Site Conditions and the Vulnerability of Norway Spruce to Bark Beetle Infestation:

A Case Study from the Bergisches Land in Western Germany

Master Thesis

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

submitted within the University Master Program for Further Education
"Geographical Information Science & Systems – (UNIGIS MSc)"
at the Department of Geoinformatics - Z_GIS,
Faculty of Digital and Analytical Sciences,
University of Salzburg
submitted by

B. Sc. Andreas Köller

Supervisor:
Dr. Christian Neuwirth

Köln, July 2026

Declaration

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

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

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

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

Andreas Köller

Abstract

Norway spruce (*Picea abies*) suffered widespread mortality in the Bergisches Land, a low mountain region in western Germany, during the 2018 to 2024 calamity caused by the European spruce bark beetle (*Ips typographus*). Because spruce will form a considerable share of the naturally regenerating forests, practitioners face the question of where spruce can be maintained as a part of future forests at an acceptable level of risk. This thesis looks at whether three site characteristics can explain the timing of bark beetle attack, namely slope, aspect, and the soil moisture level. These were selected because they are stable over time, can be derived across the whole area, and can be assessed in the field. The study area covers two forest districts in the Regionalforstamt Bergisches Land. The vulnerability of Norway spruce stands was operationalized through the timing of attack. Stands attacked earlier were considered more predisposed than stands which were attacked later.

The time of attack was the response variable and obtained by manual classification based on digital aerial imagery and provides a date of first attack for each spruce stand. Two predictor variables were derived from a digital elevation model and one from the forest site mapping (FSK 5) of the Geologischer Dienst NRW. The variables were co-registered on a 10 m raster grid and aggregated to 1470 stands, of which about 1100 entered the ordinal analysis. The relationships were tested across two spatial scales and three measurement levels: nominal (χ^2), ordinal (Kendall τ-b and Spearman ϱ), and metric (linear regression).

Slope and aspect showed no association with the timing of attack on any scale or measurement level. The soil water regime showed a weak but robust and highly significant monotone trend, with drier stands attacked earlier and wetter stands later (Kendall τ-b = 0.14, $p < 10^{-5}$; about 0.4 years of delay per moisture step, $R^2 \approx 0.02$). The large sample size supports the assumption that the absence of a slope or aspect effect is a credible negative result. The weak signal was attributed to the narrow moisture gradient, the low elevation relief of the Bergisches Land, and the static nature of the site mapping. During mass outbreaks however, the population dynamics of the bark beetle may override site predisposition. The selected forest site variables are therefore only of limited use for predicting future vulnerability, but the approach offers a solid basis for further development.

Future studies should determine attack timing at a finer temporal resolution, distinguish more clearly between site predisposition and modelling of bark beetle spread, and include dynamic water availability predictors which capture actual moisture data. Study areas covering broader environmental gradients may also help to determine whether the observed relationship with soil moisture becomes more pronounced under a broader range of site conditions.

Zusammenfassung

Die gemeine Fichte (*Picea abies*) erfuhr während der Kalamität von 2018 bis 2024 beispiellose Ausfälle durch den Buchdrucker (*Ips typographus*) im Bergischen Land, einer Mittelgebirgsregion in Westdeutschland. Da die Fichte auf den Schadflächen einen großen Anteil an den sich wieder verjüngenden Wäldern haben wird, stehen Forstpraktiker vor der Frage, wo die Fichte mit vertretbarem Risiko erhalten werden kann. Diese Arbeit prüft, ob drei Standortseigenschaften, nämlich Hangneigung, Hangexposition und Wasserhaushaltsstufe des Bodens, den Zeitpunkt des Borkenkäferbefalls erklären können. Diese wurden ausgewählt, weil sie zeitlich stabil, aus vorhandenen Datenquellen flächig gebildet werden können und im praktischen Einsatz im Gelände erkannt werden können. Das Untersuchungsgebiet sind zwei Forstbetriebsbezirke im Regionalforstamt Bergisches Land. Die Anfälligkeit von Fichtenstandorten wird über den Befallszeitpunkt operationalisiert, sodass früher befallene Bestände stärker prädisponiert gelten als später befallene.

Die Zielvariable (Zeitpunkt des Befalls) wurde durch eine manuelle Klassifikation auf Basis von digitalen Luftbildern gewonnen und liefert für jeden Bestand einen Zeitpunkt für den Erstbefall. Die Prädiktorvariablen wurden aus einem digitalem Geländemodell und der forstlichen Standortkartierung (FSK 5) des Geologischen Dienstes abgeleitet. Die Variablen wurden auf einem 10m Rastergrid zusammengeführt und auf 1470 Bestände aggregiert, wovon etwa 1100 Bestände in die ordinale Analyse eingingen. Die Zusammenhänge wurden auf zwei räumlichen Ebenen und drei Messniveaus getestet: nominal (χ^2), ordinal (Kendall τ-b und Spearman ϱ) und metrisch (lineare Regression).

Hangneigung und Hangausrichtung zeigten auf keiner Ebene und keinem Messniveau einen Zusammenhang mit dem Befallszeitpunkt. Die Wasserhaushaltsstufe zeigte einen schwachen, aber robusten und hoch signifikanten monotonen Trend, bei dem trockenere Bestände früher und feuchtere später befallen wurden (Kendall τ-b = 0,14, $p < 10^{-5}$; etwa 0,4 Jahre Verzögerung je Feuchtestufe, $R^2 \approx 0{,}02$). Bei der großen Stichprobe ist das Fehlen eines Hangneigungs- oder Expositionszusammenhangs ein belastbares Negativergebnis. Das schwache Signal wird auf den schmalen Feuchtegradienten, die geringe Höhenstufe des Bergischen Landes sowie den statischen Charakter der Standortskartierung zurückgeführt. Zudem überlagert bei Massenvermehrung die Populationsdynamik der Borkenkäfer die Standortsunterschiede. Die ausgewählten Standortsvariablen sind daher nur beschränkt geeignet, um zukünftige Anfälligkeit vorherzusagen, aber der Ansatz bietet viel Potenzial und ist flexibel ausbaubar.

Künftige Ansätze sollten eine deutlich höhere zeitliche Auflösung des Befallszeitpunktes inkludieren, versuchen Standortprädisposition vom Modelling des Ausbreitungsgeschehen zu trennen und statische Eingangsvariablen durch dynamische Prädiktoren ersetzen, die den tatsächlichen Feuchteverlauf abbilden. Um zu prüfen, ob das schwache Feuchtesignal bei breiteren Standortgradienten deutlicher wird, müsste das Untersuchungsgebiet ausgeweitet werden.

Contents

Table of Figures

1. Introduction

1.1 The severity of bark beetle infestation in Europe and Nordrhein-Westfalen

Globally, bark beetles are a major forest disturbance in all coniferous forests in the northern hemisphere. Although bark beetle outbreaks are a natural component of forest dynamics, their frequency, severity, and spatial extent are increasingly being altered by climate change (Biedermann *et al.*, 2019). Allen et al. states that global forest vulnerability is increasing and that the threat of beetles in combination with droughts is an systematically underestimated threat (Allen, Breshears and McDowell, 2015). Within Europe, insect damage is the most dominant source of biotic forest damage and twice as prevalent as pathogens and wildlife herbivory (Kautz et al., 2017). Among these insects, European spruce bark beetle *ips typographus* is considered the most important forest pest in Europe (Fettig et al., 2022) and main disturbance agent for Norway spruce (*Picea abies*) (Weed, Ayres and Bentz, 2015; Singh *et al.*, 2024). On a European scale, bark beetles doubled their share of overall forest damage compared to the historical average from 8 to 17 % percent of forest damage (Patacca *et al.*, 2023).

Forest disturbance by *ips typographus* results in serious economic damage, since spruce has long been the backbone of commercial forestry in central Europe (Huber *et al.*, 2023). For Germany, the damage has been carefully estimated to be 12,7 billion euros between 2018 and 2020 (Möhring *et al.*, 2021). Besides economic damage, disturbance events by bark beetles may pose significant problems for ecosystem health (Thorn *et al.*, 2018), ecosystem services such as carbon storage (Seidl *et al.*, 2014) but also social problems, for example reduced tourism in affected landscapes (Kortmann *et al.*, 2022). Despite these negative impacts, the bark beetle can also contribute to higher biodiversity by replacing spruce with more diverse tree species in the long term. In his 2021 review of bark beetle ecology, Tomáš Hlásny (2021) provides a comprehensive overview of economic, ecological and cultural impacts of bark beetle damages and calls for more societal awareness of climate-change driven risks in forests.

From 2018 to 2020, central Europe experienced a previously unknown level of bark beetle outbreak. For the whole of Germany, 290 million m^3 of conifers were lost, which equates to 20 % of the entire spruce timber volume as reported by the BMLEH (*Massive Schäden - Einsatz für Wälder und Waldumbau nötig*, 2025). This represents the largest forest cover loss in Germany since the timber harvesting for war reparations after World War 2. Particularly affected Norway spruce forests were located in the low mountain ranges, especially in regions such as the Sauerland, Harz Mountains, Thuringian Forest, Westerwald, parts of Bavaria and also the “Bergisches Land” in NRW, which is the region which the study area is located.

In this wider context, the following study focuses on a specific disturbance event in Germany which occurred from 2018 to 2024 in the study region of the “Bergisches Land”. This disturbance is generally simply called the “Borkenkäferkalamität” in German, which is a broad term for the bark beetle infestation as a whole, but also often also includes the initial drought and storm damages. During the peak of the calamity, Spruce cuts increased to nearly the 8 times of the timber harvest compared to the long term average (*Holzeinschlag und Schadholz | Statistik.NRW*, 2026).

The chronological damage in Nordrhein-Westfalen is presented in Figure 1, with the main years of the infestation being 2019 to 2021 by timber volume.

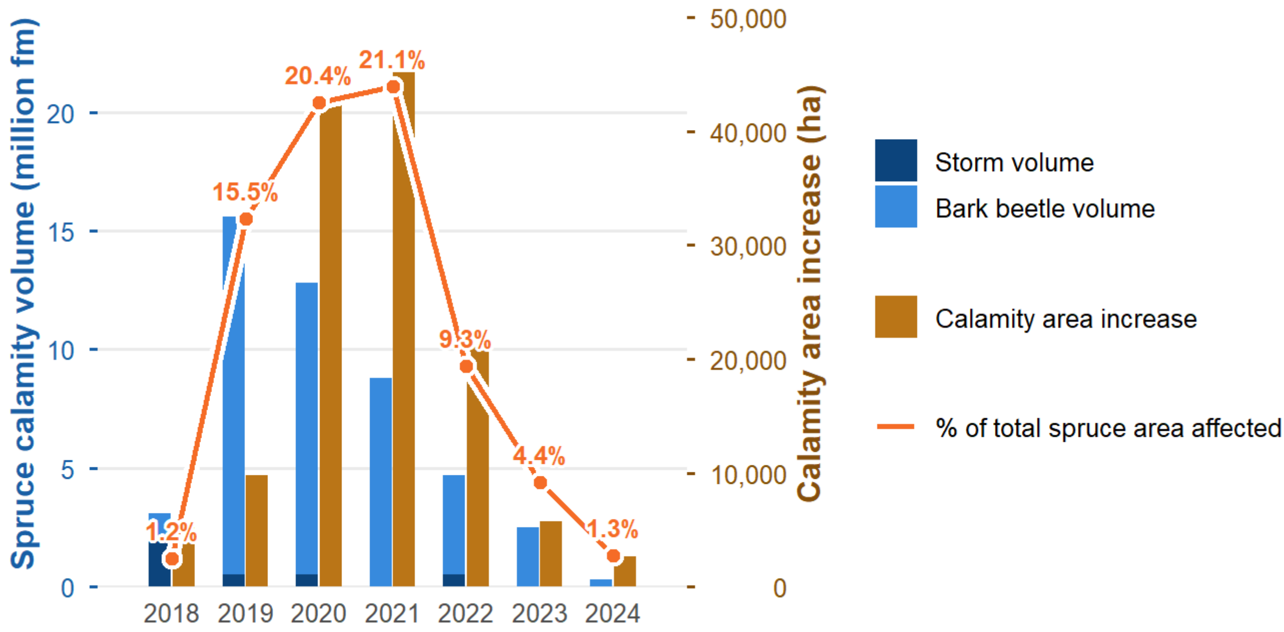

Figure 1: Temporal progression of the spruce calamity in terms of timber volume, forest area and percentage of total spruce area. Data source is the Waldzustandsbericht 2024 NRW, the visualization is my own.

These historical damages strained forest management over the last few years. However, regenerating these forest areas will present even greater challenges over the coming decades, as forest managers have to guide reforestation while facing a shortage of planting material, limited financial resources, and the uncertainties of climate change. The agricultural ministry of Nordrhein Westfalen discusses this context in their reforestation strategy paper in various facets ('Wiederbewaldungskonzept NRW', 2024).

Large forest areas affected by the bark beetle calamity will regenerate with Norway spruce as a natural component of the future forest on the same area. Spruce does already cover 33 % of the newly regenerated forest areas (*Waldzustandsbericht NRW 2024 (lang)*, 2024, p. 60). For forest managers, the key challenge is therefore not whether spruce should be retained as part of the species mixture, but where it can be maintained with an acceptable level of risk. On more predisposed areas active conversion to more resilient tree species should be prioritized. Assessing spruce vulnerability is therefore not just an analysis of the past but essential for long-term strategic forest planning and will allow for more efficient allocation of management resources.

Before introducing the research questions and hypothesis, the following sections will summarize the ecological background of bark beetle outbreaks and review previous research on remote sensing of bark beetles. Subsequently, the knowledge site factors which can predispose spruce to bark beetle infestation will be summarized. This provides the scientific basis for selecting the three site variables investigated in this study.

1.2 Current state of research

1.2.1 The ecology of the bark beetle -spruce interaction

While Bark beetle biology should not be the main focus here, some fundamentals about the population dynamics should be discussed, as they are relevant for understanding the ecology behind site conditions potentially predisposing spruce for bark beetle. For the most comprehensive review of the biology of the *ips typographus*, I refer to the review of Beat Wermelinger (2004).

Bark beetles lay their eggs in galleries under the bark of host trees. The larvae continue eating the phloem under the bark, thus cutting off the vertical transport of water and nutrient, weakening and killing the tree. After metamorphosis from larvae to beetle, they fly out, disperse for a range of a few 100m or few km (Biedermann *et al.*, 2019) and bore into the next host tree. As host trees, they normally target trees older than 60 and larger than 20 cm diameter, but under outbreak conditions they are less selective (Wermelinger, 2004). The dieback of spruce takes several weeks and progresses through three characteristic stages: the green stage, during which trees are successfully colonized but needles remain green; the red stage, when needles turn reddish-brown as the tree dies; and the gray stage, in which needles have fallen, leaving dead trees with bare branches. These phases are important for detection, regardless if by satellite, orthophoto or by a forester in the field.

Bark beetles have a population dynamic that alters between endemic (low-density) and epidemic (high-density) population phases. Biological traits that enable rapid population growth are sister broods (=multiple batches by one female beetle) (Davídková and Doležal, 2017) and voltinism (=multiple generations within one growing season)((Jönsson *et al.*, 2009). Hallas and Colleagues (2026) illustrate this in their graphic (Figure 2), which shows the typical exponential population growth in outbreak phases. Schroeder, Knape and Kärvemo (2025) documented this dynamic for spruce stands in Sweden and also showed that sanitary cuts are important to reduce outbreak severity.

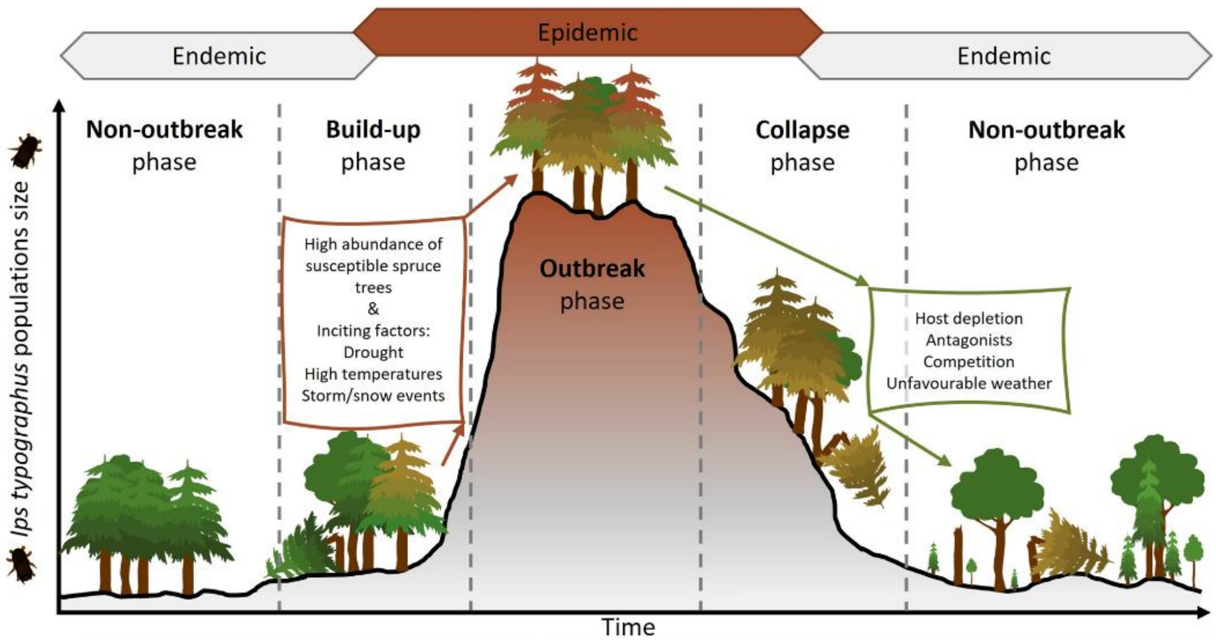

Figure 2: Typical population dynamic of bark beetle populations. Figure taken from Hallas et al. 2026. The figure was altered from the original by cropping out the lower half. The original paper and the figure are explicitly permitted for use and alterations under a creative commons license http://creativecommons.org/licenses/by/4.0/.

Intraspecific competition is one of the main drivers of bark beetle population (Toffin *et al.*, 2018). In plain words this means that as long as susceptible host species are available, competition between beetles doesn't exist and populations can grow rapidly. This highlights that for managing bark beetles in endemic phases, spruce vitality shapes susceptibility to bark beetles. In outbreak phases however, also healthy trees can succumb to the sheer masses of beetles. To analyse the vulnerability of spruce to bark beetle infestation means to look at factors which can trigger a bark beetle shift towards an outbreak phase.

In the past, the disturbance dynamic of bark beetles often followed this exponential pattern, but on a restricted scale. Okland outlines how bark beetle typically occurred in the past in a case study in Slovakia (Økland *et al.*, 2016): The classic case of bark beetle was windthrow of a small forest area, a bark beetle infestation in the toppled tree, a small outbreak at the edges of this windthrow site. This could be contained with moderate sanitary cuts of infested trees. Besides wind throw, hot and dry weather can initiate an outbreak phase (Marini *et al.*, 2017; Netherer *et al.*, 2021). These factors are strongly affected by climate change. Biedermann and colleagues reviewed bark beetle ecology on a global scale. Their synthesis is that in the Anthropocene, outbreak severity and frequency will go up, while tree health goes down (Biedermann *et al.*, 2019).

1.2.2 Literature Review of Remote sensing research of bark beetle

There are various ways in which remote sensing can help practical forest management in regard to spotting, mapping and predicting bark beetle infestations. GIS based research can inform forest management both on an operational level as well as on a strategic level of forest planning.

Early detection can help forest management in operational phases by improving bark beetle monitoring and enabling quick, targeted sanitary cuts. Timing is the priority here to prevent the exponential growth of bark beetle populations. Key advantages of Remote Sensing approaches lie in the standardized surveillance across administrative levels and gathering consistent data over the years. Satellites, DOP from flights and UAV can also help spotting damages in more remote parts of a forest district. For these reasons, a lot of remote-sensing based research about bark beetle focused on early detection (Holzwarth *et al.*, 2020), which means detecting bark beetle in the green phase or early red phase of infestation. This is especially true for the immediate years of the bark beetle calamity in Europe 2018-2020. Due to global coverage and high temporal resolution of Sentinel 2, much research focused on how to get most of the spectral resolution. Many studies tested various satellite combinations (König *et al.*, 2023) or Indices (Abdullah *et al.*, 2019). One emphasis lies on using the NIR Band /red green edge to detect loss of vitality as soon as possible (Huo, Persson and Lindberg, 2021).

Another hope is that machine-learning will help with finding patterns of ecological trends in the huge amount of available data and methods (Rammer and Seidl, 2019; Marvasti-Zadeh *et al.*, 2023), with the core methods predominantly based on either Support vector machine or Random forest classification (Sheykhmousa *et al.*, 2020). Early detection by UAV instead of plane or satellite also showed promise (Klouček *et al.*, 2024) and could be more scalable in the long term.

Kautz and Colleagues (Kautz, Feurer and Adler, 2024) critically reviewed remote sensing approaches for detecting *Ips typographus* infestations and concluded that although no single spectral index consistently outperformed others, red-edge and shortwave infrared (SWIR) wavelengths showed the greatest potential for early detection. He explicitly states that early detection is not possible with enough spatial and temporal resolution to be relevant for practical management. More recently, Candotti argued similarly, and posed that early detection is challenging, and even more so when other stress events like drought are present in the same forest area (Candotti *et al.*, 2026).

A bit later, various studies popped up which aimed at mapping the extent of the forest damage. Such Mapping and Monitoring of bark beetle has been done in various areas, e.g. Italy (Dalponte *et al.*, 2022; Francini *et al.*, 2022), Germany (Thonfeld *et al.*, 2022; Reinosch *et al.*, 2025) and Czech republic (Fernandez-Carrillo *et al.*, 2020). In order to this, they also have to classify the forest areas as detailed as possible. The main difference compared to early detection research is there is less focus on the temporal edge, but more on spatial accuracy and mostly on covering large study areas or entire regions. This trend has been also stated by Holzwarthm et al. (2023). The motivation behind these studies might be to inform policy makers by quantifying the damage done by the forest calamity in numbers (area affected, volume of timber lost).

1.2.3 Literature Review of predisposing Factors

Compared to the large body of work on detecting current infestations or mapping damage extent, few studies use spatial data to derive practical stand-level vulnerability indicators. Forest management must allocate its management resources as efficiently as possible, knowing where to replace spruce and where to grow it one more generation is decisive for this.

In our specific case this means identifying which spruce stand will be susceptible in the future. To answer this, we must know what steers bark beetle dynamics. In the literature, the questions about which factors drive bark beetle outbreaks are approached on various spatial scales and from different ecological angles. Singh states that despite that being topic with a lot of research focus for more than 100 years, there remain many gaps our understanding of the main drivers (Singh *et al.*, 2024). Nonetheless, I will discuss the main groups of variables that are often used to predict bark beetle outbreaks.

Landscape level predictors

Some studies approach bark beetle predisposition at broader landscape scales, with variables such as host-tree connectivity, host abundance, forest fragmentation or edge density (Lausch, Fahse and Heurich, 2011; Seidl *et al.*, 2016). In one study in eastern Europe, dead spruce was found to be strongest predictor in bark beetle infestation on a landscape level (Stereńczak *et al.*, 2020). These factors are useful for explaining outbreak spread across large, forested areas, but are too coarse for practical forest management in the study area which looks a the single forest stand.

Temperature

From an ecological perspective temperature is a key driver of bark beetle populations. It drives flight activity and the number of possible generations per year. Besides air temperature (Singh *et al.*, 2024), thermal sum is another potential variable influencing bark beetle infestation, as it represents the amount of heat available for beetle development. For example, the PHENIPS phenology model, uses accumulated thermal sum as its core variable to predict bark beetle flight activity and the potential number of generations (Bentz *et al.*, 2019).

However, temperature is difficult to operationalize in susceptibility mapping because it can be represented in various different ways, including air temperature, thermal sums, elevationinterpolated temperature , bark temperature, canopy surface temperature (Zakrzewska and Kopeć, 2022), and solar-radiation-based proxies (Mezei *et al.*, 2019).

Elevation

The terrain or DEM based variables are easy to obtain and often used. Elevation strongly influences mean annual temperature and thus is a great predictor, but within our study area the elevation range is not substantial enough (see 2.1.1, Study area).

Slope

Slope has been used as a predictor variable in the study by Nardi and colleagues, (Nardi *et al.*, 2023), but they don't state exactly how they derive it. With a one km grid, resolution is very low as well.

Carillo calculates slope for 25 m grid cells, which is a lot better on spatial resolution (Fernández-Carrillo *et al.*, 2024). Slope impacts spruce by increasing water run-off and more solar radiation, which can both be stressors of spruce and thus increase susceptibility. A study in Sweden (Huo, Persson and Lindberg, 2024) calculated slope for each forest stand.

Aspect

Aspect describes the direction of a sloped surface in relation to the sun. More sun means higher temperature. Aspect was used as predictor variable in the alps (Candotti and Tomelleri, 2026), Czechia (Fernández-Carrillo *et al.*, 2024) and also in Nordrhein-Westfalen (Xu *et al.*, 2024). The study in Sweden (Huo, Persson and Lindberg, 2024) also calculated aspect on a stand level.

Aspect is also good, simplified proxy variable for solar radiation, but a lot easier to obtain. Solar radiation itself is a potentially very useful predictor (Mezei *et al.*, 2019).

Water availability

Water stress can strongly affect spruce vitality and thus susceptibility. It can lead to the closing of stomata of spruce, therefore reducing photosynthesis, carbon storage and defence compounds. In short, it weakens and predisposes spruce.

Drought increases bark beetle disturbance in severity (Kärvemo *et al.*, 2023) and generally becomes an increasingly important facilitator of bark beetle under climate change (Das *et al.*, 2025).

However, similar to temperature, water stress is difficult to operationalize because it can be represented in several ways: through precipitation deficits, drought indices such as the Standardized Precipitation Evapotranspiration Index (SPEI), soil moisture or available water content, soil water storage capacity, or remote-sensing indices such as NDWI. For example, Xu et al. (2024) used soil moisture-related predictors, while Candotti and Tomelleri used NDWI and NDMI to characterize canopy water status. Huo et al uses an ordinal wetness scale of 1-dry, 2-moist, 3-wet (Huo, Persson and Lindberg, 2024). This makes water-related predictors highly relevant, but also methodologically variable across studies. Seidl et al- (2016) also states that water availability is more important than temperature for bark beetle infestation in Central Europe.

Stand Structure

Structural parameters of a forest stand are great predictors. Age of Spruce stands is a relevant factor, as bark beetle (*ips typographus*) typically only attacks mature stands older than 30-40 years

(Wermelinger, 2004). This can also be seen for Nordrhein-Westfalen, where the average forest age for spruce dropped from 70 years pre-calamity to 35 years in 2024, showing that younger stands survive (*Waldzustandsbericht NRW 2024 (lang)*, 2024, p. 27). Other structural aspects used are factors like diameter and crown closure prior infestation (Candotti *et al.*, 2026), Tree height (Schmidt *et al.*, 2024) and generally past treatment history (Schmied *et al.*, 2022).

However, since they are specific to individual stand and their history, they are not suited to predict general spruce susceptibility on the same area.

1.3 Practical perspective and selection of site variables

While many variables can be used to predict spruce susceptibility to bark beetles in principle, not all of them are suitable for practical use. Some predictors are too abstract, too difficult to calculate, change too much over time, or are simply not available for individual forest stands. For practical forest management, predictor variables should ideally be directly observable in the field or available from a simple risk-map product. I therefore judged relevance of predictor variables with four practical criteria: (1) predictive basis, meaning that there is ecological and empirical evidence that the variable influences bark beetle susceptibility; (2) spatial resolution, meaning that the data is detailed enough to distinguish individual forest stands; (3) data availability, meaning that the variable is freely accessible or practical to obtain for forest managers and (4) temporal stability, meaning that the variable remains largely unchanged over time and is therefore useful for long-term planning.

Dynamic variables such as temperature, drought indices, or current bark beetle pressure may have a strong predictive basis, but they are often difficult to use at stand level because they change over time or are not available at a sufficiently fine spatial resolution. Other site descriptors, such as elevation, nutrient or temperature balance, may be useful at larger scales, for example across NRW or Germany, but are too homogeneous within the Bergisches Land to explain local differences between stands.

For this reason, this study focuses on aspect, slope, and soil moisture regime. These variables have stable site characteristics and are available at fine spatial resolution. Aspect was selected because it influences incoming solar radiation and therefore local temperature and soil moisture conditions. Slope was selected because it can influence runoff, water availability, and terrain exposure. Soil moisture regime was included as a direct site-based indicator of potential water limitation and taken from an available map product aimed at forest manager. They will be explained in more detail in the method section. Together, these variables represent practical and stable predictors.

1.4 Research gap

Holzwarth and Colleagues provided two reviews about forest monitoring with remote sensing in 2020 and 2023 (Holzwarth *et al.*, 2020, 2023). They argued that so much research was done in the wake of the forest damage, that a second review was necessary after 3 years. She points out, however, that despite all this research no practical products for foresters were created.

Existing approaches often rely on large-scale, grid based models or detailed forest inventory data that limit their operational applicability (Nardi *et al.*, 2023; Piedallu *et al.*, 2023; Fernández-Carrillo *et al.*, 2024; Candotti and Tomelleri, 2026). The research gap is therefore not the absence of bark beetle risk models, but the limited testing of simple, stable site variables as stand-level predictors (Huo, Persson and Lindberg, 2024). Taking the concept of cell-based research from the landscape level to a small 10m grid and then aggregating it up to the stand level is the specific angle of this thesis.

This study addresses this gap by testing whether terrain aspect, slope, and soil moisture regime can explain the timing of bark beetle infestation in spruce stands for small study area.

1.5 Aim of the study

The aim of this study is to evaluate whether stable site characteristics do influence the vulnerability of Norway spruce to infestation by *Ips typographus*. The study tests bark beetle susceptibility by analyzing the order in which spruce stands were attacked during the 2018–2024 outbreak. The longterm objective is to assess whether these variables can serve as the empirical basis for an prioritization tool or map for future spruce stands. In doing so, it scientifically checks the commonly hold assumption of foresters that northern facing and wet stands are more resistant.

1.6 Hypothesis

In this study, vulnerability is operationalized by the timing of bark beetle infestation. The regional outbreak between 2018 and 2024 is treated as a single infestation event, with stands attacked earlier considered more vulnerable than stands attacked later.

The main hypothesis is that terrain **aspect**, terrain **slope**, and **soil moisture regime** influence the vulnerability of Norway spruce (*Picea abies*) to infestation by *Ips typographus*. South- and southwest-facing stands, steeper slopes, and drier sites are expected to be attacked earlier than average, reflecting a higher degree of predisposition.

For **aspect** it is assumed that south facing stands will be more vulnerable, as they receive more solar radiation and thus higher temperatures, leading to increased bark beetle activity.

For steeper **slopes**, a higher water runoff and thus increased water stress for spruce is assumed, leading to higher vulnerability.

For **Soil moisture regime**, increased water stress is assumed for drier sites, leading to higher vulnerability.

2. Material & Methods

2.1 General Workflow

The study was conducted in three separate operative steps. First, the response variable "Time of attack" was created by manually classifying a polygon layer of damaged spruce areas. In the second step, all predictor variables were processed and then co-registered with the response variable into a virtual raster layer raster of 10m cells. In the third step, two dataframes were built for a pixel and an aggregated stand level and the association between each site parameter and attack timing was tested statistically. Figure 3 illustrates the workflow. Additionally, table 1 provides an overview of all data sources used in this study design. The Link to all data sources is found in the appendix.

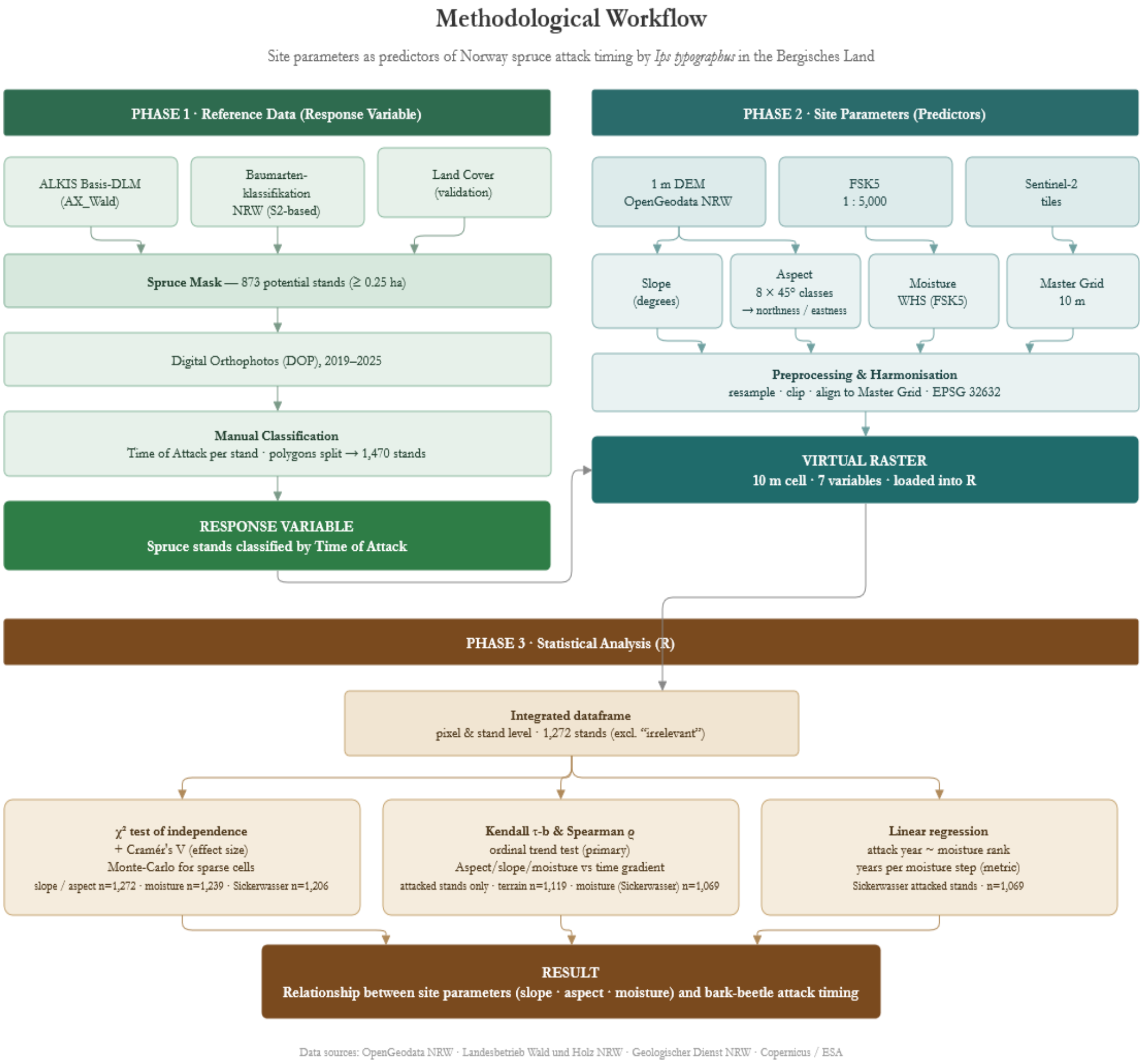

Figure 3: Conceptual Workflow of the Thesis. Each operative step is illustrated as colourblocked phase. The illustration provides an overview of data sources, process orders and analytical steps. The illustration was created using draw.io.

Table 1: Overview of Data Sources used in this study

Layer	**Purpose**	**Source**	**Time of Capture**	**Data Type**	**CRS (EPSG)**	**Spatial Reso-lution**
Forest Mask (administrative)	Creating Spruce Mask	ALKIS Basis DLM, Bezirksregierung Köln	14.4.2026 (last update)	Polygon	25832	10 m
Forest Mask (actual Land Cover)		WMS NW Landbedeckung, Tim-OnlineNRW.de	01.01.2023 - 31.12.2023	Polygon	25832	1 m
Forest Damage Classification	Assessing Spruce Mask	Landesbetrieb Wald und Holz Nordrhein-Westfalen, via OpenGeodata.NRW	06.2017 – 09/2023	Polygon	25832	10 m
Tree species Classification			3.03.2022 und 11.07.2023	Polygon	25832	10 m
Digitial Orthophotos (DOP)	**Reference** Source for classifying time of attack	Geobasis Bezirksregierung Köln & QGIS Plugin "Luftbildfinder NRW)	16, 19, 21, 23, 24, 25	DOP	25832	10/20 cm
Moisture Level	Base Data for **Soil Moisture**	Forstliche Standortskarte 1:5000, Geologischer Dienst NRW, via OpenGeoData.NRW	2023	Polygon	25832	10 m (5m Offset from Sen2)
DEM	Base Data for Calculating **Slope** and **Aspect**	OpenGeoData.NRW	2022	tif	25832	1 m
Sentinel 2	Provides **Master Grid** for Analysis	Copernicus Browser	01.06.2021	tif	32632	10 m

2.2 Study area

2.2.1 Topography and Forest Cover

The study area covers two forest districts (Forstbetriebsbezirke Kürten and Wermelskirchen, = FBB) in the natural Region "Bergisches Land", which extends to west of Düsseldorf and Cologne and is named after the Duke of Berg, and not due to its hilly topography. The location of the study area is illustrated in Figure 4. On the west, the Bergisches Land borders on the low-lying river plains of the Rhine valley, while on the east it slowly changes into the higher elevated Sauerland.

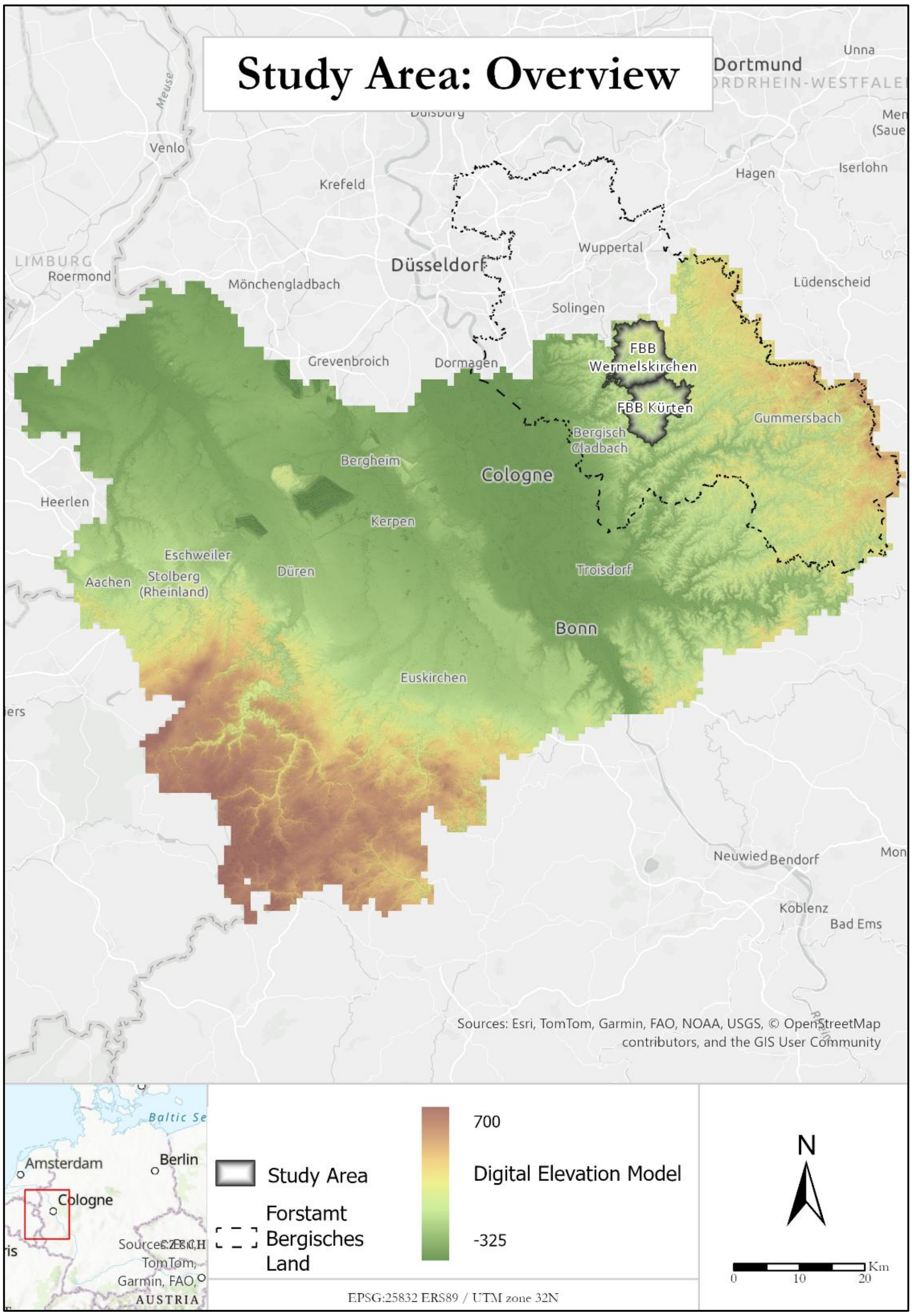

Figure 4: General Location of the Study area. The DEM extent is the administrative unit of the Regierungsbezirk Köln. The low negative elevation is due to the open pit mining of brown coal west of Cologne.

The natural geography of the study area can be characterized as a hilly highland of around 250 m, which is furrowed by small to medium river valleys. Cut across the relief, the topography resembles a rounded sawtooth profile. The altitude ranges from 90m to 360m at the highest point in the study area, with a mean of 227m. So, while the total elevation is not very high, there is still a substantial amount of pronounced elevation changes. In these steeper terrains forest cover is especially high, as these areas weren't converted as much to agriculture and pasture areas historically. Figure 5 illustrates how the forest areas are mainly situated alongside the small river valleys. In total, the area is around 20 x 10 km wide and has around 5.933 ha or 59.3 km² of forest. Forest Cover in the Study area is spread out all over the municipalities Kürten and Wermelskirchen. While there are some larger interconnected forest areas, the forest area is generally not aggregated and mostly follows the small river valleys and creeks.

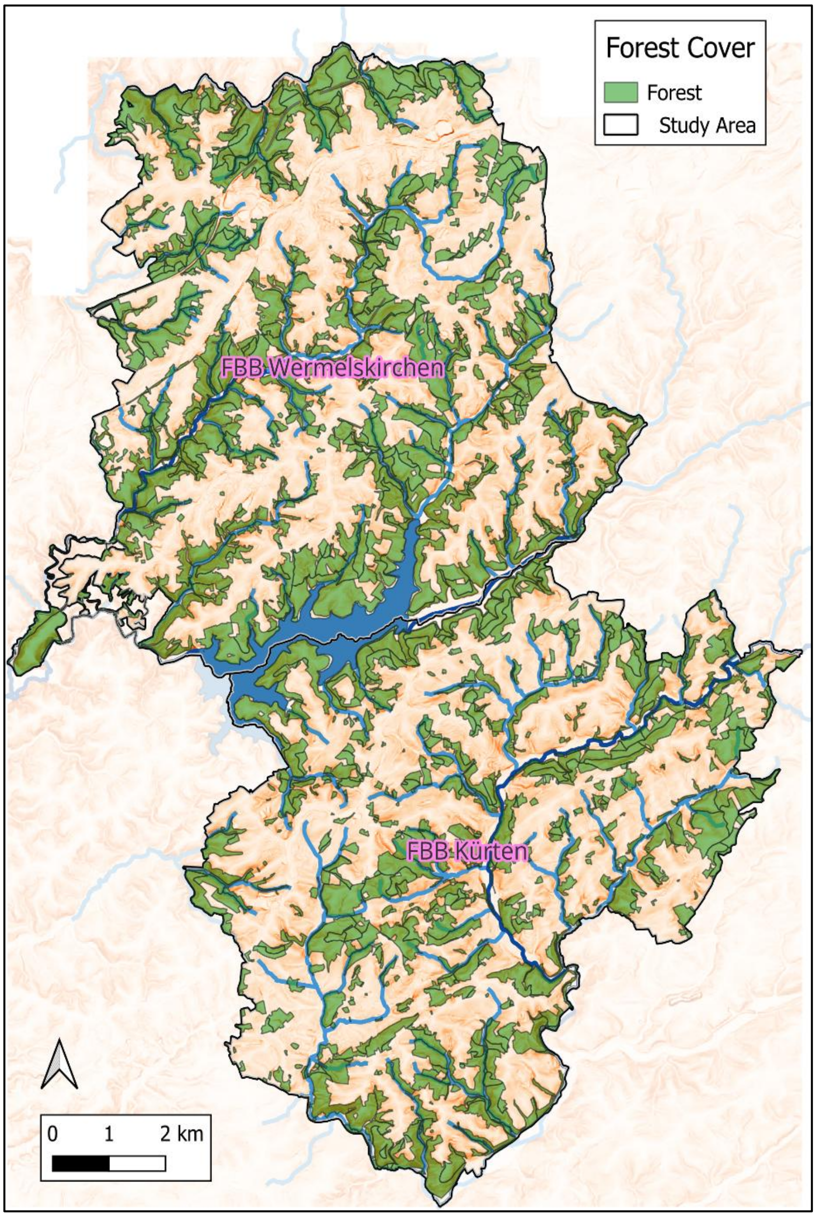

Figure 5: Detailed Overview of the two forestry districts in the study area, separated by the Dhünntalsperre

2.2.2 Climate, Soil & Forest History

The Bergisches Land sits right on the border of the maritime/continental climate divide, which diagonally splits Europe. Precipitation in this region has been historically very high, annual mean precipitation from 1980-2010 for the weather station in Wermelskirchen is 1351 mm (*dwd.de, Klimadaten Mittelwerte*, 2026). Precipitation comes from the maritime west with westerly winds. Soils are very homogenous for this study area, they are dominated by poor to average nutrient based cambisols. Temperature balance can be considered uniform for this study area as well.

After heavy deforestation since the 1800s, Norway spruce was planted in the study area since the Prussian forest administration took over in the 1840s, and especially after reforestation after WW2 in the 50s and 60s (Schmidt, 2017). Even as Spruce was always planted outside their natural elevation range and ecological niche in the Study area, the unusual high rainfall compensated for it and allowed spruce to be a major part of forest area and timber stock in the region for several decades.

A further characteristic of forests in this area is a highly divided forest ownership structure. Most forest areas are owned by private owners, with most of them only owning a few patches. In extremes, the forest lots can only be 2-5 m in width. This results in a highly fragmented and irregular forest management, which greatly adds to the mosaic structure of the forest areas. Irregular management also means that remote sensing can be an asset, as it does not rely on ownership records or keeping track of forest inventory data.

2.2.3 Administrative Area and Professional relation of the author

The two forest districts (Forstbetriebsbezirke) used in the study were those assigned to the author, who worked for the Landesbetrieb Wald und Holz Nordrhein-Westfalen. He was responsible for FBB Kürten from October 2021 -September 2024, and FBB Wermelskirchen from February 2023 to September 2024. The personal knowledge of the study area helped in the manual classification and interpretation of the results.

2.3 Creating the Time of Attack layer (Response Variable)

A layer containing the extent and time of attack information of spruce stands affected by bark beetles is the basis of the analysis, as it is the response variable in the statistical analysis. The following section will outline how 1) a layer containing all potentially relevant spruce stands was derived from freely available data sources and 2) these potential areas were classified manually to assign them a time of attack.

2.3.1 Forest (Spruce) Mask

Alkis Basis DLM

The Boundary of forests were based on the Basis DLM for the state North Rhine Westphalia (*Digitales Basis-Landschaftsmodell - Paketierung: gesamt NRW*, 2026). The class "AX_Wald" represents forests within the topographic landscape model. This DLM Dataset provides accurate forest boundaries but does not contain any information on forest composition or tree species. In this study it was used only as a clipping boundary for all other input variables to reduce computing demands.

Tree species classification NRW

The forest agency for Nordrhein Westfalen, the "Landesbetrieb Wald und Holz Nordrhein-Westfalen" conduced a remote sensing-based tree species classification (*Baumartenklassifikation NRW - Open.NRW*, 2026). It's a neural network-classification derived from Sentinel 2 and thus has 10m pixel resolution. Classification shows the dominant tree species within a cell, but not individual trees. However, for our analysis, this is sufficient in spatial resolution, as we work on stand level and not tree level. For Classifying spruce, classification accuracy shows both low errors of omission and commission, so the data is reliable with an f -measure of 97,6 (*Baumartenklassifikation NRW - Open.NRW*, 2026). The Classification also includes forest areas that were already fully cleared and currently without tree cover, which for this timeframe can be assumed to be exclusively spruce. These two classes represent both former and current spruce stands and were combined into a single spruce mask.

Validating and Filtering Spruce damage layer

The plausibility of the layer with potential spruce area was visually assessed against the actual physical Land Cover Layer class, which identifies areas covered by coniferous tree vegetation (*Landbedeckung NRW*, 2026). This classification is produced through automated analysis of Sentinel-2 imagery, orthophotos, elevation data (nDOM), and ALKIS reference data. The 2023 dataset is based on cloud-free Sentinel-2 observations collected throughout 2023 (Dec22 – Dec 23) and achieves a reported user accuracy of 90.8% with a geometric accuracy of 1–2 m.

The conifer Land cover class typically does not only cover spruce, but all conifers. The difference in the study area should be negligible, since pine (*pinus sylvestris*) is very rare, and there are only very few stands of alternative conifer species that are present in the two forestry districts. These are Douglas fir (*Pseudotsuga menziesii*) which is still restricted to young stands and Grand Fir (*abies grandis*), which only has been cultivated in very few trial stands which location is known to the author.

The comparison shows that both layers had very similar coverage, with the land cover classification showing less coverage. More importantly however, is that all potential damage areas are shown to previously be classified as conifer and thus with a high likelihood spruce.

This spruce mask was then filtered by area from 14848 to 873 spruce stands with a minimum size of 2500m^2 or 0,25 hectare. The aim was to filter out artefacts and scattered spruce area across the landscape as well as to reduce stands before visual classification. The Alkis Basis DLM was used to select only spruce stands within official forest boundaries. Otherwise, small tree groups could appear along roads, farms and in gardens in residential areas. 0,25 hectares is a good cut-off point from a practical perspective as it is a stand size where it becomes feasible to plan and actively manage the forest. Figure 6 illustrates why small stands and stands outside official forest boundaries were discarded.

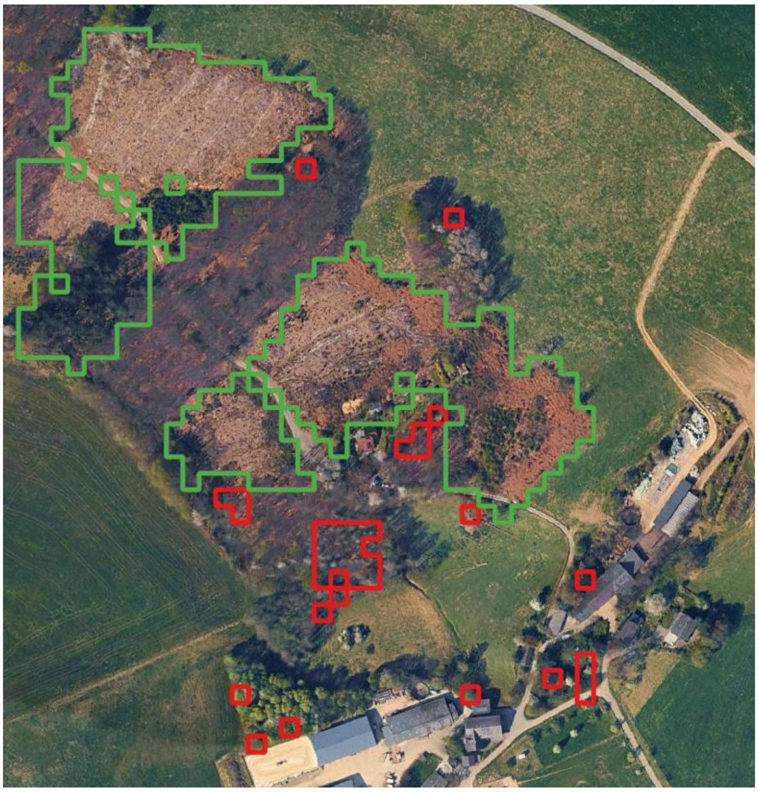

Figure 6: Green =(Former) Spruce areas higher than 0,25 ha, red= discarded areas. Reasons to discard were small stand size and if trees were outside official forest boundaries

2.3.2 Manual Classification

After successfully creating a layer containing all damaged spruce stands, each individual stand was assessed and classified through visual classification in the next step. For assessing the time of forest bark beetle attack, Digital Orthophotos (DOP) were used as a reference source. In North Rhine Westphalia, survey flights were routinely done within a three-year cycle in the past, and in 2020 the frequency was increased to two years. Since 2023 additional flights are done with a lower 20cm resolution to achieve full coverage for each year. However, there is no clear schedule within a year, so month and state of vegetation at the actual time of capture vary. Data was accessed under the OpenData NRW agreement (*WMS DOP*, 2026) and by using the Plugin "Luftbildfinder NRW" (2026). Table 2 shows the exact DOP flight dates available for the study area.

Table 2: Overview of Digital Orthophotos used for classification

Year	**DOP Time of Capture**	**DOP Season**	**DOP Bands**	**Spatial Resolution**	**Corresponding Phase / Time of Attack**
2016	20.7.2016	Summer	RGB	10cm	1
2019	16.4.2019	Spring	RGB	10cm	2
2021	1.6.2021	Early Summer	RGB	10cm	3
2023	19.4.2023	Spring	RGB	10cm	4
2024	13.4.2024	Spring	RGB + NIR False Color	20cm	5
2025	2.7.2025	Summer	RGB	10cm	6

Each Class represents a time of Attack, e.g. if bark beetle damage was first visible in 2023, it was classified into the respective phase. Next to the six possible Time of Attacks (ToA), there were two other classes: 1) irrelevant areas, which were not spruce stands, and which could be clearly discarded from further analysis 2) potential spruce stands, which were not or only minimally affected, labelled here as surviving. Table 3 shows the distribution of the classification. As expected, Time of Attack 2021 and 2023 are most frequent, while later classes are very sparse.

Table 3 :Classification scheme for manual Classification

Time of attack	**Phase ID**	**Count (Stands)**	**% of Total**
irrelevant	0	198	13.5
2016 (prior Outbreak)	1	0	0
2019	2	171	11.6
2021	3	564	38.4
2023	4	326	22.2
2024	5	44	3.0
2025	6	14	1.0
Surviving	10	153	10.4
Total		**1470**	**100**

Figure 7 shows for an exemplary area how infested spruce stands are classified using available DOPs. Besides spruce stands being cleared, it is noticeable that the 2023 flight was before the start of vegetative period.

Figure 7: Spruce stands west of Kürten, classified according to the first visible signs of stand infestation

It should be made clear that the classification was conducted at the forest **stand** level, not for individual trees. A stand is a somewhat loosely defined term but usually refers to a group of forest trees which are sufficiently uniform in species composition, age, and condition to be considered a homogeneous unit for management purposes. Single dead trees can be an indicator for bark beetle presence but is not sufficient for a classification a stand as “infested”.

Figure 8 shows the lower left forest stand from figure 7 in more detail in the 2024 False Color Infrared, in which the dead trees can be seen especially well. While there are some affected trees and there is also small clearing in the north, the stand is still homogeneous and fully present, which is why it is classified as 10 = not infested.

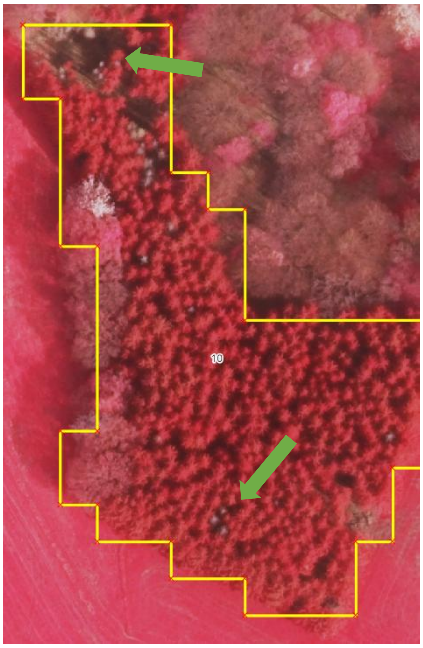

Figure 8: Example of a surviving stand, even if slight bark beetle damage is visible, the stand is still structurally sound

It was important to also split up potential stand polygons into several parts, if required. Stands were often not fully attacked at the same time. Whenever it was possible to delineate areas into polygons of different phases, they were split up, as illustrated in figure 9. The number of stands polygons thus increased from 873 potential areas to 1470 stands.

A new Stand unique ID was created for each damaged area, and the area calculated for each forest stand.

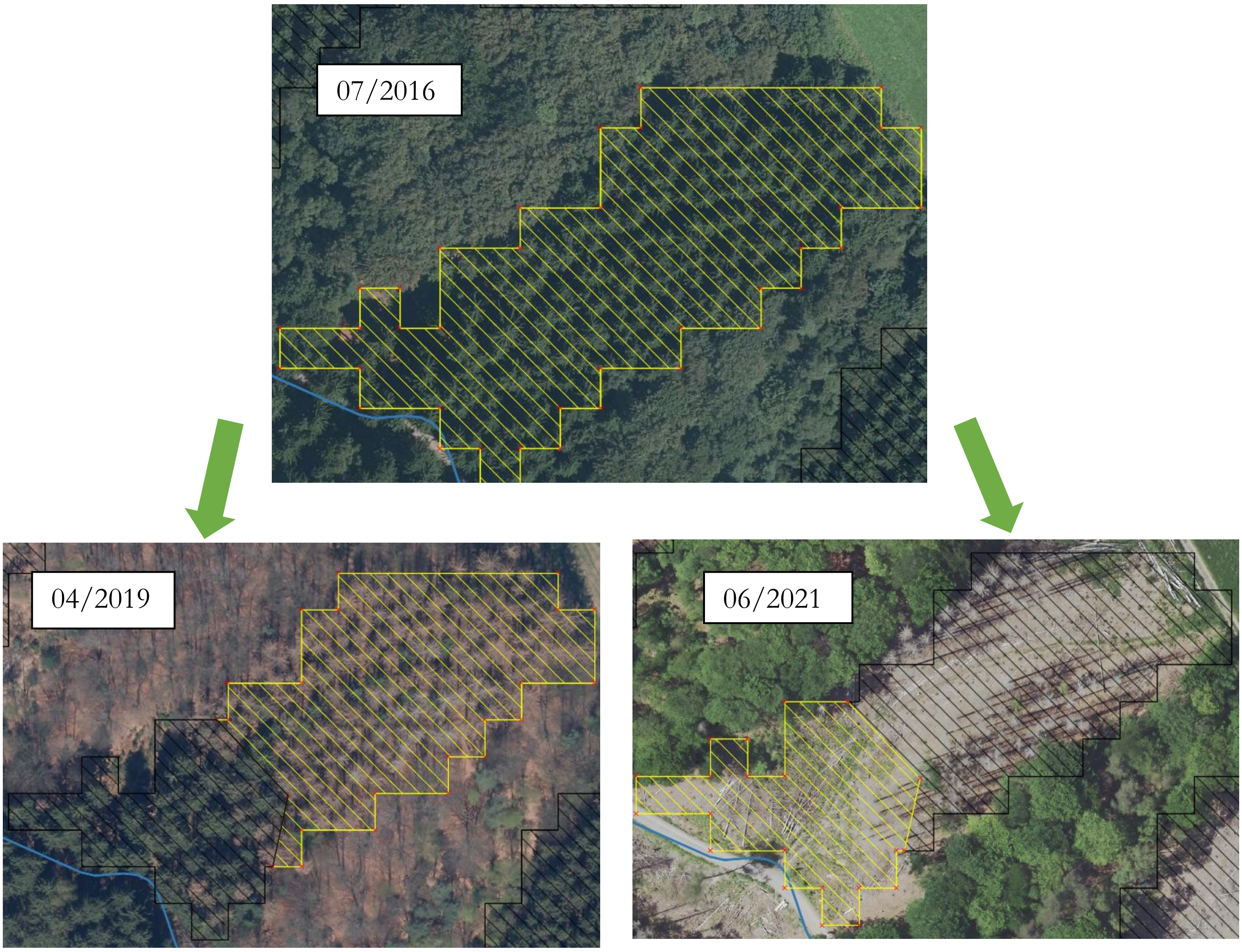

Figure 9: Example of a Stand being split due to different times of attack

2.4 Site Parameters (Predictor Variables)

2.4.1 Slope

The Slope was calculated on the basis of the 1m DEM from OpenGeodataNRW and calculated in degrees. Several processing steps were required to resample and align the final layer to the Master Grid. In short, this was a reprojection of the CRS to EPSG 32632, then the 1m Grid was resampled to a 10m DEM grid using the "average" resampling, as the mean height was required. Lastly the raster was aligned with the target grid with *gdalwarp*. From this 10m DEM slope (and Aspect) can then be directly calculated.

In order to sequence these processing steps and to be able to quickly test different target grids, a custom processing script was created. In this Process Claude Opus 4.7 was used, primarily to add the syntax for a QGIS processing tool script flow and to add convenient steps like automatical clipping or handling irregular no data values. The Script is *Script_DEM_Resampling* and in the appendix.. The tool GUI is shown in Figure 10. Slope was calculated based on Horns method in degrees.

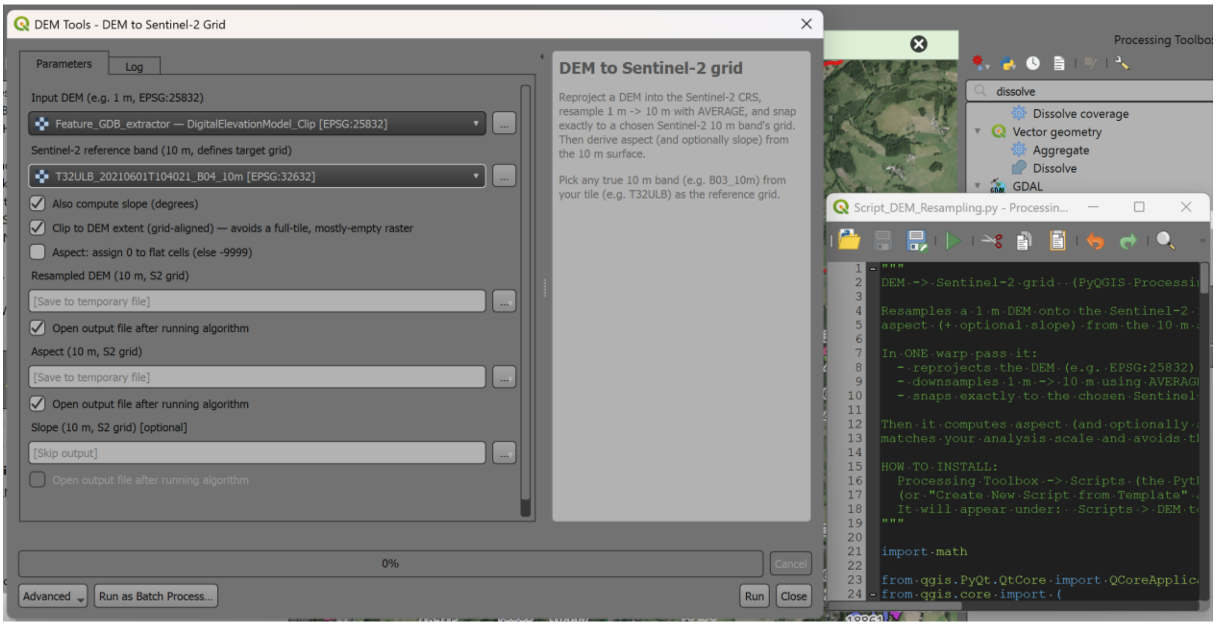

Figure 10: Input Parameters for Custom Processing script

Figure 11 shows the map of slope across the study area. It illustrates that the slopes are mostly adjacent to the river valleys and small upstream creeks. Based on the histogram of the slope distribution, slope classes which get larger in the steeper sections were chosen (see figure 12).

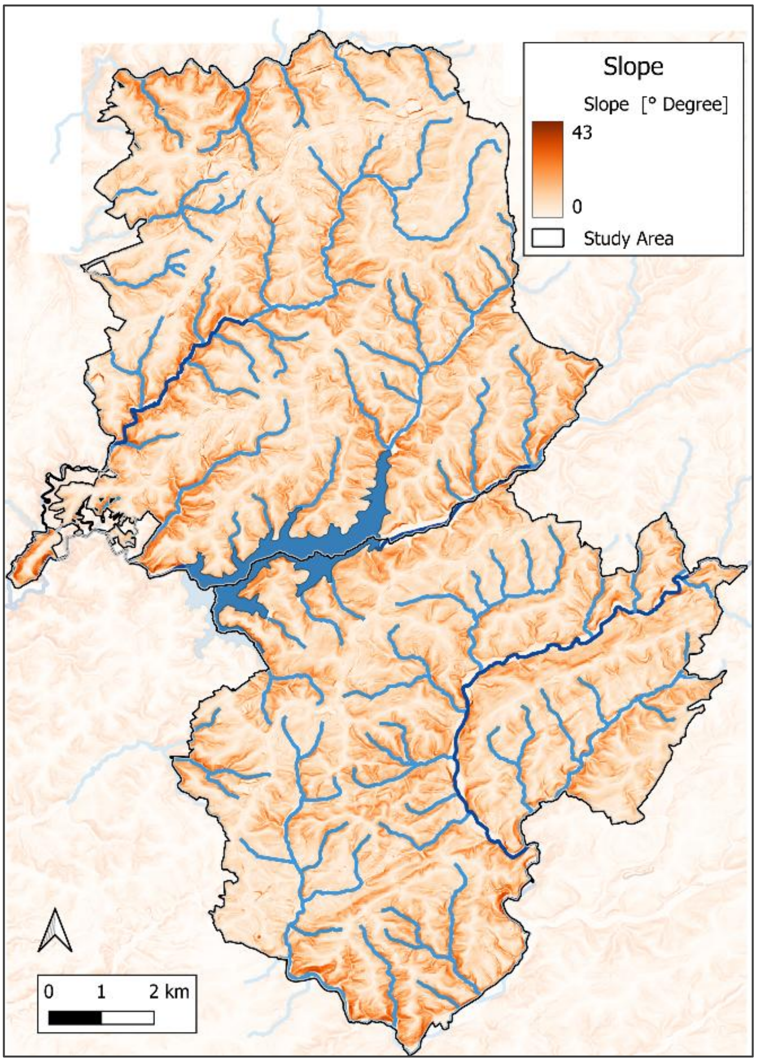

Figure 11: Map of Slope over the study area

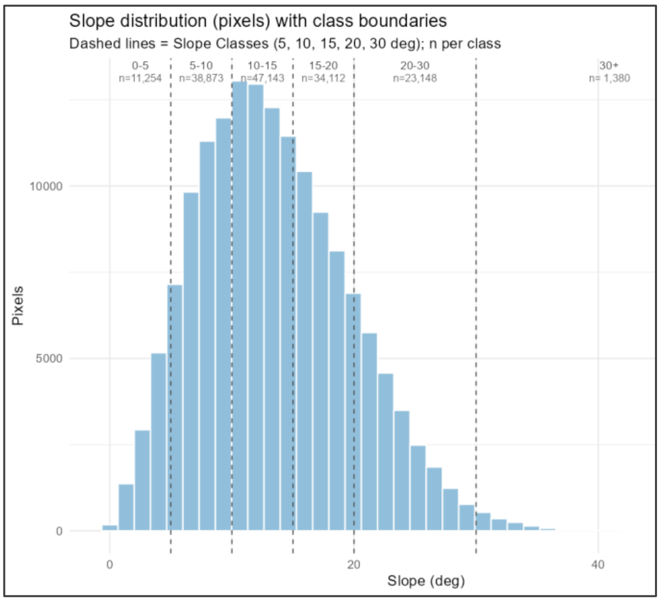

Figure 12: Histogram of Slope of the 10m pixel scale

2.4.2 Aspect

Aspect was derived from the same 10m DEM and calculated in compass degrees (0-360°). A map of the aspect in shown in Figure 13, while Figure 14 shows the distribution of aspect classes (each 45° width) per class.

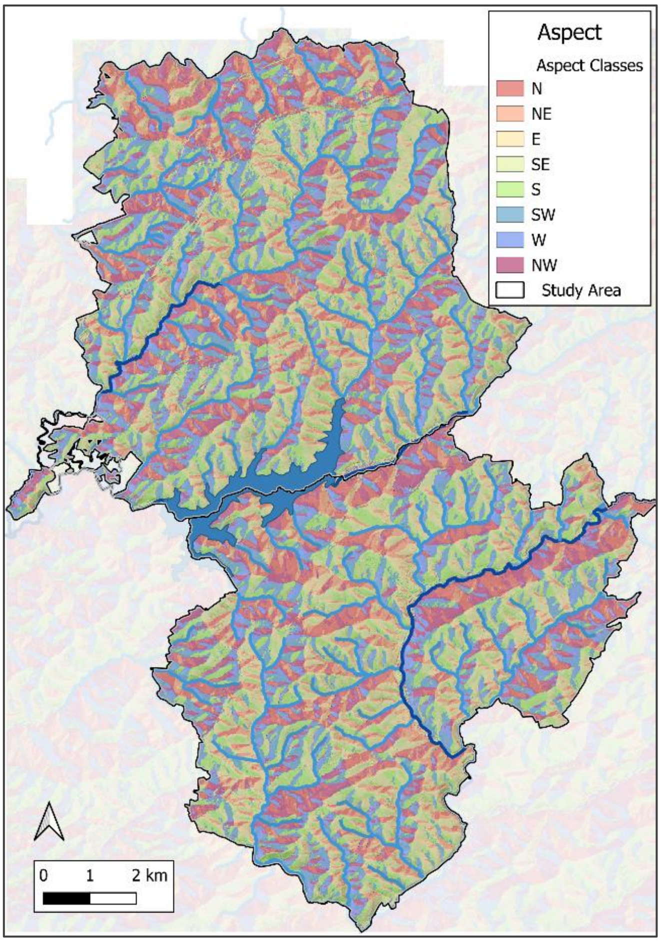

Figure 13: Map of Aspect across the Study area

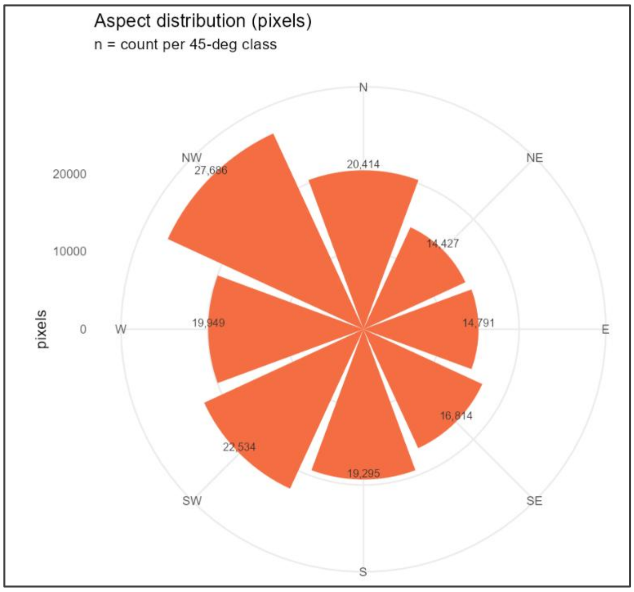

Figure 14: Distribution of pixels across Aspect classes

2.4.3 Moisture Level (Hydrology Data)

The predictor variable soil moisture in this study is based on the “Forstliche Standortkarte 1:5.000” (FSK5). This map is produced by the Geological Survey of North Rhine-Westphalia (“Geologischer Dienst NRW”) and provides detailed information on ecological forest site conditions. The unique property is its high spatial resolution. The scale of 1.5000 is specifically designed to support forest owners and practitioners in operational forest management in the field and to offer site-specific growing conditions. This spatial resolution is crucial for forest planning, as stand conditions dictate silvicultural choices available to the forest manager.

The FSK5 itself is the output of a model which aims to bring together different data source and combines them into one map product (Schulte-Kellinghaus and Weller, 2023). The site classification is based on three main environmental input components: soil, climate, and relief (topography). Soil-related parameters include soil stratification, water storage capacity, and drainage characteristics. As this study analyzes past and current stands, climatic information is derived from the DWD climate normal period 1981–2010, although alternative variations of this map product based on future climate scenarios (RCP 4.5 and RCP 8.5) are also available. Relief information is calculated from a 10-meter digital elevation model and incorporates terrain-related factors, such as subsurface lateral water flow ("Hangwasserzug"). For this reason, predictor variables are not completely independent.

The resulting site characterization is expressed through three key ecological indicators: water balance (Gesamtwasserhaushalt), nutrient availability (Nährstoffversorgung/Basengehalt), and vegetation period, represented by the average number of days per year with mean temperatures exceeding 10 °C. In this study only the overall water balance (“Gesamtwasserhaushalt”) is used. It represents the moisture regime of a site and is provided as a categorical variable describing site water availability for forest growth. Three different water sources can influence the classification. Precipitation, which passes from the top downwards and thus also referred as Seepage water (“Sickerwasser” in German). In the statistical analysis it is referred to as moisture level. Water regime can be also influenced from the bottom upwards, either due to groundwater or due to a dense layer in the soil. Figure 15 illustrates a typical water level balance distribution in the study area, while Figure 16 shows the overall distribution across classes.

Figure 15: FSK 5 north the Dhünntalsperre in the Study area, with the legends for the three main ecological indicators for forest growth nutrient avaliabiltiy, water balance, length of vegetative period (as displayed on the Website Waldinfo.NRW)

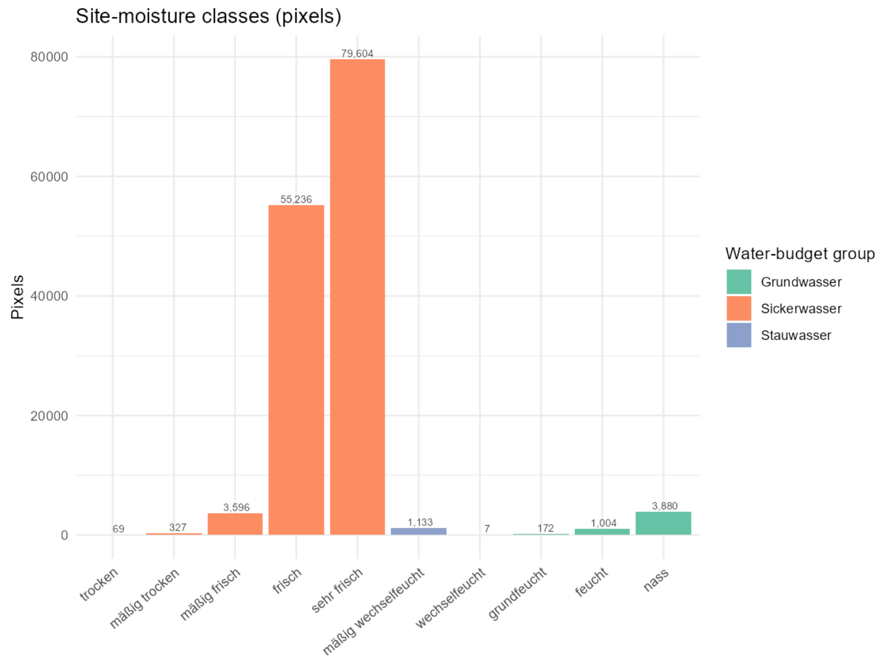

Figure 16 Distribution of moisture level classes over the study area

The FSK5 was available as a shapefile. An attribute field was created for mapping the categorical data to an integer field- Then the data was reprojected to EPSG 32632 and then rasterized. The source layer for FSK5 used a pixel-center (point) registration, whereas gdal_rasterize and the Sentinel-2 reference grid follow a pixel-corner (area) convention. To avoid a half-pixel (5 m) misalignment, the target extent was shifted by half a cell so that the burned grid's pixel edges exactly overlay with the Sentinel-2 10 m grid again.

Assumption for converting Moisture Class to ordinal data

The raw data for moisture level is a categorical variable. While it roughly follows a gradient from dry to wet, it mixes different hydrological mechanisms, which basically can be distinguished between the source of the water shaping the soil. If one only focuses on the six moisture classes which are shaped by precipitation (Sickerwasser), it is acceptable to order them along a gradient from dry to wet, thus making them an ordinal variable. This assumption enables more statistical analysis which requires this ordinal level. Table 4 shows all hydrological regimes and highlights the classes used in subsequential analysis.

Table 4: Overview of Moisture levels occurring in the data set. For ordinal analysis, only the classes in bold with green background were used. The Coulour scheme follows the original used in Waldinfo.NRW

Hydrological Regime	Colour	Class (dry → wet)	Class present in study area?
Sickerwasser / seepage-dominated (rainwater dominated)		sehr trocken	absent
		trocken	yes (pixels only)
		mäßig trocken	**yes**
		mäßig frisch	**yes**
		frisch	**yes**
		sehr frisch	**yes**
Stauwasser / perched-water-dominated		wechseltrocken	absent
		mäßig wechselfeucht	yes
		wechselfeucht	yes
		staunass	absent
Grundwasser / groundwater-dominated		grundfrisch	absent
		grundfeucht	yes
		feucht	yes
		nass	yes

2.4.4 Sentinel 2 Data

In the context of this study, the Sentinel 2 Pixel Tile structure was used as the Master Grid, onto which all other Data sources were aligned and transformed into the Coordinate Reference System EPSG 32632. To be exact, the Scene T32ULB was used as reference, as it fully covers the study area.

There were two main reasons for using this grid. Firstly, the tree species classification at the heart of the reference layer is already based on Sentinel 2 10m Pixels. Secondly, 10-meter Pixel are a good analytical level to analyze forest and Terrain Data at the stand level, as they smooth out chatter from the 1m terrain level. Additionally, it can be built upon with any further Sentinel based data for potential further research.

2.5 Statistical analysis

2.5.1 Structure of the Dataframe

For statistical analysis, all input data was rasterized and aligned to the same 10m master Grid. Then, a virtual raster layer was calculated for the study area, containing all co-registered raster layers. This Virtual Raster could then be exported and read into R. The statistical analysis was done with R, and R Studio (last version 2026.05.0+218, using the packages "terra", "dplyr", "arrow", "tidyr" and "ggplot2"). Claude Opus 4.8 was used to assist with R syntax and code documentation. Outputs and assumptions were carefully checked. Figure 17 and Table 5 show the raw data in QGIS and as an explanatory table are explanatory illustrations.

The full code of the following analysis can be found on GitHub https://github.com/andikoeller/Thesis_Barkbeetle_Koeller . A more refined version, formatted as markdown blog is found here https://andikoeller.github.io/Thesis_Barkbeetle_Koeller/ . The latter offers a more direct view of the code right next to its output.

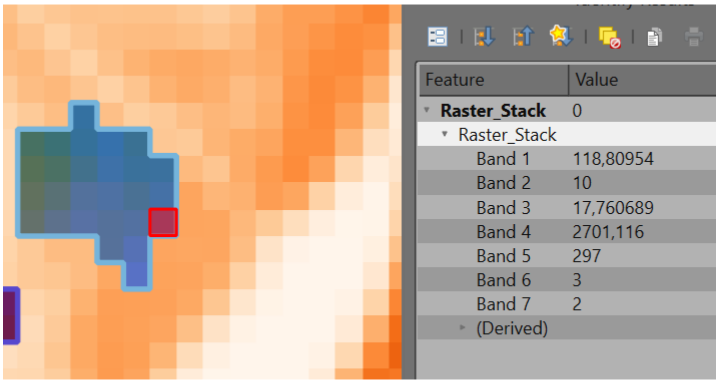

Figure 17: Screenshot of Virtual Raster Layer in QGIS

Table 5: Table of the variables contained in the Virtual Raster Layer. Only the variable in bold were used in further testing

Band	Variable	Type	Role in analysis
1	**aspect**	continuous (0-360 deg)	Basis for: aspect class, northness/eastness (predictor)
2	**phase**	categorical	Basis for: TimeOfAttack (response)
3	**slope**	continuous (0- 90 deg)	Basis for: slope class (predictor)
4	stand_area	continuous	stand attribute: not used in this study
5	**stand_id**	categorical	aggregation key (pixel to stand)
6	VisKlass2	categorical	Secondary classification for surviving stands, not used in this study
7	**moisture**	categorical	Basis for site water-budget class (predictor)

Based on this raw data, two data frames were built in R on two spatial resolutions: the pixel level and the stand level. The pixel data frame has one row per 10-meter raster cell and is built directly from the co-registered stacked raster. It stores the response, the year of first attack, together with the site predictors slope, aspect and soil moisture class. When building the dataframe, NoData values are cleaned, aspect is classed into eight 45-degree wide Aspect classes (North, Northwest, West etc.), slope is classed into six classes, and the integer moisture and attack phase codes are decoded into readable labels. The result is a table with 155.910 pixel cells (Figure 18).

	cell_id	x	y	aspect	phase	slope	stand_area	stand_id	VisKlass2	moisture	aspect_class	slope_class	moisture_text	TimeOfAttack
25195	753897	375735	5663895	303.466858	3	17.061384	25122.336	1234	NA	8	NW	15-20	sehr frisch	2021
25196	753898	375745	5663895	324.632996	3	15.093374	25122.336	1234	NA	8	NW	15-20	sehr frisch	2021
25197	753899	375755	5663895	340.061432	3	15.014377	25122.336	1234	NA	8	N	15-20	sehr frisch	2021
25198	753900	375765	5663895	350.358887	3	14.634381	25122.336	1234	NA	8	N	10-15	sehr frisch	2021
25199	753901	375775	5663895	3.208397	3	13.980709	25122.336	1234	NA	8	N	10-15	sehr frisch	2021
25200	753902	375785	5663895	15.000252	3	14.184576	25122.336	1234	NA	8	N	10-15	sehr frisch	2021
25201	754020	376965	5663895	231.066101	4	16.170408	10160.267	1434	NA	8	SW	15-20	sehr frisch	2023
25202	754021	376975	5663895	233.323639	4	22.723906	10160.267	1434	NA	8	SW	20-30	sehr frisch	2023
25203	754022	376985	5663895	228.945908	4	21.623474	10160.267	1434	NA	8	SW	20-30	sehr frisch	2023
25204	754023	376995	5663895	223.540466	4	18.868603	10160.267	1434	NA	8	SW	15-20	sehr frisch	2023
25205	754024	377005	5663895	219.245178	4	17.240618	10160.267	1434	NA	8	SW	15-20	sehr frisch	2023
25206	754025	377015	5663895	216.692184	4	16.861748	10160.267	1434	NA	8	SW	15-20	sehr frisch	2023
25207	754034	377105	5663895	169.323425	10	14.135324	15117.510	1226	4	2	S	10-15	frisch	surviving
25208	754035	377115	5663895	169.325333	10	13.808164	15117.510	1226	4	8	S	10-15	sehr frisch	surviving
25209	754036	377125	5663895	168.735504	10	13.730757	15117.510	1226	4	8	S	10-15	sehr frisch	surviving

Figure 18: The table shows an excerpt from the dataframe in R on pixel level, with all required input and output information, as described in the text.

The second dataframe is on the stand level. The stand data frame has one row per forest stand and is built by aggregating the pixel within each stand, based on their Stand_ID. The variables are aggregated depending on their data type. Categorical predictors are summarized by their modal class, which is simply the class that most of the pixels in a stand belong to. For example, if a stand has 100 pixels and 70 of them are "sehr frisch", the whole was classified as "sehr frisch". Slope was aggregated by the median value. Figure 19 shows an example of the structure and the data values.

	stand_id	TimeOfAttack	n_pixel	stand_area	slope_median	northness	eastness	slope_class	aspect_class	moisture_text
1	1	2021	61	6102.573	15.878913	-0.1158358925	-0.981255707	15-20	W	sehr frisch
2	2	surviving	30	2876.145	10.265155	-0.2549252980	-0.947200941	10-15	W	sehr frisch
3	3	2023	27	2701.246	15.808555	0.7949016302	0.347138637	10-15	NE	frisch
4	4	2021	28	2801.252	15.000443	-0.1345381621	-0.904088820	15-20	W	frisch
5	5	2019	53	5310.733	10.888275	-0.9581168948	-0.213275569	10-15	S	frisch
6	6	2021	629	62934.648	8.091779	-0.5901971625	-0.275379992	5-10	SW	sehr frisch
7	7	irrelevant	35	3501.613	18.759665	0.6465420457	-0.744765873	15-20	NW	sehr frisch
8	8	2023	82	8042.958	8.210864	-0.8037307303	0.429497605	5-10	S	sehr frisch
9	9	2019	54	5402.350	16.608835	-0.4111650683	-0.382186214	15-20	W	sehr frisch
10	10	irrelevant	30	3001.385	10.282673	-0.3884618468	0.655547895	5-10	SE	frisch

Figure 19: Dataframe in R of the Stand level, showing all relevant parameters used for later analyis. This dataframe adds eastness, northness and the slope median

Aspect cannot be averaged directly because it is a circular variable; a stand facing just east of north (10 degrees) and one facing just west of north (350 degrees) point in almost the same direction, yet their numeric values are far apart, and a plain average would wrongly place the stand facing south. To avoid this, aspect is split into two linear components, northness (the cosine of aspect) and eastness (the sine of aspect), ranging from minus one to plus one, which can be averaged across the pixels of a stand. Northness is plus one for a north facing pixel and minus one for a south facing one, and eastness is plus one for east and minus one for west. Linear gradients also allow for linear regression later. This decomposition follows the approach of Amatulli and colleagues (Amatulli *et al.*, 2018).

The stand is the actual analytical unit of the study. Pixels within share the same site conditions and are strongly spatially autocorrelated. Treating them as independent observations would be pseudo replication with excessive significance. For this reason, the pixel level is only descriptive, it shows that aggregating the pixels to stands preserves the effect sizes instead of distorting them.

The number of stands tested in each step was narrowed down stepwise and is shown in Figure 20. Of the 1470 potential spruce stands 198 were discarded as irrelevant, because they were clearly very young forest stands, which are not affected by bark beetle, or they were non deciduous forest or other vegetation, but not spruce. The 153 surviving stands were kept in the nominal chi-square as an ordinary category but left out of the primary ordinal test as they don't have a time of attack. Including them would shift the study design from when spruce die to if spruce stands die. Nonetheless, surviving stands were used to check robustness of the trend. This leaves 1119 attacked stands, which form the terrain set used for the slope and aspect tests. For the moisture tests the sample is restricted further to the ordered seepage-water (Sickerwasser) gradient, which drops 50 stands. It omits them either because they have no moisture mapping (25 stands) or because their hydrology is shaped by groundwater (20 stands) or perched water (5 stands) rather than by seepage water which is essentially rainfall. This leaves 1069 attacked stands for the ordinal and metric moisture level analysis.

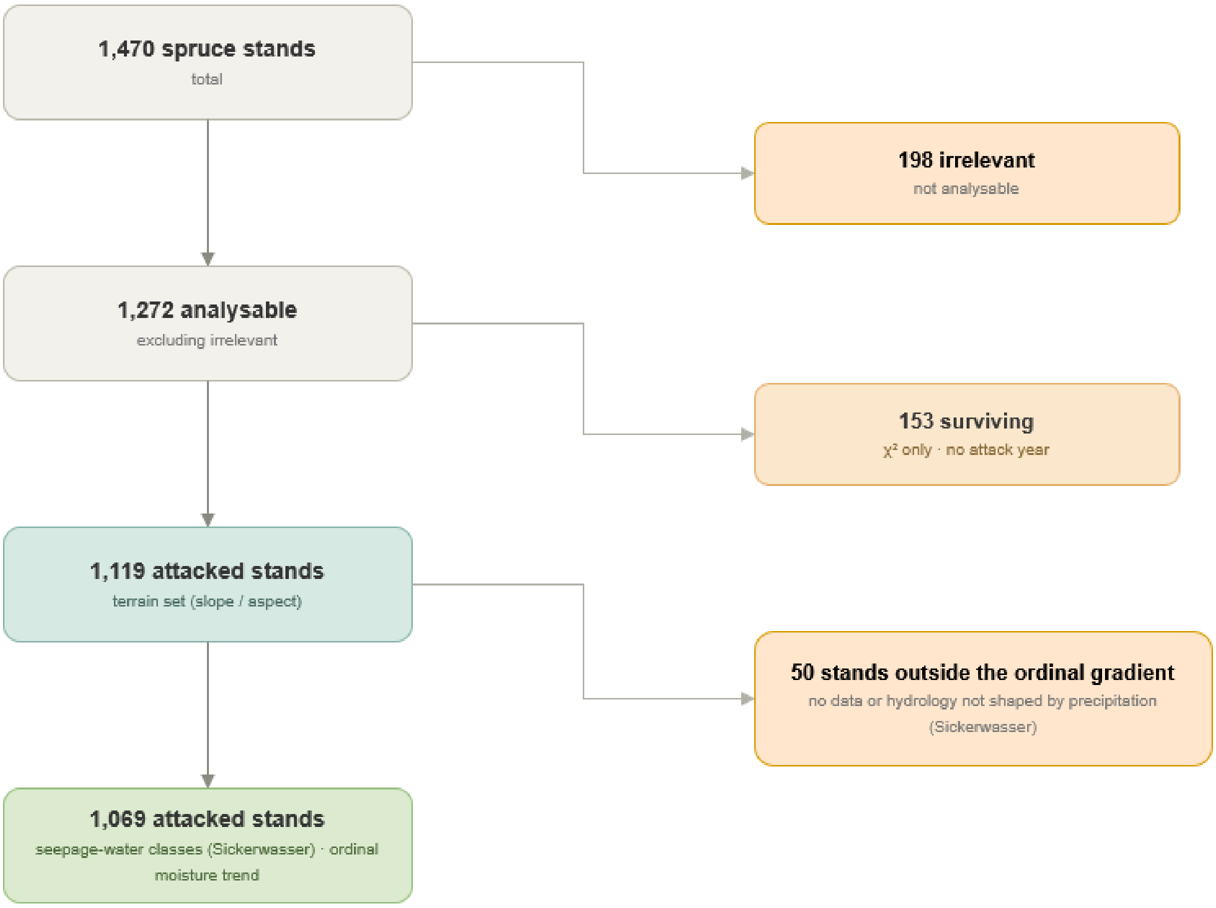

Figure 20 : Workflow of deriving the number of stands for the statistical analysis. Using moisture classes which removed 50 stands from the subset used for DEM based predictor variables.

2.5.2 Spatial level of analysis (Pixel vs stand) & nominal association

The first question is whether the timing of attack is associated with each site factor (slope class, aspect class, moisture level class) if these parameters are used as categorical variables. This nominal association is tested with the Pearson chi-square test (Chi^2). The test is build on the basis of a contingency table of attack year vs each predictor factor.

The strength of association is reported as Cramer's V, which ranges from 0 to 1. It is independent from sample size n, so pixel and stand scales can be compared. If the pixel and stand values are similar, little information is lost in the aggregation, while a large difference would mean that the aggregation to stands had distorted associations in some way. Standard residuals show which combination of time of attack and predictor variable occurs more or less often than expected and thus can give a hint of the association.

The Chi^2 test assumes independent observations and expected cell counts that are not too small. A general rule is of cells of the contingency table should have an expected count of 5 or more (McHugh, 2013). Independence is true on the stand scale, but not on the pixel level due to autocorrelation, which is why effect size is more important than the p values. The count rule isn't satisfied in several cells of the contingency table due to the few stands attacked in 2024 and 2025 and very few dry stands. This makes the Chi^2 distribution table unreliable. The values are instead obtained by Monte Carlo simulations, which builds the null distribution by repeatedly generating random tables that keep the same row and column totals as the observed table and only varies the counts inside it. Here I use n=20.000 tables. The approach follows the outline of the r documentation for chisq.test (*R: Pearson's Chi-squared Test for Count Data*, no date) and the paper in this documentation (Patefield, 1981). The same procedure is applied to all stand-level tables, so that the method is uniform. Where cells are not sparse the simulated and asymptotic p values agree closely.

The moisture factor is tested twice, once across all available moisture levels and once to the ordered seepage water gradient (Sickerwasser).

2.5.3 Ordinal tests and Regression

The nominal tests detect an association, but not the direction. The central hypothesis, that drier stands are attacked earlier, is a monotone trend and is tested with ordinal rank correlation. This is possible because both attack year (2019 < 2021 < 2023 < 2024 < 2025) and the moisture gradient (mäßig trocken < mäßig frisch < frisch < sehr frisch) can be ranked.

Rank correlation Kendall tau-b & Spearman rho:

Kendall's tau-b and Spearman's rho use ranks of the values rather than actual number values, so each variable first must be brought into an ordered sequence. Two steps are required for the terrain variables (n = 1119 stands). Slope is kept as its continuous median rather than its slope class, so the rank test can be applied directly on the median value. Aspect is a circular variable that cannot be rank ordered. It is split up into two linear components, northness and eastness (see 2.5.1), each of which is ranked independently. For the moisture test the sample is then restricted to stands on the ordered seepage water gradient (n = 1069 stands).

Ordinal trend was tested using Kendall's tau-b (τ-b) and spearman's rho (ϱ). Kendall's tau-b is the primary statistic, since by design better works with the large ties, which result from a rather coarse ordinal scale. It splits all pairs of stands into concordant pairs, where the drier stand was attacked earlier, discordant pairs, where the order is reversed. Tied pairs, which share the same moisture class or the same attack year are very pronounced due to the distribution of the variables. Kendall's tau-b is the difference between the number of concordant and discordant pairs and corrects for the large share of ties in the denominator. Figure 21 shows the many ties on the example of the 1069 stands on the moisture gradient

Figure 21: The share of concordant, discordant and tied pairs, shown here by the example of moisture level vs time of attack. The large share of tied pairs is better handled by Kendall than spearman

Spearman's rho (ϱ) is reported alongside it. It is a Pearson correlation based on the ranks of the two variables. It also measures a monotone association but through the ranked values rather than pairwise comparisons. Both are reported to check if they agree in direction, but Kendall's tau-b remains the primary statistic, since Spearman hasn't got a similar correction for ties. P value is calculated for both tests. The surviving stands were kept in the nominal chi-square as an ordinary category, because that test only needs class membership and treats "not attacked" as a valid class. They are left out of the primary ordinal test, because rank correlation needs a position on the ordered time axis, and a stand that was never attacked does not appear on this gradient. For this reason, the surviving stands are only used as a robustness check.

Numeric test and sensitivity of the time scale

At the metric level of analysis, the attack years are used as real calendar values rather than as equally spaced ranks. This serves as a sensitivity check on the ordinal time scale, because the calendar gaps are uneven. For every predictor, a Pearson correlation with the numeric attack year is computed.

For slope and aspect this is unproblematic, as slope is a real measurement in degrees and aspect has its two numeric components northness and eastness.

Moisture needs a special rule, because the classes are ordinal and are only given the ranks one to four, so both the Pearson correlation and the fitted slope treat these steps as if they were equal. Treating an ordinal site indicator as if it were numeric is common in vegetation ecology, where Ellenberg indicator values are routinely used as numbers, a practice that is useful but debated exactly because of this equal spacing assumption (Diekmann, 2003). For this reason the whole metric test should be viewed as supporting the ordinal analysis and not replacing it. On a technical level, this can be simply done by treating the rank values as real numbers (full codes in Appendix):

```
moisture_rank = as.integer(moisture_Sickerwasser)
```

The assumption is useful, however, because it is what allows a slope to be fitted by linear regression. This slope is interesting as it could serve as practical guidance for forest practitioners, if the effect were strong enough. If a clear and significant regression could be shown, it could be directly translated from the slope a practical rule of thumb. A hypothetical example would be each moisture level delays bark beetle infestation by about three weeks.

Both the linear regression and the Pearson correlation return an r value, a p value and an R^2 and are available for every predictor variable. The slope in years per step is fitted only on moisture level.

2.5.4. Reproducibility of Data

Monte Carlo Simulation was used in Chi^2 test for stand level to simulate weak classes. A fixed seed was used to make the p value deterministic. The fixed seed is 42. The script can be found in 7.4, along with all other data sources.

3. Results

3.1 Descriptive statistics and distribution of variables

Before further analysis, the basic distributions of the response and predictor variables are presented, which are the result of the manual classification and the data preparation steps. Some distribution on the pixel level were already displayed in the data description in 2.4, but I will focus on the analytical level of the stand here.

Distribution of Response Variable (Time of Attack)

The time of attack is unevenly spaced and mostly contained in 2021 and 2023 class. The shape resembles the exponential curve of the bark beetle infestation in NRW. Most stands were first attacked in 2021 (564 stands, 38 %), with fewer in 2023 (326, 22 %) and 2019 (171, 12 %), and only a handful in 2024 and 2025 (44 and 14 stands; 3 % and 1 %). The sharp drop in the most recent years leaves those classes thinly populated, which is accounted for in the χ^2 analysis.

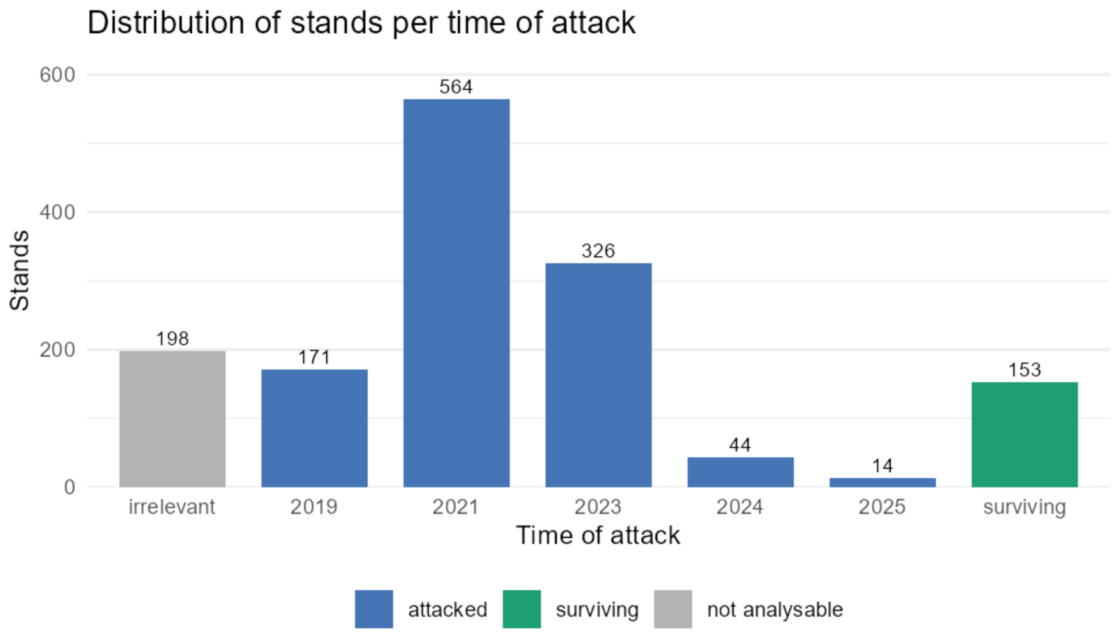

Figure 22: Distribution of the response Variable

Attack timing and irrelevant stands are relatively evenly spaced across the study area, but surviving stands seem to be more common in the northern part of Wermelskirchen. For a visual overview before the results, the reference layer and response variable time of attack is shown in figure 23.

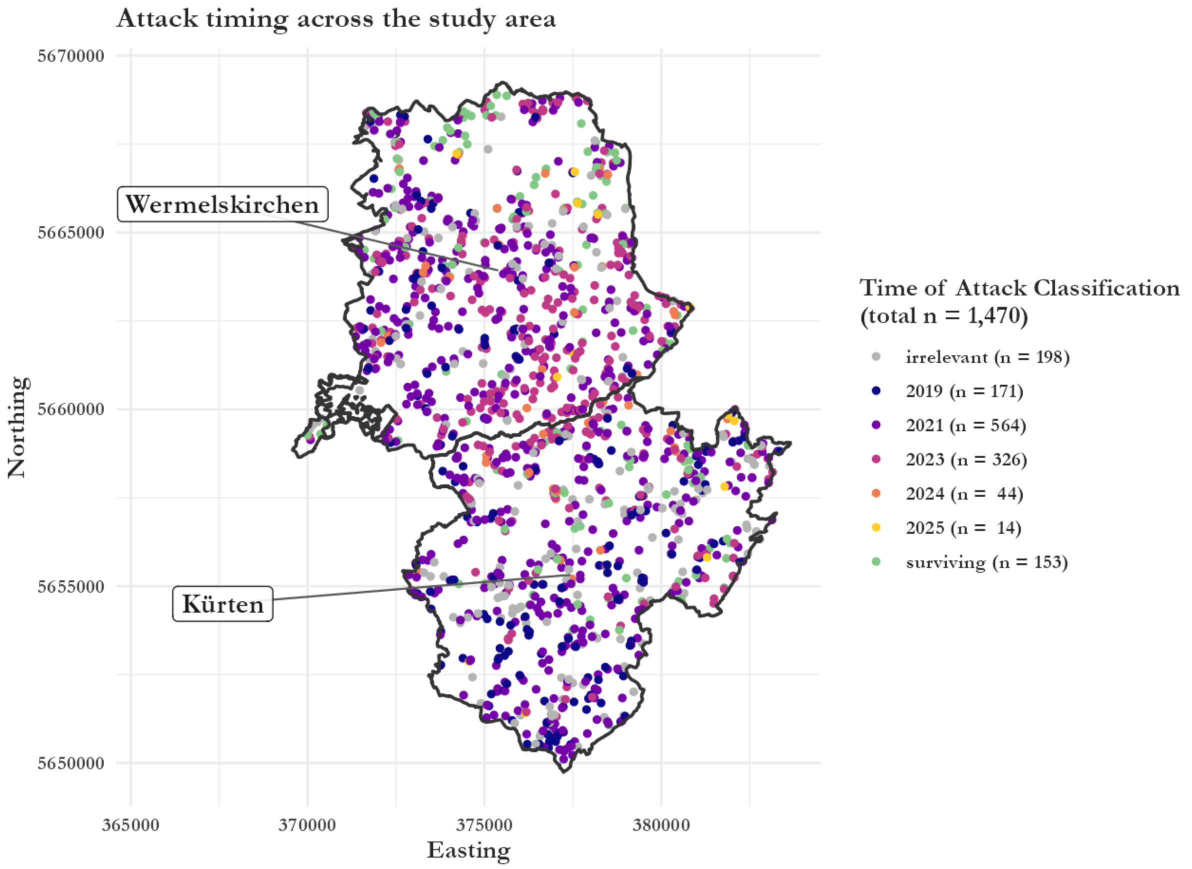

Figure 23. Map of the reference / response variable Time of Attack in the Study area. Each point is centroid of a forest stand.

Forest Stand slope is moderate, with a median of 13° and a range 1-32 degrees, (Figure 24). Most stands fall in the 5-20° classes (5-10°: 24 %, 10-15°: 35 %, 15-20°: 26 %); gentle terrain (0-5°: 63 stands) and very steep slopes (> 30°: 2 stands) are rare.

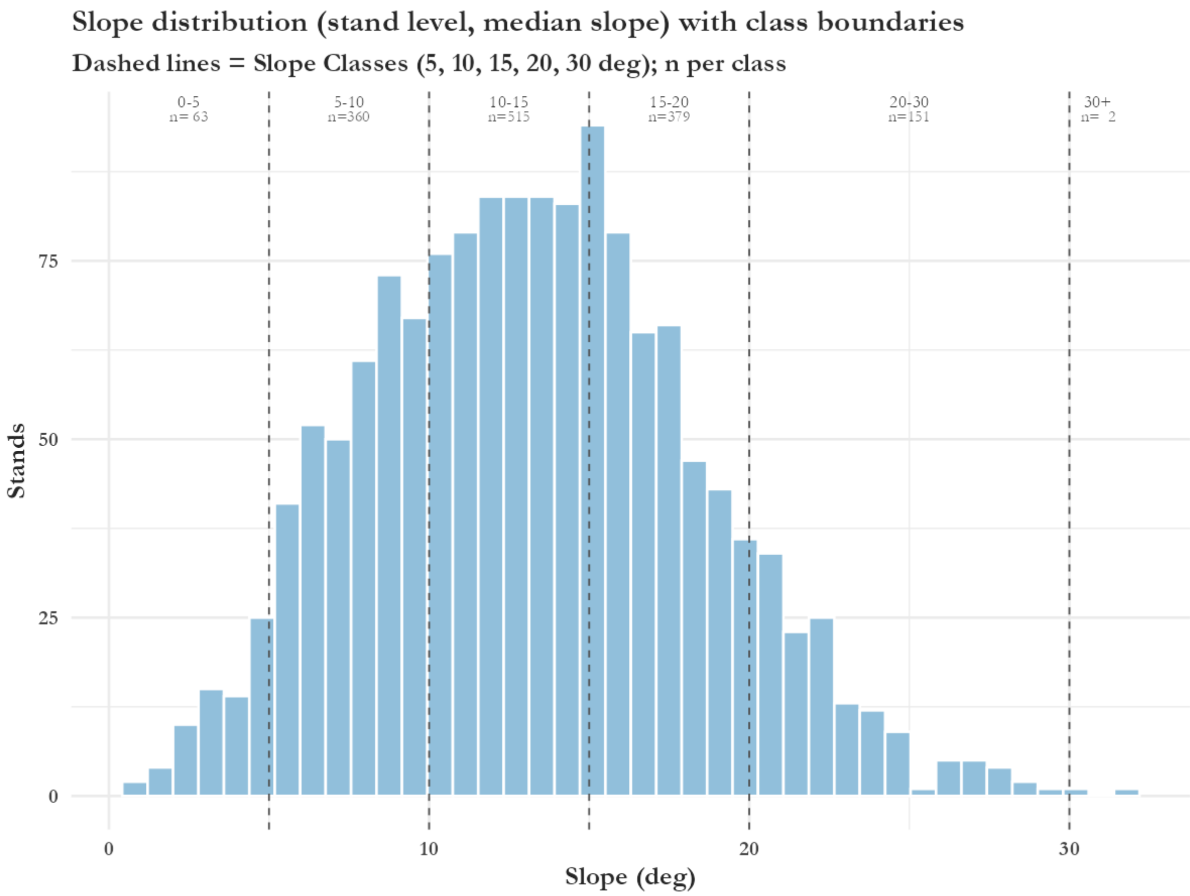

Figure 24: Distribution of Slope Classes within the study area

Aspect is spread fairly evenly across all eight directions with a mild concentration toward the north-west (NW 18 %, SW 14 %) and the fewest stands facing north-east (9 %). The mean directional components confirm no strong bias (northness 0.01, eastness - 0.12).

The high mean within-stand resultant length (0.87) shows that pixels within a stand generally share one aspect, so the directional means are reliable and not much information is lost from aggregating pixels to stand level.

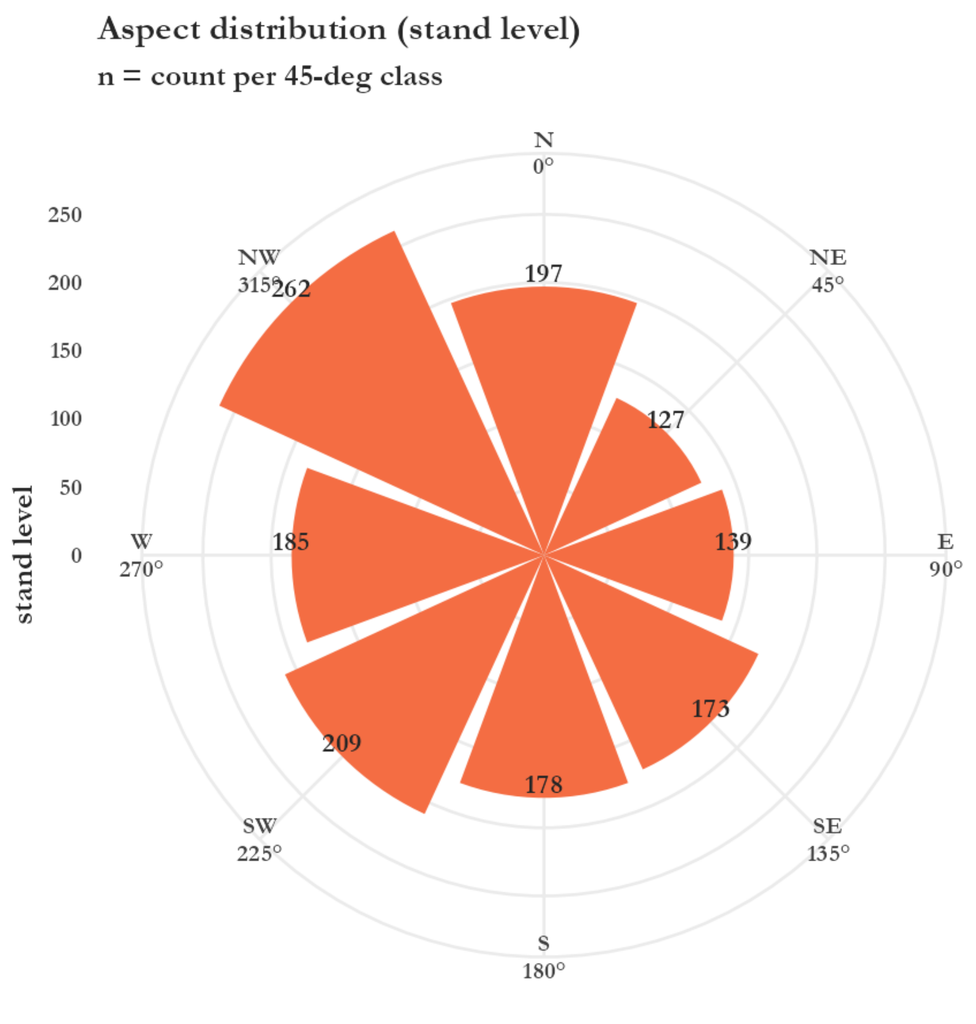

Figure 25:Distribution of Aspect classes within the study area

Almost all stands belong to the seepage-water (Sickerwasser) group (94 %), and within it the fresh end dominates with "sehr frisch" (813 stands, 55 %) and "frisch" (535, 36 %) while the drier classes are nearly absent (mäßig frisch 32, mäßig trocken 1). The remaining stands fall in the wetter Stau- and Grundwasser groups (feucht, nass, mäßig wechselfeucht; all together < 3 %). The moisture gradient therefore has a limited range, concentrated at the wetter end. For this reason, not too many stands are excluded when only looking "Sickerwasser" Classes in the ordinal analysis.

Figure 26 shows the absolute number of stands for each moisture level class, split into the hydrological groups. Figure 27however shows only the moisture levels within the ordinal gradient and shows the relative share of moisture levels for each year. This figure again illustrates the shift of drier stands being affected earlier.

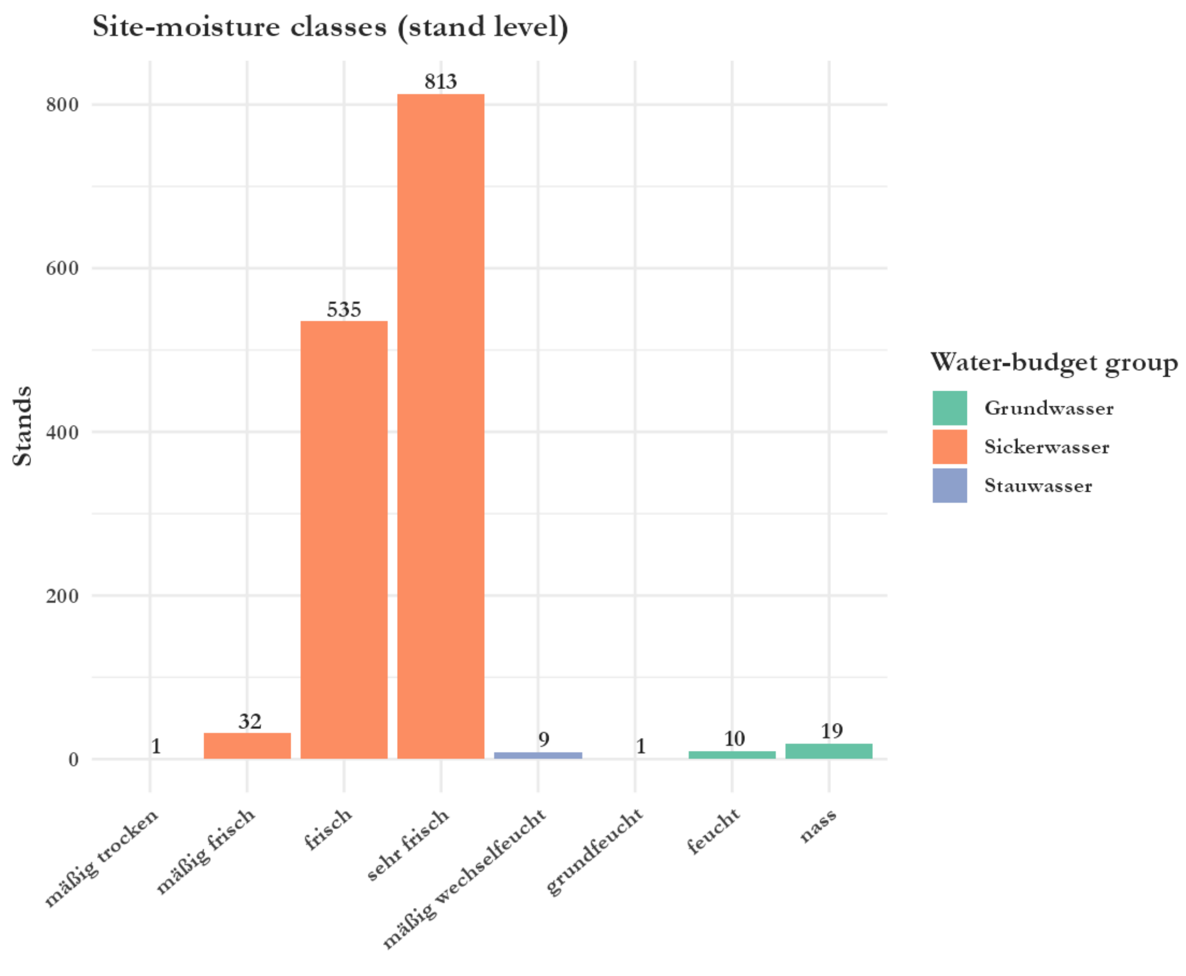

Figure 26: Distribution of the moisture classes. The colors indicate the hydrological dynamic of their water budget class.

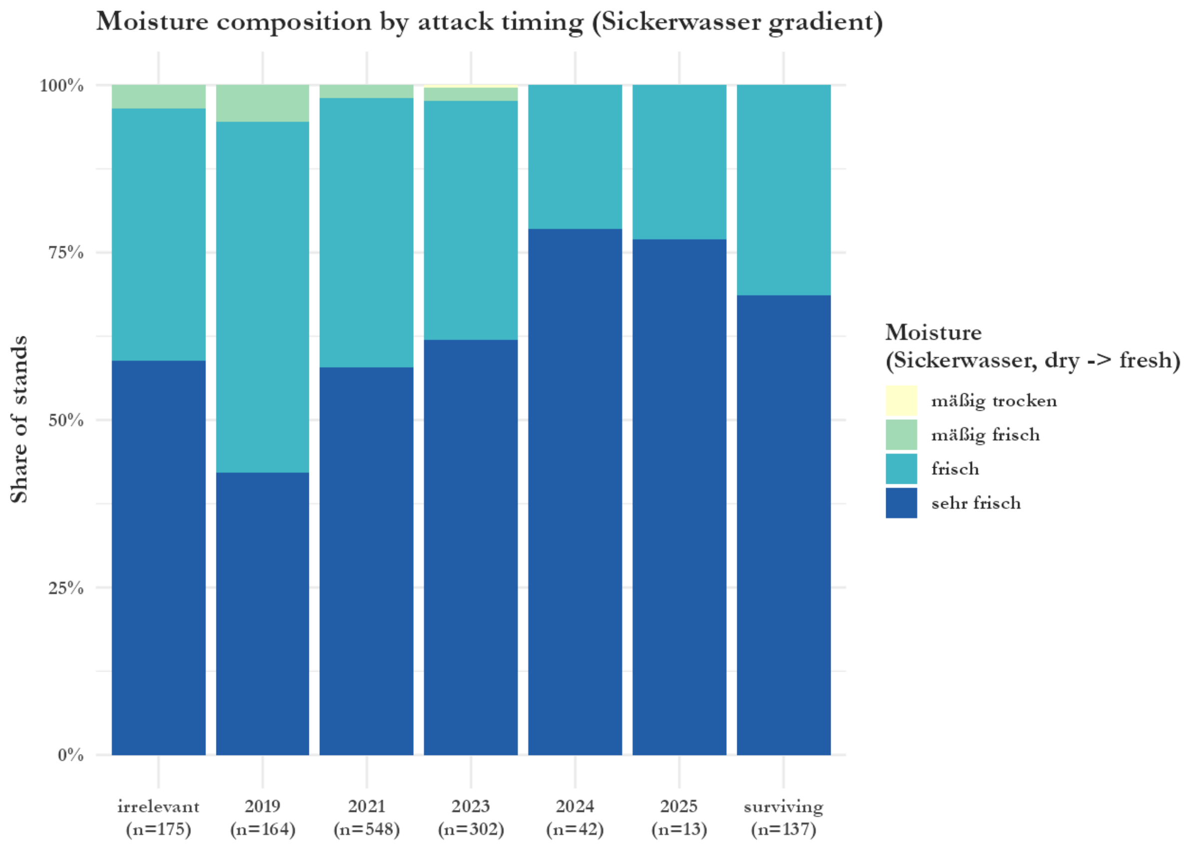

Figure 27: Distribution of moisture levels, on the basis of their hydrological group. Here, the focus is on the relative share of the moisture classes, not the total amounts.

3.2 Chi Square

At the pixel level all predictor variables have very small Cramer's V (slope 0.039, aspect 0.060, moisture 0.065). At the stand level, slope and aspect also show no association (Cramer's V 0.055 and 0.082), while moisture level shows a weak but comparatively stronger one (Cramer's V 0.099 across all classes; 0.111on the ordered seepage-water gradient). The exact n differs per factor and scale, as moisture excludes unmapped groundwater influenced stands, and are listed in Table 6. The effect sizes are almost identical across the two scales. For this reason, aggregating pixels to stands does not inflate or remove the association and analysis on stand level is valid.

Table 6: Comparison of Cramer V on Pixel and Stand level

Spatial Scale		Slope class	Aspect class	Moisture (all classes)	Moisture (Ordinal Sickerwasser gradient)
Pixel Level	Pixel n	143617	143617	134074	128214
	Pixel Cramer V	0.039	0.060	0.065	0.061
Stand Level	Stand n	1272	1272	1239	1206
	Stand Cramer V	0.055	0.082	0.099	**0.111**

The stand level thus is the analytical unit. Table 7 reports the full Chi Square test of attack timing against each site variable, so that the validity of the test can be judged by this table. Next to the effect size Cramer's V it shows the reported p value, the number of stands, the chi square statistic and the degrees of freedom. The share of cells in the contingency table with less than 5 counts are shown, as these sparsely populated table calls make the asymptotic chi-square distribution unreliable. The p values are calculated by a Monte Carlo Simulation instead. One can see the in the table that the higher the share of Cells E<5 is, the more the two p values differ. For this reason simulated value is used.

Table 7: Full results of Chi2

Predictor Variable	n	χ^2	df	Cells E<5	p (asymptotic)	p (Monte-Carlo)	Cramér's V
Slope class	1272	19.4	25	33 %	0.78	0.72	0.055
Aspect class	1272	42.9	35	21 %	0.17	0.16	0.082
Moisture (all classes)	1239	60.4	35	71 %	0.005	0.051	0.099
Moisture (Sickerwasser)	1206	44.6	15	42 %	< 0.001	0.013	0.111

The main result is the effect size and illustrated in Figure 28. Slope and aspect do not show any association with attack timing, their Cramer's V is small (0.055 and 0.082) and their Monte Carlo p values are not significant (0.72 and 0.16). Moisture is the only site parameter that carries some signal. Across all classes it is stronger than the terrain variables but is at the edge of significance (Cramer's V 0.099, p = 0.051). Restricting moisture levels of the ordered seepage-water gradient makes it clearly significant (Cramer's 0.111, p = 0.013). The nominal test therefore locates the only association in the moisture variable. The chi-square test treats the classes as unordered. It can show that an association exists but not in which direction, which the rank correlation addresses.

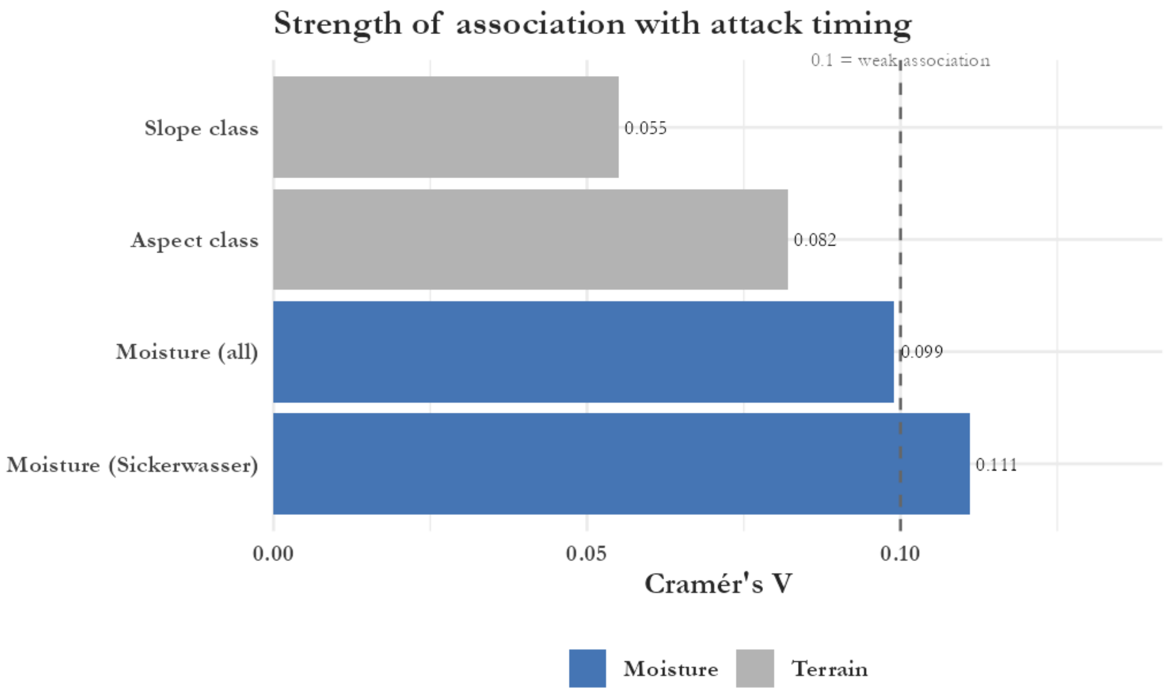

Figure 28: Cramer's V of stand level Chi2 Test. Only Moisture on the ordinal Sickerwasser gradient shows a weak association above 0,1.

However, a Standardized residuals heatmap can illustrate where this association is coming from (Figure 29).

The heatmap shows the standardized (adjusted) residuals of the stand-level contingency table of attack timing (rows) against the four occupied Sickerwasser moisture classes (columns). A standardized residual is the difference between expected and observed counts in a cell, scaled by their standard error. The Residuals show two things. First the direction of the deviation is indicated in colors. Blue shows fewer counts than expected, red means more counts. Secondly, the actual number shows the magnitude of the deviation. More than 2 would be relevant, more than 3 a strong deviation. Stand count n is below each residual.

When looking at the table, drier classes are overrepresented in early years and underrepresented in later years and surviving stands. The most important line here is that for 2019, "Frisch" is strongly overrepresented (3.8) and the wettest class "sehr Frisch" strongly underrepresented (-4.7). The inverse is true for wetter stands, where "sehr Frisch" becomes overrepresented in later years (2.6 in 2024, 2.5 for survivors). Cells with a dashed border have an expected count below five. Their standardized residual rests on very few stands and can change strongly with a single case. The table should thus be read with caution, as the pattern in "Frisch" and "sehr Frisch" is clear but is based on just a few cells of the table.

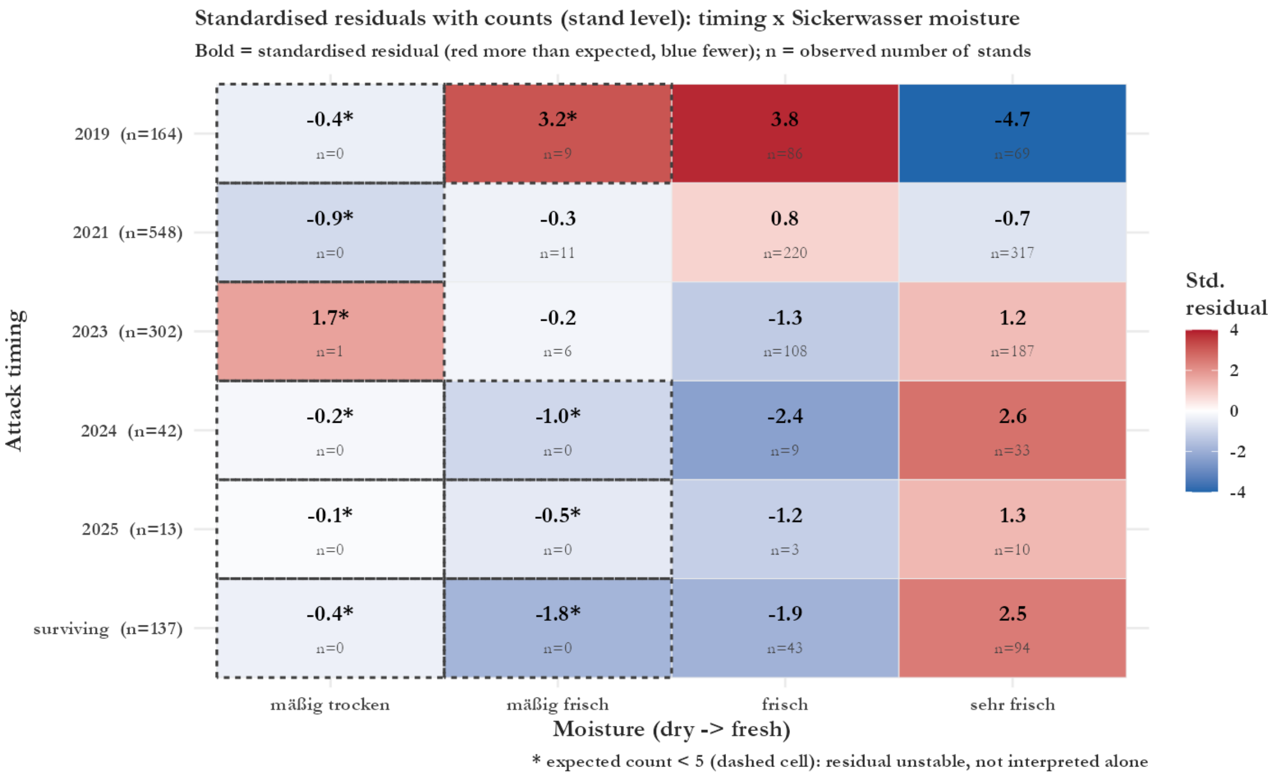

Figure 29: Residual Heatmap of stands of the Sickerwasser gradient, the only weak association of a predictor variable. Blue means fewer stands were in this class than expected, red means more stands were attacked. Focusing on the right two columns, we can see a switch of the trend from 2021 to 2023. This indicates that wetter stands could resist bark beetles in the beginning, but later were all affected nonetheless, when more predisposed stands were attacked already and thus show up blue from 2023 onwards.

3.3 Ordinal Rank correlation & Regression

The rank correlation tests the direction of the association that the chi-square test detected. Both Kendall's tau-b and Spearman's rho run from -1 to 1. A positive coefficient means wetter sites are attacked later. Tau-b is usually a little smaller than rho, so each test should be read on its own scale.

The results of the primary analysis built strictly on the ordinal scale is shown in Table 8. Moisture shows a weak but highly significant monotone trend. On the attacked seepage-water stands (Sickerwasser, n = 1,069) Kendall's tau-b is 0.136 and Spearman's rho 0.146, both with $p < 0.001$, so drier stands are attacked earlier and wetter stands later. Slope and aspect show no trend (n = 1,119): Kendall's tau-b is 0.042 for slope, 0.022 for northness and 0.019 for eastness, and none is significant. Slope, with $p = 0.065$, is the closest to significance among the terrain variables but does not reach it, and also its coefficient is negligible.

Table 8: Result of primary ordinal analysis

Predictor Variable	N (stands)	Kendall τ_b	Spearman ϱ	P (Kendall)
Moisture (Sickerwasser)	1,069	**0.136**	**0.146**	**< 0.001**
Slope (median)	1,119	0.042	0.055	0.065
Aspect northness	1,119	0.022	0.028	0.348
Aspect eastness	1,119	0.019	0.024	0.407

The supporting metric analysis supports this finding (Table 9). Treating the attack year as a real calendar value, the Pearson correlation with moisture is 0.149 ($p < 0.001$), almost identical to the rank coefficients of ordinal analysis. The linear model gives a slope of about 0.40 years per moisture step. Even so, moisture explains only about two percent of the variation in attack timing (R^2 = 2.2 %). For slope and aspect, the Pearson correlations are equally small (0.047, 0.029 and 0.032). With an R^2 as low as 0.2 percent, no linear regression is fitted for aspect and slope because there is no monotone trend to quantify.

In Summary, across the ordinal and the metric level the picture is the same. Moisture is the only variable with a directional signal, with drier earlier and wetter later, and it is highly significant, yet it is weak and accounts for only a small share of the variation. Slope and aspect show no trend on either level.

Table 9: Supporting analysis: Pearson correlation and linear regression for moisture level

Predictor Variable	N (stands)	Pearson r	Slope (lm)	R2
Moisture (Sickerwasser)	1,069	**0.149 *** **	0.40 yr/moisture level	**2.2 %**
Slope (median)	1,119	0.047	-	0.2 %
Aspect northness	1,119	0.029	-	0.08 %
Aspect eastness	1,119	0.032	-	0.10 %

*p < 0.05, ** p < 0.01, *** p < 0.001

3.4 Robustness checks of the statistical analysis

The trend does not depend on the choice of statistics or on the treatment of the surviving stands. On the attacked stands the three measurement levels agree closely: Kendall's tau-b (0.136), Spearman's rho (0.146) and the Pearson correlation on the numeric year (0.149) all have the same direction at almost the same strength, and all are highly significant ($p < 0.001$).

Adding the surviving stands at the late, never attacked end of the time axis was done to see if this trend continues if the classes are less restricted. Adding them does not weaken the trend but strengthens it slightly, from a tau-b of 0.136 to 0.143 and a rho of 0.146 to 0.157. This is consistent with the surviving stands being the wettest. The weak positive association between drier sites and earlier attacks is therefore robust to both the method and the subset of stands.

Table 10 : List of Results across Methods and Data subsets. They all converge to a similar coefficient size

Dataset	**n (stands)**	**Method**	**Coefficient**	**p**
Standard (all attacked Stands)	1069	Kendall τ-b	0.136	1.6×10^{-6}
	1069	Spearman ρ	0.146	1.5×10^{-6}
	1069	Pearson r (numeric year)	0.149	9.6×10^{-7}
Attacked stands + surviving stands	1206	Kendall τ-b	0.143	5.3×10^{-8}
	1206	Spearman ρ	0.157	4.4×10^{-8}

Because each predictor was tested individually, the site variables were checked for collinearity. The pairwise Spearman correlations among them are all low, the strongest being moisture with northness at 0.21 and all others below 0.1 The predictors are therefore effectively independent of one another, so testing them one at a time is justified and the moisture effect is not a hidden proxy for slope or aspect. This addresses confounding among the measured predictors only. However, unmeasured confounders are not excluded and are addressed in the discussion.

Table 11: Pairwise test of collinearity for all predictor variable combinations, ordered by strength

Pair	**Spearman rho**
Moisture - Northness	0.207
Moisture - Slope	0.082
Slope - Northness	0.077
Moisture - Eastness	0.031
Slope - Eastness	-0.009

4. Discussion

4.1 Principal Finding

The statistical analysis showed that Slope and Aspect had no association with the time of attack of bark beetle. For Moisture level however, drier sites were consistently attacked earlier and wetter sites later, showing a weak but robust trend (Kendall τ-b ≈ 0.14; Pearson r ≈ 0.15; R^2 ≈ 2 %)

These results were confirmed on three measurement levels: nominal (χ^2), ordinal (Kendall/Spearman) and metric (regression) and on two spatial scales (Pixel and Stand). On the stand level the sample size of 1069 individual stands was large enough to show the weak trend of moisture on the ordinal rank correlation test. Because the large sample size gave the analysis enough statistical power to detect even a weak trend, the missing effect of a slope or aspect should be viewed as robust negative result rather than a failure to detect an existing effect.

The study design operationalized vulnerability by the timing of attack. In this context, only the moisture level has a low signal of identifying predisposed forest stands.

4.2 Interpretation of results

4.2.1 Moisture level

Moisture showed a monotone trend, which is ecologically plausible. A lower moisture level in the FSK5 Mapping means that drought or warm temperatures will lead to faster water stress for spruce. Drier soils have been reported to lead to larger bark beetle outbreaks in the literature (Kärvemo *et al.*, 2023). In a study in the Vosges mountains, water-balance predictors accounted for about half of the explained model deviance for silver fir and Norway spruce dieback (Piedallu *et al.*, 2023). In the Alps, dry sites increased the probability of outbreaks (Candotti and Tomelleri, 2026). Most comparable to my results is a study from (Xu *et al.*, 2024), conducted in the Arnsberger Wald only about 100 km from the study area. It showed stands with relatively lower soil moisture were attacked earlier, and reduced soil moisture preceded infestation by one year later. Not all studies agree with this trend however. On a 1 km grid in France, Nardi et al. (Nardi *et al.*, 2023) found the opposite, with flatter and wetter sites more affected. They attribute this to the growth differentiation balance hypothesis, which means here that trees on chronically dry sites invest more in defense and are therefore more resistant. However, their coarse grid averages across many stands and explains only about 23 percent of the variation, so their findings cannot be transferred to the stand level tests in my study area.

Despite all this mostly affirming research, the measured effect is only weak. One potential reason could be that the moisture gradient in our study areas is very narrow. Firstly, moisture classes which were not seepage water dominated were excluded to allow for ordinal analysis. However, these classes are very rare anyway as they typically are special sites in wetlands or directly next to river streams. The study area is dominated by "Frisch" and "sehr Frisch", drier classes are rare. The soil and its hydrological properties in Bergisches Land is simply very homogenous. With such a small gradient, an effect might be hard to capture.

A second reason is that the moisture level is a static product of a classification created by the Geologische Dienst NRW. While it is accurately mapped, it only predicts the baseline and does not take the actual, dynamic weather into account. Piedallu tested 24 predictors for water availability, and showed that dynamic water predictors outperform static ones (Piedallu *et al.*, 2023). From this standpoint it seems plausible that the forest mapping captures the water signal, but only a part of the actual effect size.

4.2.2 Slope

Slope was included as a predictor because greater runoff and higher solar exposure could plausibly raise water stress, and testing this common assumption was itself an aim of the study. No independent slope effect was found. A likely reason is the low relief and limited elevation range of the study area, which leaves little slope variation to act on. The wider literature is itself inconsistent on the effect direction of slope: steeper stands were more affected in Sweden (Huo, Persson and Lindberg, 2024), whereas flatter and wetter sites were more affected in France (Nardi *et al.*, 2023). Fernández-Carrillo et al. (2024) found slope to be the single least important of twelve predictors in a large, partly mountainous machine learning model. Slope therefore appears to be a weak and context dependent predictor of bark beetle attack even where relief is substantial, and not useful in low relief terrain.

4.2.3 Aspect

Aspect was expected to show greater vulnerability in south or southwest facing stands, due to increased solar radiation and the resulting higher temperatures in the stand. No effect was observed in the study area. Consistent with this, aspect was weak and the effect direction was inconsistent even in the studies of Candotti and Tomelleri (Candotti and Tomelleri, 2026) and Fernández-Carrillo et al. (Fernández-Carrillo *et al.*, 2024), which both have more elevation and a more relief.

Conversely, Huo et al. (Huo, Persson and Lindberg, 2024) found south and west facing stands to be less affected. The authors attribute this to excessive radiation and bark overheating that dry out the phloem and hinder the bark beetle brood, but provide no further evidence for this interpretation.

The low elevation of the study area is a further reason, as it was for slope. Aspect only translates into meaningful differences in temperature and solar load where the terrain is varied enough to create real contrast between slopes. In the study area hillsides can be steep, but only for 100m elevation difference or so, so aspect has little variation with which to influence timing. This is supported by Mezei et al. (Mezei *et al.*, 2019), who showed that a physically modelled potential solar radiation surface was a strong predictor of infestation, and that the most sun exposed sites were attacked first. However, their study area was a mountainous landscape of 700 to 1700 m. Raw aspect captures only part of that solar signal, and in flat terrain even the full effect would be weak.

4.2.4 Effect of the population phase of Bark beetles

A further reason for the weak effect is the population phase of the outbreak. Site conditions matter most in the endemic phase, when beetles depend on already weakened trees. In the epidemic phase, mass attack and pheromone aggregation overwhelm even healthy, well supplied trees, so site differences lose importance (Wermelinger, 2004). Seidl et al. (Seidl *et al.*, 2016) show this directly: drought triggered infestations mainly when beetle pressure was low, but once pressure was high, trees were attacked regardless of water supply. While this effect was expected to some extent, the differentiation between endemic and outbreak phases was more important than assumed. In the study area, most stands were first attacked in 2021, at the peak of the calamity, when beetle pressure was extreme. A weak link between static site conditions and attack timing is therefore not surprising. At the peak, the outbreak dynamics decide which stand falls next, not site quality. The moisture effect might be clearer if only the early years were analyzed, as Nardi et al. (2023) did by restricting their study to the starting phase.

This also points to a worthwhile angle: separating site predisposition research from the modelling the spread of the outbreak itself. The first is primarily effective in endemic phases, while the latter describes the exponential phasers. Using the distance to earlier infestation could be a good variable for current outbreaks. It would show how much the attack timing is controlled by immediate beetle pressure rather than site conditions.

4.3 Strength and Limitations

Strengths

One strength is the manual classification of the response. Existing predisposition studies lack the high spatial detail because they require so much ground data. This classification was carried out by a former forester in the study area, so the orthophotos could be interpreted with firsthand field knowledge of the stands. This improved the reliability of the attack timing classification.

Analyzing the spruce damage on the stand level avoids pseudo replication pixel values due to spatial autocorrelation. The large samples size of over a thousand stands provide a robustness of the results, even though the study area was constrained to two forest administration districts. N was large enough to show with confidence that slope and aspect have no effect on attack timing.

A further strength lies in the analytical grid itself. The data analysis is built upon the Sentinel cells, which offer many possibilities of including additional spectral indices or any Sentinel product.

Limitations

This manual, knowledge-based classification is at the same time a limitation, because it does not scale. Extending the approach would require the same local expertise and manual effort, which is unlikely and hard to standardize.

Data wise, one limitation is the moisture predictor variable. As discussed, it's a static classification and doesn't capture the actual weather during the drought and bark beetle outbreak and is mostly constrained to the small gradient of effectively two moisture classes.

Of course, the response variable has also very coarse spatial resolution. Digital Orthophotos are not even captured yearly but in uneven gaps in different phases of the vegetative period. A detailed view of the actual lag effect of wetter stands being less susceptible might have been missed due to the low temporal resolution of the response variable. The classification records visible mortality rather than the moment of actual beetle colonization. This adds uncertainty to the exact timing, but this would only be relevant under stricter temporal monitoring, as it first symptoms roughly appear two weeks after the start of infestation. Stand age and structure were not included for lack of inventory data, even though the literature identifies them as important drivers.

Each predictor was tested individually, being vulnerable to confounding factors. The low explanatory power is consistent with this, since it shows that most of the variation in attack timing is driven by factors outside the model. The most important is stand age and structure, which the literature identifies as a leading driver (Fernández-Carrillo *et al.*, 2024) but which could not be included for lack of inventory data. A second candidate, once the first infestations had formed, is the distance to previously infested stands, since proximity to existing foci also drives the timing of attack. Elevation is the third, but minor because of the low topography of the study area. Because the moisture effect is already weak, even a moderate uncontrolled confounder would have a strong influence.

4.4 Outlook & Further research

This thesis could not provide a reliable risk model for practical use, but it does show several ways how future research could refine the approach used here. In the following I will outline several areas of improvement to overcome the limitations discussed above.

Improving the response Variable: Better temporal resolution

The low temporal resolution of my response variable is the clearest and weakness in my study design, as not even yearly DOPs are available. A finer and gap free temporal reference, ideally intra-annual drone imagery from April to October, would show the seasonal dynamics that yearly orthophotos miss. Drone flights would serve a dual purpose, combining active monitoring of bark beetles with creating an image archive for later analysis. Widespread Drone use would be too costly at this moment, but for a test approach the technical basis should be there. Large Forest owners such as state forest agencies could lead the way here as it has been already tested successfully in the Czech republic (Klouček *et al.*, 2024). One problem with this is that manual classification would require a lot of time and another source of validation would be required.

An additional source reference data could be forest harvester data. Modern harvester processor units record the timber volume they cut with a finely, officially calibrated sensors and carry a differential GPS, so for every stand they log where, when and how much timber was cut, which gives accurate and spatially explicit data (Kemmerer and Labelle, 2021). This data is already being used to split profits based on sold timber volume according to the share in whose land parcel it was cut. Data access and aggregation might be difficult, since the data protocol contain lots of sensitive information about private contracts between forest owners and the company providing the logging service, such as prices and property rights. But this is only the case in the context of Germany. Techniques around this are pioneered by large companies in Canada and Scandinavia, where timber companies can work on their own land. From a purely technical perspective this data source should be looked at.

Improving the predictor variables

Several steps could improve the predictor variables. Raw aspect could be replaced by a modelled potential solar radiation surface (Mezei *et al.*, 2019) and the static moisture class by dynamic, process-based water balance layers that also capture actual trends (Piedallu *et al.*, 2023). Mezei explicitly refers to the already existing ArcGIS tool for calculating solar radiation (*Area Solar Radiation (Spatial Analyst)—ArcGIS Pro | Documentation*, no date), which can quickly be integrated.

An important predictor not used in the study is stand age and structure. Stand inventory data on age and species mixture would allow such key confounders to be controlled. This inventory data, however, is not available from a homogenous data source. This is also due to the diverse and complicated forest ownership structure with hundreds of small forest owners in the study area. While many of them are organized in cooperatives, obtaining the inventory data in a digital form would be politically difficult, would not scale easily and not provide full coverage for all forest areas.

The analysis is built on the Sentinel 2 grid cells. Any spectral index or Sentinel product could be integrated directly into the study design as predictor and tested.

Improving the statistical analysis

Another approach is a rather conceptual one. Site predisposition at low population phases has a different dynamic compared to outbreak spread at a mass population phase of the beetle. One way to respect this is by including the distance to previously infested stands as a predictor variable, since proximity to active infestations is likely a strong predictor of infestation timing at the peak of an outbreak (Fernández-Carrillo *et al.*, 2024; Candotti and Tomelleri, 2026). Nardi avoided this by only looking at bark beetle damages in forest stands without mass outbreaks (Nardi *et al.*, 2023). However, this approach requires its own classification or assumptions to first decide in which phase beetle populations are.

Susceptibility to bark beetles is likely nonstationary, because it can differ based on site characteristics and microclimate. Geographically weighted ordinal regression could further test whether the moisture effect varies within the forest districts. While fundamentally interesting, this approach might be a lot of modelling effort for detecting small differences in a weak signal.

Extending the study to a larger area with a wider moisture range and greater elevation range would test whether the site effects strengthen where the gradients are broader.

Rather than testing these factors in isolation as here, they could be combined into a single stand level vulnerability product, joining a dynamic water layer, stand age and structure, and a spread variable based on distance to recent infestations. Piedallu et al. (2023) show that such maps can be built from freely available data and validated against dead wood harvest volumes, and offers a template for this next step.

4.5 Practical implication for local Forestry

Foresters should be careful with the common assumption that steep or south facing stands are hit first. This negative result is still useful, because it warns against relying on predictors or assumptions that have no empiric basis for the study area. Site moisture does carry a real signal in the expected direction, but it is weak, so the static mapping of the FSK5 is important, but should not be used as the basis for far reaching management decisions such as tree species selection.

5. Conclusion

In this thesis I tested if the three forest site characteristics slope, aspect and soil moisture can explain the order the time of infestation of a forest stand within a bark beetle outbreak. For slope and aspect this was not the case, and soil moisture only had a weak effect of drier stands being attacked earlier. Different statistical tests converged to the same result, regardless of measurement level.

The weak effect fits the setting and the timing of the outbreak. The Bergisches Land has a narrow moisture gradient, and although its terrain is hilly with many slopes, the generally low elevation keeps the climatic contrast between sites small. Most stands were first attacked in 2021 at the peak of the outbreak, when mass attack and pheromone aggregation can overwhelm even healthy trees and site quality may lose importance. The degree to which the effect of the population dynamics override stand predisposition was stronger than expected.

In this study I tried to connect remote sensing-based forest ecology research with practical forest management. For forest practitioners there are two takeaway messages: The common assumption that steep or south facing stands are attacked first is not supported here and should be treated with caution. Soil moisture does have a genuine effect, but on its own it is too weak to guide prioritization of spruce stands.

The approach used here is limited but offers considerable potential for improvement, with different angles being discussed in the thesis. The literature review in introduction illustrates the dynamic improvements in the field of mapping and modelling bark beetle outbreaks. The clearest path to build on my approach is to improve the reference data for timing the attack by increasing the temporal resolution. Drone flights could be done repeatedly within a single vegetative period and would allow to trace the outbreak dynamics within a year.

6. References

Abdullah, H. *et al.* (2019) 'Sentinel-2 accurately maps green-attack stage of European spruce bark beetle (Ips typographus, L.) compared with Landsat-8', *Remote Sensing in Ecology and Conservation*, 5(1), pp. 87–106. Available at: https://doi.org/10.1002/rse2.93.

Allen, C.D., Breshears, D.D. and McDowell, N.G. (2015) 'On underestimation of global vulnerability to tree mortality and forest die-off from hotter drought in the Anthropocene', *Ecosphere*, 6(8), p. art129. Available at: https://doi.org/10.1890/ES15-00203.1.

Amatulli, G. *et al.* (2018) 'A suite of global, cross-scale topographic variables for environmental and biodiversity modeling', *Scientific Data*, 5(1), p. 180040. Available at: https://doi.org/10.1038/sdata.2018.40.

Area Solar Radiation (Spatial Analyst)—ArcGIS Pro | Documentation (no date). Available at: https://pro.arcgis.com/en/pro-app/3.4/tool-reference/spatial-analyst/area-solar-radiation.htm (Accessed: 22 July 2026).

Baumartenklassifikation NRW - Open.NRW (2026). Available at: https://ckan.open.nrw.de/dataset/452e4e9e-a923-440e-8af4-64a05157bada (Accessed: 8 June 2026).

Bentz, B.J. *et al.* (2019) 'Ips typographus and Dendroctonus ponderosae Models Project Thermal Suitability for Intra- and Inter-Continental Establishment in a Changing Climate', *Frontiers in Forests and Global Change*, 2. Available at: https://doi.org/10.3389/ffgc.2019.00001.

Biedermann, P.H.W. *et al.* (2019) 'Bark Beetle Population Dynamics in the Anthropocene: Challenges and Solutions', *Trends in Ecology & Evolution*, 34(10), pp. 914–924. Available at: https://doi.org/10.1016/j.tree.2019.06.002.

Candotti, A. *et al.* (2026) 'Recent Advances in Remote Sensing for Early Detection, Risk Prediction, and Post-disturbance Assessment of Bark Beetle Attacks in Temperate and Boreal Forests', *Current Forestry Reports*, 12(1), p. 11. Available at: https://doi.org/10.1007/s40725-026-00272-0.

Candotti, A. and Tomelleri, E. (2026) 'Assessment of regional scale-based bark beetle disturbance predisposition in complex terrain with earth observations', *Forestry: An International Journal of Forest Research*, 99(2), p. cpaf043. Available at: https://doi.org/10.1093/forestry/cpaf043.

Dalponte, M. *et al.* (2022) 'Mapping a European Spruce Bark Beetle Outbreak Using Sentinel-2 Remote Sensing Data', *Remote Sensing*, 14(13), p. 3135. Available at: https://doi.org/10.3390/rs14133135.

Das, A.K. *et al.* (2025) 'The increasing role of drought as an inciting factor of bark beetle outbreaks can cause large-scale transformation of Central European forests', *Landscape Ecology*, 40(6), p. 108. Available at: https://doi.org/10.1007/s10980-025-02125-w.

Davídková, M. and Doležal, P. (2017) 'Sister broods in the spruce bark beetle, *Ips typographus* (L.)', *Forest Ecology and Management*, 405, pp. 13–21. Available at: https://doi.org/10.1016/j.foreco.2017.08.040.

Diekmann, M. (2003) 'Species indicator values as an important tool in applied plant ecology – a review', *Basic and Applied Ecology*, 4(6), pp. 493–506. Available at: https://doi.org/10.1078/1439-1791-00185.

Digitales Basis-Landschaftsmodell - Paketierung: gesamt NRW (2026). Available at: https://www.opengeodata.nrw.de/produkte/geobasis/lm/akt/basis-dlm/ (Accessed: 8 June 2026).

dwd.de, Klimadaten Mittelwerte (2026). Available at: https://www.dwd.de/DE/leistungen/klimadatendeutschland/mittelwerte/nieder_8110_fest_html.html;jsessionid=4003770C5705F3CE7D09368E8AA2AE5C.live11041?view=nasPublication&nn=16102 (Accessed: 15 July 2026).

Fernandez-Carrillo, A. *et al.* (2020) 'Monitoring Bark Beetle Forest Damage in Central Europe. A Remote Sensing Approach Validated with Field Data', *Remote Sensing*, 12(21), p. 3634. Available at: https://doi.org/10.3390/rs12213634.

Fernández-Carrillo, Á. *et al.* (2024) 'Predictive Model for Bark Beetle Outbreaks in European Forests', *Forests*, 15(7), p. 1114. Available at: https://doi.org/10.3390/f15071114.

Fettig, C.J. *et al.* (2022) 'Management tactics to reduce bark beetle impacts in North America and Europe under altered forest and climatic conditions', *Bark Beetle Management, Ecology, and Climate Change*. Elsevier, pp. 345–394. Available at: https://doi.org/10.1016/B978-0-12-822145-7.00006-4.

Francini, S. *et al.* (2022) 'An open science and open data approach for the statistically robust estimation of forest disturbance areas', *International Journal of Applied Earth Observation and Geoinformation*, 106, p. 102663. Available at: https://doi.org/10.1016/j.jag.2021.102663.

Hallas, T. *et al.* (2026) 'Assessing disturbance risks of the European spruce bark beetle under climate change: a review of model approaches and future perspectives', *European Journal of Forest Research*, 145(2), p. 23. Available at: https://doi.org/10.1007/s10342-025-01841-x.

Hlásny, T. *et al.* (2021) 'Bark Beetle Outbreaks in Europe: State of Knowledge and Ways Forward for Management', *Current Forestry Reports*, 7(3), pp. 138–165. Available at: https://doi.org/10.1007/s40725-021-00142-x.

Holzeinschlag und Schadholz | Statistik.NRW (2026). Available at: https://statistik.nrw/wirtschaft-und-umwelt/land-und-forstwirtschaft/forstwirtschaft/holzeinschlag-und-schadholz (Accessed: 24 June 2026).

Holzwarth, S. *et al.* (2020) 'Earth Observation Based Monitoring of Forests in Germany: A Review', *Remote Sensing*, 12(21), p. 3570. Available at: https://doi.org/10.3390/rs12213570.

Holzwarth, S. *et al.* (2023) 'Earth-Observation-Based Monitoring of Forests in Germany—Recent Progress and Research Frontiers: A Review', *Remote Sensing*, 15(17), p. 4234. Available at: https://doi.org/10.3390/rs15174234.

Huber, C. *et al.* (2023) 'Potential alternatives for Norway spruce wood: a selection based on defect-free wood properties', *Annals of Forest Science*, 80(1), p. 41. Available at: https://doi.org/10.1186/s13595-023-01206-7.

Huo, L., Persson, H. and Lindberg, E. (2024) 'Analyzing the environmental risk factors of European spruce bark beetle damage at the local scale', *European Journal of Forest Research*, pp. 1–16. Available at: https://doi.org/10.1007/s10342-024-01662-4.

Huo, L., Persson, H.J. and Lindberg, E. (2021) 'Early detection of forest stress from European spruce bark beetle attack, and a new vegetation index: Normalized distance red & SWIR (NDRS)', *Remote Sensing of Environment*, 255, p. 112240. Available at: https://doi.org/10.1016/j.rse.2020.112240.

Jönsson, A.M. *et al.* (2009) 'Spatio-temporal impact of climate change on the activity and voltinism of the spruce bark beetle, Ips typographus', *Global Change Biology*, 15(2), pp. 486–499. Available at: https://doi.org/10.1111/j.1365-2486.2008.01742.x.

Kärvemo, S. *et al.* (2023) 'Different triggers, different stories: Bark-beetle infestation patterns after storm and drought-induced outbreaks', *Forest Ecology and Management*, 545, p. 121255. Available at: https://doi.org/10.1016/j.foreco.2023.121255.

Kautz, M. *et al.* (2017) 'Biotic disturbances in Northern Hemisphere forests – a synthesis of recent data, uncertainties and implications for forest monitoring and modelling', *Global Ecology and Biogeography*, 26(5), pp. 533–552. Available at: https://doi.org/10.1111/geb.12558.

Kautz, M., Feurer, J. and Adler, P. (2024) 'Early detection of bark beetle (*Ips typographus*) infestations by remote sensing – A critical review of recent research', *Forest Ecology and Management*, 556, p. 121595. Available at: https://doi.org/10.1016/j.foreco.2023.121595.

Kemmerer, J. and Labelle, E.R. (2021) 'Using harvester data from on-board computers: a review of key findings, opportunities and challenges', *European Journal of Forest Research*, 140(1), pp. 1–17. Available at: https://doi.org/10.1007/s10342-020-01313-4.

Klouček, T. *et al.* (2024) 'Early detection of bark beetle infestation using UAV-borne multispectral imagery: a case study on the spruce forest in the Czech Republic', *Frontiers in Forests and Global Change*, 7. Available at: https://doi.org/10.3389/ffgc.2024.1215734.

König, S. *et al.* (2023) 'Assessing Combinations of Landsat, Sentinel-2 and Sentinel-1 Time series for Detecting Bark Beetle Infestations', *GIScience & Remote Sensing*, 60(1), p. 2226515. Available at: https://doi.org/10.1080/15481603.2023.2226515.

Kortmann, M. *et al.* (2022) 'Disturbance Severity and Human–Nature Relationships: A New Approach to Analyze People's Well-Being along a Bark Beetle Infestation Gradient', *Forests*, 13(11), p. 1954. Available at: https://doi.org/10.3390/f13111954.

Landbedeckung NRW (2026). Available at: https://www.opengeodata.nrw.de/produkte/geobasis/lusat/akt/lb_nrw/ (Accessed: 8 June 2026).

Lausch, A., Fahse, L. and Heurich, M. (2011) 'Factors affecting the spatio-temporal dispersion of *Ips typographus* (L.) in Bavarian Forest National Park: A long-term quantitative landscape-level analysis', *Forest Ecology and Management*, 261(2), pp. 233–245. Available at: https://doi.org/10.1016/j.foreco.2010.10.012.

Marini, L. *et al.* (2017) 'Climate drivers of bark beetle outbreak dynamics in Norway spruce forests', *Ecography. 40: 1426-1435.*, 40, pp. 1426–1435. Available at: https://doi.org/10.1111/ecog.02769.

Marvasti-Zadeh, S.M. *et al.* (2023) 'Early Detection of Bark Beetle Attack Using Remote Sensing and Machine Learning: A Review', *ACM Computing Surveys*, 56(4), p. 97:1-97:40. Available at: https://doi.org/10.1145/3625387.

Massive Schäden - Einsatz für Wälder und Waldumbau nötig (2025) *BMEL*. Available at: https://www.bmleh.de/DE/themen/wald/wald-in-deutschland/wald-trockenheit-klimawandel.html (Accessed: 27 June 2026).

McHugh, M.L. (2013) 'The Chi-square test of independence', *Biochemia Medica*, pp. 143–149. Available at: https://doi.org/10.11613/BM.2013.018.

Mezei, P. *et al.* (2019) 'Potential Solar Radiation as a Driver for Bark Beetle Infestation on a Landscape Scale', *Forests*, 10(7), p. 604. Available at: https://doi.org/10.3390/f10070604.

Möhring, B. *et al.* (2021) 'Schadenssumme insgesamt 12,7 Mrd. Euro : Abschätzung der ökonomischen Schäden der Extremwetterereignisse der Jahre 2018 bis 2020 in der Forstwirtschaft', 147(9), pp. 155–158.

Nardi, D. *et al.* (2023) 'Drought and stand susceptibility to attacks by the European spruce bark beetle: A remote sensing approach', *Agricultural and Forest Entomology*, 25(1), pp. 119–129. Available at: https://doi.org/10.1111/afe.12536.

Netherer, S. *et al.* (2021) 'Interactions among Norway spruce, the bark beetle Ips typographus and its fungal symbionts in times of drought', *Journal of Pest Science*, 94(3), pp. 591–614. Available at: https://doi.org/10.1007/s10340-021-01341-y.

Økland, B. *et al.* (2016) 'Transition from windfall- to patch-driven outbreak dynamics of the spruce bark beetle *Ips typographus*', *Forest Ecology and Management*, 363, pp. 63–73. Available at: https://doi.org/10.1016/j.foreco.2015.12.007.

Patacca, M. *et al.* (2023) 'Significant increase in natural disturbance impacts on European forests since 1950', *Global Change Biology*, 29(5), pp. 1359–1376. Available at: https://doi.org/10.1111/gcb.16531.

Patefield, W.M. (1981) 'An Efficient Method of Generating Random R × C Tables with Given Row and Column Totals', *Journal of the Royal Statistical Society Series C*, 30(1), pp. 91–97.

Piedallu, C. *et al.* (2023) 'Spatial vulnerability assessment of silver fir and Norway spruce dieback driven by climate warming', *Landscape Ecology*, 38(2), pp. 341–361. Available at: https://doi.org/10.1007/s10980-022-01570-1.

QGIS Web (2026) *Luftbildfinder NRW*. Available at: https://plugins.qgis.org/plugins/luftbildfinder-nrw/ (Accessed: 16 July 2026).

R: Pearson's Chi-squared Test for Count Data (no date). Available at: https://web.mit.edu/r/current/lib/R/library/stats/html/chisq.test.html (Accessed: 18 July 2026).

Rammer, W. and Seidl, R. (2019) 'Harnessing Deep Learning in Ecology: An Example Predicting Bark Beetle Outbreaks', *Frontiers in Plant Science*, 10. Available at: https://doi.org/10.3389/fpls.2019.01327.

Reinosch, E. *et al.* (2025) 'Detailed validation of large-scale Sentinel-2-based forest disturbance maps across Germany', *Forestry: An International Journal of Forest Research*. Edited by T. Kattenborn, 98(3), pp. 437–453. Available at: https://doi.org/10.1093/forestry/cpae038.

Schmidt, S.I. *et al.* (2024) 'The Significance of Tree Height as a Predictor of Tree Mortality during Bark Beetle Outbreaks in a Small Catchment', *Forests*, 15(5), p. 803. Available at: https://doi.org/10.3390/f15050803.

Schmidt, U.E. (2017) 'Die Fichte in der Wald- und Forstgeschichte – eine soziokulturelle Betrachtung', *Lwf_wissen_80* [Preprint].

Schmied, G. *et al.* (2022) 'The Past Matters: Previous Management Strategies Modulate Current Growth and Drought Responses of Norway Spruce (Picea abies H. Karst.)', *Forests*, 13(2), p. 243. Available at: https://doi.org/10.3390/f13020243.

Schroeder, M., Knape, J. and Kärvemo, S. (2025) 'Rise and fall of a spruce bark beetle outbreak – Importance of colonisation density and reproductive success', *Forest Ecology and Management*, 586, p. 122695. Available at: https://doi.org/10.1016/j.foreco.2025.122695.

Schulte-Kellinghaus, S. and Weller, A. (2023) 'Forstlich-standortkundliche Auswertung auf Basis der Bodenkarten 1 : 5 000 und 1 : 50 000 Update 2023 (methodisch ergänzt)'.

Seidl, R. *et al.* (2014) 'Increasing forest disturbances in Europe and their impact on carbon storage', *Nature climate change*, 4(9), pp. 806–810. Available at: https://doi.org/10.1038/nclimate2318.

Seidl, R. *et al.* (2016) 'Small beetle, large-scale drivers: how regional and landscape factors affect outbreaks of the European spruce bark beetle', *Journal of Applied Ecology*, 53(2), pp. 530–540. Available at: https://doi.org/10.1111/1365-2664.12540.

Sheykhmousa, M. *et al.* (2020) 'Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review', *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, pp. 6308–6325. Available at: https://doi.org/10.1109/JSTARS.2020.3026724.

Singh, V.V. *et al.* (2024) 'Understanding bark beetle outbreaks: exploring the impact of changing temperature regimes, droughts, forest structure, and prospects for future forest pest management', *Reviews in Environmental Science and Bio/Technology*, 23(2), pp. 257–290. Available at: https://doi.org/10.1007/s11157-024-09692-5.

Stereńczak, K. *et al.* (2020) 'Influence of selected habitat and stand factors on bark beetle *Ips typographus* (L.) outbreak in the Białowieża Forest', *Forest Ecology and Management*, 459, p. 117826. Available at: https://doi.org/10.1016/j.foreco.2019.117826.

Thonfeld, F. *et al.* (2022) 'A First Assessment of Canopy Cover Loss in Germany's Forests after the 2018–2020 Drought Years', *Remote Sensing*, 14(3), p. 562. Available at: https://doi.org/10.3390/rs14030562.

Thorn, S. *et al.* (2018) 'Impacts of salvage logging on biodiversity: A meta-analysis', *Journal of Applied Ecology*, 55(1), pp. 279–289. Available at: https://doi.org/10.1111/1365-2664.12945.

Toffin, E. *et al.* (2018) 'Colonization of weakened trees by mass-attacking bark beetles: no penalty for pioneers, scattered initial distributions and final regular patterns', *Royal Society Open Science*, 5(1), p. 170454. Available at: https://doi.org/10.1098/rsos.170454.

Waldzustandsbericht NRW 2024 (lang) (2024). Ministerium für Landwirtschaft und Verbraucherschutz des Landes Nordrhein-Westfalen. Available at: https://www.wald-und-holz.nrw.de/fileadmin/Wald_in_NRW/250102_waldzustandsbericht_nrw_2024_lang.pdf (Accessed: 27 June 2026).

Weed, A.S., Ayres, M.P. and Bentz, B.J. (2015) 'Chapter 4 - Population Dynamics of Bark Beetles', in F.E. Vega and R.W. Hofstetter (eds) *Bark Beetles*. San Diego: Academic Press, pp. 157–176. Available at: https://doi.org/10.1016/B978-0-12-417156-5.00004-6.

Wermelinger, B. (2004) 'Ecology and management of the spruce bark beetle *Ips typographus*—a review of recent research', *Forest Ecology and Management*, 202(1), pp. 67–82. Available at: https://doi.org/10.1016/j.foreco.2004.07.018.

'Wiederbewaldungskonzept NRW' (2024). Available at: https://www.mlv.nrw.de/wp-content/uploads/2024/11/wiederbewaldungskonzept_nrw.pdf (Accessed: 21 July 2026).

WMS DOP (2026). Available at: https://www.wmts.nrw.de/geobasis/wmts_nw_dop/1.0.0/WMTSCapabilities.xml?REQUEST=GetCapabilities&SERVICE=WMTS%20 (Accessed: 16 July 2026).

Xu, C. *et al.* (2024) 'Relating soil moisture and Sentinel-2 vegetation index patterns to spruce bark beetle infestations prior to outbreak', *Forestry: An International Journal of Forest Research*, 97(5), pp. 728–738. Available at: https://doi.org/10.1093/forestry/cpae007.

Zakrzewska, A. and Kopeć, D. (2022) 'Remote sensing of bark beetle damage in Norway spruce individual tree canopies using thermal infrared and airborne laser scanning data fusion', *Forest Ecosystems*, 9, p. 100068. Available at: https://doi.org/10.1016/j.fecs.2022.100068.

7. Appendix

The full repository of my code, both Python and R can be found on GitHub.

https://github.com/andikoeller/Thesis_Barkbeetle_Koeller?tab=readme-ov-file

Additionally, I also published a supplementary markdown Blog on Github pages. This is the best way to see code, results and diagrams right next to each other.

https://andikoeller.github.io/Thesis_Barkbeetle_Koeller/

Data source which could not be uploaded area linked in the following table and are freely available.

7.1 Data Sources with Links

Table 12: Overview and Access links of Data sources

Layer	**Purpose**	**Source**	**Land cover nrw**
Forest Mask (administrative)	Creating Spruce Mask	ALKIS Basis DLM, Bezirksregierung Köln	https://www.bezreg-koeln.nrw.de/geobasis-nrw/produkte-und-dienste/landschaftsmodelle/aktuelle-landschaftsmodelle/digitales-basis
Forest Mask (actual Land Cover)		WMS NW Landbedeckung, Tim-OnlineNRW.de	https://www.opengeodata.nrw.de/produkte/geobasis/lusat/akt/lb_nrw/
Wald und Holz Opendata *Forest Damage Classification* *Tree species Classification*	Assessing Spruce Mask	Landesbetrieb Wald und Holz Nordrhein-Westfalen, via OpenGeodata.NRW	https://www.opengeodata.nrw.de/produkte/umwelt_klima/wald_forst/fernerkundung
Digitial Orthophotos (DOP)	**Reference** Source for classifying time of attack	Geobasis Bezirksregierung Köln & QGIS Plugin "Luftbildfinder NRW)	https://www.bezreg-koeln.nrw.de/geobasis-nrw/produkte-und-dienste/luftbild-und-satellitenbildinformationen/aktuelle-luftbild-und-0fsk

Moisture Level	Base Data for **Soil Moisture**	Forstliche Standortskarte 1:5000, Geologischer Dienst NRW, via OpenGeoData.NRW	https://www.opengeodata.nrw.de/produkte/geologie/boden/BK/
DEM	Base Data for Calculating **Slope** and **Aspect**	OpenGeoData.NRW	https://www.opengeodata.nrw.de/produkte/geobasis/hm/dgm1_tiff/
Sentinel 2	Provides **Master Grid** for Analysis	Copernicus Browser	Copernicus Browser

7.2 Full Results Table for moisture on the Sickerwasser gradient

Table 13: All Variations of the test of time of attack/Moisture Sickerwasser with different subsets of stands and methods. All results converge on stand level converge

Scale	*Subset*	*Method*	*Coefficient*	*p*	*n (stands)*
Stand	attacked only	Kendall tau-b	0.136	1.6e-6	1,069
Stand	attacked only	Spearman rho	0.146	1.5e-6	1,069
Stand	attacked only	Pearson (numeric year)	0.149	9.6e-7	1,069
Stand	+ surviving	Kendall tau-b	0.143	5.3e-8	1,206
Stand	+ surviving	Spearman rho	0.157	4.4e-8	1,206
Pixel	attacked only	Kendall tau-b	0.106	- (descr.)	119,323
Pixel	attacked only	Spearman rho	0.112	- (descr.)	119,323
Pixel	+ surviving	Kendall tau-b	0.105	- (descr.)	128,214
Pixel	+ surviving	Spearman rho	0.112	- (descr.)	128,214

7.3 Python Script (Processing Slope and Aspect TIFs)

Claude Opus 4.7 was used in generating the script and to turn the single processing steps into a Processing Tool Script for QGIS. This copy and coloration of the code was generated using Notepad ++ NppExport Plugin, using "Copy all formats to clipboard".

```
"""
DEM -> Sentinel-2 grid  (PyQGIS Processing script)

Resamples a 1 m DEM onto the Sentinel-2 10 m grid (e.g. tile T32ULB) and derives
aspect (+ optional slope) from the 10 m surface.

In ONE warp pass it:
  - reprojects the DEM (e.g. EPSG:25832) into the S2 CRS (e.g. EPSG:32632),
  - downsamples 1 m -> 10 m using AVERAGE (correct anti-aliasing for a 10x reduction),
  - snaps exactly to the chosen Sentinel-2 band's CRS / extent / pixel grid.

Then it computes aspect (and optionally slope) FROM the 10 m DEM, so the result
matches your analysis scale and avoids the circular-aspect resampling problem.

HOW TO INSTALL:
  Processing Toolbox -> Scripts (the Python icon) -> "Open Existing Script..."
  (or "Create New Script from Template" and paste this in), then Save.
  It will appear under:  Scripts > DEM tools > DEM to Sentinel-2 grid.
"""

import math

from qgis.PyQt.QtCore import QCoreApplication
from qgis.core import (
    QgsProcessing,
    QgsProcessingAlgorithm,
    QgsProcessingParameterRasterLayer,
    QgsProcessingParameterRasterDestination,
    QgsProcessingParameterBoolean,
    QgsCoordinateTransform,
)
import processing


class DemToSentinel2Grid(QgsProcessingAlgorithm):

    INPUT_DEM = "INPUT_DEM"
    REF_S2 = "REF_S2"
    COMPUTE_SLOPE = "COMPUTE_SLOPE"
    ZERO_FOR_FLAT = "ZERO_FOR_FLAT"
    CLIP_TO_DEM = "CLIP_TO_DEM"
    OUTPUT_DEM = "OUTPUT_DEM"
    OUTPUT_ASPECT = "OUTPUT_ASPECT"
    OUTPUT_SLOPE = "OUTPUT_SLOPE"

    # --- boilerplate ---------------------------------------------------------
    def tr(self, string):
        return QCoreApplication.translate("DemToSentinel2Grid", string)

    def createInstance(self):
        return DemToSentinel2Grid()

    def name(self):
        return "dem_to_sentinel2_grid"

    def displayName(self):
        return self.tr("DEM to Sentinel-2 grid")

    def group(self):
        return self.tr("DEM tools")

    def groupId(self):
        return "demtools"

    def shortHelpString(self):
        return self.tr(
            "Reproject a DEM into the Sentinel-2 CRS, resample 1 m -> 10 m with "
            "AVERAGE, and snap exactly to a chosen Sentinel-2 10 m band's grid. "
            "Then derive aspect (and optionally slope) from the 10 m surface.\n\n"
```

```
            "Pick any true 10 m band (e.g. B03_10m) from your tile (e.g. T32ULB) "
            "as the reference grid."
        )

    # --- parameters ----------------------------------------------------------
    def initAlgorithm(self, config=None):
        self.addParameter(
            QgsProcessingParameterRasterLayer(
                self.INPUT_DEM, self.tr("Input DEM (e.g. 1 m, EPSG:25832)")
            )
        )
        self.addParameter(
            QgsProcessingParameterRasterLayer(
                self.REF_S2,
                self.tr("Sentinel-2 reference band (10 m, defines target grid)"),
            )
        )
        self.addParameter(
            QgsProcessingParameterBoolean(
                self.COMPUTE_SLOPE, self.tr("Also compute slope (degrees)"),
defaultValue=False
            )
        )
        self.addParameter(
            QgsProcessingParameterBoolean(
                self.CLIP_TO_DEM,
                self.tr("Clip to DEM extent (grid-aligned) — avoids a full-tile, mostly-empty
raster"),
                defaultValue=True,
            )
        )
        self.addParameter(
            QgsProcessingParameterBoolean(
                self.ZERO_FOR_FLAT,
                self.tr("Aspect: assign 0 to flat cells (else -9999)"),
                defaultValue=False,
            )
        )
        self.addParameter(
            QgsProcessingParameterRasterDestination(
                self.OUTPUT_DEM, self.tr("Resampled DEM (10 m, S2 grid)")
            )
        )
        self.addParameter(
            QgsProcessingParameterRasterDestination(
                self.OUTPUT_ASPECT, self.tr("Aspect (10 m, S2 grid)")
            )
        )
        self.addParameter(
            QgsProcessingParameterRasterDestination(
                self.OUTPUT_SLOPE,
                self.tr("Slope (10 m, S2 grid)"),
                optional=True,
                createByDefault=False,
            )
        )

    # --- processing ----------------------------------------------------------
    def processAlgorithm(self, parameters, context, feedback):
        dem = self.parameterAsRasterLayer(parameters, self.INPUT_DEM, context)
        ref = self.parameterAsRasterLayer(parameters, self.REF_S2, context)

        ref_crs = ref.crs()
        ext = ref.extent()
        # target pixel size taken from the reference band (should be 10 m)
        px = ref.rasterUnitsPerPixelX()
        py = ref.rasterUnitsPerPixelY()

        clip_to_dem = self.parameterAsBool(parameters, self.CLIP_TO_DEM, context)

        if clip_to_dem:
            # Snap the DEM's own footprint to the S2 grid lines, so the output is
            # cropped tight to the DEM but its pixel boundaries still coincide
            # exactly with the Sentinel-2 grid (same origin/spacing as the ref band).
            xform = QgsCoordinateTransform(dem.crs(), ref_crs, context.transformContext())
            dem_ext = xform.transformBoundingBox(dem.extent())

            # grid anchor = ref tile's lower-left corner; expand OUTWARD to grid lines
```

```
            ax, ay = ext.xMinimum(), ext.yMinimum()
            xmin = ax + math.floor((dem_ext.xMinimum() - ax) / px) * px
            xmax = ax + math.ceil((dem_ext.xMaximum() - ax) / px) * px
            ymin = ay + math.floor((dem_ext.yMinimum() - ay) / py) * py
            ymax = ay + math.ceil((dem_ext.yMaximum() - ay) / py) * py
            target_extent = f"{xmin},{xmax},{ymin},{ymax}"
            warp_extra = ""  # extent is already grid-aligned; no -tap needed
            feedback.pushInfo("Clip mode: ON (tight, grid-aligned to S2)")
        else:
            # Full S2 tile extent (output covers the whole 100x100 km tile).
            target_extent =
f"{ext.xMinimum()},{ext.xMaximum()},{ext.yMinimum()},{ext.yMaximum()}"
            warp_extra = "-tap"
            feedback.pushInfo("Clip mode: OFF (full tile extent)")

        feedback.pushInfo(f"DEM CRS:    {dem.crs().authid()}")
        feedback.pushInfo(f"Target:     {ref_crs.authid()}  |  pixel: {px} x {py}")
        feedback.pushInfo(f"Snap extent: {target_extent} [{ref_crs.authid()}]")

        # 1) ONE warp pass: reproject + average resample to 10 m + snap to S2 grid
        feedback.pushInfo("Step 1/2-3: warp (reproject + average resample + snap) ...")
        warp = processing.run(
            "gdal:warpreproject",
            {
                "INPUT": dem,
                "SOURCE_CRS": dem.crs(),
                "TARGET_CRS": ref_crs,
                "RESAMPLING": 5,  # 5 = average
                "TARGET_RESOLUTION": px,
                "TARGET_EXTENT": target_extent,
                "TARGET_EXTENT_CRS": ref_crs,
                "NODATA": None,
                "DATA_TYPE": 6,  # Float32
                "EXTRA": warp_extra,  # "-tap" for full tile; "" when clip extent is pre-
snapped
                "OUTPUT": parameters.get(self.OUTPUT_DEM) or QgsProcessing.TEMPORARY_OUTPUT,
            },
            context=context,
            feedback=feedback,
            is_child_algorithm=True,
        )
        dem10 = warp["OUTPUT"]

        # 2) aspect FROM the 10 m surface
        feedback.pushInfo("Step 2: aspect from the 10 m DEM ...")
        zero_for_flat = self.parameterAsBool(parameters, self.ZERO_FOR_FLAT, context)
        aspect = processing.run(
            "gdal:aspect",
            {
                "INPUT": dem10,
                "BAND": 1,
                "TRIG_ANGLE": False,
                "ZERO_FLAT": zero_for_flat,
                "COMPUTE_EDGES": True,
                "ZEVENBERGEN": False,
                "OUTPUT": parameters.get(self.OUTPUT_ASPECT) or
QgsProcessing.TEMPORARY_OUTPUT,
            },
            context=context,
            feedback=feedback,
            is_child_algorithm=True,
        )

        results = {
            self.OUTPUT_DEM: dem10,
            self.OUTPUT_ASPECT: aspect["OUTPUT"],
        }

        # 3) optional slope FROM the 10 m surface
        if self.parameterAsBool(parameters, self.COMPUTE_SLOPE, context):
            feedback.pushInfo("Step 3: slope from the 10 m DEM ...")
            # The slope destination is optional (createByDefault=False), so if the
            # user ticked "compute slope" but gave no path, it's absent from
            # `parameters`. Fall back to a temporary output instead of KeyError.
            slope_dest = parameters.get(self.OUTPUT_SLOPE) or QgsProcessing.TEMPORARY_OUTPUT
            slope = processing.run(
                "gdal:slope",
                {
```

```
                    "INPUT": dem10,
                    "BAND": 1,
                    "SCALE": 1,
                    "AS_PERCENT": False,
                    "COMPUTE_EDGES": True,
                    "ZEVENBERGEN": False,
                    "OUTPUT": slope_dest,
                },
                context=context,
                feedback=feedback,
                is_child_algorithm=True,
            )
            results[self.OUTPUT_SLOPE] = slope["OUTPUT"]

        feedback.pushInfo("Done. All outputs are in the S2 CRS, snapped to the reference
grid.")
        return results
```

7.4 R Scripts (Statistical Analysis, Summaries of Results)

Claude Opus 4.8 was used in refactoring the statistical analysis scripts into a coherent, sequential R script that can reproduce the results and is shown below. Reproducibility was checked and ensured by the fixed seed number in the script. This copy and coloration of the code was generated using Notepad ++ NppExport Plugin, using "Copy all formats to clipboard". The Full Code and documentation as Markdown Blog is available on GitHub (see 7.0).

https://github.com/andikoeller/Thesis_Barkbeetle_Koeller?tab=readme-ov-file

https://andikoeller.github.io/Thesis_Barkbeetle_Koeller/

```
# Appendix: R analysis code ( subset testing; figure-generation code
removed).
# Verbatim from the analysis scripts, in run order:
#   setup -> import -> chi-square -> ordinal/metric -> collinearity.
# The collinearity block is adapted from Results_reproduce.R; terrain
preprocessing
# is in Script_DEM_Resampling.py (PyQGIS); the full scripts are in the
repository.
# Author: Andreas Koeller


# --- 00_setup.R : constants + helper functions ---

#
==========================================================================
=
# 00_setup.R  --  run this FIRST in a fresh session.
#
# DATA IN      : nothing (bootstraps the session).
# PROVIDES     : the libraries used across the project, the shared helper
#                functions (mode_cat, cramers_v, kendall_tau_b) and the
shared
#                constants (Sickerwasser_levels, moist_order, moist_group)
that
#                several scripts rely on. Sourcing this once removes the
hidden
#                run-order and duplicated-definition dependencies between
scripts.
# OUTPUTS/VIZ  : none (definitions only).
#
# Pipeline (source in this order, once per fresh R session):
#   00_setup.R                helpers + libraries + constants      (RUN
FIRST)
#   01_import.R               raster -> df_pixel -> df_stand
#   02_descriptive.R          codebook, summaries, EDA figures
(optional)
#   03_chi2.R                 chi-square + Cramer's V, pixel & stand (+
figures)
#   04_correlation.R          Kendall / Spearman / Pearson trend    (+
figure)
#   Results_reproduce.R       self-contained: re-derives EVERY reported
value
#
==========================================================================
=
```

```r
# --- libraries ----------------------------------------------------------------
----
suppressMessages({
  library(terra)      # raster import
  library(dplyr)      # data wrangling
  library(tidyr)      # reshaping (EDA)
  library(arrow)      # parquet backup
  library(ggplot2)    # figures
})

# --- portable project root ------------------------------------------------
----
# Auto-detect the folder that contains the .Rproj file, walking up from the
# working directory. This makes the whole pipeline path-independent: copy
the
# folder anywhere and it still resolves its own outputs (figures, Parquet,
CSVs).
# Falls back to getwd() if no .Rproj is found.
.find_root <- function(start = getwd()) {
  d <- normalizePath(start, winslash = "/", mustWork = FALSE)
  repeat {
    if (length(list.files(d, pattern = "\\.Rproj$"))) return(d)
    parent <- dirname(d); if (parent == d) break; d <- parent
  }
  normalizePath(getwd(), winslash = "/", mustWork = FALSE)
}
PROJ_ROOT <- .find_root()
# Source raster stack. Override this if the raster lives elsewhere; if it
is not
# found, 01_import.R falls back to the bundled Import_final.parquet.
raster_path <- file.path(PROJ_ROOT, "data", "Raster_Stack.vrt")

# --- shared constants -----------------------------------------------------
----
# The ordered seepage-water ("Sickerwasser") moisture gradient (dry ->
fresh).
# "sehr trocken" is a valid level but absent in the data. This gradient is
the
# primary ordinal predictor; it was previously redefined in ~6 scripts.
Sickerwasser_levels <- c("sehr trocken", "trocken", "mäßig trocken",
                         "mäßig frisch", "frisch", "sehr frisch")

# Full moisture ordering (dry -> wet within each hydrological group) for
ordered
# axes and grouping in the descriptive figures.
moist_order <- c("sehr trocken", "trocken", "mäßig trocken", "mäßig
frisch",
                 "frisch", "sehr frisch", "wechseltrocken", "mäßig
wechselfeucht",
                 "wechselfeucht", "staunass", "grundfrisch", "grundfeucht",
                 "feucht", "nass")

# Map a moisture class to its hydrological water-budget group.
moist_group <- function(x) dplyr::case_when(
  x %in% c("sehr trocken","trocken","mäßig trocken","mäßig
frisch","frisch","sehr frisch") ~ "Sickerwasser",
  x %in% c("wechseltrocken","mäßig
wechselfeucht","wechselfeucht","staunass")                ~ "Stauwasser",
  x %in% c("grundfrisch","grundfeucht","feucht","nass")
~ "Grundwasser",
  TRUE ~ NA_character_)
```

```
# --- shared helper functions ------------------------------------------------
----

# Modal category: the most frequent level of a categorical vector, ignoring
NA.
# Ties are broken by factor order (the first level among tied maxima).
#
# IMPORTANT: this is the original definition from the project .Rhistory
PLUS a
# guard for all-NA input. The original (without the guard) returned the
FIRST
# factor level for a stand that has no data at all -- silently mislabelling
the
# ~33 stands with no site-moisture mapping as "feucht", which inflated the
# moisture-all chi-square (70.2 -> the correct 60.4). The guard returns NA
so
# those stands are correctly dropped.
mode_cat <- function(x) {
  x <- x[!is.na(x)]
  if (length(x) == 0) return(NA)
  names(sort(table(x), decreasing = TRUE))[1]
}

# Cramer's V: effect size for a chi-square contingency table (0 = none,
# ~0.1 weak, ~0.3 moderate, ~0.5 strong). The reportable measure of
# association strength, independent of sample size. Used in 03_chi2.R.
cramers_v <- function(tab) {
  chi <- suppressWarnings(chisq.test(tab))
  sqrt(as.numeric(chi$statistic) / (sum(tab) * (min(dim(tab)) - 1)))
}

# Kendall's tau-b from a contingency table (O(cells), avoids the O(n^2)
blow-up
# of cor.test(method="kendall") on ~10^5 rows). Identical value, instant.
Used
# for the descriptive pixel-level trend in 04_correlation.R.
kendall_tau_b <- function(tab) {
  tab <- as.matrix(tab)
  R <- nrow(tab); C <- ncol(tab)
  conc <- 0; disc <- 0
  for (i in 1:R) for (j in 1:C) {
    if (i < R && j < C) conc <- conc + tab[i, j] * sum(tab[(i + 1):R, (j +
1):C])
    if (i < R && j > 1) disc <- disc + tab[i, j] * sum(tab[(i + 1):R, 1:(j
- 1)])
  }
  n  <- sum(tab)
  n0 <- n * (n - 1) / 2
  n1 <- sum(rowSums(tab) * (rowSums(tab) - 1) / 2)   # ties within rows
  n2 <- sum(colSums(tab) * (colSums(tab) - 1) / 2)   # ties within columns
  (conc - disc) / sqrt((n0 - n1) * (n0 - n2))
}

cat("00_setup.R loaded: libraries + constants +
mode_cat()/cramers_v()/kendall_tau_b() ready.\n")


# --- 01_import.R : raster -> df_pixel -> df_stand ---

#
==========================================================================
=
```

```
# 01_import.R  --  build the two analysis data frames from the raster
stack.
#
# DATA IN      : Raster_Stack.vrt (10 m, 7 bands: aspect, phase, slope,
#                stand_area, stand_id, VisKlass2, moisture), path from
00_setup.R.
#                Falls back to the bundled Import_final.parquet if the
raster is
#                unreachable (e.g. the folder was copied to another
machine).
# ANALYSIS     : none -- this is data preparation only:
#                (1) clean per-band NoData, (2) classify aspect/slope,
#                (3) decode moisture and attack-phase codes to labels,
#                (4) aggregate pixels to stands (modal classes, median
slope,
#                    directional aspect components + reliability).
# OUTPUTS      : df_pixel  (155,910 rows, one per 10 m pixel)   [+ Parquet
backup]
#                df_stand  (1,470 rows, one per forest stand)
# OUTPUTS/VIZ  : none (see 02_descriptive.R for figures of these tables).
#
# PREREQUISITE : source("00_setup.R") first (libraries + mode_cat +
PROJ_ROOT).
#
=============================================================================
=

library(terra)
library(dplyr)
library(arrow)

# --- portable paths -------------------------------------------------------
----
# PROJ_ROOT / raster_path are normally provided by 00_setup.R; define
fallbacks
# so this script also works if run standalone from the project folder.
if (!exists("PROJ_ROOT"))   PROJ_ROOT   <- normalizePath(getwd(), winslash
= "/", mustWork = FALSE)
if (!exists("raster_path")) raster_path <- file.path(PROJ_ROOT, "data",
"Raster_Stack.vrt")
parquet_path <- file.path(PROJ_ROOT, "data", "Import_final.parquet")

#
=============================================================================
=
# STEP 1 -- raster -> df_pixel
#
=============================================================================
=
if (!file.exists(raster_path)) {
  # Raster not reachable -> use the cached pixel table bundled in the
project.
  message("Raster not found at ", raster_path,
          "\n  -> loading cached df_pixel from ", parquet_path)
  df_pixel <- as.data.frame(arrow::read_parquet(parquet_path))
} else {
  # a) Import
  r <- rast(raster_path)
  names(r) <- c("aspect", "phase", "slope", "stand_area",
                "stand_id", "VisKlass2", "moisture")
  NAflag(r) <- NA   # ignore the wrong NoData=0 tag; mask per band below
```

```
  df <- as.data.frame(r, xy = TRUE, cells = TRUE, na.rm = FALSE) |>
    rename(cell_id = cell) |>
    mutate(
      aspect     = na_if(aspect,     -9999),
      slope      = na_if(slope,      -9999),
      phase      = na_if(phase,       0),
      stand_area = na_if(stand_area,  0),
      stand_id   = na_if(stand_id,    0),
      VisKlass2  = na_if(VisKlass2,   0),
      moisture   = na_if(moisture,    0)
    ) |>
    filter(!is.na(phase)) |>
    mutate(across(c(phase, moisture, VisKlass2, stand_id), factor))

  # b) Klassifizieren Aspect/Slope
  df <- df |>
    mutate(
      aspect_class = cut((aspect + 22.5) %% 360,
                         breaks = seq(0, 360, 45),
                         labels = c("N","NE","E","SE","S","SW","W","NW"),
                         include.lowest = TRUE),
      slope_class  = cut(slope,
                         breaks = c(0, 5, 10, 15, 20, 30, Inf),
                         labels = c("0-5","5-10","10-15","15-20","20-
30","30+"),
                         include.lowest = TRUE, right = FALSE)
    )

  # c) Zurückkodieren der Feuchteklassen
  df$moisture_text <- factor(c("feucht","frisch","grundfeucht","mäßig
frisch",
                               "mäßig trocken","mäßig
wechselfeucht","nass",
                               "sehr
frisch","trocken","wechselfeucht")[as.integer(as.character(df$moisture))])

  # d) zurückkodieren der befallsphasen
  df <- df |>
    mutate(
      TimeOfAttack = factor(c("2"="2019","3"="2021","4"="2023","5"="2024",

"6"="2025","10"="surviving","20"="irrelevant")[as.character(phase)],
                            levels =
c("irrelevant","2019","2021","2023","2024","2025","surviving"))
    )

  # e) Export Parquet (Backup, inside the project so a copied folder is
self-contained)
  write_parquet(df, parquet_path)

  #) rename DF for pixel level
  df_pixel <- df
}

cat("df_pixel:", nrow(df_pixel), "pixels x", ncol(df_pixel), "cols\n")
str(df_pixel)

#
===========================================================================
=
# STEP 2 -- df_pixel -> df_stand   (needs mode_cat from 00_setup.R)
```

```
#
===========================================================================
=
# One row per forest stand: modal categorical classes, median slope, and
the
# directional aspect components (northness/eastness) with a reliability
flag.
df_stand <- df_pixel |>
  group_by(stand_id) |>
  summarise(
    TimeOfAttack  = first(TimeOfAttack),
    n_pixel       = n(),
    stand_area    = first(stand_area),
    slope_median  = median(slope, na.rm = TRUE),
    northness     = mean(cos(aspect * pi / 180), na.rm = TRUE),   # +1 = N,
-1 = S
    eastness      = mean(sin(aspect * pi / 180), na.rm = TRUE),   # +1 = E,
-1 = W
    slope_class   = mode_cat(slope_class),
    aspect_class  = mode_cat(aspect_class),
    moisture_text = mode_cat(moisture_text),
    .groups = "drop"
  ) |>
  mutate(
    # aspect reliability: mean resultant length of the pixel aspects (0-1).
    # High R = pixels point the same way (mean aspect trustworthy); low R =
    # aspects scattered around the compass (northness/eastness
meaningless).
    aspect_R        = sqrt(northness^2 + eastness^2),
    aspect_reliable = aspect_R >= 0.5,
    TimeOfAttack  = factor(TimeOfAttack,  levels =
levels(df_pixel$TimeOfAttack)),
    slope_class   = factor(slope_class,   levels =
levels(df_pixel$slope_class)),
    aspect_class  = factor(aspect_class,  levels =
levels(df_pixel$aspect_class)),
    moisture_text = factor(moisture_text, levels =
levels(df_pixel$moisture_text))
  )

cat("df_stand:", nrow(df_stand), "stands x", ncol(df_stand), "cols\n")


# --- 03_chi2.R : chi-square + Cramer's V ---

#
===========================================================================
=
# 03_chi2.R  --  nominal association: attack timing vs site factors (chi-
square)
#
# DATA IN      : df_pixel, df_stand   (from 01_import.R)
#                Constants Sickerwasser_levels + helper cramers_v() from
00_setup.R.
# ANALYSIS     : Pearson chi-square of TimeOfAttack against each site
factor
#                (slope class, aspect class, moisture class), reported as
the
#                effect size Cramer's V (association strength, sample-size
free)
#                plus standardised residuals (which cells drive the
association).
```

```
#                  * PIXEL level (n=155,910): descriptive effect sizes only
--
#                  * STAND level (n=1,272):; chi-square
#                    p-values are Monte-Carlo (sparse cells), seed 42,
B=20,000.
#                  * SICKERWASSER (n=1,206): chi-square on the ordered
seepage-
#                    water gradient (nominal test; the ORDINAL trend is
04).
# OUTPUTS/VIZ  : EDA/fig_effectsize_cramersV.png  (Cramer's V bar, section
3.2)
#                EDA/fig_residual_heatmap.png      (std residuals, section
3.6)
#                EDA/chi_summary.csv               (chi2 / p_MC / Cramer's
V table)
#
# PREREQUISITE : source 00_setup.R -> 01_import.R first.
#
============================================================================
=

library(dplyr); library(tidyr); library(ggplot2)

out_dir <- if (exists("PROJ_ROOT")) PROJ_ROOT else normalizePath(getwd(),
winslash = "/", mustWork = FALSE)
eda_dir <- file.path(out_dir, "EDA"); dir.create(eda_dir, showWarnings =
FALSE)
theme_set(if (exists("THEME_THESIS")) THEME_THESIS else
theme_minimal(base_size = 11))

set.seed(42)       # Monte-Carlo chi-square p-values reproducible ...
B_MC <- 20000      # ... and consistent with Results_reproduce.R
if (!exists("Sickerwasser_levels"))
  Sickerwasser_levels <- c("sehr trocken", "trocken", "mäßig trocken",
                           "mäßig frisch", "frisch", "sehr frisch")

#
============================================================================
=
# A.  PIXEL LEVEL
#
============================================================================
=
cat("\n########## A. PIXEL-LEVEL chi-square / Cramer's V (descriptive)
##########\n")
df_px <- df_pixel |>
  filter(TimeOfAttack != "irrelevant") |>
  mutate(TimeOfAttack = droplevels(TimeOfAttack))

for (v in c("slope_class", "aspect_class", "moisture_text")) {
  tab <- table(df_px$TimeOfAttack, droplevels(df_px[[v]]))
  cat("\n--- pixel:", v, "---\n")
  print(tab)
  cat("Cramer's V:", round(cramers_v(tab), 3), "\n")
}

#
============================================================================
=
# B.  STAND LEVEL  --  chi-square (MC p), residuals, Cramer's V
[inferential]
```

```
#
================================================================================
=
cat("\n########## B. STAND-LEVEL chi-square (MC p, seed 42, B =", B_MC, ")
##########\n")
df_stand_a <- df_stand |>
  filter(TimeOfAttack != "irrelevant") |>
  mutate(TimeOfAttack = droplevels(TimeOfAttack))
cat("n stands (excl. irrelevant):", nrow(df_stand_a), "\n")

chi_stand <- list()   # collect chi2 / p / V for the effect-size figure
for (v in c("slope_class", "aspect_class", "moisture_text")) {
  tab  <- table(df_stand_a$TimeOfAttack, droplevels(df_stand_a[[v]]))
  test <- chisq.test(tab, simulate.p.value = TRUE, B = B_MC)
  cat("\n--- stand:", v, "---\n")
  print(tab)
  cat("chi2 =", round(test$statistic, 1),
      "| p (MC) =", round(test$p.value, 4),
      "| Cramer's V =", round(cramers_v(tab), 3), "\n")
  cat("standardised residuals (>|2| notable):\n")
  print(round(test$stdres, 1))
  chi_stand[[v]] <- c(chi2 = as.numeric(test$statistic),
                      p = test$p.value, V = cramers_v(tab))
}

#
================================================================================
=
# C.  SICKERWASSER GRADIENT  --  chi-square on the ordered seepage-water
sites
#
================================================================================
=
cat("\n########## C. SICKERWASSER ordered chi-square ##########\n")
df_stand_Sickerwasser <- df_stand_a |>
  filter(moisture_text %in% Sickerwasser_levels) |>
  mutate(
    TimeOfAttack          = droplevels(TimeOfAttack),
    moisture_Sickerwasser = factor(moisture_text, levels =
Sickerwasser_levels, ordered = TRUE) |>
                              droplevels()
  )
cat("n Sickerwasser stands:", nrow(df_stand_Sickerwasser), "\n")

tab_Sickerwasser  <- table(df_stand_Sickerwasser$TimeOfAttack,
                           df_stand_Sickerwasser$moisture_Sickerwasser)
test_Sickerwasser <- chisq.test(tab_Sickerwasser, simulate.p.value = TRUE,
B = B_MC)
print(tab_Sickerwasser)
cat("chi2 =", round(test_Sickerwasser$statistic, 1),
    "| p (MC) =", round(test_Sickerwasser$p.value, 4),
    "| Cramer's V =", round(cramers_v(tab_Sickerwasser), 3), "\n")
cat("standardised residuals:\n"); print(round(test_Sickerwasser$stdres, 1))
#
================================================================================
=
# E.  SPARSE-CELL DIAGNOSTICS + ROBUSTNESS  (Monte-Carlo p )
================================================================================
=
# Many stand-level cells have small expected counts. Cochran's rule: the
# asymptotic chi-square is unreliable if >20% of cells have expected E<5
(or any
```

```
# E<1). This is driven by the sparse recent years (2024 n=42, 2025 n=13).
cat("\n########## E. sparse-cell diagnostics + robustness ##########\n")

diag_row <- function(tab, scale, label, report_p = TRUE) {
  a <- suppressWarnings(chisq.test(tab)); E <- a$expected
  p_as <- if (report_p) signif(a$p.value, 3) else NA_real_
  p_mc <- if (report_p) signif(chisq.test(tab, simulate.p.value = TRUE, B =
B_MC)$p.value, 3) else NA_real_
  data.frame(scale = scale, table = label, rows = nrow(tab), cols =
ncol(tab),
             cells = length(tab), N = sum(tab), E_lt5 = sum(E < 5),
             pct_E_lt5 = round(100 * mean(E < 5)), min_E = round(min(E),
2),
             chi2 = round(as.numeric(a$statistic), 1), df =
as.integer(a$parameter),
             p_asympt = p_as, p_MC = p_mc, CramersV = round(cramers_v(tab),
3),
             stringsAsFactors = FALSE)
}

# pixel Sickerwasser subset (for the pixel diagnostics row)
df_px_S <- df_px |> filter(moisture_text %in% Sickerwasser_levels) |>
  mutate(mS = factor(moisture_text, levels = Sickerwasser_levels) |>
droplevels())

# collapsed stand timing:
early(2019+2021)/mid(2023)/late(2024+2025)/surviving (robustness check)
collapse_timing <- function(x) factor(dplyr::case_when(
  x %in% c("2019", "2021") ~ "early", x == "2023" ~ "mid",
  x %in% c("2024", "2025") ~ "late", TRUE ~ "surviving"),
  levels = c("early", "mid", "late", "surviving"))
df_stand_c  <- df_stand_a            |> mutate(tc =
collapse_timing(TimeOfAttack))
df_stand_Sc <- df_stand_Sickerwasser |> mutate(tc =
collapse_timing(TimeOfAttack))

set.seed(42)   # MC p-values in the diagnostics block reproducible as a
unit
sparsity <- rbind(
  diag_row(table(df_px$TimeOfAttack,   droplevels(df_px$slope_class)),
"pixel",                "slope",         report_p = FALSE),
  diag_row(table(df_px$TimeOfAttack,   droplevels(df_px$aspect_class)),
"pixel",                "aspect",        report_p = FALSE),
  diag_row(table(df_px$TimeOfAttack,   droplevels(df_px$moisture_text)),
"pixel",                "moisture all",  report_p = FALSE),
  diag_row(table(df_px_S$TimeOfAttack, df_px_S$mS),
"pixel",                "moisture Sicker", report_p = FALSE),
  diag_row(table(df_stand_a$TimeOfAttack,
droplevels(df_stand_a$slope_class)),   "stand (6 timing)", "slope"),
  diag_row(table(df_stand_a$TimeOfAttack,
droplevels(df_stand_a$aspect_class)),  "stand (6 timing)", "aspect"),
  diag_row(table(df_stand_a$TimeOfAttack,
droplevels(df_stand_a$moisture_text)), "stand (6 timing)", "moisture all"),
  diag_row(table(df_stand_Sickerwasser$TimeOfAttack,
df_stand_Sickerwasser$moisture_Sickerwasser), "stand (6 timing)", "moisture
Sicker"),
  diag_row(table(df_stand_c$tc,  droplevels(df_stand_c$slope_class)),
"stand (4 collapsed)", "slope"),
  diag_row(table(df_stand_c$tc,  droplevels(df_stand_c$aspect_class)),
"stand (4 collapsed)", "aspect"),
  diag_row(table(df_stand_c$tc,  droplevels(df_stand_c$moisture_text)),
"stand (4 collapsed)", "moisture all"),
```

```
  diag_row(table(df_stand_Sc$tc, df_stand_Sc$moisture_Sickerwasser),
"stand (4 collapsed)", "moisture Sicker")
)
write.csv(sparsity, file.path(eda_dir, "chi2_sparsity_diagnostics.csv"),
row.names = FALSE)
cat("\n=== sparse-cell diagnostics (EDA/chi2_sparsity_diagnostics.csv)
===\n")
print(sparsity, row.names = FALSE)
cat("\nReading:",
    "\n * PIXEL: Cochran issue minor (slope/aspect min E >= 4; moisture
~17% cells E<5)",
    "   and p is omitted anyway (pseudoreplication) -> not the reason pixel
p is invalid.",
    "\n * STAND (6 timing): 21-71% of cells have E<5 -> asymptotic chi2
unreliable",
    "   (see 'moisture all': asymptotic ~0.005 vs MC ~0.05) -> MC p
reported.",
    "\n * STAND (4 collapsed): far fewer sparse cells (aspect 0%), Cramer's
V ~unchanged,",
    "   conclusions intact (only moisture associates). Near the 0.05
boundary the exact",
    "   MC digit is seed-sensitive; quote Results_values.csv for the
headline value.\n")

cat("\nDone. Figures + chi_summary.csv + chi2_sparsity_diagnostics.csv
written to:", eda_dir, "\n")


# --- 04_correlation.R : Kendall / Spearman / regression ---

#
===========================================================================
=
# 04_correlation.R  --  ordinal / metric trend:
#
# DATA IN      : df_pixel, df_stand   (from 01_import.R)
#                Constants Sickerwasser_levels + helper kendall_tau_b()
from 00_setup.R.
# ANALYSIS     : rank + linear correlation of attack timing with site
factors.
#                Attack year is an ordinal time axis
(2019<2021<2023<2024<2025);
#                "surviving" has no position in time, "irrelevant" is
excluded.
#                  A. MOISTURE, stand level (n=1,069): Kendall tau-b +
Spearman
#                     rho on ranks; attacked-only and the +surviving
variant.
#                  B. MOISTURE, numeric year (n=1,069): Pearson r +
lm(attack_year
#                     ~ moisture_rank) -> years-per-step, 95% CI, R2; mean
year
#                     per moisture class (respects the real 2020/2022
calendar gaps).
#                  C. MOISTURE, pixel level (descriptive): tau-b (table
formula,
#                     avoids O(n^2)) + Spearman; effect sizes only
(pseudoreplication).
#                  D. SLOPE & ASPECT (n=1,119, attacked, all moisture
groups):
#                     Kendall/Spearman/Pearson on continuous slope_median
and the
```

```
#                      directional aspect components (northness = cos,
eastness = sin);
#                      aspect is circular so it is NOT rank-tested directly.
# OUTPUTS/VIZ  : EDA/fig_mean_attack_year.png  (mean year +/- 95% CI,
section 3.4)
#                 EDA/mean_attack_year.csv, EDA/trend_summary.csv
#
# NOTE ON n    : moisture uses the Sickerwasser subset (n=1,069);
slope/aspect use
#                 all attacked stands (n=1,119)
# PREREQUISITE : source 00_setup.R -> 01_import.R first.
#
===========================================================================
=

library(dplyr); library(ggplot2)

out_dir <- if (exists("PROJ_ROOT")) PROJ_ROOT else normalizePath(getwd(),
winslash = "/", mustWork = FALSE)
eda_dir <- file.path(out_dir, "EDA"); dir.create(eda_dir, showWarnings =
FALSE)
theme_set(if (exists("THEME_THESIS")) THEME_THESIS else
theme_minimal(base_size = 11))
if (!exists("Sickerwasser_levels"))
  Sickerwasser_levels <- c("sehr trocken", "trocken", "mäßig trocken",
                           "mäßig frisch", "frisch", "sehr frisch")
if (!exists("kendall_tau_b")) stop("kendall_tau_b() missing -- source
00_setup.R first.")

# Shared Sickerwasser subset (attacked + surviving, ordered gradient) -----
-----
df_sick <- df_stand |>
  filter(TimeOfAttack != "irrelevant", moisture_text %in%
Sickerwasser_levels) |>
  mutate(
    TimeOfAttack           = droplevels(TimeOfAttack),
    moisture_Sickerwasser = factor(moisture_text, levels =
Sickerwasser_levels, ordered = TRUE) |> droplevels()
  )

#
===========================================================================
=
# A.  MOISTURE, STAND LEVEL  --  Kendall tau-b + Spearman rho (ranks)
#
===========================================================================
=
cat("\n########## A. MOISTURE trend, stand level ##########\n")
df_trend <- df_sick |>
  filter(TimeOfAttack != "surviving") |>
  mutate(attack_rank   = as.integer(droplevels(TimeOfAttack)),    #
2019<2021<2023<2024<2025
         moisture_rank = as.integer(moisture_Sickerwasser))       # dry ->
fresh
cat("attacked Sickerwasser stands:", nrow(df_trend), "\n")
kt_a <- cor.test(df_trend$attack_rank, df_trend$moisture_rank, method =
"kendall")
sp_a <- suppressWarnings(cor.test(df_trend$attack_rank,
df_trend$moisture_rank, method = "spearman"))
print(kt_a); print(sp_a)
```

```
# Variant including "surviving" as the far end (never attacked = beyond
2025)
df_trend_surv <- df_sick |>
  mutate(attack_rank   = as.integer(droplevels(TimeOfAttack)),   #
2019<...<2025<surviving
         moisture_rank = as.integer(moisture_Sickerwasser))
cat("\n--- incl. surviving (n =", nrow(df_trend_surv), ") ---\n")
kt_s <- cor.test(df_trend_surv$attack_rank, df_trend_surv$moisture_rank,
method = "kendall")
sp_s <- suppressWarnings(cor.test(df_trend_surv$attack_rank,
df_trend_surv$moisture_rank, method = "spearman"))
print(kt_s); print(sp_s)

# Robustness: collapse the sparse recent years to early<mid<late (defuses
the
# 2024/2025 sparsity flagged in 03_chi2.R Section E). The trend should
persist.
df_trend3 <- df_trend |>
  mutate(timing3 = as.integer(factor(dplyr::case_when(
           TimeOfAttack %in% c("2019", "2021") ~ "early",
           TimeOfAttack == "2023" ~ "mid", TRUE ~ "late"),
           levels = c("early", "mid", "late"), ordered = TRUE)))
kt3 <- cor.test(df_trend3$timing3, df_trend3$moisture_rank, method =
"kendall")
cat("\n--- robustness (3-class early/mid/late): Kendall tau-b =",
round(kt3$estimate, 3),
    ", p =", signif(kt3$p.value, 3), " (vs 5-class", round(kt_a$estimate,
3), ") ---\n")

#
===========================================================================
=
# B.  MOISTURE, NUMERIC YEAR  --  Pearson + linear model (years per step)
#
===========================================================================
=
cat("\n########## B. MOISTURE, numeric attack year ##########\n")
df_year <- df_trend |> mutate(attack_year =
as.numeric(as.character(TimeOfAttack)))
cat("n stands:", nrow(df_year), "\n")
print(cor.test(df_year$attack_year, df_year$moisture_rank, method =
"pearson"))
m <- lm(attack_year ~ moisture_rank, data = df_year)   # slope>0 => wetter
later
print(summary(m)); print(confint(m))
cat("\nMean attack year per moisture class:\n")
year_tbl <- df_year |>
  group_by(moisture_Sickerwasser) |>
  summarise(n = n(), mean_attack_year = mean(attack_year),
            se = sd(attack_year) / sqrt(n()), .groups = "drop")
print(year_tbl |> mutate(mean_attack_year = round(mean_attack_year, 2)))

#
===========================================================================
=
# C.  MOISTURE, PIXEL LEVEL  --  descriptive tau-b + Spearman (effect size
only)
#
===========================================================================
=
# formula kendall_tau_b() gives the exact tau-b in O(cells), instant on
~10^5 rows.
```

```
cat("\n########## C. MOISTURE trend, pixel level (descriptive)
##########\n")
df_pixel_Sicker <- df_pixel |>
  filter(TimeOfAttack != "irrelevant", moisture_text %in%
Sickerwasser_levels) |>
  mutate(TimeOfAttack          = droplevels(TimeOfAttack),
         moisture_Sickerwasser = factor(moisture_text, levels =
Sickerwasser_levels, ordered = TRUE) |> droplevels())

px <- df_pixel_Sicker |>
  filter(TimeOfAttack != "surviving") |>
  mutate(attack_rank = as.integer(droplevels(TimeOfAttack)),
         moisture_rank = as.integer(moisture_Sickerwasser))
tab_px <- table(px$attack_rank, px$moisture_rank)
cat("n pixels (attacked):", nrow(px),
    "| Spearman rho:", round(cor(px$attack_rank, px$moisture_rank, method =
"spearman"), 3),
    "| Kendall tau-b:", round(kendall_tau_b(tab_px), 3), "\n")

px_surv <- df_pixel_Sicker |>
  mutate(attack_rank = as.integer(droplevels(TimeOfAttack)),
         moisture_rank = as.integer(moisture_Sickerwasser))
tab_px_surv <- table(px_surv$attack_rank, px_surv$moisture_rank)
cat("n pixels (+surviving):", nrow(px_surv),
    "| Spearman rho:", round(cor(px_surv$attack_rank,
px_surv$moisture_rank, method = "spearman"), 3),
    "| Kendall tau-b:", round(kendall_tau_b(tab_px_surv), 3), "\n")

#
===========================================================================
=
# D.  SLOPE & ASPECT  --  Kendall / Spearman / Pearson (attacked, all
groups)
#
===========================================================================
=
# SLOPE is ordinal/continuous -> rank tests valid (use continuous
slope_median,
# avoiding arbitrary class breaks). ASPECT is circular -> use its
continuous
# components northness (cos) and eastness (sin) instead of an invalid rank
test.
cat("\n########## D. SLOPE & ASPECT trend ##########\n")
df_t <- df_stand |>
  filter(!TimeOfAttack %in% c("irrelevant", "surviving")) |>
  mutate(attack_year = as.numeric(as.character(TimeOfAttack)))   #
2019..2025
cat("n attacked stands (all moisture groups):", nrow(df_t), "\n")
for (v in c("slope_median", "northness", "eastness")) {
  cat("\n--- ", v, " ---\n", sep = "")
  print(cor.test(df_t$attack_year, df_t[[v]], method = "kendall"))
  print(suppressWarnings(cor.test(df_t$attack_year, df_t[[v]], method =
"spearman")))
  print(cor.test(df_t$attack_year, df_t[[v]], method = "pearson"))
}
# A near-zero, non-significant coefficient for slope and both aspect
components


# --- Collinearity + joint model : Table 11 and the Section 3.4
independence check ---
```

```
# (adapted from Results_reproduce.R). Uses df_year, the attacked
Sickerwasser stands
# (n = 1,069) built in the 04_correlation section above.
# (pairwise Spearman + VIF),
dc <- na.omit(df_year[, c("moisture_rank", "slope_median", "northness",
"eastness")])
cat("\ncollinearity n:", nrow(dc), "\n")
cat("Pairwise Spearman among predictors:\n")
print(round(cor(dc, method = "spearman"), 3))
vpred <- c("moisture_rank", "slope_median", "northness", "eastness")
vif <- sapply(vpred, function(v) {
  f <- as.formula(paste(v, "~", paste(setdiff(vpred, v), collapse = "+")))
  1 / (1 - summary(lm(f, data = dc))$r.squared)
})
cat("VIF (all ~1 => no collinearity):\n"); print(round(vif, 2))
mj <- lm(attack_year ~ moisture_rank + slope_median + northness + eastness,
data = df_year)
cat("Joint model coefficients:\n"); print(round(summary(mj)$coefficients,
4))
cat("Joint R2:", round(summary(mj)$r.squared, 3), "\n")
```