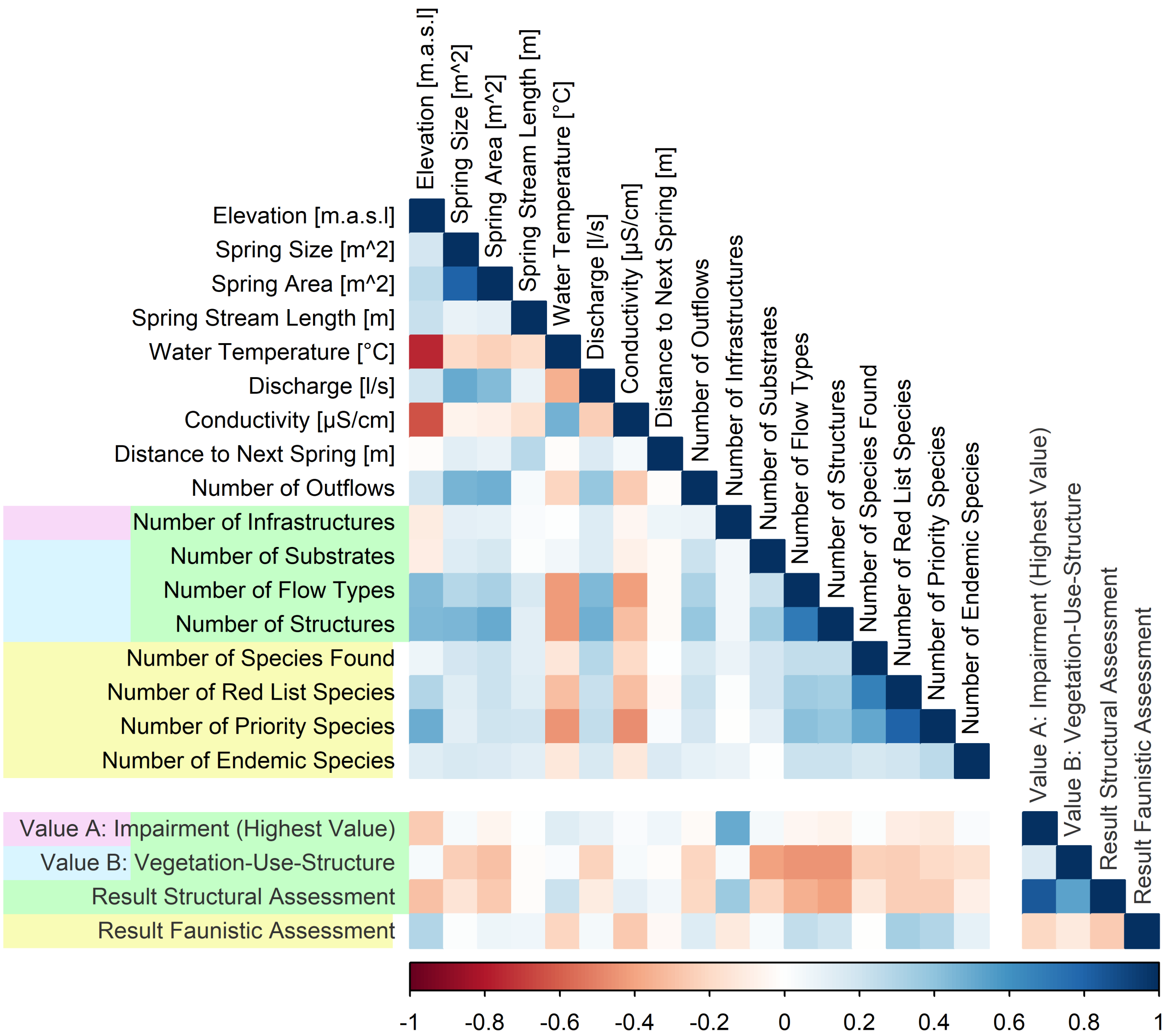


Figure 7: *Spearman correlation matrix of all numerical variables in the dataset. Blue cells indicate positive correlations, red cells indicate negative correlations, and color intensity corresponds to the absolute correlation strength. Note that some variables do not have data available for all springs, sample size can therefore vary among different correlations. Variables highlighted in yellow belong to Result Faunistic Assessment, those in green to Result Structural Assessment, and those in white represent header data.*

the average annual temperature decreases, which, as expected, is reflected in the water temperatures of the springs. One other correlation lies above the threshold of 0.70, while all remaining 132 correlations fall below it. The correlation between Number of flow structures and Number of flow types (0.71) is likewise unsurprising, as it seems logical that a large number of different flow structures in a spring area would lead to a large number of different flow types, and vice versa.

Table 3 shows a descending list of the largest mean absolute Spearman correlations (excluding the results of Value A, Value B, as well as the structural and faunistic assessments). It can be seen that the two Value B variables Number of structures and Number of flow types exhibit the highest absolute correlations. Thus, these two variables either exert a very strong influence on other variables or are strongly influenced by others (or both). Interestingly, the third numerical Value B variable, Number of substrates, has one of the lowest absolute correlation values, indicating that it is barely influenced by or barely influences other variables. The two variables Number of priority species and Number of red list species, which themselves correlate strongly, also show high absolute correlation values. In contrast, Number of species found appears to be only fairly influenced by other variables. Particularly low absolute correlation values are observed in Distance to next spring and Number of infrastructures. These variables seem to exert very little influence on other variables and, in turn, are hardly influenced by them.

Table 3: Overview of the average absolute Spearman correlations between the numerical variables in the dataset.

Average Absolute Correlations	
Number of Structures	0.340
Number of Flow Types	0.316
Number of Priority Species	0.305
Elevation [m.a.s.l]	0.285
Discharge [l/s]	0.283
Number of Red List Species	0.277
Spring Area [m^2]	0.276
Water Temperature [°C]	0.272
Spring Size [m^2]	0.254
Conductivity [µS/cm]	0.243
Number of Outflows	0.232
Number of Species Found	0.219
Number of Endemic Species	0.151
Spring Stream Length [m]	0.133
Number of Substrates	0.127
Distance to Next Spring [m]	0.070
Number of Infrastructures	0.067

It can also be clearly observed in Figure 7 that certain variables display very similar correlation values to others. For example, Spring area and Spring size often exhibit similar correlation values with the remaining variables. The same applies to Water temperature and Conductivity, which show a comparable behaviour in relation to other variables. Among Number of structures, Discharge, and Number of outflows, there is even a group of three variables that present similar correlation values with the rest of the variables.

Below the correlation matrix of the numerical variables, the matrix continues with the results of the structural and faunistic assessments. In the correlations with the variable Value A, it is noticeable that only one correlation exceeds 0.3, namely the correlation with Number of infrastructures (0.51). This is understandable insofar as Number of infrastructures is the sole numerical value derived from the calculation of Value A. In the correlations with the Value B variable, the highest correlations occur with those variables that are taken into account when calculating Value B: Number of structures (-0.45), Number of flow types (-0.44), and Number of substrates (-0.41). All other correlations with Value B are below 0.3. The overall Result structural assessment does not show any correlation above 0.4, while the highest correlation with Result faunistic assessment is 0.33. The correlation between Result faunistic assessment and Result structural assessment is -0.25. Consequently, better faunistic conditions (higher fauna values) tend to coincide with better structural conditions (lower structure values). However, this relationship is not particularly strong.

In Figure 8, a scatterplot matrix is shown, illustrating the relationships of all numerical variables except Spring area with Result structural assessment. The first two rows of eight plots (Elevation, Spring size, Spring stream length, Water temperature, Discharge, Conductivity, Distance to next spring and Number of outflows) depict the numerical header data of the structural assessment sheet. These are therefore the numerical variables that do not factor into the assessment calculation but are always recorded in the field. None of the eight plots exhibit very clear patterns, clusters, or trends suggesting clear correlations between two variables. In particular, Elevation and Water temperature show no discernible trend at all. In plots b, c, e, f, and g, a large share of the data has very small values on the x-axis, with only a few data points reaching higher values, which hinders or at least complicates the visibility of potential trends. In plots h and b, it can be seen that larger values on the respective x-axis occur only at low structural assessment values, that is, in well-rated springs. However, even in these two plots, a low structural assessment does not automatically mean larger values on the x-axis. The third row in the scatterplot matrix consists of four plots (Number of infrastructures, Number of substrates, Number of flow types, and Number of structures) and represents the numerical Value A and Value B data from the structural assessment sheet. Again, no clear clusters or distributions are discernible in these four plots. Only a few weak trends are observable. In plots j, k, and l, no very low structural assessments occur when the values on the x-axis are very low. In contrast, at very high values on the x-axis, which means a large number of different substrates, flow types, or structures, very high structural values are not observed. The fourth row in the scatterplot matrix comprises four plots (Number of species found, Number of red list species, Number of priority species, and Number of endemic species), which represent the numerical variables of the faunistic assessment sheet. In plot m, Number of species found, no trend can be identified at all. In plots n, o, and p, only a weak trend is evident, namely that very high x-axis values (many Red List, Priority, or Endemic Species) are not associated with very high structural assessments. A large number of different Red List, Priority, or Endemic Species are therefore found almost exclusively in springs that are rated very highly in structural terms.

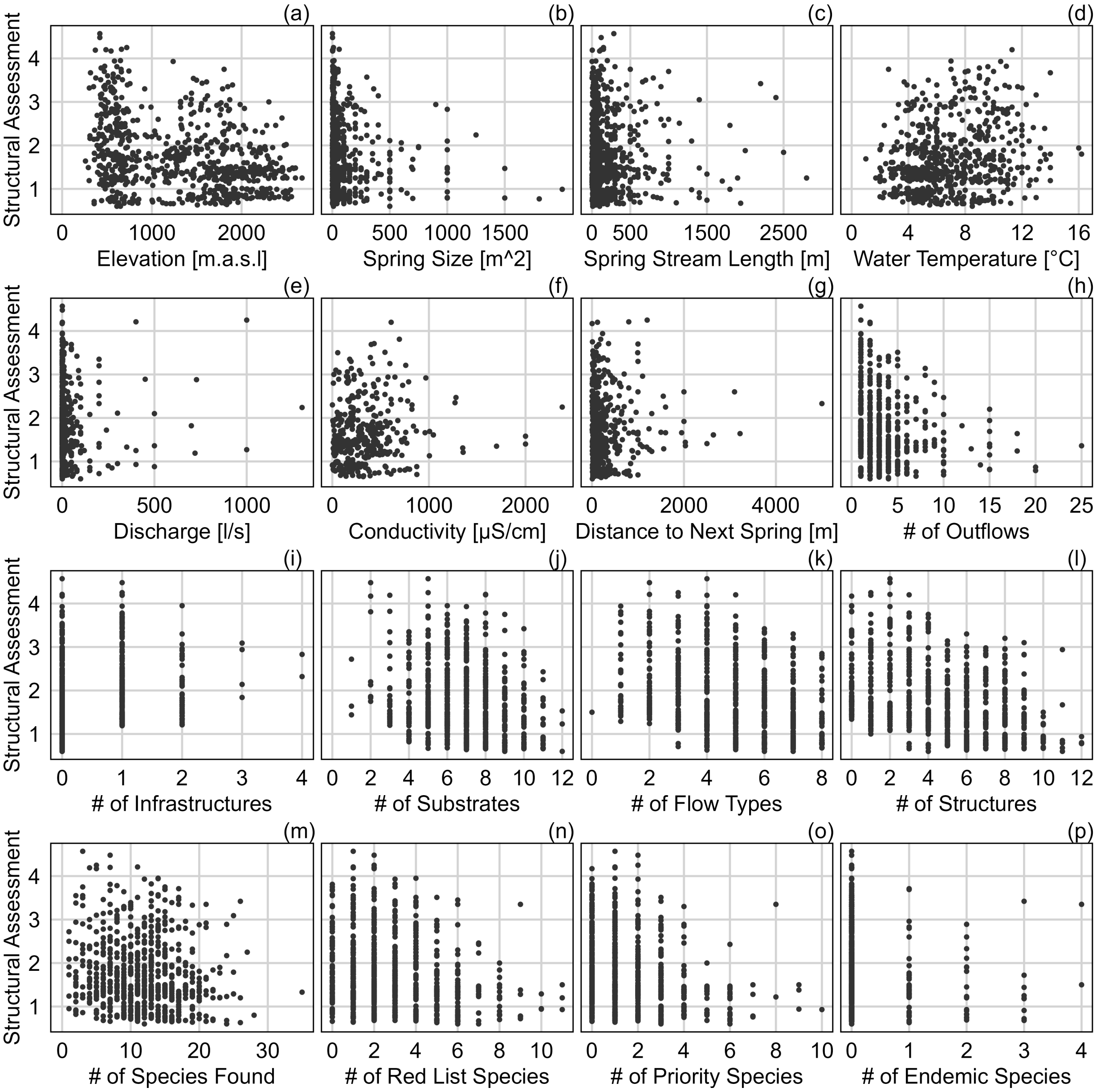


Figure 8: *Scatterplot matrix of 16 numerical variables on the x-axes plotted against Result structural assessment on the y-axes. Plots a-h show the header data of the structural assessment sheet. Plots i-l show Value A and Value B data of the structural assessment sheet. Plots m-p show numerical values of the faunistic assessment sheet. Note that some axes are truncated to highlight the main data range, resulting in a few outliers falling outside the plotted area (2 outliers in plot b, 1 in plot c, 3 in plot e and 1 in plot h).*

In Figure 9, a scatterplot matrix is shown, illustrating the relationships of all numerical variables except Spring area with the Result faunistic assessment. Plots a through g closely resemble the corresponding plots in Figure 8. Again, no clear clusters or trends are evident. Unlike plot h in Figure 8, plot h here does not show any weak tendency toward higher or lower values of the y-axis with a larger Number of outflows. In plot i, a very slight trend can be observed toward a lower faunistic assessment at higher values of Number of infrastructures. An equally slight trend appears in plot k, where a higher Number of flow types tends to lead to a better faunistic assessment. No trends can be identified in the plots j and l. The plots m and p also reveal no apparent dependencies. In plot n, there is a slight tendency toward a higher faunistic assessment with a greater Number of red list species, and a similar but even weaker trend can be seen in plot o.

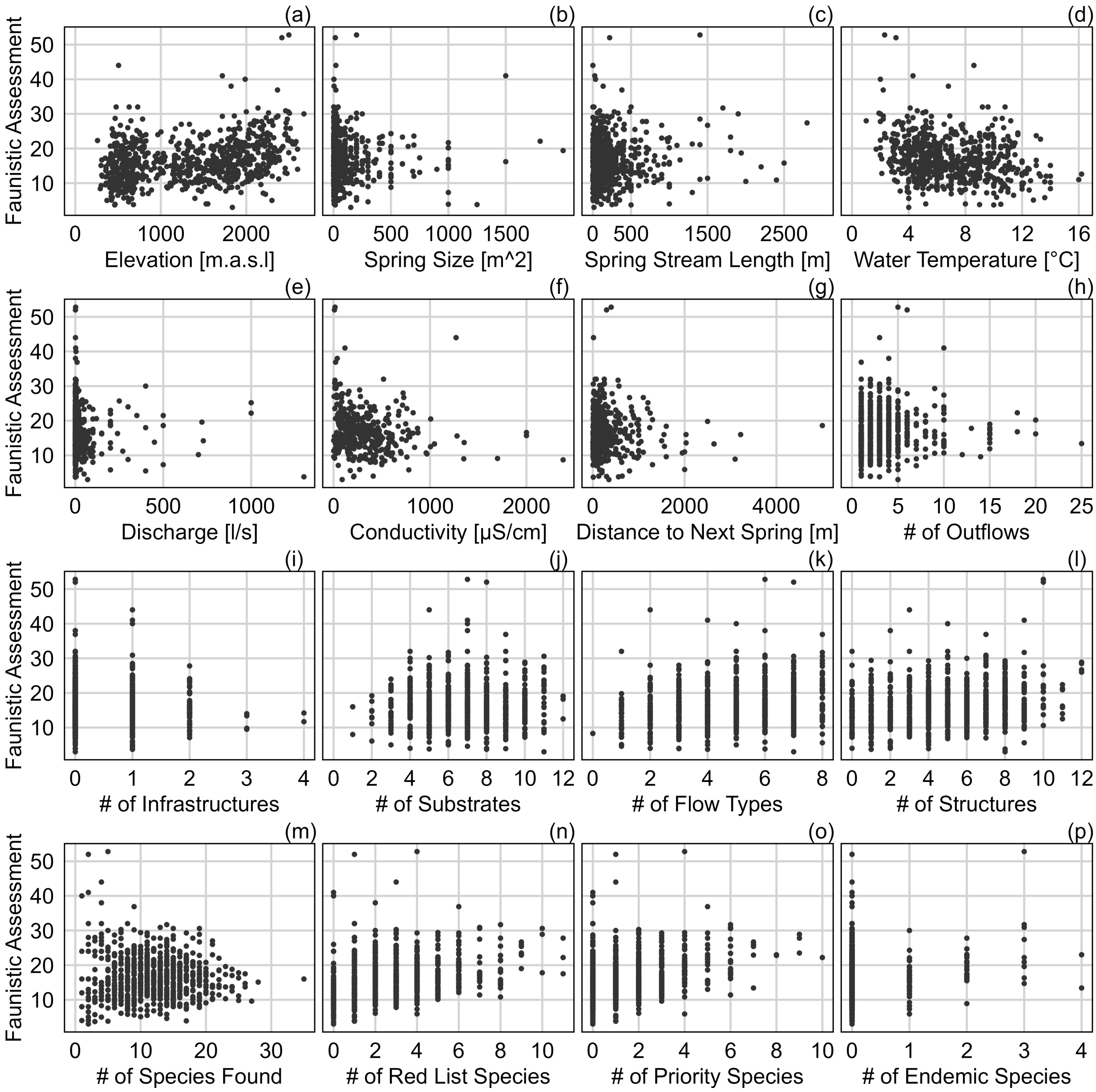


Figure 9: *Scatterplot matrix of 16 numerical variables on the x-axes plotted against Result faunistic assessment on the y-axes. Plots a-h show the header data of the structural assessment sheet. Plots i-l show Value A and Value B data of the structural assessment sheet. Plots m-p show numerical values of the faunistic assessment sheet. Note that some axes are truncated to highlight the main data range, resulting in a few outliers falling outside the plotted area (2 outliers in plot b, 1 in plot c, 3 in plot e and 1 in plot h).*

3.1.2 Categorical Data

In Figure 10, two scatterplots are shown, which illustrate the results of the t-SNE analysis for the categorical data in the Value A section of the structural assessment sheet. In the left scatterplot, the data points are coloured according to the results of the structural assessment, while in the right scatterplot they are coloured according to the results of the faunistic assessments. Note that the colour scheme here does not follow the standard pattern, as the aim was to highlight the differences among the points more clearly. In the scatterplot, two clear clusters can be identified: one dense cluster on the right and another fairly less dense cluster on the top left. A third possible cluster is located in the lower middle-left area. The points in individual clusters share similar characteristics. It can also be seen that the clusters on the right and at the top left are clearly separated from each other, meaning these groups of data points are unambiguously different. Whether or not the broad cluster on the lower middle left overlaps with the one on the middle left is not entirely clear. An overlap would mean strong similarities between the two clusters. Meanwhile, the distance between individual clusters is meaningless. It is therefore not possible to say whether the cluster at the top left differs more from the cluster on the right than from the cluster in the middle, even if the distance to the right is greater. Because identical point characteristics are plotted at the same location, statements can also be made about the data distribution within a cluster. In the cluster on the right, the data points are closely grouped, suggesting strong correlations among them. The trend is similar, though weaker, in the top-left cluster, whereas in the middle cluster there are fewer strong relationships among the data points. In the left plot, which shows the results of the structural assessment, a clear distinction can be seen in the colouring of the data points between the different clusters. The top-left cluster contains the lowest scores, and there are only very few data points with a rating greater than 2. The middle cluster has the highest ratings, although it also includes many

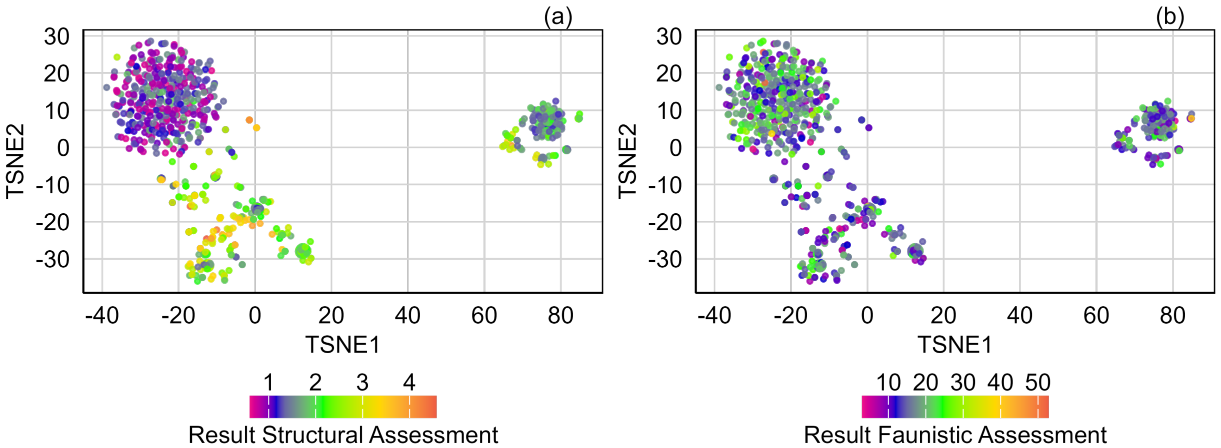


Figure 10: *t-SNE scatterplots for the categorical variables of structural assessment part A: impairment. Both scatterplots show the same datapoints, but colored according to the results of the structural assessment (plot a) and the results of the faunistic assessment (plot b). Note that the axes represent the two-dimensional embedding obtained by the t-SNE analysis, with no real-world units for direct interpretation.*

medium ratings and some relatively low ratings around 1. In the right cluster, the mix is again relatively large, with an overall slightly lower rating than in the middle cluster. In the right plot, which shows the results of the faunistic assessment, there is no clear distinction in the colouring of the data points among the different clusters. All clusters contain very low and very high ratings. The separation between the middle cluster and the top-left cluster is also less pronounced here because, unlike the structural assessment, the ratings do not clearly diverge.

In Figure 11, two scatterplots are shown, illustrating the t-SNE analysis results for the categorical data in the Value B section of the structural assessment sheet. In the left scatterplot, the data points are once again coloured according to the results of the structural assessment, while in the right scatterplot they are coloured according to the results of the faunistic assessments. Unlike in the scatterplot of the Value A data, no clear clusters can be identified here. Thus, no groups of data points can be found that are unambiguously different, and as a result there are no strong correlations between the features and the data points. In the middle-right area of the scatterplot, some data points appear closely clustered, indicating a strong correlation among those points, although this does not constitute a distinct cluster. With respect to the colouring of the points, scarcely any trend can be discerned. On the left side, where the structural assessment results are displayed, it is evident that medium and high values occur throughout the entire data range, while the lowest values appear only in the middle to left portion of the plotted area. In the right plot, which shows the results of the faunistic assessment, there is also no discernible trend in the coloring. Low, medium, and high assessments occur everywhere in the plotted area.

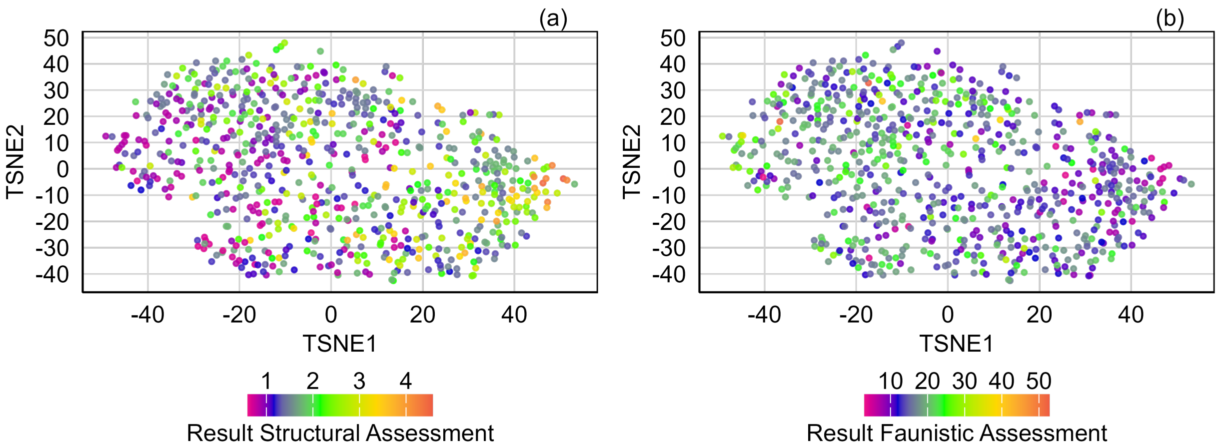


Figure 11: *t-SNE scatterplots for the categorical variables of structural assessment part B: vegetation / use - structure. Both scatterplots show the same datapoints, but colored according to the results of the structural assessment (plot a) and the results of the faunistic assessment (plot b). Note that the axes represent the two-dimensional embedding obtained by the t-SNE analysis, with no real-world units for direct interpretation.*

3.1.3 Fauna Data

In Figure 12, four histograms are shown, illustrating the results of the faunistic surveys in the dataset. The number of springs within a certain index value range is summed and plotted on the y axis, while the x axis displays the corresponding index ranges. The first row shows the plots for the Shannon Index, and the second row presents the Simpson Index. On the left, the bars are coloured according to the respective fauna categories, and on the right, according to the respective structure categories.

In the Shannon Index plots, it can be observed that most of the springs reach relatively high values. The frequency distribution is not fully normal but rather slightly skewed to the right, towards greater diversity. More than two-thirds of the springs have a Shannon Index of 1.5 or higher, while fewer than one-third register a Shannon Index below 1.5. Since higher values reflect a larger number of species and a more balanced distribution, it can be concluded that most of the springs exhibit high Shannon diversity. Only a few springs display very low diversity and similarly few show very high diversity. It is important to note that the Shannon Index is sensitive to rare species, suggesting that at least one rare species was found in many of these springs.

In plot b, it can be seen that, contrary to expectations, there is no trend towards a higher proportion of well-rated springs (blue) at higher Shannon Index values. The share of each fauna category in the total number of springs for a given Shannon value hardly changes between different index ranges. This suggests that the faunistic assessment based on the FOEN method produces results that are scarcely comparable to the indices used here. Particularly striking is that among the highest Shannon ratings (>2.4), meaning the springs with the greatest diversity, only about one sixth fall into the blue fauna category. In plot a, there is also no discernible trend towards higher or lower structural assessments at higher Shannon Index values. Here, too, the proportion of each structure category in the total number of springs changes very little across the different index ranges. However, in contrast to plot b, it is noticeable that a large majority (approximately 75%) of the springs with the highest Shannon ratings (>2.4) also fall into the highest structural assessment category.

In the Simpson Index plots, it can be seen that a large proportion of the springs reach high values. More than half of the springs exhibit a Simpson Index above 0.75, whereas only about 10% have a Simpson Index below 0.5. Since higher values indicate that fewer species are strongly dominant, and therefore diversity is lower, it must be noted that a substantial share of the springs shows low Simpson diversity. Although only a few springs have very high values (>0.9), there are also only around 30 springs with a Simpson value below 0.25. It is important to remember that the Simpson Index emphasises the dominance of the most common species. Consequently, in many springs, there is a relatively high probability that two randomly selected individuals from a sample belong to the same species.

In plot d, too, there is no discernible trend towards a higher proportion of well-rated springs (blue) at lower Simpson values. Here, also, the shares of each fauna category in the total number of springs for a given Simpson value remain largely constant throughout most of the index range. Only about one third of the approximately 30 springs with a Simpson value below 0.25 fall into the blue fauna category. This further supports the assumption that the faunistic assessment according to the FOEN method produces results that are barely comparable to the Shannon and Simpson indices. In plot c, no trend in the structural assessments at higher Simpson Index values can be identified either. However, more than half of the roughly 30 springs with a Simpson value below 0.25 fall into the blue structure category.

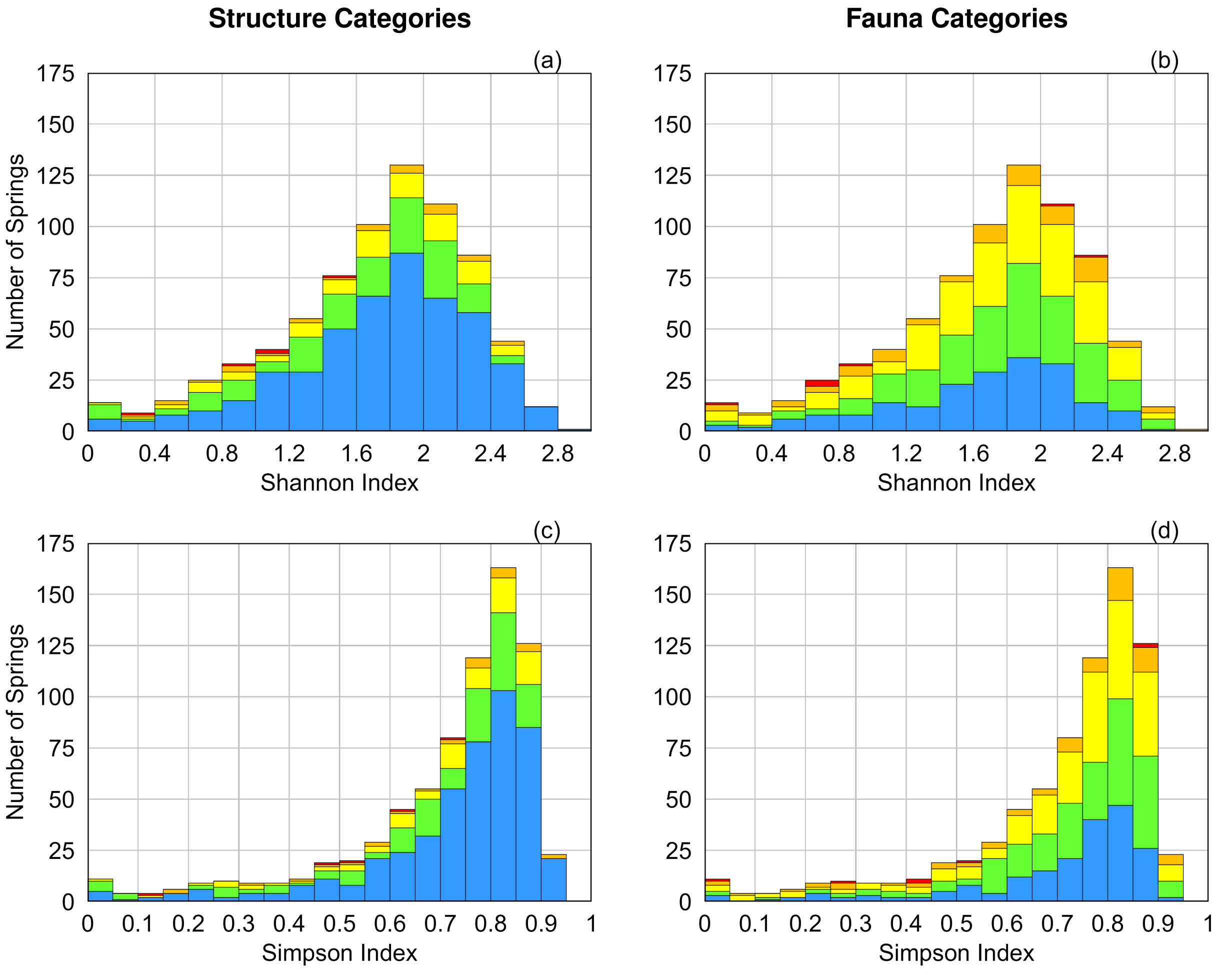


Figure 12: *Histogram plots showing the total count of springs with certain Shannon or Simpson index values. The Shannon index values are shown in the first row, while the Simpson index values are shown in the second row. The structural classes are shown on the left, while the fauna classes are shown on the right.*

3.2 Modelling Faunistic Condition Based on Spring Structure Data

3.2.1 Regression Random Forest: Modelling Faunistic Assessment Values

The results of the different Regression Random Forest variants used to answer the first research question are shown in Figure 13. In these models, data from the structural survey according to the FOEN method were used to model Result faunistic assessment. The observed faunistic assessment values are shown on the x-axis and the faunistic assessment values predicted by the models are shown on the y-axis. This results in a diagonal 1:1 line on which the data points would lie in the ideal case of a perfect model. The median of the observed faunistic assessment values is 16.0, the mean is 16.5. For each Random Forest variant, the values of RMSE (root mean squared error), R^2 (coefficient of determination), MAE (mean absolute error) and MedAE (median absolute error) are given in the figure in order to be able to compare the model quality with each other.

The xy-plot of Baseline RF in Figure 13 shows that the model always overestimates the low observed faunistic assessment values. The opposite is true for high observed faunistic assessment values: here the model underestimates the real values. This results in a relatively flat fitted line of the predicted fauna values. A slight tendency towards correct predictions can be recognised. The data are relatively heavily scattered around the fitted line. Baseline Random Forest has an RMSE value of 5.30. The average deviation of the predictions from the true value is therefore greater than the interval of a fauna class (5.00). The R^2 value of 0.24 indicates that the model can only explain approximately 24% of the variance of the target values. The MAE value of 4.01 indicates that the prediction is on average 4 units off the true value. This value is slightly lower than the RMSE value, which means that there are some outliers that worsen the RMSE result. The MedAE value of 3.51 indicates that in at least 50% of the cases the prediction error is no more than 3.5 units.

The data distribution of Regularized RF shows a very similar picture to Baseline RF with a similarly strong scattering of the data around the very similarly inclined fitted line. The RMSE and R^2 values of Regularized RF are almost identical to those of Baseline RF, but the values of MAE and MedAE are marginally lower and therefore better. In Conditional inference forest, the data points are less scattered than in Baseline RF, but the fitted line is still fairly flatter. As a result, almost all metrics show a worse performance than Baseline RF, with Conditional inference forest only matching the Baseline model in the MedAE value. Extremely randomized trees shows a very similar picture to Conditional inference forest with less scatter and a flatter fitted line than the Baseline model. Here, too, the metrics are all worse than in the Baseline model, except for MedAE (3.37), which shows a small improvement. With Quantile regression forest, the picture of less scatter, flatter fitted line and worse metrics is repeated, except for MedAE (3.50). In Stratified sampling RF, the data are similarly strongly scattered around the similarly inclined fitted line as in Baseline RF. The metrics are also very comparable to Baseline RF. The data distribution of Cost-sensitive RF is again similarly scattered as in Baseline RF, although the metrics show a slightly poorer performance.

In general, it can be concluded that the vast majority of predicted data points fall within the range of 12 to 22, whereas the observed values range from below 5 to more than 30. Identifying the best model variant is not straightforward, as no single model variant clearly exhibits superior metrics. Regularized RF is selected as the best model variant within this model group for further analysis in this thesis, primarily because relatively large differences in the metrics MAE and MedAE can be observed compared to other well-performing models like Baseline RF or Stratified sampling RF.

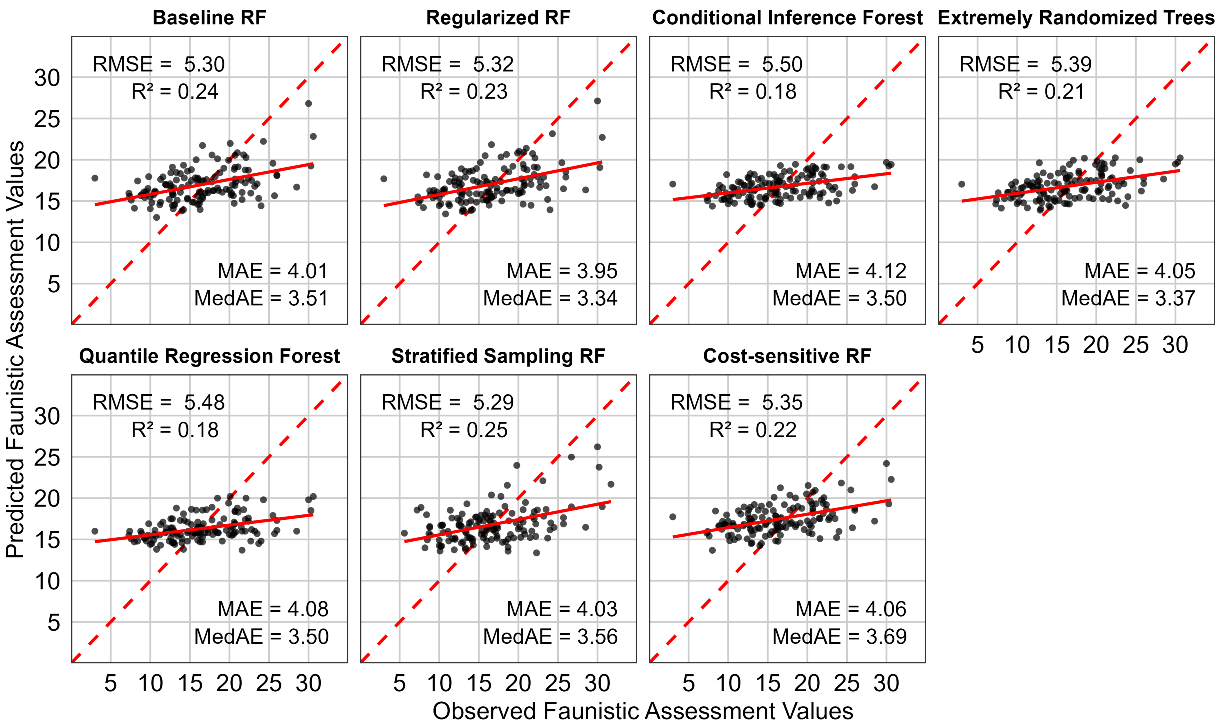


Figure 13: *Scatterplots for various random forest regression model variants predicting Result faunistic assessment based on spring structure data. The dashed red line represents the 1:1 line (perfect prediction), and the solid red line is a linear regression fit to the observed data points. Each panel reports RMSE, R^2, MAE, and MedAE as performance metrics.*

3.2.2 Classification Random Forest: Modelling Faunistic Assessment Values

The results of the different Classification Random Forest variants used to answer the first research question are shown in Figure 14. In these models, data from the structural survey according to the FOEN method were used to model classification into faunistic assessment classes. The Baseline Random Forest model has an accuracy of 0.43. This means that 43% of the instances were correctly classified in all classes, which is the best value of all model variants. The balanced accuracy of Baseline RF is 0.29, which is the tied second-best value of all models. This means that if all classes, regardless of their frequency, are included equally in the calculation of the average, only 29% of all instances are correctly classified. The F1-score for the classes *very non-spring* (1 spring) and *non-spring* (13 springs) is non-existent for the Baseline RF model, as none of these springs were assigned to the correct class by the Baseline model (see Appendix A13 for confusion matrices). The F1-score of the *spring-compliant* class (43 springs) for the Baseline model is 0.44, which is the second-best value of all classes. In the *conditionally spring-typical* class with 38 springs, the F1-score for the Baseline model is also 0.44, which is the tied second-best value of all models. In the *spring-typical* class with 40 springs a F1-score of 0.50 is identified, which is the third highest values of all models. In addition to the metrics in Figure 14, additional metrics (precision per class, recall per class) are listed in Appendix A14 for the different model variants.

	Accuracy	Balanced Accuracy	F1-Score very non-spring	F1-Score non-spring	F1-Score spring-compliant	F1-Score cond. spring-typ.	F1-Score spring-typ.
Baseline RF	0.43	0.29	–	–	0.44	0.44	0.50
Regularized RF	0.39	0.26	–	–	0.40	0.37	0.46
Cond. Inf. Forest	0.41	0.27	–	–	0.42	0.42	0.46
Extr. Random. Trees	0.41	0.28	–	–	0.40	0.44	0.45
Quantile Regr. Forest	0.42	0.29	–	–	0.49	0.48	0.29
Cost-sensitive RF	0.40	0.27	–	–	0.43	0.36	0.51
Class weighting	0.41	0.28	–	–	0.43	0.43	0.47
Dynamic weighting	0.39	0.26	–	–	0.44	0.37	0.43
weighted tree voting	0.21	0.34	0.03	0.14	0.43	0.20	0.02
threshold adjustment	0.26	0.19	–	0.12	0.44	0.28	–
up/downsampling	0.39	0.27	–	0.09	0.40	0.39	0.46

Figure 14: *Performance metrics table for various random forest classification model variants predicting fauna categories based on spring structural data. For each metric, the top three values are highlighted in dark green, green, and light green.*

The results of Regularized RF are quite similar to the Baseline model across all metrics, but always slightly worse. Regularized RF also does not recognise any of the springs in the *non-spring* and *very non-spring* classes. Conditional inference forest has the tied third best accuracy value of all model variants (0.41). Here too, the two worst fauna classes are not recognised, while all the other F1-scores are slightly worse than for the Baseline model. Extremely randomized trees also achieves the third-best accuracy value (0.41). All other metrics are also lower than the Baseline model, except for the F1-score of the *conditionally spring-typical* class, which is tied with the Baseline model value (0.44). Quantile regression forest shows very similar results to the Baseline model for accuracy and balanced accuracy. The F1-scores of the *spring-compliant* and the *conditionally spring-typical* class are the highest of all models, while the F1-score of the *spring-typical* class is very low (0.29) compared to Baseline RF. Cost-sensitive Random Forest shows slightly worse results than Baseline RF in all metrics except for the F1-score of the *spring-typical* class. Class weighting has the third best of all accuracy values at 0.41 and also shows a good balanced accuracy value (0.28). The F1-scores are only slightly worse than the scores of Baseline RF. With Dynamic weighting, all metrics show slightly worse model performance compared to the Baseline model. The Weighted tree voting model only recognises very few springs in the *spring-typical* and *conditionally spring-typical* classes, which leads to very poor accuracy. As comparatively good F1-score values can be determined in the two classes *non-spring* and *spring-compliant*, this results in the highest balanced accuracy value of all model variants. The Threshold adjustment model is not able to identify any springs in the *spring-typical* class, which leads to a low accuracy value. The F1-score of the *spring-compliant* class is tied second best for all models, but the balanced accuracy value is still comparatively low. Up/downsampling has a slightly worse performance than the Baseline model for all metrics except for the F1-score of the *non-spring* class where one of the 13 springs was correctly classified. Baseline RF is selected as the best model variant within this model group for further analysis in this thesis.

3.3 Modelling Ecological Value Based on Landscape & Topographic Variables

3.3.1 Regression Random Forest: Structural Assessment Values

The results of the different Regression Random Forest variants used to answer the first part of the second research question are shown in Figure 15. In these models, the landscape & topographic variables were used to model Result structural assessment. The observed structural assessment values are shown on the x-axis and the structural assessment values predicted by the models are shown on the y-axis. The median of the observed structural assessment values is 1.49, and the mean is 1.72.

It can be observed that all model variants exhibit more or less widely scattered point clouds, which, according to the fitted line, show a slight tendency toward accurate predictions but do not suggest particularly strong results. The vast majority of predicted data points lie between 1 and 2.5, whereas the observed values range from below 1 to over 4. Identifying the best model is relatively straightforward in this case, as Stratified sampling RF achieves the best result across all metrics. However, even this model explains only about 25% of the variance, and the mean absolute error is 0.55. With class intervals of 0.8, a mean absolute error of 0.55 means that predictions are often off by more than half a class. As a result, there is a high chance that a spring will be placed in the wrong class, possibly even more often than in the correct class.

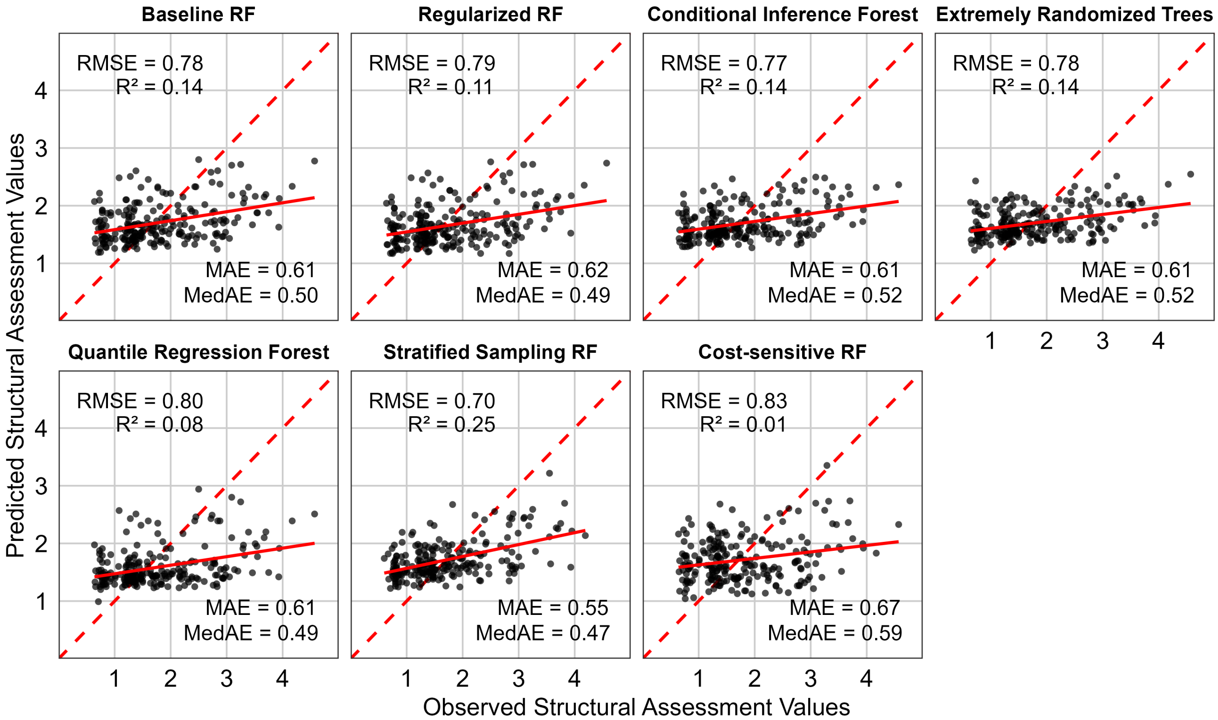


Figure 15: *Scatterplots for various random forest regression model variants predicting Result structural assessment based on landscape & topographic variables. The dashed red line represents the 1:1 line (perfect prediction), and the solid red line is a linear regression fit to the observed data points. Each panel reports RMSE, R^2 , MAE, and MedAE as performance metrics.*

3.3.2 Regression Random Forest: Faunistic Assessment Values

The results of the different Regression Random Forest variants used to answer the second part of the second research question are shown in Figure 16. In these models, the landscape & topographic variables were used to model Result faunistic assessment. The observed faunistic assessment values are shown on the x-axis and the faunistic assessment values predicted by the models are shown on the y-axis.

In this model group, also all model variants show, according to the fitted line, a slight tendency toward accurate predictions. However, none of the model variants demonstrates truly strong results in predicting Result faunistic assessment. Similarly to Figure 13, the vast majority of predicted data points lie between 12 and 25, whereas the observed values range from below 5 to over 30. Identifying the best model is again relatively straightforward in this case, as Regularized RF achieves the best metric results for R^2 , MAE, and MedAE, as well as the second-best RMSE value among all model variants. However, even this model explains only about 15% of the variance, and the mean absolute error is 4.33. With class intervals of 5, a mean absolute error of 4.33 means that predictions are often off by more than half a class. As a result, there is a very high chance that a spring will be placed in the wrong class, likely even more often than in the correct class.

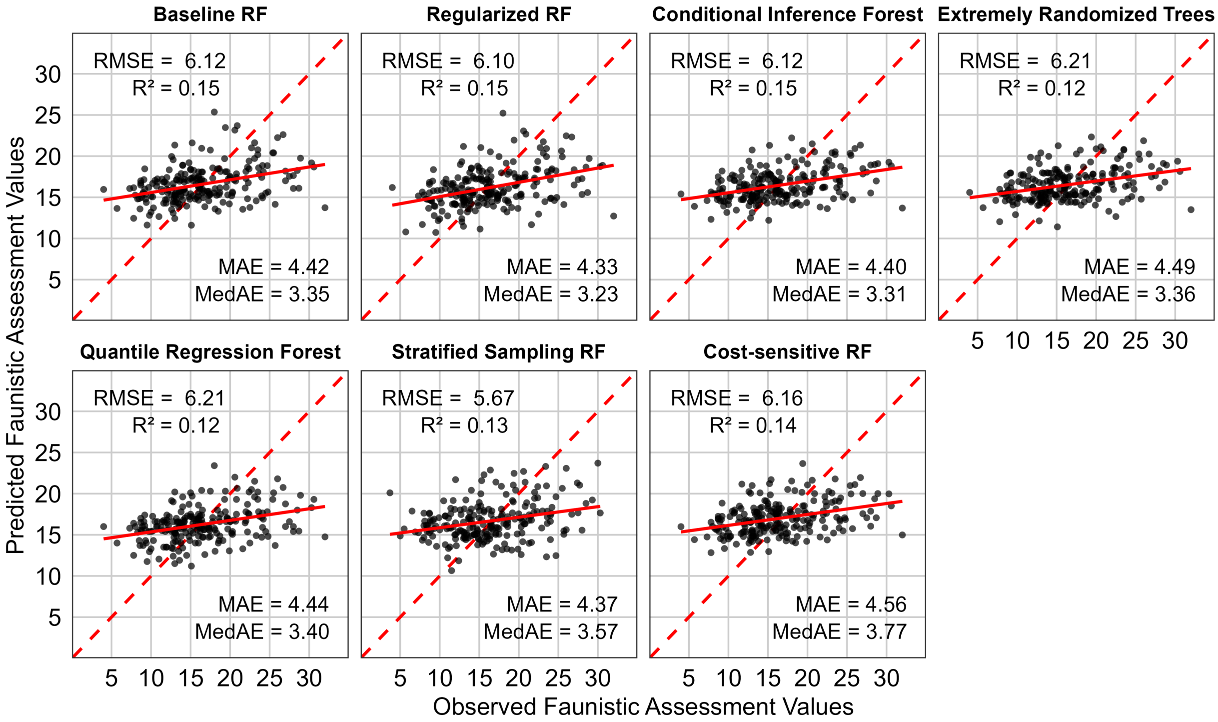


Figure 16: *Scatterplots for various random forest regression model variants predicting Result faunistic assessment based on landscape & topographic variables. The dashed red line represents the 1:1 line (perfect prediction), and the solid red line is a linear regression fit to the observed data points. Each panel reports RMSE, R^2, MAE, and MedAE as performance metrics.*

3.3.3 Classification Random Forest: Structural Assessment Values

The results of the different Classification Random Forest variants used to answer the first part of the second research question are shown in Figure 17. In these models, landscape & topographic variables were used to model the classification into structural assessment classes.

A comparison of the accuracy values for the different model variants shows that many models achieve similarly high values, with Quantile regression forest reaching the highest value (0.65). Some models, such as Class weighting or Dynamic weighting, are only able to correctly classify a very small number of springs and therefore show low accuracy values. Regarding balanced accuracy, models that are able to correctly assign all four classes (at least partially), such as Threshold adjustment and Up/downsampling, achieve good values. All other models are able to correctly assign at most three classes, leading to lower balanced accuracy values (see Appendix A15 for confusion matrices). Identifying the best model variant within this group is not entirely straightforward. While Up/downsampling achieves the highest balanced accuracy and a relatively good accuracy value, the F1-score for the largest class *natural* (0.69) is clearly lower than that of Quantile regression forest (0.80). In addition to having the highest accuracy, Quantile regression forest also achieves the third-best balanced accuracy value. For these reasons, this model variant is used as the best model within this model group in the remainder of this thesis. In addition to the metrics in Figure 17, additional metrics (precision per class, recall per class) are listed in Appendix A16 for the different model variants.

	Accuracy	Balanced Accuracy	F1-Score str. damaged	F1-Score damaged	F1-Score mod. impaired	F1-Score cond. natural	F1-Score natural
Baseline RF	0.57	0.21	–	–	0.11	0.09	0.74
Regularized RF	0.57	0.22	–	–	0.16	0.11	0.74
Cond. Inf. Forest	0.60	0.21	–	–	0.06	0.04	0.76
Extr. Random. Trees	0.57	0.20	–	–	0.06	0.06	0.74
Quantile Regr. Forest	0.65	0.25	–	–	0.17	0.29	0.80
Cost-sensitive RF	0.55	0.22	–	–	0.22	0.14	0.71
Class weighting	0.05	0.11	–	0.04	0.15	0.03	–
Dynamic weighting	0.05	0.12	–	0.05	0.13	0.04	–
weighted tree voting	0.60	0.20	–	–	–	–	0.75
threshold adjustment	0.47	0.29	–	0.17	0.21	0.40	0.60
up/downsampling	0.49	0.41	0.40	–	0.21	0.15	0.69

Figure 17: *Performance metrics table for various random forest classification model variants predicting structure categories based on landscape & topographic variables. For each metric, the top three values are highlighted in dark green, green, and light green.*

3.3.4 Classification Random Forest: Faunistic Assessment Values

The results of the different Classification Random Forest variants used to answer the second part of the second research question are shown in Figure 18. In these models, landscape & topographic variables were used to model the classification into faunistic assessment classes.

The accuracy values of all model variants fall within a relatively narrow range of 0.08. Conditional inference forest and Extremely randomized trees have the highest accuracy, both with a value of 0.40. For balanced accuracy, the values of the different model variants fall within an even narrower range of just 0.05. Looking at the F1-scores for the different classes, it is noticeable that very different model variants achieve good results in different classes, and there is not a single model variant that consistently produces strong values across the board (see Appendix A17 for confusion matrices). Identifying the best model variant within this group is again not entirely straightforward as the accuracy and balanced accuracy vary only slightly between model variants. The best model variant is considered to be Conditional inference forest. Although this model variant is only able to partially classify three out of five fauna classes correctly, the best accuracy value and the second-best balanced accuracy value speak in favour of this model. In addition to the metrics in Figure 18, additional metrics (precision per class, recall per class) are listed in Appendix A18 for the different model variants.

	Accuracy	Balanced Accuracy	F1-Score very non-spring	F1-Score non-spring	F1-Score spring-compliant	F1-Score cond. spring-typ.	F1-Score spring-typ.
Baseline RF	0.33	0.24	–	0.15	0.32	0.29	0.42
Regularized RF	0.33	0.23	–	0.08	0.32	0.28	0.44
Cond. Inf. Forest	0.40	0.27	–	–	0.46	0.27	0.53
Extr. Random. Trees	0.40	0.27	–	0.08	0.43	0.37	0.45
Quantile Regr. Forest	0.37	0.24	–	–	0.35	0.45	0.27
Cost-sensitive RF	0.38	0.26	–	–	0.47	0.16	0.48
Class weighting	0.34	0.24	–	0.08	0.34	0.29	0.45
Dynamic weighting	0.36	0.25	–	0.08	0.39	0.35	0.41
weighted tree voting	0.33	0.23	–	0.08	0.37	0.27	0.41
threshold adjustment	0.32	0.24	–	0.21	0.31	0.26	0.43
up/downsampling	0.35	0.28	–	0.22	0.36	0.19	0.51

Figure 18: *Performance metrics table for various random forest classification model variants predicting fauna categories based on landscape & topographic variables. For each metric, the top three values are highlighted in dark green, green, and light green.*

3.4 Combining Swiss Springs Data and Landscape & Topographic Variables

3.4.1 Regression Random Forest: Faunistic Assessment Values

The results of the different Regression Random Forest variants used to model Result faunistic assessment using mixed spring structure and landscape & topographic variables data are shown in Figure 19. The observed faunistic assessment values are shown on the x-axis and the faunistic assessment values predicted by the models are shown on the y-axis.

Even when the two datasets from research question 1 and research question 2 are combined, Figure 19 shows a very similar pattern to Figures 13 and 16. Here, too, the vast majority of predicted data points lie between 12 and 23, whereas the observed values range from below 5 to over 30, with no model variant showing more than a slight tendency toward accurate predictions. The best model variant in this model group is Cost-sensitive RF, mainly due to the lowest MAE values, the highest R^2 values, and the second lowest RMSE value among all model variants. While the best model variant in this group still has relatively poor RMSE and R^2 values compared to the other faunistic regression model groups, Cost-sensitive RF does show some of the best results for MAE and MedAE.

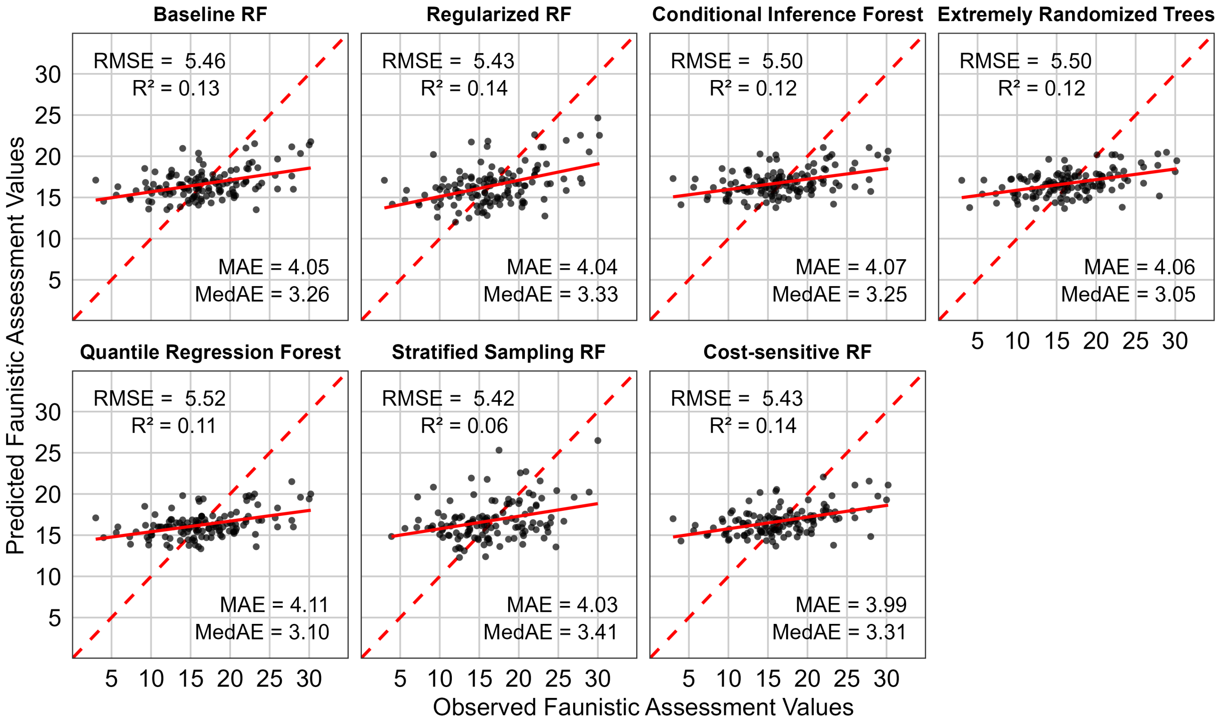


Figure 19: *Scatterplots for various random forest regression model variants predicting Result faunistic assessment based on mixed spring structure and landscape & topographic variables data. The dashed red line represents the 1:1 line (perfect prediction), and the solid red line is a linear regression fit to the observed data points. Each panel reports RMSE, R^2 , MAE, and MedAE as performance metrics.*

3.4.2 Classification Random Forest: Faunistic Assessment Values

The results of the different Classification Random Forest variants used to model Result faunistic assessment using mixed spring structure and landscape & topographic variables data are shown in Figure 20.

The accuracy values of all model variants fall within a relatively narrow range of 0.10, while for balanced accuracy, the range is larger at 0.26. Only one model variant is able to partially classify more than three fauna classes correctly, namely Threshold adjustment (see Appendix A19 for confusion matrices). The best model variant within this group is Class weighting, as this variant achieves both the highest accuracy and the highest balanced accuracy of all model variants. These strong results are due to the fact that Class weighting also achieves the highest F1-scores for the *conditionally spring-typical* and *spring-typical* classes, and additionally shows a comparatively good F1-score for the *spring-compliant* class. In addition to the metrics in Figure 20, additional metrics (precision per class, recall per class) are listed in Appendix A20 for the different model variants.

	Accuracy	Balanced Accuracy	F1-Score very nonspring	F1-Score nonspring	F1-Score spring-compliant	F1-Score cond. spring-typ.	F1-Score spring-typ.
Baseline RF	0.43	0.33	–	–	0.38	0.48	0.49
Regularized RF	0.36	0.28	–	–	0.42	0.34	0.38
Cond. Inf. Forest	0.38	0.42	–	–	0.36	0.41	0.42
Extr. Random. Trees	0.37	0.29	–	–	0.26	0.43	0.46
Quantile Regr. Forest	0.35	0.21	–	–	0.22	0.49	0.24
Cost-sensitive RF	0.40	0.31	–	–	0.39	0.44	0.43
Class weighting	0.46	0.47	–	–	0.38	0.51	0.56
Dynamic weighting	0.42	0.43	–	–	0.31	0.50	0.48
weighted tree voting	0.39	0.40	–	–	0.34	0.44	0.45
threshold adjustment	0.38	0.28	–	0.08	0.45	0.25	0.53
up/downsampling	0.43	0.28	–	–	0.36	0.49	0.55

Figure 20: *Performance metrics table for various random forest classification model variants predicting fauna categories based on mixed spring structure and landscape & topographic variables data. For each metric, the top three values are highlighted in dark green, green, and light green.*

3.5 Most Important Modelling Variables

The results of the permutation of the most important variables for each regression model group, which were used to answer the third research question, are shown in Figure 21. Plot a) within Figure 21 shows, on the x-axis, the increase in RMSE that occurs when only the respective variable is permuted. On the y-axis, the ten most important predictor variables of the best model variant (RRF) for modelling Result faunistic assessment using spring structure data are displayed and are highlighted in yellow on the plot. At the top is always the most important variable, the variable for which permutation results in the largest deterioration of the RMSE value. In the case of plot a), the most important predictor variable is Elevation, as permuting this variable increases the RMSE value of 5.32 by more than 0.5. Water temperature, Number of structures, and Discharge also show increases in RMSE of more than 0.05. All other predictor variables, such as Average Flow Velocity, Shrubs, etc., show only a very small increase in RMSE for RRF. All additional predictor variables that are not listed among these top ten predictors have even smaller increases in RMSE, or in some cases, even lead to a decrease in RMSE, a slight improvement of the model. This clearly demonstrates that only very few predictor variables have a quantifiable influence on the model. It is evident that Elevation is not only the most important predictor variable for RRF but also for all other model variants. However, for the remaining predictor variables, greater differences between the model variants become apparent, indicating that the top ten predictor variables and their ranking vary between the model variants.

Plot b) shows the permutation importance of the most important predictor variables of the best model variant (RRF) for modelling Result faunistic assessment using landscape & topographic data. Land use, followed by Lithology, Soil type, and Water permeability, are the most important predictors, each with an increase in RMSE of more than 0.1. From Slope and Soil skeleton content onwards, the increases in RMSE are again very close to zero, indicating that these variables are of little importance to the model. The most important predictor of the RRF model variant is also among the most important predictors in most of the other model variants. However, for example, Lithology is the most important predictor in Stratified sampling RF, while Water permeability is most important in Cost-sensitive RF.

Plot c) shows the permutation importance of the most important predictor variables of the best model variant (SSRF) for modelling Result structural assessment using landscape & topographic data. In this case, the x-axis had to be adjusted, as the scales of the faunistic and structural assessments are not directly comparable. Land use and Lithology, as in plot b), are also the most important predictors here. Other important variables include Distance to streets, Agricultural suitability, Slope, Soil type, and others. It is clearly evident that Stratified sampling RF found the largest increase in RMSE for nearly all predictor variables. Only in a few cases such as RRF for Distance to streets or CSRF for Plan curvature did other model variants assign a higher increase in RMSE to a given variable.

Plot d) shows the permutation importance of the most important predictor variables of the best model variant (CSRF) for modelling the Result faunistic assessment using mixed spring structure and data from landscape & topographic variables. As in plot a), Elevation is the most important predictor, with an increase in RMSE of more than 0.1. Other key predictor variables include Terrain roughness index, Water temperature, Spring stream length, and others. In general, the increases in RMSE are lower here than in plots a) and b). The heterogeneity of the increase in RMSE values between the different model variants is relatively high for all predictors. While Elevation is the most important predictor variable in most model variants, the permutation importances of the remaining variables vary considerably between models.

Comparing the most important fauna predictor variables of the mixed dataset in plot d) with the variables of the separate datasets in plots a) and b) reveals several interesting observations. The seven most important predictor variables from plot d) are also present in plots a) and b). However, it is notable that not only the most important predictor variables from plots a) and b) reappear in plot d). Moreover, the order of importance does not always match; for example, Lithology ranks ahead of Terrain roughness index in plot b), whereas the order is reversed in plot d).

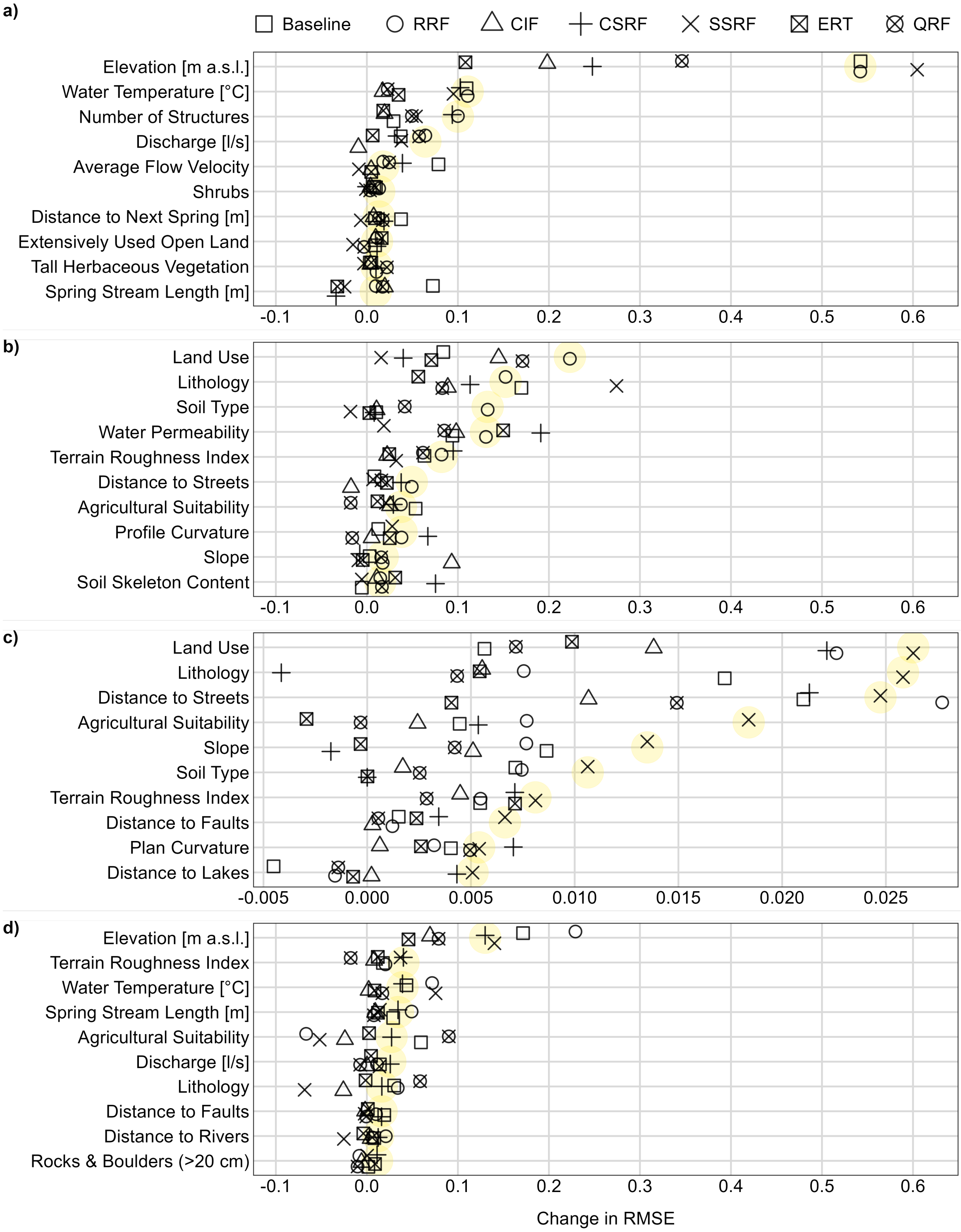


Figure 21: *Permutation importances of the top ten predictor variables in different regression model variants. Panels a) and b) show results for predicting faunistic assessment values using either spring structure data (a) or landscape & topographic variables (b). Panel c) displays the top predictors for structural assessment values based on landscape & topographic variables. Panel d) presents results for predicting faunistic assessment values using mixed spring structure and landscape & topographic variables data. Yellow highlights indicate the best model variant in each group. Note that the x-axis in c) is scaled differently. Positive values indicate an increase in RMSE (greater importance).*

The results of the permutation of the most important variables for each classification model group, which were used to answer the third research question, are shown in Figure 22. Plot a) within Figure 22 displays, on the x-axis, the decrease in balanced accuracy observed when only the respective variable is permuted. On the y-axis, the ten most important predictor variables of the best model variant (Baseline RF) for modelling fauna class membership using spring structure data are shown and highlighted in yellow in the plot. In plot a), Flow direction is the most important predictor variable, followed by Elevation, Sand, Terrain Gradient, and others. It is evident that Baseline RF found the largest decrease in balanced accuracy for almost all predictor variables. Only in one case (QRF for Elevation) did another model variant assign a higher decrease in balanced accuracy to a variable. Notably, Flow direction is not the most important predictor variable in all model variants; in some model variants, an increase in balanced accuracy is even observed when the variable is permuted.

Plot b) shows the permutation importance of the most important predictor variables of the best model variant (QRF) for modelling the classification into fauna classes using landscape & topographic variables data. Distance to streets, Terrain roughness index, and Soil skeleton content are the most important predictors. The lowest-ranking predictors within the top ten, Lithology, Slope, and Distance to lakes, show only very minor decreases in balanced accuracy values, indicating that they are of little importance for the prediction. Interestingly, in a surprisingly large number of cases, permuting the predictor variables in model variants other than QRF leads to an increase in balanced accuracy, i.e., an improvement of the model.

Plot c) shows the permutation importance of the most important predictor variables of the best model variant (CIF) for modelling the classification into structure classes using landscape & topographic data. Water permeability, Land use, and Slope are the most important predictors, while all other predictors are already very close to zero and have little influence on the model. In plot c), even the Random parameter, which was included in every Random Forest modelling as a control, appears among the top ten predictors.

Plot d) shows the permutation importance of the most important predictor variables of the best model variant (CWRF) for modelling the classification into fauna classes using mixed spring structure and landscape & topographic variables data. Here, the most important variables are Discharge, Trampling damage, and Distance to lakes. The decreases in balanced accuracy are lower here than in plot a), which was already observed in Figure 21. In particular, the lower-ranking variables within the top ten do not approach 0 but rather -0.01, suggesting that more variables have a quantifiable influence on the model compared to plot b or c.

Comparing the most important fauna predictor variables of the mixed dataset from plot d) with the variables of the separate datasets from plots a) and b) again reveals several interesting observations. In plot d), only two of the twenty predictor variables from plots a) and b), namely Distance to lakes and Distance to streets, reappear. All other top ten predictors in plot d) are not among the top ten in plots a) or b).

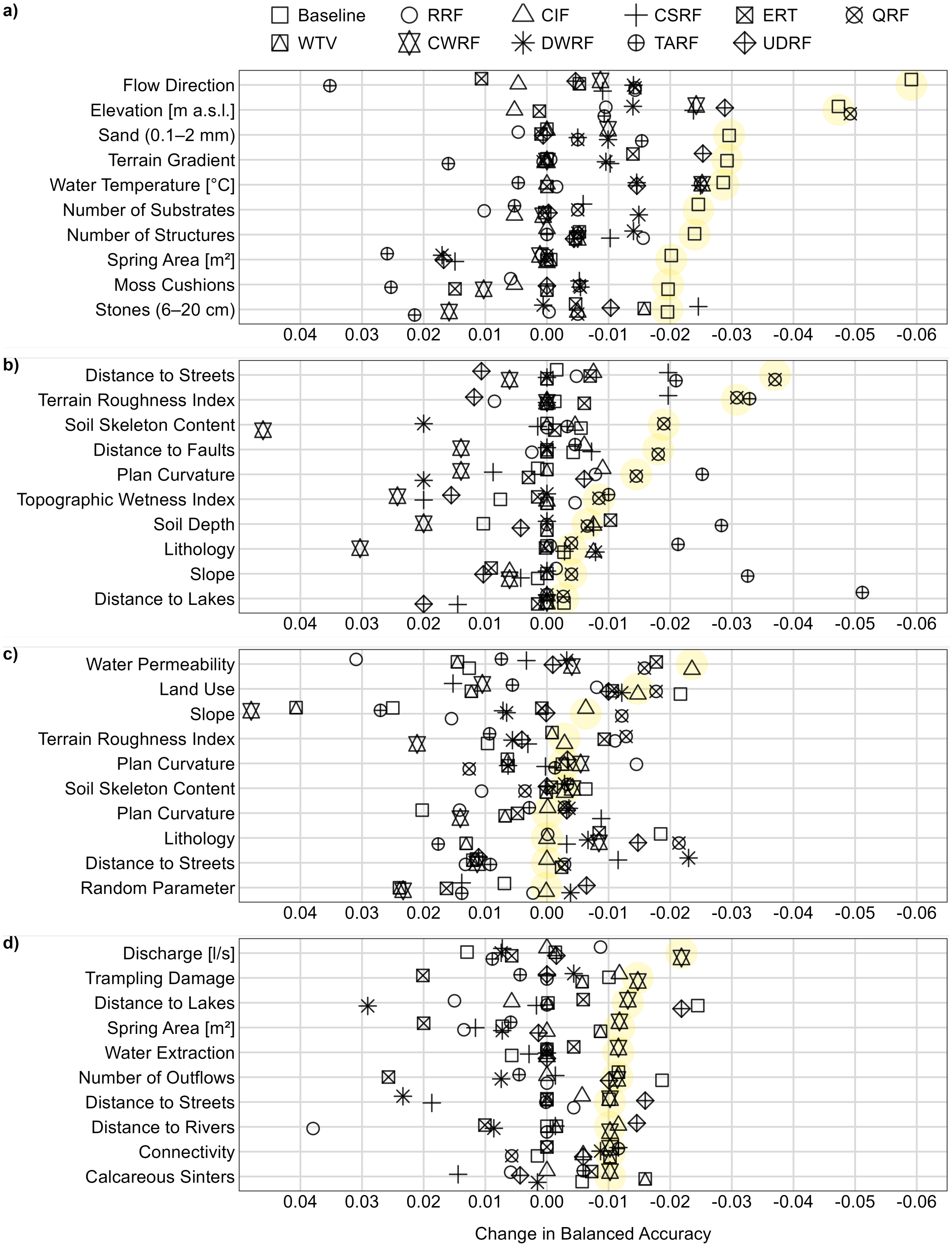


Figure 22: *Permutation importances of the top ten predictor variables in different classification model variants. Panels a) and b) show results for predicting fauna classes using either spring structure data (a) or landscape & topographic variables (b). Panel c) displays the top predictors for structure class affiliation based on landscape & topographic variables. Panel d) presents results for predicting fauna classes using mixed spring structure and landscape & topographic variables data. Yellow highlights indicate the best model variant in each group. Negative values indicate a decrease in Balanced Accuracy (greater importance).*

4. Discussion

4.1 Modelling Faunistic Condition Based on Spring Structure Data (RQ1)

A central question in this study was whether the faunistic condition of a spring can be modelled based on the results of structural assessments using the FOEN method (research question 1). The results of the different regression model variants all demonstrate that the Random Forest approach does not succeed in predicting the faunistic assessment based on structural spring investigations (see Figure 13). The predicted faunistic assessment values tend to cluster more strongly around the median or mean of the observed values for all model variants, rather than reflecting the observed distribution. As a result, the fit in all cases is very flat, with the models rarely predicting either low or high faunistic assessment values. A flat distribution of points, as observed in Figure 13, would not necessarily be problematic if models were able to correct for this, provided the points are relatively tightly grouped. Unfortunately, however, for all model variants, the data points are also relatively widely scattered, in some cases more so than in others. This scatter makes it impossible for the models to consistently assign lower predicted values to springs with lower observed faunistic assessment values, or higher predicted values to those with higher observed values. This scattered distribution results in high RMSE and MAE values, indicating large average prediction errors, along with a low R^2 value, meaning that only a small portion of the total variance is explained. On a positive note, there is at least a slight tendency towards correct predictions. This suggests that the predictions are not purely random but that certain parameters do exhibit a positive predictive effect, even if the results are not as clear as hoped.

A comparison with other Regression Random Forest studies provides insight into how the results of this thesis relate to modelling outcomes achieved in comparable ecological contexts. In the study by Olaya-Marín et al. (2013), data from 90 sampling sites in three Mediterranean rivers in Spain were used to model the diversity of native fish species using random forests. The models were based on a wide range of environmental variables, which are fairly comparable to the structural data used in this thesis, including physicochemical, hydromorphological, and biological parameters. The achieved model performances for their Random Forest approach showed a coefficient of determination of $R^2 = 0.68$. In another comparable study by Derot et al. (2020), long-term Lake Geneva data was used to model the intensity of cyanobacterial blooms with Random Forests. The data base consisted of more than 30 years of regular measurements of physicochemical parameters and cyanobacterial concentrations, thus representing a similar data structure to that used in the present thesis. Their best models reached a regression coefficient of $R^2 = 0.61$. Comparisons with the results of the studies Olaya-Marín et al. (2013) and Derot et al. (2020) demonstrate that relatively high model quality could be achieved in these cases. By comparison, the results of the present thesis suggest that considerably lower model performance can be expected. In summary, it should be noted that, in regression models, the faunistic condition of a spring cannot be predicted using structural characteristics.

The results of the different classification model variants all show that the Random Forest approach does not succeed in predicting fauna class affiliation based on structural spring investigations (see Figure 14). All model variants share the observation that Random Forest appears to be unable to predict all five fauna classes simultaneously. None of the model variants managed to correctly classify at least one spring in the five fauna classes. This is, of course, strongly related to the fact that there is only one spring in the *very non-spring* class. Only Weighted tree voting, Threshold adjustment, and Up-/downsampling manage to correctly classify at least one spring in four out of five classes. However, it appears that for these model variants, there is a trade-off when predicting the *very non-spring* or *non-spring* classes. In these models, the prediction performance for the higher fauna classes drops sharply, as the models have to concentrate their predictive power on the lower fauna classes. In all other models, correct predictions are only observed in the top three fauna classes. As these upper fauna classes contain the majority of the springs, the model variants that concentrate on predicting the upper three classes show the highest accuracy values. However, the accuracy values alone are not very meaningful due to the differing number of springs in each class, as classes with few springs are under-represented. The balanced accuracy values, which are intended to correct exactly this problem of unequal class sizes, are themselves highly dependent on whether the two lowest fauna classes can be predicted or not. The results for balanced accuracy are thus skewed by the very few springs in the *very non-spring* and *non-spring* classes. This is particularly evident in Weighted tree voting, as this model variant achieves the highest balanced accuracy value, even though only very few springs in the two largest upper classes are correctly classified. One possibility would have been to merge the *non-spring* and *very non-spring* classes in order to harmonise the class sizes, but this was not done due to lack of comparability with the data analysis and incompatibility with the FOEN method.

The following comparison with other Classification Random Forest studies in comparable ecological settings provides additional perspective to interpret the performance of the models used in this thesis. In the study by Cheng et al. (2012), data from six shallow lakes in the middle Yangtze Basin in China were used to model fish communities and their diversity using Random Forests. The basis was 117 datasets with a wide range of biotic and abiotic environmental variables, which makes the dataset comparable to the structural data used in this thesis. The achieved model performance for site assignment to fish communities had a mean accuracy of 0.74 for the Random Forest approach. These values indicate that a higher model quality was achieved in the mentioned study than in the present thesis. Thus, as already observed with the regression models, the challenges of transferring ecological modelling approaches to other ecological research questions or habitats become evident. In summary, it must be stated that, in classification models, the fauna classes cannot be predicted using structural characteristics.

4.2 Modelling Ecological Value Based on Landscape & Topographic Variables (RQ2)

Another central question in this study was whether the ecological condition (e.g. structural and faunistic condition) of a spring can be modelled based on landscape & topographic variables not addressed in the FOEN method (research question 2). The results of the different regression model variants all demonstrate that the Random Forest approach does not succeed in predicting the structural and faunistic assessment values based on landscape & topographic variables as the existing variation in the data cannot be captured by the models (see Figure 15 & Figure 16). The predicted structural and faunistic assessment values tend to cluster more strongly around the median or mean of the observed values for all model variants, rather than reflecting the observed distribution. All model variants that model structural assessment show a relatively sharp lower boundary in the predicted values, especially in contrast to the predictions of the faunistic assessment (see Figure 15 & Figure 16). This lower boundary is found, for all variants, somewhere between 1 and 1.2, with only a few predicted values falling below this range. However, towards higher values, there is no such distinct boundary. The results of the structural assessment data are also widely scattered, reflected in a low R^2 value. Although a slight tendency towards correct predictions can be observed, meaning the predictions are not entirely random.

The results of the modelling of the faunistic assessment with landscape & topographic variables (see Figure 16) and the modelling of the faunistic assessment with structural data (see Figure 13) are very similar. Here as well, many predicted values are close to the median of the observed values, and only a slight tendency toward correct predictions can be observed amidst widely scattered data. Interestingly, there are relatively large differences in the model performance between the two groups of models. For example, the RMSE values in the models using landscape & topographic variables are notably worse than in the models using structural assessment data, higher by between 0.5 and 0.8 points. The same pattern appears for the R^2 values, which are between 0.03 and 0.12 lower in the landscape & topographic variables models than in the structural assessment models, and for the MAE values, which are higher by between 0.34 and 0.50. Surprisingly, this pattern does not continue for MedAE. Almost all MedAE values are considerably lower in the landscape & topographic variables models than in the structural assessment models, indicating that as model performance decreases, predicted values are increasingly concentrated around the median of the observed values, thus reducing the median absolute error.

A comparison with other Regression Random Forest studies provides insight into how the results of this thesis relate to modelling outcomes achieved with comparable ecological data. In the study by Cabezas et al. (2016), data from 44 sampling plots in a Chilean wetland were used to model the diversity of vascular plants using Random Forests. The predictors were textural variables derived from Landsat satellite data, which are therefore quite comparable to the landscape & topographic variables used in this thesis. The final model achieved a prediction quality of R^2 = 0.60. These results show that with texture-based predictors from remote sensing data, relatively high model performance could be achieved in the studied wetland, whereas the results of the present thesis indicate lower model quality. In summary, it should be noted that, in regression models, the affiliation of springs to structure and fauna classes cannot be predicted using landscape & topographic variables.

The results of the different classification model variants all show that the Random Forest approach does not succeed in predicting structure and fauna class assignment based on landscape & topographic variables (see Figure 17 & Figure 18). None of the different Classification Random Forest approaches using landscape & topographic variables is able to correctly classify at least one spring in all five classes. It appears that, for the Random Forest models, there is once again a trade-off between poor results in the upper structure classes and the ability to predict springs in the lower structure classes. Only Threshold adjustment and up/downsampling manage to correctly classify at least one spring in four classes, while for all other model variants, this is the case for a maximum of three classes. For Class weighting and Dynamic weighting, it is noticeable that although three different classes were at least partially classified correctly, the class with by far the most springs (*natural*) was not predicted at all. The accuracy values are generally higher than for the modelling of the fauna classes in Figures 14 and 18; this is because, for the structure classes, an even higher proportion of springs are found in the highest classes compared to the fauna classes. For the model, it is essentially easier to produce good results when the focus is on the class with the most springs (*natural*). The balanced accuracy values are lower than in the modelling of the fauna classes, since this data imbalance makes the correct assignment to the various classes even more difficult.

The results of the modelling of fauna class affiliation with landscape & topographic variables (see Figure 18) and the modelling of fauna class affiliation with structural data (see Figure 14) are quite similar. When using the landscape & topographic variables, it appears that more model variants are able to at least partially correctly classify four out of five classes than was possible when using the structuralsimilarly data. In terms of accuracy, the best models differ only slightly between the two data approaches; overall, across all model variants, the accuracy values are slightly higher when using structural data. The same pattern is seen for balanced accuracy . Although when using landscape & topographic variables, many model variants could partially correctly classify four classes, the balanced accuracy values are lower than in the models using structural data. This is again due to trade-off, where the reduction in prediction quality for the higher fauna classes is necessary to at least partially correctly classify the *non-spring* class.

The following comparison with other Classification Random Forest studies using comparable ecological data provides additional perspective to interpret the performance of the models used in this thesis. In the study by Tian et al. (2025), Random Forests were used to investigate the selection of the Grey Heron colony site during the breeding season at 100 colony sites in China.

The predictors were detailed land use variables (such as water bodies, croplands, forests, and impervious surfaces), which are comparable to the landscape & topographic variables used in this thesis. The models achieved an average classification accuracy of 0.82. The results show that high model quality can be achieved to predict colony site selection when extensive and differentiated environmental data are available. In the study by Martínez-Santos et al. (2021), the occurrence of aquatic ecosystems in the western catchment of the Cauca River in Colombia was mapped using Random Forests. The data base consisted of 350 sampling points, for which various DEM-derived variables and other environmental variables were calculated, making the approach comparable to that used in this thesis. For the Random Forest model, a very high prediction quality was achieved, with an accuracy of 0.93. Comparisons with the results of the studies Olaya-Marín et al. (2013) and Derot et al. (2020) demonstrate that in both studies the quality of the model was significantly higher than in the present thesis. This confirms that good results in ecological modelling can only be achieved when well-adapted predictors are available and that it is not trivial to transfer such approaches to other research questions. In summary, it should be noted that, in classification models, the affiliation of springs to structure and fauna classes cannot be predicted using landscape & topographic variables.

4.3 Most Important Modelling Variables (RQ3)

4.3.1 Random Forest Variable Importance

The first two research questions lead to the question of which variables are most important for identifying ecologically valuable springs (research question 3). It is important to note that the selection of the most important predictor variables for the regression models (see Figure 21) is highly dependent on which model variants the predictor variables are shown for. It is very likely that each model variant would identify a different set of the most important predictor variables. Some model variants appear to be able to extract more information from the variables for certain research questions than others. This can be clearly seen in the data points for SSRF in plot c) of Figure 21, where SSRF consistently shows the highest permutation changes in RMSE for almost all predictor variables. However, it remains unclear where the boundary lies between predictors with a certain quantifiable importance and those that have ended up in the top ten more or less by chance. This can be observed very well in plots a), b), and d) of Figure 21, where many of the lower-ranking variables trigger almost no change in RMSE upon permutation, yet still make it into the top ten of the respective variable groups. This is also the case in plot c), but due to the changed scale resulting from the structural assessment, the differences simply appear larger than they actually are in comparison. Thus, it remains unclear whether the models are only operating with information from very few top predictors, or whether the moderate model results arise from the combined information of many predictor variables, each with only a very small influence. Where the line is drawn between truly important and negligible predictor variables cannot be answered. Nevertheless, there are some predictor variables that very likely have a real influence on the model results. Elevation is certainly among them, as the change in RMSE in Figure 21 plot a) is very large compared to the other predictor variables, and nearly as large in all other model variants. In addition, Elevation is by far the most important predictor variable when modelling faunistic values with the mixed data in plot d). Other predictor variables that appear in the top ten both in plot a) (modelling faunistic values with structural data) and in plot d) (modelling faunistic values with mixed data) include Water temperature, Discharge, and Spring Stream Length. In plots b) and c), where landscape & spring variables were used for modelling structure and faunistic values, Land use and Lithology must be mentioned, as they are the most and second most important predictor variables, respectively. Additional variables that appear in the top ten in both model groups are Distance to streets, Soil type, Agricultural suitability, Terrain roughness index, and Slope. Predictor variables from the landscape & topographic variables group that also appear in the top ten in plot d) with the mixed data are Lithology, Agricultural suitability, and Terrain roughness index. It is interesting to note that Land use does not appear in the top ten in plot d) with the mixed data, and Lithology is ranked much further behind and seems to have less influence in this model variant than expected.

The most important predictor variables of the structural data classification model are Flow direction and Elevation (see Figure 22a), as the change in balanced accuracy compared to all other predictor variables in the top ten is very large. Spring area is the only predictor variable from the structural data that appears in the top ten both in plot a) (modelling fauna classes with structural data) and in plot d) (modelling fauna classes with mixed data). In plots b) and c), where environmental and spring variables were used for modelling structure and fauna classes, Terrain roughness index, Soil skeleton content, Distance to streets, Plan curvature, Slope, and Lithology appear in the top ten most important predictor variables for both model groups. Distance to streets is also the only predictor variable from the landscape & topographic variables that, in addition to plots b) and c), also appears in the top ten in plot d) with the mixed data. It is particularly interesting to observe that in plot c), the random parameter, which was included as a control variable in all model variants, is listed among the ten most important predictor variables. This means that at least the remaining 14 landscape & topographic variables do not provide any information value to the model. However, since other variables show similarly low changes in balanced accuracy as the random parameter, it must be assumed that, for this model variant, the model performance depends primarily on just a few top predictors.

In conclusion, it can be stated that the identification of the most important predictor variables is not trivial and varies considerably between model variants and depending on the research question. A non-exhaustive list of the most important predictor variables from the structural data includes Elevation, Water temperature, Discharge, Spring stream length, Flow direction, and Spring area. A non-exhaustive list of the most important predictor variables from the landscape & topographic variables includes Land use, Lithology, Agricultural suitability, Distance to streets, and Terrain roughness index.

4.3.2 Comparison to Dataset Analysis

The data analysis in section 3.1 was performed to familiarise with the dataset, to identify dependencies between individual variables, and ultimately to allow a comparison between the most important predictor variables determined by the random forest models and the information that can be drawn from the exploratory data analysis.

For the predictor variables Elevation, Number of structures, and Water temperature, the Kruskal-Wallis tests revealed significantly different values between the fauna classes (see Table 2). As expected, these three variables are also among the most important predictors for modelling faunistic values using structural assessment data (see Figure 21a), as well as among the most important predictors for modelling fauna classes using structural assessment data (see Figure 22a). For these variables, it can be confirmed that significant differences between the different classes provide the model with the ability to correctly classify the springs. However, for the variables Spring stream length, Distance to next spring, and Discharge, no significant differences between the fauna classes were identified in the Kruskal-Wallis test. However, these variables are still found to be among the top ten most important predictors of the modelling of faunistic value using structural assessment data. This is surprising, as one would expect that significant differences between classes are necessary for a variable to have predictive value; however, this is obviously not always the case, and certain model variants are able to extract relevant information from the data regardless.

Substantial differences in the data distribution between the structure and fauna classes can be observed for both Elevation and Number of structures in the boxplots of Figure 5. As expected, these two predictor variables are also found among the top ten in their respective model variants using structural assesssment data (see Figure 21a & Figure 22a). However, for Number of substrates, however, the differences in data distribution between the different classes are not very pronounced according to the boxplots, yet this predictor variable still ranks among the top ten most important predictors in the regression model using structural assessment data. Number of flow types exhibits differences in the boxplots comparable to those of Elevation and Number of structures, but this variable does not appear in any of the top ten predictor lists of the Random Forest model variants.

In the correlation matrix (see Figure 7), none of the top ten variables listed in Figure 21a & Figure 22a exhibited particularly high correlation values with Result faunistic assessment. The highest correlation values were found for Elevation, Water temperature, and Number of structures, which also represent the top three most important predictor variables in the regression model using structural assessment data. For the remaining numerical structural variables in Figure 21a (Discharge, Distance to next spring, and Spring stream length) only very low correlation values with Result faunistic assessment were observed in the correlation matrix. Thus, the results of the permutation importance analyses largely confirm the patterns already apparent from the correlation matrix with remarkable accuracy. Therefore, correlation with the target variable appears to be a very good indicator to identify the most important predictor variables.

The t-SNE plot with Result faunistic assessment (see Figure 10) indicates the clustering of variables in part A of the structural assessment, but according to the colour scale, there does not appear to be a clear correlation with Result faunistic assessment. For the t-SNE plot of part B of the structural data, no clusters could be detected in the data. Therefore, it is even more remarkable that several variables from part B of the structural assessment are among the ten most important predictor variables in Figure 21a and Figure 22a, while no variable in part A appears in these top ten lists. This clearly demonstrates that, in heterogeneous datasets, t-SNE plots alone cannot reliably predict which groups of variables will ultimately be more important than others.

In general, it can be concluded that the data analysis and the Random Forest models identified only partially overlapping sets of predictor variables. Although there is a notable correspondence for certain variables, particularly those showing significant differences between classes and high correlation with the target variable, the Random Forest models also highlight predictors whose relevance was not apparent from exploratory analyses alone. This discrepancy underscores the added value of machine learning approaches in uncovering complex, multidimensional relationships that may not be detected through classical statistical methods. At the same time, the observed overlaps help confirm which predictors are truly important, as shown by both the data analysis and the models. Ultimately, a combination of both approaches provides the most comprehensive understanding of the drivers of faunistic and structural conditions in springs.

4.4 Methodological Reflection & Data Basis

4.4.1 Reliability and Robustness of RF Models

All Random Forest model variants used in this study worked reliably and could be applied to the dataset without major problems. The robustness of the results across the different model variants was also high, as all variants produced very similar results. For most model types, the computation time during model runs was very low (a few seconds), while, for example, Conditional inference forest required much more time to run (several minutes). In larger projects with multiple model runs, this could be an important criterion when selecting the model variant. Surprisingly, the Baseline Random Forest often did not produce clearly worse results than the other model variants. Despite the methodological improvements or adjustments introduced by the other variants, Baseline RF was sometimes even the best-performing model. This clearly shows that the choice of Random Forest model variant cannot be the main reason for the moderately successful modelling outcomes. If refined methods do not provide improvements over the simplest Random Forest model, the limitation is likely due to the input data used rather than the modelling approach itself.

4.4.2 Alternative Model Variants Not Used

In this study, various Random Forest model variants were applied; however, other variants or approaches could certainly have been used as well. Feature engineering is a widely used technique to unlock hidden potential in a dataset. The model results are likely to be improved by feature engineering, but this approach was not implemented in this study. Since feature engineering leads to changes in the data and variable structure, identifying the most important predictor variables would have been more difficult - an important part of the research question in this thesis. With a standard Random Forest approach, it is clear and traceable how variables and data are handled, whereas feature engineering alters the data, which should be avoided due to the importance of maintaining transparency in data origin and generation according to the FOEN method.

Other Random Forest model variants that might have produced promising results, since they extend or go beyond the approaches used in this thesis, include Totally randomized trees (such as in Geurts et al. (2006)), Tree weighting based on error for minority classes (as in Devi et al. (2019)), as well as global refinement and global pruning approaches (as in Ren et al. (2015)). Furthermore, methods such as Iterative joint feature pruning, tree number optimisation ((Paul et al. 2018)), Stacking-based weighted Random Forest ((Shahhosseini & Hu 2020)), or Modified balanced random forest ((Agusta & Adiwijaya 2019)) could have provided valuable new insights, but were not applied due to time constraints. Furthermore, an entirely different methodological approach outside the Random Forest spectrum, such as hybrid models or other approaches introduced in subsection 1.5.1, could have yielded better results than the Random Forest approach used in this thesis.

4.4.3 Reflection on the Data Used

The structural data from the Swiss springs dataset were predefined, with only a few adjustments made before the implementation of the Random Forest models. For the modifications carried out, such as the removal of a few variables due to insufficient entries and the aggregation of vegetation data to 16 instead of 52 columns, care was always taken to ensure an improvement or at least no deterioration in the predictive performance of the Random Forest models. This was consistently verified using simple baseline models.

The landscape & topographic variables were generated within the framework of this study, with selection criteria being simple availability, relatively straightforward calculation, and potential significance for the ecological value of a spring. Thus, the data to be used had to be publicly available and should not exceed a reasonable data volume. Furthermore, calculations were intended to be relatively simple in QGIS, since this was not intended to be the main methodological focus of the thesis. If more time had been available for the calculation and generation of landscape & topographic variables, more variables and different approaches could have been tested, which might have been beneficial. Consequently, the focus was placed on the most important and potentially meaningful variables. The potential significance for ecological value of a spring is, of course, fairly vague and subjective; nevertheless, boundaries were set for which variables should or should not be considered. In the studies by Chen et al. (2018), Naghibi et al. (2016), Naghibi & Pourghasemi (2015), Rahmati et al. (2018), Zabihi et al. (2016) variables such as stream density, NDVI, solar radiation, sediment transport index, stream power index, and relative slope position were used, together with other variables, to model the occurrence of springs in the landscape. Due to their lack of clear relevance for modelling the ecological value of a spring, these variables were not included in this thesis. In addition, some similar variables are already part of the predictor variables set. For the generation of DEM-derived variables, DEM data with a resolution of 2 metres were used. DEM data would also have been available in finer or coarser resolutions, which could have affected the predictor variable and thus the model performance. With a stronger focus on landscape & topographic variables, it would have been possible to calculate variables from different DEM resolutions and compare their predictive performance in the Random Forest models. For certain variables that, for example, strongly depend on very local flow paths, such as the TWI or Convergence index, differences in predictive quality might have emerged. Consequently, the extent to which DEM resolution influences the predictive power of the variables in the applied models remains unresolved.

4.5 Limitations and Future Work

4.5.1 Data and Sampling Limitations

Although the Swiss springs dataset is very extensive in terms of the number of variables included, the distribution of the springs between the structure and fauna classes represents a major limitation for this study. The investigated springs are predominantly restricted to those with good faunistic and very good structural conditions. Data were not collected exclusively for research purposes; many springs were surveyed due to the potential for subsequent protection. As a result, spring habitats considered ecologically valuable were investigated more frequently than springs that, due to poor structure and fauna assessments, would not be eligible for protection. This leads to a strong bias in the dataset, making it very difficult for the Random Forest models to predict poorly rated springs, simply because there is insufficient training data for such cases. Although the sample size in the Swiss springs dataset is generally large, with more than 750 springs, a substantial number of springs are located in alpine elevations, while only a few are from intermediate elevation ranges (see Figure 1). This imbalance makes it difficult to interpret the model results, especially for Elevation, since not all elevation ranges are equally represented in the dataset and therefore in the models. Moreover, the spring surveys on which the data are based were conducted over more than 20 years and by more than 50 different individuals, suggesting some variation in methodology and introduces a degree of subjective uncertainty, further limiting data comparability.

For the Random Forest modelling, the dataset had to be reduced to 442 springs, as only these had complete data for all variables. The alternative would have been to exclude variables with missing data from the models, which would have allowed the use of the entire dataset of more than 750 springs. It is possible that the larger training dataset, even with fewer predictor variables, might have led to better model results. Another limitation of the dataset arises from the FOEN methodology itself, which, by its nature, cannot cover all the variables that are important for the ecology of springs. For example, there are no data on the connectivity between springs, the prevailing microclimate, the age of the spring, or on flow variability and flow permanence.

4.5.2 Future Research and Practice

One of the objectives of this study was to determine whether the structural parameters of the dataset can be used to predict the faunistic quality of spring habitats. Although this goal could not be satisfactorily achieved, the results show that, given optimal conditions regarding both methods and data, it should be possible to generate some promising modelling outcomes. Another objective was to assess whether additional landscape & topographic variables could be used to predict the quality of spring habitats, i.e., to model the ecological value of spring habitats at known points in the landscape. This goal was clearly not met; however, the study did demonstrate that certain variables, which are not part of the official FOEN method, may still provide indications for the presence of ecologically valuable springs. Unfortunately, the applied conservation goal of this work to provide an additional tool for spring inventory that would simplify and make the process more efficient was not achieved. The scientific goal, namely to expand or adapt existing spring modelling approaches for probability of occurrence to instead model the ecological value of springs, was implemented in principle, but the main finding was that it remains challenging to model the ecological value of spring habitats using the available and newly generated data.

Although the ecological requirements of spring-dwelling species, their habitat needs, and many of their interactions are relatively well known, it currently does not seem possible to predict their occurrence, and, by extension, the ecological value of springs, based solely on structural information about their habitat. Future studies may help identify the most important predictor variables and the most effective modelling approaches for conservation purposes. To make this possible, it will be necessary to include more springs with poor structure and fauna results in the dataset in order to better train the models. It will also be important to ensure a more balanced dataset with respect to elevation, land use categories, and other relevant factors to enable more robust and transferable modelling.

Despite these suggestions for improvement, it should also be acknowledged that it may simply not be possible to model ecological value in the way required by Swiss legislation. Since springs are highly complex habitats with highly specialized and often isolated inhabitants, it is possible that natural processes of this complexity and at this scale may simply not be suitable for modelling.

5. Conclusion

This thesis has addressed the critical decline of spring habitats across Switzerland, highlighting that these habitats are among the most threatened aquatic ecosystems. To protect spring habitats effectively, an ecological assessment of their value is necessary. However, the existing evaluation procedure, the FOEN method, is resource-intensive and costly. Simplifying or even partially substituting this evaluation with a modelling approach could facilitate the protection process significantly. Previous research has mainly focused on modelling the occurrence of springs and identifying variables predicting their occurrence, rather than on modelling their ecological value (e.g., structural and faunistic conditions). Therefore, a new approach to modelling the ecological quality of springs is required.

The aim of this study was to determine whether structural parameters from the FOEN method can predict the faunistic quality of spring habitats and to identify the most important predictors involved. Hence, this thesis sought to model the ecological value of known spring habitats. The second objective was to investigate if landscape and topographic variables could predict the ecological value of springs and to identify the key predictors. Consequently, this study aimed to explore whether ecological value predictions can be made about spring occurrences within the landscape. Ultimately, the goal of this work was to expand existing spring modelling methods, traditionally focused on occurrence probabilities, to include the ecological value of springs, thereby providing an additional tool for spring inventory in Switzerland.

Seven regression and eleven classification Random Forest model variants were applied to model structural and faunistic assessment values according to the FOEN method and to classify springs into corresponding structure and fauna categories. The most significant predictor variables were identified using permutation importance within Random Forest models and were compared with key predictors identified in the dataset analysis. However, neither structural parameters from the FOEN method nor landscape and topographic variables yielded strong modelling results for predicting ecological values or class affiliations. All model variants produced similar outcomes, unable to surpass merely indicating correct predictive tendencies. Consequently, identifying the most influential predictors proved challenging, although some variables with relatively high predictive quality were still successfully identified.

Despite these limitations, this study successfully demonstrated the application of Random Forest models using data derived from the FOEN method to assess spring habitats. Although overall model results were not fully satisfactory, specific predictors with clear relevance for indicating ecological value were identified. Additionally, the findings suggest that successful modelling of the faunistic quality of springs requires additional, yet unknown, structural parameters, indicating the need for an expansion or adaptation of the current FOEN methodology. Several limitations should be noted, including challenges posed by imbalanced class distributions, dominated by many highly-rated springs, fewer poorly-rated springs, and a high number of springs located in alpine regions. Furthermore, the dataset contained missing values and was collected over 20 years by more than 50 different individuals. While existing spring modelling approaches were successfully expanded to incorporate ecological values, the unsatisfactory modelling outcomes prevented the creation of a practical additional tool for spring inventory in Switzerland. Thus, structural parameters from the FOEN method could not reliably predict the faunistic condition of springs, and similarly, landscape and topographic variables failed to model their ecological values accurately.

For practical implications, it is recommended that the data basis be expanded to include springs across all elevations, land-use categories, and notably lower-rated springs. Furthermore, integrating additional ecological structure factors or enhanced landscape and topographic variables may improve model performance. Future research should focus on identifying critical structural predictor variables that define the ecological value of spring habitats. Investigating and applying alternative modelling approaches, such as hybrid models or neural networks, could potentially enhance predictive accuracy. Finally, assessing the robustness and generalizability of these models across different regions or related ecosystems would be worthwhile.

Bibliography

Agusta, Z. P. & Adiwijaya (2019), 'Modified balanced random forest for improving imbalanced data prediction', *International Journal of Advances in Intelligent Informatics* **5**, 58–65.

Al-Fugara, A., Pourghasemi, H. R., Al-Shabeeb, A. R., Habib, M., Al-Adamat, R., Al-Amoush, H. & Collins, A. L. (2020), 'A comparison of machine learning models for the mapping of groundwater spring potential', *Environmental Earth Sciences* **79**.

Ayalew, L., Yamagishi, H. & Ugawa, N. (2004), 'Landslide susceptibility mapping using gis-based weighted linear combination, the case in tsugawa area of agano river, niigata prefecture, japan', *Landslides* **1**, 73–81.

BAFU (2017), Aktionsplan des bundesrates. aktionsplan strategie biodiversität schweiz, Technical report, BAFU. https://www.bafu.admin.ch/dam/bafu/de/dokumente/biodiversitaet/fachinfo-daten/aktionsplan-strategie-biodiversitaet-schweiz.pdf.download.pdf/Aktionsplan_SBS_final_Deutsch.pdf.

BAFU (2021), Quell-lebensräume erfassen – erhalten – aufwerten. eine arbeitsgrundlage für die praxis, Technical report, BAFU. https://www.bafu.admin.ch/bafu/de/home/themen/biodiversitaet/publikationen-studien/publikationen/quell-lebensraeume.html.

Blattner, L., Lucek, K., Beck, N., Berner, D. & von Fumetti, S. (2022), 'Intra-alpine islands: Population genomic inference reveals high degree of isolation between freshwater spring habitats', *Diversity and Distributions* **28**, 291–305.

Bonettini, A. M. & Cantonati, M. (1996), 'Macroinvertebrate assemblages of springs of the river sarca catchment (adamello-brenta regional park, trentino, italy)', *Crunoecia* **5**, 71–78.

Breiman, L. (2001), Random forests, Technical report.

Cabezas, J., Galleguillos, M. & Perez-Quezada, J. F. (2016), 'Predicting vascular plant richness in a heterogeneous wetland using spectral and textural features and a random forest algorithm', *IEEE Geoscience and Remote Sensing Letters* **13**, 646–650.

Cantonati, M., Füreder, L., Gerecke, R., Jüttner, I. & Cox, E. J. (2012), 'Crenic habitats, hotspots for freshwater biodiversity conservation: Toward an understanding of their ecology', *Freshwater Science* **31**, 463–480.

Cantonati, M., Gerecke, R. & Bertuzzi, E. (2006), 'Springs of the alps - sensitive ecosystems to environmental change: From biodiversity assessments to long-term studies'.

Cantonati, M., Lichtenwöhrer, K., Leonhardt, G., Seifert, L., Mustoni, A., Hotzy, R., Schubert, E., Blattner, L., Bilous, O., Lotz, A., Poschlod, B. & Gerecke, R. (2022), 'Using springs as sentinels of climate change in nature parks north and south of the alps: A critical evaluation of methodological aspects and recommendations for long-term monitoring'.

Cantonati, M. & Ortler, K. (1998), 'Using spring biota of pristine mountain areas for long term monitoring using spring biota of pristine mountain areas for long-term monitoring', *Water Resources and Ecology in Headwaters* . https://www.researchgate.net/publication/236145405.

Cantonati, M., Poikane, S., Pringle, C., Stevens, L., Turak, E., Heino, J., Richardson, J., Bolpagni, R., Borrini, A., Cid, N., Tvrtlikova, M., Galassi, D., Hajek, M., Hawes, I., Levkov, Z., Naselli-Flores, L., Saber, A., DiCicco, M., Fiasca, B., Hamilton, P., Kubecka, J., Segadelli, S. & Znachor, P. (2020c), 'Characteristics, main impacts, and stewardship of natural and artificial freshwater environments: Consequences for biodiversity conservation', *Water (Switzerland)* **12**.

Cantonati, M., Segadelli, S., Spitale, D., Gabrieli, J., Gerecke, R., Angeli, N., Nardo, M. T. D., Ogata, K. & Wehr, J. D. (2020a), 'Geological and hydrochemical prerequisites of unexpectedly high biodiversity in spring ecosystems at the landscape level', *Science of the Total Environment* **740**.

Cantonati, M., Stevens, L. E., Segadelli, S., Springer, A. E., Goldscheider, N., Celico, F., Filippini, M., Ogata, K. & Gargini, A. (2020b), 'Ecohydrogeology: The interdisciplinary convergence needed to improve the study and stewardship of springs and other groundwater-dependent habitats, biota, and ecosystems'.

Carrington, A. M., Manuel, D. G., Fieguth, P. W., Ramsay, T., Osmani, V., Wernly, B., Bennett, C., Hawken, S., Magwood, O., Sheikh, Y., Mcinnes, M. & Holzinger, A. (2023), 'Deep roc analysis and auc as balanced average accuracy, for improved classifier selection, audit and explanation', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **45**, 329–341.

Cartwright, J. M., Dwire, K. A., Freed, Z., Hammer, S. J., McLaughlin, B., Misztal, L. W., Schenk, E. R., Spence, J. R., Springer, A. E. & Stevens, L. E. (2020), 'Oases of the future? springs as potential hydrologic refugia in drying climates', *Frontiers in Ecology and the Environment* **18**, 245–253.

Chen, W., Li, H., Hou, E., Wang, S., Wang, G., Panahi, M., Li, T., Peng, T., Guo, C., Niu, C., Xiao, L., Wang, J., Xie, X. & Ahmad, B. B. (2018), 'Gis-based groundwater potential analysis using novel ensemble weights-of-evidence with logistic regression and functional tree models', *Science of the Total Environment* **634**, 853–867.

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 (Switzerland)* **10**.

Cheng, L., Lek, S., Lek-Ang, S. & Li, Z. (2012), 'Predicting fish assemblages and diversity in shallow lakes in the yangtze river basin', *Limnologica* **42**, 127–136.

Conoscenti, C., Ciaccio, M., Caraballo-Arias, N. A., Álvaro Gómez-Gutiérrez, Rotigliano, E. & Agnesi, V. (2015), 'Assessment of susceptibility to earth-flow landslide using logistic regression and multivariate adaptive regression splines: A case of the belice river basin (western sicily, italy)', *Geomorphology* **242**, 49–64.

Das, A., Abdel-Aty, M. & Pande, A. (2009), 'Using conditional inference forests to identify the factors affecting crash severity on arterial corridors', *Journal of Safety Research* **40**, 317–327.

Delarze, R., Eggenberg, S., Steiger, P., Bergamini, A., Fivaz, F., Gonseth, Y., Guntern, J., Hofer, G., Sager, L. & Stucki, P. (2016), Rote liste der lebensräume der schweiz. aktualisierte kurzfassung zum technischen bericht 2013 im auftrag des bundesamtes für umwelt bafu, Technical report, BAFU. https://www.infoflora.ch/de/assets/content/documents/download/CH_RL_Lebensraeume_2017_v171130.pdf.

Deng, H. & Runger, G. (2012), 'Feature selection via regularized trees'. http://arxiv.org/abs/1201.1587.

Derot, J., Yajima, H. & Jacquet, S. (2020), 'Advances in forecasting harmful algal blooms using machine learning models: A case study with planktothrix rubescens in lake geneva', *Harmful Algae* **99**.

Devi, D., Biswas, S. K. & Purkayastha, B. (2019), A cost-sensitive weighted random forest technique for credit card fraud detection, Technical report.

Dumnicka, E., Galas, J. & Koperski, P. (2007), 'Benthic invertebrates in karst springs: Does substratum or location define communities?', *International Review of Hydrobiology* **92**, 452–464.

Fernández-Martínez, M., Barquín, J., Bonada, N., Cantonati, M., Churro, C., Corbera, J., Delgado, C., Dulsat-Masvidal, M., Garcia, G., Margalef, O., Pascual, R., Peñuelas, J., Preece, C., Sabater, F., Seiler, H., Zamora-Marín, J. M. & Romero, E. (2024), 'Mediterranean springs: Keystone ecosystems and biodiversity refugia threatened by global change'.

Friedman, J. H. & Meulman, J. J. (2003), 'Multiple additive regression trees with application in epidemiology', *Statistics in Medicine* **22**, 1365–1381.

Fumetti, S. V. & Nagel, P. (2012), 'Discharge variability and its effect on faunistic assemblages in springs', *Freshwater Science* **31**, 647–656.

Geissler, P. (1976), 'Zur vegetation alpiner fliessgewässer', *Beiträge zur Kryptogamenflora der Schweiz* . https://www.e-periodica.ch/digbib/view?pid=cry-001:1976:14::239#113.

Gerecke, R., Cantonati, M., Spitale, D., Stur, E. & Wiedenbrug, S. (2011), 'The challenges of long-term ecological research in springs in the northern and southern alps: Indicator groups, habitat diversity, and medium-term change', *Journal of Limnology* **70**, 168–187.

Gerecke, R., Franz, H. & Cantonati, M. (2009), 'Invertebrate diversity in springs of the national park berchtesgaden (germany): relevance for long-term monitoring', *Internationale Vereinigung für theoretische und angewandte Limnologie: Verhandlungen* **30**, 1229–1233.

Geurts, P., Ernst, D. & Wehenkel, L. (2006), 'Extremely randomized trees', *Machine Learning* **63**, 3–42.

Guisan, A. & Thuiller, W. (2005), 'Predicting species distribution: Offering more than simple habitat models', *Ecology Letters* **8**, 993–1009.

Ilmonen, J., Mykrä, H., Virtanen, R., Paasivirta, L. & Muotka, T. (2012), 'Responses of spring macroinvertebrate and bryophyte communities to habitat modification: Community composition, species richness, and red-listed species', *Freshwater Science* **31**, 657–667.

Ilmonen, J., Paasivirta, L., Virtanen, R. & Muotka, T. (2009), 'Regional and local drivers of macroinvertebrate assemblages in boreal springs', *Journal of Biogeography* **36**, 822–834.

Imesch, C. & Küry, D. (2024), 'Quell-lebensräume – eine einführung und aktuelle herausforderungen', *URP Umweltrecht in der Praxis* . https://www.quell-lebensrÃď'ume.ch/images/publikationen/URP_02-24-komp.pdf.

Jaafarzadeh, M. S., Tahmasebipour, N., Haghizadeh, A., Pourghasemi, H. R. & Rouhani, H. (2021), 'Groundwater recharge potential zonation using an ensemble of machine learning and bivariate statistical models', *Scientific Reports* **11**.

Kamp, G. V. D. (1995), 'The hydrogeology of springs in relation to the biodiversity of spring fauna: A review', *Source: Journal of the Kansas Entomological Society* **68**, 4–17. https://about.jstor.org/terms.

Knoll, L., Breuer, L. & Bach, M. (2019), 'Large scale prediction of groundwater nitrate concentrations from spatial data using machine learning', *Science of the Total Environment* **668**, 1317–1327.

Kubíková, L., Simon, O. P., Tichá, K., Douda, K., Maclak, M. & Bílý, M. (2012), 'The influence of mesoscale habitat conditions on the macroinvertebrate composition of springs in a geologically homogeneous area', *Freshwater Science* **31**, 668–679.

Kumar, S. & Pati, J. (2022), 'Assessment of groundwater arsenic contamination using machine learning in varanasi, uttar pradesh, india', *Journal of Water and Health* **20**, 829–848.

Kumari, S., Kumar, D., Kumar, M. & Pande, C. B. (2023), 'Modeling of standardized groundwater index of bihar using machine learning techniques', *Physics and Chemistry of the Earth* **130**.

Küry, D. (2009), Quellen als lebensräume 81 quellen als lebensräume, Technical report, Baselbieter Heimatbuch 27: 81–93. https://www.quell-lebensrÃď'ume.ch/images/publikationen/2009QuellenLebensrBHB27_081-093.pdf.

Küry, D., Lubini-Ferlin, V. & Stucki, P. (2019), Quell-lebensräume - anleitung zur systematischen erhebung und ermittlung ihrer bedeutung im naturschutz. expertenbericht im auftrag des bundesamtes für umwelt bafu., Technical report, BAFU. https://www.quell-lebensrÃď'ume.ch/images/publikationen/2019Kury_et_al_Quell-Lebensraume_Anleitung_Erfassung_Bewertung_2019.pdf.

LfU (2023), Quellen - einzigartig und schützenswert, Technical report, Bayerisches Landesamt für Umwelt (LfU).

Lubini-Ferlin, V., Stucki, P., Vicentini, H. & Küry, D. (2014), Bewertung von quell-lebensräumen in der schweiz. entwurf für ein strukturelles und faunistisches verfahren. bericht im auftrag des bundesamtes für umwelt bafu., Technical report, BAFU. https://www.quell-lebensrÃď'ume.ch/images/publikationen/Quellbewertung_v2_D_20181012.pdf.

Martínez-Santos, P., Aristizábal, H. F., Díaz-Alcaide, S. & Gómez-Escalonilla, V. (2021), 'Predictive mapping of aquatic ecosystems by means of support vector machines and random forests', *Journal of Hydrology* **595**.

More, A. S. & Rana, D. P. (2017), *1st International Conference on Intelligent Systems and Information Management : ICISIM-2017 : proceedings : 5-6 October 2017, Aurangabad, Maharashtra, India*, IEEE.

Naghibi, S. A., Ahmadi, K. & Daneshi, A. (2017), 'Application of support vector machine, random forest, and genetic algorithm optimized random forest models in groundwater potential mapping', *Water Resources Management* **31**, 2761–2775.

Naghibi, S. A., Hashemi, H., Berndtsson, R. & Lee, S. (2020), 'Application of extreme gradient boosting and parallel random forest algorithms for assessing groundwater spring potential using dem-derived factors', *Journal of Hydrology* **589**.

Naghibi, S. A. & Pourghasemi, H. R. (2015), 'A comparative assessment between three machine learning models and their performance comparison by bivariate and multivariate statistical methods in groundwater potential mapping', *Water Resources Management* **29**, 5217–5236.

Naghibi, S. A., Pourghasemi, H. R. & Dixon, B. (2016), 'Gis-based groundwater potential mapping using boosted regression tree, classification and regression tree, and random forest machine learning models in iran', *Environmental Monitoring and Assessment* **188**, 1–27.

Nekola, J. C. (1999), 'Paleorefugia and neorefugia: The influence of colonization history on community pattern and process', *Ecology* **80**, 2459–2473.

Nhu, V. H., Hoa, P. V., Melgar-García, L. & Bui, D. T. (2023), 'Comparative analysis of deep learning and swarm-optimized random forest for groundwater spring potential identification in tropical regions', *Remote Sensing* **15**.

Olaya-Marín, E. J., Martínez-Capel, F. & Vezza, P. (2013), 'A comparison of artificial neural networks and random forests to predict native fish species richness in mediterranean rivers', *Knowledge and Management of Aquatic Ecosystems* .

Paul, A., Mukherjee, D. P., Das, P., Gangopadhyay, A., Chintha, A. R. & Kundu, S. (2018), 'Improved random forest for classification', *IEEE Transactions on Image Processing* **27**, 4012–4024.

Peterka, T., Hájková, P., Jiroušek, M., Hinterlang, D., Chytrý, M., Aunina, L., Deme, J., Lyons, M., Seiler, H., Zechmeister, H., Apostolova, I., Beierkuhnlein, C., Bischof, M., Biţă-Nicolae, C., Brancaleoni, L., Ćušterevska, R., Dengler, J., Didukh, Y., Dítě, D., Felbaba-Klushyna, L., Garbolino, E., Gerdol, R., Iemelianova, S., Jansen, F., Juutinen, R., Kamberović, J., Kapfer, J., Klímová, B., Knollová, I., Kolari, T. H., Lazarević, P., Luostarinen, R., Mikulášková, E., Đorđije Milanović, Miserere, L., Moeslund, J. E., Molina, J. A., Pérez-Haase, A., Petraglia, A., Puglisi, M., Ruprecht, E., Šmerdová, E., Spitale, D., Tomaselli, M., Vassilev, K. & Hájek, M. (2023), 'Formalized classification of the class montio-cardaminetea in europe: towards a consistent typology of spring vegetation', *Preslia* **95**, 347–383.

Pešić, V., Savić, A., Jabłońska, A., Michoński, G., Grabowski, M., Bańkowska, A. & Zawal, A. (2019), 'Environmental factors affecting water mite assemblages along eucrenon-hypocrenon gradients in mediterranean karstic springs', *Experimental and Applied Acarology* **77**, 471–486.

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: Comprising the capability and robustness of three different modeling approaches', *Journal of Hydrology* **565**, 248–261.

Reiss, M. & Chifflard, P. (2018), 'Different forest cover and its impact on eco-hydrological traits, invertebrate fauna and biodiversity of spring habitats', *Nature Conservation* **27**, 85–99.

Ren, S., Cao, X., Wei, Y. & Sun, J. (2015), Global refinement of random forest, Technical report.

Sashikkumar, M. C. & Colins, J. (2017), 'Gis based groundwater modeling study to assess the effect of artificial recharge: A case study from kodaganar river basin, dindigul district, tamil nadu', *Geological Society of India* **89**, 57–64. https://www.researchgate.net/publication/314282581.

Scarsbrook, M. R., Barquin, J. & Gray, D. (2007), *New Zealand coldwater springs and their biodiversity*, Dept. of Conservation.

Seiler, H., Küry, D., Billeter, R. & Dengler, J. (2021), 'Regional typology of spring vegetation in parc ela (grisons, switzerland)', *Vegetation Classification and Survey* **2**, 257–274.

Shahhosseini, M. & Hu, G. (2020), Improved weighted random forest for classification problems, Technical report.

Staudacher, K. & Füreder, L. (2007), 'Habitat complexity and invertebrates in selected alpine springs (schütt, carinthia, austria)', *International Review of Hydrobiology* **92**, 465–479.

Stevens, L. E., Schenk, E. R. & Springer, A. E. (2021), 'Springs ecosystem classification', *Ecological Applications* **31**.

Stucki, P., Ryser, S., Lubini-Ferlin, V., Küry, D. & Kunz, S. (2015), Quellen im fokus, Technical report, aqua viva Die Zeitschrift für Gewässerschutz. https://www.quell-lebensrÃďume.ch/images/publikationen/2015Aqua_Viva_Zeitschrift_3_2015.pdf.

Svetnik, V., Liaw, A., Tong, C., Culberson, J. C., Sheridan, R. P. & Feuston, B. P. (2003), 'Random forest: A classification and regression tool for compound classification and qsar modeling', *Journal of Chemical Information and Computer Sciences* **43**, 1947–1958.

Sáez, J. A. & Romero-Béjar, J. L. (2022), 'Impact of regressand stratification in dataset shift caused by cross-validation', *Mathematics* **10**.

Tao, H., Hameed, M. M., Marhoon, H. A., Zounemat-Kermani, M., Heddam, S., Sungwon, K., Sulaiman, S. O., Tan, M. L., Sa'adi, Z., Mehr, A. D., Allawi, M. F., Abba, S. I., Zain, J. M., Falah, M. W., Jamei, M., Bokde, N. D., Bayatvarkeshi, M., Al-Mukhtar, M., Bhagat, S. K., Tiyasha, T., Khedher, K. M., Al-Ansari, N., Shahid, S. & Yaseen, Z. M. (2022), 'Groundwater level prediction using machine learning models: A comprehensive review', *Neurocomputing* **489**, 271–308.

Taxböck, L., Linder, H. P. & Cantonati, M. (2017), 'To what extent are swiss springs refugial habitats for sensitive and endangered diatom taxa?', *Water (Switzerland)* **9**.

Tian, R., Li, D., Zhang, S., Zhao, Y. & Yi, G. (2025), 'Colony site selection of gray heron (ardea cinerea) during the breeding period at multiple spatial scales', *Ecology and Evolution* **15**.

Weigand, E. (1998), Limnologisch-faunistische charakterisierung von karstquellen, quellbächen und unterirdischen gewässern nach choriotopen und biozönotischen gewässerregionen (nationalpark o.ö. kalkalpen, Österreich), Technical report, Nationalpark o.ö. Kalkalpen.

Yang, X. F. (2025), 'Mapping desert shrub aboveground biomass in the junggar basin, xinjiang, china using quantile regression forest (qrf)', *PeerJ* **13**.

Zabihi, M., Pourghasemi, H. R., Pourtaghi, Z. S. & Behzadfar, M. (2016), 'Gis-based multivariate adaptive regression spline and random forest models for groundwater potential mapping in iran', *Environmental Earth Sciences* **75**.

Zapata-Rios, X., Brooks, P. D., Troch, P. A., McIntosh, J. & Guo, Q. (2016), 'Influence of terrain aspect on water partitioning, vegetation structure and vegetation greening in high-elevation catchments in northern new mexico', *Ecohydrology* **9**, 782–795.

Zaresefat, M. & Derakhshani, R. (2023), 'Revolutionizing groundwater management with hybrid ai models: A practical review', *Water (Switzerland)* **15**.

Zhu, M., Xia, J., Jin, X., Yan, M., Cai, G., Yan, J. & Ning, G. (2018), 'Class weights random forest algorithm for processing class imbalanced medical data', *IEEE Access* **6**, 4641–4652.

Appendix

A1 Spring Protocol Structure

Quellen Protokoll - Struktur | **Kanton :** VD | **ID :** 001_VD

Quelle:	La Diey	Datum :	08.06.2007	Koordinaten (X/Y) :	524580 / 172075
Flurname :	Romainmôtier	Höhe ü.M. :	680	BearbeiterIn (leg) :	Pascal Stucki

KOPFDATEN *(nicht bewertet, nur Infos)* *! Skizze / Bemerkungen / Gefährdung / Massnahmen => auf der Rückseite (wird gescannt) !* *Ausfüllen oder zutreffendes ankreuzen* x

Austrittsform (Liste)	Sturzquelle	**Quelle** (Grösse [m²])	100	**Vernetzung**	Einzelquelle [ ] Q-system [ ] Q-komplex [x]
Hanglage	Mittelhang	**Quellbereich** [m²]	200	Dist. zur Nachbarquelle (m)	1000 — Anz. Austritte 3
Abflussrichtung	NO	**Quellbachlänge** [m]	40	Bemerkungen	
Geländeeignung	mässig	**Wassertemperatur** [°C]	8.5		
Quellschüttung	ganzjährig	**Quellschüttung** [l/s]	10	Fotos und andere Dokumente [x]	ID 001_VD_20070608_Q_FOTO_1.jpg
mittl. Fliessgesch.	mässig	**Leitfähigkeit** [µS20/cm]		Trinkwassernutzung [ ]	Schutzstatus [ ] Kulturhistorische Bedeutung [ ]

Bewertung Teil A : Beeinträchtigung *Zutreffendes mit "1" markieren* 1

Einträge/Verbau

Fassung	neu	alt	verfallen
Brunnenstube mit Überlauf			
Rohr und Becken			
nur Rohr/Rinne		1	
keine			

Wasserentnahme	>60%	30-59%	<30% / unbekannt	keine
			1	

Bemerkung / Zweck : Wasserversorgung Fischzucht

Verlegung	10-100m	<10m	unbekannt
alt			
neu			
keine			1

Aufstau	nach <10m	nach>=10-49m	unbekannt
Hauptschluss, 1-5 m2		1	
Hauptschluss, >5 m2			
Nebenschluss			
kein			

künstlicher Absturz	Gesamtabfluss	Teilabfluss	nein
			1

Verbau* (Ufer, Sohle)	stark	mittel	gering
Holz			
Steinschüttung			
wilder Verbau			
Naturstein		1	
Beton			
Verrohrung			
kein			

* nur erste Fliessmeter

Unterhalt/Trittschäden	gering	mässig	stark	keine
	1			

Ursache : Wandertourismus

Infrastruktur

Bänke / Parkplatz [ ]	Zuwegung [1]	Trittsteine [ ]	Überdachung [ ]
Wildfutterstelle [ ]	Viehtränke [ ]	Feuerstelle [ ]	Sonstiges [ ]
Anzahl Infr. 1	Sonstiges :		

Ablagerung* Deckungsgrad :	vollständig	teilweise	vereinzelt
Haus-/ Gewerbemüll			
Holzabfall			
Pflanzenabfall			
Erdaushub/ Bauschutt			
org. Reste/ Faulschlamm			
keine			1

* nur erste Fliessmeter

Einleitungen

unverdünnt		Oberfläche / Strasse	
Rohr trocken		Drainage / Graben	
		keine	1

Distanz zum Quellaustritt (m)

Bewertung Teil B : Vegetation-Nutzung-Struktur *Zutreffendes mit "1" markieren* 1

Vegetation/Nutzung

	Einzugsgebiet	Umfeld	Quellbereich	Quellufer	Quellbach
standortyp. Vegetation		1	1	1	1
standortfrem. Vegetation					
Moosgesellschaften		1	1	1	1
Heiden					
Hochstaudenfluren					
Laubwald					
Mischwald	1	1	1		
Gebüsch				1	
standorttyp. Nadelwald					
standortfremd. Nadelwald					
extens. genutztes Freiland					
intens. genutztes Freiland					
Acker/ Sonderkultur					
unbefestigter Weg					
befestigter Weg/Strasse		1			
künstl. veg.-frei/Siedlung					

	unbeschattet	schwach	mittel	stark
Sommerbeschattung				1
stark & Überdachung oder Nadelforst				

Struktur

Substrat		stark (>50%)	mittel (>20%)	gering (>1%)
->natürlich	Fels/Blöcke (>20 cm)			1
	(Kiesel) Steine (6-20 cm)	1		
	Kies/Schotter (0.2-6 cm)		1	
	Sand (0.1 - 2 mm)			1
	Feinmaterial (<0.1 mm)			1
	Moospolster		1	
	Wurzeln			1
	Totholz			
	Pflanzen			
	Falllaub			1
	Detritus/Org.Schlamm			
	Kalksinter…*			1
Anzahl Substrate 9		stark (>50%)	mittel (>20%)	gering (>1%)
	künstlich			
->verändert (nur Infos)	Fadenalgen			

Strömungs-diversität	Spritzwasser [1]	glatt [1]	fliessend [1]	überfliessend [1]
	gerippelt [1]	plätschernd [ ]	überstürzend [ ]	fallend [1]
Anzahl Strömungen 6				
Wasser-Land-Verzahnung		gross [ ]	mittel [1]	gering [ ]
Besondere Strukturen	Laufverzweigung [ ]	Inselstruktur [1]	Quellflur [ ]	Sandwirbel [ ]
	gr. Tiefenvarianz [ ]	natürl. Pools [ ]	Kaskaden [ ]	Wasserfall [1]
	Fliesshindernisse [ ]	Wassermoos [1]	Lückensyst. [1]	Rieselflur [1]
Anzahl Strukturen 5				

Wert A : Beeinträchtigung (höchster Wert) 3

Revitalisierungsobjekt (Einschätzung) JA — JA / NEIN

Wert B : Vegetation-Nutzung-Struktur 1.30

Bonus b -0,4 Punkte bei guter Struktur -> Aufwertung - 0.40

Gesamtergebnis [(A+B)/2]-b 1.75

Klassierung / Classement :	Gesamteindruck als Bewertungsvergleich		Bewertung / Evaluation
naturnah		blau	0.6 - 1.8
bedingt naturnah		grün	1.81 - 2.6
mässig beeinträchtigt		gelb	2.61 - 3.4
geschädigt		orange	3.41 - 4.2
stark geschädigt		rot	4.21 - 5.0

Quelle nicht bewertbar : [ ] **Q. zerstört**

Zutreffendes ankreuzen [x] [ ] **kein Abfluss**

AQ/ps_ver_20160413

Spring Protocol Structure copied from Lubini-Ferlin et al. (2014)

A2 Spring Protocol Fauna

Quellen Protokoll - Fauna	Kanton :	VD	ID :	001_VD

Quelle : La Diey	Datum :	08.06.2007	Koordinaten X/Y:	524580 / 172075
Ortsname : Romainmôtier	Höhe :	680	BestimmerIn :	Pascal Stucki

TAXALISTE alpine Quelle ☐ *ankreuzen [x]*

	Taxa	Stadium	RL	NP	Endemit	öwz	öwzA	Anzahl	Klasse
1	**Crenobia alpina**					16	8	**10**	**3**
2	**Niphargus spp.**					8	8	**1**	**1**
3	**Proasellus cavaticus**					8	8	**1**	**1**
4	**Gammarus fossarum**					4	4	**1**	**1**
5	**Ancylus fluviatilis**					2	2	**1**	**1**
6	**Rhtithrogena hybrida**					2	2	**4**	**2**
7	**Isoperla rivulorum**					2	2	**3**	**2**
8	**Leuctra subalpina**		NT	4		8	2	**7**	**2**
9	**Nemoura sp.**					4	16	**2**	**1**
10	**Protonemura nitida**					8	4	**3**	**2**
11	**Protonemura risi**					8	8	**32**	**4**
12	**Drusus mixtus**				E	8	8	**10**	**3**
13	**Hydropsyche sp.**					2	2	**1**	**1**
14	**Plectrocnemia geniculata**		NT			16	16	**5**	**2**
15	**Plectrocnemia brevis**		NT			16	16	**3**	**2**
16	**Rhyacophila vulgaris**					4	4	**3**	**2**
17	**Rhyacophila pubescens**					8	8	**1**	**1**
18	**Synagapetus dubitans**		NT			16	16	**10**	**3**
19	**Philopotamus variegatus**					1	1	**1**	**1**
20									
21									
22									
23									
24									
25									
26									
27									
28									
29									
30									
31									
32									
33									
34									
35									
36									

	Taxa	Stadium	RL	NP	Endemit	öwz	öwzA	Anzahl	Klasse
37									
38									
39									
40									
41									
42									
43									
44									
45									
46									
47									
48									
49									
50									
	zusätzliche Taxa								
51									
52									
53									
54									
55									
56									
57									
58									
59									
60									
61									
62									
63									
64									
65									
66									
67									
68									
69									
70									
71									

ERGEBNIS :

Klassierung :	ÖWS		
quelltypisch	>20	Anzahl Arten	19
bedingt quelltypisch	15.0-19.9	Rote Listen Arten	4
quellverträglich	10.0-14.9	Prioritäre Arten	1
quellfremd	5.1-9.9	Endemiten	0
sehr quellfremd	<5	ÖWZ 16 Arten	4
		ÖWZ 8 Arten	7
		ÖWS	**15.8**

Abundanzklassen : 1 => 1 - 2 Ind. • 2 => 3 - 7 Ind. • 3 => 8 - 15 Ind. • 4 => 16 - 50 Ind. • 5 => >50 Ind. oder [x] ☐ nur genaue Anzahl

FeldbearbeiterIn (leg) ändern falls anders	**Protokoll - Struktur** (dazugehörig) **ID** ändern falls anders	ausgefüllt am (Datum)
Pascal Stucki	001_VD	08.06.2007

AQ/ps_ver_20160307

Spring Protocol Fauna copied from Lubini-Ferlin et al. (2014)

A3 QGIS Python Code: Filled DEMs

```
import os
import processing
from qgis.core import QgsProcessingFeedback

input_folder = r"your\input\folder"
output_folder = os.path.join(input_folder, "filled_outputs")
os.makedirs(output_folder, exist_ok=True)

raster_list = [f for f in os.listdir(input_folder) if f.endswith(".tif")]

feedback = QgsProcessingFeedback()

for raster_name in raster_list:
    raster_path = os.path.join(input_folder, raster_name)
    id = os.path.splitext(raster_name)[0].split("_")[-1]

    filled_raster_path = os.path.join(output_folder, f"dem_filled_{id}.tif")
    direction_raster_path = os.path.join(output_folder, f"flowdir_tmp_{id}.tif")

    print(f"Processing {raster_name}...")

    params = {
        'input': raster_path,
        'areas': 'None',
        'output': filled_raster_path,
        'direction': direction_raster_path,
        'format': 0,
        'overwrite': True
    }

    processing.run("grass7:r.fill.dir", params, feedback=feedback)
```

A4 QGIS Python Code: slope and aspect 3x3

```
import os
import processing
from qgis.core import QgsProcessingFeedback

input_folder = r"your\input\folder"
output_folder = os.path.join(input_folder, "slopeandaspect")
os.makedirs(output_folder, exist_ok=True)

raster_list = [f for f in os.listdir(input_folder) if f.startswith("dem_filled_") and
    f.endswith(".tif")]

feedback = QgsProcessingFeedback()

for raster_name in raster_list:
    raster_path = os.path.join(input_folder, raster_name)
    id = raster_name.replace("dem_filled_", "").replace(".tif", "")

    slope_out = os.path.join(output_folder, f"slope_{id}.tif")
    aspect_out = os.path.join(output_folder, f"aspect_{id}.tif")

    print(f"Processing slope and aspect for: {raster_name}")

    params = {
        'elevation': raster_path,
        'slope': slope_out,
        'aspect': aspect_out,
        'format': 0,
        'overwrite': True
    }

    processing.run("grass7:r.slope.aspect", params, feedback=feedback)
```

A5 QGIS Python Code: plan curvature and profile curvature 3x3

```
import os
import processing
from qgis.core import QgsProcessingFeedback

input_folder = r"your\input\folder"
output_folder = os.path.join(input_folder, "morphometrie_rparam")
os.makedirs(output_folder, exist_ok=True)

raster_list = [f for f in os.listdir(input_folder) if f.startswith("dem_filled_") and
    f.endswith(".tif")]

feedback = QgsProcessingFeedback()

for raster_name in raster_list:
    id = raster_name.replace("dem_filled_", "").replace(".tif", "")
    raster_path = os.path.join(input_folder, raster_name)

    pcurv_out = os.path.join(output_folder, f"pcurv_{id}.tif")
    tcurv_out = os.path.join(output_folder, f"tcurv_{id}.tif")

    print(f"profile curvature for: {raster_name}")
    processing.run("grass7:r.param.scale", {
        'input': raster_path,
        'output': pcurv_out,
        'method': 6, # profile curvature
        'size': 7,
        'format': 0,
        'overwrite': True
    }, feedback=feedback)

    print(f"plan curvature for: {raster_name}")
    processing.run("grass7:r.param.scale", {
        'input': raster_path,
        'output': tcurv_out,
        'method': 7, # plan curvature
        'size': 7,
        'format': 0,
        'overwrite': True
    }, feedback=feedback)
```

A6 QGIS Python Code: Slope, Aspect, Plan and Profile Curvature 13x13 & 25x25

```
import os
import processing
from qgis.core import QgsProcessingFeedback

input_folder = r"your\input\folder"
slope_aspect_folder = os.path.join(input_folder, "slopeandaspect")
curvatures_folder = os.path.join(input_folder, "morphometrie_rparam")
os.makedirs(slope_aspect_folder, exist_ok=True)
os.makedirs(curvatures_folder, exist_ok=True)

raster_list = [f for f in os.listdir(input_folder) if f.startswith("dem_filled_") and
    f.endswith(".tif")]

window_sizes = {
    "26m": 13,
    "50m": 25
}

feedback = QgsProcessingFeedback()

for raster_name in raster_list:
    id = raster_name.replace("dem_filled_", "").replace(".tif", "")
    raster_path = os.path.join(input_folder, raster_name)

    for label, size in window_sizes.items():
        print(f"{raster_name} | Skala: {label} (size={size})")

        # SLOPE
        slope_out = os.path.join(slope_aspect_folder, f"slope_{label}_{id}.tif")
        processing.run("grass7:r.param.scale", {
            'input': raster_path,
            'output': slope_out,
            'method': 1, # slope
            'size': size,
            'format': 0,
            'overwrite': True
        }, feedback=feedback)

        # ASPECT
        aspect_out = os.path.join(slope_aspect_folder, f"aspect_{label}_{id}.tif")
        processing.run("grass7:r.param.scale", {
            'input': raster_path,
            'output': aspect_out,
            'method': 0, # aspect
            'size': size,
            'format': 0,
            'overwrite': True
        }, feedback=feedback)

        # PROFILE CURVATURE
        pcurv_out = os.path.join(curvatures_folder, f"pcurv_{label}_{id}.tif")
        processing.run("grass7:r.param.scale", {
            'input': raster_path,
            'output': pcurv_out,
            'method': 6, # profile curvature
            'size': size,
            'format': 0,
```

```
    'overwrite': True
}, feedback=feedback)

# PLAN CURVATURE
tcurv_out = os.path.join(curvatures_folder, f"tcurv_{label}_{id}.tif")
processing.run("grass7:r.param.scale", {
    'input': raster_path,
    'output': tcurv_out,
    'method': 7, # plan curvature
    'size': size,
    'format': 0,
    'overwrite': True
}, feedback=feedback)
```

A7 QGIS Python Code: Terrain Roughness Index TRI

```
import os
import numpy as np
from osgeo import gdal

def compute_tri(in_path, out_path, window_size):
    ds = gdal.Open(in_path)
    band = ds.GetRasterBand(1)
    arr = band.ReadAsArray().astype(np.float32)

    pad = window_size // 2
    arr_padded = np.pad(arr, pad_width=pad, mode='reflect')

    try:
        from numpy.lib.stride_tricks import sliding_window_view
        windows = sliding_window_view(arr_padded, (window_size,
            window_size))
        std = np.std(windows, axis=(2,3))
    except ImportError:
        nrows, ncols = arr.shape
        std = np.zeros_like(arr, dtype=np.float32)
        for i in range(nrows):
            for j in range(ncols):
                win = arr_padded[i:i+window_size, j:j+window_size]
                std[i, j] = np.std(win)

    driver = gdal.GetDriverByName('GTiff')
    out_ds = driver.Create(
        out_path, ds.RasterXSize, ds.RasterYSize, 1, gdal.GDT_Float32
)
    out_ds.SetGeoTransform(ds.GetGeoTransform())
    out_ds.SetProjection(ds.GetProjection())
    out_band = out_ds.GetRasterBand(1)
    out_band.WriteArray(std)
    out_band.GetStatistics(True, True)
    out_ds = None
    ds = None

input_folder = r"your\input\folder"
output_folder = os.path.join(input_folder, "tri_numpy")
os.makedirs(output_folder, exist_ok=True)

tri_scales = {
    "6m": 3,
    "26m": 13,
    "50m": 25
}

for fname in os.listdir(input_folder):
    if not (fname.startswith("dem_filled_") and fname.endswith(".tif")):
         continue
    dem_id = fname.replace("dem_filled_", "").replace(".tif", "")
    in_path = os.path.join(input_folder, fname)

    for label, window in tri_scales.items():
        out_path = os.path.join(output_folder, f"tri_numpy_{label}_{
            dem_id}.tif")
        print(f"Computing TRI {label} for {fname} (window {window}x{
            window})...")
        compute_tri(in_path, out_path, window)
```

A8 QGIS Python Code: Slope Length & TWI

```
import os
import processing
from qgis.core import QgsProcessingFeedback

input_folder = r"your\input\folder"
output_folder = os.path.join(input_folder, "watershed_outputs")
os.makedirs(output_folder, exist_ok=True)

raster_list = [f for f in os.listdir(input_folder) if f.startswith("dem_filled_") and
    f.endswith(".tif")]

thresholds = {
    "500": 500,
    "10000": 10000,
    "80000": 80000
}

feedback = QgsProcessingFeedback()

for raster_name in raster_list:
    id = raster_name.replace("dem_filled_", "").replace(".tif", "")
    raster_path = os.path.join(input_folder, raster_name)

    for label, threshold in thresholds.items():
        print(f"Watershed {label} | {raster_name}")

        twi_out = os.path.join(output_folder, f"twi_{label}_{id}.tif")
        slopelen_out = os.path.join(output_folder, f"slopelen_{label}_{id}.tif")

        processing.run("grass7:r.watershed", {
            'elevation': raster_path,
            'accumulation': None,
            'drainage': None,
            'basin': None,
            'stream': None,
            'tci': twi_out,
            'spi': None,
            'length_slope': slopelen_out,
            'blocking': None,
            'depression': None,
            'threshold': threshold,
            'convergence': 5,
            'memory': 300,
            'flags': '',
            'GRASS_REGION_PARAMETER': None,
            'GRASS_RASTER_FORMAT_OPT': '',
            'GRASS_RASTER_FORMAT_META': '',
            'GRASS_OUTPUT_TYPE_PARAMETER': 0,
        }, feedback=feedback)
```

A9 QGIS Python Code: Convergence Index Proxy

```
import os
import processing
from qgis.core import QgsProcessingFeedback

input_folder = r"your\input\folder"
output_folder = os.path.join(input_folder, "convergence_proxy")
os.makedirs(output_folder, exist_ok=True)

raster_list = [f for f in os.listdir(input_folder) if f.startswith("dem_filled_") and
    f.endswith(".tif")]

window_sizes = {
    "6m": 3,
    "26m": 13,
    "50m": 25
}

feedback = QgsProcessingFeedback()

for raster_name in raster_list:
    id = raster_name.replace("dem_filled_", "").replace(".tif", "")
    raster_path = os.path.join(input_folder, raster_name)

    for label, size in window_sizes.items():
        conv_out = os.path.join(output_folder, f"conv_{label}_{id}.tif")
        print(f"Convergence (Proxy) {label} for: {raster_name}")

        processing.run("grass7:r.param.scale", {
            'input': raster_path,
            'output': conv_out,
            'method': 5,
            'size': size,
            'format': 0,
            'overwrite': True
        }, feedback=feedback)
```

A10 Calculation of Performance Metrics

RMSE (Root Mean Squared Error)

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

Here, y_i is the true target value for observation i, $\hat{y}_i$ is the corresponding model prediction, and n is the total number of observations. RMSE measures the average deviation of the predictions from the true values, with errors squared to weigh larger deviations more heavily. The lower the RMSE, the better the predictive accuracy of the model. RMSE is expressed in the same units as the target variable y. Due to squaring, RMSE is sensitive to outliers: a single large error can substantially increase its value.

R^2 (Coefficient of Determination)

$$\mathrm{R}^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}$$

Here, y_i is the true target value for observation i, $\hat{y}_i$ the corresponding model prediction, $\bar{y}$ the average of all target values in the dataset, and n the total number of observations. R^2 measures the proportion of variance in the target variable that is explained by the model. It is dimensionless: values closer to 1 indicate a better fit (with $\mathrm{R}^2 = 1$ corresponding to a perfect prediction). Although R^2 can be compared across different datasets, its interpretation can be misleading for nonlinear relationships or highly heterogeneous data.

MAE (Mean Absolute Error)

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right|$$

MAE measures the average absolute error in the same units as the target variable y. Each error contributes linearly; large outliers are not squared and therefore are less heavily penalized than in RMSE.

MedAE (Median Absolute Error)

$$\mathrm{MedAE} = \mathrm{median}\left(\{|y_i - \hat{y}_i|\}_{i=1}^{n}\right)$$

The median of the absolute errors indicates the threshold below which 50% of the absolute errors fall, providing a robust measure of typical error magnitude without being skewed by a few large prediction errors. It ignores the distribution above the 50th percentile and does not convey information about the size of the largest errors.

Accuracy
Accuracy measures the proportion of instances that the model classified correctly. It is most meaningful when the classes are roughly balanced and the cost of errors in each class is similar.

Balanced Accuracy
Balanced Accuracy computes the average of the per-class recall scores. By weighting each class equally, it gives a more robust assessment than plain Accuracy on imbalanced datasets.

F1-Score
The F1-Score is the harmonic mean of Precision and Recall and provides a single metric that balances the trade-off between the two. It is defined as

$$F1 - Score = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

with $\text{Precision} = \frac{TP}{TP+FP}$ and $\text{Recall} = \frac{TP}{TP+FN}$, where TP denotes true positives, FP false positives, and FN false negatives. The F1-Score ranges from 0 to 1, with 1 representing perfect Precision and Recall. It is particularly useful in situations with imbalanced datasets or when both false positives and false negatives carry significant consequences. Unlike Accuracy, the F1 Score remains informative even if class distributions are skewed.

A11 R Code Baseline Regression Random Forest: Predicting Faunistic Assessment Values Using Spring Structure Data

```
# Baseline Regression Random Forest predicting ...
# ... Faunistic Assessment Values using spring structure data)

library(tidyverse)
library(randomForest)
library(ggplot2)

structure_file <- "your\\file\\path.csv"
fauna_file <- "your\\file\\path.csv"

structure_data <- read_delim(structure_file, delim = ";", col_types = cols())
fauna_data <- read_delim(fauna_file, delim = ";", col_types = cols())

# select only certain columns
fauna_subset <- fauna_data %>%
  select(Numbering, `Result Faunistic Assessment`)

# Merge both datasets
merged_data <- inner_join(structure_data, fauna_subset, by = "Numbering")

# Rename target variable and convert to numeric
merged_data <- merged_data %>%
  rename(FaunisticAssessment = `Result Faunistic Assessment`) %>%
  mutate(FaunisticAssessment = as.numeric(as.character(FaunisticAssessment)))

# Remove unwanted predictors
cols_to_exclude <- c("Conductivity", "Distance to Spring Outflow [m]",
                     "Value A: Impairment (Highest Value)",
                     "Value B: Vegetation-Use-Structure",
                     "Bonus b -0.4 Points for Good Structure -> Upgrade",
                     "Result Structural Assessment")
model_data <- merged_data %>% select(-one_of(cols_to_exclude))

# Remove rows with missing data
model_data <- model_data[complete.cases(model_data), ]

# Convert all character columns to factors
model_data <- model_data %>% mutate(across(where(is.character), as.factor))

# Aggregate vegetation variables:
veg_types <- c("Site-Specific.Vegetation", "Non-Native.Vegetation", "Moss.Communities"
    ,
               "Dwarf.Shrub.Heaths", "Tall.Herbaceous.Vegetation", "Deciduous.Forest",
               "Mixed.Forest", "Shrubland", "Site-Specific.Conifer.Forest", "Non-Native.
                   Conifer.Forest",
               "Extensively.Used.Open.Land", "Intensively.Used.Open.Land", "Farmland.
                   Special.Crops",
               "Unpaved.Path", "Paved.Path.Road", "Artificial.Vegetation-Free.Settlement
                   ")

aggregated_veg <- sapply(veg_types, function(vt) {
  vt_cols <- grep(paste0("^", vt), names(model_data), value = TRUE)
  if (length(vt_cols) == 0) {
    return(rep(0, nrow(model_data)))
  } else {
    mat <- model_data[, vt_cols, drop = FALSE]
    mat <- as.data.frame(lapply(mat, function(x) as.numeric(as.character(x))))
```

```r
    apply(mat, 1, function(row) ifelse(sum(row, na.rm = TRUE) > 0, 1, 0))
  }
})
aggregated_veg <- as.data.frame(aggregated_veg)
names(aggregated_veg) <- veg_types

# Remove original vegetation columns and add aggregated ones
veg_pattern <- paste(veg_types, collapse = "|")
model_data <- model_data[, !grepl(veg_pattern, names(model_data))]
model_data <- bind_cols(model_data, aggregated_veg)
model_data <- model_data %>% select(-Numbering)
names(model_data) <- make.names(names(model_data))

# Split according to the saved indices
test_index <- read.csv("your/file/path.csv")$x
test_data <- model_data[test_index, ]
train_data <- model_data[-test_index, ]

# Create Random Forest regression model
rf_model <- randomForest(FaunisticAssessment ~ ., data = train_data, importance = TRUE
    )
print(rf_model)

# Make predictions on the test dataset
predictions <- predict(rf_model, newdata = test_data)

################################################################################
# Calculate comparative metrics for all models
MedAE <- function(obs, pred) median(abs(obs - pred))
MBE <- function(obs, pred) mean(pred - obs)
rRMSE <- function(obs, pred) sqrt(mean((obs - pred)^2)) / mean(obs)
NRMSE <- function(obs, pred) sqrt(mean((obs - pred)^2)) / (max(obs) - min(obs))
MAPE <- function(obs, pred) mean(abs((obs - pred) / obs)) * 100
sMAPE <- function(obs, pred) mean(2 * abs(pred - obs) / (abs(obs) + abs(pred))) * 100
MedSE <- function(obs, pred) median((obs - pred)^2)
CCC <- function(obs, pred) {
  mean_obs <- mean(obs)
  mean_pred <- mean(pred)
  s_obs <- var(obs)
  s_pred <- var(pred)
  covar <- mean((obs - mean_obs) * (pred - mean_pred))
  2 * covar / (s_obs + s_pred + (mean_obs - mean_pred)^2)
}

obs <- test_data$FaunisticAssessment
pred <- predictions

metrics <- data.frame(
  model = "Baseline RF",
  MSE = mean((obs - pred)^2),
  Percent_Var_Explained = 100 * (1 - sum((obs - pred)^2) / sum((obs - mean(obs))^2)),
  RMSE = sqrt(mean((obs - pred)^2)),
  R2 = 1 - sum((obs - pred)^2) / sum((obs - mean(obs))^2),
  MAE = mean(abs(obs - pred)),
  MedAE = MedAE(obs, pred),
  MBE = MBE(obs, pred),
  rRMSE = rRMSE(obs, pred),
  NRMSE = NRMSE(obs, pred),
  MAPE = MAPE(obs, pred),
  sMAPE = sMAPE(obs, pred),
  MedSE = MedSE(obs, pred),
```

```
  CCC = CCC(obs, pred),
  LinsCCC = CCC(obs, pred)
)

results_df <- data.frame(
  observed = obs,
  predicted = pred,
  model = "Baseline RF"
)

# Export results
results_dir <- "your/directory/path"
dir.create(results_dir, showWarnings = FALSE)
write.csv2(results_df, file = file.path(results_dir, "rf_baseline_results.csv"), row.
    names = FALSE)
write.csv2(metrics, file = file.path(results_dir, "rf_baseline_metrics.csv"), row.
    names = FALSE)

###############################################################################
# Calculate permutation importance for all predictors
library(dplyr)
library(stringr)

# List of all predictor variables
predictor_vars <- setdiff(names(test_data), "FaunisticAssessment")

get_metrics <- function(obs, pred) {
  data.frame(
    RMSE = sqrt(mean((obs - pred)^2)),
    R2 = 1 - sum((obs - pred)^2) / sum((obs - mean(obs))^2),
    MAE = mean(abs(obs - pred)),
    MedAE = median(abs(obs - pred))
  )
}

obs <- test_data$FaunisticAssessment
pred_original <- predict(rf_model, newdata = test_data)
metrics_original <- get_metrics(obs, pred_original)
metrics_original$variable <- "ALL"
metrics_original$type <- "original"

# Permute each variable
results <- lapply(predictor_vars, function(var) {
  td_perm <- test_data
  td_perm[[var]] <- sample(td_perm[[var]])
  pred <- predict(rf_model, newdata = td_perm)
  m <- get_metrics(obs, pred)
  m$variable <- var
  m$type <- "permuted"
  m
})

results_df <- do.call(rbind, results)
results_df <- rbind(metrics_original, results_df)
results_df$model_name <- "Baseline RF"
results_df$target_variable <- "FaunisticAssessment"

# Group binary-coded variables
results_df$base_variable <- case_when(
  str_starts(results_df$variable, "Special.Structures_") ~ "Special.Structures",
  str_starts(results_df$variable, "Infrastructure_") ~ "Infrastructure",
```

```r
  str_starts(results_df$variable, "Flow.Diversity_") ~ "Flow.Diversity",
  TRUE ~ results_df$variable
)

results_grouped <- results_df %>%
  group_by(model_name, target_variable, type, base_variable) %>%
  summarize(
    RMSE = mean(RMSE, na.rm = TRUE),
    R2 = mean(R2, na.rm = TRUE),
    MAE = mean(MAE, na.rm = TRUE),
    MedAE = mean(MedAE,na.rm = TRUE),
    .groups = "drop"
  ) %>%
  select(model_name, target_variable, type, base_variable, RMSE, R2, MAE, MedAE)

# Export grouped metrics
write.csv2(
  results_grouped,
  file = file.path(results_dir, "rf_per_variable_metrics_Grouped.csv"),
  row.names = FALSE,
  quote = FALSE
)
```

A12 R Code Baseline Classification Random Forest: Predicting Fauna Classes Using Spring Structure Data

```
# Baseline Classification Random Forest predicting ...
# ... fauna classes using spring structure data)

library(tidyverse)
library(randomForest)
library(ggplot2)

structure_file <- "your\\file\\path.csv"
fauna_file <- "your\\file\\path.csv"

structure_data <- read_delim(structure_file, delim = ";", col_types = cols())
fauna_data <- read_delim(fauna_file, delim = ";", col_types = cols())

# select only certain columns
fauna_subset <- fauna_data %>%
  select(Numbering, `Result Faunistic Assessment`)

# Merge both datasets
merged_data <- inner_join(structure_data, fauna_subset, by = "Numbering")

# Rename target variable, convert to numeric, and categorize into classes
merged_data <- merged_data %>%
  rename(FaunisticAssessment = `Result Faunistic Assessment`) %>%
  mutate(FaunisticAssessment = as.numeric(as.character(FaunisticAssessment)),
         FaunisticClass = cut(FaunisticAssessment,
                              breaks = c(-Inf, 5, 10, 15, 20, Inf),
                              labels = c("very non spring", "non-spring",
                                         "spring-compliant", "conditionally spring-typical
                                             ",
                                         "spring-typical"),
                              right = FALSE))

# Remove unwanted predictors
cols_to_exclude <- c("Conductivity",
                     "Value A: Impairment (Highest Value)",
                     "Value B: Vegetation-Use-Structure",
                     "Bonus b -0.4 Points for Good Structure -> Upgrade",
                     "Result Structural Assessment")
model_data <- merged_data %>% select(-one_of(cols_to_exclude))

# Remove rows with missing data
model_data <- model_data[complete.cases(model_data), ]

# Convert all character columns to factors
model_data <- model_data %>% mutate(across(where(is.character), as.factor))

# Aggregate vegetation variables
veg_types <- c("Site-Specific.Vegetation", "Non-Native.Vegetation", "Moss.Communities"
    ,
               "Dwarf.Shrub.Heaths", "Tall.Herbaceous.Vegetation", "Deciduous.Forest",
               "Mixed.Forest", "Shrubland", "Site-Specific.Conifer.Forest", "Non-Native.
                   Conifer.Forest",
               "Extensively.Used.Open.Land", "Intensively.Used.Open.Land", "Farmland.
                   Special.Crops",
               "Unpaved.Path", "Paved.Path.Road", "Artificial.Vegetation-Free.Settlement
                   ")
aggregated_veg <- sapply(veg_types, function(vt) {
```

```
  vt_cols <- grep(paste0("^", vt), names(model_data), value = TRUE)
  if (length(vt_cols) == 0) {
    rep(0, nrow(model_data))
  } else {
    mat <- as.data.frame(lapply(model_data[, vt_cols, drop = FALSE],
                              function(x) as.numeric(as.character(x))))
    apply(mat, 1, function(row) ifelse(sum(row, na.rm = TRUE) > 0, 1, 0))
  }
})
aggregated_veg <- as.data.frame(aggregated_veg)
names(aggregated_veg) <- veg_types

# Remove original vegetation columns and add aggregated ones
veg_pattern <- paste(veg_types, collapse = "|")
model_data <- model_data[, !grepl(veg_pattern, names(model_data))]
model_data <- bind_cols(model_data, aggregated_veg)
model_data <- model_data %>% select(-Numbering, -FaunisticAssessment)
names(model_data) <- make.names(names(model_data))

# Split according to the saved indices
test_index <- read.csv("your/file/path.csv")[,1]
all_idx <- seq_len(nrow(model_data))
train_idx <- setdiff(all_idx, test_index)
train_data <- model_data[train_idx, ]
test_data <- model_data[test_index, ]

# Train Random Forest classification model
rf_model <- randomForest(FaunisticClass ~ ., data = train_data, importance = TRUE)
print(rf_model)
predictions <- predict(rf_model, newdata = test_data)

################################################################################
# Calculate comparative metrics for all models
conf_matrix <- table(test_data$FaunisticClass, predictions)
accuracy <- sum(diag(conf_matrix)) / sum(conf_matrix)
recall_per_class <- diag(conf_matrix) / rowSums(conf_matrix)
precision_per_class <- diag(conf_matrix) / colSums(conf_matrix)
f1_per_class <- 2 * recall_per_class * precision_per_class /
                     (recall_per_class + precision_per_class)
balanced_accuracy <- mean(recall_per_class, na.rm = TRUE)
calc_kappa <- function(conf) {
  total <- sum(conf)
  p0 <- sum(diag(conf)) / total
  pe <- sum(rowSums(conf) * colSums(conf)) / (total^2)
  (p0 - pe) / (1 - pe)
}
kappa <- calc_kappa(conf_matrix)

metrics_df <- data.frame(
  Metric = c("Accuracy", "Balanced Accuracy", "Kappa",
             paste0("Recall_", names(recall_per_class)),
             paste0("Precision_", names(precision_per_class)),
             paste0("F1_", names(f1_per_class))),
  Value = c(mean(test_data$FaunisticClass == predictions),
             balanced_accuracy, kappa,
             as.numeric(recall_per_class),
             as.numeric(precision_per_class),
             as.numeric(f1_per_class)),
  Model = "Baseline RF"
)
write.table(metrics_df, file="baseline_rf_metrics.csv", sep=";", dec=",",
```

```r
          row.names=FALSE, quote=FALSE)

results_out <- data.frame(observed=test_data$FaunisticClass,
                         predicted=predictions, Model="Baseline RF")
write.table(results_out, file="baseline_rf_observed_predicted.csv",
          sep=";", dec=",", row.names=FALSE, quote=FALSE)

################################################################################
# Calculate permutation importance for all predictors
library(dplyr); library(stringr); library(tidyr); library(purrr)

# List of all predictor variables
predictor_vars <- setdiff(names(test_data), "FaunisticClass")
all_classes <- levels(test_data$FaunisticClass)

accuracy_fun <- function(obs,pred) mean(obs==pred)
balanced_fun <- function(obs,pred) {
  cm <- table(obs,pred); mean(diag(cm)/rowSums(cm), na.rm=TRUE)
}
recall_fun <- function(obs,pred) { cm<-table(obs,pred); diag(cm)/rowSums(cm) }
precision_fun<- function(obs,pred) { cm<-table(obs,pred); diag(cm)/colSums(cm) }
f1_fun <- function(obs,pred) { r<-recall_fun(obs,pred); p<-precision_fun(obs,pred);
                                 2*r*p/(r+p) }

get_row <- function(obs,pred,model,type,base) {
  tibble(model_name=model, target_variable="FaunisticClass", type=type,
        base_variable=base, Accuracy=accuracy_fun(obs,pred),
        Balanced_Accuracy=balanced_fun(obs,pred),
        !!!set_names(as.list(recall_fun(obs,pred)), paste0("Recall_", all_classes)),
        !!!set_names(as.list(precision_fun(obs,pred)),paste0("Precision_", all_classes
            )),
        !!!set_names(as.list(f1_fun(obs,pred)), paste0("F1_", all_classes)))
}

obs <- test_data$FaunisticClass
pred_original <- predictions
metrics_list <- list(get_row(obs,pred_original,"Baseline RF","original","ALL"))

for(var in predictor_vars) {
  td_perm <- test_data; td_perm[[var]] <- sample(td_perm[[var]])
  pred <- predict(rf_model, newdata=td_perm)
  metrics_list[[length(metrics_list)+1]] <- get_row(obs,pred,"Baseline RF","permuted",
      var)
}
metrics_df <- bind_rows(metrics_list)
metrics_df$base_variable <- case_when(
  str_starts(metrics_df$base_variable,"Special.Structures_")~"Special.Structures",
  str_starts(metrics_df$base_variable,"Infrastructure_") ~"Infrastructure",
  str_starts(metrics_df$base_variable,"Flow.Diversity_") ~"Flow.Diversity",
  TRUE ~metrics_df$base_variable
)
metrics_df_grouped <- metrics_df %>%
  group_by(model_name, target_variable, type, base_variable) %>%
  summarize(across(where(is.numeric), ~mean(.x, na.rm = TRUE)), .groups = "drop")

# Export grouped metrics
write.csv2(metrics_grouped, file="rf_per_variable_metrics_GROUPED.csv", row.names=
    FALSE, quote=FALSE)
```

A13 Confusion Matrices of Random Forest Models Predicting Result Faunistic Assessment Values Based on Spring Structure Data

predicted

observed

Baseline RF

	very non-spring	non-spring	spring-compl.	cond. spring-typ.	spring-typical
very non-spring	0	0	0	1	0
non-spring	0	0	7	5	1
spring-compl.	0	1	20	20	2
cond. spring-typ.	0	0	9	22	7
spring-typical	0	0	7	14	19

Regularized RF

	very non-spring	non-spring	spring-compl.	cond. spring-typ.	spring-typical
very non-spring	0	0	0	1	0
non-spring	0	0	6	6	1
spring-compl.	0	1	17	21	4
cond. spring-typ.	0	0	11	19	8
spring-typical	0	0	7	17	16

Conditional Inference Forest

	very non-spring	non-spring	spring-compl.	cond. spring-typ.	spring-typical
very non-spring	0	0	0	1	0
non-spring	0	0	8	4	1
spring-compl.	0	0	18	24	1
cond. spring-typ.	0	0	10	23	5
spring-typical	0	0	6	20	14

Extremely Randomized Trees

	very non-spring	non-spring	spring-compl.	cond. spring-typ.	spring-typical
very non-spring	0	0	0	0	1
non-spring	0	0	6	7	0
spring-compl.	0	1	16	22	4
cond. spring-typ.	0	0	8	24	6
spring-typical	0	0	7	18	15

Quantile Regression Forest

	very non-spring	non-spring	spring-compl.	cond. spring-typ.	spring-typical
very non-spring	0	0	0	1	0
non-spring	0	0	2	11	0
spring-compl.	0	0	15	28	0
cond. spring-typ.	0	0	4	32	2
spring-typical	0	0	4	27	9

Cost-sensitive RF

	very non-spring	non-spring	spring-compl.	cond. spring-typ.	spring-typical
very non-spring	0	0	0	1	0
non-spring	0	0	6	6	1
spring-compl.	0	0	19	21	3
cond. spring-typ.	0	0	14	18	6
spring-typical	0	0	7	16	17

predicted

observed

class weighting

	very non-spring	non-spring	spring-compl.	cond. spring-typ.	spring-typical
very non-spring	0	0	0	0	0
non-spring	0	0	0	0	0
spring-compl.	0	0	0	0	0
cond. spring-typ.	0	0	0	0	0
spring-typical	0	0	0	0	0

dynamic weighting

	very non-spring	non-spring	spring-compl.	cond. spring-typ.	spring-typical
very non-spring	0	0	0	0	0
non-spring	0	0	0	0	0
spring-compl.	0	0	0	0	0
cond. spring-typ.	0	0	0	0	0
spring-typical	0	0	0	0	0

threshold adjustment ROC

	very non-spring	non-spring	spring-compl.	cond. spring-typ.	spring-typical
very non-spring	0	0	1	0	0
non-spring	1	2	9	0	1
spring-compl.	2	9	22	3	7
cond. spring-typ.	0	5	17	5	11
spring-typical	0	3	8	7	22

up/downsampling

	very non-spring	non-spring	spring-compl.	cond. spring-typ.	spring-typical
very non-spring	0	0	0	1	0
non-spring	1	1	8	3	0
spring-compl.	0	5	18	12	8
cond. spring-typ.	0	1	12	16	9
spring-typical	0	2	8	13	17

weighted tree voting

	very non-spring	non-spring	spring-compl.	cond. spring-typ.	spring-typical
very non-spring	0	0	0	1	0
non-spring	0	1	9	3	0
spring-compl.	0	0	20	20	3
cond. spring-typ.	0	0	12	20	6
spring-typical	0	0	10	16	14

A14 Additional Performance Metrics for Random Forest Classification Models Predicting Fauna Categories Based on Spring Structure Data

	Precision very non-spring	Precision non-spring	Precision spring-compliant	Precision cond. spring-typ.	Precision spring-typ.	Recall very non-spring	Recall non-spring	Recall spring-compliant	Recall cond. spring-typ.	Recall spring-typ.
Baseline RF	–	0.00	0.47	0.35	0.66	0.00	0.00	0.47	0.58	0.47
Regularized RF	–	0.00	0.41	0.30	0.55	0.00	0.00	0.40	0.50	0.40
Cond. Inf. Forest	–	–	0.43	0.32	0.67	0.00	0.00	0.42	0.61	0.35
Extr. Random. Trees	–	0.00	0.43	0.34	0.58	0.00	0.00	0.37	0.63	0.38
Quantile Regr. Forest	–	–	0.65	0.33	0.78	0.00	0.00	0.40	0.87	0.17
Cost-sensitive RF	–	–	0.41	0.29	0.63	0.00	0.00	0.44	0.47	0.42
Class weighting	–	0.00	0.45	0.33	0.62	0.00	0.00	0.42	0.61	0.38
Dynamic weighting	–	–	0.49	0.28	0.56	0.00	0.00	0.40	0.55	0.35
weighted tree voting	–	1.00	0.39	0.02	0.26	1.00	0.08	0.47	0.16	0.02
threshold adjustment	0.00	0.11	0.39	0.00	0.27	0.00	0.15	0.51	0.29	0.00
up/downsampling	0.00	0.11	0.39	0.36	0.50	0.00	0.08	0.42	0.42	0.42

A15 Confusion Matrices of Random Forest Models Predicting Result Structural Assessment Values Based on Landscape & Topographic Variables

predicted

observed

Baseline RF

	str. damaged	damaged	mod. impaired	cond. natural	natural
str. damaged	0	0	0	0	1
damaged	0	0	0	4	6
mod. impaired	0	0	2	5	26
cond. natural	0	0	1	3	43
natural	0	0	2	8	124

Regularized RF

	str. damaged	damaged	mod. impaired	cond. natural	natural
str. damaged	0	0	0	0	1
damaged	0	0	0	4	6
mod. impaired	0	0	3	6	24
cond. natural	0	1	0	4	42
natural	0	0	2	11	121

Conditional Inference Forest

	str. damaged	damaged	mod. impaired	cond. natural	natural
str. damaged	0	0	0	0	1
damaged	0	0	0	1	9
mod. impaired	0	0	1	4	28
cond. natural	0	0	0	1	46
natural	0	0	0	1	133

Extremely Randomized Trees

	str. damaged	damaged	mod. impaired	cond. natural	natural
str. damaged	0	0	0	0	1
damaged	0	0	0	2	8
mod. impaired	0	0	1	5	27
cond. natural	0	0	0	2	45
natural	0	0	1	8	125

Quantile Regression Forest

	str. damaged	damaged	mod. impaired	cond. natural	natural
str. damaged	0	0	0	1	0
damaged	0	0	1	4	6
mod. impaired	0	0	2	7	10
cond. natural	0	0	0	12	29
natural	0	0	2	17	132

Cost-sensitive RF

	str. damaged	damaged	mod. impaired	cond. natural	natural
str. damaged	0	0	0	1	0
damaged	0	0	0	2	8
mod. impaired	0	0	5	4	24
cond. natural	0	2	1	5	39
natural	1	0	6	14	113

predicted

observed

class weighting

	str. damaged	damaged	mod. impaired	cond. natural	natural
str. damaged	0	0	1	0	0
damaged	0	3	7	0	0
mod. impaired	0	20	7	6	0
cond. natural	0	34	12	1	0
natural	0	90	34	10	0

dynamic weighting

	str. damaged	damaged	mod. impaired	cond. natural	natural
str. damaged	0	1	0	0	0
damaged	0	4	6	0	0
mod. impaired	0	23	6	4	0
cond. natural	0	36	10	1	0
natural	0	95	35	4	0

threshold adjustment ROC

	str. damaged	damaged	mod. impaired	cond. natural	natural
str. damaged	0	0	0	1	0
damaged	0	1	1	8	0
mod. impaired	0	0	5	21	7
cond. natural	0	0	1	34	12
natural	0	1	7	60	66

up/downsampling

	str. damaged	damaged	mod. impaired	cond. natural	natural
str. damaged	1	0	0	0	0
damaged	1	0	4	2	3
mod. impaired	1	5	6	6	15
cond. natural	0	5	4	6	32
natural	1	7	11	18	97

weighted tree voting

	str. damaged	damaged	mod. impaired	cond. natural	natural
str. damaged	0	0	0	0	1
damaged	0	0	0	0	10
mod. impaired	0	0	0	0	33
cond. natural	0	0	0	0	47
natural	0	0	0	0	134

A16 Additional Performance Metrics for Random Forest Classification Models Predicting Structure Categories Based on Landscape & Topographic Variables Data

	Precision str. damaged	Precision damaged	Precision mod. impaired	Precision cond. natural	Precision natural	Recall str. damaged	Recall damaged	Recall mod. impaired	Recall cond. natural	Recall natural
Baseline RF	–	–	0.40	0.15	0.62	0.00	0.00	0.06	–	0.93
Regularized RF	–	0.00	0.60	0.16	0.62	0.00	0.00	0.09	–	0.90
Cond. Inf. Forest	–	–	1.00	0.14	0.61	0.00	0.00	0.03	–	0.99
Extr. Random. Trees	–	–	0.50	0.12	0.61	0.00	0.00	0.03	–	0.93
Quantile Regr. Forest	–	–	0.40	0.29	0.75	0.00	0.00	0.11	–	0.87
Cost-sensitive RF	0.00	0.00	0.42	0.19	0.61	0.00	0.00	0.15	–	0.84
Class weighting	–	0.02	0.11	0.06	–	0.00	0.30	0.21	–	0.00
Dynamic weighting	–	0.03	0.11	0.11	–	0.00	0.40	0.18	–	0.00
weighted tree voting	–	–	–	–	0.60	0.00	0.00	0.00	–	1.00
threshold adjustment	–	0.50	0.36	0.27	0.78	0.00	0.10	0.15	–	0.49
up/downsampling	0.25	0.00	0.24	0.19	0.66	1.00	0.00	0.18	–	0.72

A17 Confusion Matrices of Random Forest Models Predicting Result Faunistic Assessment Values Based on Landscape & Topographic Variables

predicted

observed

Baseline RF

	very non-spring	non-spring	spring-compl.	cond. spring-typ.	spring-typical
very non-spring	0	0	0	1	0
non-spring	0	2	11	7	3
spring-compl.	0	0	24	32	17
cond. spring-typ.	0	0	25	21	24
spring-typical	0	2	15	14	27

Regularized RF

	very non-spring	non-spring	spring-compl.	cond. spring-typ.	spring-typical
very non-spring	0	0	0	1	0
non-spring	0	1	12	10	0
spring-compl.	0	0	24	31	18
cond. spring-typ.	0	0	25	21	24
spring-typical	0	0	15	15	28

Conditional Inference Forest

	very non-spring	non-spring	spring-compl.	cond. spring-typ.	spring-typical
very non-spring	0	0	1	0	0
non-spring	0	0	16	5	2
spring-compl.	0	0	43	18	12
cond. spring-typ.	0	0	38	16	16
spring-typical	0	0	18	8	32

Extremely Randomized Trees

	very non-spring	non-spring	spring-compl.	cond. spring-typ.	spring-typical
very non-spring	0	0	0	1	0
non-spring	0	1	11	7	4
spring-compl.	0	0	34	22	17
cond. spring-typ.	0	1	25	26	18
spring-typical	0	1	14	15	28

Quantile Regression Forest

	very non-spring	non-spring	spring-compl.	cond. spring-typ.	spring-typical
very non-spring	0	0	0	1	0
non-spring	0	0	10	13	0
spring-compl.	0	0	23	50	0
cond. spring-typ.	0	0	15	50	5
spring-typical	0	0	10	38	10

Cost-sensitive RF

	very non-spring	non-spring	spring-compl.	cond. spring-typ.	spring-typical
very non-spring	0	0	1	0	0
non-spring	0	0	17	1	5
spring-compl.	0	0	45	15	13
cond. spring-typ.	0	0	37	8	25
spring-typical	0	0	17	9	32

predicted

observed

class weighting

	very non-spring	non-spring	spring-compl.	cond. spring-typ.	spring-typical
very non-spring	0	0	0	2	0
non-spring	0	2	22	16	6
spring-compl.	0	2	54	62	28
cond. spring-typ.	0	0	62	42	36
spring-typical	0	0	32	30	54

dynamic weighting

	very non-spring	non-spring	spring-compl.	cond. spring-typ.	spring-typical
very non-spring	0	0	0	2	0
non-spring	0	2	24	16	4
spring-compl.	0	2	60	64	20
cond. spring-typ.	0	0	44	56	40
spring-typical	0	0	32	38	46

threshold adjustment ROC

	very non-spring	non-spring	spring-compl.	cond. spring-typ.	spring-typical
very non-spring	0	0	0	2	0
non-spring	0	8	20	14	4
spring-compl.	2	10	48	60	26
cond. spring-typ.	0	2	60	36	42
spring-typical	0	10	32	22	52

up/downsampling

	very non-spring	non-spring	spring-compl.	cond. spring-typ.	spring-typical
very non-spring	0	0	2	0	0
non-spring	2	12	20	6	6
spring-compl.	12	28	50	12	44
cond. spring-typ.	12	14	40	18	56
spring-typical	6	8	16	10	76

weighted tree voting

	very non-spring	non-spring	spring-compl.	cond. spring-typ.	spring-typical
very non-spring	0	0	0	2	0
non-spring	0	2	22	16	6
spring-compl.	0	0	58	58	30
cond. spring-typ.	0	0	52	38	50
spring-typical	0	0	32	32	52

A18 Additional Performance Metrics for Random Forest Classification Models Predicting Fauna Categories Based on Landscape & Topographic Variables Data

	Precision very non-spring	Precision non-spring	Precision spring-compliant	Precision cond. spring-typ.	Precision spring-typ.	Recall very non-spring	Recall non-spring	Recall spring-compliant	Recall cond. spring-typ.	Recall spring-typ.
Baseline RF	–	0.50	0.32	0.28	0.38	0.00	0.09	0.33	0.30	0.47
Regularized RF	–	1.00	0.32	0.27	0.40	0.00	0.04	0.33	0.30	0.48
Cond. Inf. Forest	–	–	0.37	0.34	0.52	0.00	0.00	0.59	0.23	0.55
Extr. Random. Trees	–	0.33	0.40	0.37	0.42	0.00	0.04	0.47	0.37	0.48
Quantile Regr. Forest	–	–	0.40	0.33	0.67	0.00	0.00	0.32	0.71	0.17
Cost-sensitive RF	–	–	0.38	0.24	0.43	0.00	0.00	0.62	0.11	0.55
Class weighting	–	0.50	0.32	0.28	0.44	0.00	0.04	0.37	0.30	0.47
Dynamic weighting	–	0.50	0.38	0.32	0.42	0.00	0.04	0.41	0.40	0.40
weighted tree voting	–	1.00	0.35	0.26	0.38	0.00	0.04	0.40	0.27	0.45
threshold adjustment	0.00	0.27	0.30	0.27	0.42	0.00	0.17	0.33	0.26	0.45
up/downsampling	0.00	0.19	0.39	0.39	0.42	0.00	0.26	0.34	0.13	0.66

A19 Confusion Matrices of Random Forest Models Predicting Result Faunistic Assessment Values Based on Mixed Spring Structure and Landscape & Topographic Variables Data

predicted (columns); observed (rows)

Baseline RF

	very non-spring	non-spring	spring-compl.	cond. spring-typ.	spring-typical
very non-spring	0	0	1	1	0
non-spring	0	0	8	1	1
spring-compl.	0	0	15	15	5
cond. spring-typ.	0	1	15	24	6
spring-typical	0	0	6	13	15

Regularized RF

	very non-spring	non-spring	spring-compl.	cond. spring-typ.	spring-typical
very non-spring	0	0	0	2	0
non-spring	0	0	6	2	2
spring-compl.	0	1	19	9	6
cond. spring-typ.	0	1	21	15	9
spring-typical	0	0	9	13	12

Conditional Inference Forest

	very non-spring	non-spring	spring-compl.	cond. spring-typ.	spring-typical
very non-spring	0	0	1	1	0
non-spring	0	0	4	5	1
spring-compl.	0	0	14	20	1
cond. spring-typ.	0	0	17	23	6
spring-typical	0	0	6	17	11

Extremely Randomized Trees

	very non-spring	non-spring	spring-compl.	cond. spring-typ.	spring-typical
very non-spring	0	0	1	1	0
non-spring	0	0	5	4	1
spring-compl.	0	0	9	22	4
cond. spring-typ.	0	1	13	24	8
spring-typical	0	0	6	14	14

Quantile Regression Forest

	very non-spring	non-spring	spring-compl.	cond. spring-typ.	spring-typical
very non-spring	0	0	1	1	0
non-spring	0	0	4	6	0
spring-compl.	0	0	7	27	1
cond. spring-typ.	0	0	12	33	1
spring-typical	0	0	6	23	5

Cost-sensitive RF

	very non-spring	non-spring	spring-compl.	cond. spring-typ.	spring-typical
very non-spring	0	0	1	1	0
non-spring	0	0	7	2	1
spring-compl.	0	1	15	13	6
cond. spring-typ.	0	0	14	22	10
spring-typical	0	0	5	15	14

predicted

observed

class weighting

	very non-spring	non-spring	spring-compl.	cond. spring-typ.	spring-typical
very non-spring	0	0	1	1	0
non-spring	0	0	7	2	1
spring-compl.	0	0	15	16	4
cond. spring-typ.	0	0	14	25	7
spring-typical	0	0	7	9	18

dynamic weighting

	very non-spring	non-spring	spring-compl.	cond. spring-typ.	spring-typical
very non-spring	0	0	1	1	0
non-spring	0	0	6	3	1
spring-compl.	0	0	11	21	3
cond. spring-typ.	0	0	12	28	6
spring-typical	0	0	7	13	14

threshold adjustment ROC

	very non-spring	non-spring	spring-compl.	cond. spring-typ.	spring-typical
very non-spring	0	1	1	0	0
non-spring	1	1	7	0	1
spring-compl.	0	4	23	5	3
cond. spring-typ.	0	8	24	8	6
spring-typical	0	0	12	6	16

up/downsampling

	very non-spring	non-spring	spring-compl.	cond. spring-typ.	spring-typical
very non-spring	0	0	1	1	0
non-spring	0	0	5	3	2
spring-compl.	0	3	13	14	5
cond. spring-typ.	0	3	12	24	7
spring-typical	0	0	6	10	18

weighted tree voting

	very non-spring	non-spring	spring-compl.	cond. spring-typ.	spring-typical
very non-spring	0	0	1	1	0
non-spring	0	0	7	1	2
spring-compl.	0	0	13	15	7
cond. spring-typ.	0	0	15	22	9
spring-typical	0	0	5	14	15

A20 Additional Performance Metrics for Random Forest Classification Models Predicting Fauna Categories Based on Mixed Spring Structure and Landscape & Topographic Variables Data

	Precision very nonspring	Precision nonspring	Precision spring-compliant	Precision cond. spring-typ.	Precision spring-typ.	Recall very nonspring	Recall nonspring	Recall spring-compliant	Recall cond. spring-typ.	Recall spring-typ.
Baseline RF	0.00	0.00	0.43	0.52	0.44	–	0.00	0.33	0.44	0.56
Regularized RF	0.00	0.00	0.54	0.33	0.35	–	0.00	0.35	0.37	0.41
Cond. Inf. Forest	0.00	0.00	0.40	0.50	0.32	–	–	0.33	0.35	0.58
Extr. Random. Trees	0.00	0.00	0.26	0.52	0.41	–	0.00	0.26	0.37	0.52
Quantile Regr. Forest	–	–	0.23	0.37	0.71	0.00	0.00	0.20	0.72	0.15
Cost-sensitive RF	0.00	0.00	0.43	0.48	0.41	–	0.00	0.36	0.42	0.45
Class weighting	0.00	0.00	0.43	0.54	0.53	–	–	0.34	0.47	0.60
Dynamic weighting	0.00	0.00	0.31	0.61	0.41	–	–	0.30	0.42	0.58
weighted tree voting	0.00	0.00	0.37	0.48	0.44	–	–	0.32	0.42	0.45
threshold adjustment	0.00	0.07	0.34	0.42	0.62	0.00	0.10	0.66	0.17	0.47
up/downsampling	–	0.00	0.35	0.46	0.56	0.00	0.00	0.37	0.52	0.53

Personal Declaration

Personal declaration: I hereby declare that the submitted thesis is the result of my own, independent work. All external sources are explicitly acknowledged in the thesis.

Nicholas von Holzen

Zürich, 26.08.2025