# Unlocking Pattern Extraction from Geospatial Big Data - Methodological Innovations in Remote Sensing and Geospatial Analysis for Environmental and Geoarchaeological Research

Von der Fakultät für Georessourcen und Materialtechnik der

Rheinisch-Westfälischen Technischen Hochschule Aachen

zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigte Dissertation

vorgelegt von

Bruno Boemke, M.Sc.

**Berichter:** Univ.-Prof. Dr. rer. nat. Frank Lehmkuhl

Jun.-Prof. Dr. phil. Andreas Maier

Tag der mündlichen Prüfung: 12.01.2024

# Abstract

In recent decades, the amount of openly available geodata has increased exponentially. To a large part, this can be attributed to the technological progress in satellite remote sensing, producing world-wide coverage from a large variety of sensors on a sub-daily basis. In addition, improved accessibility and a growing user base have favoured methodological advances in geospatial analysis as well as an increase in the number of derived products such as, e.g., land cover classifications. While this opens up new opportunities for geospatial applications, the vast amounts of geodata also challenge studies in terms of selection, filtering, harmonizing, processing and interpretation. To fully utilize the opportunities that the increasing amount of geodata offers while tackling its challenges, new and innovative methodologies for different geospatial applications are needed. This cumulative dissertation is a contribution to this goal in the fields of environmental science and geoarchaeology. It presents three novel and experimental approaches on how to effectively utilize a large variety and/or long time series of geodatasets and analyse them sensibly within the framework of a specific research question.

The first approach explores the possibilities and limitations of assessing complex aeolian dune field morphology and evolution using synthetic aperture radar (SAR) satellite data. This study relies on the Sentinel-1 mission, which acquires data in volumes of approximately 600 gigabytes per day. To analyse this continuous stream of geospatial big data, the study examines the key interaction mechanism between C-Band radar and sand dunes and introduces a visual pattern extraction method based on continuous wavelet transfer. This novel method is applied to the Western Mongolian dune field Bor Khyar. The results give new insights into the temporal and spatial dynamics of dune scales and their response to aeolian activity, revealing local differences as well as inter- and intra-annual variations in the dune morphology.

The second approach is a methodological contribution to the field of archaeological predictive modelling. The main challenge of this study is the extraction of a thematic pattern from a small sample of 23 available Upper Palaeolithic sites in Lower Austria. This is achieved using a novel approach combining a classical deductive method with the capabilities of machine learning. This way, ten spatial predictors representing morphological, hydrological, and sedimentological factors of the paleo-environment are analysed for optimal, viable, and non-viable value ranges and combined mathematically. The resulting predictive model reveals several spatial dynamics of site probability and shows high compliance with known sites in the study area.

In stark contrast to this study, which is challenged by the small number of available sites, the third study conducts geoarchaeological pattern extraction based on a substantially bigger dataset of close to 4200 European Upper and Final Palaeolithic sites. The main aim of this study is to explore whether

the site distribution is representative of human distribution in the paleo-landscape or if sampling biases obscure this information. To this goal, eight Pan-European geodatasets representing both settlement-relevant factors of the paleo-environment and discovery-relevant biases of the modern to contemporary landscape are analysed using a combination of geospatial and geostatistical methods. The results show that the actual distribution of sites seems to be most strongly influenced by sampling biases. The influence of the settlement factor, however, is still significant when comparing site subsets from different regions, different Upper Palaeolithic periods, and, especially, between open-air and cave sites. The implications of this study are substantial for geoarchaeological approaches, as the sampling bias is often overlooked or underestimated as a factor actively influencing the distribution of known sites.

All three approaches present a novel methodological approach in their respective field of study and outline a workflow that can be adapted and built on. For the availability to a broader audience, all studies are published as open access. In addition, the results of both geoarchaeological approaches are distributed as open data in universally usable geodata formats. As such, they can serve as foundation and inspiration for many future studies that utilize geospatial big data for environmental and geoarchaeological research.

# Zusammenfassung

In den letzten Jahrzehnten ist die Menge frei verfügbarer Geodaten exponentiell gestiegen. Dies ist zu einem großen Teil auf den technologischen Fortschritt in der Satellitenfernerkundung zurückzuführen, die eine weltweite Abdeckung durch eine Vielzahl von Sensoren auf einer sub-täglichen Basis ermöglicht. Darüber hinaus haben die bessere Zugänglichkeit und eine wachsende Nutzerbasis methodische Fortschritte bei der Geodatenanalyse sowie eine Zunahme der Zahl abgeleiteter Produkte wie z. B. Klassifizierungen der Bodenbedeckung begünstigt. Dies eröffnet zwar neue Möglichkeiten für raumbezogene Anwendungen, die riesigen Mengen an Geodaten stellen jedoch auch eine Herausforderung für Studien in Bezug auf Auswahl, Filterung, Harmonisierung, Verarbeitung und Interpretation dar. Um die Möglichkeiten, die die zunehmende Menge an Geodaten bietet, voll auszuschöpfen und gleichzeitig die damit verbundenen Herausforderungen zu bewältigen, werden neue und innovative Methoden für verschiedene Geodatenanwendungen benötigt. Diese kumulative Dissertation ist ein Beitrag zu diesem Ziel in den Bereichen Umweltwissenschaften und Geoarchäologie. Sie stellt über drei neuartige und experimentelle Ansätze vor, wie man eine große Vielfalt und/oder lange Zeitreihen von Geodatensätzen effektiv nutzen und im Rahmen einer spezifischen Forschungsfrage sinnvoll auswerten kann.

Der erste Ansatz untersucht die Möglichkeiten und Limitierungen der Erfassung von Morphologie und Morphodynamik komplexer äolischer Dünenfelder anhand von Radarsatellitendaten. Diese Studie stützt sich auf die Sentinel-1-Mission, die Daten in einer Größenordnung von etwa 600 Gigabyte pro Tag sammelt. Zur Analyse dieses kontinuierlichen Stroms von Geodaten werden in der Studie die wichtigsten Interaktionsmechanismen zwischen C-Band-Radar und Sanddünen untersucht und eine Methode zur visuellen Musterextraktion auf der Grundlage eines kontinuierlichen Wavelet-Transfers eingeführt. Diese neuartige Methode wird auf das westmongolische Dünenfeld Bor Khyar angewendet. Die Ergebnisse geben neue Einblicke in die zeitliche und räumliche Dynamik der äolischen Formen auf unterschiedlicher Skalenebene und ihre Reaktion auf äolische Aktivitäten, indem sie lokale Unterschiede sowie inter- und intra-jährliche Variationen in der Dünenmorphologie aufzeigen.

Der zweite Ansatz ist ein methodischer Beitrag im geoarchäologischen Fachbereich des *predictive modelling*. Die größte Herausforderung dieser Studie ist die Extraktion eines thematischen Musters aus einer kleinen Stichprobe von 23 verfügbaren oberpaläolithischen Fundstellen in Niederösterreich. Dies wird durch einen neuartigen Ansatz erreicht, der eine klassische deduktive Methode mit den Möglichkeiten des maschinellen Lernens kombiniert. Auf diese Weise werden zehn räumliche Prädiktoren, die morphologische, hydrologische und sedimentologische Faktoren der Paläo-Umgebung repräsentieren, auf optimale, tragbare und nicht tragbare Wertebereiche untersucht und

mathematisch kombiniert. Das daraus resultierende Vorhersagemodell beleuchtet verschiedene räumliche Dynamiken der Standortwahrscheinlichkeit und weist eine hohe Übereinstimmung mit bekannten Standorten im Untersuchungsgebiet auf.

Im Gegensatz zu dieser Studie, die primär durch die geringe Anzahl verfügbarer Fundstellen limitiert ist, wird in der dritten Studie eine geoarchäologische Musterextraktion auf der Grundlage eines wesentlich größeren Datensatzes von fast 4200 europäischen jung- und spätpaläolithischen Fundstellen durchgeführt. Das Hauptziel dieser Studie ist es, zu untersuchen, ob die Verteilung der Fundstellen repräsentativ für die Verteilung der Menschen in der Paläolandschaft ist oder ob Stichprobenverzerrungen diese Informationen überlagern. Zu diesem Zweck werden acht paneuropäische Geodatensätze, die sowohl siedlungsrelevante Faktoren der Paläoumwelt als auch entdeckungsrelevante Verzerrungen der modernen bis zeitgenössischen Landschaft repräsentieren, mit einer Kombination aus räumlichen und geostatistischen Methoden analysiert. Die Ergebnisse zeigen, dass die tatsächliche Verteilung der Fundstellen am stärksten von Stichprobenverzerrungen beeinflusst zu sein scheint. Der Einfluss des Siedlungsfaktors ist jedoch immer noch signifikant beim Vergleich zwischen Teilmengen von Fundstellen aus verschiedenen Regionen, verschiedenen oberpaläolithischen Perioden und insbesondere zwischen Freiland- und Höhlenfundstellen. Die Auswirkungen dieser Studie sind für zukünftige geoarchäologische Ansätze von großer Bedeutung, da der Stichprobenfehler als Faktor, der die Verteilung der bekannten Fundstellen aktiv beeinflusst, oft übersehen oder unterschätzt wird.

Alle drei Ansätze stellen einen neuartigen methodischen Ansatz in ihrem jeweiligen Fachgebiet dar und vermitteln einen klar definierten Arbeitsablauf, der in zukünftigen Studien angepasst und weiter ausgebaut werden kann. Um einem breiten Publikum zugänglich zu sein, sind alle Studien frei verfügbar als *Open Access* veröffentlicht. Darüber hinaus stehen die Ergebnisse der zwei geoarchäologischen Ansätze als *Open Data* in universell nutzbaren Geodatenformaten zum Download bereit. Dadurch können sie als Grundlage und Inspiration für viele künftige Studien dienen, die *geospatial big data* für Umwelt- und geoarchäologische Forschungsansätze nutzen.

# Table of contents

# List of figures

# List of tables

# 1 Introduction

## 1.1 Research framework and outline

Environmental science and geoarchaeology were amongst the first disciplines to implement geographical information systems (GIS) for the processing of spatial data (see e.g. Allen et al. 1990; Blumberg 1998). While both shared the enthusiasm for aerial imagery as early as in the 1960s, a broad implementation was only reached in the 1980s with the introduction of commercially available GIS software. In this timeframe, GIS applications in geoarchaeology focussed mostly on artefact inventory and distribution and predicting locations of yet undiscovered sites (González-Tennant 2016). GIS for environmental approaches in these early stages was primarily used by environmental agencies and forestry companies, managing their spatial inventories with administrative data and aerial imagery (Goodchild 2003).

The 1980s also marked the beginning of the continuous mapping of the earth's surface from space, starting with the Thematic Mapper (TM) class of Landsat sensors by National Aeronautics and Space Administration (NASA) (Markham et al. 2004; Senthil Kumar et al. 2013). Since then, numerous additional earth observation missions have been launched (Rast and Painter 2019), and both the public and private sector have compiled geospatial datasets of unprecedented sizes. While the public sector primarily collects geospatial data that serves an administrative purpose (Lansley et al. 2017), the private sector monetarises geospatial information by e.g. launching satellites and selling the imagery (Fu et al. 2019). The term geospatial data describes digital information with a spatial reference such as coordinates. While access to a subset of geospatial data is reserved for certain institutions or locked behind a paywall, a clear trend towards open access can be observed in recent decades (Mobasheri et al. 2020). Most of these open geodata are distributed in a plethora of different formats via a vast field of platforms. However, some effort has been made in order to harmonize geodata within spatial data infrastructures, e.g., within the European INSPIRE directive (Vancauwenberghe and van Loenen 2018; Minghini et al. 2021).

The first approaches that utilized this new source of earth surface information in geoarchaeology were mostly focussing on the manual extraction of archaeological features (Giardino 2011; Leisz 2013). In environmental science, early implementations of Landsat imagery primarily aimed at mapping the earth's surface by e.g. quantifying land surface vegetation using simple vegetation indices (Goward and Williams 1997). Since then, in order to cope with the increasing volumes of geospatial data, the methodological agendas in environmental science and geoarchaeology have shifted from labour-intensive manual procedures towards supervised and automated processes, in some cases assisted by machine learning algorithms (McCoy and Ladefoged 2009; Gibert et al. 2018; Opitz and Herrmann

2018). In addition, as datasets become too big and methodologies too complex to be handled by regular local hardware, geospatial processing is outsourced towards powerful web-based platforms. The most popular platform for remote sensing applications is the Google Earth Engine (GEE) (Amani et al. 2020), which promises “planetary-scale geoprocessing for everyone” (Gorelick et al. 2017).

The large supply of openly accessible geodata and the wide field of available methods result in an unprecedented number of opportunities for local, regional, national, and international environmental and geoarchaeological studies. Researchers who want to seize these opportunities, however, are faced with entirely new challenges in terms of data selection, handling, processing, and interpretation, each requiring an informed decision and thus setting new requirements for trained professionals. These challenges are addressed in the research question of this cumulative dissertation:

*How can geospatial big data be utilized effectively in the framework of environmental and geoarchaeological research questions*?

This research question was the common denominator of all projects carried out during the authors three-year doctoral programme in the Chair of Physical Geography and Geoecology, Department of Geography, RWTH Aachen University. During this programme, the author published three studies in international peer-reviewed journals, which form the main body of this dissertation thesis. The first study presented in this thesis is an exploratory approach in the field of environmental science, assessing the possibilities and limitations of dune field observation and monitoring with radar imagery and applying the approach to a study area in Western Mongolia (Boemke et al. 2023b). The second study presents a geoarchaeological approach that combines a classical deductive method with machine learning to create a predictive model for Upper Palaeolithic sites from very limited archaeological evidence in Lower Austria (Boemke et al. 2022). The third study aims at differentiating between the influence of settlement-relevant factors of the paleo-landscape and modern to contemporary sampling biases on the Upper and Final Palaeolithic record based on a large Pan-European archaeological dataset (Boemke et al. 2023a).

The main commonplace between the three studies is that they all present interdisciplinary, novel, and experimental approaches on how to implement and analyse a large variety and/or long time series of geospatial datasets within the framework of an environmental or geoarchaeological research question. In addition, all studies place their main emphasis on the methodology, focussing on reproducibility and defining a foundation for a continuation of research within the topic. This is additionally emphasised by the publication as open access and the distribution of both geoarchaeological datasets as open data. Another similarity between both studies in the field of geoarchaeology is the type of geospatial data that they are based on, as both use a large variety of morphological and sedimentological datasets to

approximate the paleo-environment. Both studies also aim at extracting spatial and thematic patterns from the used environmental datasets via point-based archaeological site databases. In contrast, the first study on dune assessment extracts multi-temporal patterns via profiles. Figure 1 shows a visual representation of the different forms of pattern extraction that were conducted in the three studies.

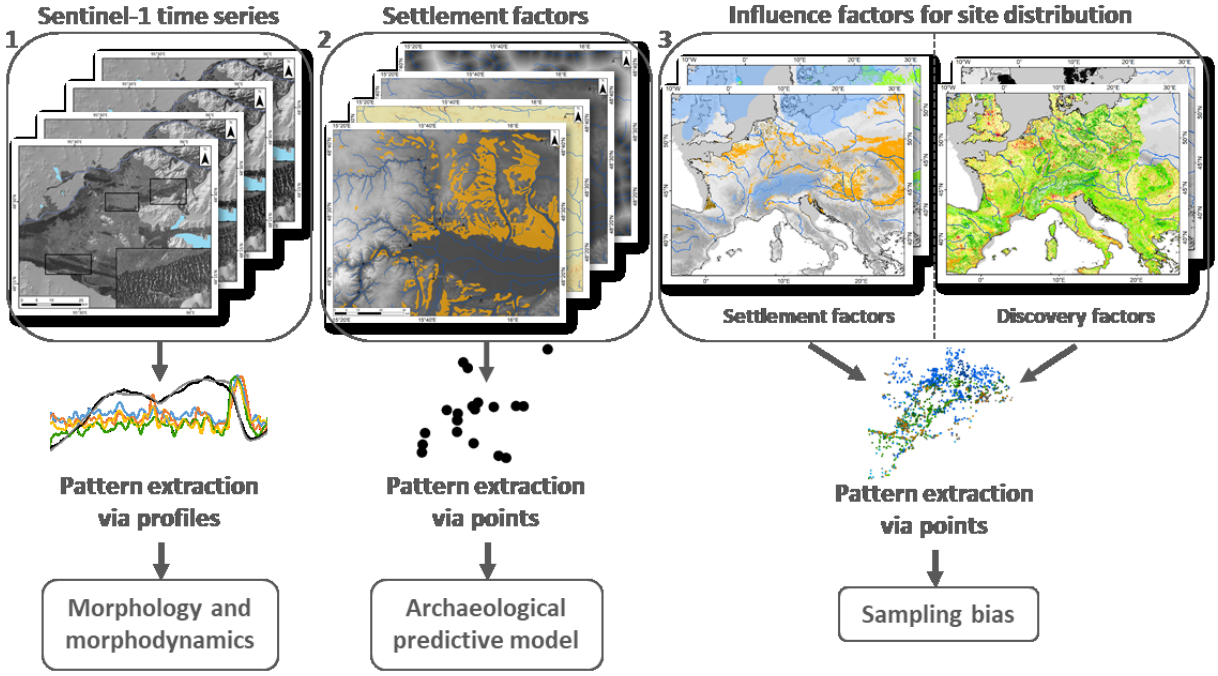



*Figure 1: Comparison of the main data types and the form of pattern extraction applied in the three different studies. Please refer to the respective study for a detailed description.*

One of the main differences between the two geoarchaeological studies, however, is the purpose of the pattern extraction. Whereas one study relies on factors depicting the paleo-landscape to predict the location of unknown sites, the other study additionally includes discovery-relevant factors of the modern to contemporary landscape to assess the potential sampling bias. Another main difference can be found in the size and location of the different study areas. As the second study focusses on the limited evidence of Upper Palaeolithic sites in Lower Austria, the study area is also limited to this part of central Europe. The third study is based on a Pan-European archaeological dataset and covers a larger study area accordingly. In stark contrast, the methodology developed in the first study was applied to a complex dune field in Western Mongolia. This deviating study area was chosen due to previous studies on aeolian sediments conducted in the area by the department (Grunert and Lehmkuhl 2004; Klinge et al. 2017). This environmental study also differs from the other studies as it attempts pattern extraction from a vast time series of a single dataset while the geoarchaeological studies are both based on a large collection of temporally one-dimensional datasets.

## 1.2 Geospatial big data – challenges and opportunities

The term *big data* emerged from the field of data science in the 1990s and was used to describe massive and unstructured datasets that posed a challenge to traditional processing techniques (Chen et al. 2014). The use of the term was ambiguous until Laney (2001) used volume, velocity, and variety, known today as 3Vs, to characterize it. Using this concept, data is considered 'big data' when it reaches a size in terms of storage capacity (volume), speed in terms of generation (velocity) and/or diversity in terms of data types and sources (variety) where traditional methods fail to provide effective solutions. As such, big data requires special means in terms of storage, transmission, curation, analysis, and visualization (Chen and Zhang 2014). While numerous researchers suggested additional Vs like veracity, variability, and value (e.g. Marr 2015; ur Rehman et al. 2016), the base concept remains unchanged as of today.

*Geodata* or *geospatial data* describes nothing else than data with a spatial reference. This can be something as simple as a coordinate or an address. The arguable phrase *80% of data is geographic* indicates that a majority of worldwide data can be georeferenced, which makes geospatial big data the logical result of the emergence of big data (Li et al. 2016). In addition to these regular datasets with secondary spatial information, primarily spatial datasets with secondary informational attributes can also be considered geospatial big data when they fulfil one or more of the 3V-characteristics. Lansley et al. (2017) differentiate between three categories of geospatial big data based on their origin:

- **Human-sourced data:** This category includes actively generated primary geodata such as the biggest geospatial crowd-sourced project Open Street Map (OSM, Haklay and Weber 2008) but also passively generated secondary geodata such as georeferenced social-media posts and automated mobile device tracking via integrated GPS.
- **Process-mediated data:** This category includes both administrative and commercial data. The common denominator is, that both were not primarily designed to represent the real world but rather to support administrative and commercial functions. Such functions include, e.g., assessment of taxation, electoral allocation, or personalized advertisement.
- **Machine-generated data:** This category describes data that is continuously generated by different types of sensors. The biggest contributor to this category is satellite remote sensing, generating terabytes of readings from the earth's surface and atmosphere per day (Ma et al. 2015; Soille et al. 2018). Other geodata that would fall into this category are, e.g., climate sensors, river gauges, or even traffic measuring sensors.

While this is most certainly not the only way to categorize geospatial big data, it illustrates how diverse the field of geodata is and how well it fits the concept of big data. However, one category that plays an important role in this dissertation is overlooked in these categories:

- **Science-generated data:** This fourth category, proposed by the author, describes geospatial datasets that were created for a scientific purpose and represent real-world objects or areas. It includes datasets that were directly created for a scientific purpose (e.g., through surveys) or where the scientifically relevant information was derived from other datasets (such as administrative or machine-generated data). Examples of this category are land use/land cover classifications (LULC), archaeological inventories, and thematic spatial data such as digital maps of geology or soil. While some of these datasets don't reach the required volume to be considered big data, the scientific community generates such a large variety of datasets at a high velocity that at least two of the Vs are accounted for.

Each of these categories of geospatial big data opens up a wide range of new opportunities for researchers to gain a better understanding of, e.g., urban mobility, environmental processes, or human-environment interactions. However, due to the mentioned characteristics of geospatial big data, extracting the relevant information from them is not an easy task. This is reflected in the term *geodata mining*, which was introduced by Miller (2007) and is still used today to describe the challenging process of extracting patterns from geospatial big data (Pei et al. 2020). Depending on the type of geospatial big data and the research question, the challenges that this process poses can differ fundamentally. This dissertation aims to extract patterns from science-generated geospatial big data to answer research questions in the fields of environmental science and geoarchaeology. These types of datasets are primarily characterized by a high variety and results in these fields of study often leave a lot of room for discussion. As such, challenges connected to pattern extraction in this dissertation are primarily connected to data selection, handling, processing, and interpretation. While these are the basic challenges connected to all types of data science, their application for geospatial big data can be summarized as follows:

1. **The challenge of data selection:** Due to the vast field of available geodatasets, the decision of which ones to include in a study is of fundamental importance. Especially when there are multiple geodatasets representing similar information. A good example for this are international and global LULC datasets of which García-Álvarez and Nanu (2022) alone list more than 100. When deciding which one of these to use, one has to consider multiple factors such as the spatial resolution, thematic resolution/classification scheme, source dataset/classification method, accuracy as well as the timeframe that it represents. The

decision should, however, not be based solely on the highest spatial and thematic resolution and accuracy but, most importantly, on the information value that the dataset contains for the individual research question. This might differ significantly, even between studies in the same field of research where, e.g., one primarily requires a sharp differentiation between different types of forest, and the other needs an accurate differentiation between vegetated, bare, and built-up areas. As such, the main question to ask when selecting a fitting geodataset can be broken down to: *Which geodatasets contain the information needed for the study at a sufficient spatial and thematic resolution and accuracy*?

2. **The challenge of data handling:** As already mentioned, geodatasets come in a plethora of different formats and resolutions. In order to successfully extract the needed information, the handling of the datasets should include filtering and harmonization. While filtering describes the selection and extraction of only relevant subsets from geodatasets and the aim of harmonization is to achieve comparability between different datasets, both processes are interdependent and closely related to the specific research question. To stick with the example of LULC datasets, filtering might include selecting and extracting only relevant classes and time frames, whereas harmonizing would require rescaling and reclassification into a common denominator in terms of thematic units. In some cases, this might even require a reduction of spatial or thematic resolution in favour of improved comparability. When handling geospatial big data, the main question should therefore be: *How can a certain collection of geodatasets be filtered and harmonized for optimal informational content and comparability at a minimal loss of relevant information?*
3. **The challenge of data processing:** The main aim of processing geospatial big data is the extraction of humanly comprehensible data, also called small data. The wide field of available and well-documented methodologies for this task makes selecting a suitable method a difficult decision. This is equally true for geospatial and geostatistical processing as no consensus has been reached concerning the right set of tools for specific tasks (Gibert et al. 2010). As such, researchers have to make an informed decision for each approach individually based on similar studies, scientific debates and/or their own working experience. This emphasizes the importance of scientific exchange between different institutions and fields of research. After deciding which method to use, it is equally relevant to evaluate strengths and weaknesses of said method for each application and to consider them in the interpretation of results. In addition, as different processing of the same dataset can provide contradictory conclusions, a very important aspect of the processing is the open communication of the used methodology for reproducibility and open discussion (Gibert et al. 2018). As such, the main question related

to the processing of geospatial big data should be: *Which methodological tools are suitable for extracting the wanted information from a certain collection of geodatasets, what are their strengths and weaknesses, and how can they be communicated adequately for optimal reproducibility and open discussion?*

4. **The challenge of result interpretation:** An appropriate interpretation of geospatial big data processing results requires the interpreter to understand and master the challenges of data selection, handling, and processing. From the data selection process, it is relevant to consider to which degree the chosen geospatial datasets are abstracted from the research-question-relevant real-life objects that they represent. The degree of filtering and harmonizing during the handling has to be additionally considered to adequately address the question of how representative the geodatasets are of their real-life equivalents. A comprehensive assessment of the chosen methodological approach, including related strengths and weaknesses, is the last prerequisite for a profound interpretation of geospatial big data processing results. In this regard, it is important to consider how possible cross-correlations between chosen geodatasets might influence the results to avoid confusion of correlation and causality. The results additionally have to be assessed for significance to separate between signal and noise. To ensure that these requirements are met, researchers have to address the following question: *Which research questions can be adequately answered based on the combination of geodatasets and methods, how well can meaningful results be differentiated from noise, and how do they translate into the real world?*

Only if these challenges and related questions are adequately addressed, the required information can be successfully extracted with minimal potential biases. Then, however, the extraction of patterns not only from single spatial big datasets but even from combinations and long time series of geodatasets is possible.

## 1.3 State of research

As already mentioned in the outline of this dissertation, the recent decades were characterized by an unprecedented increase in open source software and open access spatial data. This is not only true for science-generated geodata but also for governmental and even commercial spatial data (Mobasheri et al. 2020). This trend has been widely recognized by researchers all over the scientific spectrum, leading to a plethora of new studies utilizing geospatial big data. This section, therefore, aims at giving an overview of recent approaches and trends in pattern extraction from geospatial big data in the

fields of environmental science and geoarchaeology. As environmental science is a diverse field of research, this section focuses on dune-related studies.

A source of geospatial big data that has a long history of application in both environmental science and geoarchaeology is satellite remote sensing data. While worldwide imagery of the earth's surface has been publicly available since the 1980s, the recent decade marks an explosive increase in the number of active earth observation missions and their amount of stored volume (Soille et al. 2018, see Figure 2). In this timeframe, the focus of researchers has shifted from satellites of the NASA Landsat mission towards the Sentinel mission by the European Space Agency (ESA) in favour of the increased spatial and temporal resolution (Zhao et al. 2022). In addition, commercial stakeholders like the companies Planet Labs or Spire Global are joining the once governmentally dominated field of satellite remote sensing and setting new standards in terms of spatial and temporal resolution, although locked behind a paywall (Roy et al. 2021). One of the major advances towards an increased accessibility and easier handling of remotely sensed geospatial big data are web-based processing platforms, the most popular being the GEE (Gorelick et al. 2017). The use of this platform is widely recognized in both environmental science and geoarchaeology and has shown a steep increase in the recent decade (Amani et al. 2020; Tamiminia et al. 2020; Herndon et al. 2023).

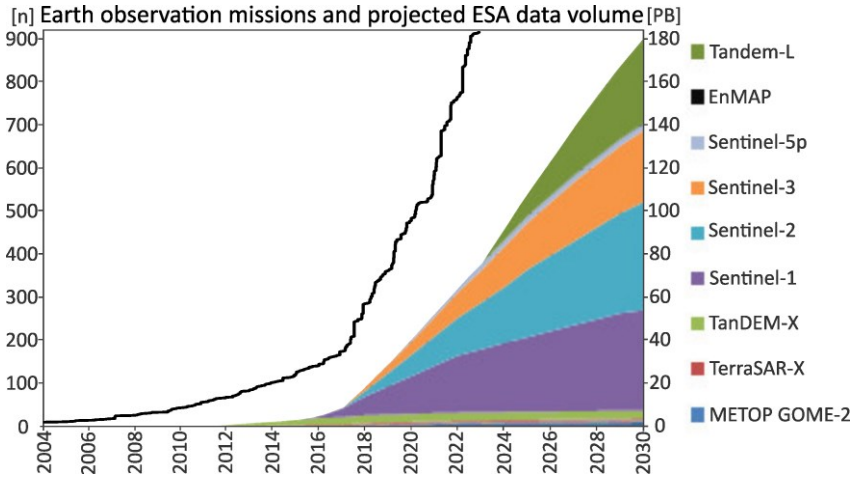



*Figure 2: The recent development of active earth observation satellites (black line, left vertical axis, Grimwood 2022) and the projected data volume of the nine European Space Agency (ESA) earth observation satellites (coloured bars, right vertical axis, DLR 2018).*

Studies in the field of aeolian dune detection and monitoring have historically been limited in scale due to the remoteness and difficult accessibility of the study areas. This drastically changed with the launch of the first earth observation satellites, allowing worldwide studies of dune fields (McKee 1979). Until today, optical satellite imagery remains the main source of geospatial big data in this field (Hugenholtz et al. 2012). While Landsat imagery is still a popular choice due to the long time series of comparable data, recent years have shown a trend towards missions with higher spatial resolution

such as Sentinel-2 or SPOT (Zheng et al. 2022). As the spatial resolution is of high importance for, e.g., the classification of dune types and the assessment of dune activity, many recent studies also use platforms with compiled high-resolution imagery such as Google Earth or ESRI imagery basemap (see, e.g., Pradhan et al. 2018; Herzog et al. 2021). Digital elevation models (DEMs) present another popular source of geospatial big data in dune studies (Zheng et al. 2022). Their main advantage over optical imagery is the possibility to assess dune morphology directly. The Shuttle Radar Topography Mission (SRTM, Farr et al. 2007) achieved the first globally available DEM, which since then has been used in numerous studies to assess dune morphology (see, e.g., Blumberg 2006; Effat et al. 2011). While newer global open access DEMs show no improvements in terms of horizontal spatial resolution, comparative studies have proven that they outperform the SRTM in terms of vertical accuracy (Bubenzer and Bolten 2008; Hugenholtz and Barchyn 2010), allowing the assessment of even smaller dunes (Shumack et al. 2020). For significant improvements in spatial resolution, recent studies have deployed unmanned aerial vehicles (UAVs) to retrieve optical imagery and DEMs (Solazzo et al. 2018; Luo et al. 2020; Fabbri et al. 2021). Due to the labour-intensive retrieval of UAV-Data, however, these studies are limited to smaller study areas. Another recent development in the field is the use of synthetic aperture radar (SAR) data. While early studies saw great potential in this type of data (Blumberg 1998; Qong 2000), the requirements for continuous monitoring in terms of data quality as well as spatial and temporal resolution were only met in 2014 with the launch of Sentinel-1. Since then, Sentinel-1 SAR has been applied successfully to identify dunes (Havivi et al. 2018; Delgado Blasco et al. 2020) and monitor their activity (Mahmoud et al. 2020; Manzoni et al. 2021).

From a methodological perspective, pattern extraction from geospatial big data for aeolian dunes is a very diverse field. While many studies that use optical imagery to monitor dune migration still rely on labour-intensive manual mapping of dune features (Hamdan et al. 2016; Yang et al. 2019; Dörwald et al. 2023), DEM-based approaches allow for an easier automated derivation of patterns and mobility (Cazenave et al. 2013; Dong 2015; Shumack et al. 2020). Therefore, high hopes are associated with the anticipated public release of high-resolution global DEMs such as, e.g., TanDEM-X (Zink et al. 2014). Another promising recent development is the improvement of the temporal resolution of dune-related geospatial big datasets. Many studies in the past were limited to mono-temporal data for, e.g., the classification of dune types (e.g., Dong et al. 2013; White et al. 2015) and studies on dune migration often only use less than a handful of images to estimate long-term mean migration rates (e.g. Hamdan et al. 2016; Dörwald et al. 2023). The currently possible daily to sub-daily return period of combined satellite sensors, however, allow insights into single aeolian events. This allows for a deeper understanding of the frequency and magnitude of dune-forming aeolian processes, which researchers have only been able to address in time-consuming field studies to date (see e.g. Poortinga et al. 2015).

In geoarchaeology, the most frequently implemented remote sensing applications are the identification of archaeological sites and features from optical imagery (Herndon et al. 2023). Howey et al. (2020) explain this with the similarity to traditional archaeological surveying practices. While some authors explore automated processes for feature detection (Liss et al. 2017; Soroush et al. 2020; Altaweel et al. 2022), most approaches still rely on visual interpretation (Dana Negula et al. 2020; Abate et al. 2022; Lasaponara et al. 2022). This highlights the importance of expert knowledge in this scientific field, which is often emphasised by researchers (see, e.g., Casana 2014). DEMs are another source of geospatial big data that has been successfully applied for this kind of analysis (Freeland et al. 2016; Bonhage et al. 2021; Štular et al. 2021). As the spatial resolution is the main limiting factor of both manual and automated feature extraction, the implementation of high-resolution UAV imagery and DEMs marks one of the most promising recent trends (Orengo and Garcia-Molsosa 2019; Agapiou et al. 2021). All feature extraction approaches, however, can only be applied to identify features that are close enough to the surface to directly impact the morphology and/or vegetation. They also rely on ground-truthing for the validation of results (Herndon et al. 2023). This, however, makes them an ideal tool for fieldwork planning and pre-analysis.

Another application of geospatial big data in geoarchaeology is cultural heritage site assessment. The main aim of this practice is to monitor anthropogenic and environmental changes and events at heritage sites to assess possible damages and risks. The main anthropogenic risk for site preservation is land use change and especially urbanization, which has been monitored around sites in numerous studies (Noronha Vaz et al. 2012; Yu et al. 2016; Agapiou 2017, 2021; Rayne et al. 2020). The main environmental risks to cultural heritage sites, on the other hand, are natural hazards. While riverine flooding is one of the most researched natural risks to heritage conservation (see e.g. El-Behaedi and Ghoneim 2018; Fattore et al. 2021; Elfadaly et al. 2022a), other researched hazards include forest fires (Mallinis et al. 2016; Salazar et al. 2021) coastal flooding (Reeder-Myers 2015) and heavy precipitation events (Carmichael et al. 2023). Moreno et al. (2022) and Carmichael et al. (2023) expect natural hazard-related risks to increase in the future due to the effects of climate change. This emphasizes the increasing relevance of cultural heritage monitoring and preservation in the future.

Both of these geoarchaeological practices of geospatial big data application, however, are rarely applied in prehistoric archaeology. This is due to the fact that they rely on surface visibility, while prehistoric features and sites seldom leave a visible morphological footprint and are often covered by thick layers of sediment (Campana and Piro 2008; Alday et al. 2018). Geospatial big data approaches in prehistoric archaeology, therefore, mostly focus on environmental reconstruction or predictive modelling. Practices in environmental reconstruction aim at recreating past landscape features and climatic factors that were of fundamental importance to prehistoric settlement choice or mobility.

Individual approaches differ widely in dimension and scope, ranging from the reconstruction of single landscape features based on optical imagery or radar data (Orengo and Petrie 2017; Brandolini et al. 2021; Elfadaly et al. 2022b; Elfadaly et al. 2020) to superregional studies using climatic and morphological big data to estimate human existence potential and demography (Maier et al. 2016; Maier et al. 2020; Klein et al. 2021; Maier et al. 2022). Predictive modelling, on the other hand, uses geospatial big data as an abstraction of the paleo-environment to anticipate the location of yet undiscovered sites. This prediction is either based on expert knowledge on past human preferences (deductive) or on the distribution of known sites (inductive) (Verhagen and Whitley 2012). Malaperdas and Zacharias (2019) as well as Howey et al. (2020) are recent examples for methodological advances within this field. Especially the high variety of datasets used in these studies make them classifiable as geospatial big data approaches.

While all studies presented by now implement geospatial big data for some kind of pattern extraction, it is important to also consider geoarchaeology as a source of big data. In addition to the recent positive trend of dataset publication as supplemental material along scientific articles, substantial effort has been made in order to compile and distribute archaeological information within large databases. McCoy (2017) classifies these databases into *Data Repositories, Location Indexes, Radiocarbon Databases, Project Websites, and Academic Sources*. *Data Repositories* describe collections of various archaeological data types, such as, e.g., articles, spatial datasets, images etc., within a browsable and searchable web environment (e.g., the Digital Archaeological Record (tDAR) for the United States (McManamon et al. 2017) or ARIADNE for Europe (Meghini et al. 2017)). Only a subset of these datasets, however, has a geospatial component. The categories *Location Indexes* and *Radiocarbon Databases*, on the other hand, contain primarily spatial information with complementary information about, e.g., cultural attribution or measured age. Such databases present a formidable source of archaeological geospatial big data for, e.g., predictive modelling or geostatistical approaches. Examples of such databases are the Digital Index of North American Archaeology (DINAA) with close to 900,000 entries (Kansa et al. 2018), or the Radiocarbon Palaeolithic Europe database with more than 13,000 georeferenced and radiometrically dated sites (Vermeersch 2020). While the categories *Project Websites* and *Academic Sources* mostly contain regionally limited or specific geodata, some exemptions, like the English Landscape and Identities Project (Cooper and Green 2016) or the Collaborative Research Centre 806 (Willmes 2016), compile larger datasets that qualify as big data based on their overall volume. Many of the presented databases are expanded continuously, allowing future geoarchaeological big data studies to build on even bigger datasets.

# 2 Materials and Methods

## 2.1 Data selection and preprocessing

The first step in all three approaches presented in this dissertation was the careful selection of geodatasets required to answer each research question adequately. As mentioned in section 1.2, the main challenge of this task is to identify the geodatasets that are best suited to represent the object of investigation. The second step was the preprocessing, which aims at preparing the geodatasets for geospatial and geostatistical analyses through filtering and harmonizing. In section 1.2, the main aim of this step is defined as extracting the informational content and ensuring optimal comparability between datasets. The following section briefly summarizes which measures were taken in the studies presented in this dissertation to ensure that these challenges of geospatial big data selection and preprocessing were successfully addressed. For an overview of all used geospatial datasets, their reference, the object of investigation that they represent, and the preprocessing steps that were conducted in order to extract relevant information and harmonize the datasets for optimal comparability, see Table 1. For more detailed information on data selection and preprocessing, please refer to the respective section in the studies themselves.

For the first study on dune assessment, data selection was not an issue as the research question directly addresses the Sentinel-1 archive for the assessment of dune morphology. However, as Sentinel-1 does not contain elevation information but rather the roughness of ground material (Williams and Greeley 2004) and incidence angle of electromagnetic radiation (Blumberg 1998; Delgado Blasco et al. 2020), it can only be seen as an indirect representation of dune morphology. The filtering process included identifying informative polarization settings and time slices from the vast time series of Sentinel-1 imagery. Based on an extensive analysis of wind patterns in the study area via the ERA5 reanalysis dataset (C3S 2022, see Figure 4), four scenes from two different wind regimes were selected. As such, the data indirectly represents dune morphology during the north-westerly dominated winter and the south-easterly dominated summer. As Sentinel-1 SAR imagery is temporally heterogeneous and contains terrain- and pixel-based artefacts (Truckenbrodt et al. 2019), the four scenes were additionally harmonized for optimal comparability. This process included multi-temporal speckle filtering, border noise correction, and radiometric terrain normalization based on the preprocessing approach proposed by Mullissa et al. (2021).

*Table 1: Overview of the different datasets utilized in the three studies of this dissertation including the name of the dataset, the real-world research object that they represent, the filtering and harmonization that was conducted and their reference.*

| *Study* | *Geodataset* | *Represents* | *Filtering* | *Harmonizing* | *Reference* |
|---|---|---|---|---|---|
| Sentinel-1 dune field assessment | Sentinel-1 | - Dune morphology | - Polarization (VV)<br>- Scene selection | - Border noise correction<br>- Speckle filtering<br>- Terrain normalization | Copernicus Sentinel data [2015-2021] |
| Upper Palaeolithic site probability in Lower Austria | DEM10 of Austria | - 5 Late Pleistocene terrain parameters | - Low-pass filter | | www.data.gv.at |
| | CCM21 | - 3 Late Pleistocene hydrology parameters | | | De Jager and Vogt 2007 |
| | Loess map | - Late Pleistocene environment and site preservation | | - Rasterization<br>- Rescaling to DEM resolution | Lehmkuhl et al. 2021 |
| Upper and Final Palaeolithic sampling bias in Europe | NASADEM | - 3 Late Pleistocene terrain parameters | | - Rasterization<br>- Rescaling | NASA JPL 2020 |
| | 1:5 million Geological map of Europe | - Late Pleistocene cave probability and availability of lithic raw material | - Extraction of certain stratigraphic units | - Uncertainty radius<br>- Rasterization<br>- Rescaling | Asch 2003 |
| | Quaternary glaciations | - Late Pleistocene availability of lithic raw material | - Extraction of penultimate accumulation area | - Uncertainty radius<br>- Rasterization<br>- Rescaling | Ehlers et al. 2011 |
| | Pleistocene aeolian deposits (topsoil) | - Late Pleistocene environment, site preservation and sampling bias | | - Uncertainty radius<br>- Rasterization<br>- Rescaling | Bertran et al. 2016 and 2021 |
| | Pleistocene aeolian deposits (geology and soil) | - Late Pleistocene environment, site preservation and sampling bias | | - Uncertainty radius<br>- Rasterization<br>- Rescaling | Lehmkuhl et al. 2021 |
| | Corine Land Cover | - Contemporary land use and anthropogenic impact | - Extracting different levels of anthropogenic impact | | land.copernicus.eu/paneuropean/corine-land-cover |
| | HYDE land use model | - Modern land use and anthropogenic impact | - Selecting only population density and built-up area | - Rescaling | Klein Goldewijk et al. 2017 |

In the second study on Upper Palaeolithic site probability in Lower Austria, the data selection aimed at compiling geospatial evidence and predictors for the settlement choice of past humans. For the evidence, a point-based dataset was compiled from known archaeological excavations and studies in

the area. As the study aims to create a predictive model for open-air sites only, further filtering of the site dataset was necessary. Therefore, cave sites and sites with missing or ambiguous spatial and chronological attribution were filtered out. Initial to the selection of predictive geodatasets, two regional experts aided in deductively identifying paleo-environmental features with an assumed influence on the settlement choice of past humans. This resulted in 10 predictors related to terrain, hydrology, and geology/sedimentology (see Table 5). These predictors were derived from three geospatial datasets, namely the 10m DEM of Austria (available at www.data.gv.at), the CCM21 European river dataset (De Jager and Vogt 2007), and the European Loess map by Lehmkuhl et al. (2021). As these geodatasets were created based on recent measurements and, e.g., the terrain has undergone both erosive as well as accumulative changes since the Late Pleistocene, they can only be considered an indirect representation of the paleo-environment. Filtering of the geodatasets included a smoothing of the DEM to remove anthropogenic impacts such as e.g. roads. For an optimal comparability, all parameterised predictors were rasterized and rescaled to the spatial resolution of the DEM. This harmonisation was essential for the implementation of the machine learning pattern extraction software MaxEnt, which was developed for modelling species distribution and environmental niches (Phillips and Dudík 2008).

The third study on sampling bias of the European Upper and Final Palaeolithic record aims to compare influences of past settlement choices and modern to contemporary likelihood of discovery based on the spatial distribution of known sites. As such, geospatial predictors were needed, not only for decision-influencing features of the paleo-environment (settlement factors), but also for discovery-relevant features of the modern to contemporary landscape (discovery factors). The archaeological evidence that was used to extract patterns from these factors was represented by a point dataset, combined from chrono-cultural site collections which were compiled within the framework of the cologne protocol (Schmidt et al. 2021b). As not all of these datasets contained the same auxiliary information per site, the dataset was filtered for the most important common denominators, namely coordinates, site type (cave or open air), and chrono-cultural attribution. Geospatial settlement factors were identified in the terrain, geology, and sedimentology, represented by six Pan-European geodatasets, while discovery factors were found in the modern to contemporary LULC, represented by two geodatasets. During the preprocessing, these geodatasets were filtered to extract the main information related to the influence on site distribution. For the geological datasets, this included the extraction of stratigraphic units, which influence cave formation and the availability of lithic raw material such as flint. The LULC datasets were filtered according to the intensity of anthropogenic impact and the possibility of deep soil intervention. Two different harmonization approaches were applied to account for inaccuracies of the influence factors as well as the site dataset and to ensure

optimal comparability in both the geostatistical and the MaxEnt analysis. For the geostatistical approach, an uncertainty radius of 500 meters around each site was applied in the spatial intersections with influence factors. In preparation for the MaxEnt software, all geospatial influence factors were rasterized and rescaled to the spatial resolution of the contemporary LULC dataset.

## 2.2 Geospatial and geostatistical analysis

The aim of geospatial big data analysis is to break it down into humanly comprehensible small data. As discussed in section 1.2, the main challenge of this process is finding a suitable methodological toolset to extract the information needed. As there might be multiple possible approaches for this, it is important to consider their strengths and weaknesses and communicate them openly for an assessment of the validity and the reproducibility of results. The following section, therefore, gives an insight into the chosen methodologies to extract patterns from the geospatial big datasets presented in 2.1. For an overview of these methods for pattern extraction, a short description, their aim, and the software that they were based on, see Table 2. For more detailed description of the methodologies, please refer to the respective sections in the studies themselves.

The main aim of the first study was to assess the capabilities of Sentinel-1 for multi-temporal dune morphology analysis. As this is a pilot study on the usage of raw Sentinel-1 ground range detected (GRD) data, it does not try to propose a fully developed best practice processing algorithm but rather presents possibilities for future studies. To reduce the geospatial big data of the Sentinel-1 time series into humanly comprehensible small data, a visual approach based on continuous wavelet transform (CWT) was applied. CWT was originally invented as a macroeconomic tool for the decomposition of economic time series (Ramsey and Lampart 1998; Aguiar-Conraria and Soares 2014). Simply put, it measures how fluctuations in a complex signal (e.g., 100 years of gross national product of a country) correlate to different frequencies (e.g., quarterly, yearly, decadal,…). As dune morphology in a cross section along the main wind direction also presents a complex signal, it is applicable for CWT (as shown in, e.g., Turki et al. 2021). In this case, however, the x-axis does not represent time (frequency) but distance (wavelength). From one time slice of Sentinel-1 data, CWT can thereby extract overlaying aeolian features with different wavelengths (e.g., underlying paleo-dunes, smaller recent dunes, and inter-dune ripples). A comparison of different time frames of Sentinel-1 data can show how these differently scaled features change through time and, thereby, how they respond to aeolian drivers. The main limiting factor of the application of CWT in this approach is that it cannot be applied to the whole three-dimensional dune field but only to selected two-dimensional cross sections. As such, the selection of representative cross-sections is of great importance. To this end, an extensive visual

analysis of the dune field was conducted beforehand based on optical imagery, elevation models, and wind data in order to select three areas of interest with different conditions in terms of aeolian activity, sediment characteristics, and sub-dune morphology. Through the application of CWT in these areas over a certain time frame, the complex multi-temporal surface information was broken down into visually interpretable small data. As such, the combined approach allowed for testing the suitability of Sentinel-1 for the assessment of dune morphology and evolution.

*Table 2: Overview of the different methods applied in the three studies of this dissertation including the name of the method, a short description, their main aim and the software that they are based on.*

| *Study* | *Method* | *Description* | *Main aim* | *Software* |
|---|---|---|---|---|
| Sentinel-1 dune field assessment | Continuous wavelet transfer | Visual pattern extraction method. Breaks down a complex signal into different frequencies/wavelengths. | Differentiating between aeolian forms at different scales. Visualize morphodynamics through change-detection. | MATLAB |
| | Profile analysis | Visual comparison of Sentinel-1, NASADEM and GLO30DEM profiles extracted from the dune field. | Preanalysis of fitting profile locations. Validation of the CWT results. | ArcGIS, version 10.7.1<br>Excel 2016 |
| Upper Palaeolithic site probability in Lower Austria | MaxEnt response curves | 2D-Visualization of the probability associated to the values of an environmental predictor. | Getting an inductive *opinion* on the statistical connection between environmental predictors and site probability. | MaxEnt, version 3.4.4 |
| | Deductive model | Assessing optimal, viable and unviable value ranges from the response curves and mathematically adding them up. | Assessing the plausibility of response curves and thereby the causality of environmental predictors and sites. Creating a spatial model from the result. | ArcGIS, version 10.7.1 |
| Upper and Final Palaeolithic sampling bias in Europe | Deviation from expected mean/share | Comparison between expected (mean/share in/of the habitable area) and observed (mean/share in/of the site location) value | Assessing conditions that are favourable or unfavourable for the settlement choice or discovery probability. | ArcGIS, version 10.7.1<br>Excel 2016 |
| | MaxEnt jackknife variable importance | Measurement of separability between the values associated to the presence and absence of sites of each single factor | Assessing the strength of the different settlement and discovery factors at predicting the presence/absence of sites. | MaxEnt, version 3.4.4 |
| | Additional geostatistical approaches | Statistical queries with the archaeological class (e.g. chrono-cultural attribution, cave/open air) as independent variable and factor values as dependent variables | Assessing possible cross-correlations between the factors and testing if the factors show significant differences between the archaeological classes | IBM SPSS Statistics |

The second study aimed to predict the probability of Upper Palaeolithic sites in Lower Austria. This study was primarily motivated by the limited evidence of sites from this time frame in this region, which is explained by the thick loess cover, hiding possible Late Pleistocene artefacts and making

random discovery less likely (Einwögerer et al. 2014). As already mentioned, archaeological predictive models are either based on expert knowledge of past human preferences (deductive) or on the distribution of known sites (inductive) (Verhagen and Whitley 2012). Due to the very limited evidence of only 23 sites, a purely inductive approach was not applicable. Instead, a combined approach with special emphasis on the causality between the paleo-landscape and human activity was developed. The first step in this approach was the extraction of a spatial pattern from the environmental predictors based on the site dataset. For this process, the software MaxEnt was used, which is a predictive tool based on the maximum entropy principle (Elith et al. 2011). The main advantage of this tool over more traditional approaches, such as logistic regression, is the fact that it utilizes presence-only data, which is what most archaeological datasets consist of (Wachtel et al. 2018). For each input predictor, MaxEnt calculates two probability densities (one from presence data and one from background data) and minimizes the relative entropy between them. The result of this pattern extraction is visualized in response curves (Merow et al. 2013, see Figure 13). In a deductive endeavour and based on regional archaeological expertise, these response curves were then evaluated on their thematic plausibility. The aim of this evaluation was to verify the causality between input predictors and site probability and correct response curves where the probability contradicted common archaeological assumptions. Based on the corrected curves, optimal, viable, and unviable value ranges of each predictor were assessed and mathematically combined into a simple additive model. This methodology underlines the importance of expert knowledge in evaluating automatically extracted patterns from geospatial big data. This is especially true for approaches with limited training data where an adequate representation of the statistical population cannot be safely assumed.

The main aim of the third study on the Upper and Final Palaeolithic sampling bias in Europe was to differentiate between two influences on the distribution of sites; the past settlement choice and the modern to contemporary discovery probability. As some of the geodatasets selected for this task could not be easily attributed to only one of these influence factors and cross-correlations between some of the geodatasets were assumed, extracted patterns were handled with caution. In order to achieve the best possible differentiation, multiple geospatial and geostatistical analyses were conducted in an attempt to extract humanly comprehensible small data. The first of these analyses was the deviation from the expected mean/share. By extracting the values from the selected geodatasets at the location of the sites (observation) and comparing them to the background values of the potentially habitable area of interest (expectation), different surface conditions can be assessed as being favourable or unfavourable for settlement or discovery. As this assessment can be visualized in a manageable number of charts, it can be considered an effective reduction of the geospatial big input. Another approach for pattern extraction was the application of the MaxEnt model. In addition to the response

curves, which are explained in the previous paragraph, MaxEnt allows the assessment of the predictive strength of geospatial input datasets via the jackknife variable importance. This statistic is an indication of how clearly the probability densities of presence and background data can be differentiated. Geodatasets with easy differentiation have high predictive strength and, thereby, variable importance, while geodatasets where both presence data and background data show a similar value range have low variable importance. As this allows for a direct comparison between geodatasets, it is an effective tool for pattern extraction within this research question. The third and final approach for pattern extraction was a combined statistical workflow for the assessment of cross-correlations as well as similarities and differences of geodataset values within the archaeologically predefined groups (e.g., chrono-cultural attribution or cave/open air). This approach was conducted within IBMM SPSS and included the calculation of contingency coefficients, classification based on the two-step cluster analysis, and testing based on a discriminant analysis and Naïve Bayes. While the assessment of contingency coefficients can help to identify cross-correlations and thereby prevent possible misinterpretations of causality, the other three assessments show how well the archaeologically defined classes are represented by natural grouping and how well they can be differentiated based on the geodataset values associated with them. This complex combination of geospatial and geostatistical approaches reflects the complexity of settlement and discovery factors. Due to the unambiguous attribution of some geospatial datasets to one of the investigated factors, all extracted patterns were interpreted with caution.

# 3 Assessing complex aeolian dune field morphology and evolution with Sentinel-1 SAR imagery – possibilities and limitations

Bruno Boemke[1,4], Imen Turki[2], Catrina Brüll[3], Frank Lehmkuhl[1]

[1] Department of Geography, RWTH Aachen University, Aachen, Germany

[2] Univ Rouen Normandie, Univ Caen Normandie, CNRS, M2C, UMR 6143, Rouen, France

[3] Institute of Hydraulic Engineering and Water Resources Management, RWTH Aachen University, Aachen, Germany

[4] Corresponding author. E-Mail: bruno.boemke@geo.rwth-aachen.de

This chapter was published as the following article: Boemke, B.; Turki, I.; Brüll, C.; Lehmkuhl, F. (2023b): Assessing complex aeolian dune field morphology and evolution with Sentinel-1 SAR imagery – Possibilities and limitations. In *Aeolian Research* 62. DOI: 10.1016/j.aeolia.2023.100876.

As the first and corresponding author, BB was responsible for the main investigation, methodology, data analysis, visualization, data curation, writing of the manuscript, and revision. The second author, IT, contributed in the fields of writing, revision, conceptualization, methodology, and visualization. The third author, CB, contributed in the areas of conceptualization, supervision, and funding acquisition. The last author, FL, contributed in the areas of conceptualization, revision, supervision, and funding acquisition.

## Abstract

Aeolian dune movement poses a threat to critical infrastructure, urban areas, water resources as well as agriculture. This threat is expected to increase in the coming years due to land degradation, desertification and climate change. Several approaches have been used to investigate the evolution of dune fields. Satellite remote sensing can be considered one of the most accurate tools for the continuous monitoring of global sand covered surfaces. Although early studies found a great potential in synthetic aperture radar (SAR) for dune assessment, the full potential has not been explored as of yet. Therefore, in this study, we present a novel method for assessing complex dune field morphology based on the easily accessible and globally available Sentinel-1 ground range detected (GRD) SAR dataset. In this application, dune features are extracted based on backscatter properties related to the local incidence angle. This provides a clear identification of (1) active dune sand, (2) dune ridges and (3) inter-dune ripples. By extracting these features through profiles, the multi-timescale evolution of the Western Mongolian dune field Bor Khyar was analysed through three areas of interest (AOIs) based on the spectral technique of continuous wavelets. The result of this investigation gives new insights into the temporal and spatial dynamics of dunes scale and their response to aeolian activity, revealing differences in aeolian activity as well as inter- and intra-annual variations in the dune morphology. These results are promising and highlight the potential in using satellite SAR imagery for dune monitoring.

## 3.1 Introduction

15% of the world's surface is covered by arid regions, of which approximately one third is mantled by sand deposits (Thomas 2011). In these regions, aeolian transport processes favour dune movement, posing a potential threat to critical infrastructure, urban areas, water resources as well as agriculture. Recent well documented examples for the hazardous aspects of aeolian transport can be found in the Nile region (El Gammal and El Gammal 2010; Eljack et al. 2010; Verstraeten et al. 2014; Saad et al. 2018), Western Asia (Al-Ghamdi and Hermas 2015; Pradhan et al. 2018; Amin and Seif 2019) as well as Western Africa (Ikazaki 2015; Benjaminsen and Hiernaux 2019; Yang et al. 2022). Due to land degradation, desertification and ongoing climate change, this threat is expected to increase dramatically and to occur in additional regions worldwide in the coming years (Davies et al. 2015; Reed and Stringer 2015; Shukla et al. 2019).

A continuous monitoring of sand covered surfaces is crucial to access the risk that desertification and related processes pose and ensure sustainable development in arid regions. Due to difficulties in accessibility of dune fields, field studies in these landscapes are often limited in their spatial coverage and associated with high costs. Satellite remote sensing, however, can cover large areas systematically, repetitively, and at a very low cost, making it the optimal tool for this task. In the last years, openly accessible remote sensing data have improved significantly in terms of spatial resolution and return rate. In combination with the recent progress in processing techniques, this has led to significant advances in the assessment and monitoring of aeolian sand dunes worldwide (Hugenholtz et al. 2012; Zheng et al. 2022).

A majority of these dune monitoring approaches are based on optical satellite data. The potential of synthetic aperture radar (SAR) data, however, has not been fully explored as of yet, despite the positive reception in early studies (Blumberg 1998; Qong 2000). The main advantages of SAR over optical imagery are the lack of dependence on atmospheric conditions and lighting and the easy discrimination of dunes due to the low volume scattering on sand in most associated microwave frequencies (Nashashibi et al. 2012). Therefore, the analysis and monitoring of single dunes or small aggregations which are easily distinguishable from coarser or vegetated surroundings have been the focus of recent SAR-based studies such as Havivi et al. (2018) and Delgado Blasco et al. (2020). While these studies show promising results in detecting and tracking dune features, they can only be applied where aeolian dunes are surrounded by high-backscatter surfaces. Complex dune fields, however, are much more challenging to analyse, as inter-dune boundaries show no significance differences in backscatter. The few studies that do access complex dune fields such as Mahmoud et al. (2020) and Manzoni et al. (2021), are using interferometric methodologies or offset tracking. While these methods are well-

suited to assess surface stability as well as displacement rates and directions, they only measure relative and absolute changes while neglecting dune morphology. However, as the dune morphology reflects past and present aeolian activity as well as sediment availability, we don't think that this factor should not be omitted in dune field analysis. Therefore, the main aim of this study is to explore the potentials and limitations of SAR imagery for assessing dune morphology in complex aeolian dune fields.

For this, we aim at developing a new method based on Sentinel-1 satellite imagery coupled with the spectral approach of continuous wavelets to assess the morphology of complex dune fields and investigate their temporal and spatial evolution. The new method will be applied and tested on the large Western Mongolian dune field Bor Khyar where the aeolian processes and dune morphology are not thoroughly investigated as of yet.

## 3.2 Materials and methods

### 3.2.1 The study area

The dune field Bor Khyar Els was selected as the study area due to its complexity in time and space and previous studies by the authors in this region. It is located in the so-called Valley of the Great Lakes in Western Mongolia, which consists of several large catchments, ultimately draining into the endorheic lakes Uvs Nuur, Khyargas Nuur and Khar Nuur. Each of these lakes marks the starting point of a large dune field as can be seen in Figure 3. The large basin has a minimum elevation of 800 meters above sea level (asl) in the north and 1200 m asl in the south and is surrounded by the Altai and Khangai mountains, where summits reach more than 4000 m asl. While the central parts of these mountains are made up of Palaeozoic granites and gneiss and the margins by metamorphic and sedimentary rocks (Lehmkuhl 1999), the basin itself is mostly covered by thick layers of fanglomerates, dune sands and lacustrine sediments (Walther and Naumann 1997). Within the basin, landforms such as vast dune fields, widespread alluvial fans and beach ridges up to 130m above current lake level hint at large climatic variations in the past, which were investigated in numerous studies (Grunert et al. 2000; An et al. 2008; Klinge et al. 2017; Lehmkuhl et al. 2018; Klinge and Sauer 2019).

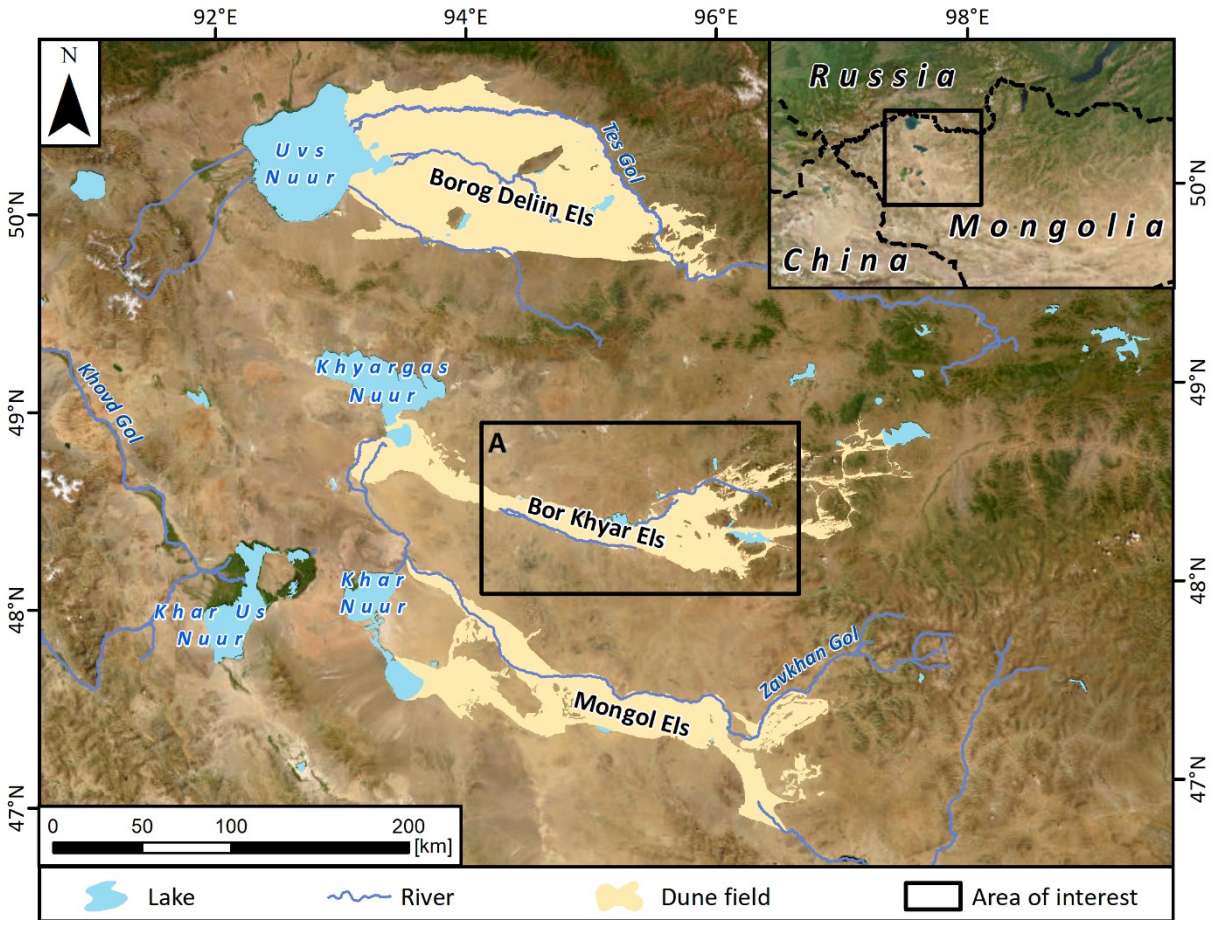



*Figure 3: Overview map of the broader study area of the valley of the great lakes. It includes the big endorheic lakes, large rivers and the outlines of the three major dune fields, all framed by high resolution optical imagery (ESRI 2022). The central dune field of Bor Khyar (A) is the main target of this study.*

The recent climate is mainly characterized by extreme continental conditions. As such, temperatures reach less than -20°C in winter but close to 20°C in summer. The precipitation in the area has its main gradient between the arid basin, reaching an annual precipitation of 50mm in some parts, and the semi-arid mountain ranges with an annual precipitation of 200-400mm (Klinge 2001). Annual precipitations vary strongly over the seasons with 70-80% falling during the early summer months (Lehmkuhl and Klinge 2000). The wind regime is mainly influenced by westerly to north-westerly winds, reaching the highest wind speeds in the winter months. In the summer, far weaker winds from the south-east prevail (C3S 2022).

The Bor Khyar Els dune field itself was formed by westerly winds during the Pleistocene and Holocene and stretches from the Khyargas Nuur lake in the west high into the Khangai mountains in the east (Grunert and Lehmkuhl 2004). It covers a distance of more than 200 kilometres with a maximum width of about 40 kilometres. In its mid-section, the dune field dams off the Bayan Nuur lake, as the westward advance of the water is blocked by vertical offset from a north-south running tectonic fault line (Enkhbold et al. 2021). The two neighbouring dune fields, Borog Deliin Els in the north and Mongol Els

in the south, both display a cyclic sand system. This means that sand is transported upslope and westwards through the wind where it is bound by the river system and transported downslope and eastwards again. The Bor Khyar Els, however, lacks a strong adjacent fluvial system that can act as a natural barrier to the aeolian processes and transport sand downslope into the source lake (Grunert and Lehmkuhl 2004). Instead, after traveling 800 metres vertically up-slope, the sand disperses into several smaller dune fields situated within mountain valleys.

Preliminary visual analyses of the dune field based on high resolution optical satellite imagery reveal a large variety of dune types and related sandy surfaces including sandy plains, transverse dunes, grid dunes, longitudinal dunes, parabolic dunes as well as barchans. A preliminary analysis of the ERA5 climate dataset reveals large seasonal differences in wind with the strongest winds in winter from west to northwest (C3S 2022). Based on this preliminary analysis, three areas of interest (AOIs) with differences in complexity, dune types and wind conditions were selected for a detailed analysis. The AOIs include; (A) The Bor Khyar dune field as a whole, (B) A system of dense large transverse dunes, superposed and surrounded by smaller transverse dunes and grid dunes in the south of the dune field, (C) A system of presumably inactive separated large transverse dunes, surrounded by sandy plains in the north of the dune field and (D) A system of transverse and grid dunes in the north-east of the dune field. Figure 4 offers a first overview over these AOIS as well as ERA5 wind directions and speeds over AOI A.

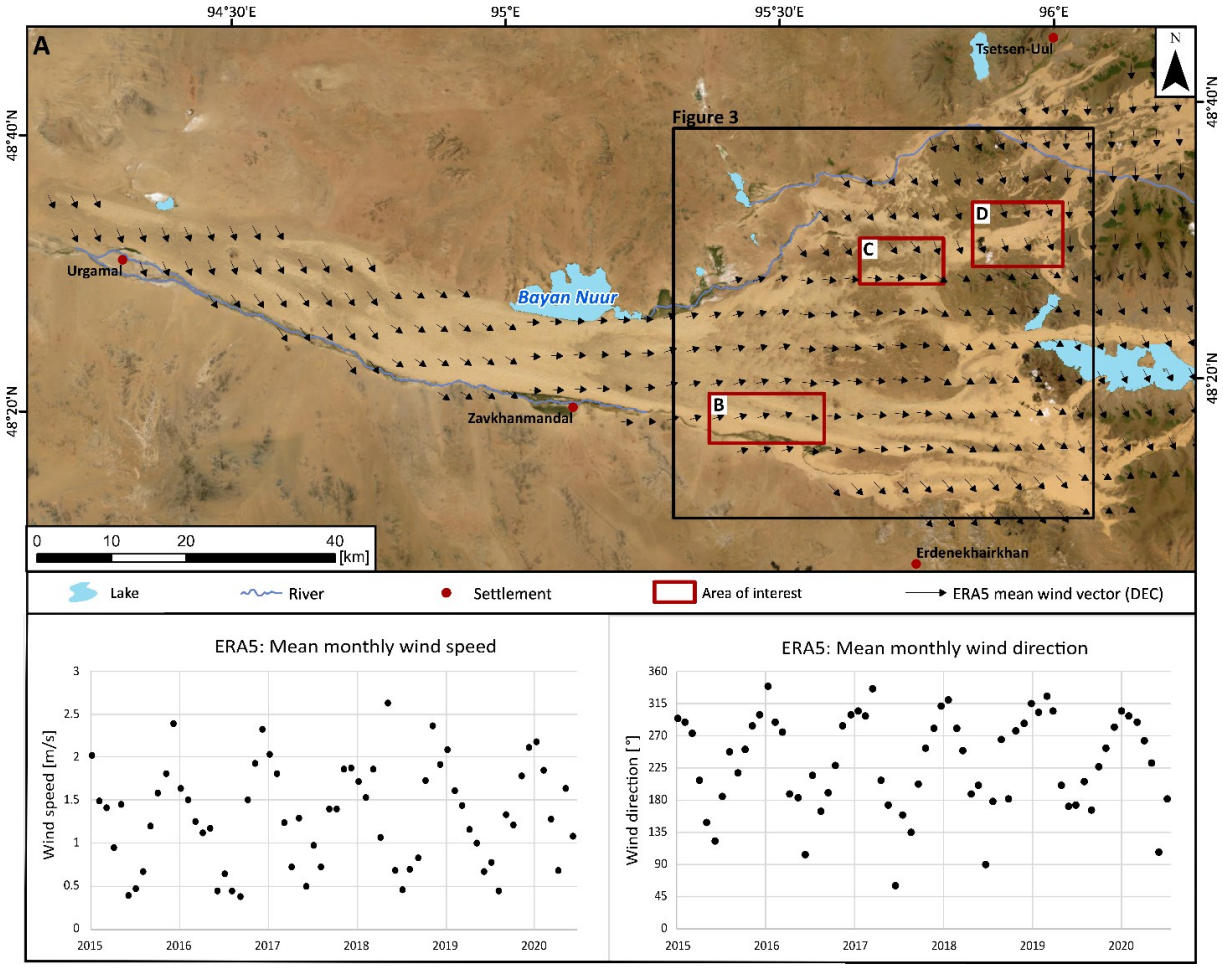



*Figure 4* Upper half: Map of the dune field Bor Khyar including high resolution optical imagery (ESRI 2022) and mean ERA5 wind vectors for the month December calculated from daily data between 2015 and 2021 (C3S 2017). December was chosen for wind vector display as the highest wind speeds occur here. The outline of figure 5 as well as the areas of interest (AOIs) are added for orientation purposes. Lower half: Monthly mean assessment of ERA5 wind speed and wind direction over the whole dune field. Note that the highest mean wind speeds occur during winter from west to north while lowest wind speeds occur during summer from east to south.

### 3.2.2 Sentinel-1 SAR imagery

Although early studies saw great potential in SAR data for the extraction of dune characteristics (Blumberg 1998; Qong 2000), requirements for a continuous global monitoring at high spatial resolution have only been met since the launch of Sentinel-1 in 2014 (Torres et al. 2012). The main advantage that Sentinel-1 has over its predecessors are the high spatial resolution of 10*10 meters in the main acquisition mode (Interferometric Wide Swath, IDW), a short repeat cycle of 1-6 days when combining both S-1A and S-1B satellites and the possibility of single and dual polarisations (Geudtner et al. 2014). In combination with the broad availability and easy accessibility, this makes Sentinel-1 the ideal tool for SAR-based assessment as well as continuous monitoring of dunes.

Sentinel-1 operates within the C-Band with a wavelength of 5.6 cm. At this wavelength, even thin layers of sand cannot be penetrated. In combination with the very low backscatter on sand due to the relative

surface smoothness, this wavelength allows for a clear differentiation between sand and surrounding surfaces (Williams and Greeley 2004). Within active dunes fields, differences in backscatter can mainly be attributed to the local incidence angle, as the influence of vegetation and grain size differences can be neglected here. As such, slopes that are oriented towards the sensor show slightly elevated backscatter than those that are oriented away from the sensor (Blumberg 1998). With sufficient spatial resolution, this allows for the detection of smaller aeolian forms, superposing the main dune forms. The most prominent dune feature that can be easily extracted with C-band SAR is the ridge, as its morphology leads to double-bounce backscattering (similar to a corner reflector), resulting in a backscatter far higher than the surroundings (Blumberg 1998; Delgado Blasco et al. 2020). These general assumptions towards the SAR-interaction of dunes can be visually confirmed in Figure 5.

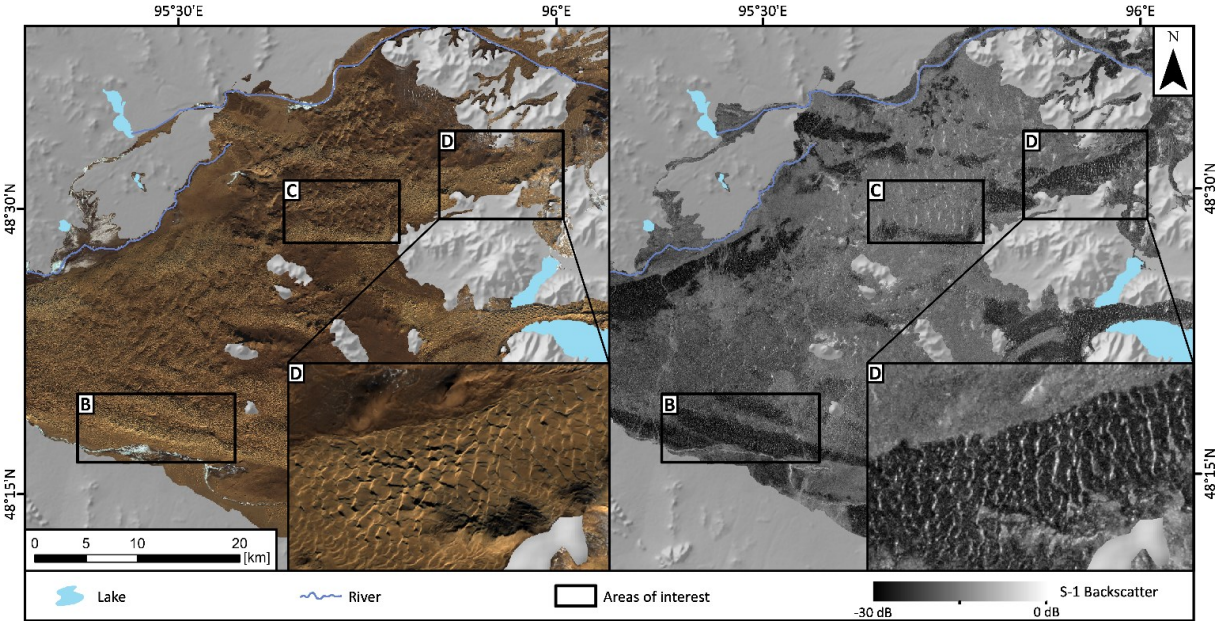



*Figure 5:* Comparison between Sentinel-2 true colour composite (left) and Sentinel-1 backscatter (right). Non-sandy surfaces are masked out and replaced by SRTM hillshade. Dark colours on the right image indicate aeolian sediments. More details can be seen in the cut-out in the lower right. The areas of interest B, C and D mark the areas that were selected for detailed analysis. See section 3.2.3 for further context.

### 3.2.3 Geospatial processing

As morphological dune parameters can be directly extracted from digital elevation models (DEM), the first attempt was to create a DEM from Sentinel-1 scenes. For this, the workflow for Sentinel-1 based DEM extraction proposed by Braun (2021) was applied to two scenes from the dates 2021-02-10 and 2021-02-22. The main criteria for the identification of these two scenes was the short temporal baseline of 12 days and a large perpendicular baseline distance of 152 meters, which were identified using the online baseline tool by the Alaskan Satellite Facility (ASF 2022). The resulting DEM showed promising results for inactive dune fields and the surrounding mountain ranges. Within active dune

fields, however, elevations showed large negative deviations and an overall implausible result. This can be attributed to the very low backscatter on sand and the large variation in backscatter within small distances resulting in a very low coherence, which is challenging for interferometric SAR analysis. As such, the DEM-approach was dismissed. Instead, we focused on the indirect derivation of morphological features from Sentinel-1 ground range detected (GRD) data.

The pre-processing and extraction of Sentinel-1 GRD scenes as well as the ERA5 wind analysis was conducted within the Google Earth Engine (GEE), an openly accessible cloud-based platform for geospatial analysis, powered by the Google server infrastructure (Gorelick et al. 2017). For Sentinel-1, we chose to work with the VV-polarized dataset, as it shows the highest dependency on the local incidence angle. To achieve analysis-ready-data (ARD), the pre-processing included additional border noise correction, speckle filtering and radiometric terrain normalization. For this, we used a slightly modified version of the openly accessible ARD script by Mullissa et al. (2021), applying a 10-scene multi-temporal improved Lee sigma speckle filter with a kernel size of 15 cells (Lee et al. 2009). For the terrain normalization, the NASA SRTM global digital elevation model was used (Farr et al. 2007). Based on a preliminary visual interpretation, pre-processed Sentinel-1 scenes from 4 dates were selected and exported for further processing. The dates include 2015-01-31, 2018-06-14, 2018-10-12 and 2021-12-07, each for which a Sentinel-1 GRD mosaic of the whole dune field was created. The GEE was also used to calculate wind direction and speed from ERA5 wind u-component and v-component. Monthly data was used to aggregate speed and direction over the full timespan (2015-2021) while wind vectors for visualization were created from daily data.

For the indirect derivation of morphological features as well as an analysis of their evolution in time, a frequency analysis in form of the continuous wavelet transform (CWT) was applied to these mosaics. To compare these backscatter values to a directly derived morphology, SRTM and FABDEM elevation model profiles were also extracted (Farr et al. 2007; Hawker et al. 2022). Both elevation models have a spatial resolution of 25*25 meters in the study area. While the SRTM represents a surface from the year 2000, the FABDEM is an improved version of the Copernicus GLO-30 DEM that was surveyed between 2011 and 2015. As such, the comparison between the two DEMs offers a first assessment of dune activity. For the profile extraction, the already mentioned AOIS were used, representing different dune types and different levels of complexity (see Figure 4 and Figure 5). For the spatial processing and extraction of profiles, we used ArcGIS, version 10.7.1. The pre-processing for profile analysis included reprojection into a metric coordinate reference system (WGS84, UTM Zone 46N, EPSG:32646) and resampling into the same spatial resolution of 10*10 meters. In each AOI, a central tracking axes was defined manually, representing the centreline of the local dune field along the wind direction. Four additional profiles parallel to each central tracking axis were extracted in distances of 200 and

400 meters left and right. The resulting lines were used to extract S-1 and SRTM surfaces to profiles. These profiles were then extracted as csv text files via the 3D Analyst toolbar, ready for the visualization in Excel 2016 and the CWT analysis. To enhance the understanding of this multi-step workflow, it is visualized in Figure 6.

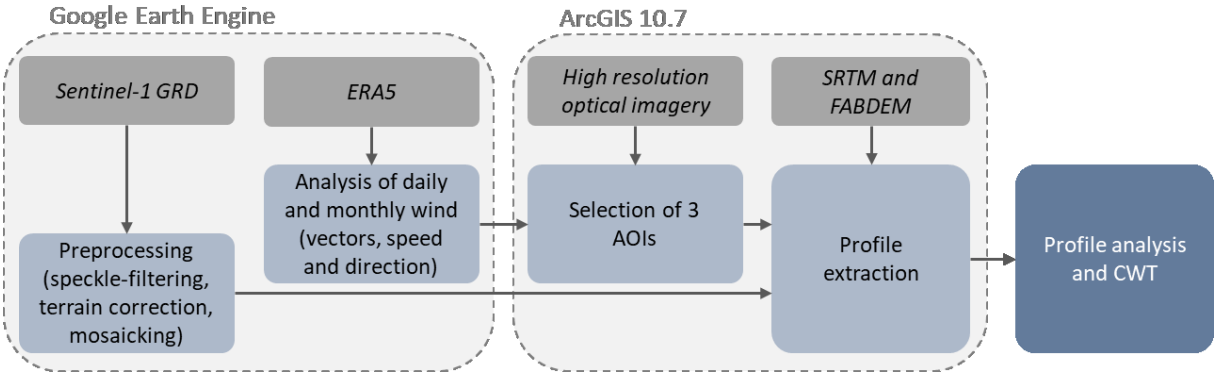



*Figure 6:* Workflow diagram showing the different processing steps and the environments they were conducted in. Datasets are displayed in grey, processing steps are displayed in light blue and the central part of the study, the profile analysis and continuous wavelet transform (CWT) are displayed in dark blue.

### 3.2.4 Spectral analysis based on wavelet transform for the morphological decomposition

From a technical point of view, dune fields, such as the ones investigated in this research, can be seen as a set of spatially variable morphological features at different scales. This association gets clear when looking at the profiles portrayed in Figure 7, Figure 8, and Figure 9. The spacing and size of these features is a function of wind interactions with a surface and sand availability. The variability of such dunes in response to external forces such as wind follows a non-linear pattern. All changes result in a combination of several morphological modulations at various ranges of spatial scales. Therefore, we used an approach to investigate the spatial frequencies of morphological changes by decomposing the total variability to a series of scales. This method is useful to identify the evolution of dunes and their organization in response to the external drivers of aeolian energy and internal drivers associated to the characteristics (texture, age, …) of sediments composing these morphological forms.

To do so, we used the techniques of Continuous Wavelet Transform (CWT), which are well documented in Labat et al. (2005) works and well-known for hydrological, meteorological and climate applications (e.g. Turki et al. 2015; Massei et al. 2017). The CWT has been explored by Turki et al. 2021) for other applications related to the morphology of intertidal dunes and their migration under the wave-tide interactions, case of the Baie de Somme (France). These publications have demonstrated the relevance of this method to gain more insights into the evolution of morphological structures at different temporal and spatial scales.

When applying a CWT analysis, a complex signal (e.g. Sentinel-1 profiles through dune fields) is scanned using different wavelengths. While scanning with a wavelength, high rhythmicity results in a high power of this wavelength, while absent rhythmicity is reflected by low power. This way, the rhythmicity and thereby relevance of each tested wavelength is collected along the signal/profile. The resulting CWT-Diagram gives a good overview of which wavelengths of dune features are relevant along the profile. Spatial and temporal differences in the CWT diagram allow an interpretation of the morphological forces. Taking advantage of this approach, changes in dune features, detected via Sentinel-1 SAR imagery, as well as the possible rhythmic structures of dunes associated to the sediment transport dynamics and the aeolian activity were explored.

The CWT has been carried out for the four Sentinel-1 scenes mentioned under 3.2.3 at different profiles along the AOIs B, C and D. For the assessment of cross-axis variability, the CWT was applied to each central axis as well as four parallel axes 200 and 400 meters north and south of the central axis (see Figure 7, Figure 8, and Figure 9 for localization of the central axes).

The CWT diagrams contain: (1) the contour diagram with space along the profile in meters*10 on the x-axis; (2) the wavelength of morphological features (dunes) in meters*10 on the y-axis; and (3) the power or variance on the z-axis.

## 3.3 Results

In section 3.3.1, the Sentinel-1 images are visually interpreted in comparison to optical imagery and a DEM. In addition, the extracted profiles along the central tracking axes are presented and analysed. In section 3.3.2 we the present the results of the CWT frequency analysis. Table 3 gives an overview over which methods were applied in the different AOIs.

*Table 3: Short description of the different areas of interest (AOIs) and which methods were applied in them.*

| AOI | Description | Visual interpretation | Profile analysis | CWT |
|---|---|---|---|---|
| **A** | The Bor Khyar dune field as a whole | **+** | **-** | **-** |
| **B** | A system of dense large transverse dunes, superposed and surrounded by smaller transverse dunes and grid dunes | **+** | **+** | **+** |
| **C** | A system of presumably inactive separated large transverse dunes, surrounded by sandy plains in the north of the dune field | **+** | **+** | **+** |
| **D** | A system of transverse and grid dunes in the north-east of the dune field | **+** | **+** | **+** |

### 3.3.1 Visual interpretation and profile analysis

In the visual interpretation of the whole dune field (AOI A), Sentinel-1 scenes, Sentinel-2 true colour composites and the SRTM DEM were compared. This comparison confirms the expectations about the C-Band sensitive backscatter characteristics of active and inactive dunes, stated in 3.2.2. As such, it could be observed that active dune fields can be separated from inactive dune fields and surrounding areas by their lower backscatter.

For a simple thresholding approach, a backscatter of <-20dB yielded the best results in differentiating active dunes across all scenes. The only other surfaces in the study area that fall below this threshold are lakes. Confusion between these two surface types, however, can be easily avoided by the complementary use of optical imagery and/or surface water datasets.

The only features within active dune fields, that do not fall below the threshold, are dune ridges. Instead, they show strongly elevated backscatter values, reaching above 0dB at bend points within dune slip surfaces (see Figure 5 and/or Figure 9). This supports the double-bounce backscatter interaction and corner-reflector-similarity mentioned in 3.2.2. A direct comparison to optical imagery shows the importance of the direction of illumination: ridges perpendicular to the illumination direction are enhanced while ridges parallel to the illumination direction only show slightly elevated backscatter or no difference in backscatter at all (see Figure 5 and/or Figure 9).

Within the south-western AOI B, the visual interpretation of SAR, optical imagery and the DEMs in comparison show the main limitations of Sentinel-1-based complex dune field analysis. As can be seen in Figure 7, the Sentinel-1 scene still allows for an easy differentiation between inactive and active dune fields as well as an easy identification of dune ridges perpendicular to the illumination direction. This information, however, can only be obtained for the superposing grid-dune structures with inter-ridge-distances of about 100-200 meters. The underlying large transverse forms with wavelengths of about 1000m, which can be clearly seen in both the optical imagery as well as the DEMs, are not visible in the Sentinel-1 scenes.

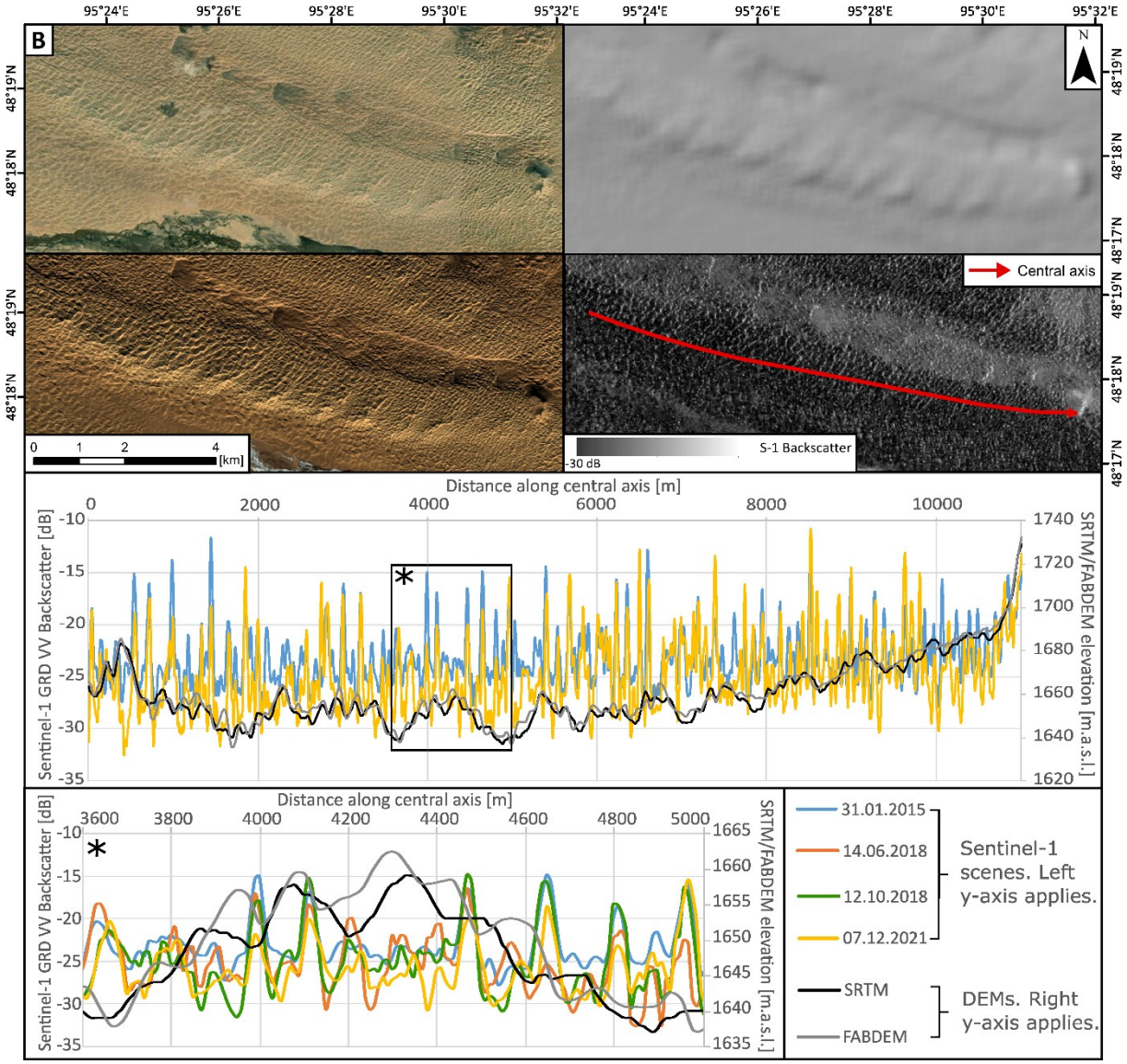



*Figure 7:* Comparison of the source data used to analyse AOI B (upper half), including high resolution optical imagery (upper left), the SRTM DEM (upper right), Sentinel-2 optical imagery (lower left) and Sentinel-1 GRD SAR imagery (lower right). On the lower half of the figure, we see the Sentinel-1, SRTM and FABDEM profiles at the central axis (see lower right map for localization). A subsets of the profile is shown in detail to highlight the interaction between the Sentinel-1 backscatter and the morphology as well as the temporal changes in backscatter and DEM. As can be seen in the lower left diagram, the highest SAR backscatter peaks migrate downwind while the smaller peaks and valleys in between vary strongly. This indicates active conditions, which is supported by the changes in the DEM.

In the Sentinel-1 profile graphs, no differences in backscatter between the underlying large transverse dune slope facing the sensor and facing away from the sensor can be detected, either. As such, only information about the superposing grid dunes can be extracted from these profiles. In comparison to the DEM profiles, showing only the underlying transverse forms and 2-4 superposing grid dune ridges on each of these, the Sentinel-1 profiles provide surface information in far higher detail, including several peaks and valleys in between the main superposing grid dune ridges. These changes in backscatter between ridges are most probably triggered by differences in the local incidence angle,

indicating ripple features. When comparing the Sentinel-1 profiles from different dates, the highest peaks, indicating main dune ridges, show a high overlap. However, the magnitude of the peaks differs widely between the scenes, showing differences of up to 15 dB. In addition, a shift of several meters along the profile axis through time can be observed in many cases, indicating wind-induced dune movement and as such active aeolian conditions. This is additionally supported by the differences between the SRTM and FABDEM elevation, showing a windward dune migration between the years 2000 and 2011-2015.

The northern AOI C shows large visual differences in comparison to the other AOIs. The main ridges show far higher distances of 700 to 1100 meters. In the Sentinel-1 scenes, only slight variations in backscatter can be observed between these main ridges. In addition to that, the backscatter is generally higher, not falling under the threshold for finely grained active dune sand of -20 dB. The comparison between the SRTM and FABDEM also show a high overlap. Within the optical Sentinel-2 imagery, no indications for vegetation can be found in these areas. As such, the higher backscatter is most likely caused by larger grain sizes, indicating inactive conditions where the finely grained sand is eroded.

In the Sentinel-1 profile graphs of AOI C, the overlap of peaks and valleys along the profile axis is the highest among all AOIs. This supports the assumption of inactivity based on the visual interpretation. When assuming temporal stability, this part of the dune field offers the opportunity to locate the double-bounce-related high backscatter in the Sentinel-1 profiles in comparison to the FABDEM morphology. As Figure 8 shows, this peak is located at the steepest part of the slip face, supporting the double-bounce assumption. Therefore, when trying to extract ridges via Sentinel-1, a certain offset between the highest backscatter and the actual ridge has to be considered.

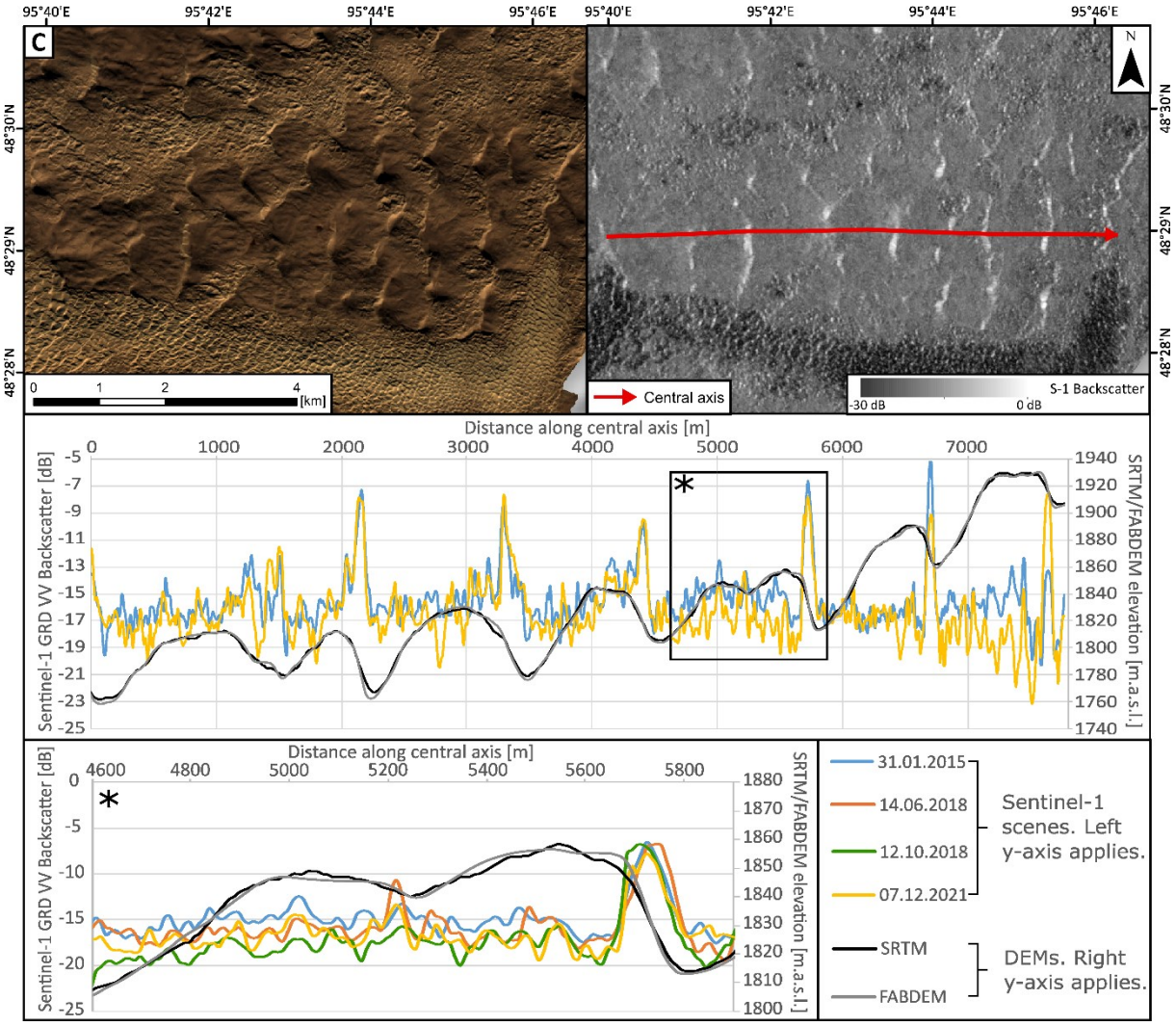



*Figure 8: Comparison of optical imagery (upper left) and SAR imagery (upper right), showing a mostly inactive part of the dune field. On the lower half of the figure, we see the Sentinel-1 and DEM profiles at the central axis (see upper right map for localization). A subset of the profile is shown in detail to highlight the interaction between the Sentinel-1 backscatter and the DEM morphology as well as the temporal changes in backscatter. As this diagram shows, the highest SAR backscatter peaks are located at the DEM ridges and no clear direction of dune migration can be seen. The inter-ridge variations are also comparatively low and the DEMs show a very high alignment. This indicates inactive conditions.*

The north-eastern AOI D shows a complex system of transverse, grid and barchanoid dunes. The mean inter-ridge-distance is between 200 and 300 meters, showing a high variability within the AOI. Within both the optical and the SAR imagery, differences within the AOI, especially between the northern and southern part, can be observed. While most dunes in the south seem to be formed by westerly winds as indicated by their north to south-alignment, dunes in the northern parts of the AOI shift towards a north-east to south-west alignment, indicating stronger north-westerly winds. This fits the ERA5 wind data, showing the strongest winds from north-west during winter in this part of the dune field (see Figure 4). Due to the importance of the illumination direction, some dune ridges showing the strongest east-to-west alignment are not represented in the Sentinel-1 GRD imagery (see Figure 9).

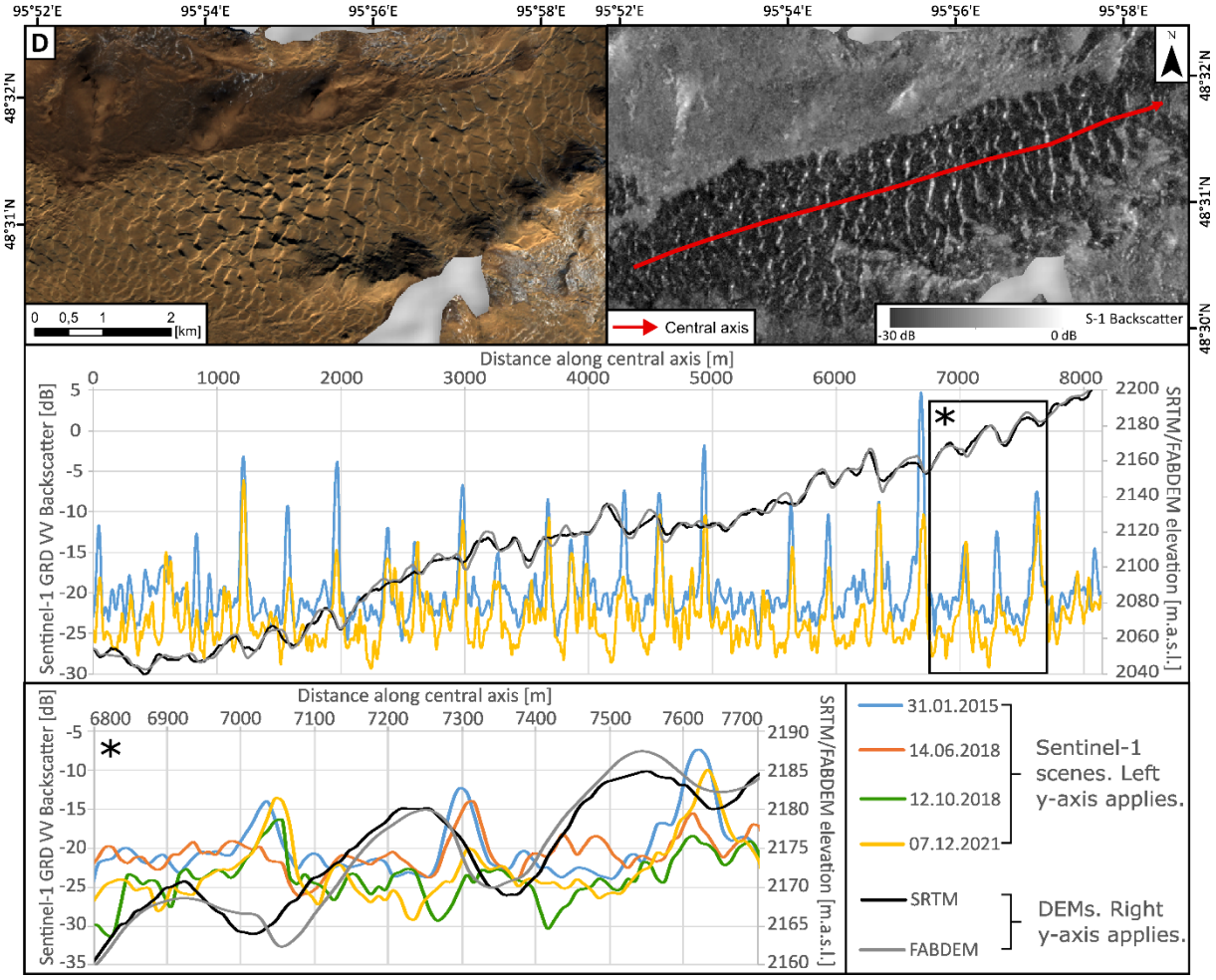



*Figure 9:* Comparison of optical imagery (upper left) and SAR imagery (upper right), showing an active part of the dune field. On the lower half of the figure, we see the Sentinel-1 and DEM profiles at the central axis (see upper right map for localization). A subset of the profile is shown in detail to highlight the interaction between the Sentinel-1 backscatter and the SRTM morphology as well as the temporal changes in backscatter. As this diagram shows, the highest SAR backscatter peaks are located at DEM dune ridges and migrate downwind while the smaller peaks and valleys in between vary strongly between each date. This indicates active conditions, which is supported by the changes in the DEM.

The central tracking axis profile graphs of AOI D show many similarities to AOI B. Comparing the different Sentinel-1 scenes, the main ridges show a high overlap with a slight shift along the tracking axis through time, indicating wind-induced dune movement. A far higher variability can be seen in between the main ridges, indicating vast changes in superposing sand forms such as ripples. These assumptions towards aeolian activity are additionally supported by the comparison between the SRTM and FABDEM, showing windward dune migration.

### 3.3.2 Continuous wavelet transform

The spectral approach of continuous wavelet transform (CWT) has been used comparatively in the AOI's B, C and D to investigate the present morphological forms as well as their inter-annual and intra-annual variability with the aim to identify the morphological modulations of dunes. Based on absolute and relative differences between these modulations, the aeolian activity and its influence on the dune changes are discussed. In this section, we are presenting the results for three cross sections of each AOI, namely north (Figure 10), central (Figure 11) and south (Figure 12), represented by the central tracking axis and the parallels in 200 meters distance. This approach is useful to identify the different rhythmic patterns controlling the dune variation (erosional vs depositional areas) and the wavelength of the associated morphological forms. The application of the CWT to the different AOI's sections has highlighted the presence of dune modulations and their associated morphological dune features as well as their evolution through time.

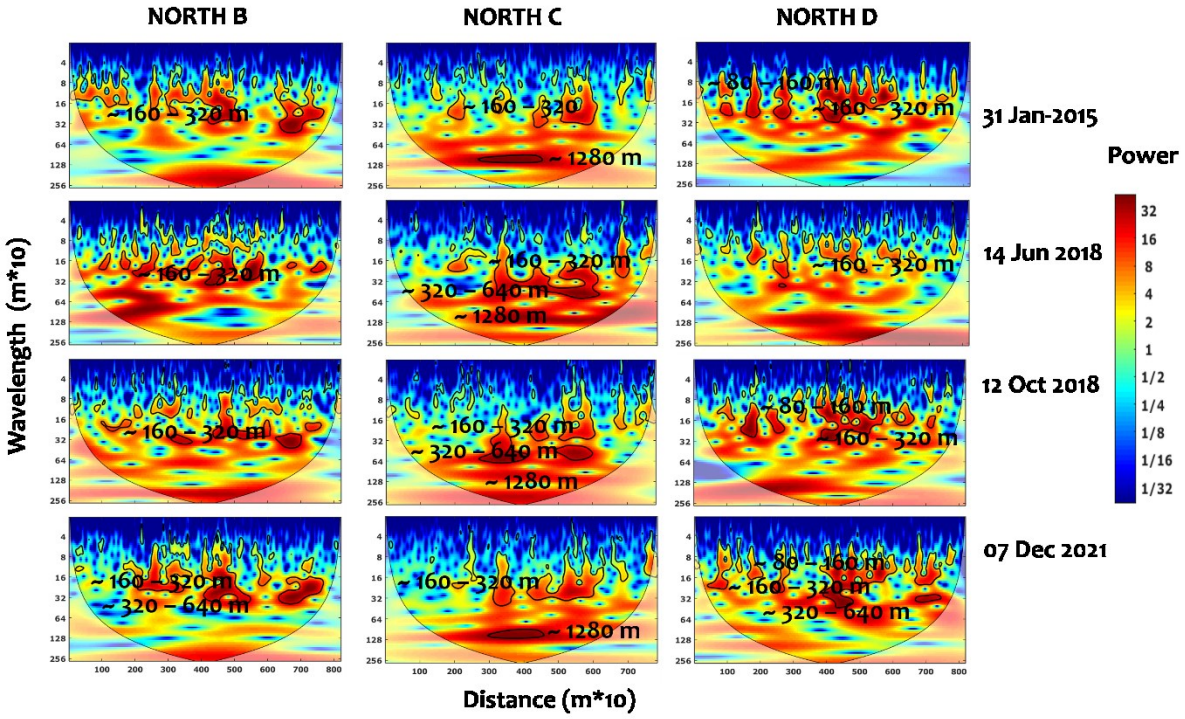



*Figure 10: Continuous wavelet transform (CWT) diagrams of dune morphological changes in the different areas B, C, and D during the different Sentinal-1 scenes: 31 Jan 2015; 14 Jun 2018; 12 Oct 2018; 7 Dec 2021 in the north of the AOIs. The wavelengths of multi-space-scale features (y-axis of the CWT diagram) identified are: ~8-16, ~ 16-32, ~32-64 and ~ 128 m*10. x-axis: the distance; y-axis: the frequency (spatial scale) or equivalent wavelength; colour scale: the power or variance (which quantify the correlation between the signal and the wavelet basis) from blue (low) to red (high).*

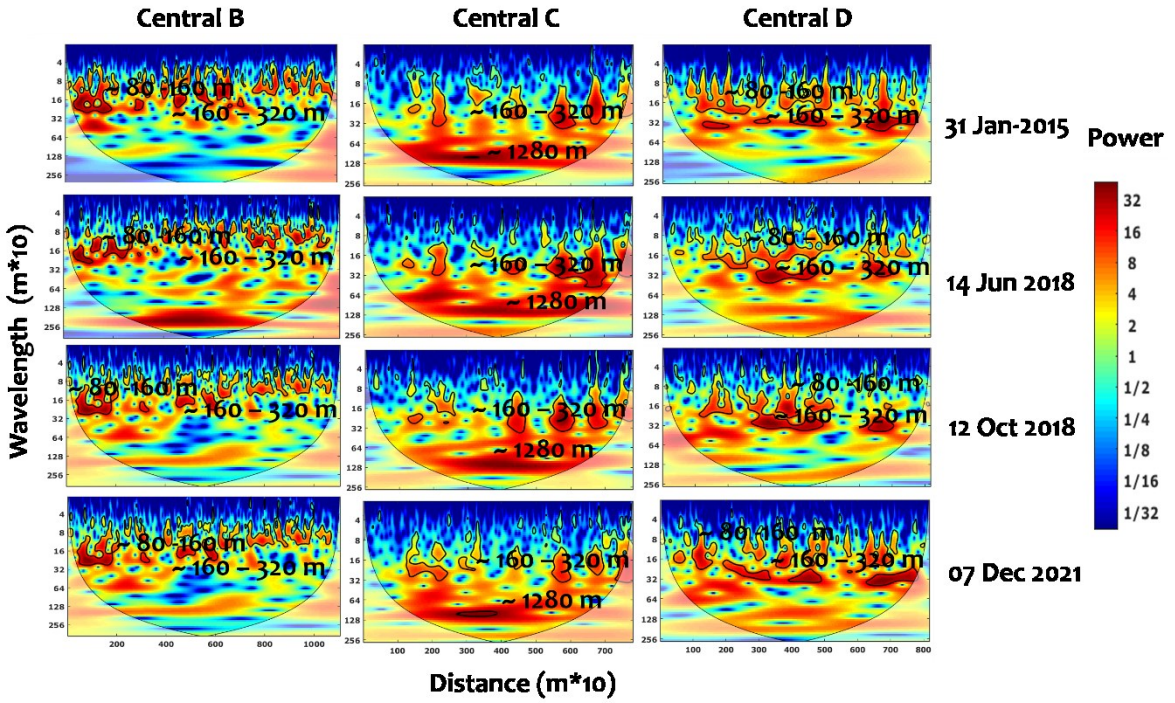



*Figure 11: Continuous wavelet transform (CWT) diagrams of dune morphological changes in the different areas B, C, and D during the different Sentinal-1 missions: 31 Jan 2015; 14 Jun 2018; 12 Oct 2018; 7 Dec 2021 at the central axis of the AOIs.*

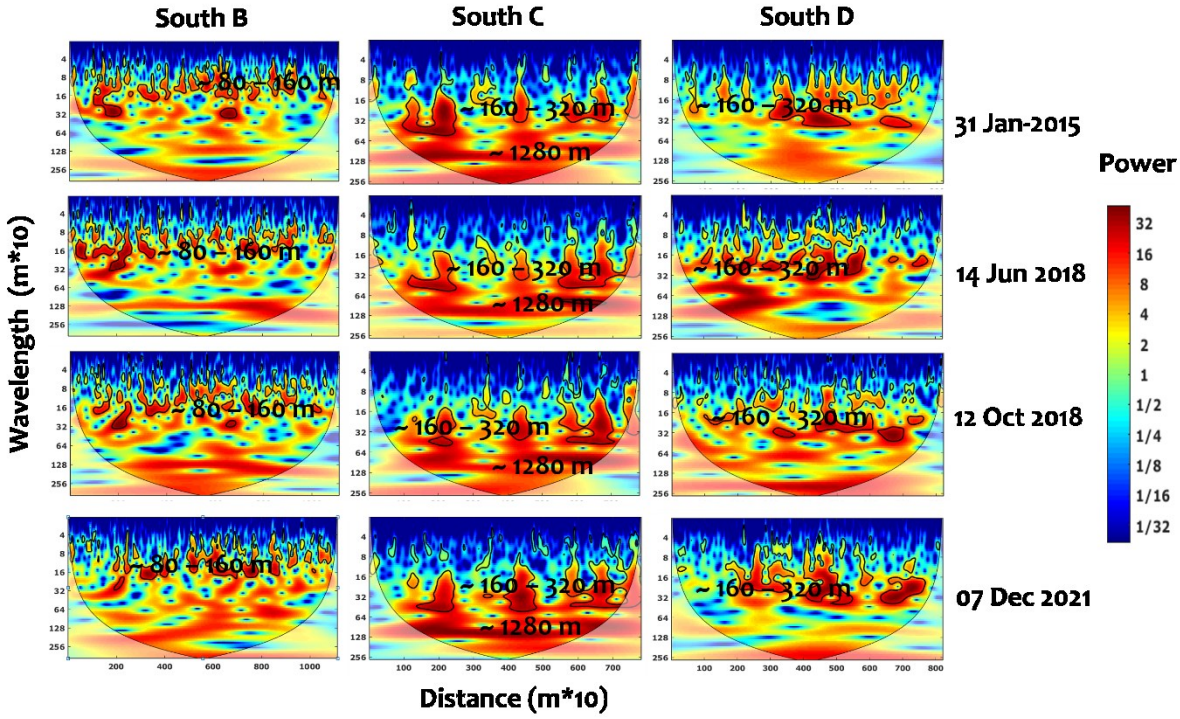



*Figure 12: Continuous wavelet transform (CWT) diagrams of dune morphological changes in the different areas B, C, and D during the different Sentinal-1 missions: 31 Jan 2015; 14 Jun 2018; 12 Oct 2018; 7 Dec 2021 at the south of the dune AOIs.*

As can be seen in Figure 10, Figure 11, and Figure 12, the distribution of energy in the CWT spectrum is non-uniform with altering bands of high and low power, represented by red and blue colours, respectively. This distribution highlights the existence of several modulations of morphological changes structured at a series of spatial periods with different wavelengths, divided into the categories ~8-16 (shortest wavelengths, highest spatial frequencies), ~160-320, ~320-640and ~1280 m (longest wavelengths, lowest spatial frequencies).

The high power of the ~1280 m wavelength seems to be similarly manifested during all years in AOI C at all profiles from the north to the south. Such modulations are associated with the large and inactive dunes in this AOI. The small dune features, shown by the wavelength of ~8-16 m*10, are often connected to larger structures (~160-320 and ~320-640 m) as illustrated in AOI B and D. This connection seems to be significantly manifested in the north of the AOIs with regular structures, however, it is relatively limited in the centre and the south.

The smaller morphological modulations of ~80-160 m exhibit non-organized forms according to the scale range; most of them are observed from the centre to the south of AOI B and vanish in the north of the AOI where the morphological modulations are well structured and strongly connected.

According to the spectral analysis of the morphological dune changes, a series of findings can be formulated as follows:

1. The temporal variation of dune morphology is highlighted by an inter-annual evolution of morphological modulations from 2015 to 2018 which is significantly manifested in AOI D where the dynamics are highly active. This dynamic seems to be extremely reduced for dunes of AOI C.
2. The seasonal patterns of dune changes, reflected from the comparison of dune fields between June and October of 2018, are manifested in the different AOIs and exhibit a pronounced sedimentary connection in winter periods when the sediment transport induced by the increasing aeolian energy is important. However, this connection seems to be reduced during summer periods, most probably due to the seasonal shift in the wind regime (section 3.2.1).
3. The morphological forms of dunes are strongly related to the energetic conditions of wind. The distribution of morphological forms from homogenous structures in AOI B and AOI D (clear connection between small and large dune features) to non-homogenous ones in AOI C (inactive paleo-dunes with limited connections to small dune features) where the sediment transport induced by wind regime seems to be controlled by different directions.

## 3.4 Discussion

### 3.4.1 Limitations and sources for error

In comparison to optical imagery, Sentinel-1 GRD data requires an extensive pre-processing procedure before it is in an analysis-ready state. The aim of this procedure is to remove any noise in form of radar speckle, terrain effects and border noise. When applying a speckle filter, however, there is a chance that not only noise is removed but also a part of the signal is smoothed out. The choice of the filtering algorithm and parameters should therefore be considered with caution. Based on this study, we recommend the use of the improved sigma filter (Lee et al. 2009), effectively filtering speckle while preserving the signal. As applying this filter to larger scenes requires a lot of processing power, we additionally recommend the ARD-implementation into the GEE by Mullissa et al. (2021).

Due to the areal low coherence between Sentinel-1 scenes of active dune fields, interferometric approaches and DEM-derivation are not applicable. Therefore, dune morphology and dune features have to be extracted indirectly from GRD scenes. Backscatter differences in active dune fields are mainly influenced by the local incidence angle. Therefore, the direction of dune features in comparison to the illumination direction is one of the main limitations. As could be seen in AOI B, another main limitation is the fact that subsurface forms like the underlying large transverse dunes cannot be detected in GRD imagery. Therefore, it is necessary to include optical imagery or DEMs to assess the dune field comprehensively.

The Sentinel-1 GRD profile analysis in this study has shown that while the backscatter values of the scenes show a high spatial overlap, the absolute backscatter values vary widely between them. For automated thresholding and feature extraction processes, this means that a variable approach has to be applied, evaluating each scene separately.

### 3.4.2 Potentials and perspective

From Sentinel-1 GRD imagery, three main features relevant for complex dune field dynamics can be extracted. The first is finely grained active dune sand, which can be easily differentiated from coarsely grained or vegetated inactive dunes as well as other adjacent surfaces by its very low backscatter of lower than -20dB. The second feature is dune ridges, which have a far higher backscatter than the surrounding dune sand due to the double-bounce backscatter effect. This effect, however, is highly dependent on the illumination direction and the curvature and slope of the ridge, leading to large differences in backscatter between scenes. The third type of dune features that can be extracted via Sentinel-1 GRD imagery are smaller scale ripples between the main ridges.

As we have shown in this study, the CWT frequency is a fitting method to analyse these indirectly derived dune features. As such, it provides useful information related to the changes in spatial and temporal morphological modulations of dunes in response to aeolian energy conditions. This approach has given new insights related to the morphological modulations of dunes in time and space. A series of hypotheses have been highlighted; (i) strongest aeolian activity is found in the north of the AOIs which are mainly controlled by an exposure to north-westerly winds, (ii) multi-timescale evolution of dunes can be observed in AOI B and D in response to active aeolian conditions from seasonal to inter-annual scales, while they are weakly active to inactive in AOI C, (iii) strong morphological connections between small and large dune features can be observed during winter periods while these connections are weakly manifested during summer periods when changes in the energy and the direction of wind are significant. As such, this technique has given some insights into the external (climate conditions) and internal (sediment characteristics) drivers controlling the morphological migration of dunes.

The main advantage of Sentinel-1 SAR GRD over optical imagery for complex dune field analysis is the easy identification of the mentioned morphological features as well as the easy discrimination of inactive and active dunes. This is due to the fact that these different surfaces and features show very little differences in their spectral reflectance, making them difficult to distinguish in optical imagery. In C-band SAR, however, the backscatter is mainly influenced by surface roughness and local incidence angle, allowing an easy differentiation.

The main advantage of Sentinel-1 SAR GRD over openly accessible global DEMs is the higher spatial resolution and multi-temporality. Although the dune morphology could be directly extracted from the SRTM DEM in this study, it lacked the spatial resolution to take smaller forms into account. Due to the possibility of direct extraction of dune morphology, the first attempt in this study was to create a DEM from Sentinel-1 SLC imagery. Due to very low coherence, however, this was not possible in the active parts of the dune field. Instead, the dune morphology was extracted indirectly from GRD imagery. Although this methodology brings some challenges in the analysis and interpretation, the multi-temporality was favoured over the direct dune form extraction.

As the advantages of complex dune field analysis with Sentinel-1 GRD are mainly the easy differentiation of the mentioned features and high spatial and temporal resolution, we see its main value in the automated derivation of dune features. For example, the continuous automated derivation of dune ridges and active dune sand in combination with wind datasets could bring new results for the annual variability of dune movement and the response to storm events.

## 3.5 Conclusion

In this study, we have shown the possibilities as well as limitations in using the currently most easily accessible SAR dataset, Sentinel-1 GRD, to access the morphology of a complex dune field and changes therein. As the very low areal coherence within active complex dune fields does not allow for interferometric approaches such as the calculation of a DEM, we used the GRD backscatter values to extract morphological features indirectly. Using this method, three main dune features can be extracted based on surface roughness and local incidence angle. These are namely active dune sand, dune ridges and inter-dune ripples.

Based on this information, a comparative analysis to optical Sentinel-2 imagery and the SRTM DEM as well as a profile-based continuous wavelet transform frequency analysis were conducted for the complex dune field Bor Khyar Els in western Mongolia. Using these methodologies, three areas of interest within the dune field were analysed, revealing significant differences in aeolian activity, wind direction and wind seasonality.

Based on the results of this study, we see a good suitability of Sentinel-1 for complex dune field analysis. This suitability is additionally amplified by the comparison to similar approaches based on optical imagery and DEMs. In comparison to these datasets, Sentinel-1 allows for a far better detection of dune-relevant morphological features. In comparison to optical imagery where these features only show slight differences in their reflectivity, SAR-based analysis offers an easy discrimination. The main advantage over globally available DEMs is the comparably high spatial resolution and especially the multi-temporality.

Due to these advantages of SAR-based complex dune field analysis, we see great potential in its further implementation. For future studies, we recommend its implementation in automated analyses of continuous time series. Due to the high temporal resolution of Sentinel-1, this would offer new insights into the temporal and spatial evolution of dune fields as well as the related frequencies and magnitudes of the aeolian driving forces.

# 4. Upper Palaeolithic site probability in Lower Austria – a geoarchaeological multi-factor approach

Bruno Boemke[1,3], Thomas Einwögerer[2], Marc Händel[2], Frank Lehmkuhl[1]

[1] Department of Geography, RWTH Aachen University, Aachen, Germany

[2] Austrian Archaeological Institute, Austrian Academy of Sciences, Vienna, Austria

[3] Corresponding author. E-Mail: bruno.boemke@geo.rwth-aachen.de

This chapter was published as the following article: Boemke, B.; Einwögerer, T.; Händel, M.; Lehmkuhl, F. (2022): Upper Palaeolithic site probability in Lower Austria – a geoarchaeological multi-factor approach. In Journal of Maps 18 (4), pp. 610–618. DOI: 10.1080/17445647.2021.2009926.

As the first and corresponding author, BB was responsible for the main investigation, methodology, data analysis, visualization, data curation, writing of the manuscript and revision. The second author, TE, contributed in the fields of revision, conceptualization, and data curation. The third author, MH, contributed in the areas of writing, revision, conceptualization, supervision and data curation. The last author, FL, contributed in the areas of conceptualization, revision, supervision and funding acquisition.

## Abstract

In archaeology, predictive models play a key role in understanding the interactions between humans and the paleo-environment. They are also of great value for cultural heritage management and planning purposes. This is particularly true for Palaeolithic sites in the east Austrian loess landscape, which are often deeply embedded in sediment sequences. In this study, we analyse the geospatial behaviour of 23 Upper Palaeolithic sites in Lower Austria. Hereby, we apply a new approach, which combines the advantages of a classical deductive method with the capabilities of machine learning, implemented via the MaxEnt software. The result is a predictive model for an area of 7850 km², exploring the potential for the presence of Upper Palaeolithic sites. The model highlights several spatial dynamics of site probability in the study area. Possible sources of inaccuracies within the source data and the methodology are critically discussed.

## 4.1 Introduction

Archaeological predictive modelling (APM) tries to predict "the location of archaeological sites or materials in a region, based either on a sample of that region or on fundamental notions concerning human behaviour" (Kohler and Parker 1986). It is based on the assumption that human spatial behaviour is predictable and can be extrapolated from samples to larger areas (Verhagen 2007). As 90-99% of all archaeological remains are presumably yet undiscovered, APM is based on the 1-10%

explored and documented archaeological sites (Verhagen et al. 2010). As such, an APM offers the chance of preserving the 90%-99% undiscovered sites from destruction by e.g. construction projects while being based on only 1-10% discovered archaeological sites, raising the question of representability. In addition to cultural heritage management implementations, APM can help us in understanding the interaction between our ancestors and the paleo-environment (Kamermans & Niccolucci 2010). For a more in-depth introduction to APM as well as a thorough discussion of its strengths and weaknesses, see van Leusen et al. (2005) and references therein.

This study is based on 23 Upper Palaeolithic sites in Lower Austria and presents a first attempt at creating an APM for a Late Pleistocene timeframe in this area. Despite the limited number of sites, the model, in its current state, can already be used as a regional geospatial tool for archaeological research. We recommend implementation into cultural heritage management only after additional empirical data is added and the model is thoroughly validated. To create the model, we use a novel approach, combining the explanatory power of a deductive method with the statistical reliability of an inductive approach. The result can be seen as part of the "Middle Range Theory", as it breaks up the boundaries between deductive and inductive methodologies (Verhagen & Whitley 2012).

## 4.2 Study Area

The study area covers more than 7800 km² and stretches from the Bohemian Massif to the Eastern Alpine forelands. Geologically, it marks the transition between massif metamorphic rocks to molasses, covered by Pleistocene sediments. Hydrologically, the Danube leaves the narrow valleys of the Bohemian Massif and enters the hilly landscape of the Alpine foreland. For the Late Pleistocene paleo-environment, this translates to the transition from a turbulent, impassable river into a traversable braided river system, documented by the Late Pleistocene fluvial deposits of the lower terrace (Geologische Bundesanstalt 2013). This part of Lower Austria is widely considered a loess landscape (Lehmkuhl et al. 2021). Regarding the disputed definition of loess, mentions in this study include windblown aeolian dust as well as results of post-formational activities as proposed by Smalley et al. (2011). Up to 40 m thick loess-paleosoil sequences (LPS) can be found in the study area, several of which enabled in-depth paleo-environment reconstructions (e.g. Haesaerts et al. 1996; Terhorst et al. 2011; Sprafke et al. 2014; Sprafke et al. 2020). At the same time, numerous LPS preserved evidence for Palaeolithic occupation. Today's topography, however, is considerably impacted by historic and recent anthropogenic activity, such as large-scale terracing for viticulture, loam extraction and constructions for transportation. Many archaeological sites in the study area were discovered during such activities, potentially causing a considerable sampling bias.

## 4.3 Materials and methods

### 4.3.1 Upper Palaeolithic sites

Several dozen Palaeolithic sites have been discovered in the study area in course of ~150 years of research. A number of these such as Willendorf, Krems-Hundssteig, Krems-Wachtberg, Kammern-Grubgraben and Langmannersdorf are internationally well-known as they provide significant contributions to our understanding of early modern human behaviour and substantial evidence for human-environment interaction (e.g. Einwögerer et al. 2006; Händel et al. 2020; Neugebauer-Maresch 2008; Nigst et al. 2014; Teschler-Nicola et al. 2020). Not all sites are suitable for consideration in an APM. In some cases, the material evidence allows neither for an unambiguous techno-cultural placement nor for the production of numerical ages (i.e. lack of organic remains for radiocarbon dating and heat-influenced lithic objects for luminescence measurements). For other sites, the exact geographic position remains unknown, because the archaeological material was either collected from the surface (i.e. devoid of sedimentary context), or the precise spatial context was not documented. Cave sites were also not considered, as they underlie entirely different formation processes. Hence, 23 Upper Palaeolithic open-air sites remained to be included in this study (Table 4). Techno-culturally, these sites range from the Aurignacian to the Magdalenian. 19 sites provided numerical ages in the range between 43 and 17 ka cal BP; the other four allow for unambiguous techno-cultural placement.

*Table 4: List of all utilized Upper Palaeolithic sites within the study area.*

| Site name | Lat | Long | Elevation | Techno-culture | Age (ka cal BP)* |
|---|---|---|---|---|---|
| Aggsbach | 48.2933 | 15.4002 | 224 | Gravettian | 30 |
| Alberndorf | 48.6833 | 16.1120 | 252 | Aurignacian | 36-32 |
| Getzersdorf | 48.3280 | 15.6925 | 243 | Aurignacian | N/A |
| Gobelsburg-Zeiselberg | 48.4564 | 15.7003 | 229 | Gravettian | N/A |
| Gösing-Setzergraben | 48.4658 | 15.8068 | 316 | Gravettian | 31 |
| Großweikersdorf I/III | 48.4684 | 15.9756 | 215 | Aurignacian & Epigravettian/LGM | 37 / 24.5 |
| Kammern-Grubgraben | 48.4825 | 15.7175 | 267 | Epigravettian/LGM | 23.5-22 |
| Kamegg | 48.6128 | 15.6596 | 277 | Magdalenian | 17.5-17 |
| Krems-Wachtberg 1930 | 48.4153 | 15.5992 | 258 | Gravettian | 32-31 |
| Krems-Wachtberg 2005-2015 | 48.4150 | 15.5995 | 257 | Gravettian | 34-30.5 |
| Krems-Hundssteig 2000-2002 | 48.4146 | 15.6015 | 237 | Gravettian | 34-31 |
| Krems-Wachtberg East | 48.4154 | 15.5999 | 254 | Gravettian | 34-30.5 |

| | | | | | |
|---|---|---|---|---|---|
| Krems-Hundssteig 1893-1904 | 48.4146 | 15.6015 | 237 | EUP (Early Upper Palaeolithic) / Aurignacian | 41 |
| Langenlois A/B | 48.4681 | 15.6863 | 222 | Gravettian | 31-29 |
| Langmannersdorf | 48.2805 | 15.8340 | 220 | Epigravettian/LGM | 25-23 |
| Rosenburg | 48.6340 | 15.6342 | 264 | Epigravettian/LGM | 24 |
| Rupperstal | 48.4693 | 15.9331 | 307 | nd | 25.5 |
| Saladorf | 48.2724 | 15.8768 | 204 | Epigravettian/LGM | 22 |
| Senftenberg | 48.4537 | 15.5411 | 290 | Aurignacian | 39 |
| Spitz-Singerriedl | 48.3667 | 15.4182 | 211 | Gravettian | N/A |
| Steinaweg | 48.3711 | 15.5990 | 274 | Gravettian? | N/A |
| Stratzing/Krems-Rehberg | 48.4407 | 15.6029 | 356 | Aurignacian | 39-33 |
| Willendorf II | 48.3232 | 15.4042 | 232 | (Middle Palaeolithic) - EUP - Aurignacian - Gravettian | (46)43-29 |
| *Radiocarbon data sources: Einwögerer et al. (2014); Händel et al. (2020); Jöris et al. (2010) and references therein; Nigst et al. (2014). Calibration with IntCal20 (Reimer et al. 2020). Ages given represent mean values rounded to 0.5 ka. | | | | | |

### 4.3.2 Input predictors

In APM, environmental variables with implications for site probability are called predictors, as they are used to predict the dependent variables (Wheatley 1996). When assessing the potential for archaeological sites, not only factors influencing the settlement choice have to be considered. Natural factors influencing the preservation of sites such as sedimentation processes are at least equally important (Schiffer 1983). In particular, for the often sparse remains of hunter-gatherer societies, a separation of anthropogenic and natural factors is crucial not only for understanding the formation of a single site (Schiffer 1983) but also for assessing settlement and/or occupational patterns (Binford 1980). At the same time, the archaeological record can be indicative for natural processes, i.e. paleo-environmental conditions, leading to the site's preservation (e.g. Händel et al. 2021). For our study we identified 10 geospatial predictors. From a thematic perspective, these can be divided into three fields: terrain, water and geology. Table 5 lists all predictors together with implications for settlement choice and preservation of records.

*Table 5: Chosen predictors with implications for settlement choice and preservation as well as implementation method.*

| Predictor | Implications for settlement choice | Implications for preservation | Implementation via |
|---|---|---|---|
| Elevation | Higher altitudes mean lower temperatures. Thus, lower altitudes are favoured. | Elevations below the erosional base level are more likely to have been disturbed. There is also an upper boundary for loess deposition. | Digital elevation model (DEM) |
| Slope | Gentle slopes mean higher sunshine duration. Steep slopes do not allow for settlement. | Geomorphological slope processes increase sedimentation dynamics, potentially leading to local erosion but also to locally higher sedimentation rates and thus improved preservation. | Calculated from DEM |
| Aspect | Leeward slopes are preferred as they offer wind protection. | Leeward slopes favour sedimentation of aeolian sediments. | Calculated from DEM |
| Topographic position | Ridges and plateaus are avoided as these are exposed to the wind. Lower slope sections are more protected. | Ridges and plateaus favour aeolian erosion. Lower slope sections favour sedimentation. | Calculated from DEM |
| Sunshine hours | Longer sunshine duration increases the local temperature. | None | Calculated from DEM |
| Distance to river | Rivers were important water (and raw material) sources. Short distances are preferred. | Braided river systems with large seasonal differences in discharge act as main source for aeolian sediments. | Euclidean distance to river datasets |
| Distance to river junction | Locations close to river junctions were favoured. | None | Euclidean distance to extracted junctions of river datasets |
| Height above water level | Pleistocene rivers posed a constant threat of flooding. A certain height above the water level offered protection. The seasonal character of sites diminishes this risk. | Sites within reach of fluvial activity are not preserved. This is also relevant for the distance to river predictor. | Subtraction of interpolated water levels from DEM |
| Loess sediments | Loess landscapes supported a rich flora and fauna crucial for hunter-gatherer subsistence. | Deposition of loess sediments favours preservation. | Thematic maps |
| Late Pleistocene alluvial plains | Late Pleistocene rivers posed a constant threat of flooding. | Possible remains are displaced or destroyed by fluvial processes. No sites were found here. | Thematic maps |

All predictors were parameterised using ArcGIS 10.7.1. The main source for most terrain- and water predictors is a digital elevation model (DEM) with a spatial resolution of 10 m (available at www.data.gv.at). As anthropogenic landscape modifications (see section 4.2) imply necessity for terrain corrections, this dataset was smoothed prior to additional processing. The goal was to create a DEM which represents a natural landscape as closely as possible. The resulting DEM still differs from a Late Pleistocene paleo-surface, as climatic, geomorphological and hydrological processes have since fundamentally changed; Holocene sediments covered past stream channels, steep hills wore down, rivers migrated, etc. Late Pleistocene landscape elements cannot be easily recreated and the smoothed DEM can therefore only be considered a best guess. For the smoothing algorithm, a focal mean within a 5-cell radius circle kernel was calculated using the '*Focal Statistics'* tool.

The first predictor, elevation, is represented by the smoothed DEM. Slope and aspect were determined (tools *'Slope'* and *'Aspect'*). For the topographic position, we calculated the topographic position index (TPI) as defined by Gallant & Wilson (2000). Within ArcGIS, this was done by calculating the average elevation around each pixel (tool *'Focal Statistics'* and subtracting the resulting raster from the original DEM (tool *'Raster Calculator'*). As large radii mainly reveal major landscape units (Reu et al. 2013), a focal mean within a 10-cell radius kernel was used for this calculation. The last terrain predictor, annual sunshine duration, was calculated from the DEM (tool *'Area Solar Radiation'*).

For all predictors connected to water, we used two datasets, representing different scales. Streamlines from the CCM21 European river dataset represent larger rivers (version 2.1, De Jager & Vogt, 2007). As smaller streams also might have been vital for freshwater supply in the Upper Palaeolithic, an additional drainage was derived from the smoothed DEM. This was done by filling the DEM (tool *'Fill'*) and calculating flow direction (tool '*Flow Direction'*) and flow accumulation (tool '*Flow Accumulation'*). To include smaller streams, we used a low threshold of 5000 cells within the flow accumulation when extracting streamlines. To parameterise the distance to these river datasets, Euclidean distance was calculated (tool *'Euclidean Distance'*). Cost distance was not chosen for several reasons. Firstly, cost distance is not associated with a unit, which is disadvantageous for comprehension. Secondly, due to short distances to potential freshwater sources, a cost raster only has minor impact, especially concerning individual mobility on foot. Parameterisation of the predictor 'height above water level' was based on points extracted along streamlines in an interval of 50 meters (tool *'Generate Points Along Lines'*). In a second step, the elevation value was added to streamline points, representing the water level at each respective point (tool *'Extract Values to Points'*). For extrapolation of these point values, we used the inverse distance weighting interpolation algorithm with a search radius of 100 points (tool *'IDW'*). In a final step, the resulting raster was subtracted from the DEM to gain the effective height above water level in meters (tool *'Raster Calculator'*).

For the distribution of loess and Late Pleistocene alluvial plains, the open source dataset of Lehmkuhl et al. (2021) was used. To ensure that it meets the required accuracy for this small-scale application, the dataset was compared to a 1:50,000 geological map (Geologische Bundesanstalt, 2013). Minor corrections were made by adjusting the area covered by Late Pleistocene alluvial plains to fit the respective terrace identified in the DEM.

### 4.3.3 Implementation of MaxEnt

MaxEnt was developed as a software for species distribution and environmental niche modelling (Phillips and Dudík 2008; Phillips et al. 2017). The main output of the MaxEnt tool is a raster containing the rate of occurrence (ROR), which, in this case, translates to the predictive probability for Upper Palaeolithic sites. This probability is calculated individually for each cell from the RORs assigned to the underlying predictor-values. For an in-depth explanation of the statistics behind the MaxEnt software, see Elith et al. (2011) and Merow et al. (2013). Due to its high predictive accuracy and easy-to-use application, it has been used in many fields, including archaeology. One of the main advantages of MaxEnt over the predominant technique, the *logistic regression*, is the fact that MaxEnt utilized presence-only data (Wachtel et al. 2018). Within logistical regression approaches, this can lead to difficulties in estimating an adequate sample size and site density (Verhagen 2008). By now, many APM case studies have successfully implemented the MaxEnt software (Galletti et al. 2013; Gillespie et al. 2016; Jones et al. 2019; Alwi Muttaqin et al. 2019) and some have made direct comparisons, showing that MaxEnt outperforms logistic regression (Wachtel et al. 2018; Yaworsky et al. 2020). Despite all advantages of MaxEnt, it still is an inductive model at its core. Therefore, an indiscriminate implementation of the model leads to a loss of the explanatory value the same predictors would provide in a deductive model (Ebert 2004). To address this issue, MaxEnt includes response curves in the output, which show the predictive probability that is associated with the values of each predictor. These can be used to evaluate the thematic plausibility and thereby assess the causality within the model (Merow et al. 2013). As there is no way to alter the workflow of the model based on the plausibility of the response curves, these were the only outputs extracted for further use as a statistical substructure of a deductive approach. This is the only way to preserve causality between input predictors and the dependent variables and thereby uphold the explanatory value of the model. For a more in-depth comparison of weaknesses and strengths of inductive vs. deductive APMs, see van Leusen et al. (2005), Verhagen & Whitley (2012) and references therein.

To fittingly integrate loess and Late Pleistocene alluvial plains into MaxEnt, these predictors were transformed into binary format (1=present/0=not present). As such, these predictors were integrated

as categorical predictors while all other predictors are continuous. To determine fitting input parameters, the practical guide to MaxEnt by (Merow et al. 2013) was used. As the goal of this approach was to identify plausible optimal value ranges within the response curve of each predictor separately, cumulative output was selected. This output rescales the ROR from lowest to highest value on a scale between 0 and 100, which allows for unambiguous interpretation and easy comparison (Merow et al. 2013). The loss of raw probability values is acceptable as these are of negligible importance in a deductive approach.

### 4.3.4 Deductive method

Subsequently, we evaluated the plausibility of the statistical connection between each predictor and the site probability, represented by the response curves. In cases where the response curves fit the deductive expectation, no modification was made before reclassifying the value of the predictor to a predefined score value. In cases where parts of the response curve were deemed implausible, modifications were made accordingly. All response curves and modifications are presented in section 4.4. After obtaining plausible response curves for each predictor, these were reclassified into score values to simplify the model and thereby enhance its explanatory value. To achieve this, the values of each predictor were reclassified into three classes according to the cumulative value within the response curve. The **optimal** value range for each predictor is defined by a cumulative value equal to or larger than 80 (score value = 3). This corresponds to the top 20% of relative probability within the predictor. Cumulative values equal to or larger than 50 define the **viable** value range of each predictor (score value = 1). Values below this relative probability threshold are classified as **unviable** (score value = 0). The large difference between the score value of the optimal and viable class (3 vs 1) was chosen to highlight the large difference in cumulative relative probability (top 20% vs top 50%). For the binary categorical predictors loess and Late Pleistocene, a different approach was used. For loess, presence was defined as optimal (score value = 3) and absence as unviable (score value = 0). As Late Pleistocene alluvial plains were defined as exclusion criterion, presence was defined as unviable (score value = 0) and absence as viable (score value = 1). Through this simplifying score value method, the value range was reduced from 0-900 to 0-27, reducing the complexity immensely.

## 4.4 Results

Here, we present the results of the MaxEnt implementation and how the response curves were evaluated and integrated into the deductive approach (see Figure 13). As the TPI yielded very low model importance values (permutation importance = 0.0098, Jackknife standalone training gain =

0.0235), implying the absence of a statistical connection to the probability of sites, this predictor was excluded. All other predictors were kept as they showed satisfactory importance values.

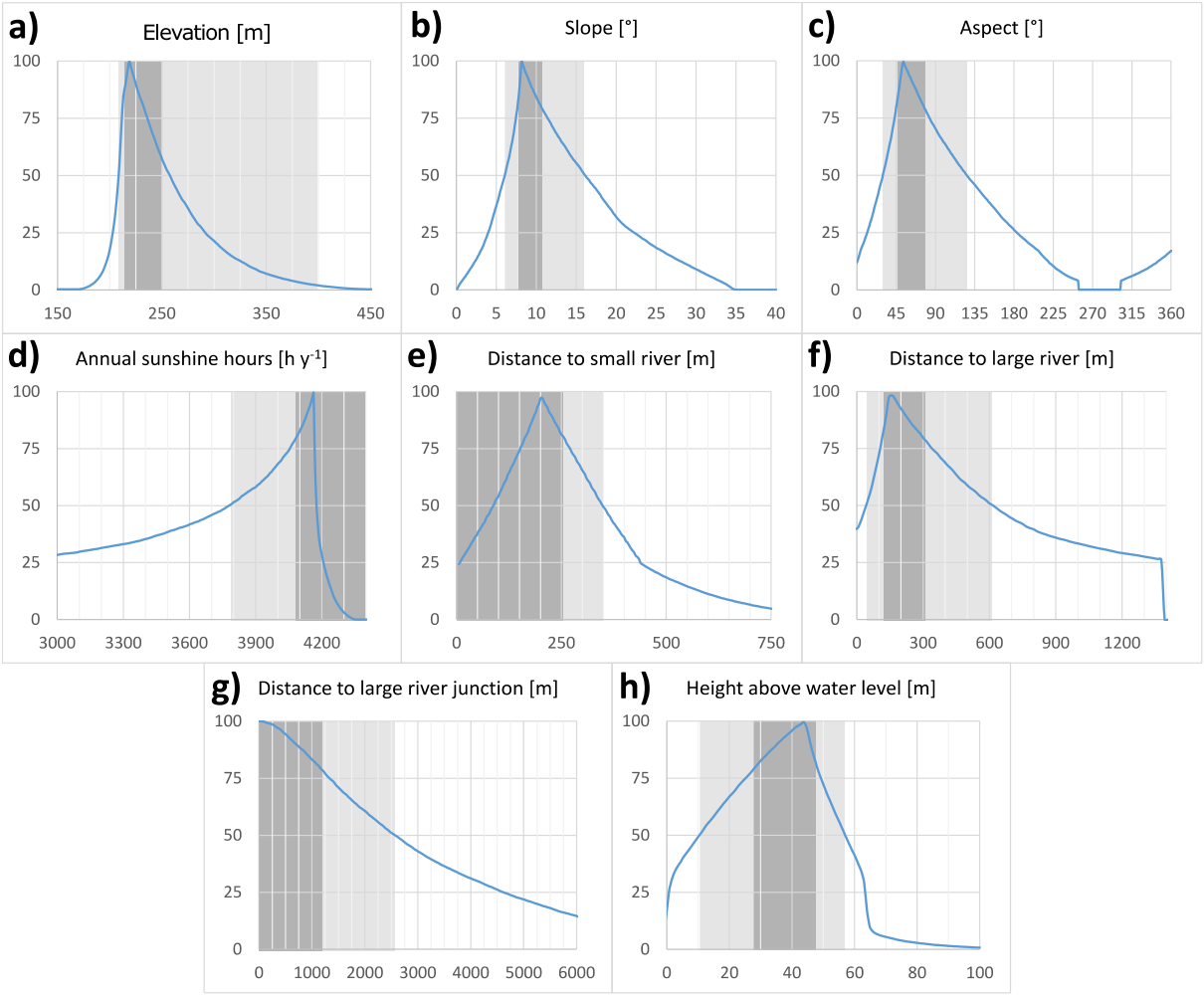



*Figure 13: Response curves received from MaxEnt, version 3.4.4, cumulative output. Optimal (dark grey) and viable (light grey) value ranges of each predictor, based on the archaeological evaluation of the response curves, are marked in grey.*

The response curve of the elevation predictor **a)** was evaluated as being partly implausible. The sharp decline in relative probability below the elevation of 220 m can be explained geomorphologically, as this value corresponds to the erosional base level. The steep decline of probability above 220 m, however, was deemed implausible, as no implications for settlement choice or preservation support this. The generally narrow optimal and viable value ranges resulting from the unmodified response curve can be attributed to the small sample size of only 23 sites, 80% of which are found on an elevation between 204 and 277 m. Factors influencing the upper boundary of viability for settlement are temperature from the settlement choice perspective and the upper boundary of loess deposition from a preservation perspective (see Table 5). As such, the viable value range was set to ≤400 m, which represents the upper boundary for loess accumulation according to Lehmkuhl et al. (2021) and the

optimal range to ≤250 m based on empirical data. The statistical connection between relative probability and slope, represented by response curve **b)** was deemed plausible. The optimal and viable slope values represent hillsides where geomorphological processes favour the preservation of sites. It is important to note, that the implications for preservation overshadow the implications for settlement choice. The response curve of the aspect predictor **c)** was deemed plausible, as the highest relative probabilities are found in northeast/east aspects, i.e. in the lee of the prevailing westerly winds (Sebe et al. 2015). As such, it fulfils all implications for settlement choice and preservation, listed in Table 5. The exponential incline of relative probability with increasing annual sunshine hours in response curve **d)** was evaluated as plausible. Only the rapid decline above 4150 annual sunshine hours was deemed implausible, as no implications for settlement choice or preservation support this. This implausibility in the response curve might again be based on the small sampling size, including no sites with more than 4250 yearly sun hours, although close to 4400 hours are theoretically possible in these latitudes. As such, the optimal value range was extended upwards, ignoring this possibly biased decline in relative probability. Response curve **e)**, representing distances to small rivers was evaluated as partly implausible, as a decline of relative probability below 200 m distance is not expected. A possible explanation for this inconsistency might be the aforementioned sampling bias (4.2), as areas in direct vicinity of rivers are less likely utilized for viticulture, loam extraction or construction. Therefore, optimal conditions were assumed for distances below 200 m. However, depending on catchment size, large variations of discharge, justifying this lower threshold, cannot be completely ruled out. Although it bears a great resemblance to the partly implausible response curve of small rivers, response curve **f)**, representing distance to large rivers, was deemed plausible. The decline in probability below 200 m distance can be explained by the large variations of discharge of these rivers, posing a threat of flooding to camp sites and even more importantly providing unfavourable conditions for site preservation. The response curve for the distance to large river junctions **g)** was deemed plausible, as it shows the expected steady decline of relative probability with increasing distance. Response curve **h)** was also evaluated as plausible, as both incline and decline of the curve can be explained in the context of the paleo-environment. The incline with increasing distance expresses the safety from flooding on the one hand while the decline at distance values above 45 meters shows decreasing accessibility on the other hand. On the basis of this evaluation, all predictors were reclassified into optimal, viable and unviable value ranges.

An additive approach was chosen to combine all reclassified predictors into a predictive model. To acknowledge the absolute exclusive criterion of Late Pleistocene alluvial plains, this predictor was included multiplicatively at the end of the equation. The reason that this geological predictor is treated as an absolute exclusive criterion is empirical evidence that no sites are documented. As nine

reclassified predictors with score values between 0 and 3 are included additively, the resulting additive score value can range from 0-27. Within this score range, very high probability for a presence of Upper Palaeolithic sites is assumed when more than half of the predictors lay within their optimal value range (additive score value ≥15); high probability when all predictors are at least within their viable value range (≥9); medium probability when more than half of the predictors lay within their viable value range (≥5); and low probability beneath this threshold (<5).

The map itself shows multiple trends in probability (see Appendix B). On a larger scale, a difference between the Bohemian Massif in the west and the Eastern Alpine forelands in the east is apparent, showing generally higher probabilities in the east. On a smaller scale, medium steep slopes near rivers and river junctions show highest probabilities. Within this trend, probabilities on eastern and north-eastern aspects are also elevated (Figure 14).

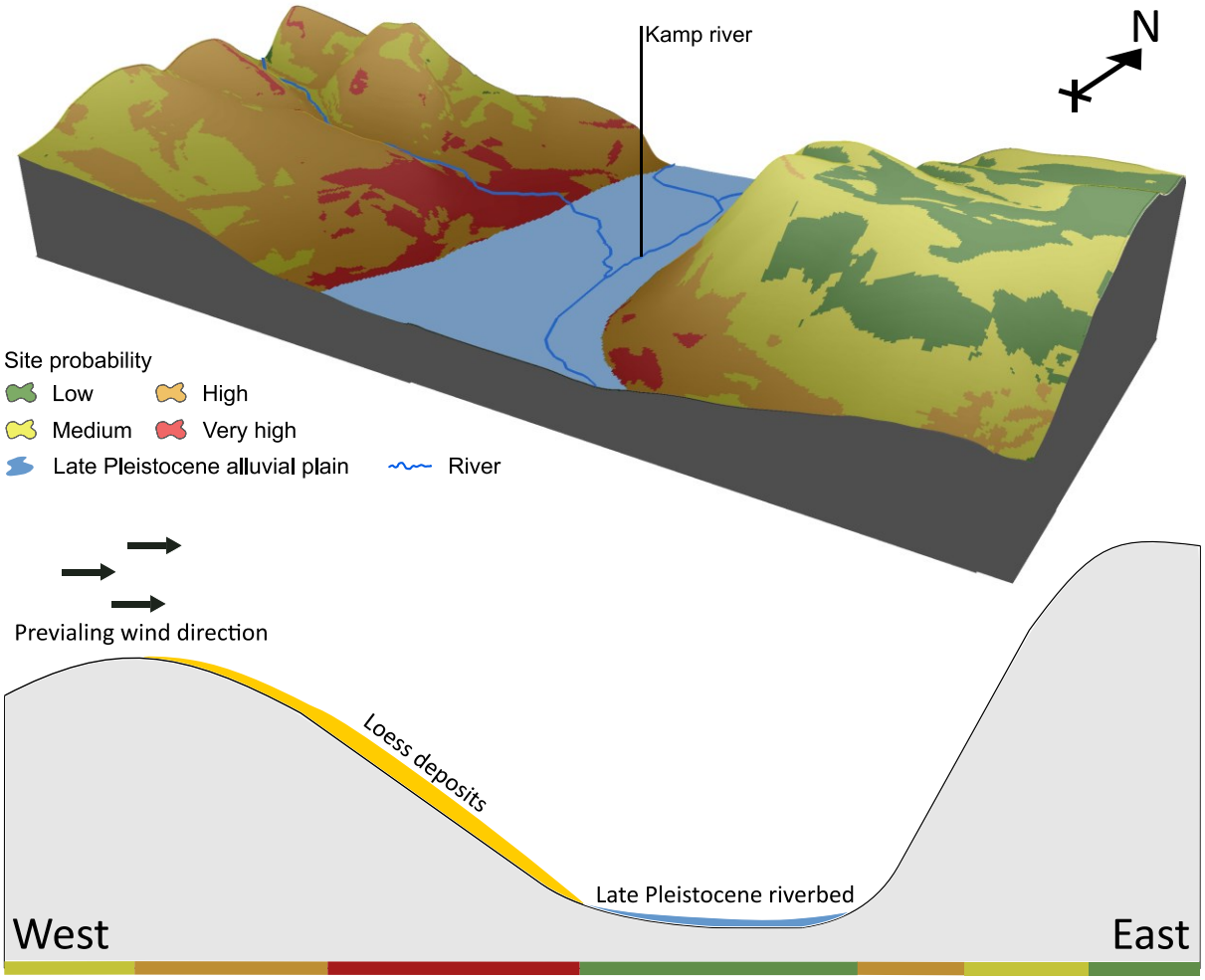



*Figure 14: Representative 3D image (upper) and idealized cross-section (lower), showing Upper Palaeolithic site probability dynamics in river valleys. The 3D image represents the area north of Langenlois at the intersection between Kamp river and Fahnbach river (See main map for localization)*

The resulting model was validated with the input Upper Palaeolithic sites, showing that more than 80% of the sites fall within very high and high probabilities (48% very high, 34% high, 18% medium) and are thus represented by the predictive model. A cross-validation with test sites could not be conducted as the sample size is too small to be split into training and validation datasets.

## 4.5 Discussion

A predictive model can only be as accurate and representative as the data it is based on. Therefore, possible sources for inaccuracies and errors require careful consideration and critical discussion. All predictors trying to depict the Late Pleistocene environment can only be an estimate, as geomorphological and especially hydrological processes have changed fundamentally. Smoothing the DEM helps removing small scale anthropogenic features, while some Late Pleistocene landforms cannot be recreated in this way as mentioned above (section 4.3.2). Even larger differences are expected regarding hydrology, in particular due to modern canalization of large rivers. Present-day streamlines are thus not representative for Upper Palaeolithic settings. To account for potential river locations, Late Pleistocene alluvial plains were integrated and treated as possible streamline location when creating the distance to river predictors.

Possible inaccuracies can also be expected from the data representing loess and Late Pleistocene alluvial plains, as these cover a European scale (Lehmkuhl et al. 2021). To minimize scale-related inaccuracies, the mentioned DEM and regional geological map were used as reference for visual validation and adjustment. Nevertheless, only half of the sites were found covered by loess shapefiles. This stands in contrast to the sedimentological context of the sites, as all are embedded in loess. It is therefore safe to say, that loess is underrepresented in geological maps of the study area. This may be caused by the exclusion of sediment covers <2 m during geological mapping (Lehmkuhl et al. 2021).

Some possible limitations of the model are also introduced by the archaeological dataset. On one hand, the dataset does not always allow for unambiguous assessment of potential duration of the sites' occupation (e.g. recurring brief or seasonal occupation vs. longer stay), although this is expected to have an impact on the preferred position within the paleo-environment. On the other hand, sites with several occupations spanning up to 25,000 years were treated as one, ignoring possible changes in settlement preferences. The only way to address these issues is to enlarge the dataset, so that it can be split into type- and chronology-specific groups without reducing the sample size below a representative minimum.

Possible sources for sampling bias regarding the site locations should also be considered to assess the representability of the dataset. This is especially important in the context of a loess landscape. As potential archaeological sites are often deeply embedded in loess sequences, discoveries are less likely to be random, but instead often connected to construction, viticulture or quarries. This leads to a significant sampling bias, which was addressed by careful selection of only well-documented and representative sites. However, this reduced the sample size to only 23, leading to difficulties in validation. As such, the model was validated with the training sample at the cost of weakening the statistical strength of the validation result.

Despite these limitations, we are able to show that predictive modelling for Upper Palaeolithic sites in the study area provides plausible results.

## 4.6 Conclusion

We introduce a new methodology for APM, which allows deductive models to profit from inductive analysis. As such, it represents a good addition to the "Middle Range Theory" in archaeology, breaking up boundaries between deductive and inductive approaches. In detail, the statistical connections between 10 environmental predictors and Upper Palaeolithic site probability in Lower Austria were assessed, using the modelling software MaxEnt. The selected predictors have implications for the settlement choice and the site preservation. Instead of letting the MaxEnt "black box" run its course, losing track of causality in the process, intermediate results (response curves) were evaluated for plausibility from an archaeological perspective before further processing. All predictors were then reclassified to optimal, viable and nonviable value ranges and combined into a predictive model in an additive approach. The resulting map highlights spatial dynamics both on a larger and smaller scale. The small sample size, however, does not allow for complex validation and raises the question of adequate representability. As such, the model, in its current state, is applicable for scientific applications only. When more empirical data is added and the model is thoroughly validated, an implementation into cultural heritage management is possible.

## 4.7 Software

For mapping, processing and statistical analysis of spatial data, we used ESRI ArcGIS 10.7.1. For the inductive parts of the APM, MaxEnt, version 3.4.4 was utilized. For the post-processing of maps and creation of graphics, we used Inkscape, version 1.1.

## 4.8 Data availability

The complete dataset raised in this study is publically available at the Collaborative Research Centre 806 (CRC806) database (DOI: 10.5880/SFB806.71).

# 5. Approaching sampling biases of Upper and Final Palaeolithic sites – a geospatial analysis of a European dataset

Bruno Boemke[1,3], Andreas Maier[2], Isabell Schmidt[2], Wolfang Römer[1], Frank Lehmkuhl[1]

[1] Department of Geography, RWTH Aachen University, Aachen, Germany

[2] Institute for Prehistoric Archaeology, University of Cologne, Cologne, Germany

[3] Corresponding author. E-Mail: bruno.boemke@geo.rwth-aachen.de

This chapter was published as the following article: Boemke, B.; Maier, A.; Schmidt, I.; Römer, W.; Lehmkuhl, F. (2023a): Testing the representativity of Palaeolithic site distribution: The role of sampling bias in the European Upper and Final Palaeolithic record. In Quaternary Science Reviews 316. DOI: 10.1016/j.quascirev.2023.108220.

As the first and corresponding author, BB was responsible for the main investigation, methodology, data analysis, visualization, data curation, writing of the manuscript and revision. The second author, AM, contributed in the fields of writing, revision, conceptualization, validation and supervision. The third author, IS, contributed in the areas of writing and data curation. The fourth author, WR, contributed in the areas of methodology, software and data analysis. The last author, FL, contributed in the areas of conceptualization, revision, supervision and funding acquisition.

## Abstract

Archaeological sites are not distributed evenly throughout the landscape. For the Palaeolithic record, signals derived from the inhomogeneous spatial patterns are used to infer spatial decision-making processes or ecological preferences of our ancestors. However, to date it is still largely unclear how sampling biases affect the large-scale distribution of sites and whether the observable spatial patterns are actually representative of the distribution of humans in the paleo-landscape. To answer this question, this study assesses the spatial distribution of 4200 Upper and Final Palaeolithic occupations from two different perspectives, i.e., past settlement choice and likelihood of discovery. On the one hand, site distribution is thus examined for settlement-relevant factors such as topography, geology and sedimentology. On the other, discovery-relevant biases, such as recent land cover and building activity are analysed. The comprehensive spatial and statistical assessments show that the actual distribution of sites seems to be most strongly influenced by sampling biases. The assessed environmental variables representing the settlement factors show a far lower statistical association to the distribution of sites. They do, however, still support several common archaeological assumptions. For all approaches using site distribution as input, such as predictive modelling, the results of this study suggests that the sampling bias must be addressed. To this end, we suggest including environmental variables addressing discovery-relevant factors to quantify the potential biases. For further studies on the sampling bias of Upper and Final Palaeolithic sites, we recommend building on this pilot study by

adding more occupations to the dataset and more environmental variables to the settlement and discovery factors. Due to the current positive trend in openly available geodatasets, we see great future potential in this.

## 5.1 Introduction

The Late Pleistocene history of human settlement in Europe is assumed to be reflected in the distribution of sites left by Palaeolithic hunter-gatherer groups in the landscape. Mapping known sites shows patterns of clusters and voids of human presence across the subcontinent. This patterning – observable for both Neanderthals and Anatomically Modern Humans – has attracted interest of archaeologists, anthropologists and ecologists alike. Since the beginning of archaeological research, scientists try to find explanations for the heterogeneous distribution of sites and to identify decisive factors (Kohler and Parker 1986; Rowland 1989; Balla et al. 2014; Noviello et al. 2018; Verhagen 2018). Over the recent decades, geographic information systems (GIS) have played an increasingly important role in these efforts (Wescott and Brandon 2003; Ebert 2004; Scianna and Villa 2011). Explanatory models developed in prehistoric research often centre around climatic, cultural, and economic factors (Maier et al. 2016, Burke et al. 2017) but also address biases resulting from research history or taphonomic loss, including sediment cover and recent land use (Surovell and Brantingham 2007; Surovell et al. 2009). The evidence from Europe holds a key role in this discussion, since the available record combines a long and intense archaeological research history with a dense set of contextual data (Bocquet-Appel et al. 2005; Stephens et al. 2019).

Human settlement history at larger spatial scales is frequently regarded as being strongly influenced by climatic factors, particularly for prehistoric foragers (Binford 2019). Consequently, paleo-environmental parameters are seen as being highly important for the spatial distribution of Palaeolithic sites (Straus 1995; Maier et al. 2022). Conversely, largescale distribution patterns are used to infer past ecological niches or climatic preferences (Banks et al. 2008; Maier et al. 2016). Since the extensive dispersal and continued survival of Anatomically Modern Humans (AMH) in Europe somewhat after 45 ky ago (Hublin et al. 2020; Shao et al. 2021a), the distribution of humans during the late Pleistocene was confined by the glaciation and deglaciation of mountainous areas and the higher latitudes, as well as changes in the spatial distribution and availability of resources. Therefore, many recent studies on Late Pleistocene forager societies draw from recently implemented paleo-climate models to identify and map areas of high settlement potential or ecological niches suitable for human survival (Banks et al. 2009; Banks et al. 2013; Hauck et al. 2018; Wren and Burke 2019; Shao et al. 2021c; Klein et al. 2021).

However, these studies have to rely on the baseline assumption that the distribution of archaeological sites is more or less representative for the actual distribution of humans in the paleo-landscape, and thus reflects either ecological niches or settlement preferences of past foragers. Mismatches observable between the models and the archaeological record already indicate potential conflicts with this assumption (Shao et al. 2021b). Moreover, studies on site formation and the growing integration of geoarchaeological analyses have raised awareness towards geological processes as a filter for archaeological observations. Besides the time-dependent loss of archaeological sites (Surovell and Brantingham 2007; Surovell et al. 2009; Coco and Iovita 2020), regional taphonomic loss due to large scale erosion events (e.g., for the LGM in Iberia, Aura et al. 2012) or invisibility and inaccessibility of sites due to massive build-up of sediments (as for the Carpathian basin, Chu 2018; Nett et al. 2021) become increasingly relevant in the general discussion.

In this study, we examine the distribution of archaeological sites in relation to factors relevant for past settlement choices (e.g., a southward exposure of site location or availability of high-quality lithic raw material) and modern site discovery (e.g., agricultural use of an area or presence of mining pits, Figure 15).

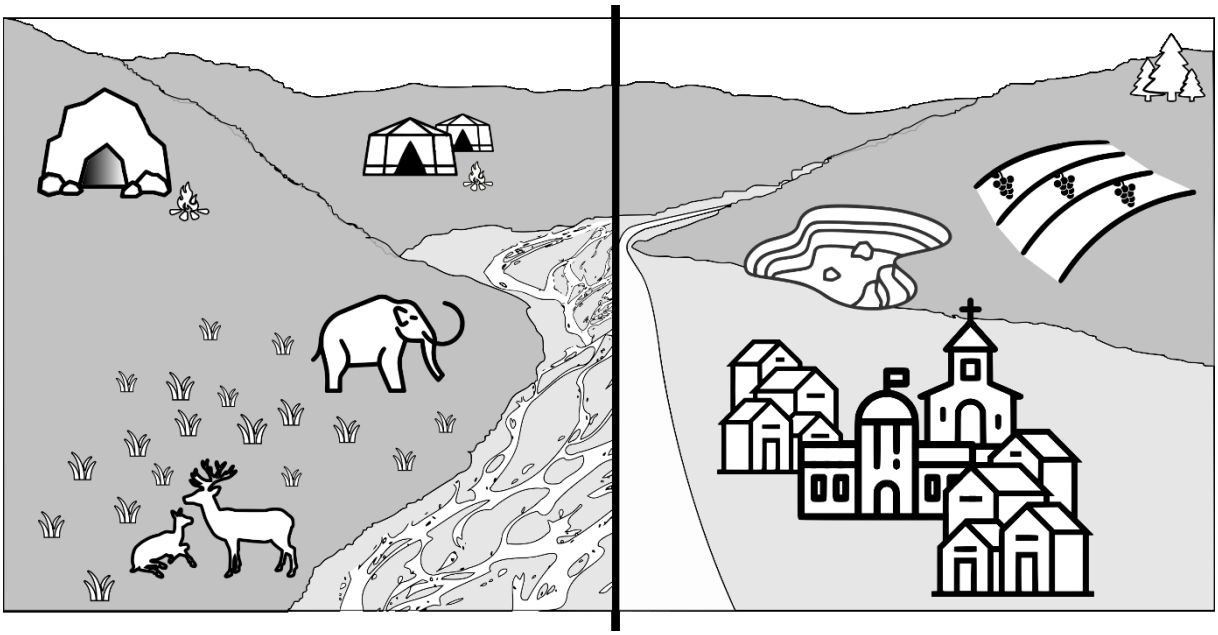


*Figure 15: Graphical illustration of factors relevant for past settlement choices (left) and modern site discovery (right). While the settlement choices influence whether or not archaeological material is present, the modern discovery context influences the chance of discovery. Together, both determine the distribution of known Upper and Final Palaeolithic sites in the current landscape.*

Building on an extensive database that is spatially intersected with a set of environmental variables representing both settlement- and discovery-relevant factors, geospatial and statistical investigations were carried out to answer the following main research questions on a European scale:

1. To what extent do different factors show site frequencies diverging from expected values given the factors' share of the investigated area?
2. Which environmental variables are best suited to predict the presence of Upper and Final Palaeolithic sites?
3. Is it possible to safely differentiate between different taxonomic units, site types (natural shelter/open-air), or regions based on the environmental variables?
4. What is the relationship between the distribution of known sites and environmental variables attributed to settlement and/or discovery factors?

Other questions of equal interest and importance cannot be addressed directly using the given dataset and methods and thus need to be postponed to future studies. For instance, since environmental variables were only assessed at the intersection with known sites, only presence and not absence of archaeological material was investigated. Inferences about whether areas void of sites reflect actual absence of humans or rather a low probability of discovery therefore cannot be drawn conclusively. Predicting areas with a high potential for the discovery of new sites is also out of the scope of this study, because the environmental variables are not suitable to produce valid results for the whole study area. Analysing how sites are distributed within the different environmental settings would require a different method, such as point pattern analysis. It is also important to stress that the above-sated questions are addressed on a European scale. Zooming in on smaller scales might alter the picture.

## 5.2 Materials and methods

### 5.2.1 Upper Palaeolithic sites

The study focuses on archaeological evidence of human occupation assigned to the Upper and Final Palaeolithic of Western and Central Europe, located between the Atlantic and the Black Sea and roughly dated between 42 to 11.6 ka. It uses available datasets on archaeological and paleo-anthropological sites, initially compiled for paleo-demographic studies according to the Cologne Protocol (Zimmermann et al. 2009; Kretschmer 2015; Schmidt et al. 2021b). In these datasets, only published and uncontested sites were considered. The datasets provide spatial information on the occurrence of sites (longitude and latitude, WGS 84), a classification of the type of site (cave, rock shelter, and open-air), as well as a commonly accepted chrono-cultural attribution to six consecutive periods:

a) 42-33 ka, containing sites assigned to the Aurignacian technocomplex (Schmidt 2021);

b) 33-29 ka containing sites assigned to an early phase of the Gravettian technocomplex (Maier and Zimmermann 2016);

c) 29-25 ka, with sites assigned to a later phase of the Gravettian (ibid.);

d) 25-20 ka, which comprises sites of the Solutrean and Epigravettian technocomplexes, dated to the LGM (Maier and Zimmermann 2015);

e) 20-14 ka, comprising sites mainly assigned to the Upper Magdalenian technocomplex (Kretschmer 2015); and

f) 14-11.6 ka, which comprises Final Palaeolithic sites (Schmidt et al. 2021a).

In the database, any archaeological occurrence that can be linked to one of these periods by either absolute dating or typo-chronological attribution is counted as an element of this unit. In stratified sites, different occupations can thus be attributed to different periods. Since we are interested in evaluation the large-scale pattern of presence and absence of occupations, differences in length of the individual occupations are not considered. The database includes close to 4200 occupations. The individual paleo-demographic studies (see: AUR: Schmidt and Zimmermann 2019, GRA: Maier and Zimmermann 2017, LGM: Maier et al. 2016, MAG: Kretschmer 2015, FPAL: Schmidt et al. 2021a) discuss potential biases and limitations of each dataset to assess the reliability of the demographic estimates. Some areas show structural biases for all periods, such as the Po Plain, where important Holocene sediment deposits render potential archaeological sites invisible or inaccessible (Peresani et al. 2021). The situation in the Pannonian Basin or on the Balkan Peninsula seems also highly likely to be influences by taphonomic biases and research intensity in comparison to other areas (Maier et al. 2021). To warrant comparative conditions for a meaningful discussion of observable differences in expected and observed site frequencies per period and setting throughout the investigated area, these regions are not considered further. The remaining map section thus comprises the study area or area of interest (AOI).

Based on a pre-assessment of the dominant sedimentary contexts of sites, the AOI was divided into two sections: The north-eastern section (NE), where the dominating sedimentary context consists of loess and the south-western section (SW), where loess plays little to no role as sedimentary context. The border between these sections was defined based on the loess domains by Lehmkuhl et al. (2021). The border between the sections was aligned with the large European catchments of the Loire, Rhone, and Po rivers on the SW side and Seine, Rhine, and Danube on the NE side. Figure 16 gives an overview of the study area with all included occupations. It is important to note that the number of occupations can differ considerably among the different subsets.

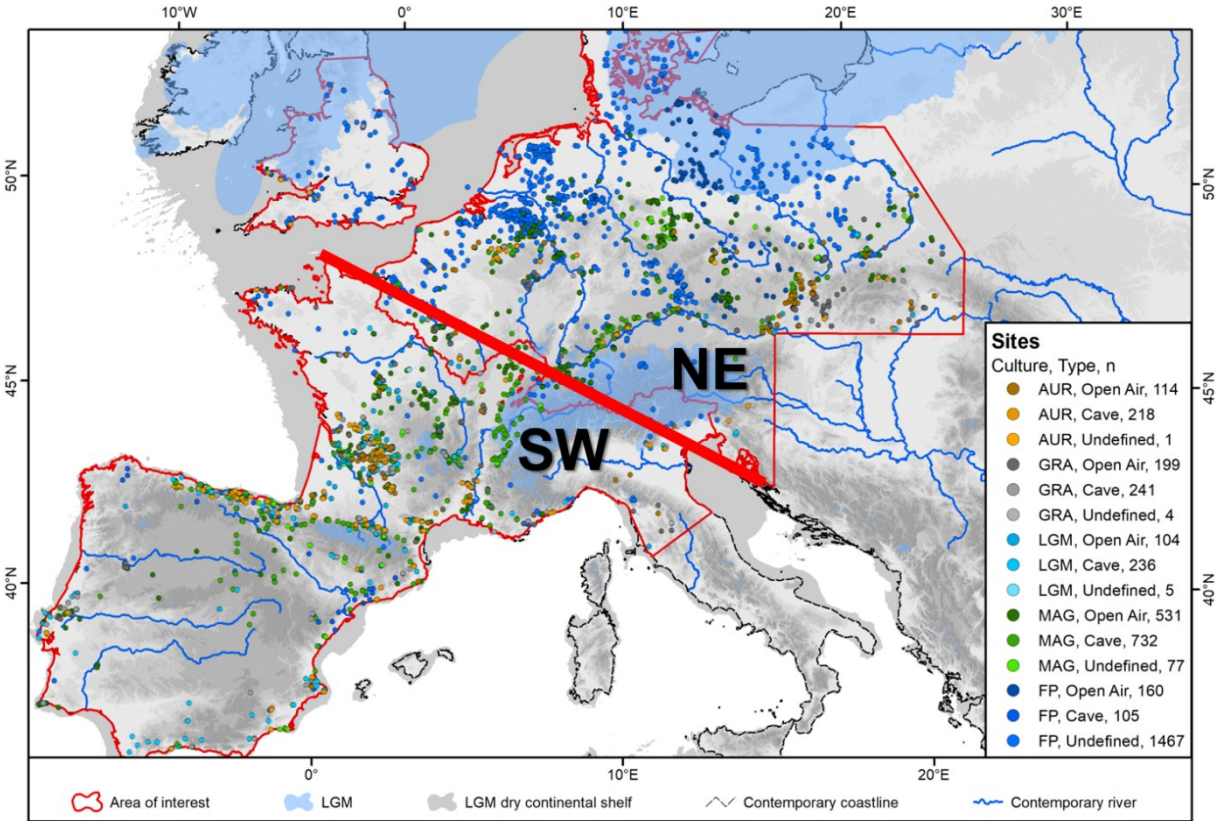



*Figure 16: Late Pleistocene map of Europe, showing both the study areas encircled by red lines as well as the archaeological site database. The big red line is used to highlight the underlying main border line between the sections. Dry continental shelf modified after Willmes 2015, glacial extent after Ehlers et al. 2011.*

### 5.2.2 Environmental variables and their role as settlement and/or discovery factors

To assess the extent to which the currently known distribution of sites is representative for the distribution of hunter-gatherers during the Upper and Final Palaeolithic of Europe, we calculate those shares that can be attributed to factors influencing the likelihood of site discovery on the one hand, and settlement choice of Palaeolithic hunter-gatherers on the other. Generally, a strong influence of discovery factors is interpreted as potentially causing a strongly biased distribution pattern with low representativity, while a strong influence of settlement factors would indicate lower biases and higher representativity.

For our assessment of the settlement and discovery factors, we use eight environmental geodatasets:

- The global elevation model (DEM) NASADEM (NASA JPL 2020)
- The 1 : 5 million international geological map of Europe and adjacent areas (Asch 2003)
- The extent and chronology of Quaternary glaciations (Ehlers et al. 2011)
- Two European maps of Pleistocene aeolian deposits based on topsoil properties (Bertran et al. 2021; Bertran et al. 2016)
- The map of loess landscapes of Europe, based on geological maps (Lehmkuhl et al. 2021)
- The Corine Land Cover (CLC) map of Europe for the year 2018 (available online at https://land.copernicus.eu/pan-european/corine-land-cover/clc2018)
- The HYDE land use model, version 3.2.1 (Klein Goldewijk et al. 2017)

Some geodatasets are only relevant for discovery factors. The CLC and HYDE, for instance, only contain information on modern to contemporary land use and are thus only relevant for the likelihood of site discovery. All other variables can be mainly attributed to the settlement factor, as they can be expected to have influenced the Upper and Final Palaeolithic settlement choice.

The CLC dataset consists of land cover classifications based on satellite imagery, coordinated by the European Environment Agency (EEA) under the framework of the Copernicus programme. From this dataset, the land cover classification for the year of 2018 was used to assess the contemporary land cover and thereby the discovery context of recently discovered sites. At many site locations, land use has probably changed since the date of discovery. Since an individual assessment is not possible for all 4200 sites, current land cover is taken as an approximation.

As many discoveries of archaeological sites in Europe were made at the end of the 19$^{th}$ and the start of the 20$^{th}$ century (Surovell et al. 2017), this timeframe should also be covered by a geospatial query with the archaeological dataset. To this end, the HYDE land use model, version 3.2.1 by Klein Goldewijk et al. (2017) was used. This dataset contains globally modelled land use estimates in generalised classes back to 12,000 ka in a coarse spatial resolution of 5 arc minutes. From the three modelled scenarios, the baseline was selected. For this study, we look at the built-up area in km² per grid cell (UOPP) and the population density in persons per grid cell (POPD). Changes between the years 1800 and 2000 were calculated, as increases in the built-up area or a higher population density imply interventions in the upper soil, possibly uncovering archaeological material.

An attribution of each geodataset to but one of the two opposing fields of settlement choice and discovery likelihood is often impossible and their effects must therefore be discussed and weighted. Here, geological settings are a case in point. Jurassic and Cretaceous units in Europe have the highest potential for high-quality lithic raw materials (Duke and Steele 2010). These materials were consumed

by hunter-gatherers on a daily basis and likely played a role in spatial decision-making processes, although to varying degrees in different periods. While the quality of lithic raw material seems to have been quite important during the Magdalenian (Maier 2015), for instance, it seems to have been of lesser importance during the Final Palaeolithic (Holzkämper et al. 2013). Moreover, thanks to their calcareous matrix, Jurassic and Cretaceous rocks are also prone to the formation of caves and rock-shelters. Containing large numbers of these "welcome structures" (*Empfangsstrukturen* according to Hahn 1995), they are also prone to positively influencing hunter-gatherer settlement choices. The availability of high-quality raw material and natural shelter thus might have been factors influencing settlement choices of hunter-gatherers. However, sediment preservation in caves and below (former) rock-shelters is usually good and for a long time, these settings have been targeted preferentially by archaeologists. These two aspects thus contribute to the discovery factors. Therefore, Jurassic and Cretaceous units are an ambiguous proxy potentially influencing both past settlement choice and likelihood of archaeological discovery.

Further complicating matters are factors which do not only affect both fields of interest, but within a single field can contribute positive and negative effects depending on (potentially shifting) threshold values. Here, a case in point is the distribution of loess. On the one hand, loess seems to be indicative for the distribution of Late Pleistocene steppe environments (Gerasimenko and Rousseau 2008; Fitzsimmons et al. 2012) and thus for habitats likely attractive for hunter-gatherers, potentially influencing settlement choice positively. During the deposition of loess, however, the related dust-storms might have had negative effects on settlement choice. On the other hand, loess plays a central role in the preservation of archaeological sites (Händel et al. 2009). The low-energy deposition of loess preserves archaeological material with minimal spatial redistribution, making it an excellent repository of Pleistocene archaeological material (Chu and Nett 2021). It therefore fosters archaeological discoveries. However, thick loess cover can also render sites inaccessible and thus hamper the discovery of sites.

To address this issue of opposing influence, we use varying combinations of three loess geodatasets, as they each have a unique approach in defining loess and related aeolian sediments and thus can be used to assess different aspects. In Bertran et al. (2016 and 2021), loess is defined by topsoil properties based on the LUCAS topsoil database (Orgiazzi et al. 2018). However, both maps differ in their definition of loess. While the 2016 publication focusses on a narrow definition of loess, mainly based on French and Belgian localities, the 2021 publication implements a broader definition of loess, taking into account the different forms of loess throughout Europe. As such, these two maps can be seen as a lower and an upper estimate of loess thickness based on topsoil data. The loess map of Lehmkuhl et al. (2021), in contrast, is compiled mainly from regional and national geological maps. In these maps,

only loess layers thicker than 2 metres are taken into account. By using the three datasets in comparison, both the topsoil properties and parent substrate indicating loess and related aeolian sediments can be incorporated into the spatial analysis.

All information on topography, namely elevation, slope and aspect, were derived from the NASADEM global digital elevation model (DEM) with a spatial resolution of 30m (NASA JPL 2020). This DEM is a reprocessing of STRM data, with improved vertical accuracy by incorporating auxiliary data from ASTER GDEM, ICESat GLAS, and PRISM datasets (Buckley et al. 2020). The original DEM and all derived products were calculated and exported using the Google Earth Engine (Gorelick et al. 2017).

All environmental variables are visualised as maps in Figure A26 to Figure A35 in the appendix.

### 5.2.3 Geospatial analysis

To reduce the complexity of the geospatial analysis, we implemented an extensive pre-processing workflow, aimed at harmonizing all used environmental variables and preparing them for geospatial query in ArcGIS, version 10.7.1. For the tabular datasets of Upper Palaeolithic sites, this includes internal harmonisation of columns and transformation into a single vector-based geodata file (shapefile, .shp). For easy handling of the resulting file, only columns relevant for individual identification and categorical selection were preserved.

The pre-processing of vector-based environmental variables included merging of the pre-defined classes from each dataset separately before dissolving them into classes selected for this study. To account for possible inaccuracies in the archaeological dataset and/or in the environmental variables, we use a search radius of 500m around each location when an archaeological site does not intersect the variable. The spatial query for raster-based environmental variables was carried out by extracting values based on spatial intersection with the site locations (ArcGIS-tool: *Extract Multi Values to Points*). As some of the environmental variables contain categorical data, cell values at site locations were not interpolated. To enable a better statistical analysis of the CLC dataset, the 44 land cover classes were reclassified and aggregated to 10 classes of related surface types based on preliminary observation on comparable site frequencies per class. Single classes that show noticeably increased or decreased site frequencies were preserved. The result of these two steps is a vector database containing an entry for each archaeological assemblage. Apart from archaeological properties, such as assignment to one of the six periods, natural shelters vs. open-air sites, etc., the attribute table contains the value of each environmental variable at the respective location.

Parallel to the intersection of sites and environmental variables, we extracted reference data based on the AOI (see 5.*2.1 Upper Palaeolithic sites*) to define the expected values and shares of each environmental variable (see 5.2.4 Statistical analysis). We only include areas where human existence was potentially possible throughout the entire Upper and Late Palaeolithic, therefore excluding elevations of >1200 m.a.s.l. based on an empirical exploration of the site dataset. This way, 19 of 4194 sites were excluded. We are aware that this is a simplification and probably leaves periodically uninhabitable regions in the reference data. However, the AOI has to fit all phases of the investigation period and thus requires compromises. Based on the polygons representing the AOI at elevations lower than 1200 m.a.s.l., reference values and shares of the environmental variables were gathered. This process was carried out three times: (1) For the entire AOI, and individually for (2) for the NE-Section and (3) for the SW-Section, (5.*2.1 Upper Palaeolithic sites*). A simplified workflow-diagram can be seen in Figure 17. For more details on the geospatial analysis, refer to Appendix A.

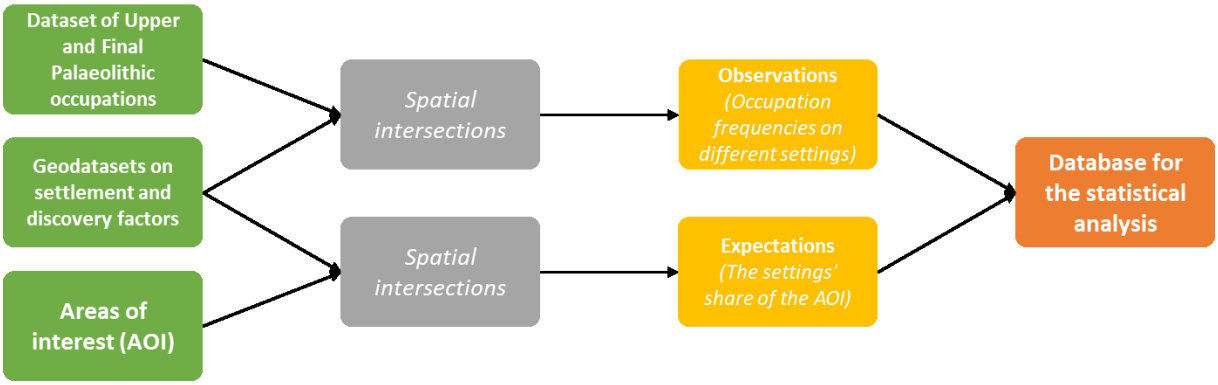



*Figure 17: Workflow diagram of the geospatial analysis, showing how the spatial datasets were processed to allow for a comprehensive statistical assessment. From left to right: The environmental variables are spatially intersected with the areas of interest to assess expected values and spatially intersected with the archaeological dataset to assess observed values. The results are compiled in a tabular database on which all statistical approaches are based. Colour-coding: Green: Input, Grey: Spatial processing, Yellow: Intermediate result, Orange: End result*

### 5.2.4 Statistical analysis

We utilised the site properties to divide the database into chronological (AUR-FP), spatial (NE/SW) and type-specific (natural shelters vs. open-air) subsets. When considering every subset and aggregation of these combinations, this results in site 49 classes. In the next step, we utilised the values of continuous environmental variables to calculate mean, STD, median and quartiles for each Upper Palaeolithic site class. From categorical environmental variables, frequencies of occurrences (n) were calculated and set into perspective to the overall class numbers, resulting in percentages of each of the 49 site classes.

To assess whether these means and class shares represent an under- or over-representation of the respective environmental variable or class in the site distribution, we compared them to their reference value, calculated based on the AOI. To simplify the interpretation, comparison, and visualisation, this value is displayed as a factor of the expected value ranging from 0 to a theoretically open end. The values can be interpreted as the following:

- <1 = Sites are underrepresented (Lower mean or class share in the site-database than in the AOI)
- 1 = No over- or under-representation (Same mean or class share in the site-database as in the AOI)
- >1 = Sites are overrepresented (Higher mean or class share in the site-database than in the AOI)

The resulting 'deviation from expected mean' or 'deviation from expected share' thus represent the statistical relation between the expected and observed occurrences.

To determine how the environmental variables influence the presence or absence of sites, the database was additionally analysed based on the maximum entropy principle using the software MaxEnt, version 3.4.4. This software was originally developed for species distribution and environmental niche modelling (Phillips and Dudík 2008; Phillips et al. 2017), but has also been successfully utilised in archaeological predictive modelling (Galletti et al. 2013; Gillespie et al. 2016; Jones et al. 2019; Alwi Muttaqin et al. 2019). In this raster-based spatial approach, absence data is generated automatically from raster cells where no site is present. The accuracy of the resulting predictive model is assessed in a receiver operating curve (ROC) which depicts the rate between true positives and false positives at different classification thresholds. The area under curve (AUC) is calculated from this ROC and can be used to evaluate the performance of the model. A detailed description of the statistics behind the software can be found in (Elith et al. 2011; Merow et al. 2013).

For processing within this software, all environmental variables were rasterised and resampled to a spatial resolution of 100*100 metres using the nearest neighbour method. For the best possible comparison between settlement and discovery factors, all environmental variables were included in each model run. As a predictive modelling is not the main aim of this study, we extracted only response curves and jack-knife variable importance from the result. The response curves show how the predicted probability of a response variable (in this case the presence of archaeological sites) changes as the values of a predictor variable changes. As the predicted probability of sites is increased on surfaces where sites are overrepresented in comparison to the background data, the response curves are comparable to the conventionally calculated 'deviation from expected share' and 'deviation from expected mean'. The jack-knife variable importance can be used to estimate the predictive value of an environmental variable as a whole, which is comparable to conventionally calculated determination measurements based on presence/absence data.

For additional internal statistical queries and tests, the tabular database was further analysed using the statistics software IBM SPSS statistics. In order to provide additional information on the association between environmental variables and archaeological classes (period, cave, open-air, etc.), the database was internally tested using the contingency coefficient. The data set was additionally analysed with a Two-Step Cluster analysis in order to reveal different groups/clusters. The strength of the Two-Step Cluster analysis is its ability to distinguish groups without predefining them. Therefore, the number of clusters is not based on the subjective choice of the researcher, but on a statistical measure of fit (Kayri, 2007, Rundle-Thiele et al., 2015). The method also permits the simultaneous analysis of categorical and continuous data by applying the Log-Likelihood method as a distance measure to separate groups, followed by the probabilistic method BIC (Bayesian Information Criterion, c.f. Schwarz, 1978) to gain an optimal grouping (Zhang et al., 1996).

In addition, the archaeologically predefined classes were tested by applying a discriminant analysis (Davis, 1986, Kovarovic et al., 2011) and the Naïve Bayes method, a probabilistic classification approach (Armero et al., 2020, Monna et al., 2020). All statistical methods applied in this study are listed in Table 6 for additional context. For more details, see Appendix A.

*Table 6: Summary of all statistical approaches conducted in this study stating the name, a short description, and the main aim of each method.*

| Statistic | Description | Main aim |
|---|---|---|
| **Deviation from expected mean/share** | Expected: Mean/share of continuous/categorical environmental variables in the area of interest (AOI).<br>Observed: Mean/share of continuous/categorical environmental variables at site locations.<br>Deviation: Comparison between expected and observed values. | Assessing the main statistical association between site locations and environmental variables. An over-representation of sites indicates favourable conditions for settlement or discovery while under-representation indicates unfavourable conditions. A comparison between settlement and discovery factors allows for a relative assessment of associative strength. |
| **MaxEnt predictive modelling** | Predictive presence/absence modelling based on the maximum entropy principle. Presence samples are taken from known site locations and absence samples are generated from non-site locations. | Assessing the strength of environmental variables in predicting the presence and absence of sites via the jack-knife variable importance, which shows the predictive strength for each environmental variable separately and cumulative. |
| **Contingency coefficient crosstabs** | A coefficient of association based on chi-squared statistics. It is used to assess if two variables are independent or dependent of each other. | Identifying possible associations between environmental variables. This is important for the comparison of settlement and discovery factors as dependencies between them might lead to misinterpretations. |

| | | |
|---|---|---|
| **Two-Step unsupervised classification** | An unsupervised classification algorithm for both continuous and categorical variables that divides datasets into clusters. Clusters are assigned with respect to variable similarity which is measured by distance to a cluster centroid. | Experimentally dividing the database into two natural clusters based solely on similarities/differences between environmental variables. By comparing the result to the archaeologically predefined classes (period, type, and region), one can assess how well they are represented by the natural grouping and how strongly the variables vary in between the classes. |
| **Discriminant analysis** | Based on a known group membership, a predictive model is built. The model consists of a set of discriminant functions based on linear combinations of environmental variables that provide the best discrimination between groups. An accuracy is assigned based on the comparison between the predefined archaeological class and the predicted class. | Assessing the environmental similarities and differences within and between archaeological classes (period, type, and region). High accuracies indicate that a class has high environmental similarities and can be easily discriminated from other classes while low predictive accuracies indicate a more environmentally heterogeneous class that cannot be discriminated easily from other classes. Environmental variables attributed to the settlement and discovery factors can be tested separately for comparison. |
| **Naïve Bayes** | Also builds a predictive model based on known group membership. Naïve Bayes, however, works with conditional probabilities. A probability for class membership is assigned for each possible value of an environmental variable. Environmental variables are additionally weighted based on predictive power. | Assessing (1) the environmental similarities and differences between the archaeological classes, similar to the discriminant analysis, and (2) the power that each environmental variable has in predicting class affiliation. Both results can aid the identification of environmental similarities and differences between the predefined classes. |

## 5.3 Results

The results of the spatial and statistical analyse are (i) the over- or under-representation of site frequency on the selected settings in comparison to their share of the AOI and (ii) statistical measures of correlation, determination and predictive power. In the following section, we present these results for the settlement and discovery factors regarding the entire database and - where large deviations from the expectations were observed – for subsets. Note that not all results can be discussed in the main text due to the large number of variables, site classes, and statistical queries, but can be found in Appendix A.

Regarding cave and open-air sites, sub-setting is permissible for all technocomplexes but the Final Palaeolithic, where information for reliable discrimination is not sufficient. For the remaining periods, we see the same pattern in virtually all cases (Table 7). The total number of cave sites dominates over open-air sites. Only in the late Gravettian, the ratio is even. Regarding subsets, however, cave sites always dominate in the SW part, while open-air sites always dominate in the NE part.

*Table 7: Class sizes (n) of archaeological classes. These classes are assigned based on period (AUR, GRA1, GRA2, LGM, MAG, FP), type (natural shelter, open-air) and region (NE, SW). Note that subsetting by type is not reliable for FP (numbers in grey).*

| | All | AUR | GRA1 | GRA2 | LGM | MAG | FP |
|---|---|---|---|---|---|---|---|
| **All** | 4194 | 333 | 309 | 135 | 345 | 1340 | 1732 |
| **All Open-Air** | 1108 | 114 | 113 | 66 | 104 | 531 | 160 |
| **All Cave** | 1532 | 218 | 175 | 66 | 236 | 732 | 105 |
| **NE** | 2288 | 107 | 105 | 53 | 47 | 586 | 1390 |
| **NE Open-Air** | 705 | 66 | 72 | 42 | 31 | 335 | 159 |
| **NE Cave** | 334 | 40 | 33 | 9 | 16 | 226 | 20 |
| **SW** | 1901 | 226 | 203 | 82 | 296 | 754 | 340 |
| **SW Open-Air** | 403 | 48 | 61 | 24 | 73 | 196 | 1 |
| **SW Cave** | 1185 | 178 | 141 | 57 | 218 | 506 | 85 |

### 5.3.1 Over- and under-representation on settlement and discovery factors

The over- and under-representation of occupations in environmental variables for the settlement factors has been calculated from the difference between the observed and expected frequencies. The highest deviations from the expectation can be observed on the **geological** settings and within the different **aeolian deposits**. Large differences are also displayed between the NE and SW section and especially between cave and open-air sites. For all occupations, the classes with highest over-representations of environmental variables for the settlement factors are:

- *Cretaceous geological units* (mean factor of 1.75, up to 4 for SW cave sites)
- *Jurassic geological units* (mean factor of 1.5, up to 6 for NE cave sites)
- *Loess and related sediments* (mean of 1.5, up to 4.5 for NE open-air sites)
- *Southern aspects* (mean factor of 1.1, up to 1.6 for NE open-air sites)

Regional examples of these over-representation are indicated in Figure 18. In contrast, occupations are mostly underrepresented in areas with glacial flintstone potential, sandy surfaces (except FP sites), flat surfaces, and areas with north-west aspect.

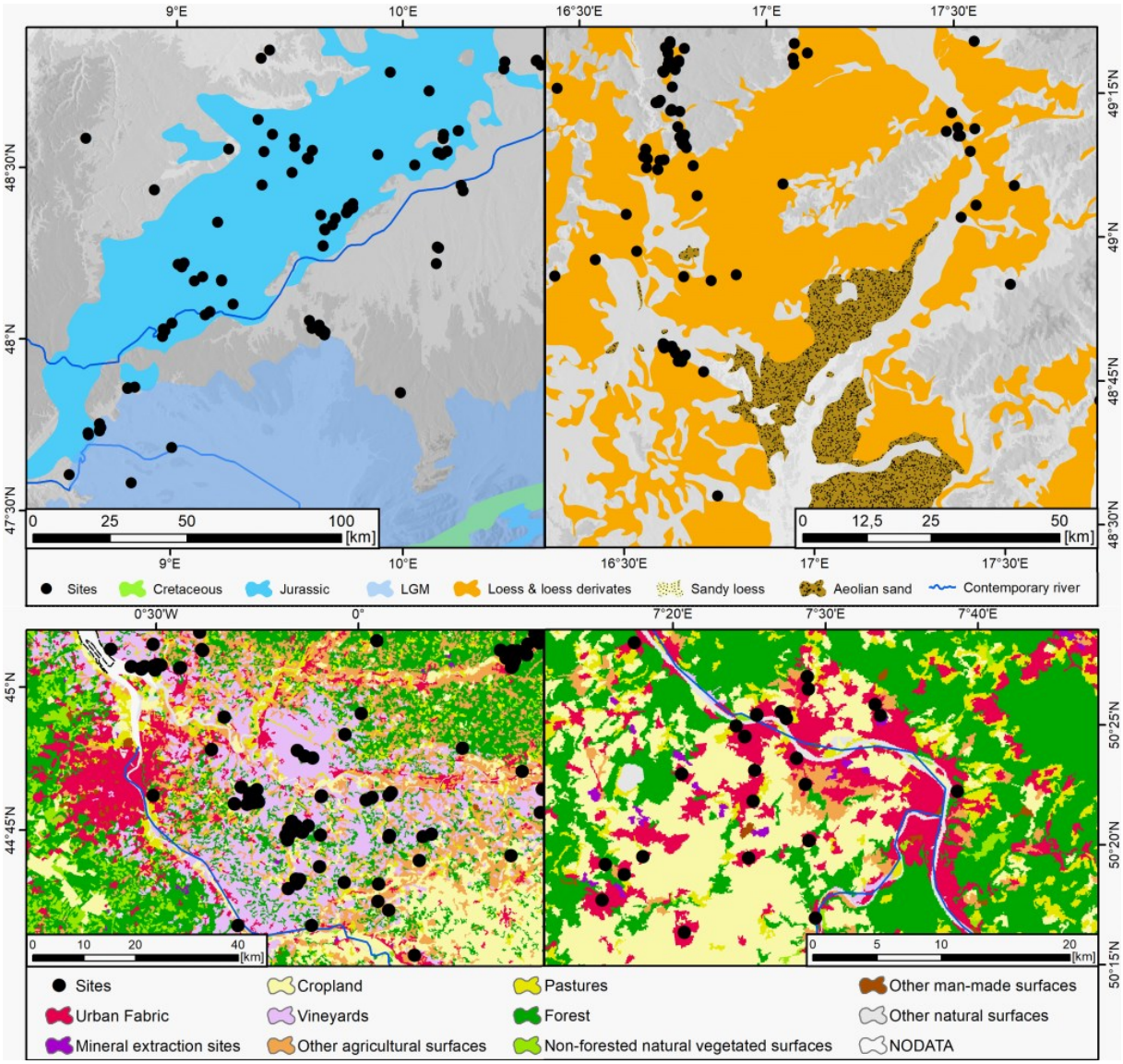



*Figure 18: Regional examples for the over-representation of sites on specific geological, sedimentological and land use contexts. Maps show the South German Scarplands (upper left many sites on Jurassic geological units), the border region between Austria, Czech Republic and Slovakia (upper right, many sites on loess and loess derivates as defined by Lehmkuhl et al. 2021) the wine region Bordeaux, France (lower left, many sites on vineyards) and the city of Koblenz, Germany (lower right, many sites on urban fabric).*

The highest deviation from the expected share in the environmental variables for the discovery factors can be found in the **Corine Land Cover** dataset. The differences between the NE and SW section and cave and open-air sites are much lower than in the settlement factors. When assessing all occupations, the classes with the highest over-representations are:

- Continuous urban fabric and discontinuous urban fabric (factor between 4 and 7.7)
- *Mineral extraction sites* (mean factor of 5.25, exception: LGM with 0)
- G*reen urban areas* (mean factor of 3.45, exception: AUR with 0)
- V*ineyards* (mean factor of 2.25, exception: FP with 1.85)

- Complex cultivation patterns (mean factor of 2.15)
- *Broad-leaved forest* (mean factor of 1.9, exception: FP with 0.5)
- *Water courses* (mean factor of 3, exception: LGM with 0)

Regional examples for these over-representations are displayed in Figure 19. Occupations are mostly **underrepresented** on CLC classes representing man made barren land, cropland, related agricultural areas and natural vegetation. All listed over-representations on the settlement and discovery factors are displayed as a graph in Figure 19. For an overview of all results of the over- and under-representations in a similar format, see Figure A36 to Figure A86 in Appendix A.

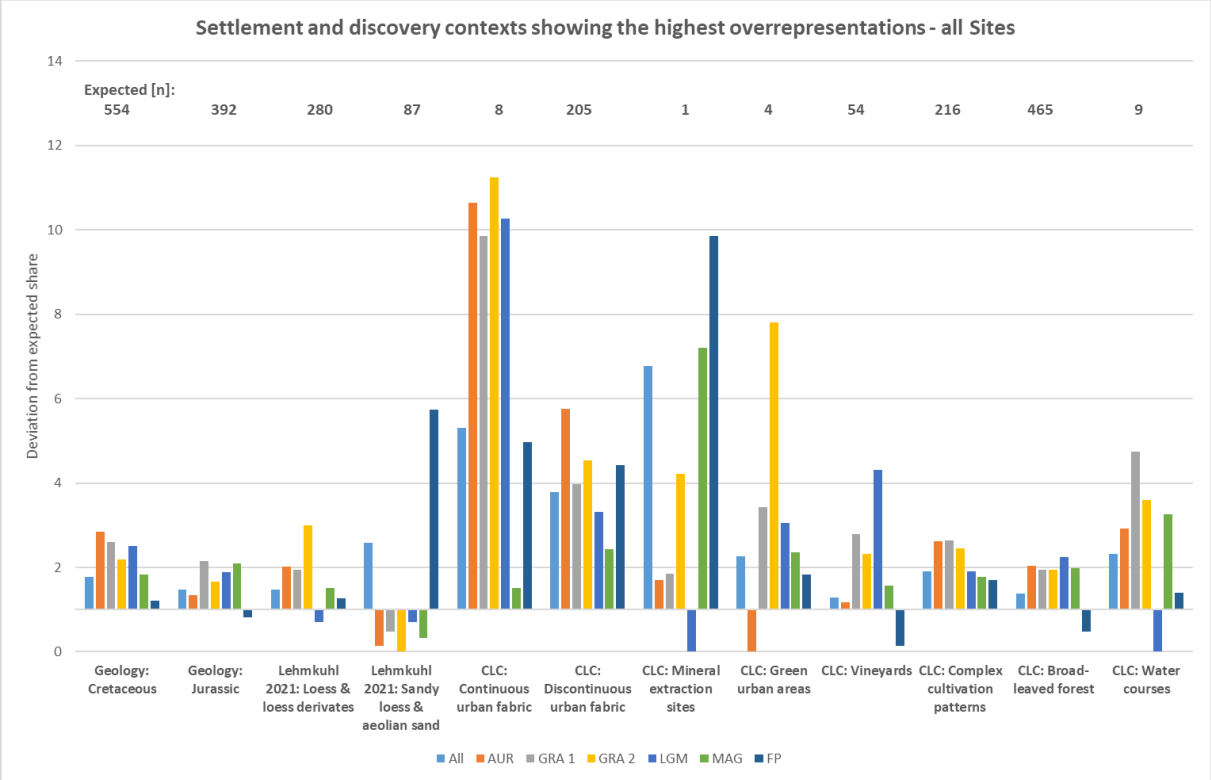



*Figure 19: Chart on the over-representation of sites on specific geological, sedimentological and land use contexts. These settlement and discovery contexts were selected as they display the highest over-representation of Upper and Final Palaeolithic sites. The expected values represent the share of all sites that would equal the area share of each respective surface. The over-representation is displayed as a factor of the expected value. Charts on all different combinations of environmental variables and archaeological classes can be found in Appendix A (Figure A36 to Figure A80).*

### 5.3.2 Predicting the presence/absence of Upper and Final Palaeolithic sites

The presence-absence modelling based on MaxEnt provides very promising results (Figure 20). All areas under curve (AUC) in the receiver operating characteristic curve (ROC) show acceptable, excellent or even outstanding accuracies. This indicates that the presence or absence of sites can be predicted at a high accuracy when using all environmental variables. The predicted accuracy is generally lower in combined, heterogeneous archaeological classes (lowest for *Sites_All*) and higher for more specific, homogeneous archaeological classes (e.g., *Sites_AUR_SW_Cave*).

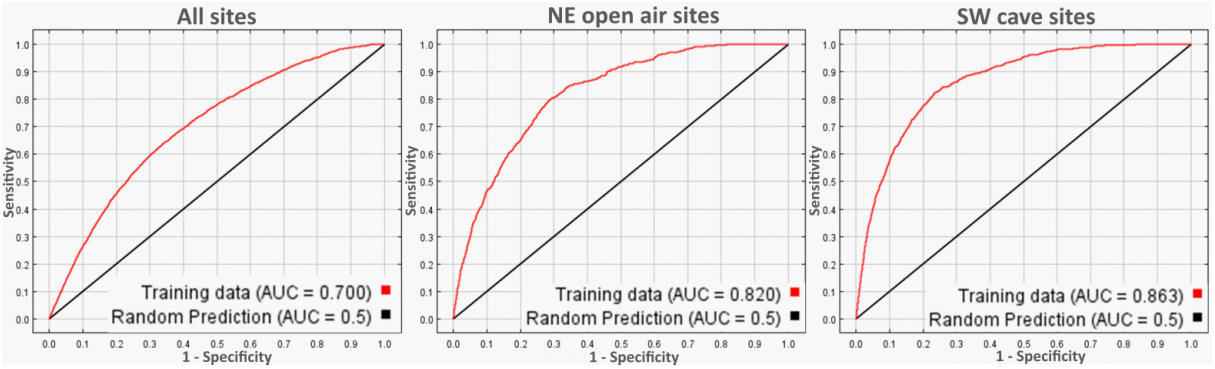



*Figure 20: MaxEnt predictive model accuracy for model runs with all environmental variables but different subsets of the archaeological database. Left: All sites, Middle: NE open-air sites, Right: SW cave sites. The red line shows the receiver operating curve (ROC), which depicts the rate between true positives and false positives at different classification thresholds. The area under curve (AUC) is calculated based on the ROC and compared to a random distribution (black line). AUC's of 0,5 suggest no discrimination, 0.7 to 0.8 is considered acceptable, 0.8 to 0.9 is considered excellent and more than 0.9 is considered outstanding.*

As the goal of this study is not a predictive model, the more relevant information is displayed in the response curves and the variable importance. The response curves are strongly associated with the over- and under-representation calculated for the environmental variables for both the settlement and the discovery factors. As such, the response curves show a high predicted probability for parameters with a high over-representation of sites and a low predicted probability for parameters with an under-representation. An example for this is the CLC response curve for all sites (Figure 21). In this curve, the CLC classes 1, 2, 7, 10, 15, 20, 23 and 40 (*continuous, urban fabric, discontinuous urban fabric, mineral extraction sites, green urban areas, vineyards, complex cultivation patterns, broad-leaved forests, water courses*) show a high cumulative predicted probability. These are the same classes that also display a high over-representation of sites.

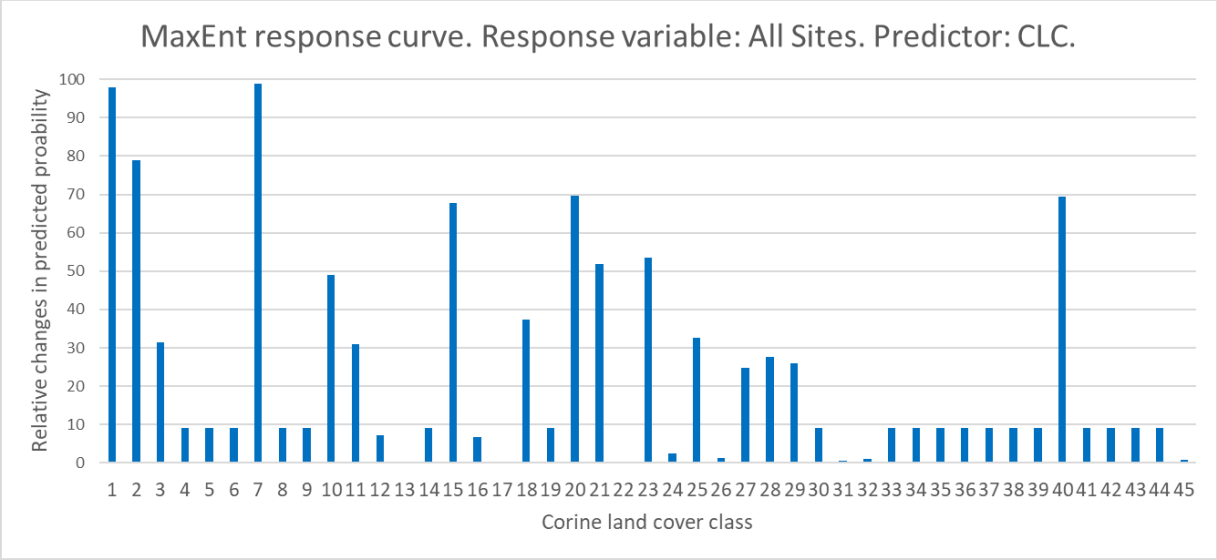



*Figure 21: Response curve showing the changes in predicted probability of archaeological sites attributed to the different values of the Corine Land Cover variable. These results can be compared to Figure A45 in Appendix A, showing the over- and under-representation of sites on the different CLC classes. Note the high conformity between these two charts.*

A high predicted probability is also indicated for loess with respect to NE open-air sites and for geological settings with respect to SW cave sites. Due to this high conformity between the deviation from the expected share/mean and the MaxEnt response curves, and to avoid redundancy, we have decided not to present further response curves in this study.

The real added value of the MaxEnt approach compared to the over/under-representation is the jack-knife variable importance. This shows the training gain that each environmental variable has in predicting the presence/absence of a site. The statistical analysis provides a very clear picture: The single most important variable to predict site presence in all archaeological classes is the CLC (see Figure 22). Only in some archaeological subsets this is contested by the SRTM elevation and/or slope. This may result from the manner MaxEnt treats background data, assuming that all elevation and slope values are theoretically viable. As has been mentioned in section 5.2.3, this is not the case, as elevations uninhabitable for most of the Upper and Final Palaeolithic were included. For NE and open-air sites, the built up area and loess has an additionally increased variable importance, whereas the geology and slope has an increased importance for predicting SW and cave sites.

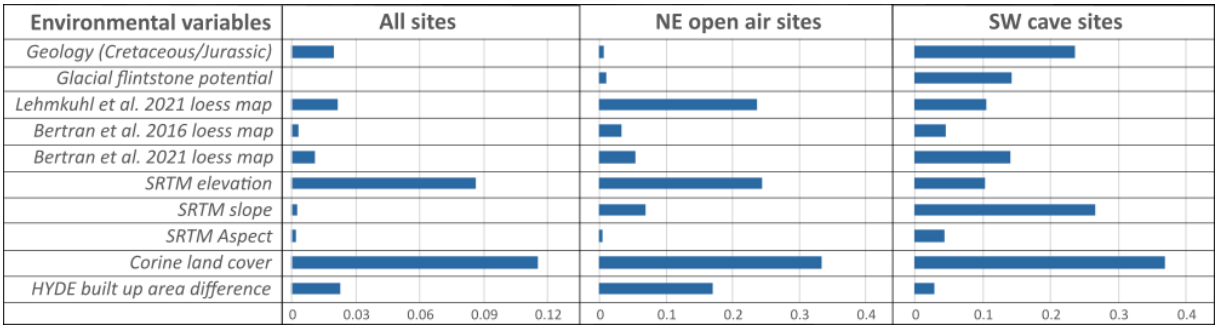



*Figure 22: MaxEnt jack-knife variable importance for model runs with all environmental variables but different subsets of the archaeological classes. Left: All sites, Middle: NE open-air sites, Right: SW cave sites. Only the isolated training gain is displayed, which indicates the predictive power of each environmental variable by itself.*

### 5.3.3 Distinguishing between archaeological classes based on environmental variables

To analyse how well archaeological classes (period, type and region) can be distinguished using the environmental variables, we determined the contingency coefficient (see Table 8), did a clustering of data via the Two-Step classifier, and conducted a discriminant analyses and Naïve Bayes classification using the commercial statistic packages of SPSS. For some analyses, continuous variables were classified as stated in Appendix A.

The contingency coefficient shows that associations between the environmental variables are mostly weak and very weak. Due to the large sample size, all associations between the variables are significant at the level of $\alpha = 0.0001$. Unsurprisingly, moderate to strong associations are found between the different definitions of loess, as these variables represent the same sedimentary complex. Moderate associations between the CLC dataset and terrain parameters, such as elevation and slope, indicate that the land use depends on the terrain to a certain degree. Therefore, influence exerted by these variables cannot be completely distinguished.

*Table 8: SPSS contingency coefficient crosstabs between all environmental variables. The contingency coefficient assesses the dependence between categorical variables. Colour-coding: Heat map from 0 (blue, no dependence) to 1 (red, perfect dependence).*

| | Elevation | Slope | Aspect | Uopp | Corine | LL_2021 | BL_2016 | BL_2021 | Geology | Flint |
|---|---|---|---|---|---|---|---|---|---|---|
| Elevation | | 0.381 | 0.187 | 0.252 | 0.482 | 0.386 | 0.306 | 0.472 | 0.283 | 0.305 |
| Slope | 0.381 | | 0.213 | 0.181 | 0.519 | 0.283 | 0.159 | 0.357 | 0.269 | 0.310 |
| Aspect | 0.187 | 0.213 | | 0.141 | 0.319 | 0.112 | 0.093 | 0.090 | 0.109 | 0.146 |
| Uopp | 0.252 | 0.181 | 0.141 | | 0.436 | 0.246 | 0.202 | 0.190 | 0.168 | 0.107 |
| Corine | 0.482 | 0.519 | 0.319 | 0.436 | | 0.417 | 0.265 | 0.387 | 0.305 | 0.286 |
| LL_2021 | 0.386 | 0.283 | 0.112 | 0.246 | 0.417 | | 0.424 | 0.541 | 0.193 | 0.267 |
| BL_2016 | 0.306 | 0.159 | 0.093 | 0.202 | 0.265 | 0.424 | | 0.623 | 0.177 | 0.088 |
| BL_2021 | 0.472 | 0.357 | 0.090 | 0.190 | 0.387 | 0.541 | 0.623 | | 0.261 | 0.475 |
| Geology | 0.283 | 0.269 | 0.109 | 0.168 | 0.305 | 0.193 | 0.177 | 0.261 | | 0.190 |
| Flint | 0.305 | 0.310 | 0.146 | 0.107 | 0.286 | 0.267 | 0.088 | 0.475 | 0.190 | |

The unsupervised Two-Step Cluster analysis with Log-Likelihood as distance measure and the Schwarz's Bayesian Criterion as clustering criterion (Backhaus et al., 2018) brought the following results. A smaller class 1 contains 30-40% of the sites, typically non-cave lowland sites on a flat or gentle slope with a pronounced loess-context and high built-up area differences. The most frequent land cover contexts of this class are urban areas, mineral extraction sites or non-irrigated arable land. The larger class 2 contains 60-70% of the sites, typically cave sites located at higher elevations and moderate to steep slopes with a pronounced Cretaceous or Jurassic geological context. The most frequent land cover contexts of this class are vineyards, complex cultivation patterns, and forest.

In class 2 (see Table 9), the archaeological classes show a close correspondence of up to 94% with cave sites, the SW section, and periods older than the Final Palaeolithic. In contrast, class 1 predominantly contains NE sites of undefined type from the Final Palaeolithic, though the correspondence is lower than in class 2 and ranges up to only 66%. However, this categorisation depends strongly on environmental variables which are attributed to the settlement factors. When considering only discovery-relevant environmental variables, the attribution does not deviate far from a random 50/50 distribution.

*Table 9: Percentage of sites from different archaeological classes attributed to SPSS two-step classification classes 1 and 2. Refer to the previous text for an explanation on the two classes. Colour-coding: Heat map from 0% (blue, no sites within this class) to 100% (red, all sites within this class).*

| | | | Culture | | | | | Type | | | Region | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Variables | Class | Class size | Aur | Gra | LGM | Mag | FP | Undef. | Cave | Open-air | NE | SW |
| All variables | Class 1 | 38.4 | 23.7 | 22.5 | 9.3 | 20.1 | 65.4 | 66.9 | 6.3 | 43.0 | 63.5 | 8.4 |
| | Class 2 | 60.6 | 74.2 | 75.9 | 88.4 | 79.6 | 33.8 | 32.2 | 92.2 | 56.7 | 36.4 | 89.9 |
| Settlement | Class 1 | 30.2 | 7.5 | 9.5 | 6.7 | 13.2 | 57.8 | 59.2 | 4.2 | 25.6 | 48.0 | 8.9 |
| | Class 2 | 68.8 | 90.4 | 89.0 | 91.0 | 86.4 | 41.4 | 40.0 | 94.3 | 74.0 | 51.8 | 89.4 |
| Discovery | Class 1 | 36.1 | 40.8 | 34.5 | 32.5 | 28.4 | 42.3 | 41.2 | 27.4 | 40.9 | 39.1 | 32.4 |
| | Class 2 | 63.9 | 59.2 | 65.5 | 67.5 | 71.6 | 57.7 | 58.8 | 72.6 | 59.1 | 60.9 | 67.6 |

To test the degree to which the archaeological classes (period, type, and region) can be distinguished with respect to the environmental variables, a discriminant analysis was carried out (Table 10). The additionally conducted Naïve Bayes analysis returned higher percentages of correctly classified data and thus appears to be better suited to predict the influence exerted by environmental variables on the archaeological class affiliation. The results of the discriminant analysis and of the Naïve Bayes method confirm the impression that the regions as well as the cave sites are closely associated with the environmental variables as suggested by the unsupervised Two-Step Cluster analysis. Again, environmental variables attributed to the settlement context show the archaeological classes than those attributed to the discovery context. The archaeological periods, however, show a different tendency, displaying a decreasing attribution correspondence with increasing age. Very low correspondences are also indicated for open-air sites. This points to a very heterogeneous class.

*Table 10: Percentage of accurately assigned/predicted archaeological class affiliation based on the discriminant analysis and Naïve Bayes. Dependent variable: archaeological class. Independent variable: Environmental variables. Colour-coding: Heat map from 0% (blue, no sites assigned accurately) to 100% (red, all sites assigned accurately).*

| | | Culture | | | | | | Type | | | | Region | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Category | Variables | AUR | GRA | LGM | MAG | FP | Total | Undef. | Cave | Open-air | Total | NE | SW | Total |
| Discriminant analysis | All variables | 30.4 | 18.8 | 37.4 | 35.4 | 60.4 | 43.8 | 55.4 | 73.3 | 54.7 | 61.7 | 72.1 | 82.9 | 77 |
| | Settlement | 18.4 | 26.5 | 37.4 | 34.8 | 60.4 | 43.4 | 55.8 | 69.9 | 26.1 | 61.1 | 71.1 | 82.8 | 76.4 |
| | Discovery | 9.6 | 10.4 | 20.6 | 30.7 | 43.6 | 31.4 | 30.6 | 73.3 | 38.8 | 48.4 | 52.2 | 76 | 63.4 |
| Naive Bayes | Training (80%) | 9.3 | 20.3 | 15.2 | 74.3 | 70.2 | 56.6 | 64.7 | 87.8 | 38 | 66.2 | 91.6 | 92.1 | 91.9 |
| | Test (20%) | 3.1 | 10.6 | 12.9 | 71.9 | 75.4 | 57.5 | 62.9 | 91 | 37.1 | 65.7 | 90.2 | 89.2 | 89.7 |

In addition, the Naïve Bayes method was used to determine the strength of each environmental variable in predicting the archaeological class affiliation (see Table 11). Again, the results confirm that the power to predict the affiliation to different periods based on environmental variables is very weak, while the affiliation to site types and especially regions is stronger. In contrast to the previous results, however, discovery factors, such as the difference in built up area and CLC are also ranking highly in terms of relative predictive strength between the environmental variables. Besides, the highest predictive strengths are displayed for elevation, aspect, and loess according to the loess classification of Lehmkuhl et al. 2021. The results also support several common assumptions about the archaeological classes like the fact that the slope ranks higher for distinguishing between different site types but not between the different periods and regions. Another example is the geological context, which ranks highest for distinguishing between regions, supporting the observations made for the over- and under-representation on Cretaceous and Jurassic units.

*Table 11: Power of the environmental variables in predicting the archaeological class affiliation according to Naïve Bayes. Colour-coding: Heat map from blue (low predictive power or relative rank 10) to red (high predictive power or relative rank 1).*

| Arch. class | Statistical value | Elev. | Slope | Aspect | Uopp | Corine | LL_2021 | BL_2016 | BL_2021 | GEO | Flint |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Culture | Rank | 5 | 9 | 2 | 1 | 6 | 3 | 4 | 8 | 7 | 10 |
| | Pseudo-BIC | 1.071 | 1.115 | 1.058 | 1.056 | 1.075 | 1.061 | 1.063 | 1.063 | 1.087 | 1.214 |
| | Average Log-Likelihood | -1.061 | -1.103 | -1.053 | -1.05 | -1.073 | -1.058 | -1.055 | -1.055 | -1.077 | -1.2 |
| Type | Rank | 7 | 5 | 2 | 1 | 4 | 3 | 6 | 10 | 8 | 9 |
| | Pseudo-BIC | 0.794 | 0.77 | 0.76 | 0.76 | 0.767 | 0.762 | 0.778 | 0.919 | 0.812 | 0.848 |
| | Average Log-Likelihood | -0.783 | -0.767 | -0.754 | -0.755 | -0.759 | -0.758 | -0.768 | -0.905 | -0.8 | -0.835 |
| Region | Rank | 3 | 7 | 1 | 2 | 4 | 8 | 5 | 10 | 6 | 9 |
| | Pseudo-BIC | 0.224 | 0.244 | 0.224 | 0.224 | 0.227 | 0.248 | 0.289 | 0.29 | 0.235 | 0.261 |
| | Average Log-Likelihood | -0.216 | -0.232 | -0.216 | -0.218 | -0.222 | -0.246 | -0.287 | -0.276 | -0.224 | -0.248 |

## 5.4 Discussion

For an interpretation of the effect of over- and under-representation of settlement and discovery factors on the representativity of site distribution patterns, the following rationale is important. First, since our approach uses a variety of datasets with different formats, types, and resolutions, and

although we have implemented a thorough pre-processing to harmonise the datasets, this variety still can lead to some potential inaccuracies. Information on how these potential inaccuracies are addressed can be found in Appendix A3. Second, a diachronic comparison of periods is meaningful when the general spatial distribution of sites within these periods is similar. This is broadly the case for all periods from the Aurignacian to the Magdalenian. Site distribution during the Final Palaeolithic, however, differs markedly with a clear emphasise on the NE part of the study area. We will thus often observe differences between the Final Palaeolithic and all other periods, which are mainly a function of the different distributions of sites. However, synchronous over- and under-representations within each period remain meaningful, since the likelihood for a site to be located at a certain position still depends on the share that these factors have in the landscape.

Having said that, the first observation that can be derived from the largescale distribution of all sites is that the influence of the joint effects of all biasing factors is not strong enough to override the general spatial differences between site location during the Upper and Final Palaeolithic. So at this very large scale, the observable site distribution seems to be representative for a shift in settlement activities from the southwest towards the northeast. However, when looking at smaller scales and individual factors, the picture is getting more complicated.

For the **geological setting** (see Figure A36 to Figure A44 in Appendix A), we observe that occupations are moderately overrepresented on **Cretaceous and Jurassic geological settings** through all technocomplexes (mean factor of 1.9) but the Final Palaeolithic. When different types of sites are considered individually, it is hardly surprising that cave sites show a high over-representation (mean factor of 2.5). At the moment, it cannot be decided how much of this over-representation might be related to past settlement preferences in areas where natural cavities were available and how much is due to the fact that cavities are sediment traps and a primary target of archaeological research. It is however worth mentioning that open-air sites also show an over-representation on these bedrocks, although to a much lower degree (mean factor of 1.25). This might indicate a slight settlement preference, maybe related to the availability of high-quality raw material. Another explanation might be that open-air sites have been preferentially detected in the vicinity of cave sites as a result of intense research in their surroundings. If we take the signal from open-air sites as a conservative base-line for all types of sites, we might attribute 25% of the general over-representation to settlement preferences and 65% to discovery biases. When comparing the two regions, another interesting pattern emerges. While cave sites in the NE section are highly overrepresented on Jurassic settings (mean factor of 5.6), they are underrepresented on Cretaceous settings (mean factor of 0.6). Cave sites in the SW section, in contrast, show a pronounced over-representation on Cretaceous settings (mean factor of 3.6) and a far lower over-representation on Jurassic settings (mean factor of 1.45). This difference cannot be

interpreted easily. For further investigations, a dataset with a better thematic resolution would be helpful, as the 1:5 million geological map of Europe (Asch 2003) is highly generalised and does not include information about abundance and quality of lithic raw materials within these main geological units.

Looking at the entire database of sites in comparison to **aeolian deposits** (see Figure A36 to Figure A44 in Appendix A), no clear pattern of over- or under-representation can be identified. Within regional and type-specific subsets, however, this changes considerably. All three definitions of loess (Bertran 2016, Bertran 2021, Lehmkuhl et al. 2021) show a very high over-representation of open-air sites from all technocomplexes with a trend of increasing over-representation with increasing age (up to a factor of >5 for AUR open-air sites). This is especially true for open-air sites in the NE-section, which show the highest over-representations within the three loess settings. The clear temporal trend in the deviation from expected shares suggests an increasing taphonomic bias towards older periods, and the positive effects of preservation overweigh potential archaeological invisibility due to extensive loess cover. This general statement, however, must be weighed for regional differences. A comparison between the loess-rich NE region (8-13% loess cover) and almost loess-free SW region (1-3% loess cover) shows that only 14% of all discovered sites in the NW are associated to LGM and earlier technocomplexes, while these make up 42% in the SW. Given that loess accumulation took place mainly during and before the LGM, and given further that the thickness of the loess cover is generally more important in the NE, this finding suggests that the overall increase of discovered sites in relation to the length of the periods towards younger technocomplexes – and particularly from the LGM to the Magdalenian – can partly be attributed to higher discovery rates due to decreasing loess cover, especially in the NE-section. However, the increase of site numbers could also be related to the postulated re-dispersal into the NE-section and related population growth. Sites from the Final Palaeolithic differ from this trend inasmuch as they show mostly equal representation or even under-representation on loess in all three definitions. High over-representations of up to a factor of 5, in contrast, are found on sandy aeolian deposits, such as silty sand (Bertran 2016), sandy loess, coversands (both Bertran 2021), and sandy loess & aeolian sand (Lehmkuhl 2021). The diachronic differences can be attributed to the northbound advance of settlement into areas closer to the former glacial front with coarser materials. The general over-representation during the Final Palaeolithic itself, however, remains a valid observation. Maybe a specific kind of habitat, formally indicated by loess, has shifted north-eastwards and is now related to sandy aeolian deposits. This would indicate that parts of the observable pattern can be related to settlement preferences. It thus seems that for pre-Magdalenian periods, the presence of loess biases the distribution of known sites towards a better preservation in relation to other sedimentation context to up to a factor of 5, while in the NE, it reduces the archaeological visibility at the same time

by at least 25%. For the Magdalenian and Final Palaeolithic, in contrast, the over-representation of sites on loessic and sandy aeolian deposits might reflect habitat preferences and thus settlement choice, indicating a more reliable representation of hunter-gatherer presence in the landscape.

When considering **aspect** (see Figure A72 to Figure A80 in Appendix A) in the entire database, it stands out that over-representations are mostly found on south, south-west and west aspects while the highest under-representation is found on northern and adjacent aspects. Interestingly, this trend is stronger in cave sites than in open-air sites. For caves in the NE-section, over-representations are found on the whole southern half of the compass. A clear exposure towards the sun indicates settlement choice as important factor due to favourable microclimatic conditions. This is in line with the observation that the period with the highest consistency in highest over-representation on southern aspects is the LGM, as a good microclimate is more important under macro-climatically unfavourable conditions. However, the positive deviations from expected shares are generally very low and a very low jack-knife importance indicates that the influence of the aspect parameter on the site's locations is far less pronounced than the influence of other parameters. Also, in addition to southern aspects, pronounced over-representations of open-air sites can be found on northern and north-eastern aspects in the NE section. And in all subsets, occupations of the late Gravettian are overrepresented on north-eastern and adjacent aspects. The fact that open-air sites vary stronger than naturally shelter sites might indicate that they were of shorter use and that aspect therefore played a less important role. Shorter stays at the same location could also explain the deviating aspect during the late Gravettian, a period of deteriorating climatic conditions. Over- and under-representations of aspect thus might be mainly related to settlement choice, but the influence of this factor on the entire database is very low.

As no representative expectations for the topographical parameters **elevation** and **slope** can be extracted from the AOI, the results for these parameters can only be discussed within each period. To this end, boxplots were created (Figure A81 to Figure A86 in Appendix A). For all occupations, a narrow **elevation** range of 100 to 250m within the 25$^{th}$ to 75$^{th}$ percentile can be observed for all periods from the Aurignacian to the LGM, while sites of the Magdalenian and Final Palaeolithic show a higher variability with 75$^{th}$ percentiles ranging up to 500m elevation. The low number of sites in higher latitudes during the Upper and Final Palaeolithic in general, and pre-Magdalenian sites in particular, might reflect less research in higher altitudes. However, the trend of increasing latitudes after the LGM indicates that the general limit in latitude might be a reliable observation reflecting settlement choice rather than preservation or research bias. As can be expected, the largest differences in **slope** can be found between open-air and cave sites, where cave sites have more than triple the median (all open-air sites: 3°, all cave sites: 11°) and almost double the range between the 25$^{th}$ to 75$^{th}$ percentile,

indicating a higher variability of slopes. Interestingly, when assessing only open-air sites, a trend throughout the technocomplexes can be observed: Within all regional subsets, median slopes increase from the Aurignacian over the Gravettian and reach their relative maximum in the LGM. From LGM over Magdalenian to the Final Palaeolithic they decrease again, reaching their minimum in the latter phase. Given that steeper slopes provide more sheltered situations, this trend can indeed indicate settlement choice rather than discovery bias. The only exception to this trend are LGM open-air sites in the SW-section, which show a lower slope median than older or younger archaeological units. No plausible explanation for this exception can be found. Slope situations are prone to erosion and thus might strongly bias our picture of site distribution. However, at least for the NE-section, the clear temporal trend speaks against a systematic under-representation of sites with increasingly slope. Whether this factor is stronger in the SW-section needs to be explored in future studies.

Since modern to contemporaneous land use influences only the likelihood of discovery, the following parameters are more straight-forward to interpret.

In archaeological practice, many open-air sites are found on arable land, an impression supported by the high share of 22% of open-air sites found on cropland classes. However, our analysis shows that occupations are mostly underrepresented on **CLC classes** representing man made barren land, cropland, related agricultural areas and natural vegetation (see Figure A45 to Figure A53 in Appendix A). This finding is thus counterintuitive at first, but as cropland takes up 32% of the AOI, we observe a relative under-representation. However, the relative share of open-air sites increases considerably on the classes *discontinuous urban fabric, mineral extraction sites*, *green urban areas* and *vineyards* (factor 4.4 to 10.15), but decreases to an under-representation in areas covered by broad-leaved forests. This shows that probability of discovery of buried open-air sites is strongly increased on land cover classes that are related to deep interventions in the soil. This general observation holds also true within the two map sections, although with changing magnitudes. In the NE-section, site shares on *continuous* and *discontinuous urban fabric* are reduced, while NE open-air sites show considerably increased shares on *green urban areas* and *vineyards* (more than factor 12 with some exceptions). In the SW-section, the site over-representation on *continuous urban areas* is the highest (mean factor of 12.4, exception: MAG with 0.75). Vineyards (mean factor of 4.1, exception: FP with 0) show similar but decreased site shares in comparison to the NE section. The site shares on *mineral extraction sites* and *green urban areas*, however, are considerably decreased, non-existent and/or highly inconsistent through the technocomplexes. The relative share of **cave sites** on these CLC classes, on the other hand, shows the exact opposite trend: Considerably decreased shares of occupations on *discontinuous urban fabric*, *green urban areas* and *vineyards* (factor of 2.2 to 0) but considerably increased over-representation on *broad-leaved forests* (mean factor of 2.75; NE-section mean factor of 3.3), probably

because of a statistical connection between steeper slopes and forest land cover. This cross-correlation is also supported by the medium high contingency coefficient between slope and the CLC of 0.519.

When using the **aggregated CLC dataset** with 10 classes instead of 45, some general trends are easier to recognise (see Figure A54 to Figure A62 in Appendix A). Occupations from all technocomplexes are overrepresented on *urban fabric*. The over-representation is higher for open-air sites and in the SW section and lower for cave sites and in the NE section. High over-representations are also found on *mineral extraction sites*. The highest site shares within this class are found in the NE region. The aggregated classes *cropland, pastures* and *non-forested natural vegetated surfaces* show a considerable under-representation of sites across most technocomplexes and class subsets. For the unchanged class *vineyards*, high over-representations with exceptions can be observed for open-air sites while cave sites are generally underrepresented. Occupation shares on *forest* classes show the direct opposite, with over-representation of cave sites but under-representation of open-air sites. Within all occupation class subsets, a slight over-representation can be observed on the aggregated class other *agricultural surfaces*. The aggregated CLC classes *other man-made surfaces* and *other natural surfaces* show no regional, site- or archaeological unit-specific trend of over- or under-representation of Upper and Final Palaeolithic occupations.

Eventually, it becomes evident that the likelihood of site detection, particularly of open-air sites, is strongly increased on land cover classes which are highly frequented by humans and/or show deep interventions in the soil. Within the loess dominated NE open-air site subsets, the highest over-representations are found on soil-invasive land cover classes, which underlines the importance of the discovery context for deeply covered archaeological material.

Far lower but still informative deviations were found for the environmental variables of historical changes in built-up areas as defined by the **HYDE land use** model (see Figure A63 to Figure A71 in Appendix A). The use of the modelled multi-temporal built up area can be seen as a way of addressing discovery-related representability, since this parameter can be seen as a multi-temporal substitution to the urban CLC classes. Importantly, the statistical deviation from the expectation of this variable is based on mean values. When assessing all occupations, a general higher than expected mean can be observed (mostly between 1.5-2), indicating that sites are overrepresented in areas with increased construction activity in the timeframe from 1800 to 2000. Far higher positive deviations from the expected means (factor of up to 5.5) can be observed for open-air sites and in the NE section. Many negative deviations, however, can be observed for cave sites and in the SW section. A possible explanation might be the loess context, requiring a soil-invasive discovery context for buried sites in the NE region, while archaeological material in the SW region might have a higher probability of

random discovery. Positive or negative deviations are mostly consistent through the HYDE classification dates (1800-2000). When assessing differences between the technocomplexes, a trend of increasing positive deviation from the expected mean with increasing age and vice versa can be observed. As such, the highest positive deviations from the expected mean are found for the Aurignacian and Gravettian, while the lowest positive and highest negative deviations are found for the Magdalenian and Final Palaeolithic. The HYDE land use model highlight once more that with increasing intervention into the soil, buried sites have a higher probability of being discovered.

## 5.5 Conclusion

Both the over-representations of archaeological sites and the variable importance in the presence/absence modelling via MaxEnt show a clear result: The recent and sub-recent land use - clearly a discovery factor – has the by far strongest influence on the distribution of known sites. An example for surfaces that increase the probability of discovery are urban areas, vineyards and mineral extraction sites, showing an over-representation of sites by a factor of up to 13. The assessed land use dataset also shows the best performance in predicting the presence/absence of Palaeolithic sites according to the MaxEnt jack-knife variable importance. This means that land use has the highest power in predicting the presence/absence of sites. Another important finding is the chronological trend of an increasing over-representation on loess with increasing age. These observations show that the analysed site database is potentially strongly biased.

The under-/over-representation of sites on environmental variables relevant to the settlement factors as well as the statistical approaches of two-step classification and discriminant analysis show that the influence of the Palaeolithic settlement choice can still be detected in the biased database. These include the preference of microclimatically favourable gentle southern slopes, Jurassic and Cretaceous geological units for cave locations, and loess plains due to their attractiveness for hunter-gatherers. Chronological trends that show how the settlement behaviour changes through the archaeological units are not easily found. Only the preferences for microclimatic conditions with the highest over-representations on gentle southern slopes during the coldest archaeological units can be observed.

The two-step classification, discriminant analysis, and Naïve Bayes approach all indicate that the environmental differences between some of the predefined archaeological classes are strong enough to securely differentiate between them. This is especially true for site type (cave/open-air) and region (NE/SW). When trying to differentiate between the Upper Palaeolithic periods (AUR, GRA, LGM, MAG), however, a very low percentage of sites (<<50%) was accurately predicted by both the discriminant analysis and Naïve Bayes. For the settlement factors, this either means that the differences in

settlement choice between the earlier technocomplexes are too small for a clear differentiation, or that the distribution is biased to a degree where the differences are overwritten. The environmental differences between Upper Palaeolithic sites on the one hand and Final Palaeolithic sites on the other, however, seem to be marked enough to allow for a clear differentiation. With all necessary caution arising from this study, the general differences in Upper and Final Palaeolithic site distribution thus seem to be reflect at least coarsely the presence of past hunter-gatherers in the landscape.

The fundamental sampling bias revealed in this study is primarily problematic for inductive approaches, where environmental similarities are directly extracted from the distribution of known sites. This is particularly relevant for predictive modelling, where the presence of sites is interpreted as a proxy for Palaeolithic decision making. Ignoring land use might lead to misinterpretations caused by cross-correlation between this discovery factor and other environmental variables. Pleistocene fluvial terraces in mountainous regions, for instance, show increased Palaeolithic site frequencies. This could be interpreted as the result of past settlement strategies. However, the fact that discovery-relevant modern human activity, such as urbanisation, construction of infrastructure, and even mineral extraction is mainly taking place on these easily accessible morphological units, advise against such a reading of the record. We thus recommend assessing the under-/over-representation of sites on surfaces relevant to the discovery factors parallel to the assessment of settlement factors to assess potential sampling biases individually for each study area and archaeological site dataset.

For further studies on the sampling bias of Upper and Final Palaeolithic sites, we recommend building on this pilot study by adding more occupations to the dataset and more environmental variables to the settlement and discovery factors and suggest to conduct case studies on regional scales where more environmental variables are available at a higher spatial resolution. Due to the current positive trend in openly available geodatasets, we see great potential for such approaches in the future.

## 5.6 Data availability

As we think that there is even greater potential in the database that we have compiled, we are making it openly accessible for further studies. The vector-based site dataset (shapefile, .shp) with all 4194 occupations and the result of the spatial queries with the environmental variables as attributes is available for download under the following link: https://doi.org/10.18154/RWTH-2023-06762. All used environmental geodatasets are already publicly available, and we refer to the respective website for individual download.

# 6 Synthesis

Each of the three studies presented in the framework of this dissertation is a scientific contribution towards a better understanding of the respective study area and a methodological advancement within the respective field of research. In this chapter, the main results of these studies are discussed within the framework of the research question of this dissertation.

In the first study of this dissertation, visual pattern extraction from Sentinel-1 data using continuous wavelet transform (CWT) revealed several spatial and temporal dynamics within the Western Mongolia dune field of Bor Khyar. This methodology allowed the differentiation of active and inactive parts of the dune field as well as the identification of seasonal and temporal differences in morphology, which fit the known wind regimes. These results were validated using multi-temporal digital elevation models (DEMs). As the dune field and its surroundings have been the object of investigation in numerous past and ongoing efforts on paleo-climate reconstruction (see, e.g., Klinge et al. 2017; Lehmkuhl et al. 2018; Klinge and Sauer 2019), a better understanding of it can aid future studies. As such, the results were already applied in the planning and pre-analysis of a field campaign in the summer of 2023, aimed at extracting samples from inactive paleo-dunes for luminescence dating.

In the second study of this dissertation, the pattern extraction was based on a combination of a deductive approach and machine learning. This methodology allowed the creation of an explanatory archaeological predictive model with clearly formulated causalities between paleo-environmental predictors and the probability of Upper Palaeolithic occupation. Due to the explanatory nature of the model, it highlights spatial dynamics which match the archaeological consensus on human decision making in the region. However, due to the low availability of training data, the results were only verified with the input data, resulting in good but most probably overestimated accuracies. As such, an application in heritage management is only recommended after additional validation. However, as this is the first study on the probability of Upper Palaeolithic sites in an area where already the very limited evidence has provided significant contributions to the understanding of human behaviour and human-environment interactions, a scientific application of the resulting model is very promising.

Through a complex combination of geospatial and geostatistical methods, the third study of this dissertation managed to extract several patterns from geospatial big data representing Upper and Final Palaeolithic settlement factors and modern to contemporary discovery factors. The rather surprising main result is that recent land use has the by far strongest influence on the distribution of known sites. This means that the analysed site database is potentially strongly biased. The influence of settlement factors, however, can still be cautiously traced within the dataset. This is especially true for differences between the Upper and Final Palaeolithic site distribution. While these results, first and foremost, have

an impact on how the analysed database is handled in the future, the determined high influence of the sampling bias on a European scale also has important implications for other fields of research within geoarchaeology. This is especially true for inductive approaches, where the representativity of site-based training datasets is a prerequisite for valid results.

In addition to these individual results, the three studies can also be seen as methodological contributions as they present novel and innovative approaches on how to effectively implement geospatial big data in environmental and geoarchaeological research.

The main contribution of the first study in this regard is the result that the easily accessible Sentinel-1 ground range detected (GRD) product can be used to extract complex dune field morphology – a feat that previously only was performed based on less approachable derived products such as interferometry. While one possible approach on how to extract a pattern from this source of remote sensing big data is presented in the study, this result also opens up new opportunities for other forms of extraction such as e.g. deep learning dune pattern mapping or similar.

The second study makes an important contribution towards the extraction of viable information from environmental big data based on limited training data. This limited evidence is a common problem in regions where the archaeological record is invisible and/or inaccessible due to massive built-up of sediments (such as e.g. in the Carpathian basin, Chu 2018; Nett et al. 2021). For these situations, the study presents an approach that does not rely fully on inductive analysis and thereby produces a comprehensible and explanatory predictive model.

The main methodological contribution of the third study is the logical addition of sampling biases into the field of geoarchaeological site distribution analysis. While numerous studies have analysed site distributions within the paleo-environment, this is the first study that compares this to geospatial datasets on the probability of discovery. Due to the surprisingly clear result and the associated implications, the topic of sampling biases might receive additional attention in the research community and the addition of discovery factors might be adapted in future geoarchaeological studies.

These three studies are combined in this dissertation to answer the research question on how to effectively utilize geospatial big data in the fields of environmental science and geoarchaeology. To this end, the dissertation presents not only novel methodologies that effectively utilize geospatial big data, but also suggestions and implications for future studies in the field. In addition, section 1.2 outlines the four main challenges that researchers face when utilizing geospatial big data and gives suggestions on how to address them. These challenges include (I) the selection of a fitting geospatial representation of the object of investigation, (II) a data handling that includes filtering and

harmonization, (III) the implementation and communication of a fitting methodology for the extraction of patterns, and (IV) an interpretation where these challenges are considered. This rough guideline for the utilization of geospatial big data is additionally visualized in Figure 23.

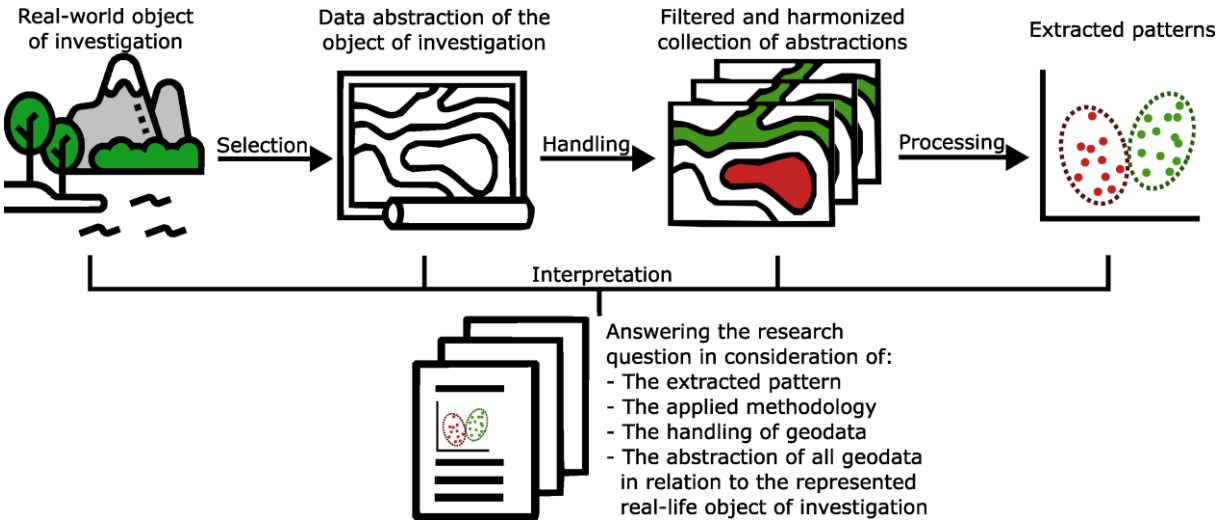



*Figure 23: Visual representation of the challenges of geospatial big data utilization. For a detailed description of the single challenges and a guideline of how they should be addressed, see section 1.2.*

As such, this dissertation can be seen as food for thought and a guideline within the rapidly evolving fields of environmental science and geoarchaeology. While it might serve as a foundation and inspiration for future studies, the utilization of geospatial big data in these fields underlies greater mechanics. The most promising developments in these regards are (I) the positive trend towards open accessibility of governmental, commercial, and scientific geospatial big datasets, (II) the efforts towards international data compilation in harmonized databases, and (III) the outsourcing of expensive geospatial processing onto powerful web-based processing platforms. In consideration of these current trends, the utilization of geospatial big data faces a bright future in environmental research and geoarchaeology.

# 7 Acknowledgements / Danksagung

The last three years of work towards this dissertation have been special to me. Not only from a scientific perspective but also personally. I have met a lot of friendly faces and had the opportunity to go to new and exciting places. For this, I want to thank a lot of people that directly or indirectly aided me in this time.

An erster Stelle möchte ich mich bei meinem Doktorvater **Prof. Dr. Frank Lehmkuhl** bedanken. Er hat mich bereits früh als studentische Hilfskraft ins Boot geholt und mich für die Möglichkeit einer Promotion begeistert. Auch hat er schon in HiWi-Zeiten den Weg für die ersten Veröffentlichungen gebahnt und mich auf das wissenschaftliche Arbeiten vorbereitet. Während der Promotion hat er viele spannende Projekte für mich an Land gezogen und mir die Freiheit gelassen, diese nach meinen Vorstellungen zu bearbeiten. Auf seine Meinung zu Ergebnissen und Abbildungen konnte man sich immer verlassen. Ohne seine Betreuung hätte ich die Promotion sicherlich nicht in der kurzen Zeit bewältigen können.

Als zweites danke ich **Prof. Dr. Andreas Maier**, nicht nur für die Zweitbegutachtung, sondern auch dafür, dass er mich für die Archäologie begeistert hat. In gemeinsamen Projekten hat er mir die archäologischen Fragestellungen und fachlichen Hintergründe immer auf eine für Geographen verständliche Ebene herunterbrechen können. Außerdem hat er sehr wertvolle Denkanstöße geliefert und war immer für Zwischenfragen und Korrekturlesungen erreichbar. In diesem Rahmen möchte ich auch **Dr. Isabell Schmidt** danken, die mir den wertvollen Fundstellendatensatz zur Verfügung gestellt hat und bei den vielen Rückfragen dazu stets geholfen hat. Auch bin ich dankbar dafür, dass aus unserem initialen Projekt noch weitere Interdisziplinäre Zusammenarbeiten entstanden sind.

Ein besonderer Dank gilt auch meinen aktuellen und ehemaligen Kolleg:innen am Geographischen Institut der RWTH Aachen. Ich habe das angenehme Arbeitsklima immer genossen und mich über den teils fachlichen Austausch gefreut. Diese positive Arbeitsumgebung hat dazu beigetragen, dass ich die Zeit hier in guter Erinnerung behalten werde. An chronologisch erster Stelle gilt mein besonderer Dank **Dr. Stephan Pötter**, mit dem ich in der ersten Zeit meiner Promotion das Büro, den Musikgeschmack und die Vorliebe für Kaassoufflé teilen durfte. Er hat mich fachlich besonders bei der Kartenerstellung unterstützt und erfolgreich für alternative Formen der Kaffeezubereitung begeistert. Einen besonderen Dank möchte ich außerdem **Alexandra Weber** aussprechen, mit der ich das Glück hatte, knappe zwei Jahre das Büro zu teilen. Sie hat meine Essensgewohnheiten nie verurteilt und auf Anfragen zum Korrekturlesen immer mit „without hesitation, sir“ geantwortet. Auch hat sie sich nicht von meinen vielen Formulierungsfragen ermüden lassen und sich auch in stressigen Phasen für meine fachlichen Fragen, weniger fachlichen Erzählungen und Mittagessenswünsche Zeit genommen. Auch

wenn ihre fachliche Kompetenz manchmal etwas einschüchternd sein kann, habe ich mich immer auf die Zeit mit ihr im Büro gefreut. Bei **Dr. Philipp Schulte** und **Johannes Keßels** möchte ich mich für die vielen erfolgreichen Ablenkungen von der Arbeit bedanken. Die manchmal mehr und manchmal weniger fachlichen Gespräche und Wortwitze haben mir die Zeit bis um 16:10 versüßt. **Apl. Prof. Dr. Wolfgang Römer** möchte ich ganz herzlich dafür danken, dass er sich für meine statistischen Fragen Zeit genommen hat und für den es ein Anliegen war, dass ich die Verfahren nicht nur anwende sondern auch verstehe. Für die schöne Zeit in der Mongolei und das Aushalten meiner damit verbundenen überschwänglichen Schlag- und Buddelfertigkeit danke ich **apl. Prof. Dr. Georg Stauch** und **Dennis Wolf**. Für die technische und administrative Hilfe während meiner Zeit am Lehrstuhl danke ich **Anja Knops** und **Gernot Frommelt**, ohne die ich meine Urlaubstage bestimmt vergessen hätte und niemals aus dem Homeoffice hätte arbeiten können. Danke auch an **Max**, **Julian**, **Viktor** und die anderen studentischen Hilfskräfte, durch deren methodischen Fragen ich mich sinnvoll gefühlt habe und die mich das eine oder andere Mal tatkräftig unterstützt haben.

I also want to thank, if not already mentioned, my co-authors for the teamwork on our three papers. I want to thank **Imen Turki** for the introduction into the world of the continuous wavelet transform, **Catrina Brüll** for the civil engineering perspective and **Marc Händel** and **Thomas Einwögerer** for the expert knowledge on the Palaeolithic in Lower Austria – both in the field and digitally.

Von Herzen danke ich auch meiner **Familie**, die meine Erzählungen über Geoarchäologie und Dünenfelder immer sehr spannend fand. Ihr habt meine Entscheidung zur Promotion nie hinterfragt und mir durch interessiertes Nachfragen immer die Möglichkeit gegeben, mich öffentlich für mein Fach zu begeistern. Mein ganz besonderer Dank gilt natürlich meinen Eltern **Uta** und **Jürgen**, von denen ich mich in jeder Entscheidung unterstützt gefühlt habe. Abgesehen davon sind sie die besten Eltern und haben mir eine tolle Kindheit beschert, die die perfekte Grundlage für ein selbstbewusstes und sorgloses Leben ist. Meiner tollsten Schwester **Hilde** möchte ich außerdem dafür danken, dass sie immer stolz auf mich ist und mit ihrer ehrfürchtigen Art des Zuhörens dafür sorgt, dass ich mich bei meinen fachlichen Ausführungen besonders schlau fühle. Meinen Katern **Bruno** und **Berlioz** möchte ich an dieser Stelle auch für ihre unfreiwillige Flauschigkeit und stressregulierende Wirkung danken.

Mein allergrößter Dank gilt natürlich **Hjördis**, die mich in guten so wie schlechten Zeiten unterstützt und immer an mich geglaubt hat. Besonders in den stressigen letzten Wochen der Promotion hast du mir den Rücken freigehalten und meine teilweise schlechte Laune ertragen. Dank dir freue ich mich jeden Tag auf den Feierabend und jede Woche auf das Wochenende.

## Acknowledgements for chapter 3: Assessing complex aeolian dune field morphology and evolution with Sentinel-1 SAR imagery – possibilities and limitations

This work was carried out in the frame of the exploratory research space (ERS), funded by the Federal Ministry of Education and Research (BMBF) and the Ministry of Culture and Science of the German State of North Rhine-Westphalia (MKW) under the Excellence Strategy of the Federal Government and the Länder [project ID G:(DE-82)EXS-SF-OPSF612]. We thank Alexandra Weber for the valuable help in reviewing the figures and structure, Damian Stawinoga for testing the interferometric approach and Martha Wingen for the conceptual and administrative help. We would also like to thank the reviewers for the time and effort invested into reviewing the manuscript. Their comments and suggestions helped in improving the quality of the manuscript significantly.

## Acknowledgements for chapter 4: Upper Palaeolithic site probability in Lower Austria – a geoarchaeological multi-factor approach

All investigations were carried out in the framework of the CRC 806 ‘Our way to Europe’, Projektnummer 57444011 – SFB 806, funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation). We thank Stephan Pötter for his helpful comments and internal proof reading. Finally, we are grateful to the reviewers Michael Kempf, Thomas Whitley, Steven Bernard and Chris Orton as well as the editor Piraye Hacigüzeller for their helpful comments, which greatly improved the quality of the map and the manuscript.

## Acknowledgements for chapter 5: Approaching sampling biases of Upper and Final Palaeolithic sites – a geospatial analysis of a European dataset

We thank our two anonymous reviewers for their valuable comments. All investigations were carried out in the framework of the CRC 806 ‘Our way to Europe’, Projektnummer 57444011 – SFB 806, funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation).

# 8 References

Abate, N.; Roubis, D.; Vitale, V.; Sileo, M.; Sogliani, F.; Masini, N.; Lasaponara, R. (2022): Integrated use of multi-temporal multi-sensor and multiscale Remote Sensing data for the understanding of archaeological contexts: the case study of Metaponto, Basilicata. In *J. Phys.: Conf. Ser.* 2204 (1), p. 12020. DOI: 10.1088/1742-6596/2204/1/012020.

Agapiou, A. (2017): Remote sensing heritage in a petabyte-scale: satellite data and heritage Earth Engine© applications. In *International Journal of Digital Earth* 10 (1), pp. 85–102. DOI: 10.1080/17538947.2016.1250829.

Agapiou, A. (2021): Multi-Temporal Change Detection Analysis of Vertical Sprawl over Limassol City Centre and Amathus Archaeological Site in Cyprus during 2015-2020 Using the Sentinel-1 Sensor and the Google Earth Engine Platform. In *Sensors* 21 (5), p. 1884. DOI: 10.3390/s21051884.

Agapiou, A.; Vionis, A.; Papantoniou, G. (2021): Detection of Archaeological Surface Ceramics Using Deep Learning Image-Based Methods and Very High-Resolution UAV Imageries. In *Land* 10 (12), p. 1365. DOI: 10.3390/land10121365.

Aguiar-Conraria, L.; Soares, M. J. (2014): The Continuous Wavelet Transform: Moving Beyond Uni- and Bivariate Analysis. In *Journal of Economic Surveys* 28 (2), pp. 344–375. DOI: 10.1111/joes.12012.

Alday, A.; Domingo, R.; Sebastián, M.; Soto, A.; Aranbarri, J.; González-Sampériz, P. et al. (2018): The silence of the layers: Archaeological site visibility in the Pleistocene-Holocene transition at the Ebro Basin. In *Quaternary Science Reviews* 184, pp. 85–106. DOI: 10.1016/j.quascirev.2017.11.006.

Al-Ghamdi, K.; Hermas, E. (2015): Assessment of dune migration hazards against landuse northwest Al-lith City, Saudi Arabia, using multi-temporal satellite imagery. In *Arab J Geosci* 8 (12), pp. 11007–11018. DOI: 10.1007/s12517-015-1947-8.

Allen, K.; Green, S. W.; Zubrow, E. B. W. (1990): Interpreting space : GIS and archaeology. London, UK: Taylor & Francis.

Altaweel, M.; Khelifi, A.; Li, Z.; Squitieri, A.; Basmaji, T.; Ghazal, M. (2022): Automated Archaeological Feature Detection Using Deep Learning on Optical UAV Imagery: Preliminary Results. In *Remote Sensing* 14 (3), p. 553. DOI: 10.3390/rs14030553.

Alwi Muttaqin, L.; Heru Murti, S.; Susilo, B. (2019): MaxEnt (Maximum Entropy) model for predicting prehistoric cave sites in Karst area of Gunung Sewu, Gunung Kidul, Yogyakarta. In. Sixth Geoinformation Science Symposium: International Society for Optics and Photonics (11311), 113110B. DOI: 10.1117/12.2543522

Amani, M.; Ghorbanian, A.; Ahmadi, S. A.; Kakooei, M.; Moghimi, A.; Mirmazloumi, S. M. et al. (2020): Google Earth Engine Cloud Computing Platform for Remote Sensing Big Data Applications: A Comprehensive Review. In *IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing* 13, pp. 5326–5350. DOI: 10.1109/jstars.2020.3021052.

Amin, A.; Seif, E.-S. S. A. (2019): Environmental Hazards of Sand Dunes, South Jeddah, Saudi Arabia: An Assessment and Mitigation Geotechnical Study. In *Earth Syst Environ* 3 (2), pp. 173–188. DOI: 10.1007/s41748-019-00100-5.

An, C.B.; Chen, F.H.; Barton, L. (2008): Holocene environmental changes in Mongolia: A review. In *Global and Planetary Change* 63 (4), pp. 283–289. DOI: 10.1016/j.gloplacha.2008.03.007.

Armero, C.; García-Donato, G.; Jiménez-Puerto, J.; Pardo-Gordó, S.; Bernabeu, J. (2020): A Bayesian naïve Bayes classifier for dating archaeological sites. In: Irigoien, I; Dae-Jin Lee, D-J.; Martínez-Minaya, J.; María Xosé Rodríguez-Álvarez; M.X.(Eds.). Proceedings of the 35th International Workshop on Statistical Modelling (IWSM). July 20-24, 2020 - Bilbao, Basque Country, Spain.

Asch, K. (2003): The 1:5 Million International Geological Map of Europe and Adjacent Areas. Development and Implementation of a GIS-enabled Concept. Stuttgart: Schweizerbart'sche Verlagsbuchhandlung. ISBN: 978-3-510-95903-7.

ASF (2022): Alaskan Satellite Facility baseline tool [Website]. Available at: search.asf.alaska.edu

Aura, J. E.; Tiffagom, M.; Jordá Pardo, J. F.; Duarte, E.; La Fernández de Vega, J.; Santamaria, D. et al. (2012): The Solutrean–Magdalenian transition: A view from Iberia. In *Quaternary International* 272-273, pp. 75–87. DOI: 10.1016/j.quaint.2012.05.020.

Backhaus, K.; Erichson, B.; Plinke, W.; Weiber, R. (2018): Multivariate Analysemethoden. Berlin, Heidelberg: Springer Berlin Heidelberg. ISBN 978-3-662-56654-1.

Balla, A.; Pavlogeorgatos, G.; Tsiafakis, D.; Pavlidis, G. (2014): Recent Advances in Archaeological Predictive Modeling for Archeological Research and Cultural Heritage Management. In *Mediterranean archaeology & Archaeometry* 14 (4).

Banks, W. E.; d'Errico, F.; Peterson, A. T.; Vanhaeren, M.; Kageyama, M.; Sepulchre, P. et al. (2008): Human ecological niches and ranges during the LGM in Europe derived from an application of eco-cultural niche modeling. In *Journal of Archaeological Science* 35 (2), pp. 481–491. DOI: 10.1016/j.jas.2007.05.011.

Banks, W. E.; d'Errico, F.; Zilhão, J. (2013): Human-climate interaction during the Early Upper Paleolithic: testing the hypothesis of an adaptive shift between the Proto-Aurignacian and the Early Aurignacian. In *Journal of Human Evolution* 64 (1), pp. 39–55. DOI: 10.1016/j.jhevol.2012.10.001.

Banks, W. E.; Zilhão, J.; d'Errico, F.; Kageyama, M.; Sima, A.; Ronchitelli, A. (2009): Investigating links between ecology and bifacial tool types in Western Europe during the Last Glacial Maximum. In *Journal of Archaeological Science* 36 (12), pp. 2853–2867. DOI: 10.1016/j.jas.2009.09.014.

Benjaminsen, T. A.; Hiernaux, P. (2019): From Desiccation to Global Climate Change: A History of the Desertification Narrative in the West African Sahel, 1900-2018. In *Global Environment* 12 (1), pp. 206–236. DOI: 10.3197/ge.2019.120109.

Bertran, P.; Bosq, M.; Borderie, Q.; Coussot, C.; Coutard, S.; Deschodt, L. et al. (2021): Revised map of European aeolian deposits derived from soil texture data. In *Quaternary Science Reviews* 266, p. 107085. DOI: 10.1016/j.quascirev.2021.107085.

Bertran, P.; Liard, M.; Sitzia, L.; Tissoux, H. (2016): A map of Pleistocene aeolian deposits in Western Europe, with special emphasis on France. In *J. Quaternary Sci.* 31 (8), e2909. DOI: 10.1002/jqs.2909.

Binford, L.R. (1980): Willow Smoke and Dogs' Tails: Hunter-Gatherer Settlement Systems and Archaeological Site Formation. In *American Antiquity* 45 (1), 4-20.

Binford, L. R. (2019): Constructing frames of reference. An analytical method for archaeological theory building using hunter-gatherer and environmental data sets. First paperback printing. Berkeley [u.a.]: University of California Press.

Blumberg, D. G. (1998): Remote Sensing of Desert Dune Forms by Polarimetric Synthetic Aperture Radar (SAR). In *0034-4257* 65 (2), pp. 204–216. DOI: 10.1016/S0034-4257(98)00028-5.

Blumberg, D. G. (1998): Remote Sensing of Desert Dune Forms by Polarimetric Synthetic Aperture Radar (SAR). In *Remote Sensing of Environment* 65 (2), pp. 204–216. DOI: 10.1016/S0034-4257(98)00028-5.

Blumberg, D. G. (2006): Analysis of large aeolian (wind-blown) bedforms using the Shuttle Radar Topography Mission (SRTM) digital elevation data. In *Remote Sensing of Environment* 100 (2), pp. 179–189. DOI: 10.1016/j.rse.2005.10.011.

Bocquet-Appel, J.-P.; Demars, P. Y.; Noiret, L.; Dobrowsky, D. (2005): Estimates of Upper Palaeolithic meta-population size in Europe from archaeological data. In *Journal of Archaeological Science* 32 (11), pp. 1656–1668. DOI: 10.1016/j.jas.2005.05.006.

Boemke, B.; Einwögerer, T.; Händel, M.; Lehmkuhl, F. (2022): Upper Palaeolithic site probability in Lower Austria – a geoarchaeological multi-factor approach. In *Journal of Maps* 18 (4), pp. 610–618. DOI: 10.1080/17445647.2021.2009926.

Boemke, B.; Maier, A.; Schmidt, I.; Römer, W.; Lehmkuhl, F. (2023a): Testing the representativity of Palaeolithic site distribution: The role of sampling bias in the European Upper and Final Palaeolithic record. In *Quaternary Science Reviews* 316. DOI: 10.1016/j.quascirev.2023.108220.

Boemke, B.; Turki, I.; Brüll, C.; Lehmkuhl, F. (2023b): Assessing complex aeolian dune field morphology and evolution with Sentinel-1 SAR imagery – Possibilities and limitations. In *Aeolian Research* 62. DOI: 10.1016/j.aeolia.2023.100876.

Bonhage, A.; Eltaher, M.; Raab, T.; Breuß, M.; Raab, A.; Schneider, A. (2021): A modified Mask region-based convolutional neural network approach for the automated detection of archaeological sites on high-resolution light detection and ranging-derived digital elevation models in the North German Lowland. In *Archaeological Prospection* 28 (2), pp. 177–186. DOI: 10.1002/arp.1806.

Brandolini, F.; Domingo-Ribas, G.; Zerboni, A.; Turner, S. (2021): A Google Earth Engine-enabled Python approach for the identification of anthropogenic palaeo-landscape features. In *Open Research Europe* 1 (22). DOI: 10.12688/openreseurope.13135.2.

Braun, A. (2021): Retrieval of digital elevation models from Sentinel-1 radar data – open applications, techniques, and limitations. In *Open Geosciences* 13 (1), pp. 532–569. DOI: 10.1515/geo-2020-0246

Bubenzer, O.; Bolten, A. (2008): The use of new elevation data (SRTM/ASTER) for the detection and morphometric quantification of Pleistocene megadunes (draa) in the eastern Sahara and the southern Namib. In *Geomorphology* 102 (2), pp. 221–231. DOI: 10.1016/j.geomorph.2008.05.003.

Buckley, S. M.; Agram, P. S.; Belz, J. E.; Crippen, R. E.; Gurolla, E. M.; Hensley, S. et al. (2020): NASADEM User's Guide. Pasadena, California.

Burke, A.; Kageyama, M.; Latombe, G.; Fasel, M.; Vrac, M.; Ramstein, G.; James, P. M. (2017): Risky business: The impact of climate and climate variability on human population dynamics in Western Europe during the Last Glacial Maximum. In *Quaternary Science Reviews* 164, pp. 217–229. DOI: 10.1016/j.quascirev.2017.04.001.

C3S (2022): ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate [Dataset]. In *Copernicus Climate Change Service*. Available at: cds.climate.copernicus.eu

Campana, Stefano; Piro, Salvatore (2008): Seeing the unseen. Geophysics and landscape archaeology. With assistance of Stefano Campana, Salvatore Piro. Boca Raton: CRC Press Taylor & Francis Group (A Balkema Book).

Carmichael, B.; Daly, C.; Fatorić, S.; Macklin, M.; McIntyre-Tamwoy, S.; Pittungnapoo, W. (2023): Global Riverine Archaeology and Cultural Heritage: Flood-Risk Management and Adaptation for the Anthropogenic Climate Change Crisis. In *Climate* 11 (10), p. 197. DOI: 10.3390/cli11100197.

Casana, J. (2014): Regional-scale archaeological remote sensing in the age of big data: Automated site discovery vs. brute force methods. In *Advances in Archaeological Practice*, pp. 222–233. DOI: 10.7183/2326-3768.2.3.222.

Cazenave, P. W.; Dix, J. K.; Lambkin, D. O.; McNeill, L. C. (2013): A method for semi-automated objective quantification of linear bedforms from multi-scale digital elevation models. In *Earth Surface Processes and Landforms* 38 (3), pp. 221–236. DOI: 10.1002/esp.3269.

Chen, M.; Mao, S.; Liu, Y. (2014): Big Data: A Survey. In *Mobile Netw Appl* 19 (2), pp. 171–209. DOI: 10.1007/s11036-013-0489-0.

Chen, P. C. L.; Zhang, C.-Y. (2014): Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. In *Information Sciences* 275, pp. 314–347. DOI: 10.1016/j.ins.2014.01.015.

Chu, W. (2018): The Danube Corridor Hypothesis and the Carpathian Basin: Geological, Environmental and Archaeological Approaches to Characterizing Aurignacian Dynamics. In *J World Prehist* 31 (2), pp. 117–178. DOI: 10.1007/s10963-018-9115-1.

Chu, W.; Nett, J. J. (2021): The past in dust: current trends and future directions in Pleistocene geoarcheology of European loess. In *J. Quaternary Sci.*, Article jqs.3388. DOI: 10.1002/jqs.3388.

Coco, E.; Iovita, R. (2020): Time-dependent taphonomic site loss leads to spatial averaging: implications for archaeological cultures. In *Humanit Soc Sci Commun* 7 (1). DOI: 10.1057/s41599-020-00635-3.

Cooper, A.; Green, C. (2016): Embracing the Complexities of 'Big Data' in Archaeology: the Case of the English Landscape and Identities Project. In *J Archaeol Method Theory* 23 (1), pp. 271–304. DOI: 10.1007/s10816-015-9240-4.

Dana Negula, I.; Moise, C.; Lazăr, A. M.; Rișcuța, N. C.; Cristescu, C.; Dedulescu, A. L. et al. (2020): Satellite Remote Sensing for the Analysis of the Micia and Germisara Archaeological Sites. In *Remote Sensing* 12 (12), p. 2003. DOI: 10.3390/rs12122003.

Davies, J.; Ogali, C.; Laban, P.; Metternicht, G. (2015): Homing in on the range: enabling investments for sustainable land management. In *Technical brief* (29.01.2015), IUCN and CEM.

Davis, J.C., (1986): Statistics and data analysis in geology. 2nd Ed., John Wiley and Sons, New York.

De Jager, Alfred; Vogt, Jürgen (2007): Rivers and Catchments of Europe - Catchment Characterisation Model (CCM). European Commission, Joint Research Centre (JRC) [Dataset]. Available at: http://data.europa.eu/89h/fe1878e8-7541-4c66-8453-afdae7469221

Delgado Blasco, J. M.; Chini, M.; Verstraeten, G.; Hanssen, R. F. (2020): Sand Dune Dynamics Exploiting a Fully Automatic Method Using Satellite SAR Data. In *Remote Sensing* 12 (23), p. 3993. DOI: 10.3390/rs12233993.

DLR (2018): 60 Petabytes for the German Satellite Data Archive D-SDA [Article]. Available at: https://www.dlr.de/eoc/en/desktopdefault.aspx/tabid-12632/22039_read-51751

Dong, P. (2015): Automated measurement of sand dune migration using multi-temporal lidar data and GIS. In *International Journal of Remote Sensing* 36 (21), pp. 5426–5447. DOI: 10.1080/01431161.2015.1093192.

Dong, Z.; Qian, G.; Luo, W.; Zhang, Z.; Lv, P. (2013): Dune types and their distribution in the Kumtagh Sand Sea, northwestern China. In *Zeitschrift für Geomorphologie* 57 (2). DOI: 10.1127/0372-8854/2012/0096

Dörwald, L.; Lehmkuhl, F.; Walk, J.; Delobel, L.; Boemke, B.; Baas, A. et al. (2023): Dune movement under climatic changes on the north-eastern Tibetan Plateau as recorded by long-term satellite observation versus ERA-5 reanalysis. In *Earth Surface Processes and Landforms* 48 (13), pp. 2613–2629. DOI: 10.1002/esp.5651.

Duke, C.; Steele, J. (2010): Geology and lithic procurement in Upper Palaeolithic Europe: a weights-of-evidence based GIS model of lithic resource potential. In *Journal of Archaeological Science* 37 (4), pp. 813–824. DOI: 10.1016/j.jas.2009.11.011.

Ebert, David (2004): Applications of archaeological GIS: Canadian Journal of Archaeology/Journal Canadien d'Archéologie.

Effat, H. A.; Hegazy, M. N.; Haack, B. (2011): Mapping sand dunes risk related to their terrain characteristics using SRTM data and cartographic modeling. In *Journal of Land Use Science* 6 (4), pp. 231–243. DOI: 10.1080/1747423X.2010.511680.

Ehlers, Jürgen; Gibbard, Philip L.; Hughes, Philip D. (Eds.) (2011): Quaternary glaciations - extent and chronology. A closer look. Amsterdam, Boston: Elsevier (Developments in quaternary sciences, v. 15).

Einwögerer, T.; Friesinger, F.; Händel, M.; Neugebauer-Maresch, C.; Simon, U.; Teschler-Nicola, M. (2006). Upper Palaeolithic infant burials. In *Nature* 444, 285.

Einwögerer, T.; Händel, M.; Simon, U.; Masur, A.; Neugebauer-Maresch, C. (2014): Upper Palaeolithic occupation in the Wachtberg area of Krems: The evidence of surveys, sections and core samples. In *Quaternary International* 351, pp. 50–66. DOI: 10.1016/j.quaint.2014.01.011.

Einwögerer, T.; Händel, M.; Simon, U.; Neugebauer-Maresch, C. (2014). Upper Palaeolithic occupation in the Wachtberg area of Krems: The evidence of surveys, sections and core samples. In *Quaternary International* 351: 50-66. DOI: 10.1016/j.quaint.2014.01.011.

El Gammal, E. S. A.; El Gammal, A. E. D. A. (2010): Hazard impact and genetic development of sand dunes west of Samalut, Egypt. In *The Egyptian Journal of Remote Sensing and Space Science* 13 (2), pp. 137–151. DOI: 10.1016/j.ejrs.2010.02.001.

El-Behaedi, R.; Ghoneim, E. (2018): Flood risk assessment of the Abu Simbel temple complex (Egypt) based on high-resolution spaceborne stereo imagery. In *2352-409X* 20, pp. 458–467. DOI: 10.1016/j.jasrep.2018.05.019.

Elfadaly, A.; Abate, N.; Masini, N.; Lasaponara, R. (2020): SAR Sentinel 1 Imaging and Detection of Palaeo-Landscape Features in the Mediterranean Area. In *Remote Sensing* 12 (16), p. 2611. DOI: 10.3390/rs12162611.

Elfadaly, A.; Abutaleb, K.; Naguib, D. M.; Lasaponara, R. (2022a): Detecting the environmental risk on the archaeological sites using satellite imagery in Basilicata Region, Italy. In *1110-9823* 25 (1), pp. 181–193. DOI: 10.1016/j.ejrs.2022.01.007.

Elfadaly, A.; Shams, A. H.; Elbehery, W.; Elftatry, M.; Wafa, O.; Hiekl, A. M. A. et al. (2022b): Revealing the paleolandscape features around the archaeological sites in the northern Nile Delta of Egypt using radar satellite imagery and GEE platform. In *Archaeological Prospection* 29 (3), pp. 369–384. DOI: 10.1002/arp.1860.

Elith, J.; Phillips, S. J.; Hastie, T.; Dudík, M.; Chee, Y. E.; Yates, C. J. (2011): A statistical explanation of MaxEnt for ecologists. In *Diversity and Distributions* 17 (1), pp. 43–57. DOI: 10.1111/j.1472-4642.2010.00725.x.

Eljack, E.; Csaplovics, E.; Adam, H. (2010): Mapping and assessment of sand encroachment on the Nile River northern Sudan, by means of remote Sensing and GIS. In *Proceedings of the Conference on International Research on Food Security, Natural Resource Management and Rural Development*.

Enkhbold, A.; Khukhuudei, U.; Kusky, T.; Tsermaa, B.; Doljin, D. (2021): Depression morphology of Bayan Lake, Zavkhan province, Western Mongolia: implications for the origin of lake depression in Mongolia. In *Physical Geography*, pp. 1–26. DOI: 10.1080/02723646.2021.1899477.

Fabbri, S.; Grottoli, E.; Armaroli, C.; Ciavola, P. (2021): Using High-Spatial Resolution UAV-Derived Data to Evaluate Vegetation and Geomorphological Changes on a Dune Field Involved in a Restoration Endeavour. In *Remote Sensing* 13 (10), p. 1987. DOI: 10.3390/rs13101987.

Farr, T. G.; Rosen, P. A.; Caro, E.; Crippen, R.; Duren, R.; Hensley, S. et al. (2007): The Shuttle Radar Topography Mission. In *Reviews of Geophysics* 45 (2), Article 2005RG000183. DOI: 10.1029/2005RG000183.

Farr, T. G.; Rosen, P. A.; Caro, E.; Crippen, R.; Duren, R.; Hensley, S. et al. (2007): The Shuttle Radar Topography Mission. In *Rev. Geophys.* 45 (2). DOI: 10.1029/2005RG000183.

Fattore, C.; Abate, N.; Faridani, F.; Masini, N.; Lasaponara, R. (2021): Google Earth Engine as Multi-Sensor Open-Source Tool for Supporting the Preservation of Archaeological Areas: The Case Study of Flood and Fire Mapping in Metaponto, Italy. In *Sensors* 21 (5), p. 1791. DOI: 10.3390/s21051791.

Fitzsimmons, K. E.; Marković, S. B.; Hambach, U. (2012): Pleistocene environmental dynamics recorded in the loess of the middle and lower Danube basin. In *Quaternary Science Reviews* 41, pp. 104–118. DOI: 10.1016/j.quascirev.2012.03.002.

Freeland, T.; Heung, B.; Burley, D. V.; Clark, G.; Knudby, A. (2016): Automated feature extraction for prospection and analysis of monumental earthworks from aerial LiDAR in the Kingdom of Tonga. In *0305-4403* 69, pp. 64–74. DOI: 10.1016/j.jas.2016.04.011.

Fu, W.; Ma, J.; Chen, P.; Chen, F. (2019): Remote Sensing Satellites for Digital Earth. In H. Guo, M. F. Goodchild, A. Annoni (Eds.): Manual of Digital Earth. Singapore: Springer Nature, pp. 55–123. DOI: 10.1007/978-981-32-9915-3_3.

Gallant, J.C.; Wilson, J.P. (2000): Primary topographic attributes. In Terrain Analysis: Principles and Applications, Wiley, New York (2000), pp. 51-85.

Galletti, C. S.; Ridder, E.; Falconer, S. E.; Fall, P. L. (2013): Maxent modeling of ancient and modern agricultural terraces in the Troodos foothills, Cyprus. In *Applied Geography* 39, pp. 46–56. DOI: 10.1016/j.apgeog.2012.11.020.

Galletti, C. S.; Ridder, E.; Falconer, S. E.; Fall, P. L. (2013): Maxent modeling of ancient and modern agricultural terraces in the Troodos foothills, Cyprus. In Applied Geography 39, pp. 46–56. DOI: 10.1016/j.apgeog.2012.11.020.

García-Álvarez, D.; Nanu, S. F. (2022): Land Use Cover Datasets: A Review. In David García-Álvarez (Ed.): Land Use Cover Datasets and Validation Tools. Validation Practices with QGIS. Cham: Springer, pp. 47–66. DOI: 10.1007/978-3-030-90998-7_4.

Geologische Bundesanstalt (2013): Geological map of Austria 1:50,000, Mapsheet 38: Krems, Federal Geological Institute of Austria, Vienna, 2015 [Dataset]. Available at: www.data.gv.at

Gerasimenko, N.; Rousseau, D.-D. (2008): Stratigraphy and Paleoenvironments of the Last Pleniglacial in the Kyiv Loess Region (Ukraine). In *Quaternaire. Revue de l'Association française pour l'étude du Quaternaire* (vol. 19/4), pp. 293–307. DOI: 10.4000/quaternaire.4592.

Geudtner, D.; Torres, R.; Snoeij, P.; Davidson, M.; Rommen, B. (2014): Sentinel-1 System capabilities and applications. In : 2014 IEEE Geoscience and Remote Sensing Symposium: IEEE.

Giardino, M. J. (2011): A history of NASA remote sensing contributions to archaeology. In *Journal of Archaeological Science* 38 (9), pp. 2003–2009. DOI: 10.1016/j.jas.2010.09.017.

Gibert, K.; Horsburgh, J. S.; Athanasiadis, I. N.; Holmes, G. (2018): Environmental Data Science. In *Environmental Modelling & Software* 106, pp. 4–12. DOI: 10.1016/j.envsoft.2018.04.005.

Gibert, K.; Sànchez-Marrè, M.; Codina, V. (2010): Choosing the right data mining technique: classification of methods and intelligent recommendation. In *International Congress on Environment Modelling and Software Modelling for Environment's Sake* Fifth Biennial Meeting, Ottawa, Canada.

Gillespie, T. W.; Smith, M. L.; Barron, S.; Kalra, K.; Rovzar, C. (2016): Predictive Modelling for Archaeological Sites:Ashokan Edicts from the Indian Subcontinent. In *Current Science* 110 (10), p. 1916. DOI: 10.18520/cs/v110/i10/1916-1921.

González-Tennant, E. (2016): Recent Directions and Future Developments in Geographic Information Systems for Historical Archaeology. In *Hist Arch* 50 (3), pp. 24–49. DOI: 10.1007/BF03377332.

Goodchild, M. F. (2003): Geographic Information Science and Systems for Environmental Management. In *Annu. Rev. Environ. Resour.* 28 (1), pp. 493–519. DOI: 10.1146/annurev.energy.28.050302.105521.

Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. (2017): Google Earth Engine: Planetary-scale geospatial analysis for everyone. In *0034-4257* 202, pp. 18–27. DOI: 10.1016/j.rse.2017.06.031.

Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. (2017): Google Earth Engine: Planetary-scale geospatial analysis for everyone. In *Remote Sensing of Environment* 202, pp. 18–27. DOI: 10.1016/j.rse.2017.06.031.

Goward, S. N.; Williams, D. L. (1997): Landsat and earth systems science: development of terrestrial monitoring. In *Photogrammetric engineering and remote sensing* 63 (7).

Grimwood, T. (2023): UCS Satellite Database [Dataset]. Available at: https://www.ucsusa.org/media/11491

Grunert, J.; Lehmkuhl, F. (2004): Aeolian sedimentation in arid and semi-arid environments of Western Mongolia. In *Paleoecology of Quaternary Drylands. Springer, Berlin.* 102, pp. 195–218. DOI: 10.1007/978-3-540-44930-0_11.

Grunert, J.; Lehmkuhl, F.; Walther, M. (2000): Paleoclimatic evolution of the Uvs Nuur basin and adjacent areas (Western Mongolia). In *Quaternary International* 65-66, pp. 171–192. DOI: 10.1016/S1040-6182(99)00043-9.

Haesaerts, P.; Damblon, F.; Bachner, M.; Trnka, G. (1996). Revised stratigraphy and chronology of the Willendorf II sequence, Lower Austria. In *Archaeologia Austriaca* 80, 25-42.

Hahn, J. (1995): Die Buttentalhöhle. Eine spät-jungpaläolithische Abristation im Oberen Donautal. 13-158 Seiten / Fundberichte aus Baden-Württemberg, Bd. 20 (1995): Fundberichte aus Baden-Württemberg. In *Fundberichte* 20, pp. 13–158. DOI: 10.11588/fbbw.1995.0.48412.

Haklay, M.; Weber, P. (2008): OpenStreetMap: User-Generated Street Maps. In *IEEE Pervasive Comput.* 7 (4), pp. 12–18. DOI: 10.1109/mprv.2008.80.

Hamdan, M. A.; Refaat, A. A.; Abdel Wahed, M. (2016): Morphologic characteristics and migration rate assessment of barchan dunes in the Southeastern Western Desert of Egypt. In *Geomorphology* 257, pp. 57–74. DOI: 10.1016/j.geomorph.2015.12.026.

Händel, M.; Simon, U.; Einwögerer, T.; Neugebauer-Maresch, C. (2009): Loess deposits and the conservation of the archaeological record—The Krems-Wachtberg example. In *Quaternary International* 198 (1-2), pp. 46–50. DOI: 10.1016/j.quaint.2008.07.005.

Händel, M.; Simon, U.; Maier, A.; Brandl, M.; Groza-Săcaciu, S.M.; Timar-Gabor, A.; Einwögerer, T. (2020): Kammern-Grubgraben revisited - First results from renewed investigations at a well-known LGM site in east Austria. In *Quaternary International*, https://doi.org/10.1016/j.quaint.2020.06.012

Händel, M.; Thomas, R.; Sprafke, T.; Schulte, P.; Brandl, M.; Simon, U.; Einwögerer, T. (2021): Using archaeological data and sediment parameters to review the formation of the Gravettian layers at Krems-Wachtberg. In *Journal of Quaternary Sciences*, DOI: 10.1002/jqs.3293

Hauck, T. C.; Lehmkuhl, F.; Zeeden, C.; Bösken, J.; Thiemann, A.; Richter, J. (2018): The Aurignacian way of life: Contextualizing early modern human adaptation in the Carpathian Basin. In *Quaternary International* 485, pp. 150–166. DOI: 10.1016/j.quaint.2017.10.020.

Havivi, S.; Amir, D.; Schvartzman, I.; August, Y.; Maman, S.; Rotman, S. R.; Blumberg, D. G. (2018): Mapping dune dynamics by InSAR coherence. In *Earth Surface Processes and Landforms* 43 (6), pp. 1229–1240. DOI: 10.1002/esp.4309.

Havivi, S.; Amir, D.; Schvartzman, I.; August, Y.; Maman, S.; Rotman, S. R.; Blumberg, D. G. (2018): Mapping dune dynamics by InSAR coherence. In *Earth Surf. Process. Landforms* 43 (6), pp. 1229–1240. DOI: 10.1002/esp.4309.

Hawker, L.; Uhe, P.; Paulo, L.; Sosa, J.; Savage, J.; Sampson, C.; Neal, J. (2022): A 30 m global map of elevation with forests and buildings removed. In *Environ. Res. Lett.* 17 (2), p. 24016. DOI: 10.1088/1748-9326/ac4d4f.

Herndon, K. E.; Griffin, R.; Schroder, W.; Murtha, T.; Golden, C.; Contreras, D. A. et al. (2023): Google Earth Engine for archaeologists: An updated look at the progress and promise of remotely sensed big data. In *2352-409X* 50, p. 104094. DOI: 10.1016/j.jasrep.2023.104094.

Herzog, M.; Henselowsky, F.; Bubenzer, O. (2021): Geomorphology of the Tafilalt Basin, South-East Morocco – implications for fluvial–aeolian dynamics and wind regimes. In *Journal of Maps* 17 (2), pp. 682–689. DOI: 10.1080/17445647.2021.1990805.

Holzkämper, J.; Maier, A.; Richter, J. (2013): „Dark Ages" illuminated–Rietberg and related assemblages possibly reducing the hiatus between the Upper and Late Palaeolithic in Westphalia: Licht im „Dunklen Zeitalter "–Rietberg und verwandte Inventare verkürzen möglicherweise den Hiatus zwischen Jung-und Spätpaläolithikum in Westfalen. In *Quartär–Internationales Jahrbuch zur Erforschung des Eiszeitalters und der Steinzeit* (60), pp. 115–136.

Howey, M. C. L.; Sullivan, F. B.; Burg, M. B.; Palace, M. W. (2020): Remotely Sensed Big Data and Iterative Approaches to Cultural Feature Detection and Past Landscape Process Analysis. In *Journal of Field Archaeology* 45 (sup1), S27-S38. DOI: 10.1080/00934690.2020.1713435.

Hublin, J.-J.; Sirakov, N.; Aldeias, V.; Bailey, S.; Bard, E.; Delvigne, V. et al. (2020): Initial Upper Palaeolithic Homo sapiens from Bacho Kiro Cave, Bulgaria. In *Nature* 581 (7808), pp. 299–302. DOI: 10.1038/s41586-020-2259-z.

Hugenholtz, C. H.; Barchyn, T. E. (2010): Spatial analysis of sand dunes with a new global topographic dataset: new approaches and opportunities. In *Earth Surface Processes and Landforms* 35 (8), pp. 986–992. DOI: 10.1002/esp.2013.

Hugenholtz, C. H.; Levin, N.; Barchyn, T. E.; Baddock, M. C. (2012): Remote sensing and spatial analysis of aeolian sand dunes: A review and outlook. In *Earth-Science Reviews* 111 (3-4), pp. 319–334. DOI: 10.1016/j.earscirev.2011.11.006.

Ikazaki, K. (2015): Desertification and a new countermeasure in the Sahel, West Africa. In *Soil Science and Plant Nutrition* 61 (3), pp. 372–383. DOI: 10.1080/00380768.2015.1025350.

Jones, P. J.; Williamson, G. J.; Bowman, D. M. J. S.; Lefroy, E. C. (2019): Mapping Tasmania's cultural landscapes: Using habitat suitability modelling of archaeological sites as a landscape history tool. In *J Biogeogr* 46 (11), pp. 2570–2582. DOI: 10.1111/jbi.13684.

Jöris, O.; Neugebeuer-Maresch, C.; Weninger, B.; Street, M. (2010). The radiocarbon chronology of the Aurignacian to Mid-Upper Paleolithic transition along the upper and middle Danube. In Neugebauer-Maresch, C. and Owen, L.R. (eds.): New aspects of the central and eastern European Upper Palaeolithic – methods, chronology, technology and subsistence. Mitteilungen der Prähistorischen Kommission 72, Vienna, 101-150.

Kamermans, H., & Niccolucci F, H. S. (2010): The application of predictive modelling in archaeology: problems and possibilities. In Beyond the artefact–Digital Interpretation of the Past-Proceedings of CAA2004-Prato 13-17 April 2004 (pp. 273-277). Archaeolingua.Kohler, Timothy A.; Parker, Sandra C. (1986): Predictive Models for Archaeological Resource Location. In Advances in Archaeological Method and Theory, pp. 397–452. DOI: 10.1016/B978-0-12-003109-2.50011-8.

Kansa, E. C.; Kansa, S. W.; Wells, J. J.; Yerka, S. J.; Myers, K. N.; DeMuth, R. C. et al. (2018): The Digital Index of North American Archaeology: networking government data to navigate an uncertain future for the past. In *Antiquity* 92 (362), pp. 490–506. DOI: 10.15184/aqy.2018.32.

Kayri, M., (2007): Two-Step Clustering Analysis in Researches: A Case Study. In *Eurasian Journal of Educational Research*, 28, pp, 89-99.

Klein Goldewijk, K.; Beusen, A.; Doelman, J.; Stehfest, E. (2017): Anthropogenic land use estimates for the Holocene – HYDE 3.2. In *Earth Syst. Sci. Data* 9 (2), pp. 927–953. DOI: 10.5194/essd-9-927-2017.

Klein, K.; Wegener, C.; Schmidt, I.; Rostami, M.; Ludwig, P.; Ulbrich, S. et al. (2021): Human existence potential in Europe during the Last Glacial Maximum. In *Quaternary International* 581-582, pp. 7–27. DOI: 10.1016/j.quaint.2020.07.046.

Klinge, M. (2001): Glazialgeomorphologische Untersuchungen im Mongolischen Altai als Beitrag zur jungquartären Landschafts-und Klimageschichte der Westmongolei (Aachener Geographische Arbeiten, 35).

Klinge, M.; Lehmkuhl, F.; Schulte, P.; Hülle, D.; Nottebaum, V. (2017): Implications of (reworked) aeolian sediments and paleosols for Holocene environmental change in Western Mongolia. In *Geomorphology* 292, pp. 59–71. DOI: 10.1016/j.geomorph.2017.04.027..

Klinge, M.; Sauer, D. (2019): Spatial pattern of Late Glacial and Holocene climatic and environmental development in Western Mongolia - A critical review and synthesis. In *Quaternary Science Reviews* 210, pp. 26–50. DOI: 10.1016/j.quascirev.2019.02.020.

Kohler, T. A.; Parker, S. C. (1986): Predictive Models for Archaeological Resource Location. In *Advances in Archaeological Method and Theory*, pp. 397–452. DOI: 10.1016/B978-0-12-003109-2.50011-8.

Kovarovic, K.; Aiello, L. C.; Cardini, A., Lockwood, C. A. (2011): Discriminant function analyses in archaeology: are classification rates too good to be true? Journal of Archaeological Science Volume 38, Issue 11, November 2011, Pages 3006-3018. https://doi.org/10.1016/j.jas.2011.06.028

Kretschmer, I. (2015): Demographische Untersuchungen zu Bevölkerungsdichten, Mobilität und Landnutzungsmustern im späten Jungpaläolithikum. In *Kölner Studien zur Prähistorischen Archäologie* (6).

Laney, D. (2001): 3D data management: Controlling data volume, velocity and variety. In *META group research note* (6).

Lansley, G.; Smith, M. de; Goodchild, M.; Longley, P. (2017): Big Data and Geospatial Analysis. In Igor Ivan, Alex Singleton, Jiří Horák, Tomáš Inspektor (Eds.): The Rise of Big Spatial Data. Cham: Springer International Publishing; Imprint; Springer (Lecture Notes in Geoinformation and Cartography).

Lasaponara, R.; Abate, N.; Masini, N. (2022): On the Use of Google Earth Engine and Sentinel Data to Detect “Lost” Sections of Ancient Roads. The Case of Via Appia. In *IEEE Geosci. Remote Sensing Lett.* 19, pp. 1–5. DOI: 10.1109/lgrs.2021.3054168.

Lee, J.-S.; Wen, J.-H.; Ainsworth, T. L.; Chen, K.-S.; Chen, A. J. (2009): Improved Sigma Filter for Speckle Filtering of SAR Imagery. In *IEEE Transactions on Geoscience and Remote Sensing* 47 (1), pp. 202–213. DOI: 10.1109/tgrs.2008.2002881.

Lehmkuhl, F. (1999): Rezente und jungpleistozane Formungs- und Prozeβregionen im Turgen-Kharkhiraa, Mongolischer Altai. In *Die Erde* 130, pp. 151–172.

Lehmkuhl, F.; Grunert, J.; Hülle, D.; Batkhishig, O.; Stauch, G. (2018): Paleolakes in the Gobi region of southern Mongolia. In *Quaternary Science Reviews* 179, pp. 1–23. DOI: 10.1016/j.quascirev.2017.10.035.

Lehmkuhl, F.; Klinge, M. (2000): Measurements of soil temperatures in the northern Mongolian Altai as indicators for periglacial geomorphodynamic in mountain areas. In *Zeitschrift für Geomorphologie* 44 (1), pp. 75–102. DOI: 10.1127/zfg/44/2000/75.

Lehmkuhl, F.; Nett, J. J.; Pötter, S.; Schulte, P.; Sprafke, T.; Jary, Z., Antoine, P.; Wacha, L.; Wolf, D.; Zerboni, A.; Hosek, J.; Markovic, S. B.; Obreht, I.; Sügemi, P.; Veres, D.; Zeeden, C.; Boemke, B.; Schaubert, V.; Vieweger, J.; Hambach, U. (2021): Loess landscapes of Europe – Mapping, geomorphology, and zonal differentiation. In *Earth-Science Reviews* 215, p. 103496. DOI: 10.1016/j.earscirev.2020.103496.

Leisz, S. J. (2013): An Overview of the Application of Remote Sensing to Archaeology During the Twentieth Century. In *Mapping Archaeological Landscapes from Space* 5, pp. 11–19. DOI: 10.1007/978-1-4614-6074-9_2.

Li, S.; Dragicevic, S.; Castro, F. A.; Sester, M.; Winter, S.; Coltekin, A. et al. (2016): Geospatial big data handling theory and methods: A review and research challenges. In *ISPRS Journal of Photogrammetry and Remote Sensing* 115, pp. 119–133. DOI: 10.1016/j.isprsjprs.2015.10.012.

Liss, B.; Howland, M. D.; Levy, T. E. (2017): Testing Google Earth Engine for the automatic identification and vectorization of archaeological features: A case study from Faynan, Jordan. In *2352-409X* 15, pp. 299–304. DOI: 10.1016/j.jasrep.2017.08.013.

Luo, W.; Shao, M.; Che, X.; Hesp, P. A.; Bryant, R. G.; Yan, C.; Xing, Z. (2020): Optimization of UAVs-SfM data collection in aeolian landform morphodynamics: a case study from the Gonghe Basin, China. In *Earth Surface Processes and Landforms* 45 (13), pp. 3293–3312. DOI: 10.1002/esp.4965.

Ma, Y.; Wu, H.; Wang, L.; Huang, B.; Ranjan, R.; Zomaya, A.; Jie, W. (2015): Remote sensing big data computing: Challenges and opportunities. In *Future Generation Computer Systems* 51, pp. 47–60. DOI: 10.1016/j.future.2014.10.029.

Mahmoud, A. M. A.; Novellino, A.; Hussain, E.; Marsh, S.; Psimoulis, P.; Smith, M. (2020): The Use of SAR Offset Tracking for Detecting Sand Dune Movement in Sudan. In *Remote Sensing* 12 (20), p. 3410. DOI: 10.3390/rs12203410.

Maier, A. (2015): The Central European Magdalenian. Regional diversity and internal variability. Dordrecht/New York: Springer.

Maier, A., Stojakowits, P., Mayr, C., Pfeifer, S., Preusser, F., Zolitschka, B., Anghelinu, M., Bobak, D., Duprat-Oualid, F., Einwögerer, T., Hambach, U., Händel, M., Kaminská, L., Kämpf, L., Łanczont, M., Lehmkuhl, F., Ludwig, P., Magyari, E., Mroczek, P., Nemergut, A., Nerudová, Z., Niţă, L., Polanská, M., Połtowicz-Bobak, M., Rius, D., Römer, W., Simon, U., Škrdla, P., Úivári, G., Veres, D. (2021). Cultural evolution and environmental change in Central Europe between 40.000 and 15.000 years ago. Quaternary International 581–582, 225–240. https://doi.org/10.1016/j.quaint.2020.09.049

Maier, A.; Lehmkuhl, F.; Ludwig, P.; Melles, M.; Schmidt, I.; Shao, Y. et al. (2016): Demographic estimates of hunter–gatherers during the Last Glacial Maximum in Europe against the background of palaeoenvironmental data. In *Quaternary International* 425, pp. 49–61. DOI: 10.1016/j.quaint.2016.04.009.

Maier, A.; Liebermann, C.; Pfeifer, S. J. (2020): Beyond the Alps and Tatra Mountains—the 20–14 ka Repopulation of the Northern Mid-latitudes as Inferred from Palimpsests Deciphered with Keys from Western and Central Europe. In *J Paleo Arch* 3 (3), pp. 398–452. DOI: 10.1007/s41982-019-00045-1.

Maier, A.; Ludwig, P.; Zimmermann, A.; Schmidt, I. (2022): The Sunny Side of the Ice Age: Solar Insolation as a Potential Long-Term Pacemaker for Demographic Developments in Europe Between 43 and 15 ka Ago. 35-51 Pages / PaleoAnthropology, 2022: PaleoAnthropology. In *pa*, pp. 35–51. DOI: 10.48738/2022.iss1.100.

Maier, A.; Zimmermann, A. (2015): CRC806-E1 LGM-Sites Database V-20150313. CRC806-Database. DOI: 10.5880/SFB806.12.

Maier, A.; Zimmermann, A. (2016): CRC806-E1 Gravettian-Sites Database V-20160219. CRC806-Database. DOI: 10.5880/SFB806.18.

Maier, A.; Zimmermann, A. (2017): Populations headed south? The Gravettian from a palaeodemographic point of view. In *Antiquity* 91 (357), pp. 573–588. DOI: 10.15184/aqy.2017.37.

Malaperdas, G.; Zacharias, N. (2019): The habitation Model Trend Calculation (MTC): A new effective tool for predictive modeling in archaeology. In *Geo-spatial Information Science* 22 (4), pp. 314–331. DOI: 10.1080/10095020.2019.1634320.

Mallinis, G.; Mitsopoulos, I.; Beltran, E.; Goldammer, J. (2016): Assessing Wildfire Risk in Cultural Heritage Properties Using High Spatial and Temporal Resolution Satellite Imagery and Spatially Explicit Fire Simulations: The Case of Holy Mount Athos, Greece. In *Forests* 7 (2), p. 46. DOI: 10.3390/f7020046.

Manzoni, M.; Molinari, M. E.; Monti-Guarnieri, A. (2021): Multitemporal InSAR Coherence Analysis and Methods for Sand Mitigation. In *Remote Sensing* 13 (7), p. 1362. DOI: 10.3390/rs13071362.

Markham, B. L.; Storey, J. C.; Williams, D. L.; Irons, J. R. (2004): Landsat sensor performance: history and current status. In *IEEE Trans. Geosci. Remote Sensing* 42 (12), pp. 2691–2694. DOI: 10.1109/tgrs.2004.840720.

Marr, Bernard (2015): Big data. Using SMART big data, analytics and metrics to make better decisions and improve performance. Hoboken: John Wiley & Sons.

Massei, N.; Dieppois, B.; Hannah, D. M.; Lavers, D. A.; Fossa, M.; Laignel, B.; Debret, M. (2017): Multi-time-scale hydroclimate dynamics of a regional watershed and links to large-scale atmospheric circulation: Application to the Seine river catchment, France. In *Journal of Hydrology* 546, pp. 262–275. DOI: 10.1016/j.jhydrol.2017.01.008.

McCoy, M. D. (2017): Geospatial Big Data and archaeology: Prospects and problems too great to ignore. In *Journal of Archaeological Science* 84, pp. 74–94. DOI: 10.1016/j.jas.2017.06.003.

McCoy, M. D.; Ladefoged, T. N. (2009): New Developments in the Use of Spatial Technology in Archaeology. In *J Archaeol Res* 17 (3), pp. 263–295. DOI: 10.1007/s10814-009-9030-1.

McKee, Edwin Dinwiddie (1979): A Study of Global Sand Seas: U.S. Government Printing Office.

McManamon, F. P.; Kintigh, K. W.; Ellison, L. A.; Brin, A. (2017): tDAR: a cultural heritage archive for twenty-first-century public outreach, research, and resource management. In *Advances in Archaeological Practice* 5 (3), pp. 238–249. DOI: 10.1017/aap.2017.18.

Meghini, C.; Scopigno, R.; Richards, J.; Wright, H.; Geser, G.; Cuy, S. et al. (2017): ARIADNE: A research infrastructure for archaeology. In *J. Comput. Cult. Herit.* 10 (3), pp. 1–27. DOI: 10.1145/3064527.

Merow, C.; Smith, M. J.; Silander, J. A. (2013): A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. In *Ecography* 36 (10), pp. 1058–1069. DOI: 10.1111/j.1600-0587.2013.07872.x.

Miller, H. J. (2007): Geographic Data Mining and Knowledge Discovery. In John P. Wilson, A. Stewart Fotheringham (Eds.): The handbook of geographic information science. 1st ed. Malden: Blackwell.

Minghini, M.; Cetl, V.; Kotsev, A.; Tomas, R.; Lutz, M. (2021): INSPIRE: The Entry Point to Europe's Big Geospatial Data Infrastructure. In M. Werner, Y. Chaing (Eds.): Handbook of Big Geospatial Data: Springer, Cham, pp. 619–641. DOI: 10.1007/978-3-030-55462-0_24.

Mobasheri, A.; Mitasova, H.; Neteler, M.; Singleton, A.; Ledoux, H.; Brovelli, M. A. (2020): Highlighting recent trends in open source geospatial science and software. In *Transactions in GIS* 24 (5), pp. 1141–1146. DOI: 10.1111/tgis.12703.

Monna, F., Magailb, J. , Rollanda, T. , Navarro, N., Wilczeka, J., Gantulgag, J.O. , Esin, Y. , Granjoni L. , Allard, A-C. , Chateau-Smith, C. (2020): Machine learning for rapid mapping of archaeological structures made of dry stones - Example of burial monuments from the Khirgisuur culture, Mongolia - Journal of Cultural Heritage Volume 43, May–June 2020, Pages 118-128. DOI: 10.1016/j.culher.2020.01.002

Moreno, M.; Bertolín, C.; Ortiz, P.; Ortiz, R. (2022): Satellite product to map drought and extreme precipitation trend in Andalusia, Spain: A novel method to assess heritage landscapes at risk. In *International Journal of Applied Earth Observation and Geoinformation* 110, p. 102810. DOI: 10.1016/j.jag.2022.102810.

Mullissa, A.; Vollrath, A.; Odongo-Braun, C.; Slagter, B.; Balling, J.; Gou, Y. et al. (2021): Sentinel-1 SAR Backscatter Analysis Ready Data Preparation in Google Earth Engine. In *Remote Sensing* 13 (10), p. 1954. DOI: 10.3390/rs13101954.

NASA JPL (2020): NASADEM Merged DEM Global 1 arc second V001 [Dataset]. NASA EOSDIS Land Processes DAAC. DOI: 10.5067/MEaSUREs/NASADEM/NASADEM_HGT.001.

Nashashibi, A. Y.; Sarabandi, K.; Al-Zaid, F. A.; Alhumaidi, S. (2012): Characterization of Radar Backscatter Response of Sand-Covered Surfaces at Millimeter-Wave Frequencies. In *IEEE Transactions on Geoscience and Remote Sensing* 50 (6), pp. 2345–2354. DOI: 10.1109/tgrs.2011.2172619.

Nett, J. J.; Chu, W.; Fischer, P.; Hambach, U.; Klasen, N.; Zeeden, C. et al. (2021): The Early Upper Paleolithic Site Crvenka-At, Serbia–The First Aurignacian Lowland Occupation Site in the Southern Carpathian Basin. In *Front. Earth Sci.* 9, Article 599986. DOI: 10.3389/feart.2021.599986.

Neugebauer-Maresch, C. (2008). Krems-Hundssteig-Mammutjägerlager der Eiszeit. Ein Nutzungsareal paläolithischer Jäger- und Sammler(innen) vor 41.000-27.000 Jahren. Mitteilungen der Prähistorischen Kommission, Vienna, p. 347.

Nigst, P.R.; Haesaerts, P.; Damblon, F.; Frank-Fellner, C.; Mallol, C.; Viola, B.; Götzinger, M.; Niven, L.; Trnka, G.; Hublin, J.-J. (2014). Europeans occupied a cold steppe 43,500 years ago. In *Proceedings of the National Academy of Sciences* 111(40), 14394–9.

Noronha Vaz, E. de; Cabral, P.; Caetano, M.; Nijkamp, P.; Painho, M. (2012): Urban heritage endangerment at the interface of future cities and past heritage: A spatial vulnerability assessment. In *Habitat International* 36 (2), pp. 287–294. DOI: 10.1016/j.habitatint.2011.10.007.

Noviello, M.; Cafarelli, B.; Calculli, C.; Sarris, A.; Mairota, P. (2018): Investigating the distribution of archaeological sites: Multiparametric vs probability models and potentials for remote sensing data. In *Applied Geography* 95, pp. 34–44. DOI: 10.1016/j.apgeog.2018.04.005.

Opitz, R.; Herrmann, J. (2018): Recent Trends and Long-standing Problems in Archaeological Remote Sensing. In *Journal of Computer Applications in Archaeology* 1 (1), pp. 19–41. DOI: 10.5334/jcaa.11.

Orengo, H. A.; Garcia-Molsosa, A. (2019): A brave new world for archaeological survey: Automated machine learning-based potsherd detection using high-resolution drone imagery. In *Journal of Archaeological Science* 112, p. 105013. DOI: 10.1016/j.jas.2019.105013.

Orengo, H.; Petrie, C. (2017): Large-Scale, Multi-Temporal Remote Sensing of Palaeo-River Networks: A Case Study from Northwest India and its Implications for the Indus Civilisation. In *Remote Sensing* 9 (7), p. 735. DOI: 10.3390/rs9070735.

Orgiazzi, A.; Ballabio, C.; Panagos, P.; Jones, A.; Fernández-Ugalde, O. (2018): LUCAS Soil, the largest expandable soil dataset for Europe: a review. In *Eur J Soil Sci* 69 (1), pp. 140–153. DOI: 10.1111/ejss.12499.

Pei, T.; Song, C.; Guo, S.; Shu, H.; Liu, Y.; Du, Y. et al. (2020): Big geodata mining: Objective, connotations and research issues. In *J. Geogr. Sci.* 30 (2), pp. 251–266. DOI: 10.1007/s11442-020-1726-7.

Peresani, M., Monegato, G., Ravazzi, C., Bertola, S., Margaritora, D., Breda, M., Fontana, A., Fontana, F., Janković, I., Karavanić, I., Komšo, D., Mozzi, P., Pini, R., Furlanetto, G., Maria De Amicis, M.G., Perhoč, Z., Posth, C., Ronchi, L., Rossato, S., Vukosavljević, N., Zerboni, A. (2021). Hunter-gatherers across the great Adriatic-Po region during the Last Glacial Maximum: Environmental and cultural dynamics. In *Quaternary International*, 581-582,128-163. https://doi.org/10.1016/j.quaint.2020.10.007

Phillips, S. J.; Anderson, R. P.; Dudík, M.; Schapire, R. E.; Blair, M. E. (2017): Opening the black box: an open-source release of Maxent. In *Ecography* 40 (7), pp. 887–893. DOI: 10.1111/ecog.03049.

Phillips, S. J.; Dudík, M. (2008): Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. In *Ecography* 31 (2), pp. 161–175. DOI: 10.1111/j.0906-7590.2008.5203.x.

Poortinga, A.; Keijsers, J.; Visser, S. M.; Riksen, M.; Baas, A. (2015): Temporal and spatial variability in event scale aeolian transport on Ameland, The Netherlands. In *GeoResJ* 5, pp. 23–35. DOI: 10.1016/j.grj.2014.11.003.

Pradhan, B.; Moneir, A. A. A.; Jena, R. (2018): Sand dune risk assessment in Sabha region, Libya using Landsat 8, MODIS, and Google Earth Engine images. In *Geomatics, Natural Hazards and Risk* 9 (1), pp. 1280–1305. DOI: 10.1080/19475705.2018.1518880.

Qong, M. (2000): Sand Dune Attributes Estimated from SAR Images. In *Remote Sensing of Environment* 74 (2), pp. 217–228. DOI: 10.1016/S0034-4257(00)00112-7.

Ramsey, J. B.; Lampart, C. (1998): Decomposition of Economic Relationships by Timescale Using Wavelets. In *Macroecon. Dynam.* 2 (1), pp. 49–71. DOI: 10.1017/s1365100598006038.

Rast, M.; Painter, T. H. (2019): Earth Observation Imaging Spectroscopy for Terrestrial Systems: An Overview of Its History, Techniques, and Applications of Its Missions. In *Surv Geophys* 40 (3), pp. 303–331. DOI: 10.1007/s10712-019-09517-z.

Rayne, L.; Gatto, M.; Abdulaati, L.; Al-Haddad, M.; Sterry, M.; Sheldrick, N.; Mattingly, D. (2020): Detecting Change at Archaeological Sites in North Africa Using Open-Source Satellite Imagery. In *Remote Sensing* 12 (22), p. 3694. DOI: 10.3390/rs12223694.

Reed, M.; Stringer, L. C. (2015): Climate change and desertification: Anticipating, assessing & adapting to future change in drylands. Montpellier: Agropolis International.

Reeder-Myers, L. A. (2015): Cultural Heritage at Risk in the Twenty-First Century: A Vulnerability Assessment of Coastal Archaeological Sites in the United States. In *The Journal of Island and Coastal Archaeology* 10 (3), pp. 436–445. DOI: 10.1080/15564894.2015.1008074.

Reimer, P.J., Austin, W.E.N., Bard, E., Bayliss, A., Blackwell, P.G., Bronk Ramsey, C., Butzin, M., Cheng, H., Edwards, R.L., Friedrich, M., Grootes, P.M., Guilderson, T.P., Hajdas, I., Heaton, T.J., Hogg, A.G., Hughen, K.A., Kromer, B., Manning, S.W., Muscheler, R., Palmer, J.G., Pearson, C., van der Plicht, J., Reimer, R.W., Richards, D.A., Scott, E.M., Southon, J.R., Turney, C.S.M., Wacker, L., Adolphi, F., Büntgen, U., Capano, M., Fahrni, S.M., Fogtmann-Schulz, A., Friedrich, R., Köhler, P., Kudsk, S.,

Miyake, F., Olsen, J., Reinig, F., Sakamoto, M., Sookdeo, A., Talamo, S. (2020). The IntCal20 northern hemisphere radiocarbon age calibration curve (0-55 cal kBP). In *Radiocarbon,* 1-33. DOI: 10.1017/RDC.2020.41.

Reu, Jeroen de; Bourgeois, Jean; Bats, Machteld; Zwertvaegher, Ann; Gelorini, Vanessa; Smedt, Philippe de et al. (2013): Application of the topographic position index to heterogeneous landscapes. In Geomorphology 186, pp. 39–49. DOI: 10.1016/j.geomorph.2012.12.015.

Rowland, M. J. (1989): Population increase, intensification or a result of preservation?: Explaining site distribution patterns on the coast of Queensland. In *Australian Aboriginal Studies* (2), pp. 32–42.

Roy, D. P.; Huang, H.; Houborg, R.; Martins, V. S. (2021): A global analysis of the temporal availability of PlanetScope high spatial resolution multi-spectral imagery. In *0034-4257* 264, p. 112586. DOI: 10.1016/j.rse.2021.112586.

Rundle-Thiele, S.; Kuback, K.; Tkaczynski, A.; Parkinson, J., (2015): Using two-stepcluster analysis to identify homogeneous physical activity group. In *Marketing Intelligence & Planning* 33(4):522-537. DOI:10.1108/MIP-03-2014-0050

Saad, S. A. M.; Seedahmed, A. M. A.; Ahmed, A.; Ossman, S. A. M.; Eldoma, A. M. A. (2018): Combating Desertification in Sudan: Experiences and Lessons Learned. In *Outlook* 10. Public private partnerships for the implementation of the 2030 Agenda for Sustainable Development.

Salazar, L. G. F.; Romão, X.; Paupério, E. (2021): Review of vulnerability indicators for fire risk assessment in cultural heritage. In *International Journal of Disaster Risk Reduction* 60, p. 102286. DOI: 10.1016/j.ijdrr.2021.102286.

Schiffer M.B.( 1983): Toward the identification of site formation processes. In *American Antiquity* 48, 675-706.Sprafke, T.; Schulte, P.; Meyer-Heintze, S.; Händel, M.; Einwögerer, T.; Simon, U.; Peticzka, R.; Schäfer, C.; Lehmkuhl, F.; Terhorst, B. (2020). Paleoenvironments from robust loess stratigraphy using high-resolution color and grain-size data of the last glacial Krems-Wachtberg record (NE Austria). In *Quaternary Science Reviews* 248, doi.org/10.1016/j.quascirev.2020.106602.

Schmidt, I. (2021): CRC806_E1_AUR_Sites_Database_20210331. CRC806-Database. DOI: 10.5880/SFB806.63.

Schmidt, I., Gehlen, B., Zimmermann, A., (2021a): Population estimates for the Final Palaeolithic (14,000 to 11,600 years cal. BP) of Europe – challenging evidence and methodological limitations. In: Société préhistorique francaise, p. 221-237. ISBN: 2-913745-86-5.

Schmidt, I.; Hilpert, J.; Kretschmer, I.; Peters, R.; Broich, M.; Schiesberg, S. et al. (2021b): Approaching prehistoric demography: proxies, scales and scope of the Cologne Protocol in European contexts. In *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 376 (1816), p. 20190714. DOI: 10.1098/rstb.2019.0714.

Schmidt, I.; Zimmermann, A. (2019): Population dynamics and socio-spatial organization of the Aurignacian: Scalable quantitative demographic data for western and central Europe. In *PLOS ONE* 14 (2), e0211562. DOI: 10.1371/journal.pone.0211562.

Schwarz, G. (1978): Estimating the Dimension of a Model." Ann. Statist. 6 (2) 461 – 464. https://doi.org/10.1214/aos/1176344136

Scianna, A.; Villa, B. (2011): GIS applications in archaeology. In *Archeologia e Calcolatori* 22, pp. 337–363.

Sebe, Krisztina; Roetzel, Reinhard; Fiebig, Markus; Lüthgens, Christopher (2015): Pleistocene wind system in eastern Austria and its impact on landscape evolution. In CATENA 134, pp. 59–74. DOI: 10.1016/j.catena.2015.02.004.

Senthil Kumar, S.; Arivazhagan, S.; Rangarajan, N. (2013): Remote Sensing and GIS Applications in Environmental Sciences - A Review. In *J. Environ. Nanotechnol.* 2 (2), pp. 92–101. DOI: 10.13074/jent.2013.06.132025.

Shao, Y.; Hense, A.; Klein, K.; Ludwig, P.; Maier, A.; Richter, J. et al. (2021a): Modelling Human Dispersal in Space and Time. In T. Litt, J. Richter, F. Schäbitz (Eds.): The Journey of Modern Humans from Africa to Europe. Culture-Environmental Interaction and Mobility: Schweizerbart'sche Verlagsbuchhandlung. ISBN: 978-3-510-65534-2.

Shao, Y.; Limberg, H.; Klein, K.; Wegener, C.; Schmidt, I.; Weniger, G.-C. et al. (2021b): Human-existence probability of the Aurignacian techno-complex under extreme climate conditions. In *Quaternary Science Reviews* 263, p. 106995. DOI: 10.1016/j.quascirev.2021.106995.

Shao, Y.; Limberg, H.; Klein, K.; Wegener, C.; Schmidt, I.; Weniger, G.-C. et al. (2021c): Human-existence probability of the Aurignacian techno-complex under extreme climate conditions. In *Quaternary Science Reviews* 263, p. 106995. DOI: 10.1016/j.quascirev.2021.106995.

Shukla, P. R.; Skea, J.; Calvo Buendia, E.; Masson-Delmotte, V.; Pörtner, H.-O.; Roberts, D. C. et al. (2019): IPCC, 2019: Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems. With assistance of Engineering & Physical Science Research Council (EPSRC). Intergovernmental Panel on Climate Change (IPCC).

Shumack, S.; Hesse, P.; Farebrother, W. (2020): Deep learning for dune pattern mapping with the AW3D30 global surface model. In *Earth Surface Processes and Landforms* 45 (11), pp. 2417–2431. DOI: 10.1002/esp.4888.

Smalley, Ian; Marković, Slobodan B.; Svirčev, Zorica (2011): Loess is [almost totally formed by] the accumulation of dust. In Quaternary International 240 (1-2), pp. 4–11. DOI: 10.1016/j.quaint.2010.07.011.

Soille, P.; Burger, A.; Marchi, D. de; Kempeneers, P.; Rodriguez, D.; Syrris, V.; Vasilev, V. (2018): A versatile data-intensive computing platform for information retrieval from big geospatial data. In *0167-739X* 81, pp. 30–40. DOI: 10.1016/j.future.2017.11.007.

Solazzo, D.; Sankey, J. B.; Sankey, T. T.; Munson, S. M. (2018): Mapping and measuring aeolian sand dunes with photogrammetry and LiDAR from unmanned aerial vehicles (UAV) and multispectral satellite imagery on the Paria Plateau, AZ, USA. In *Geomorphology* 319, pp. 174–185. DOI: 10.1016/j.geomorph.2018.07.023.

Soroush, M.; Mehrtash, A.; Khazraee, E.; Ur, J. A. (2020): Deep Learning in Archaeological Remote Sensing: Automated Qanat Detection in the Kurdistan Region of Iraq. In *Remote Sensing* 12 (3), p. 500. DOI: 10.3390/rs12030500.

Sprafke, Tobias; Schulte, Philipp; Meyer-Heintze, Simon; Händel, Marc; Einwögerer, Thomas; Simon, Ulrich et al. (2020): Paleoenvironments from robust loess stratigraphy using high-resolution color and grain-size data of the last glacial Krems-Wachtberg record (NE Austria). In Quaternary Science Reviews 248, p. 106602. DOI: 10.1016/j.quascirev.2020.106602.

Sprafke, Tobias; Thiel, Christine; Terhorst, Birgit (2014): From micromorphology to palaeoenvironment: The MIS 10 to MIS 5 record in Paudorf (Lower Austria). In CATENA 117, pp. 60–72. DOI: 10.1016/j.catena.2013.06.024.

Stephens, L.; Fuller, D.; Boivin, N.; Rick, T.; Gauthier, N.; Kay, A. et al. (2019): Archaeological assessment reveals Earth's early transformation through land use. In *Science (New York, N.Y.)* 365 (6456), pp. 897–902. DOI: 10.1126/science.aax1192.

Straus, L. G. (1995): The upper paleolithic of Europe: An overview. In *Evol. Anthropol.* 4 (1), pp. 4–16. DOI: 10.1002/evan.1360040103.

Štular, B.; Lozić, E.; Eichert, S. (2021): Airborne LiDAR-Derived Digital Elevation Model for Archaeology. In *Remote Sensing* 13 (9), p. 1855. DOI: 10.3390/rs13091855.

Surovell, T. A.; Brantingham, P. J. (2007): A note on the use of temporal frequency distributions in studies of prehistoric demography. In *Journal of Archaeological Science* 34 (11), pp. 1868–1877. DOI: 10.1016/j.jas.2007.01.003.

Surovell, T. A.; Byrd Finley, J.; Smith, G. M.; Brantingham, P. J.; Kelly, R. (2009): Correcting temporal frequency distributions for taphonomic bias. In *Journal of Archaeological Science* 36 (8), pp. 1715–1724. DOI: 10.1016/j.jas.2009.03.029.

Surovell, T. A.; Toohey, J. L.; Myers, A. D.; LaBelle, J. M.; Ahern, J. C. M.; Reisig, B. (2017): THE END OF ARCHAEOLOGICAL DISCOVERY. In *Am. Antiq.* 82 (2), pp. 288–300. DOI: 10.1017/aaq.2016.33.

Tamiminia, H.; Salehi, B.; Mahdianpari, M.; Quackenbush, L.; Adeli, S.; Brisco, B. (2020): Google Earth Engine for geo-big data applications: A meta-analysis and systematic review. In *ISPRS Journal of Photogrammetry and Remote Sensing* 164, pp. 152–170. DOI: 10.1016/j.isprsjprs.2020.04.001.

Terhorst, Birgit; Thiel, Christine; Peticzka, Robert; Sprafke, Tobias; Frechen, Manfred; Fladerer, Florian A. et al. (2011): Casting new light on the chronology of the loess/paleosol sequences in Lower Austria. In E&G Quaternary Sci. J. 60 (2/3), pp. 270–277. DOI: 10.3285/eg.60.2-3.04.

Teschler-Nicola, M.; Fernandes, D.; Händel, M.; Einwögerer, T.; Simon, U.; Neugebauer-Maresch, C.; Tangl, S.; Heimel, P.; Dobsak, T.; Retzmann, A.; Prohaska, T.; Irrgeher, J.; Kennett, D. J.; Olalde, I.; Reich, D.; & Pinhasi, R. (2020): Ancient DNA reveals monozygotic newborn twins from the Upper Palaeolithic. In *Nature Communications Biology*. DOI: 10.1038/s42003-020-01372-8.

Thomas, D. S. G. (2011): Arid zone geomorphology. Process, form and change in drylands. 3. ed., 1. impr. Chicester: Wiley-Blackwell.

Torres, R.; Snoeij, P.; Geudtner, D.; Bibby, D.; Davidson, M.; Attema, E. et al. (2012): GMES Sentinel-1 mission. In *Remote Sensing of Environment* 120, pp. 9–24. DOI: 10.1016/j.rse.2011.05.028.

Truckenbrodt, J.; Freemantle, T.; Williams, C.; Jones, T.; Small, D.; Dubois, C. et al. (2019): Towards Sentinel-1 SAR Analysis-Ready Data: A Best Practices Assessment on Preparing Backscatter Data for the Cube. In *Data* 4 (3), p. 93. DOI: 10.3390/data4030093.

Turki, I.; Laignel, B.; Chevalier, L.; Costa, S.; Massei, N. (2015): On the Investigation of the Sea-Level Variability in Coastal Zones Using SWOT Satellite Mission: Example of the Eastern English Channel (Western France). In *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 8 (4), pp. 1564–1569. DOI: 10.1109/JSTARS.2015.2419693.

Turki, I.; Le Bot, S.; Lecoq, N.; Shafiei, H.; Michel, C.; Deloffre, J. et al. (2021): Morphodynamics of intertidal dune field in a mixed wave-tide environment: Case of Baie de Somme in Eastern English Channel. In *Marine Geology* 431, p. 106381. DOI: 10.1016/j.margeo.2020.106381.

ur Rehman, M. H.; Liew, C. S.; Abbas, A.; Jayaraman, P. P.; Wah, T. Y.; Khan, S. U. (2016): Big Data Reduction Methods: A Survey. In *Data Sci. Eng.* 1 (4), pp. 265–284. DOI: 10.1007/s41019-016-0022-0.

Van Leusen, M., Deeben, J., Hallewas, D., Kamermans, H., Verhagen, P., & Zoetbrood, P. (2005). A baseline for predictive modelling in the Netherlands. In M. VanLeusen & H. Kamermans (Eds.), Predicitive modelling for archaeological heritage management: A research agenda (pp. 25–29). Nederlandse Archeologische Rapporten (NAR), Vol. 29.

Vancauwenberghe, G.; van Loenen, B. (2018): Exploring the Emergence of Open Spatial Data Infrastructures: Analysis of Recent Developments and Trends in Europe. In Saqib Saeed, T. Ramayah, Zaigham Mahmood (Eds.): User Centric E-Government: Springer, Cham, pp. 23–45. DOI: 10.1007/978-3-319-59442-2_2.

Verhagen, P. (2018): Predictive Modeling. In S. L. Varela, J. Thomas (Eds.): The Encyclopedia of Archaeological Sciences: John Wiley & Sons, Ltd, pp. 1–3.

Verhagen, P.; Whitley, T. G. (2012): Integrating Archaeological Theory and Predictive Modeling: a Live Report from the Scene. In *J Archaeol Method Theory* 19 (1), pp. 49–100. DOI: 10.1007/s10816-011-9102-7.

Verhagen, P. (2007): Case studies in archaeological predictive modelling (Vol. 14). Amsterdam University Press.

Verhagen, P. (2008): Testing archaeological predictive models: a rough guide. In: Posluschny A, Lambers K, Herzog I (eds) Layers of perception. In *Proceedings of the 35th international conference on Computer Applications and Quantitative Methods in Archaeology (CAA)*, pp 285–291, Berlin, Germany, April 2–6, 2007.

Verhagen, P.& Whitley, T. G. (2012): Integrating archaeological theory and predictive modeling: a live report from the scene. Journal of Archaeological Method and Theory, 19(1), 49-100.

Verhagen, P.; Kamermans, H.; van Leusen, M.; Ducke, B. (2010): New developments in archaeological predictive modelling. Edited by Tom Bloemers, Henk Kars, Arnold van der Valk: Amsterdam University Press (The Cultural Landscape & Heritage Paradox. Protection and Development of the Dutch Archaeological-Historical Landscape and its European Dimension).

Vermeersch, P. M. (2020): Radiocarbon Palaeolithic Europe database: A regularly updated dataset of the radiometric data regarding the Palaeolithic of Europe, Siberia included. In *2352-3409* 31, p. 105793. DOI: 10.1016/j.dib.2020.105793.

Verstraeten, G.; Mohamed, I.; Willems, H.; Laet, V. de; Delgado Blasco, J. M. (2014): Analysis of Aeolian-Fluvial-Human Interactions in the Nile Valley (Central Egypt) by Combining Field-Based Geomorphology with Remote Sensing. In : *34th EARSeL Symposium. 5th European Remote Sensing – New Opportunities for Science and Practice,* p. 55.. Warsaw, 16-20 June 2014.

Wachtel, I.; Zidon, R.; Garti, S.; Shelach-Lavi, G. (2018): Predictive modeling for archaeological site locations: Comparing logistic regression and maximal entropy in north Israel and north-east China. In *Journal of Archaeological Science* 92, pp. 28–36. DOI: 10.1016/j.jas.2018.02.001.

Walther, M.; Naumann, S. (1997): Beobachtungen zur Fußflächenbildung im ariden bis semiariden Bereich der West-und Südmongolei (Nördliches Zentralasien) (Stuttgarter geographische Studien, 126).

Wescott, Konnie; Brandon, R. Joe (2003): Practical applications of GIS for archaeologists. A predictive modeling toolkit. London, New York: Taylor and Francis.

Wheatley, D. (1996): Between the lines: the role of GIS-based predictive modelling in the interpretation of extensive survey data. Analecta Praehistorica Leidensia 28:275–292

White, K.; Bullard, J.; Livingstone, I.; Moran, L. (2015): A morphometric comparison of the Namib and southwest Kalahari dunefields using ASTER GDEM data. In *Aeolian Research* 19, pp. 87–95. DOI: 10.1016/j.aeolia.2015.09.006.

Williams, K. K.; Greeley, R. (2004): Laboratory and field measurements of the modification of radar backscatter by sand. In *Remote Sensing of Environment* 89 (1), pp. 29–40. DOI: 10.1016/j.rse.2003.09.006.

Willmes, C. (2015): LGM sealevel change (HiRes). In *Collab. Res. Centre 806 (CRC806)*.

Willmes, C. (2016): CRC806-Database: A semantic e-Science infrastructure for an interdisciplinary research centre. Dissertation.

Wren, C. D.; Burke, A. (2019): Habitat suitability and the genetic structure of human populations during the Last Glacial Maximum (LGM) in Western Europe. In *PLOS ONE* 14 (6), e0217996. DOI: 10.1371/journal.pone.0217996.

Yang, J.; Dong, Z.; Liu, Z.; Shi, W.; Chen, G.; Shao, T.; Zeng, H. (2019): Migration of barchan dunes in the western Quruq Desert, northwestern China. In *Earth Surface Processes and Landforms* 44 (10), pp. 2016–2029. DOI: 10.1002/esp.4629.

Yang, Z.; Gao, X.; Lei, J.; Meng, X.; Zhou, N. (2022): Analysis of spatiotemporal changes and driving factors of desertification in the Africa Sahel. In *CATENA* 213, p. 106213. DOI: 10.1016/j.catena.2022.106213.

Yaworsky, P. M.; Vernon, K. B.; Spangler, J. D.; Brewer, S. C.; Codding, B. F. (2020): Advancing predictive modeling in archaeology: An evaluation of regression and machine learning methods on the Grand Staircase-Escalante National Monument. In *PLOS ONE* 15 (10), e0239424. DOI: 10.1371/journal.pone.0239424.

Yu, H.; Verburg, P. H.; Liu, L.; Eitelberg, D. A. (2016): Spatial Analysis of Cultural Heritage Landscapes in Rural China: Land Use Change and Its Risks for Conservation. In *Environmental Management* 57 (6), pp. 1304–1318. DOI: 10.1007/s00267-016-0683-5.

Zhang, T.; Ramakrishnon, R.; Livny, M. (1996): BIRCH: Method for very large databases. In *Proceedings of the ACM, Management of Data*, pp. 103-114. Montreal, Canada.

Zhao, Q.; Le Yu; Du, Z.; Peng, D.; Hao, P.; Zhang, Y.; Gong, P. (2022): An Overview of the Applications of Earth Observation Satellite Data: Impacts and Future Trends. In *Remote Sensing* 14 (8), p. 1863. DOI: 10.3390/rs14081863.

Zheng, Z.; Du, S.; Taubenböck, H.; Zhang, X. (2022): Remote sensing techniques in the investigation of aeolian sand dunes: A review of recent advances. In *Remote Sensing of Environment* 271, p. 112913. DOI: 10.1016/j.rse.2022.112913.

Zimmermann, A.; Hilpert, J.; Wendt, K.P. (2009): Estimations of population density for selected periods between the Neolithic and AD 1800. In Hum. Biol. 81. DOI: 10.3378/027.081.0313

Zink, M.; Bachmann, M.; Brautigam, B.; Fritz, T.; Hajnsek, I.; Moreira, A. et al. (2014): TanDEM-X: The New Global DEM Takes Shape. In *IEEE Geosci. Remote Sens. Mag.* 2 (2), pp. 8–23. DOI: 10.1109/mgrs.2014.2318895.

# Appendix A

## Supplementary material for chapter 5: Testing the representativity of Palaeolithic site distribution: The role of sampling bias in the European Upper and Final Palaeolithic record

### A1 Geospatial analysis (extended version)

To reduce the complexity of the geospatial analysis, we implemented an extensive pre-processing workflow, aimed at harmonising all used environmental geodatasets and preparing them for geospatial query in ArcGIS, version 10.7.1. For the tabular datasets of Upper Palaeolithic sites, this includes internal harmonisation of columns and transformation into a single vector-based geodata file (shapefile, .shp). For easy handling of the resulting file, only columns relevant for individual identification and categorical selection, namely *period*, *temporal subdivision*, *coordinate precision*, *quality* and *site type* (cave/open air) were preserved.

The pre-processing of vector-based environmental geodatasets included merging of all relevant classes from each dataset separately and then dissolving them into categories selected for this study. In this form, only one spatial query per environmental geodataset is required. The spatial query between the site dataset and the vector-based geodataset itself was then carried out via spatial joins (ArcGIS-tool: *spatial join*). To account for possible inaccuracies in the archaeological dataset and/or in the environmental geodatasets, we used a search radius of 500m around each point. It is important to note that this search radius only is applied when the archaeological site does not intersect the geodataset. By use of the spatial join method, values from each environmental geodataset are written into the properties (attributes) of each site point, enabling later analysis in tabular form.

The spatial query for raster-based environmental geodatasets was carried out by extracting values based on spatial intersection with the site locations (ArcGIS-tool: *Extract Multi Values to Points*). This way, the information of each raster cell intersecting with a site is written into the properties (attributes) for later analysis. As some of the environmental geodatasets contain categorical data, cell values at site locations were not interpolated.

Additional steps were implemented to enable a better statistical analysis of the CLC dataset. This dataset contains 44 land-cover classes, which makes statistical analyses prone to outliers in combination with small comparison datasets. Therefore, it was reclassified and aggregated based on preliminary results. The reclassification reduces the number of classes to 10, aggregating related surface types that show similar site frequencies while preserving single classes that show noticeably increased or decreased site frequencies. The reclassification is shown in Table A12. For comparison, all spatial queries (the original and the aggregation) were conducted with both CLC-rasters.

*Table A12: Corine Land Cover reclassification table.*

| Original CLC classes (n = 44) | New aggregated class (n = 10) |
|---|---|
| 111, 112, 121, 122, 123, 124 | Urban fabric |
| 131 | Mineral extraction site (unchanged) |
| 132, 133, 141, 142 | Other man-made surfaces |
| 211, 212, 213 | Cropland |
| 221 | Vineyards (unchanged) |
| 222, 223, 241, 242, 243 | Other agricultural surfaces |
| 231 | Pastures (unchanged) |
| 244, 311, 312, 313 | Forest |
| 321, 322, 323, 324 | Non-forest natural vegetation cover |
| 331-999 | Other natural surfaces |

The result of these two steps is a vector database containing a point for each archaeological site. Apart from site properties, such as period, site type etc., the attribute table contains the value of each environmental geodataset at the respective location. For further statistical analysis, this tabular database was exported as a .csv-file.

Parallel to the intersection of sites and environmental geodatasets, we extracted reference data based on the defined area of interest (AOI), presented under *2.1 Upper Palaeolithic sites*. The purpose of this reference is to define the expected values and shares of each environmental variable to subsequently analyse whether the distribution of sites deviates from this expectation. In a first step, we narrowed down the AOI to only include areas where human existence was potentially possible during the Upper and Late Palaeolithic. We chose the elevation as the parameter for this exclusion and selected the value >1200 m.a.s.l. based on an empirical exploration of the site dataset. This way, only 19 of 4194 sites are not represented by the reference data. We are aware that this is a simplification and probably leaves periodically uninhabitable regions in the reference data. However, the AOI has to fit all phases of the investigation period and thus requires compromises.

Based on the polygons representing the AOI at elevations lower than 1200 m.a.s.l., reference values and shares of the environmental geodatasets were gathered. This process was carried out three times: (1) For all of the AOI, (2) for the NE-Section and (3) for the SW-Section, as defined under *5.2.1 Upper Palaeolithic sites*. For vector-based environmental geodatasets, this included clipping to the AOI and calculating area statistics per class. From these absolute areas per class, class shares of the AOI were calculated. For categorical raster datasets, a similar approach was carried out by use of the *zonal histogram*-tool. In this case, AOI-shares were calculated based on pixel counts. For environmental geodatasets represented by continuous values such as the DEM, summarising statistics including mean, standard deviation (STD), median and percentiles were gathered based on the respective AOI polygon using the *zonal statistics as table*-tool. All area calculations in this study were carried out using the lambert azimuthal equal area coordinate reference system (CRS), based on the ETRS89 Datum (ETRS89-LAEA, EPSG code: 3035). The purpose of this CRS is to ensure maximum accuracy in Pan-European statistical mapping (More information available online at epsg.io/3035).

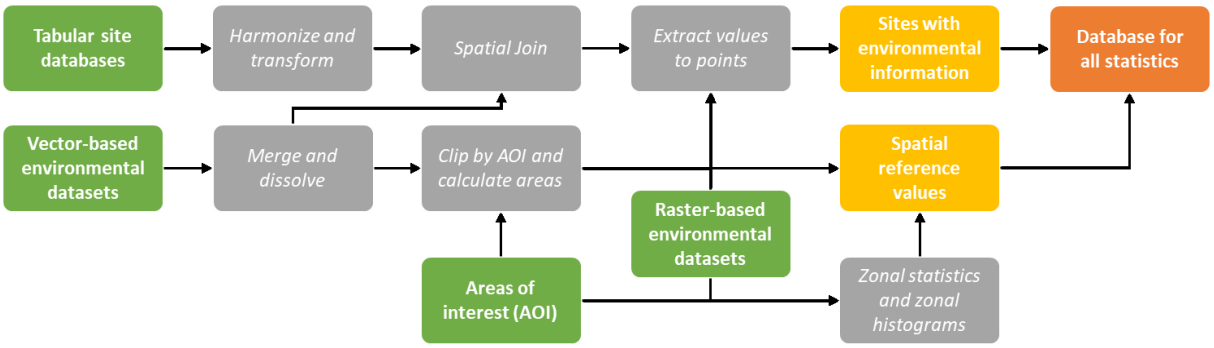



*Figure A24: Workflow diagram of the geospatial analysis. Green: Input, Grey: Spatial processing, Yellow: Intermediate result, Orange: End result*

## A2: Statistical analysis (extended version)

All results from the geospatial analysis were exported in tabular form for a comprehensive statistical analysis in Excel 2016 and IBM SPSS. For these analyses, the attribute table containing Upper Palaeolithic sites, site properties and information about underlying environmental data, was used as main database. In a first step, we utilised the site properties to divide the dataset into chronological

(AUR-FP), spatial (NE/SW) and type-specific (cave/open air) subsets. When considering every subset and aggregation of these combinations, this results in site 49 classes.

In the next step, we utilised the values of continuous environmental variables to calculate mean, STD, median and quartiles for each Upper Palaeolithic site class in Excel. These were calculated directly from the main database for each class separately, using if-then-else-clauses to only include sites attributed to the respective class. The only exception to this workflow was the aspect, which was classified into the 8 main compass directions (north, north-east, east…) and treated as a categorical environmental variable. From categorical environmental variables, frequencies of occurrences (n) were calculated. These were also calculated directly from the main database for each class separately, using count-if-clauses to only include sites attributed to the respective class. The resulting frequencies were then set into perspective to the overall class numbers, resulting in percentages of each of the 49 site classes.

To assess whether these means and class shares represent an under- or overrepresentation of the respective environmental variable or class in the site distribution, we compared them to their reference value, calculated based on the AOI. To simplify the interpretation, comparison and visualisation, this value was transformed to a percentage and shifted to:

- <0 = Sites are underrepresented (Lower mean or class share in the site-database than in the AOI)
- 0 = No over- or underrepresentation (Same mean or class share in the site-database as in the AOI)
- >0 = Sites are overrepresented (Higher mean or class share in the site-database than in the AOI)

This scale theoretically has no upper limit, as means and class shares can be any multiples of the reference AOI values. Within the text and the charts, this value is named ‘derivation from expected mean’ or ‘derivation from expected share’. The term ‘expected’ is used as the AOI-based means and class shares represent the statistical expectation, while the site-based means and class shares represent the observed values.

To determine how the environmental variables influence the presence or absence of sites, the dataset was additionally analysed based on the maximum entropy principle. For this analysis, we used the software MaxEnt, version 3.4.4. This software was originally developed for species distribution and environmental niche modelling (Phillips and Dudík 2008; Phillips et al. 2017), but has also been successfully applied in archaeological predictive modelling (Galletti et al. 2013; Gillespie et al. 2016; Jones et al. 2019; Alwi Muttaqin et al. 2019). In this raster-based spatial approach, absence data is generated automatically from raster cells where no site is present. A detailed description of the statistics behind the software can be found in (Elith et al. 2011; Merow et al. 2013).

For processing within this software, all environmental geodatasets were rasterised, resampled to a spatial resolution of 100*100 metres using the nearest neighbour method and saved in ASCII format. The site dataset was transformed into .csv format, only preserving coordinates and class information based on period, section and site type (e.g. *AUR_SW_Cave*). As a predictive model is not the aim of this study and the environmental variables are not sufficient to support such predictive analysis on a European scale, only intermediate statistical results were extracted for further analysis. These intermediate statistical results are namely response curves and jackknife variable importance. The response curves show the predictive probability that is associated to the values of each environmental variable, allowing for a comparison to the conventionally calculated ‘derivation from expected share’ and ‘derivation from expected mean’. The jackknife variable importance can be used to estimate the predictive value of an environmental variable as a whole, which is somewhat comparable to conventionally calculated determination measurements based on presence/absence data.

For additional internal statistical queries and tests, the tabular database was further analysed using the statistics software IBM SPSS statistics. This internal analysis was conducted to determine how the environmental variables influence each other and if/how they differ between the archaeological classes (based on culture, type and section). To this end, we conducted the following tests and approaches:

- Crosstabs with measures of determination to test for cross-correlations between the different environmental variables
- Two-Step unsupervised classification to test the similarities and differences between the archaeological classes and the environmentally-dictated unsupervised classification result
- Discriminant analysis to test the differences and similarities of environmental variables within the archaeological classes
- Naïve Bayes to test the strength that certain environmental variables have in predicting archaeological classes.

As some of these classification algorithms and statistical tests only work with categorical/nominal data, the continuous environmental variables were classified using different approaches. For aspect and slope, predefined topographical classes such as *337.5 < aspect > 22.5 = north* or *4° < slope > 9° = gentle slope* were used. For elevation, build-up area and population density, equal intervals were used for reclassification. The used intervals were 100 for elevation, 50 for population density and 2 for built up area, resulting in 5 to 12 classes.

## A3: Potential sources for errors and misinterpretations

As all spatial queries with the environmental geodatasets are based on the archaeological point-dataset, inaccuracies within the latter can have a large influence on the results. As the coordinates for all 4200 sites were extracted from literature as mentioned in 2.1, the accuracy is fully dependent on the precision in the original publications. Within these original publications, the main source for inaccuracies is low coordinate precision caused by few decimals used (down to 2) and/or indirect spatial reference (e.g., using a landmark close to the site). In the worst case, this can lead to inaccuracies of single occupation points in the range of several hundred meters. Another potential source for the inaccuracy in the archaeological point-dataset is the relocation of archaeological material after its deposition. This inaccuracy is, however, only relevant for environmental geodatasets representing the settlement factor, as the coordinate does not represent the point of deposition/occupation in this case. As the relocation of archaeological material through geomorphological slope processes normally spans a maximum of several tenths of metres, the inaccuracy induced by this process can be considered far lower than the influence of inaccurate coordinates. Although we have compensated for inaccuracies in the archaeological point-dataset by using a search radius in the spatial queries with environmental geodata, larger inaccuracies in the occupation dataset can still have an influence on the result of the query. For environmental geodatasets in raster format, the magnitude of possible inaccuracy is additionally dependent on the resolution of the raster, as it has a stronger influence on rasters with high spatial resolution (e.g., the DEM or CLC dataset) and makes little to no difference for rasters with low spatial resolution (e.g., the HYDE dataset). See figure S2 for a graphical illustration of the used scales. However, due to the large sample size of close to 4200 occupations in this study, inaccuracies of single points can be expected to have a far smaller influence on the end result than in smaller studies.

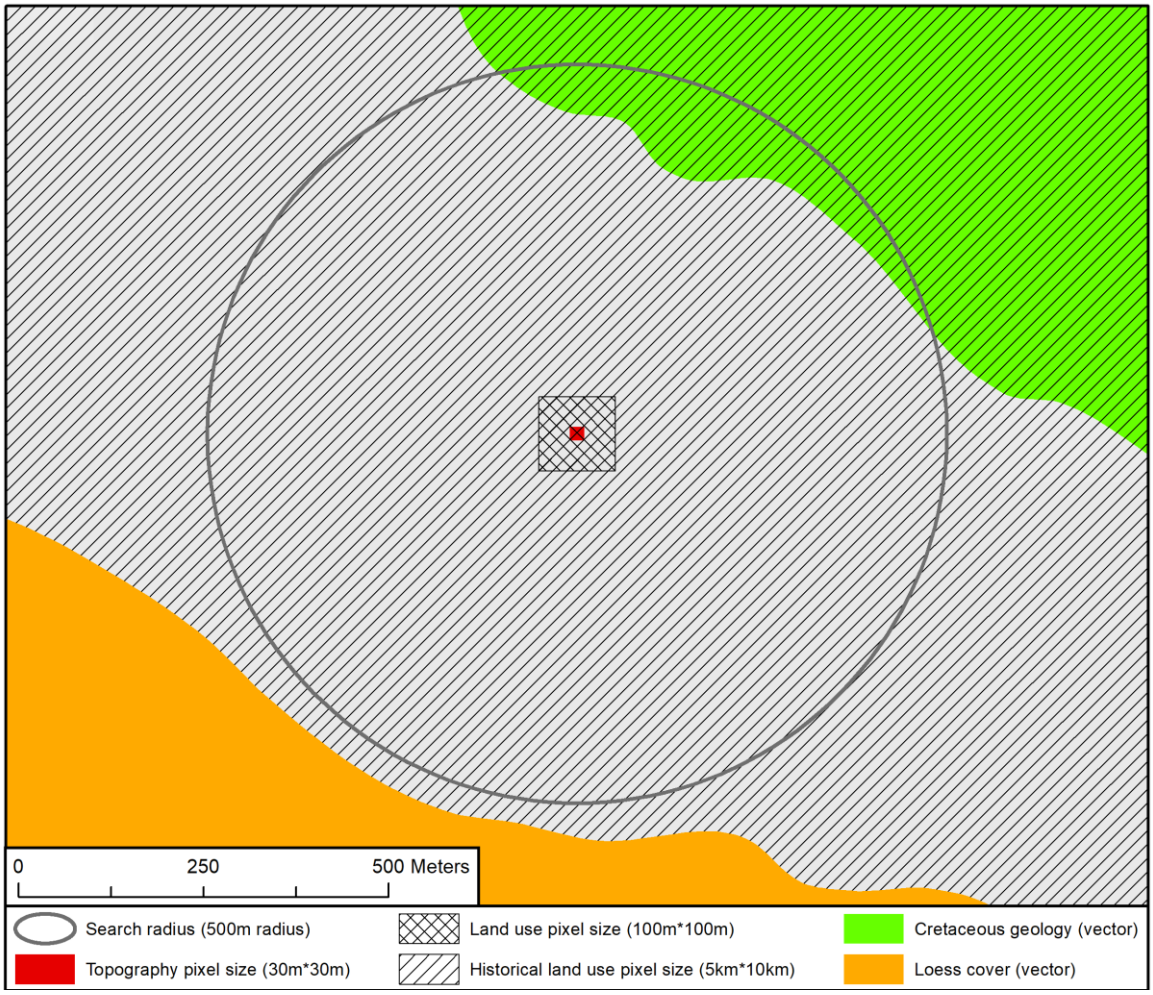



*Figure A25: Visualisation of the different scales used in this approach. Note that the cretaceous geological unit will be set as valid for an archaeological occupation at the centre of the map while the loess cover will not due to intersection with the search radius.*

Another potential for inaccuracies lies in the environmental geodatasets, as the spatial queries with these can only be as accurate as the datasets themselves. For a detailed assessment of these dataset-specific sources for inaccuracies, please refer to the individual source literature that can be found in 2.2. Specific to the approach in this study, however, the environmental geodatasets can also be a source for potential misinterpretations. This is due to the fact that they are used to represent factors of the settlement or discovery context. These factors, however, are a combination of very complex conditions, that cannot be fully represented by the used environmental geodatasets. In addition, the influence that some geodatasets have on the respective factor can be indirect or multifaceted. An example for this is the influence of loess, which can be discussed for the settlement factor for representing past steppe environments as well as the discovery factor due to the taphonomic bias and reduced visibility. This is why statistical correlation/determination measurements as well as over- and underrepresentation of occupations on surfaces like these always have to be discussed from these different viewpoints.

A different question of representability is how well each environmental geodataset represents the respective aspect of the settlement or discovery factor. For datasets representing the settlement factor, the time lag between the Upper and Late Palaeolithic settlement has to be considered. This

may not be important for the geology, as the timespan between the deposition and today can be seen as geologically short. The DEM, on the other hand, might differ from the paleo-surface, as climatic, geomorphological and hydrological processes were fundamentally different in the Late Pleistocene. As paleo-surface-DEMs are not available on a Pan-European scale, the available modern DEM has to be considered the closest available approximation. For the distribution of Late Pleistocene aeolian sediments, an important information would have been the accumulation before and after the deposition of archaeological material. This could help to differentiate the impact that loess and related sediments have on the settlement choice on the one hand and the preservation and taphonomic bias on the other hand. However, all available Pan-European loess datasets only contain a temporally one-dimensional state of the distribution to date. This distribution is most likely also influenced by post-depositional erosional or other morphological processes. Therefore, it is not clear whether or not the used datasets on Late Pleistocene aeolian deposits fittingly represent both expected influences on the settlement and discovery factor.

For the environmental geodatasets representing the discovery factor, the time lag between the discovery of archaeological material and the date that the dataset represents can also lead to potential misinterpretations. Although the maximum expected timespan between the discovery and the CLC dataset is about 200 years, the anthropogenic impact on the surface has been extensive in this timespan (Jepsen et al. 2015). As the archaeological occupation dataset does not include a discovery date, it is not possible to assess how well the land use dataset represents the conditions at the time of discovery. Therefore, it has to be assumed that the current day land use either is representative for the discovery or that the change to the current land use has led to the discovery (e.g., the change to urban area has led to the discovery due to construction activity). Multiple changes in land use since the discovery, however, cannot be accounted for.

For the HYDE dataset on built up area, the main problem is the very coarse spatial resolution of 5 arc minutes, which corresponds to a pixel area of close to 50 km² in the study area. With such a coarse spatial resolution, the representativity of each raster cell value for the respective site is not guaranteed, as sites could potentially be located several kilometres away from the built-up area. However, as this dataset is the only source for a multi-temporal assessment of one aspect of the discovery factor, it can still be considered the best currently possible estimation in the study area.

As already mentioned in 2.2, each chosen environmental geodataset can only represent a single or a few aspects of the settlement and/or discovery probability. Even a combination of all datasets cannot fully represent these immensely complex fields. Therefore, this study can be seen as a pilot study, assessing the currently available best approximations on a European scale. For further studies, we suggest to expand on this by supplementing datasets representing additional aspects of the settlement and/or discovery factor. To further test the influences of settlement and discovery factors, we also suggest to conduct case studies on regional scales where more environmental geodatasets are available at a higher spatial resolution.

## A4: Maps of Upper and Final Palaeolithic occupations and environmental geodatasets

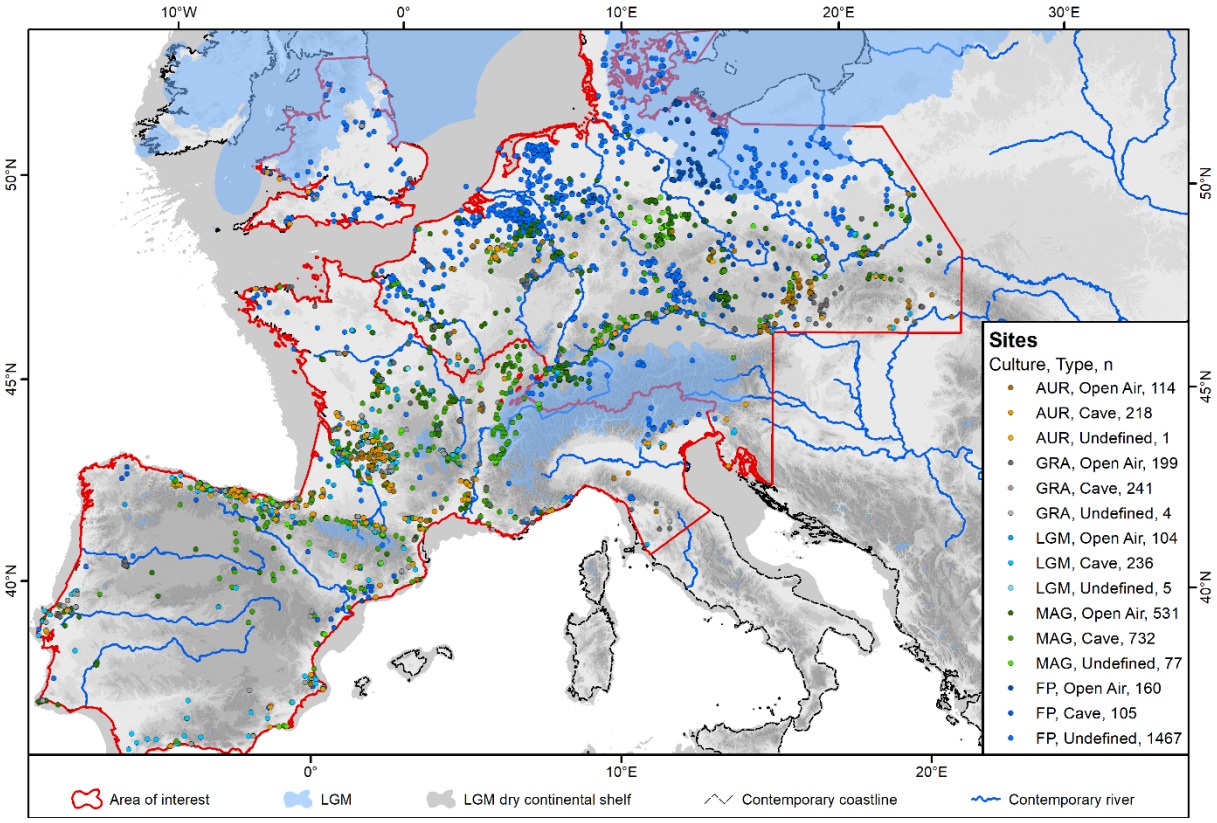



*Figure A26: Map on site distribution and the area of interest*

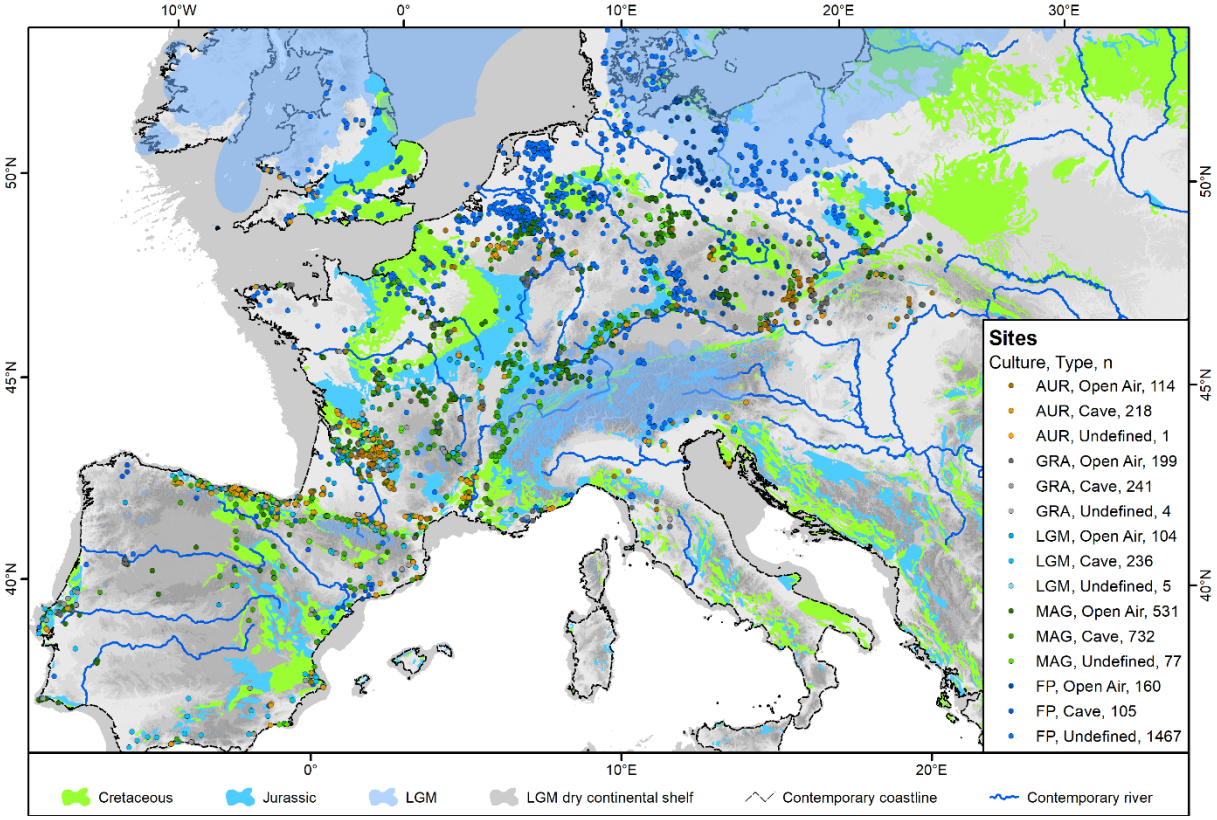



*Figure A27: Map on site distribution and Cretaceous and Jurassic geological units*

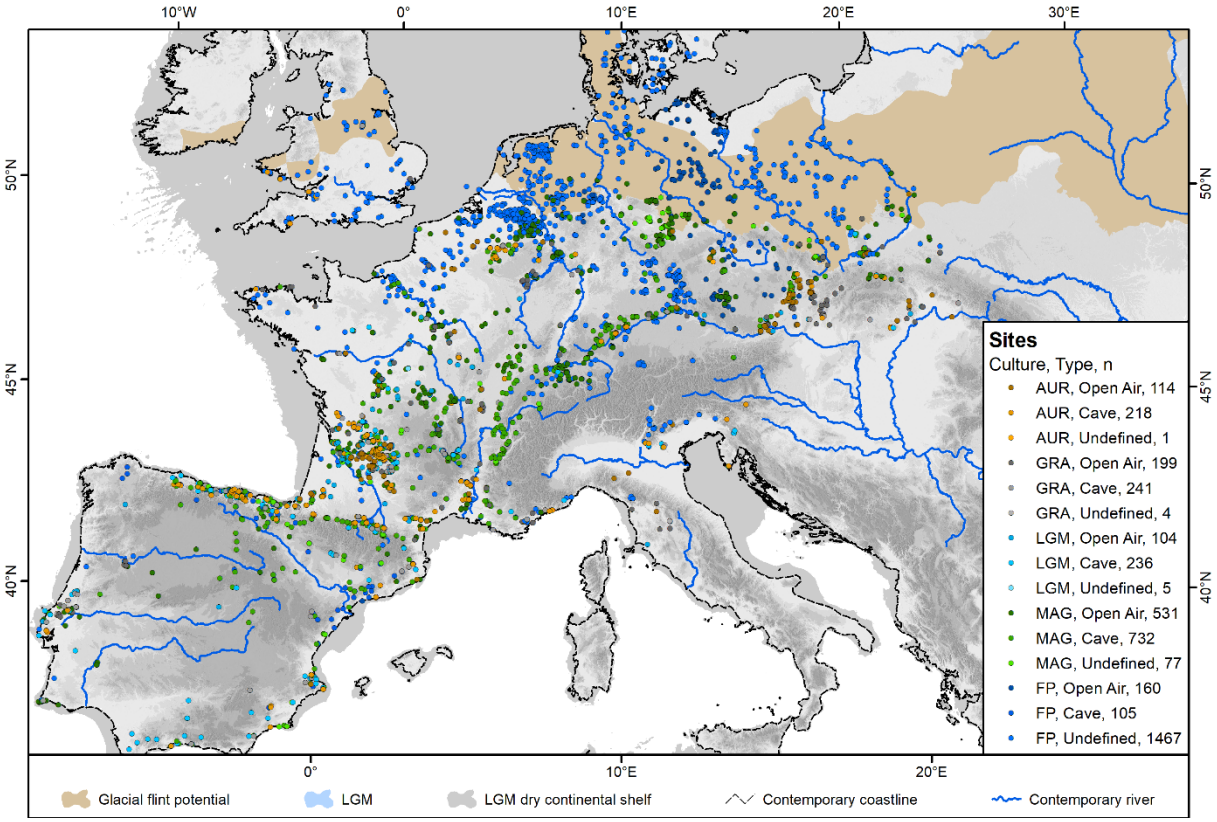



*Figure A28: Map on site distribution and the area marked as glacial flint potential. It corresponds to the accumulation area of the LGM and penultimate glaciation*

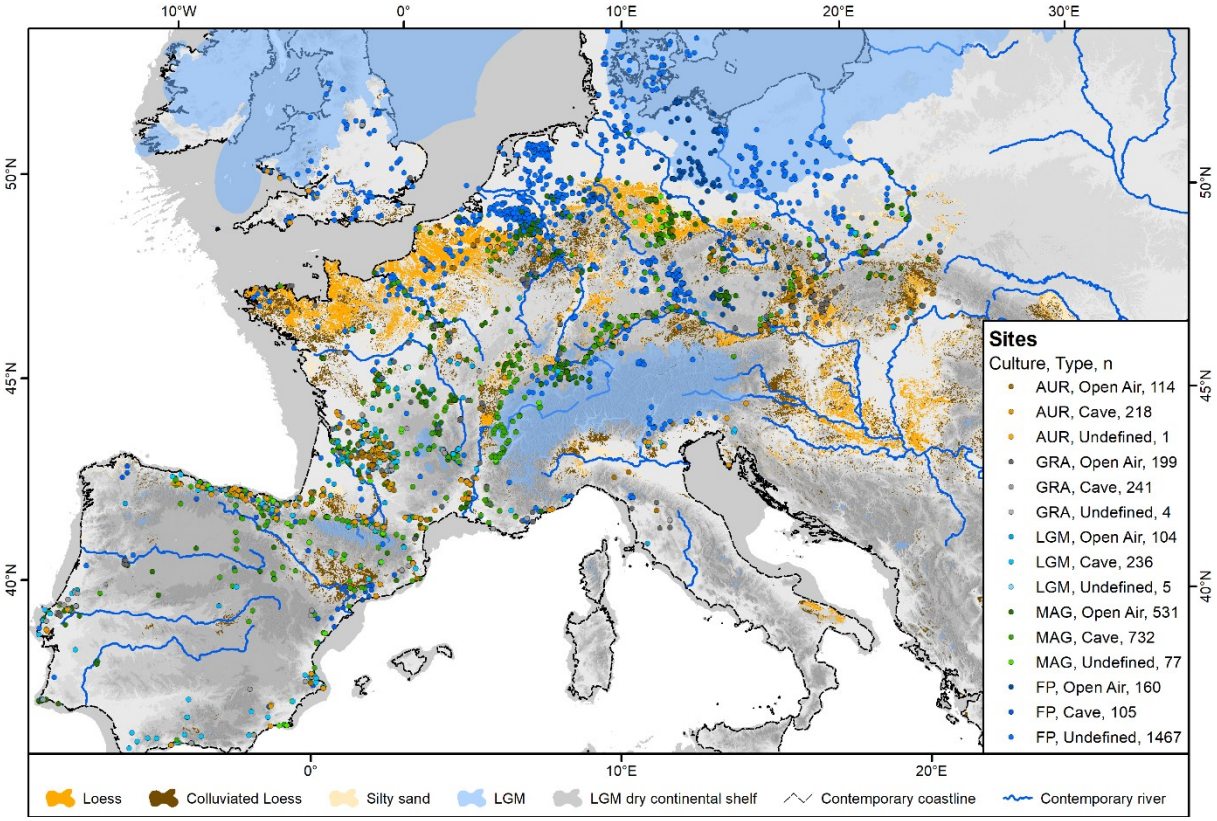



*Figure A29: Map on site distribution and loess and related sediments according to Bertran 2016*

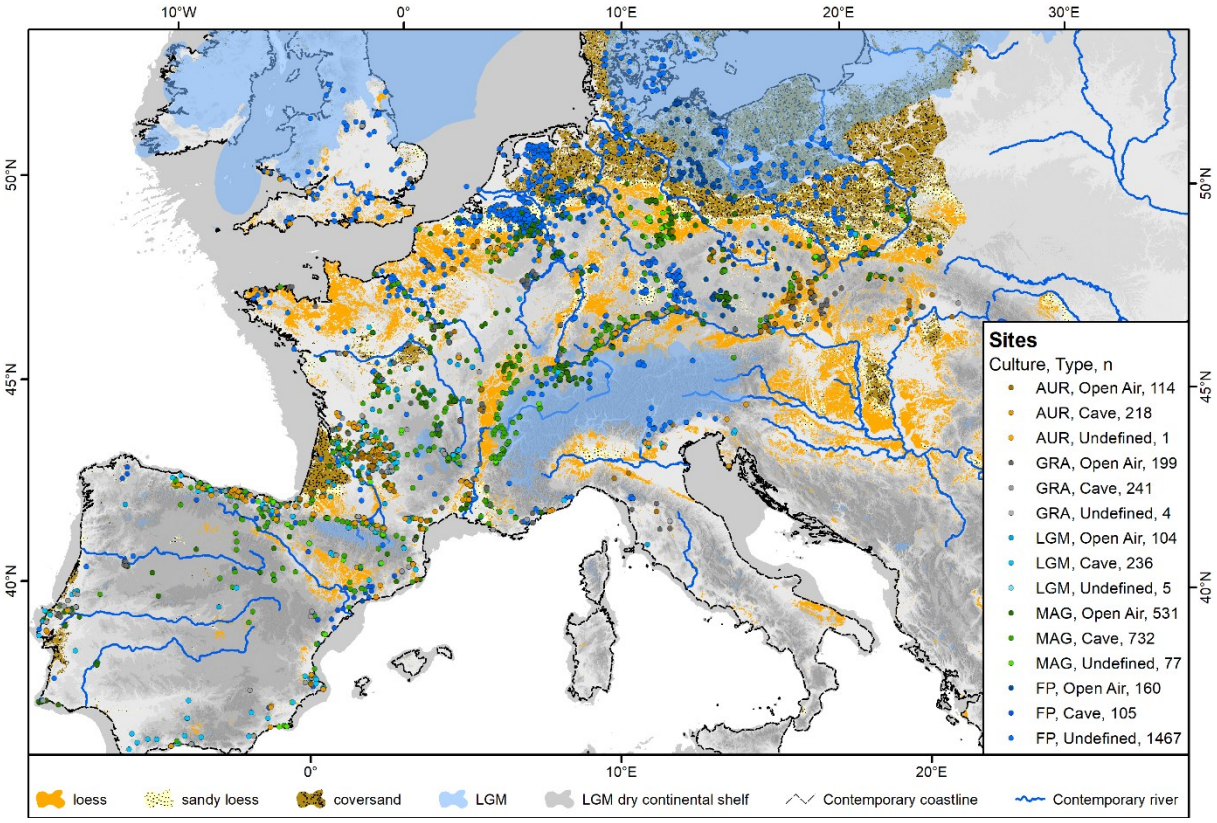



*Figure A30: Map on site distribution and loess and related sediments according to Bertran 2021*

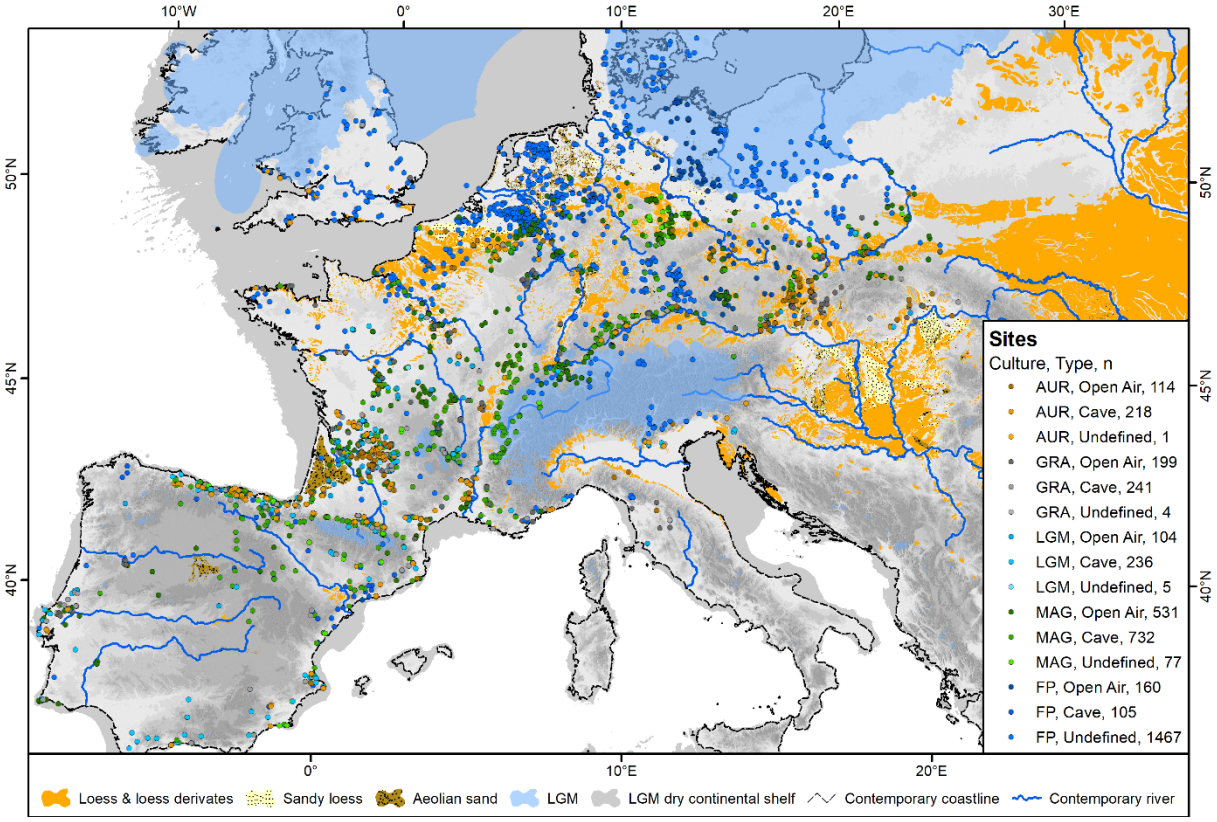



*Figure A31: Map on site distribution and loess and related sediments according to Lehmkuhl 2021*

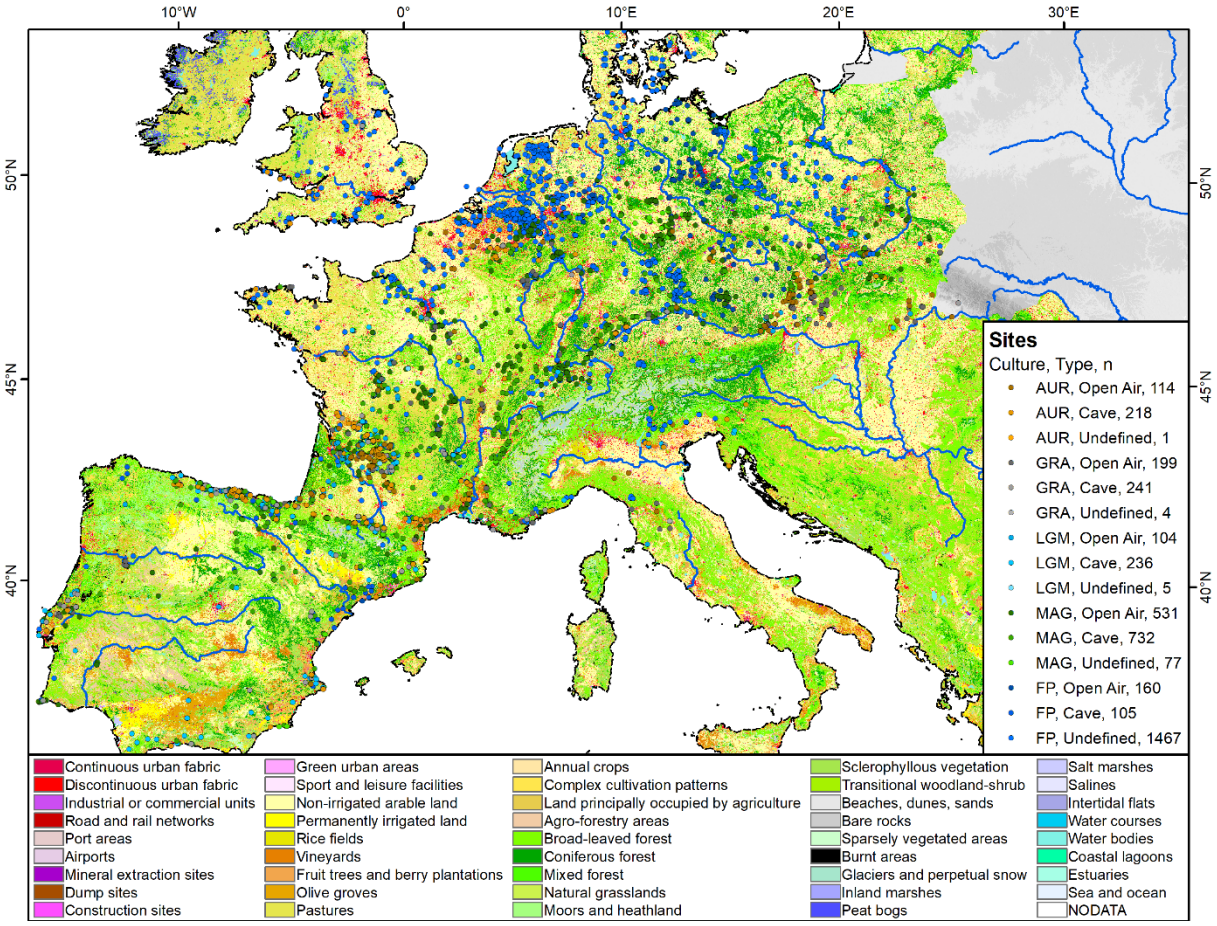



*Figure A32: Map on site distribution and Corine Land Cover dataset*

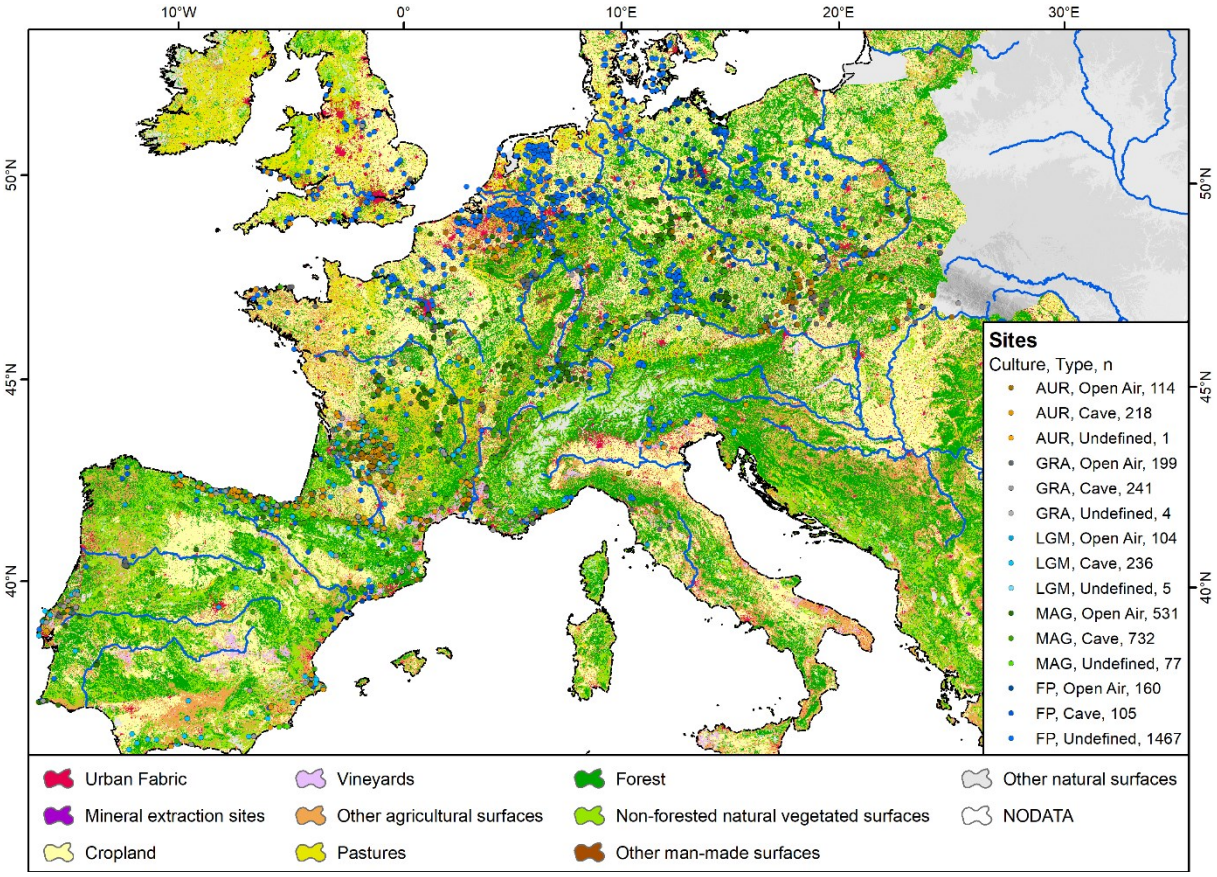



*Figure A33: Map on site distribution and the Corine Land Cover dataset, aggregated to 10 classes*

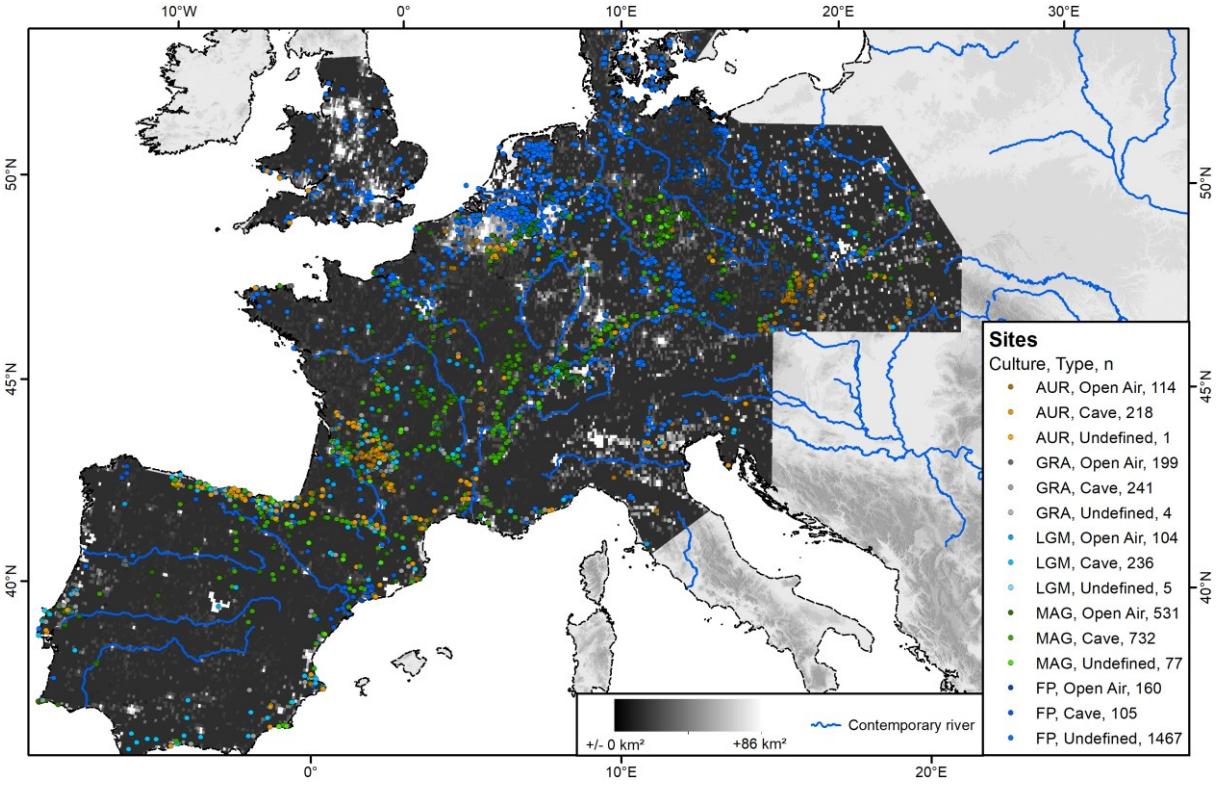



*Figure A34: Map on site distribution and the differences in built up area between 1800 and 2000 according to the HYDE land use model (Klein Goldewijk et al. 2017)*

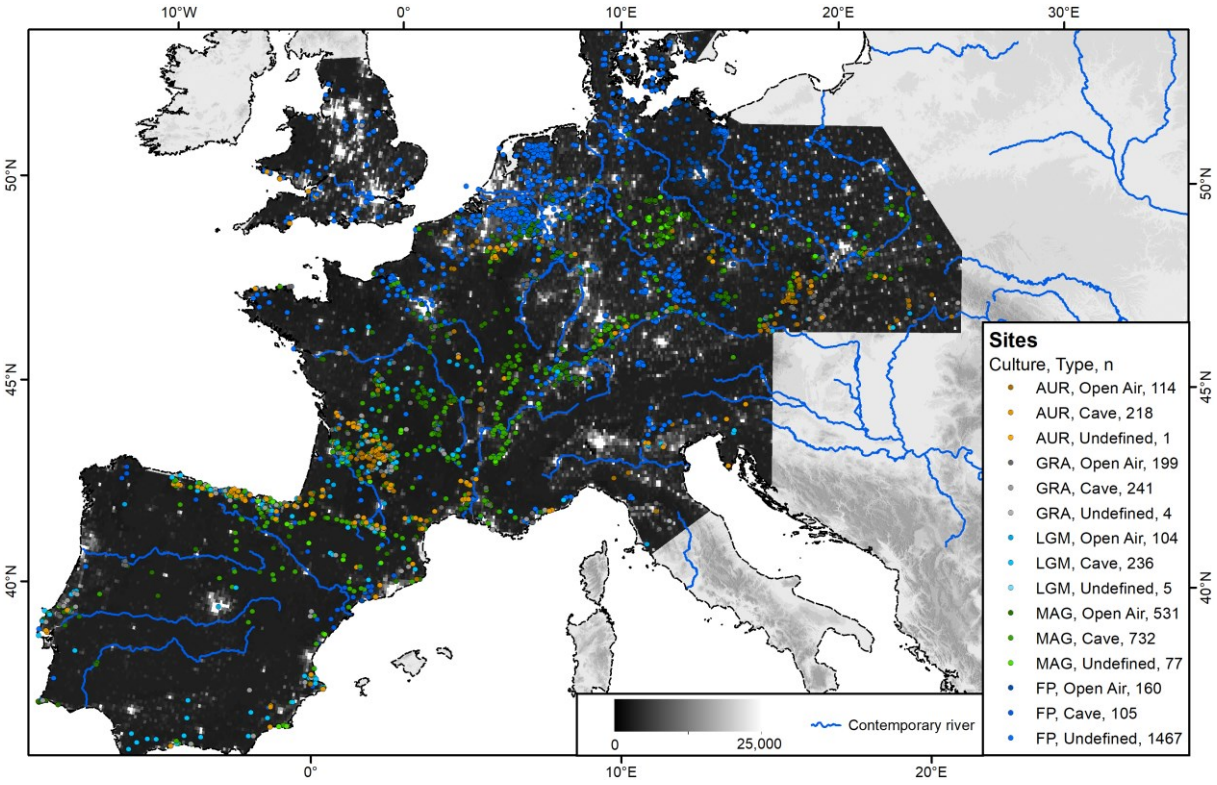



*Figure A35: Map on site distribution and the differences in population density between 1800 and 2000 according to the HYDE land use model (Klein Goldewijk et al. 2017)*

## A5: Charts of the different statistical assessments

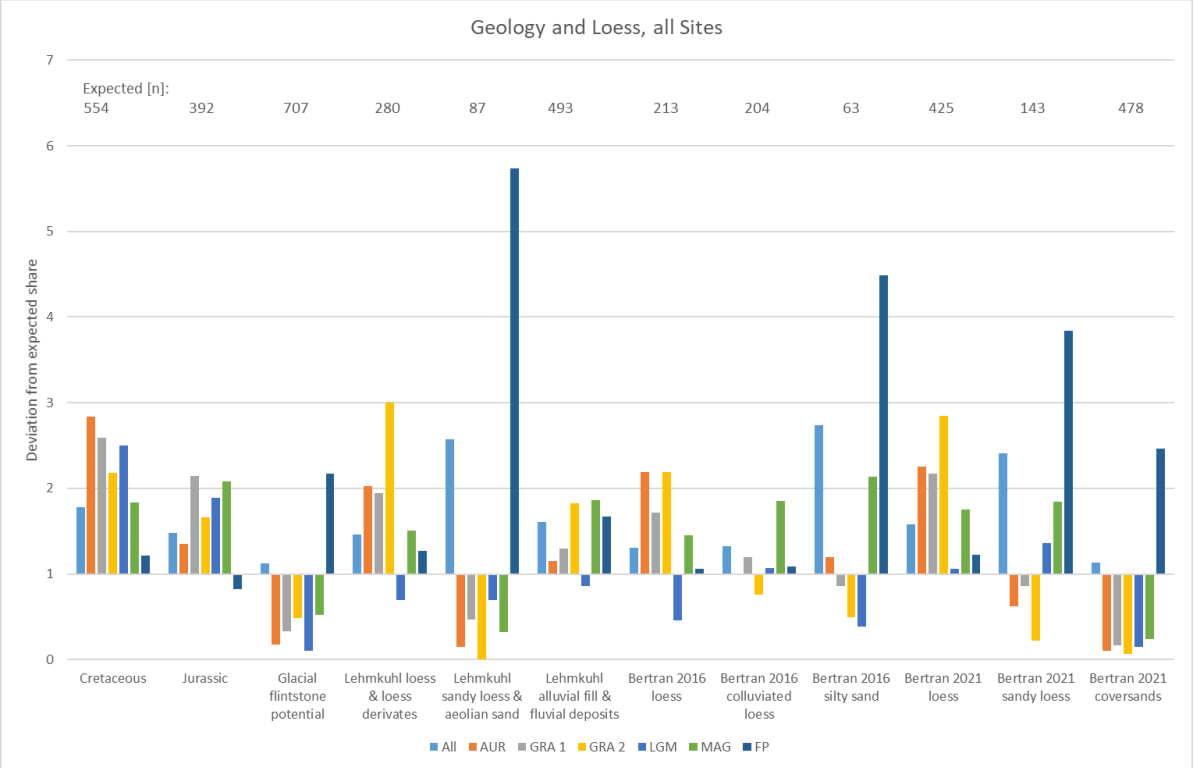



*Figure A36: Bar chart of the over- and underrepresentation. The expected values represent the share of all sites that would equal the area share of each respective surface. The overrepresentation is displayed as a factor of the expected value. Sites: All, Variables: Geology and loess*

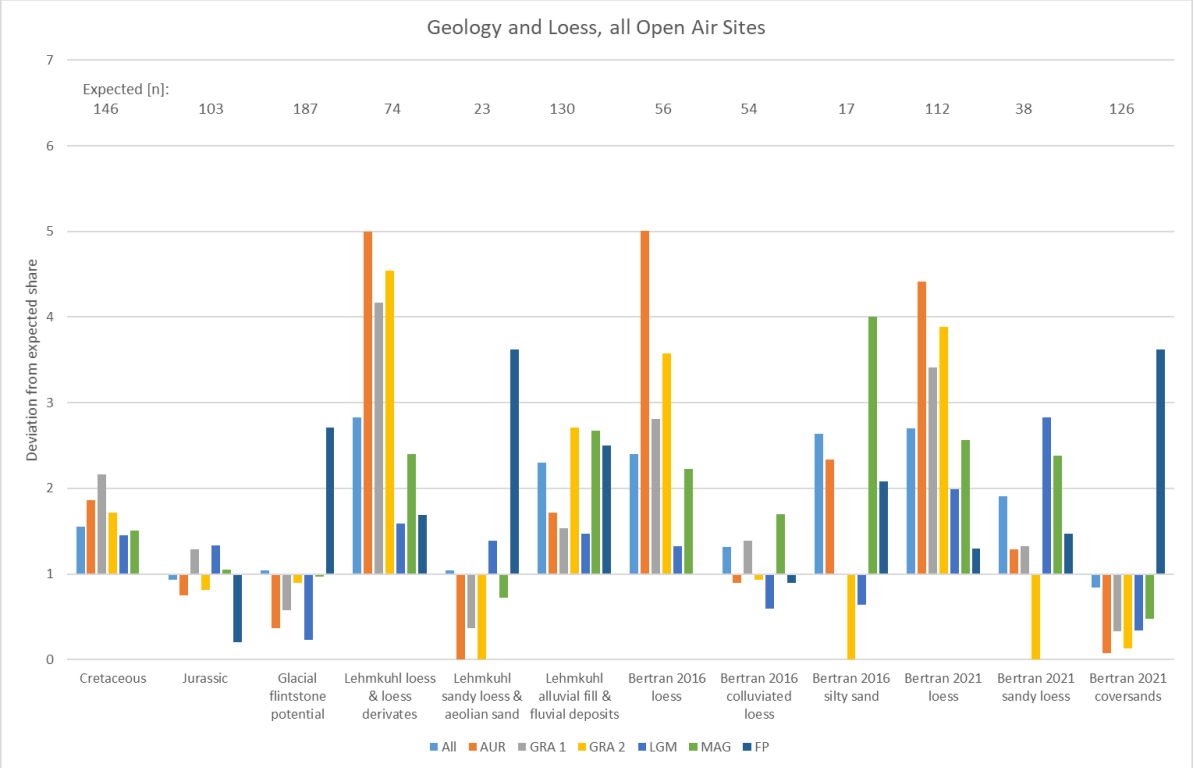



*Figure A37: Bar chart of the over- and underrepresentation. The expected values represent the share of all sites that would equal the area share of each respective surface. The overrepresentation is displayed as a factor of the expected value. Sites: All open air, Variables: Geology and loess*

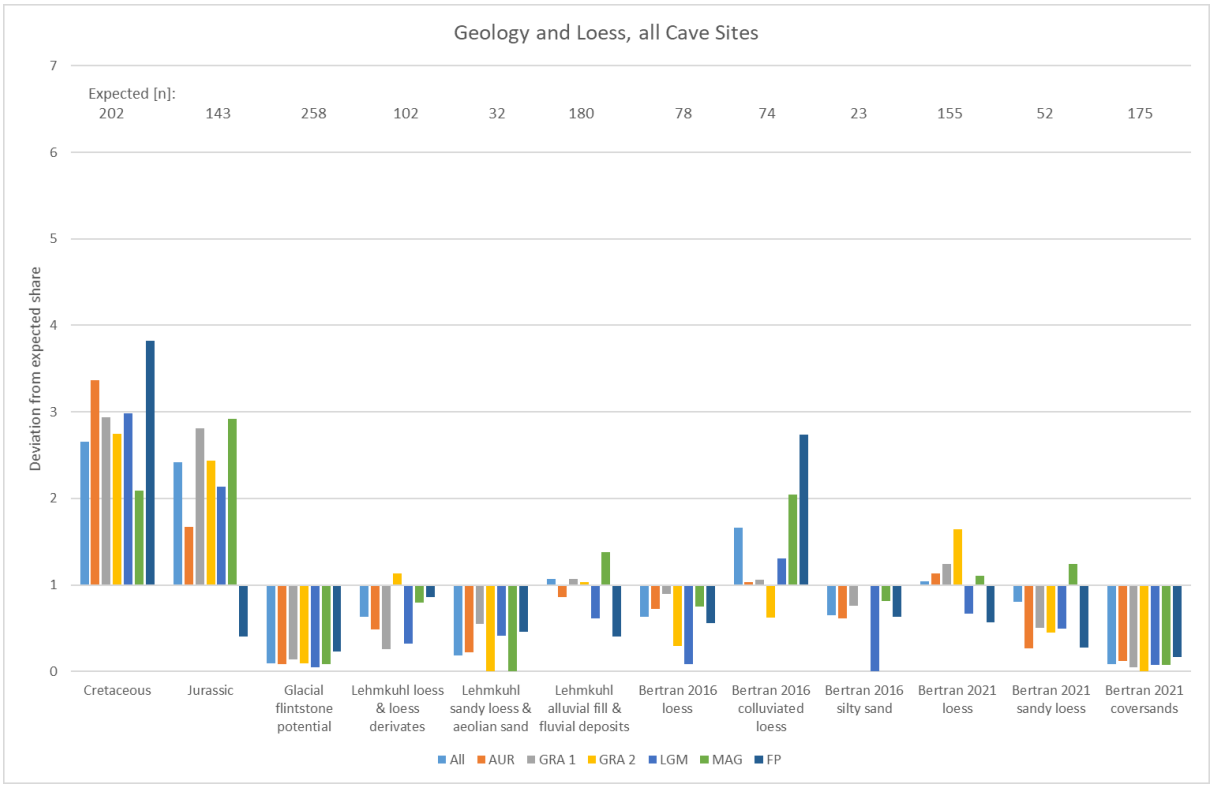



*Figure A38: Bar chart of the over- and underrepresentation. The expected values represent the share of all sites that would equal the area share of each respective surface. The overrepresentation is displayed as a factor of the expected value. Sites: All cave, Variables: Geology and loess*

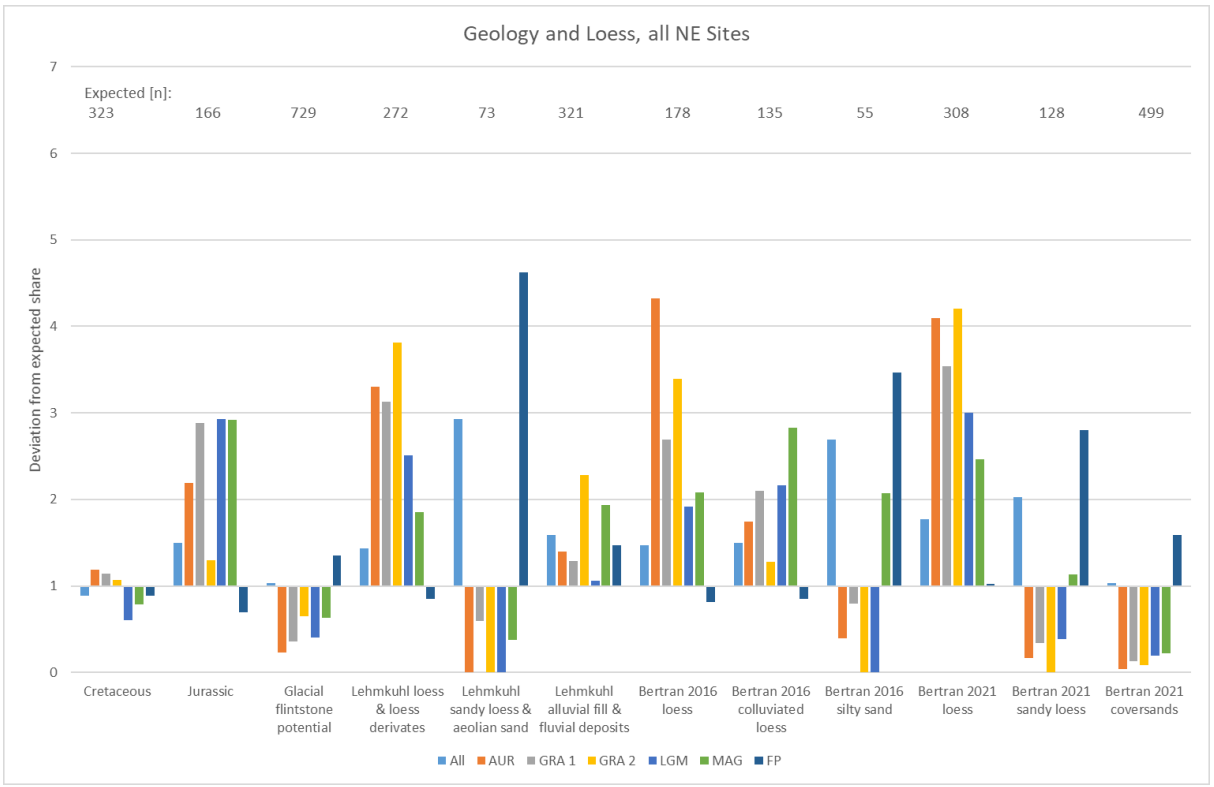



*Figure A39: Bar chart of the over- and underrepresentation. The expected values represent the share of all sites that would equal the area share of each respective surface. The overrepresentation is displayed as a factor of the expected value. Sites: NE, Variables: Geology and loess*

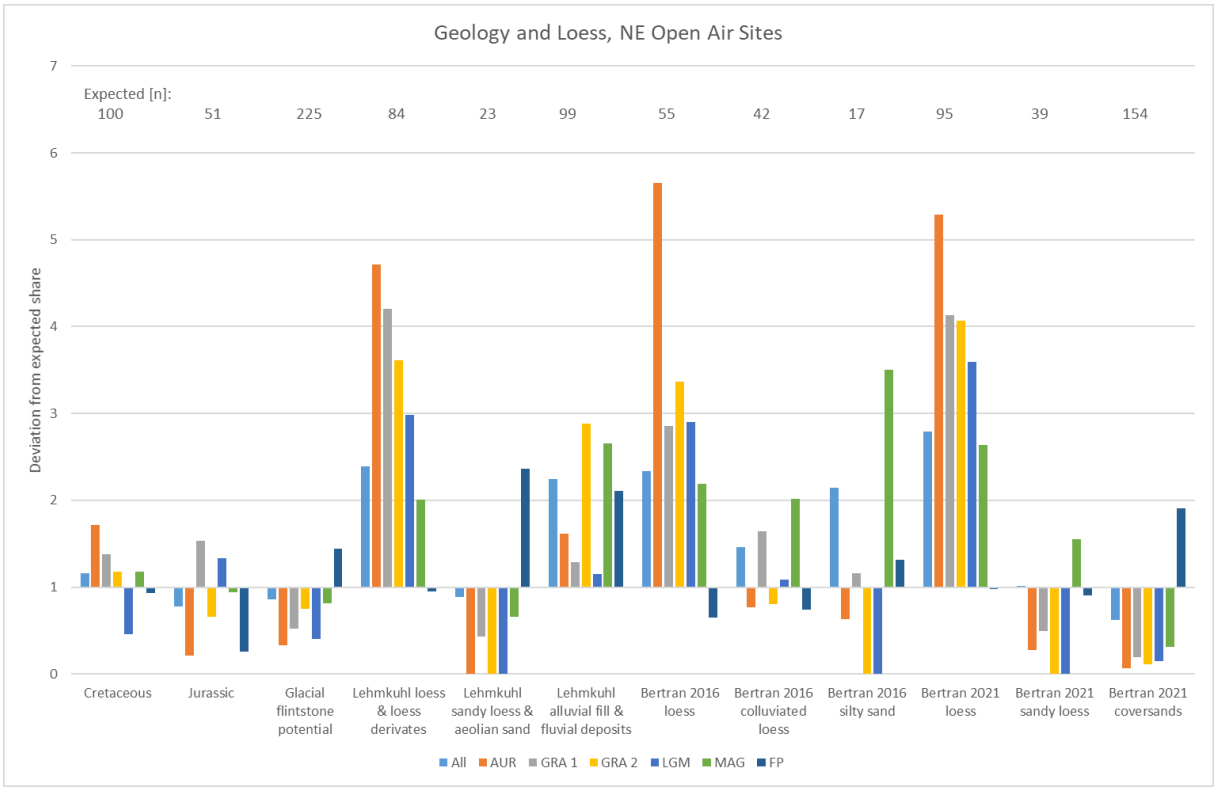



*Figure A40: Bar chart of the over- and underrepresentation. The expected values represent the share of all sites that would equal the area share of each respective surface. The overrepresentation is displayed as a factor of the expected value. Sites: NE open air, Variables: Geology and loess*

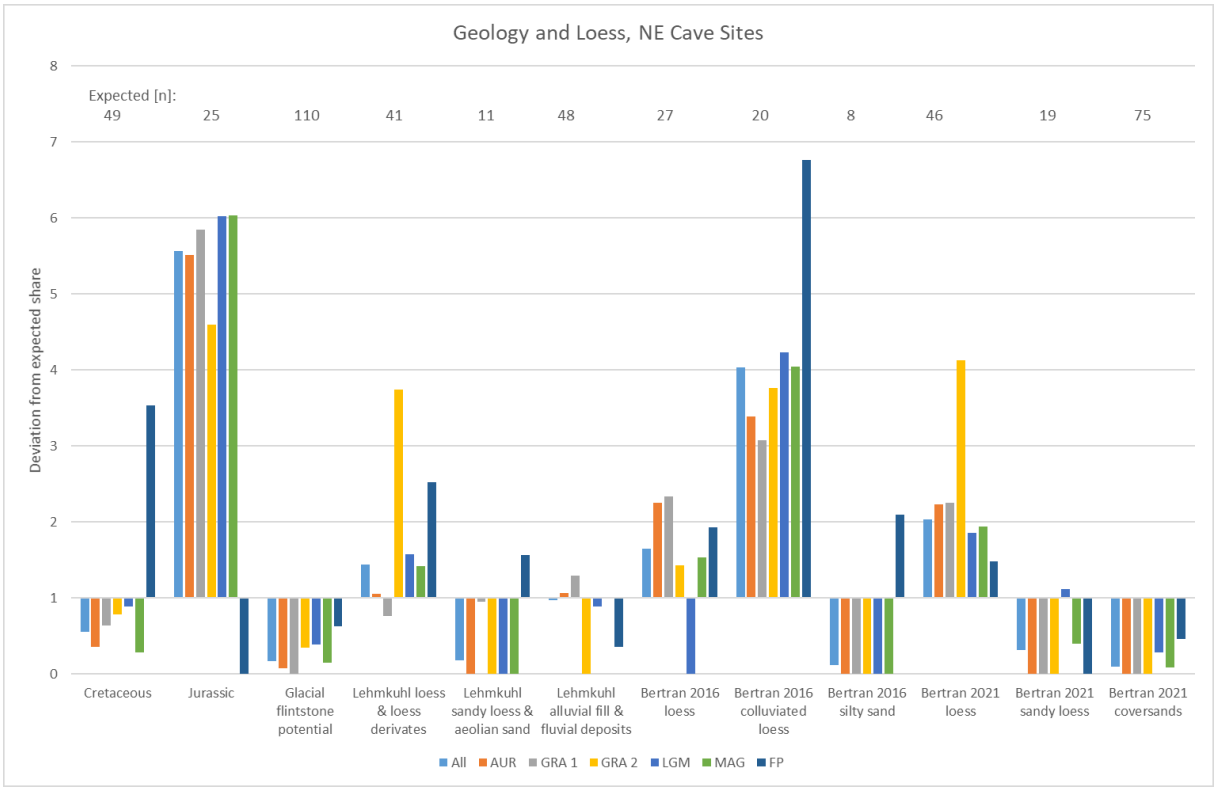



*Figure A41: Bar chart of the over- and underrepresentation. The expected values represent the share of all sites that would equal the area share of each respective surface. The overrepresentation is displayed as a factor of the expected value. Sites: NE cave, Variables: Geology and loess*

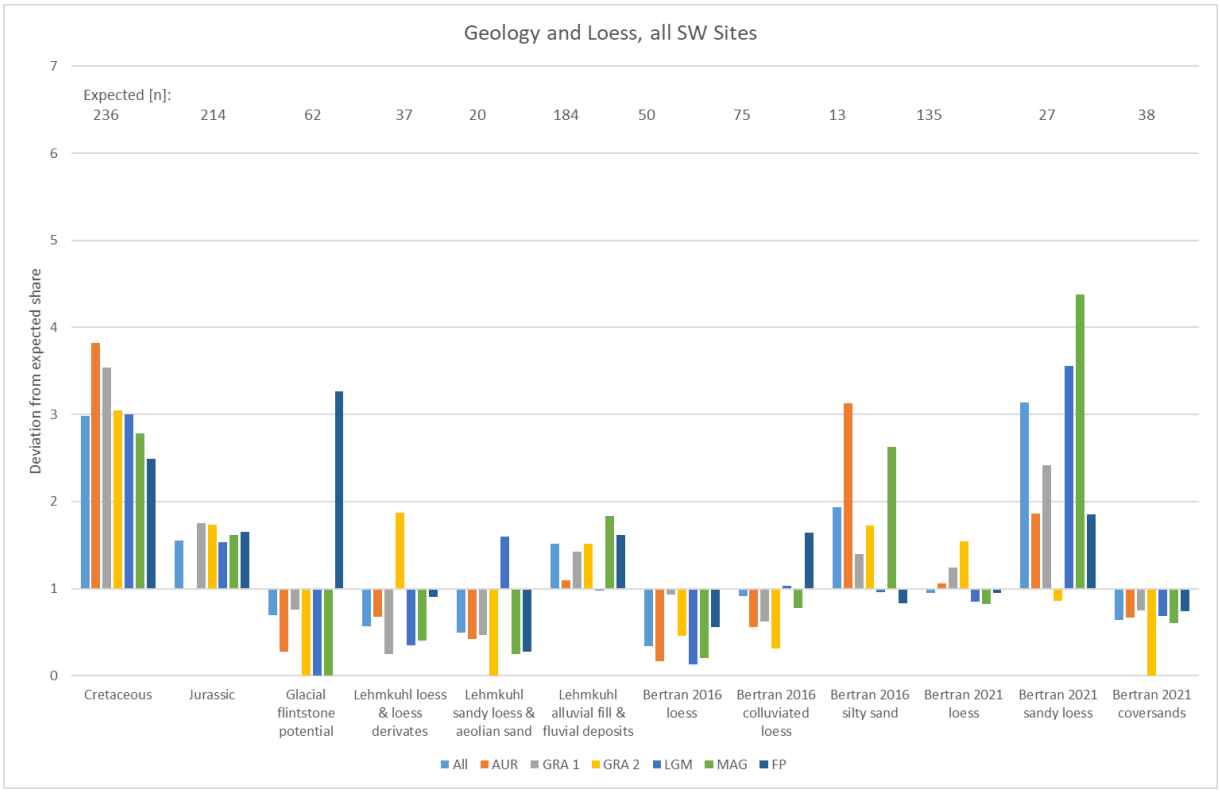



*Figure A42: Bar chart of the over- and underrepresentation. The expected values represent the share of all sites that would equal the area share of each respective surface. The overrepresentation is displayed as a factor of the expected value. Sites: SW, Variables: Geology and loess*

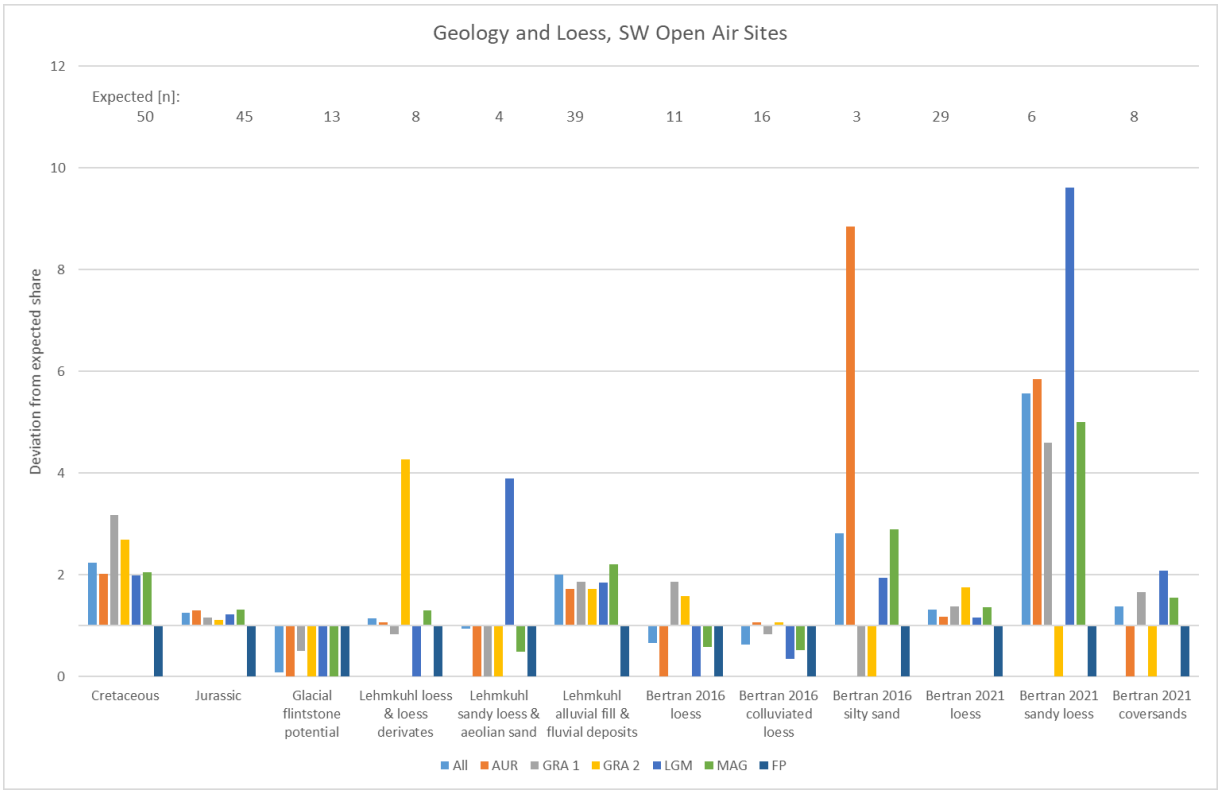



*Figure A43: Bar chart of the over- and underrepresentation. The expected values represent the share of all sites that would equal the area share of each respective surface. The overrepresentation is displayed as a factor of the expected value. Sites: SW open air, Variables: Geology and loess*

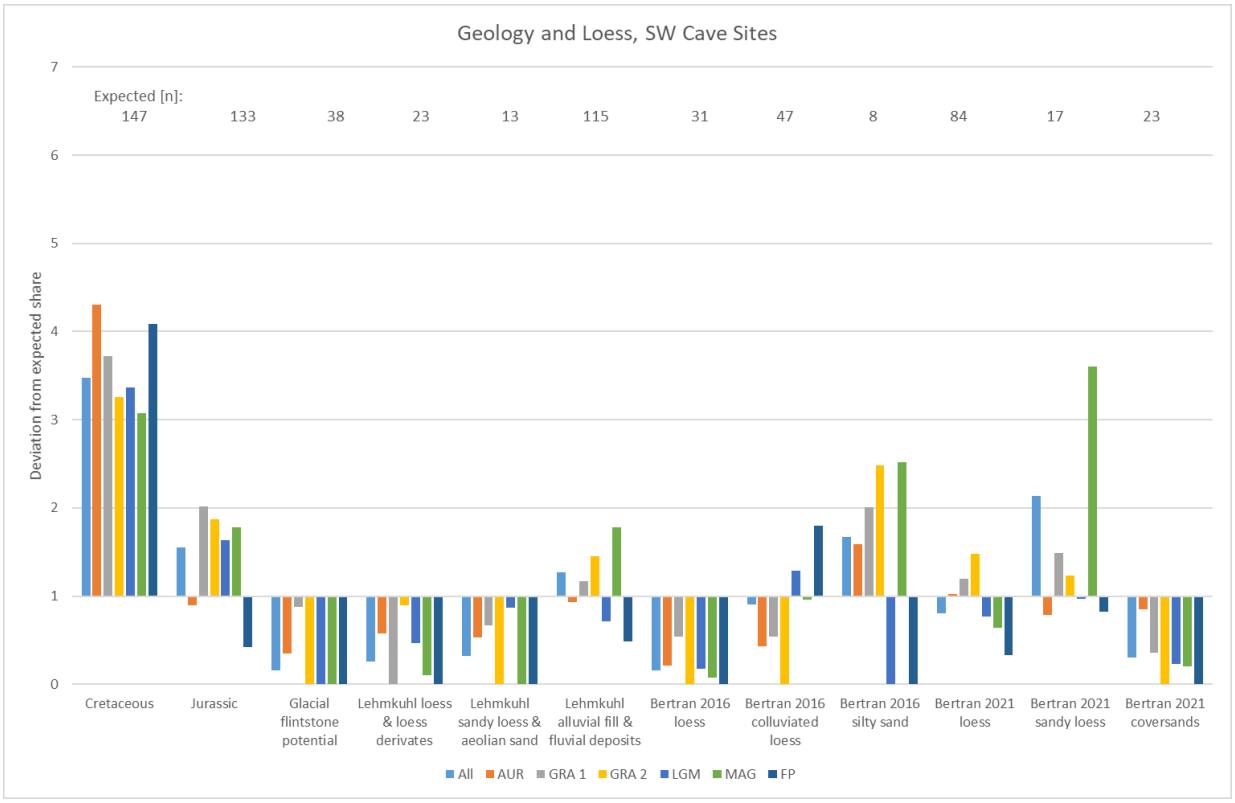



*Figure A44: Bar chart of the over- and underrepresentation. The expected values represent the share of all sites that would equal the area share of each respective surface. The overrepresentation is displayed as a factor of the expected value. Sites: SW cave, Variables: Geology and loess*

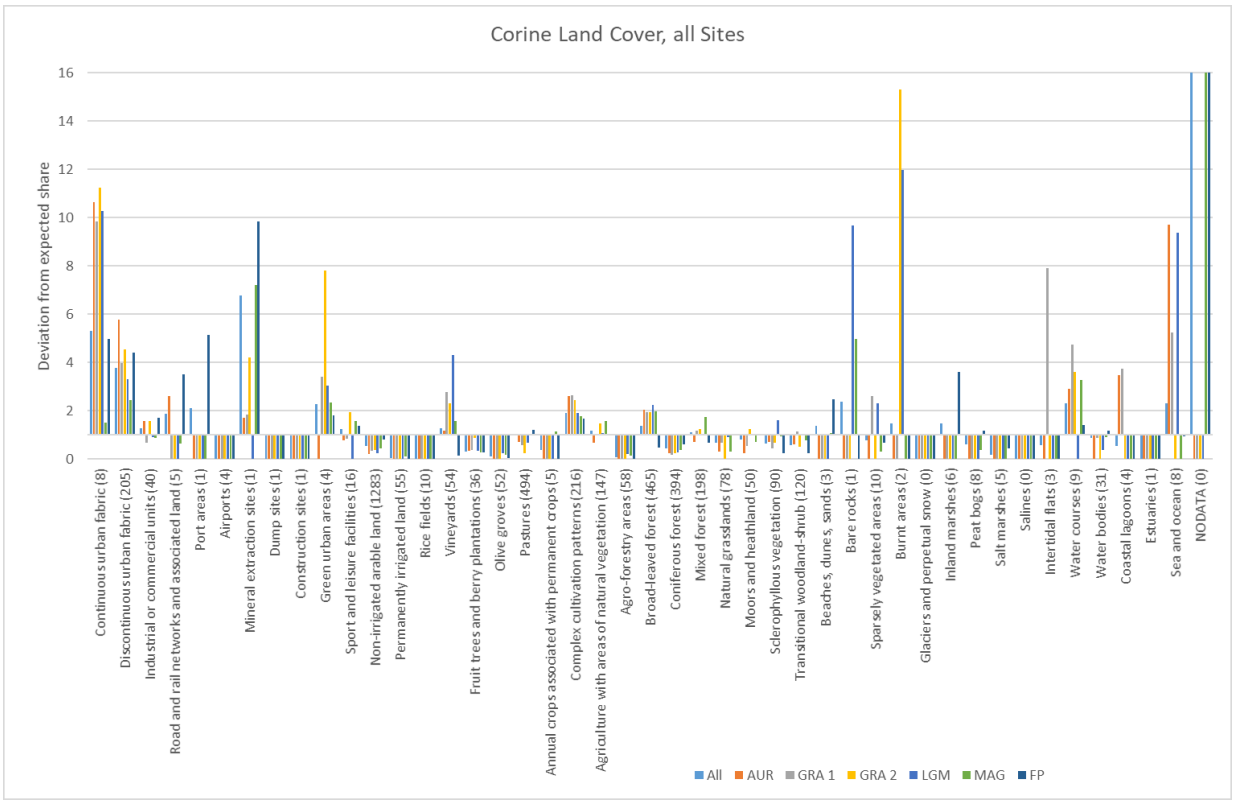



*Figure A45: Bar chart of the over- and underrepresentation. The expected values (displayed in brackets behind each variable name) represent the share of all sites that would equal the area share of each respective surface. The overrepresentation is displayed as a factor of the expected value. Sites: All, Variable: Corine Land Cover*

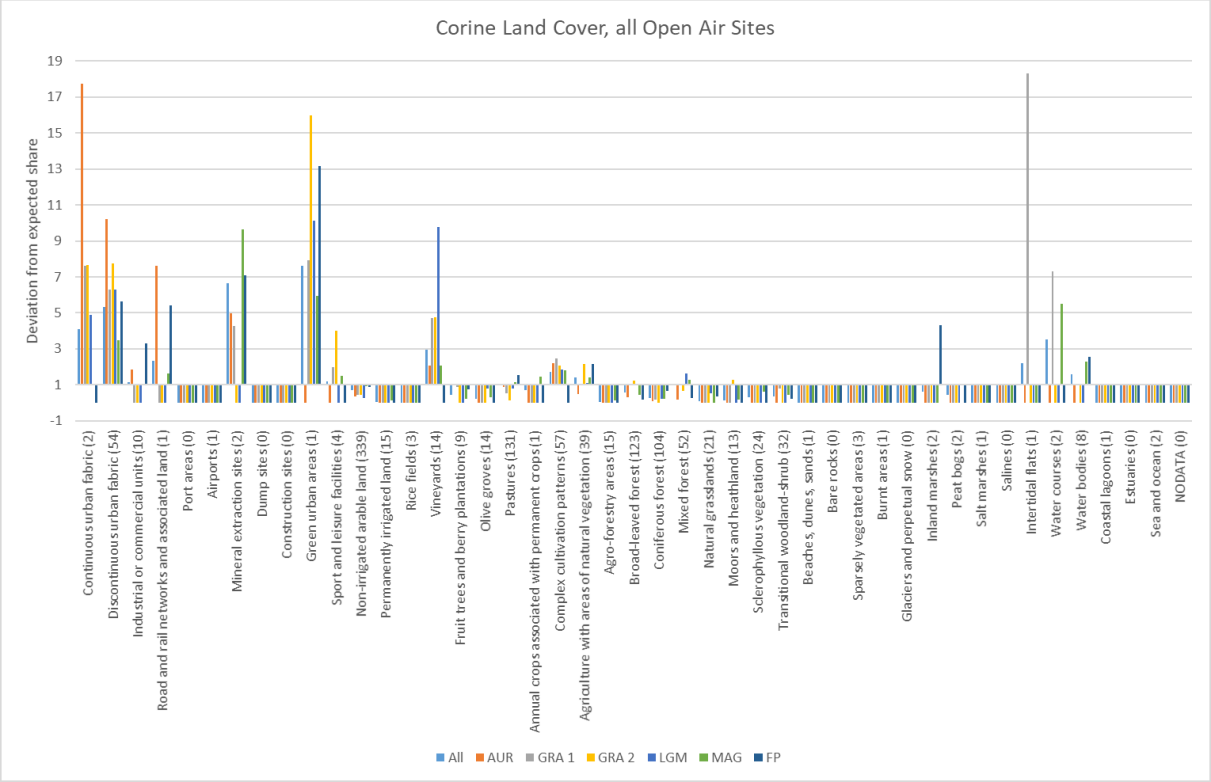



*Figure A46: Bar chart of the over- and underrepresentation. The expected values (displayed in brackets behind each variable name) represent the share of all sites that would equal the area share of each respective surface. The overrepresentation is displayed as a factor of the expected value. Sites: All open air, Variable: Corine Land Cover*

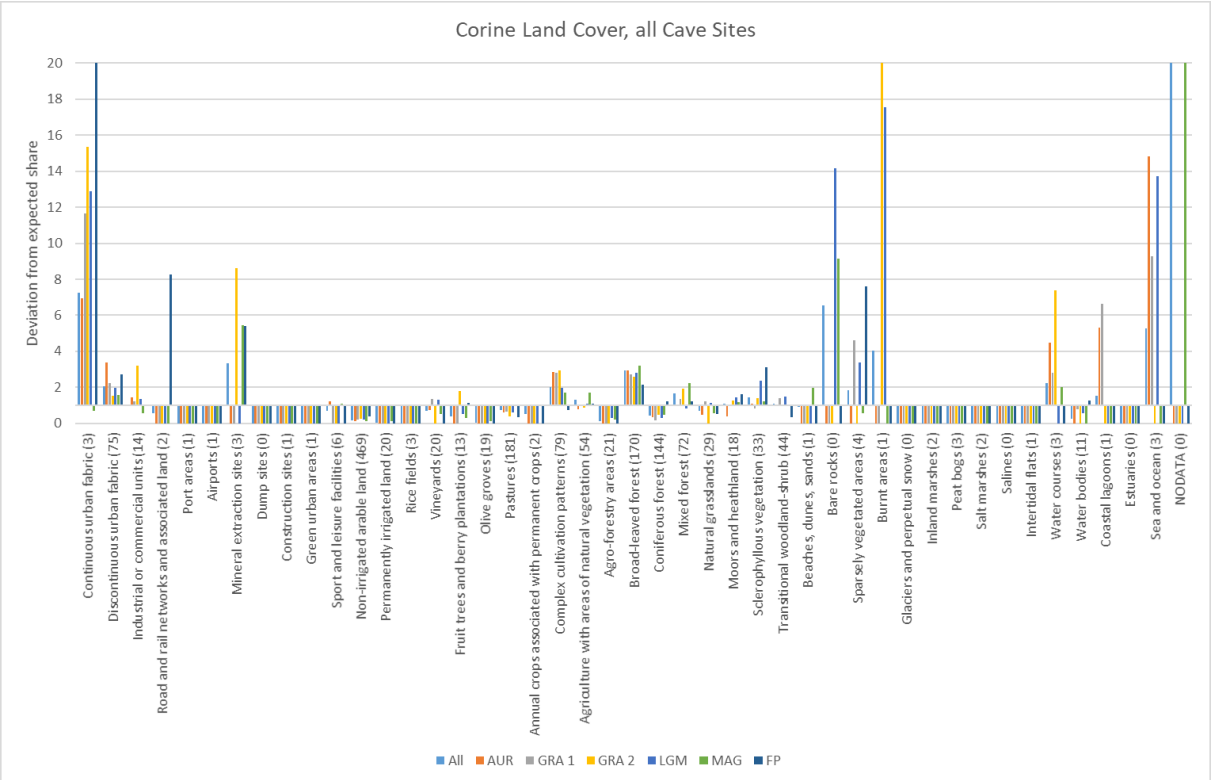



*Figure A47: Bar chart of the over- and underrepresentation. The expected values (displayed in brackets behind each variable name) represent the share of all sites that would equal the area share of each respective surface. The overrepresentation is displayed as a factor of the expected value. Sites: All cave, Variable: Corine Land Cover*

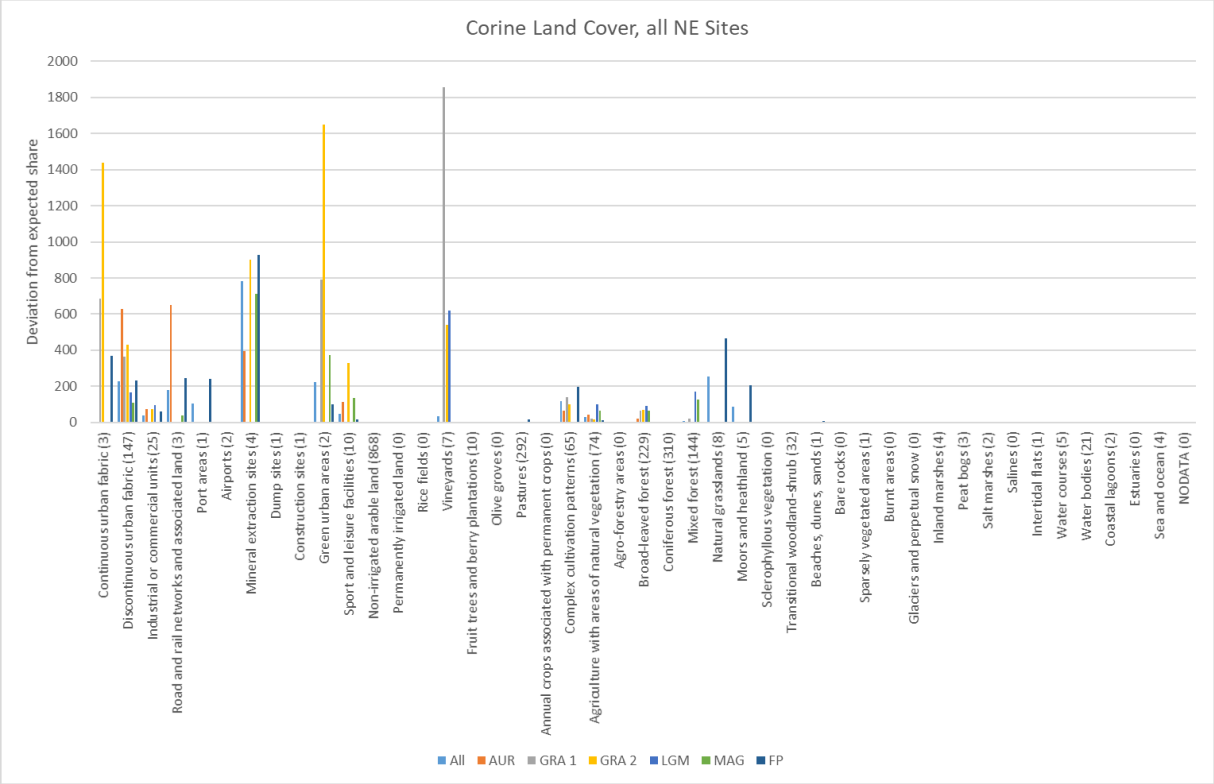



*Figure A48: Bar chart of the over- and underrepresentation. The expected values (displayed in brackets behind each variable name) represent the share of all sites that would equal the area share of each respective surface. The overrepresentation is displayed as a factor of the expected value. Sites: NE, Variable: Corine Land Cover*

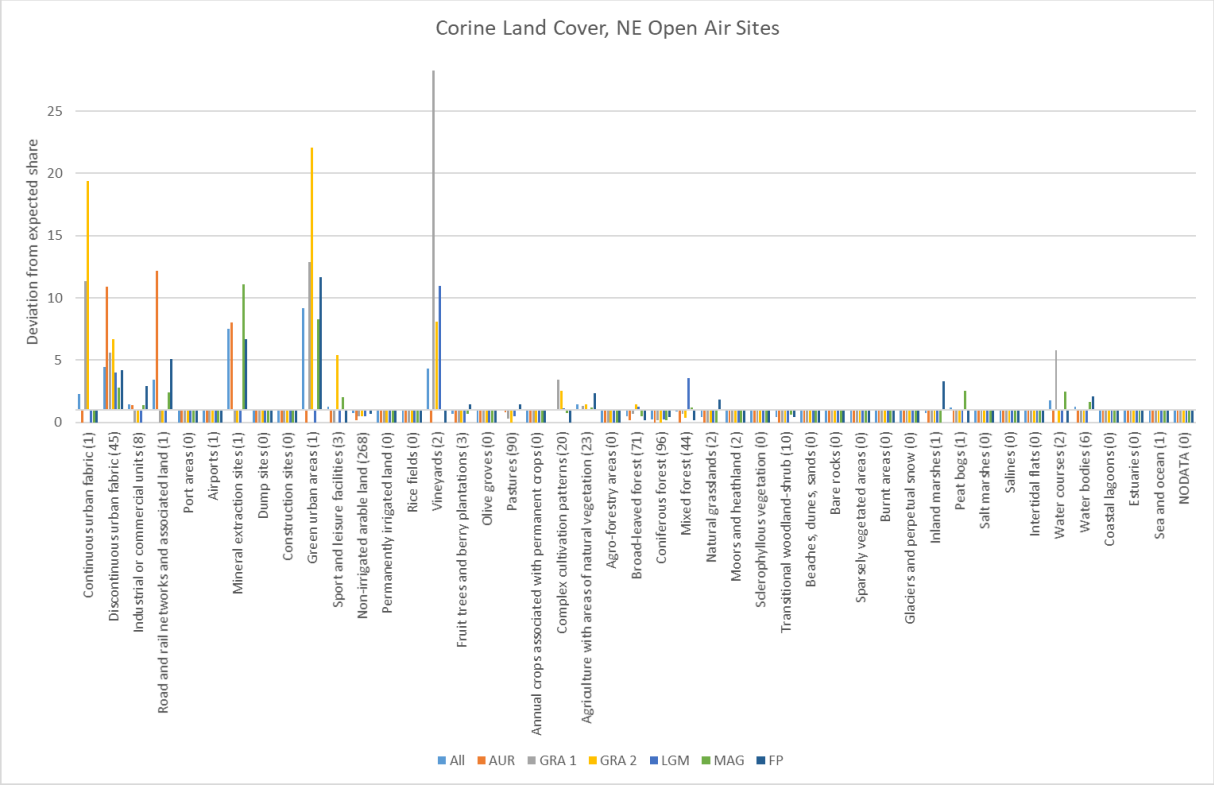



*Figure A49: Bar chart of the over- and underrepresentation. The expected values (displayed in brackets behind each variable name) represent the share of all sites that would equal the area share of each respective surface. The overrepresentation is displayed as a factor of the expected value. Sites: NE open air, Variable: Corine Land Cover*

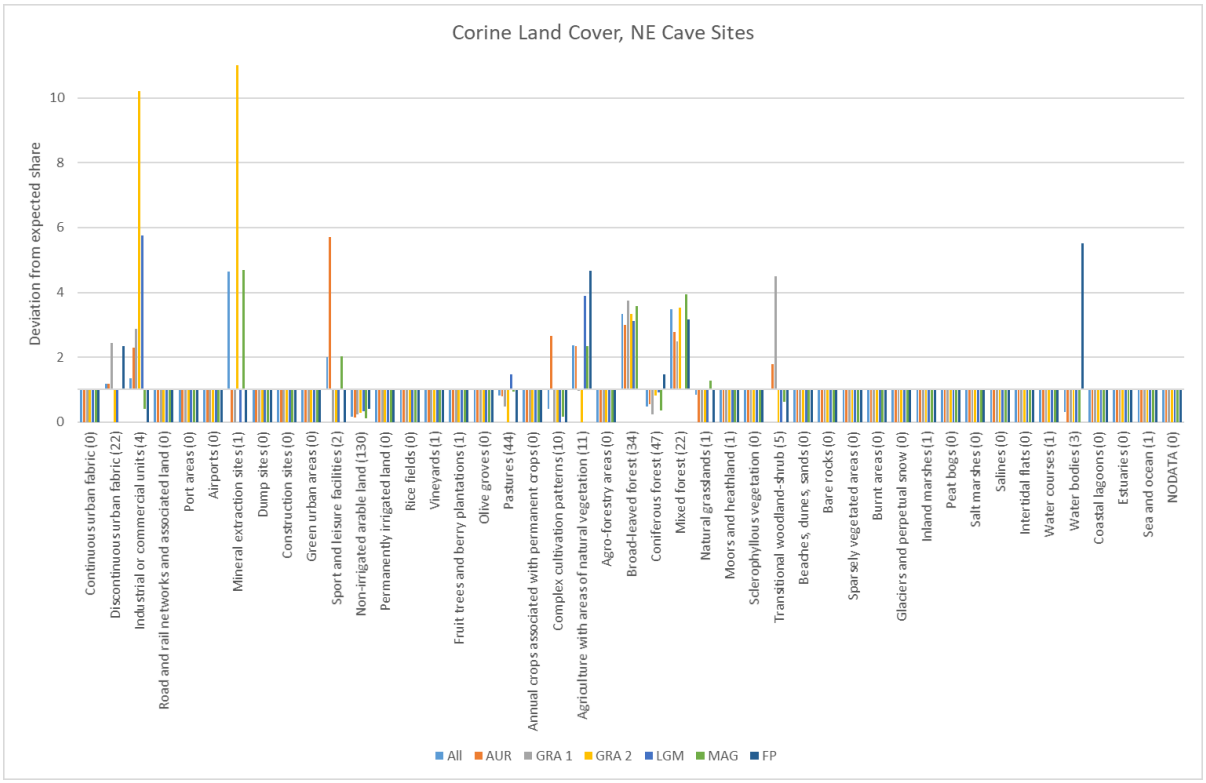



*Figure A50: Bar chart of the over- and underrepresentation. The expected values (displayed in brackets behind each variable name) represent the share of all sites that would equal the area share of each respective surface. The overrepresentation is displayed as a factor of the expected value. Sites: NE cave, Variable: Corine Land Cover*

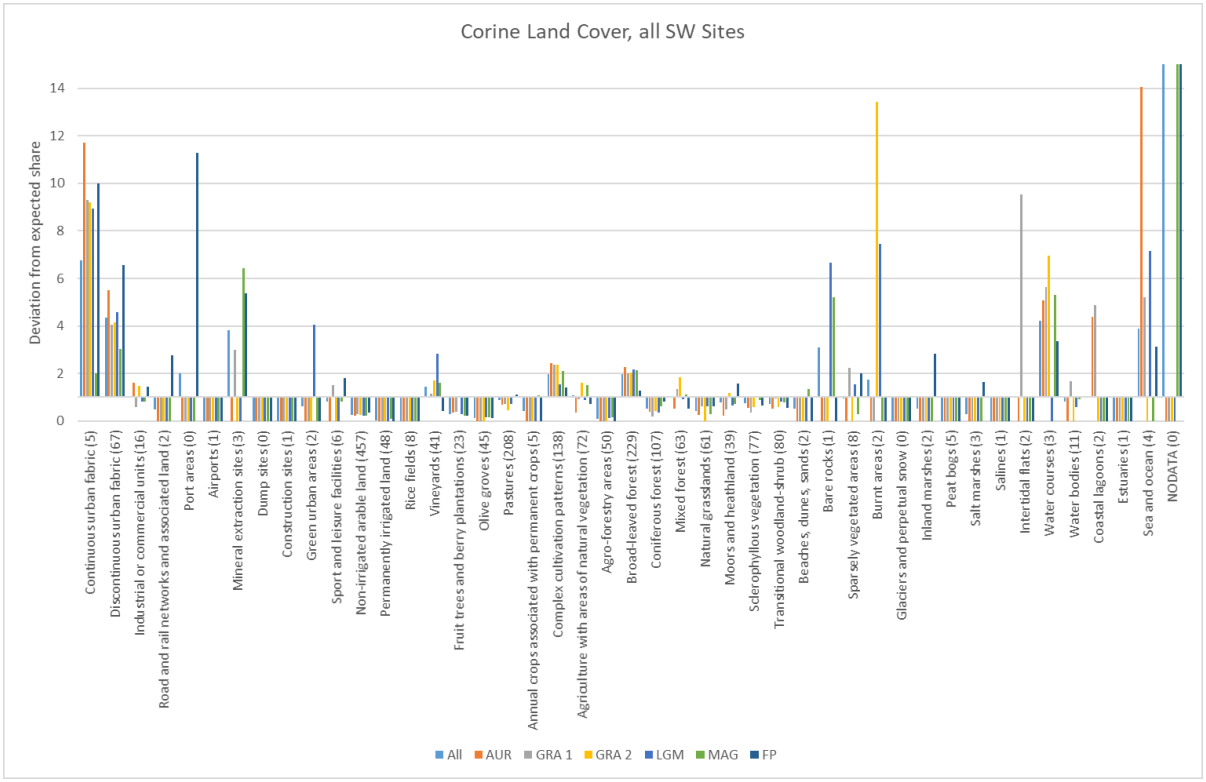



*Figure A51: Bar chart of the over- and underrepresentation. The expected values (displayed in brackets behind each variable name) represent the share of all sites that would equal the area share of each respective surface. The overrepresentation is displayed as a factor of the expected value. Sites: All SW, Variable: Corine Land Cover*

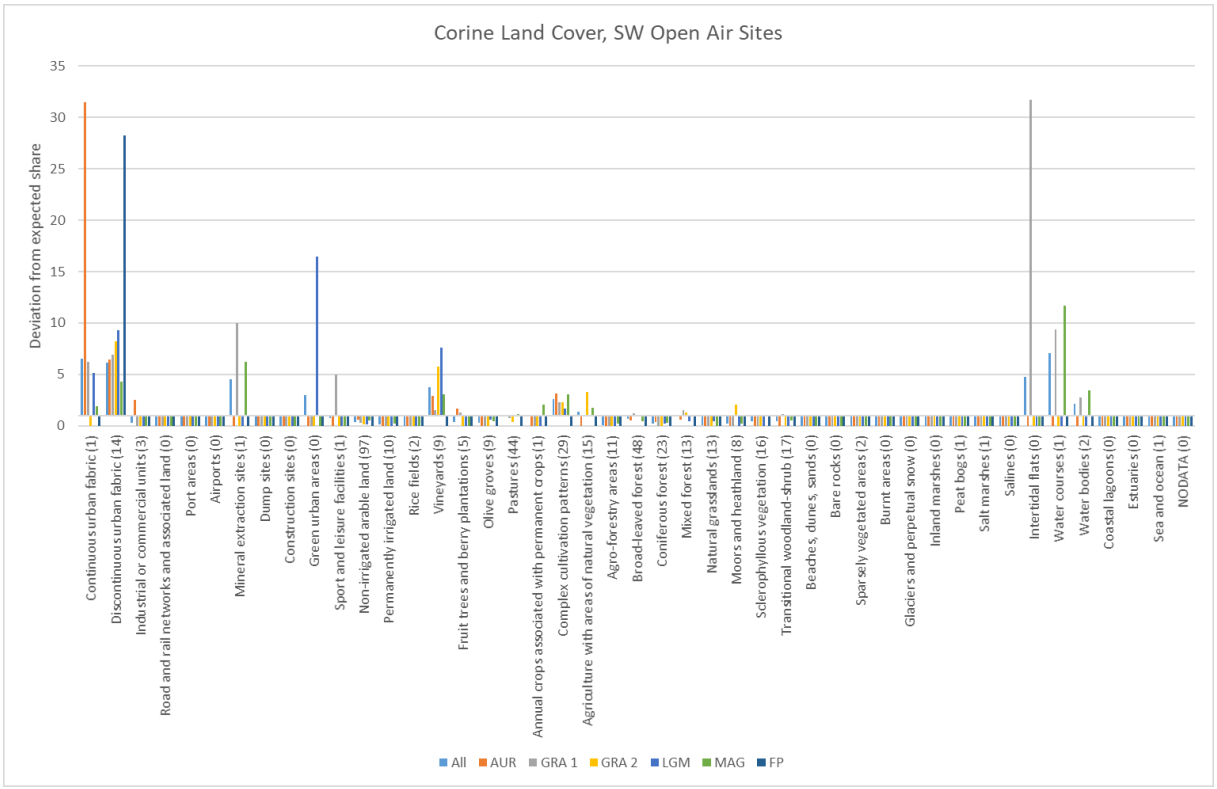



*Figure A52: Bar chart of the over- and underrepresentation. The expected values (displayed in brackets behind each variable name) represent the share of all sites that would equal the area share of each respective surface. The overrepresentation is displayed as a factor of the expected value. Sites: SW open air, Variable: Corine Land Cover*

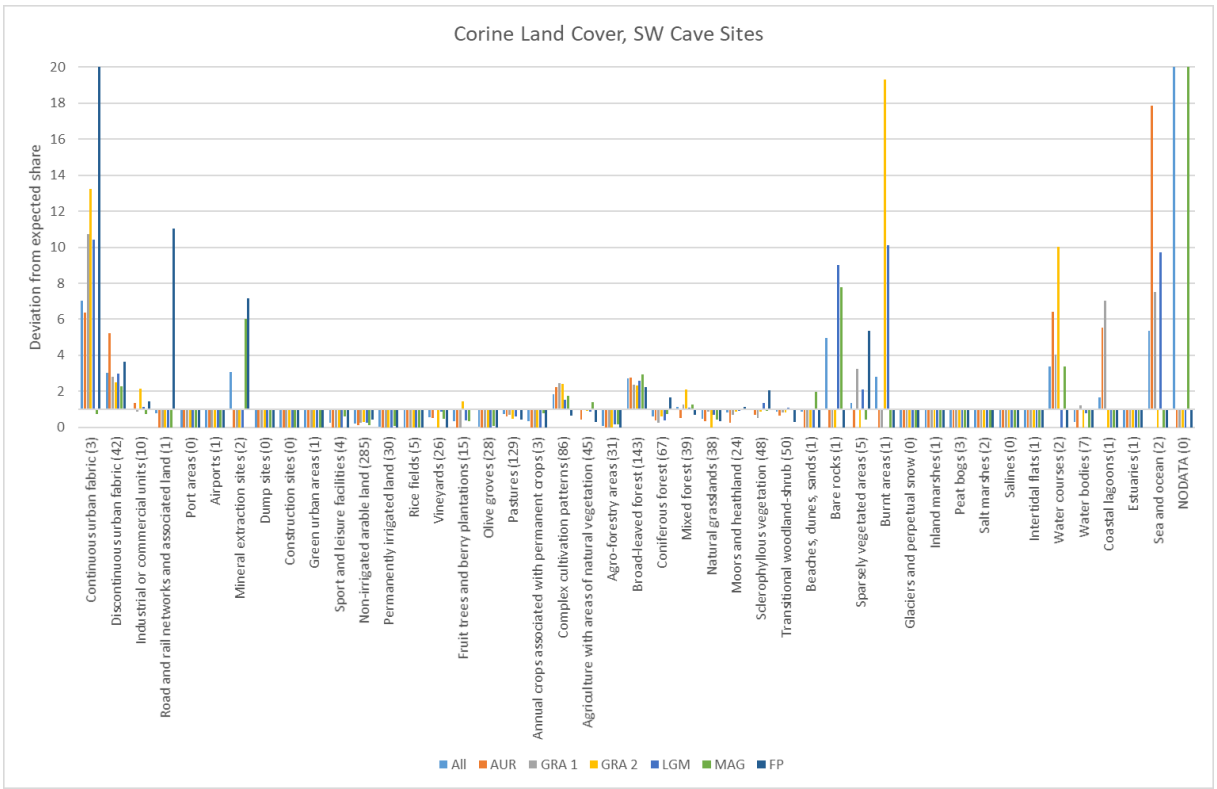



*Figure A53: Bar chart of the over- and underrepresentation. The expected values (displayed in brackets behind each variable name) represent the share of all sites that would equal the area share of each respective surface. The overrepresentation is displayed as a factor of the expected value. Sites: SW cave, Variable: Corine Land Cover*

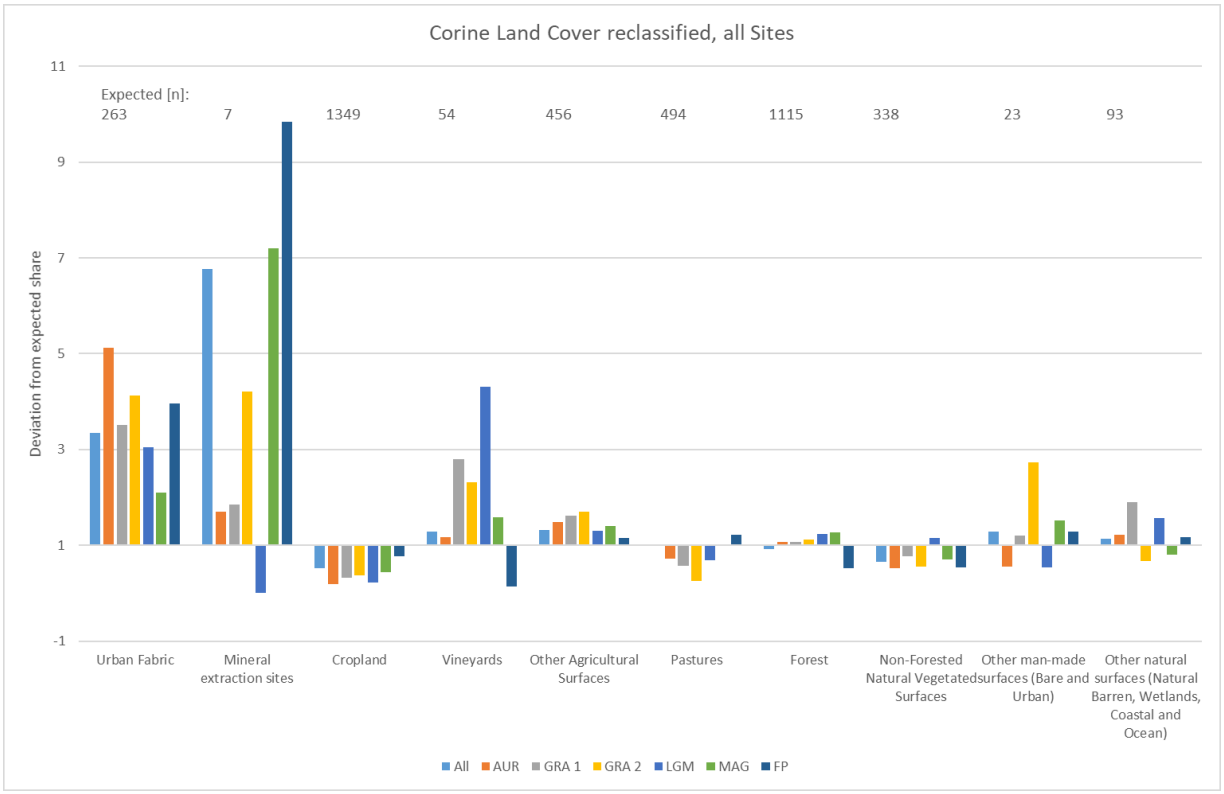



*Figure A54: Bar chart of the over- and underrepresentation. The expected values represent the share of all sites that would equal the area share of each respective surface. The overrepresentation is displayed as a factor of the expected value. Sites: All, Variable: Aggregated Corine Land Cover*

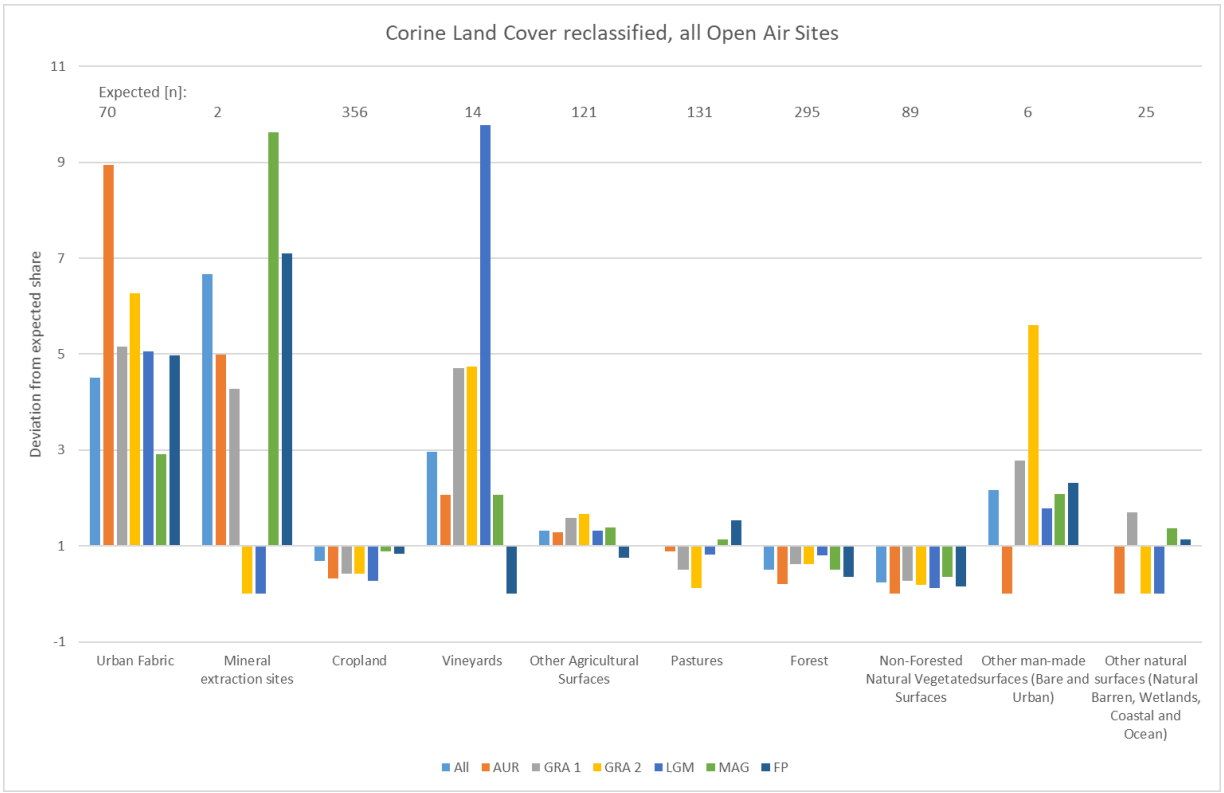



*Figure A55: Bar chart of the over- and underrepresentation. The expected values represent the share of all sites that would equal the area share of each respective surface. The overrepresentation is displayed as a factor of the expected value. Sites: All open air, Variable: Aggregated Corine Land Cover*

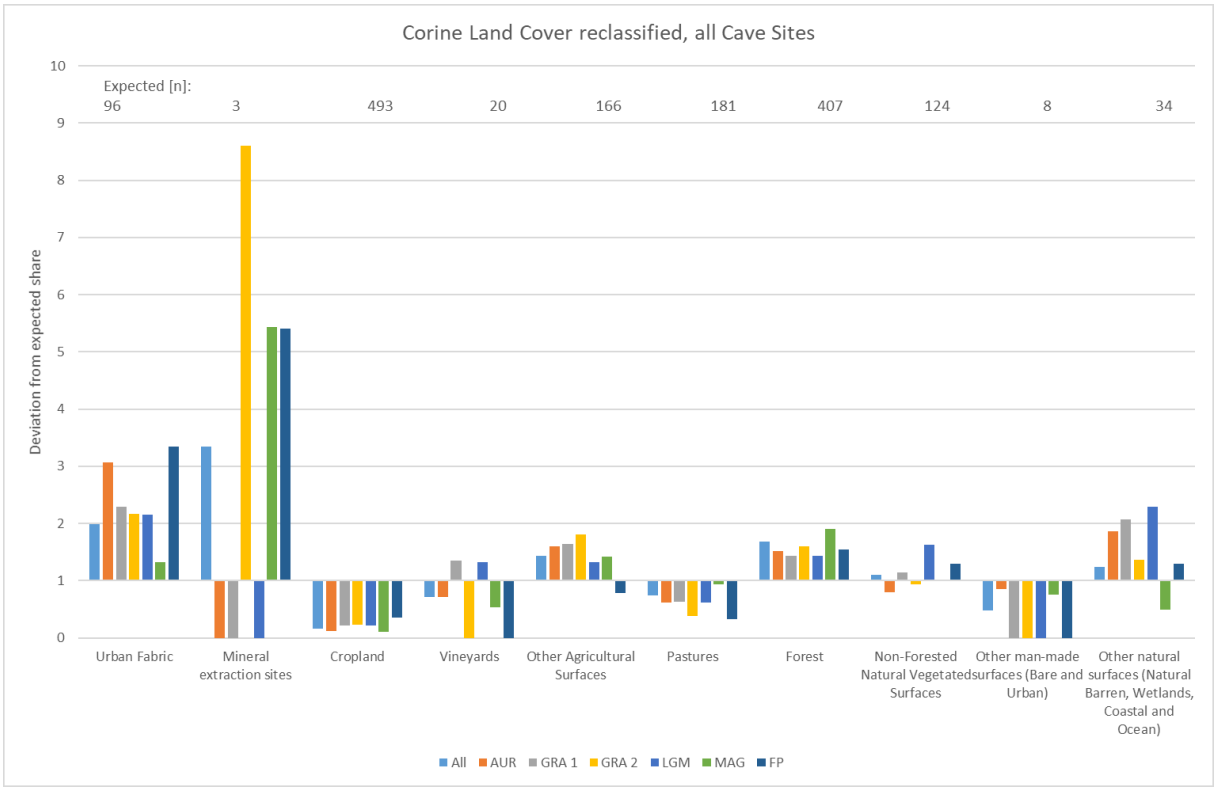



*Figure A56: Bar chart of the over- and underrepresentation. The expected values represent the share of all sites that would equal the area share of each respective surface. The overrepresentation is displayed as a factor of the expected value. Sites: All cave, Variable: Aggregated Corine Land Cover*


*Figure A57: Bar chart of the over- and underrepresentation. The expected values represent the share of all sites that would equal the area share of each respective surface. The overrepresentation is displayed as a factor of the expected value. Sites: NE, Variable: Aggregated Corine Land Cover*

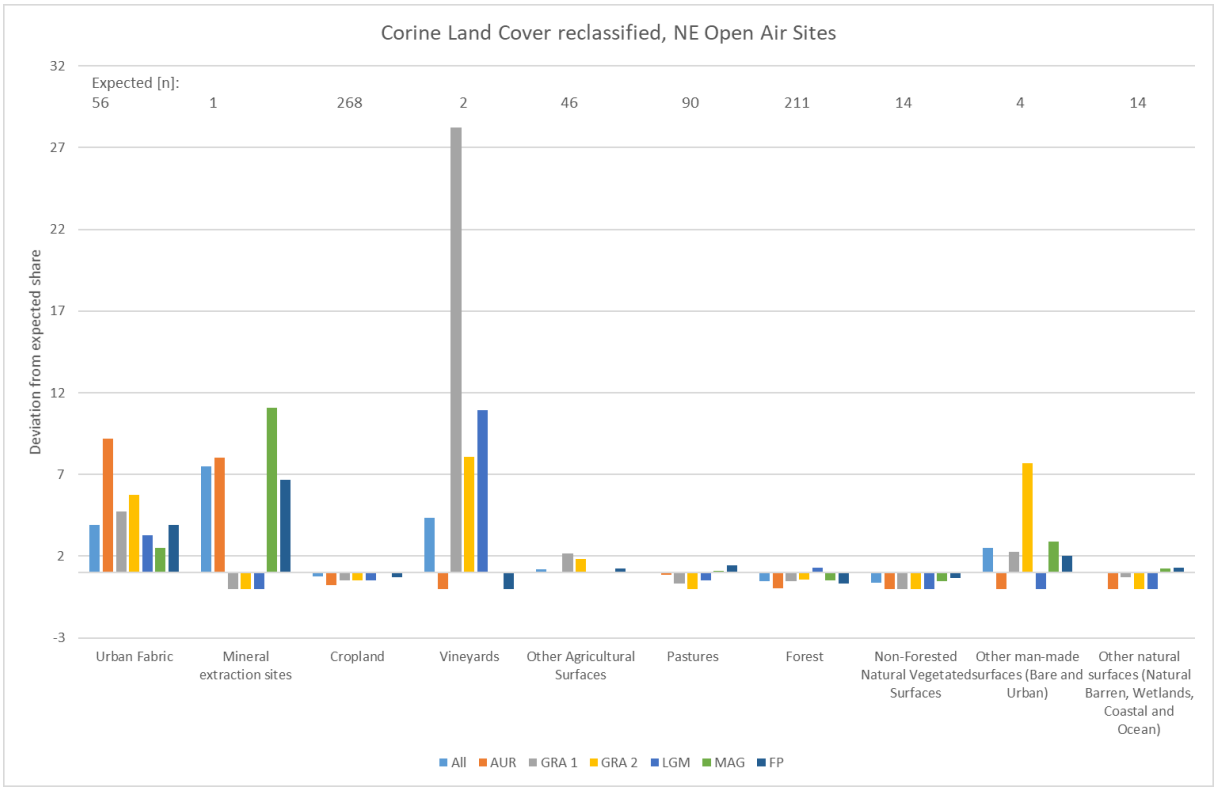



*Figure A58: Bar chart of the over- and underrepresentation. The expected values represent the share of all sites that would equal the area share of each respective surface. The overrepresentation is displayed as a factor of the expected value. Sites: NE open air, Variable: Aggregated Corine Land Cover*

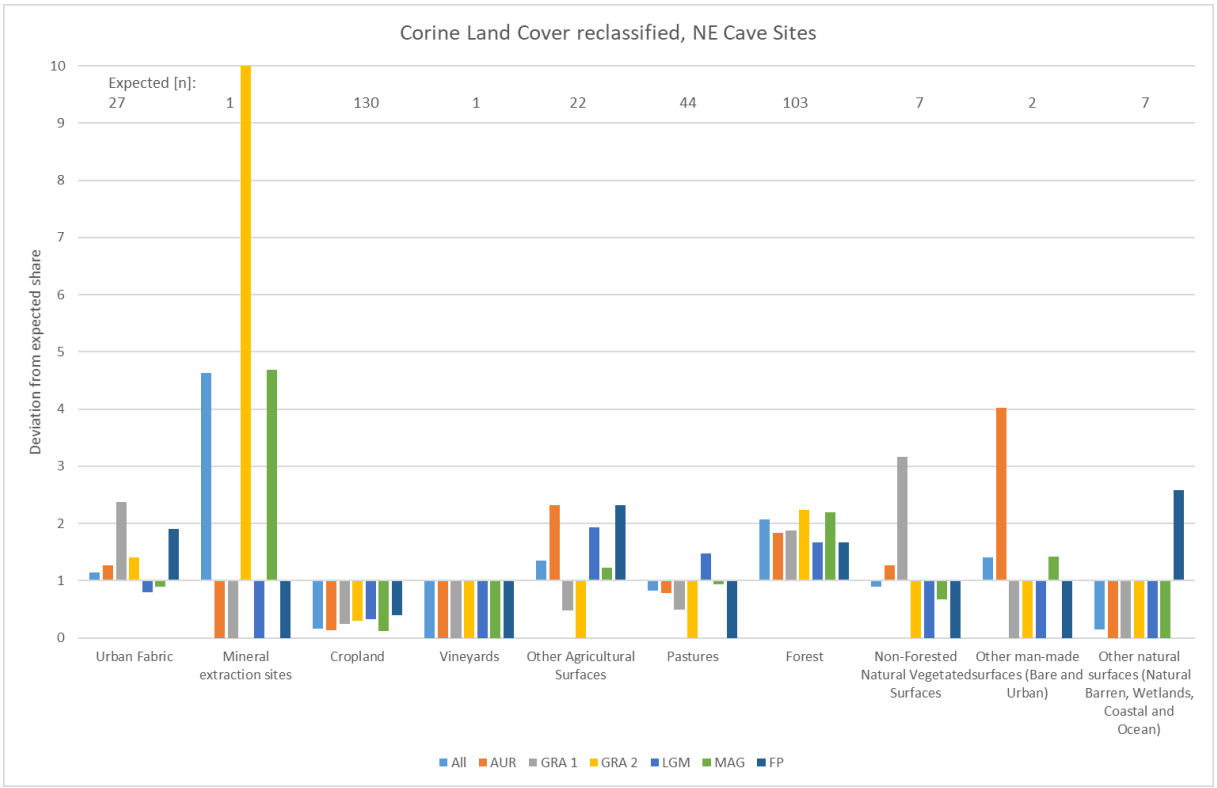



*Figure A59: Bar chart of the over- and underrepresentation. The expected values represent the share of all sites that would equal the area share of each respective surface. The overrepresentation is displayed as a factor of the expected value. Sites: NE cave, Variable: Aggregated Corine Land Cover*

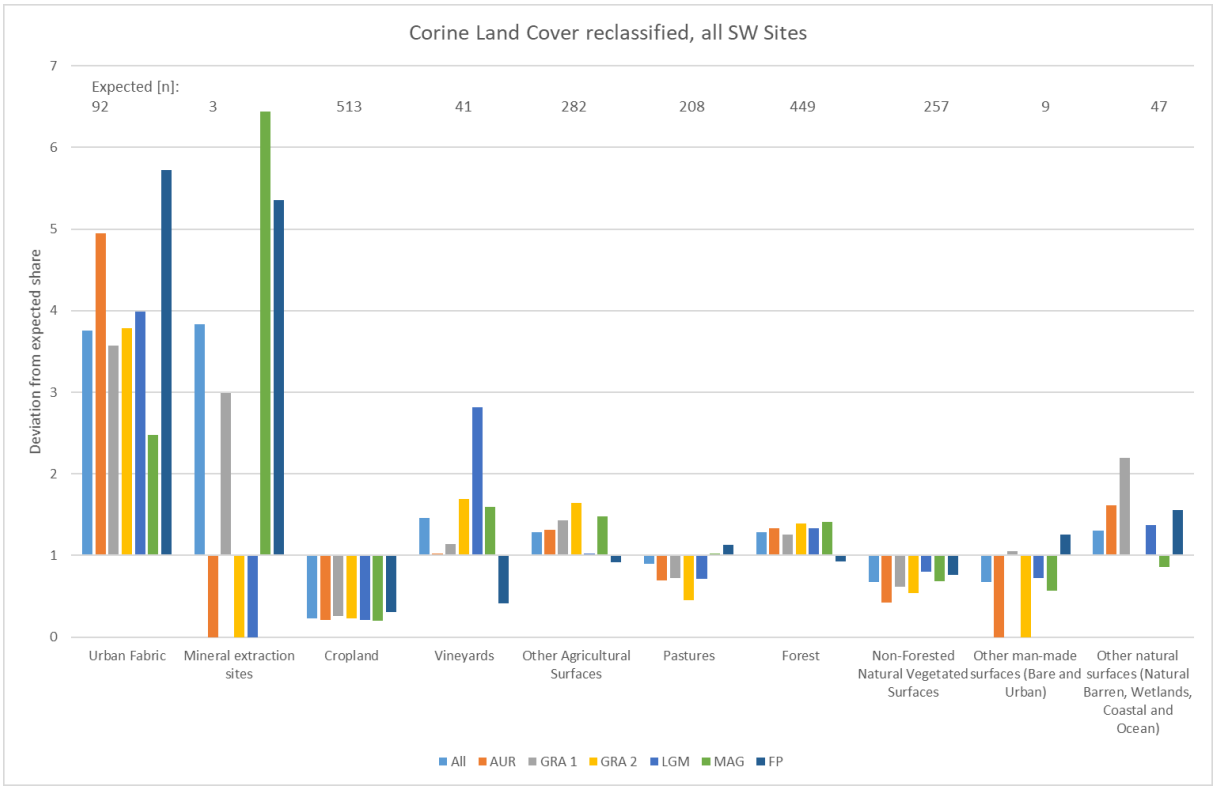



*Figure A60: Bar chart of the over- and underrepresentation. The expected values represent the share of all sites that would equal the area share of each respective surface. The overrepresentation is displayed as a factor of the expected value. Sites: SW, Variable: Aggregated Corine Land Cover*

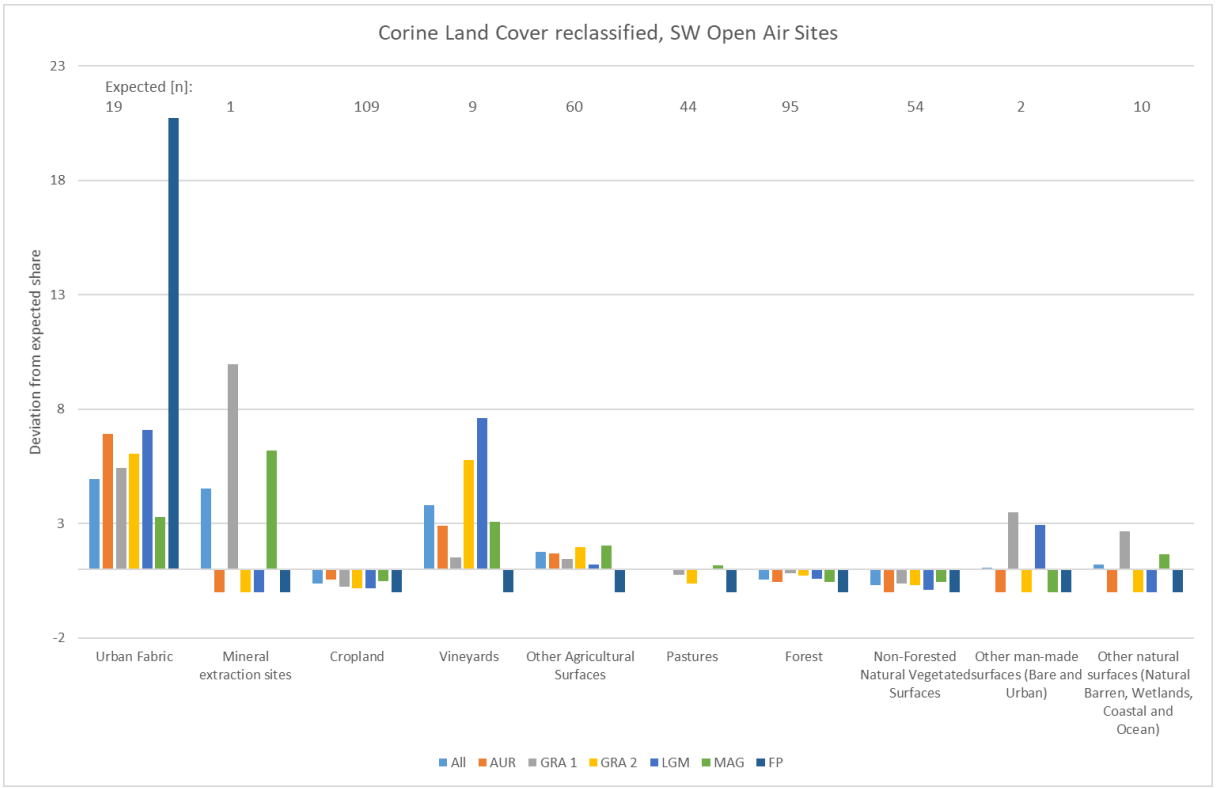



*Figure A61: Bar chart of the over- and underrepresentation. The expected values represent the share of all sites that would equal the area share of each respective surface. The overrepresentation is displayed as a factor of the expected value. Sites: SW open air, Variable: Aggregated Corine Land Cover*

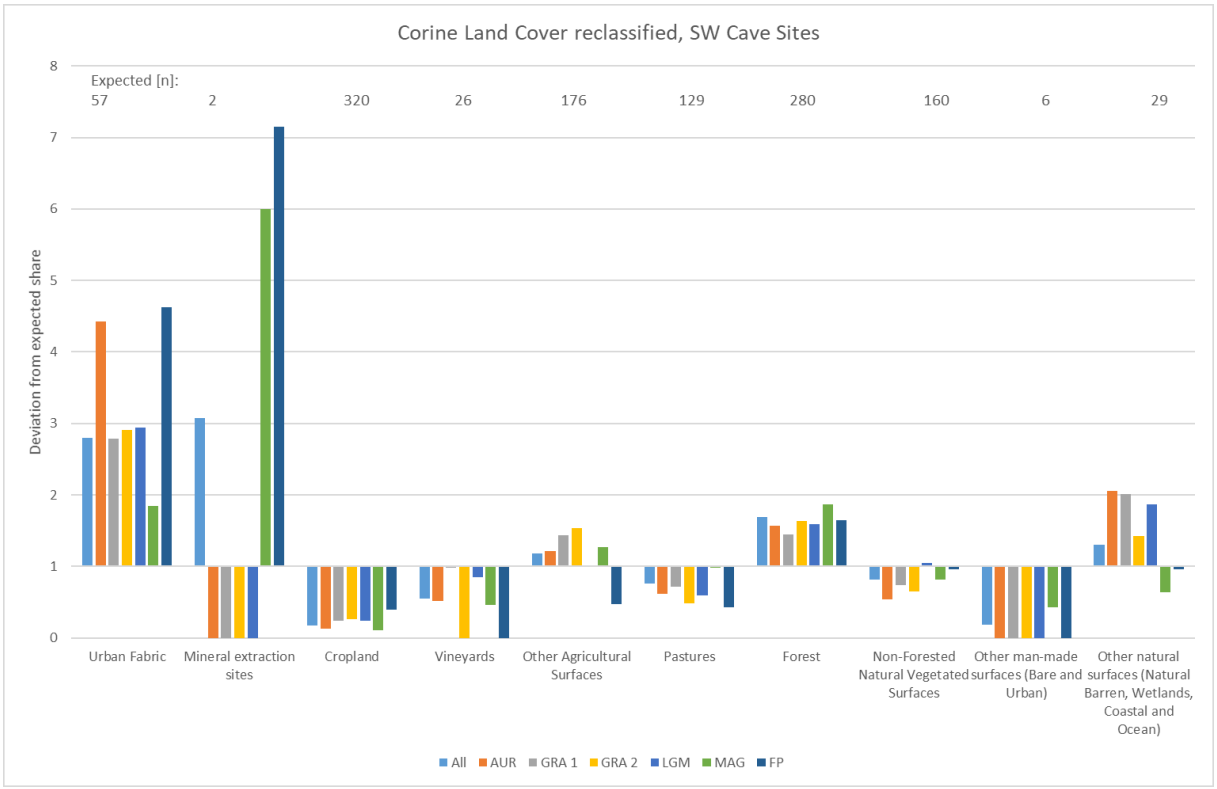



*Figure A62: Bar chart of the over- and underrepresentation. The expected values represent the share of all sites that would equal the area share of each respective surface. The overrepresentation is displayed as a factor of the expected value. Sites: SW cave, Variable: Aggregated Corine Land Cover*

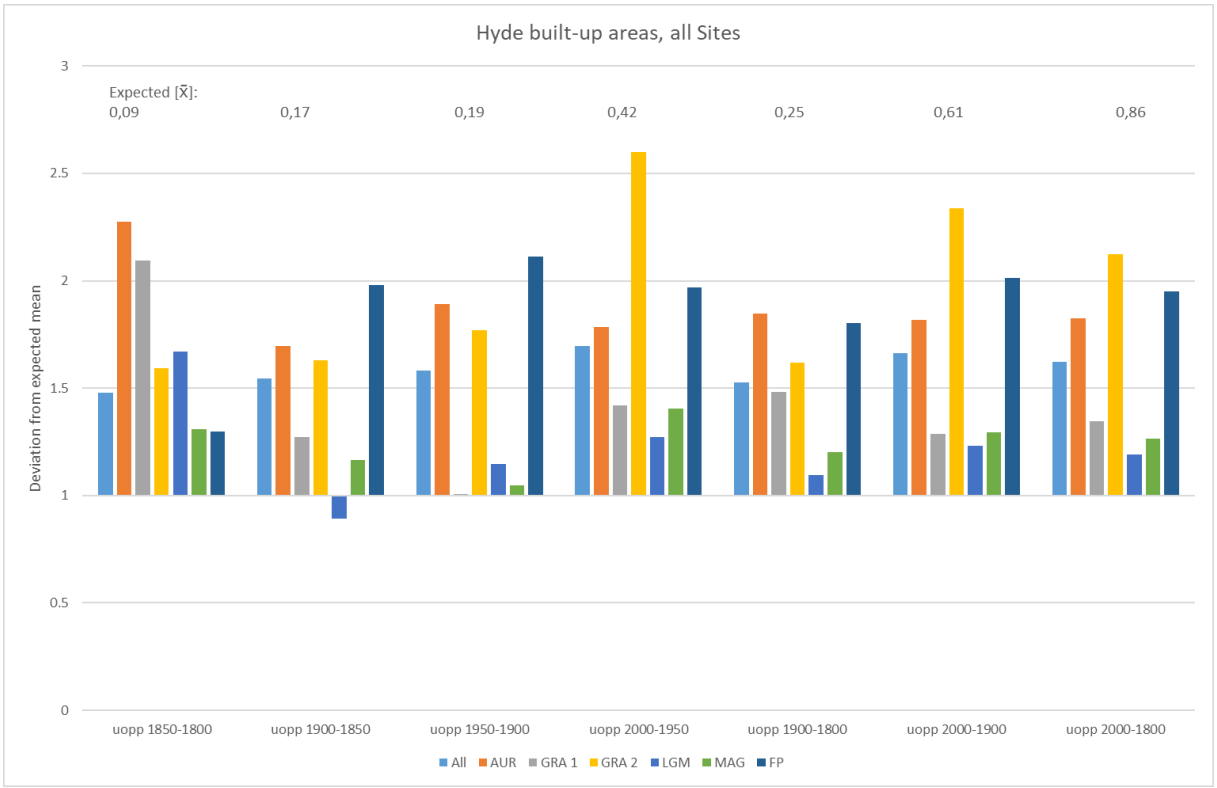



*Figure A63: Bar chart of the over- and underrepresentation. The expected values represent the mean of the environmental variable. The overrepresentation is displayed as a factor of the expected value. Sites: All, Variable: HYDE built up area difference*

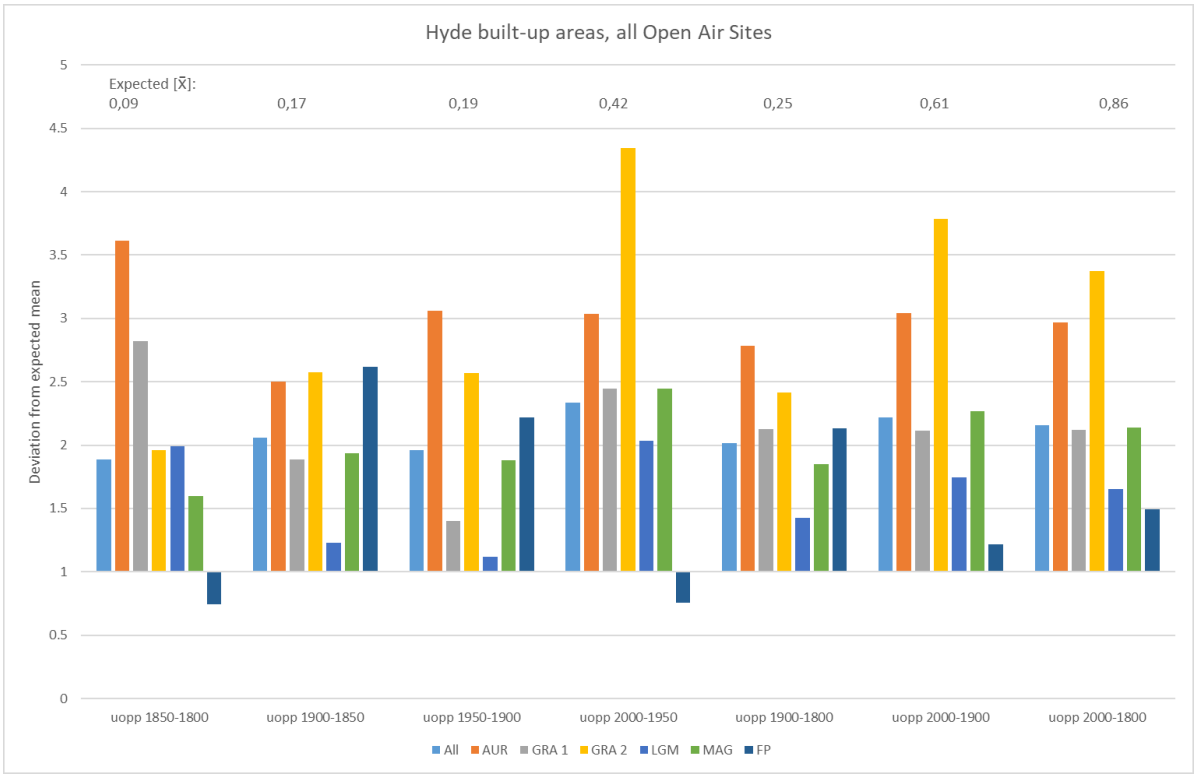



*Figure A64: Bar chart of the over- and underrepresentation. The expected values represent the mean of the environmental variable. The overrepresentation is displayed as a factor of the expected value. Sites: All open air, Variable: HYDE built up area difference*

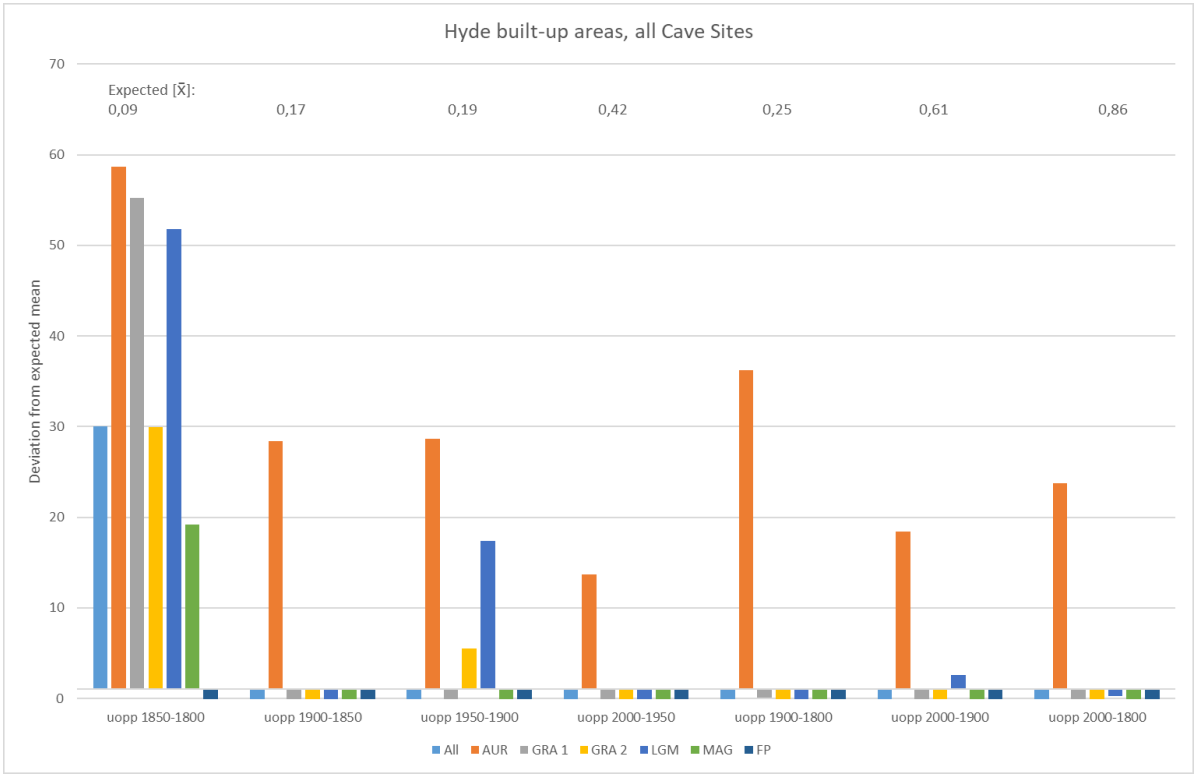



*Figure A65: Bar chart of the over- and underrepresentation. The expected values represent the mean of the environmental variable. The overrepresentation is displayed as a factor of the expected value. Sites: All cave, Variable: HYDE built up area difference*

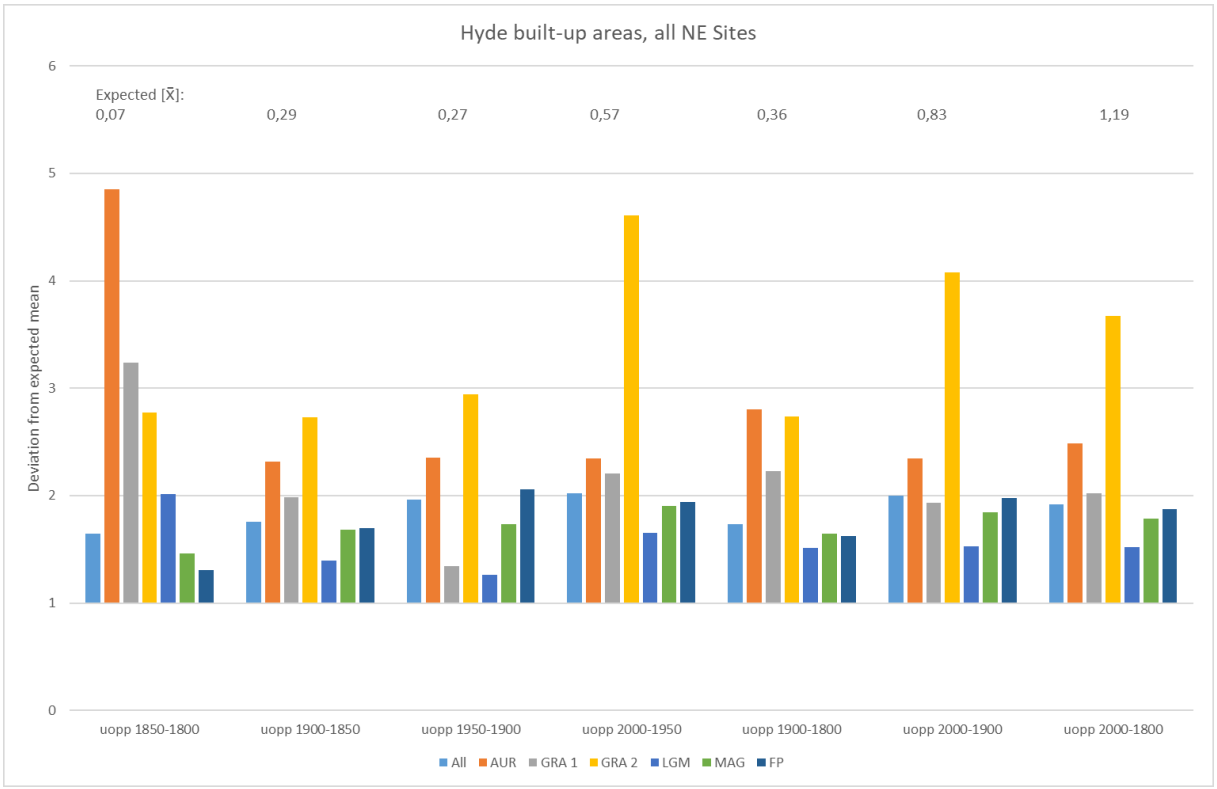



*Figure A66: Bar chart of the over- and underrepresentation. The expected values represent the mean of the environmental variable. The overrepresentation is displayed as a factor of the expected value. Sites: NE, Variable: HYDE built up area difference*

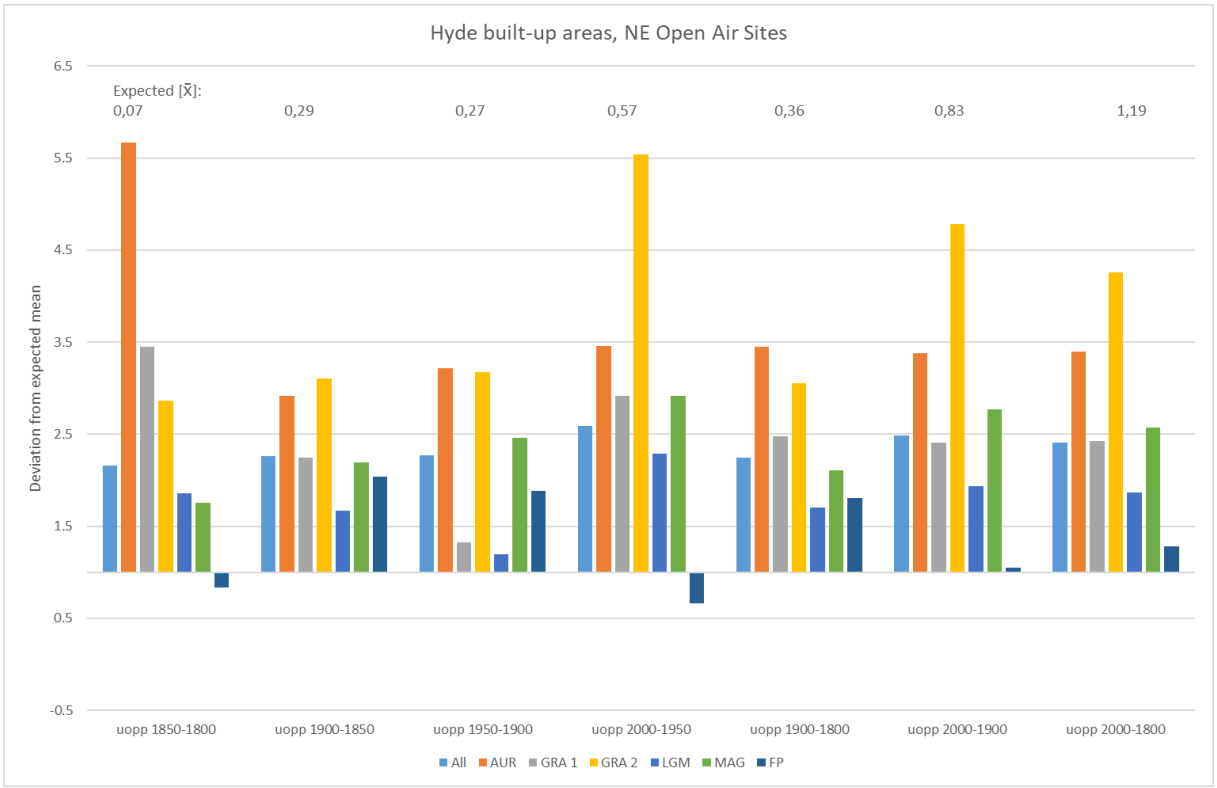



*Figure A67: Bar chart of the over- and underrepresentation. The expected values represent the mean of the environmental variable. The overrepresentation is displayed as a factor of the expected value. Sites: NE open air, Variable: HYDE built up area difference*

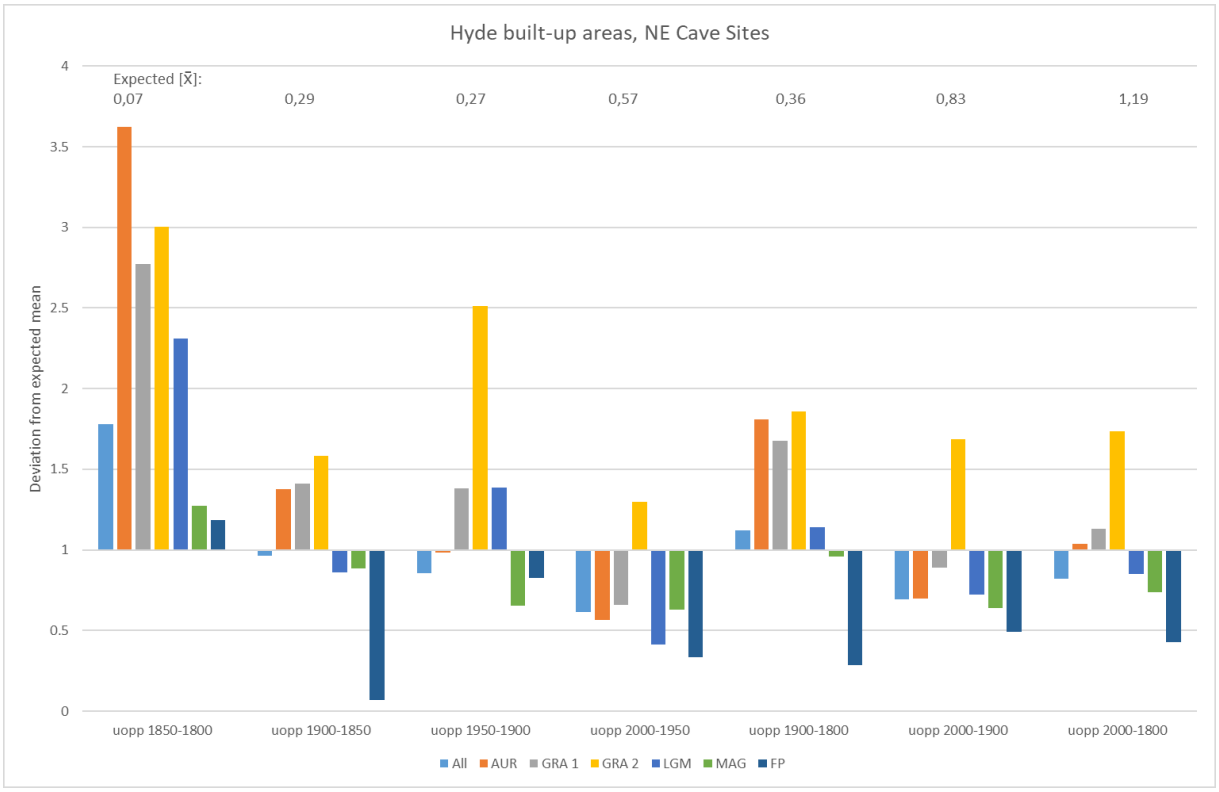



*Figure A68: Bar chart of the over- and underrepresentation. The expected values represent the mean of the environmental variable. The overrepresentation is displayed as a factor of the expected value. Sites: NE cave, Variable: HYDE built up area difference*

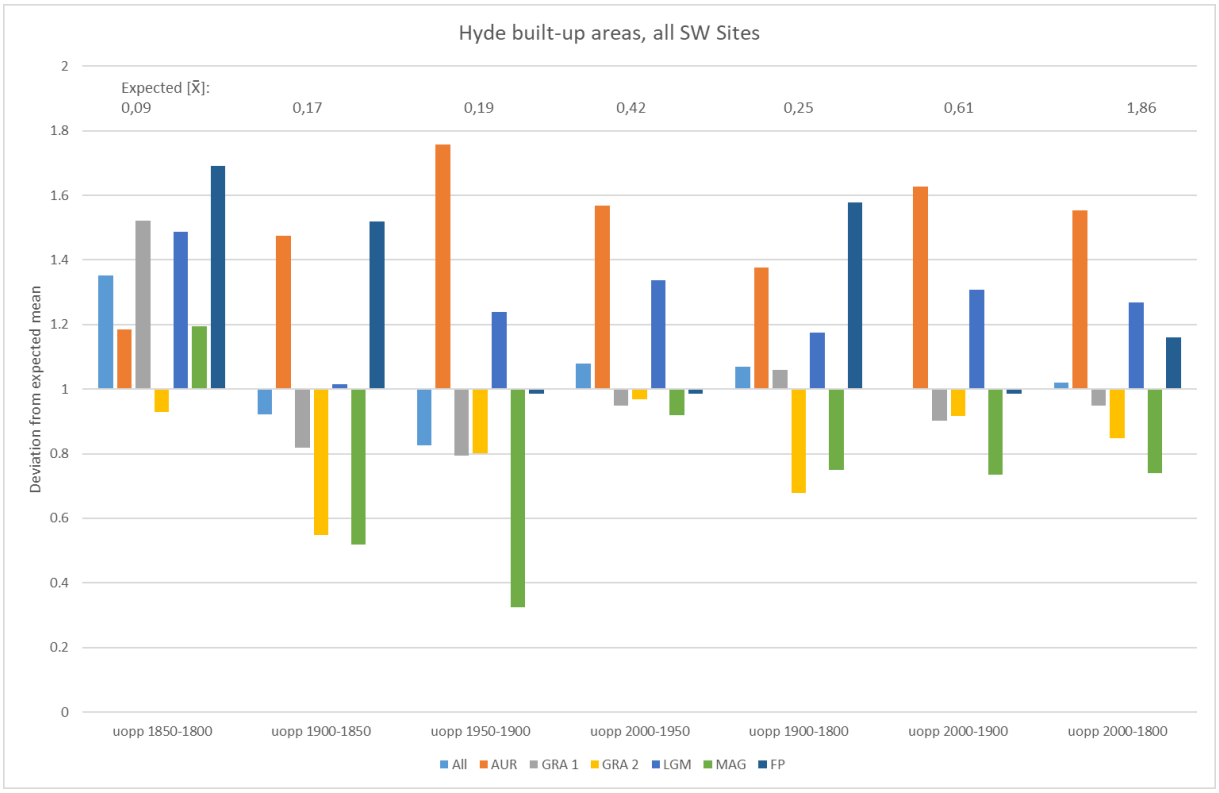



*Figure A69: Bar chart of the over- and underrepresentation. The expected values represent the mean of the environmental variable. The overrepresentation is displayed as a factor of the expected value. Sites: SW, Variable: HYDE built up area difference*

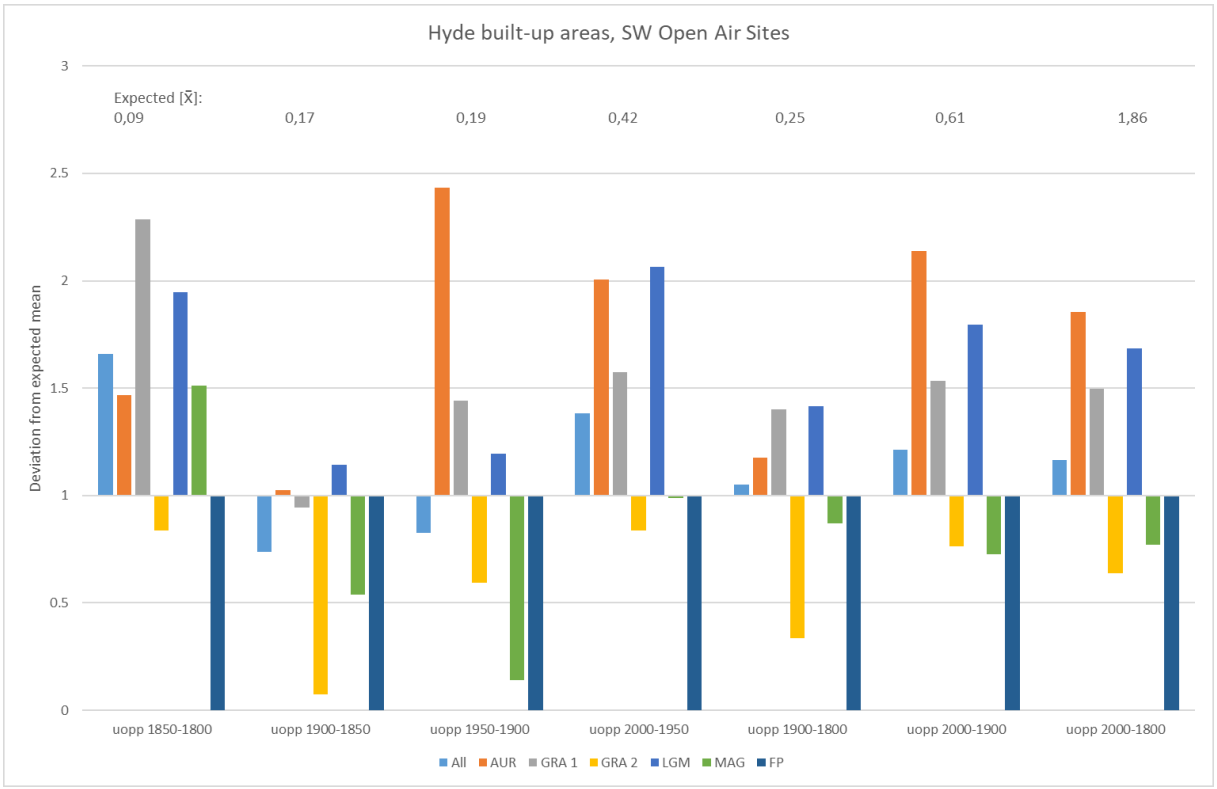



*Figure A70: Bar chart of the over- and underrepresentation. The expected values represent the mean of the environmental variable. The overrepresentation is displayed as a factor of the expected value. Sites: SW open air, Variable: HYDE built up area difference*

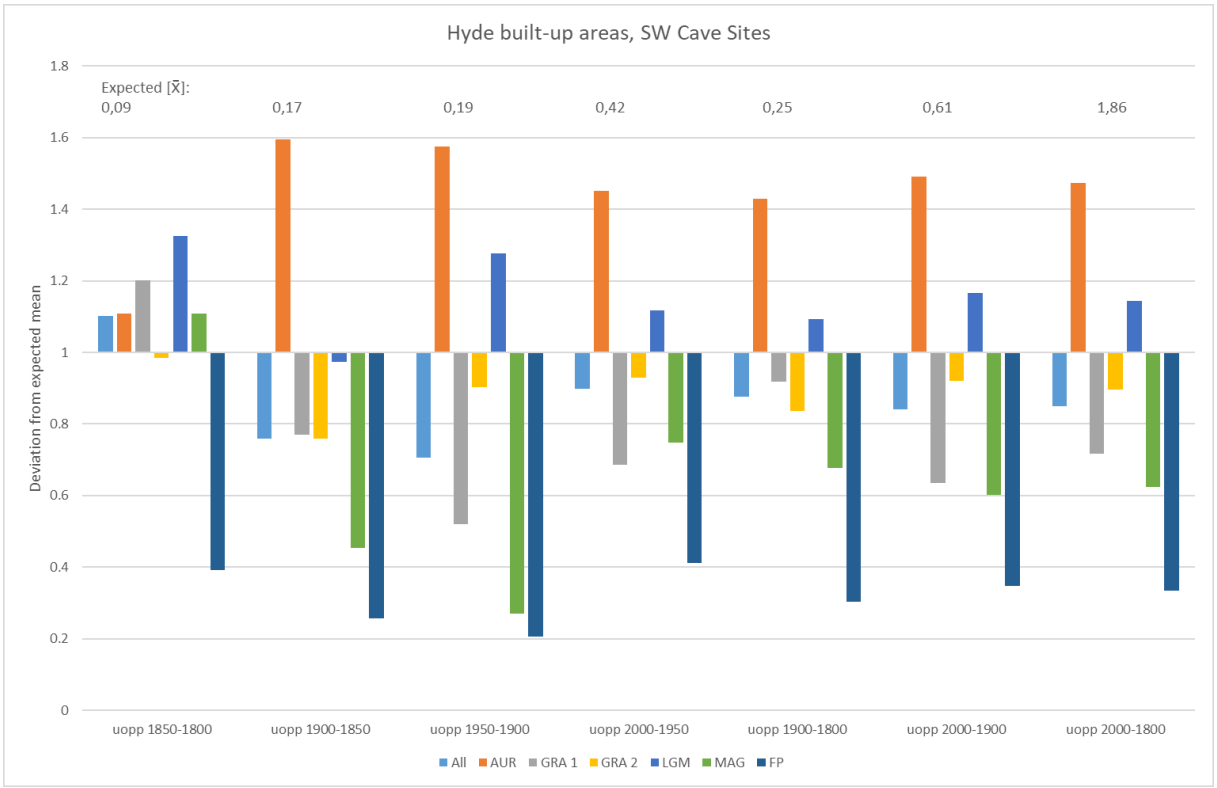



*Figure A71: Bar chart of the over- and underrepresentation. The expected values represent the mean of the environmental variable. The overrepresentation is displayed as a factor of the expected value. Sites: SW cave, Variable: HYDE built up area difference*

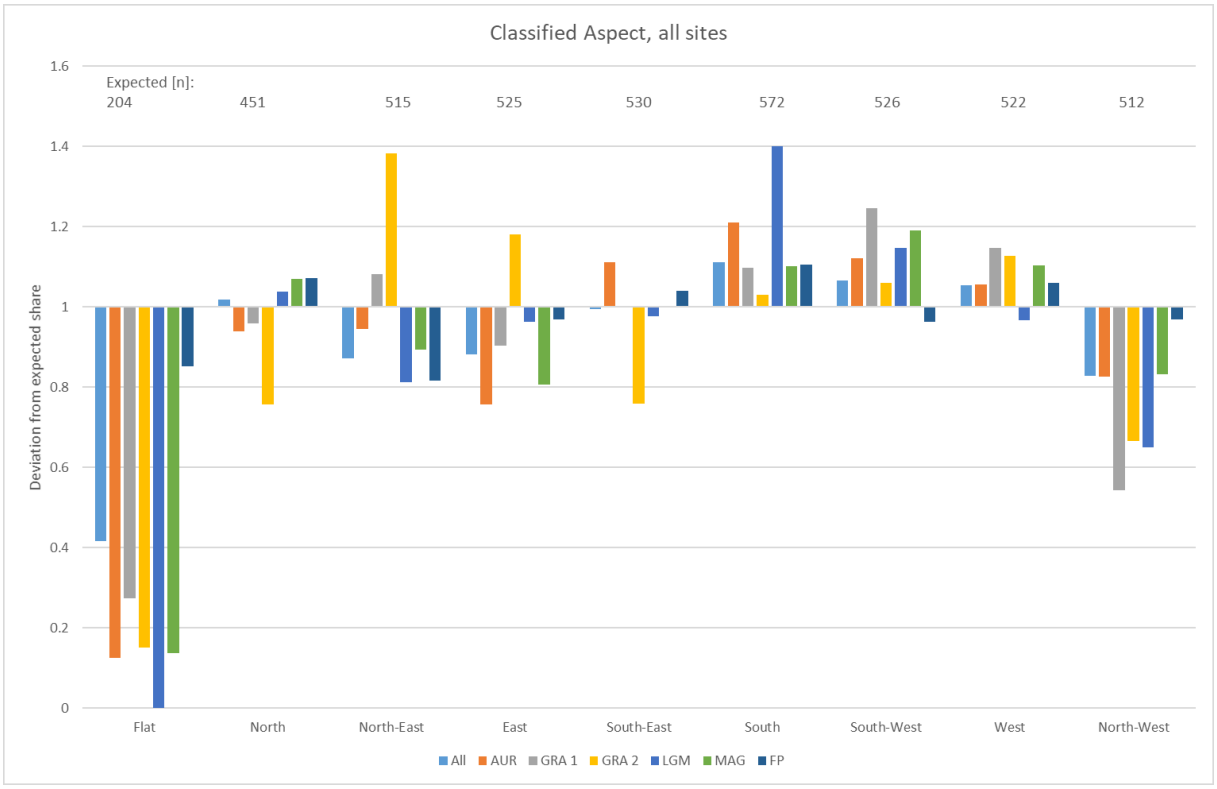



*Figure A72: Bar chart of the over- and underrepresentation. The expected values represent the share of all sites that would equal the area share of each respective surface. The overrepresentation is displayed as a factor of the expected value. Sites: All, Variable: Classified aspect*

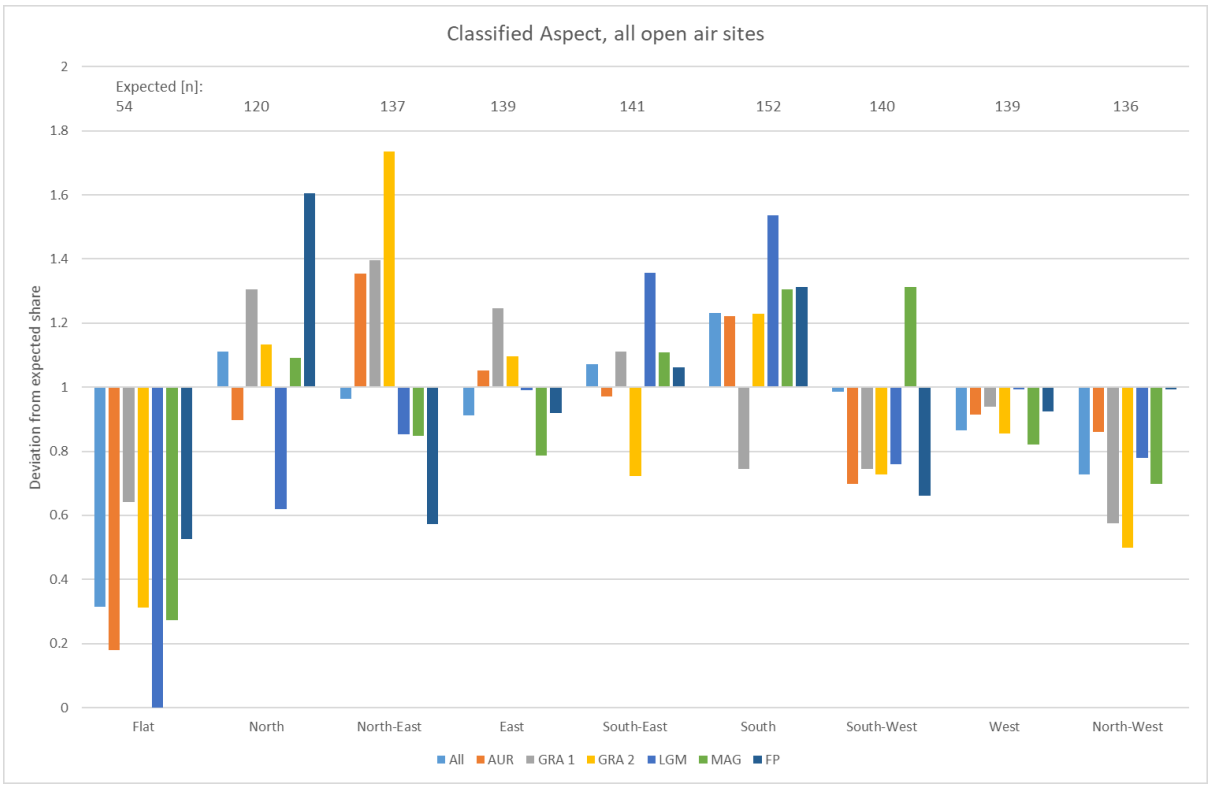



*Figure A73: Bar chart of the over- and underrepresentation. The expected values represent the share of all sites that would equal the area share of each respective surface. The overrepresentation is displayed as a factor of the expected value. Sites: All open air, Variable: Classified aspect*

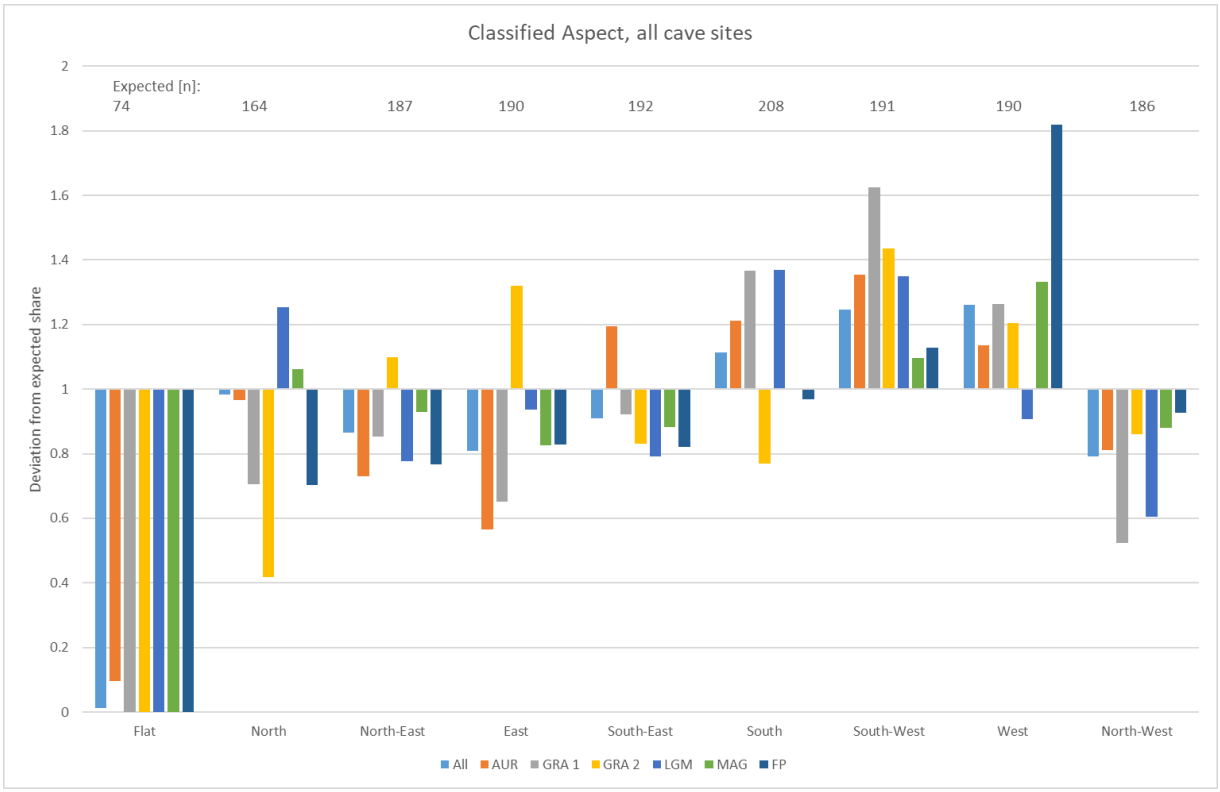



*Figure A74: Bar chart of the over- and underrepresentation. The expected values represent the share of all sites that would equal the area share of each respective surface. The overrepresentation is displayed as a factor of the expected value. Sites: All cave, Variable: Classified aspect*

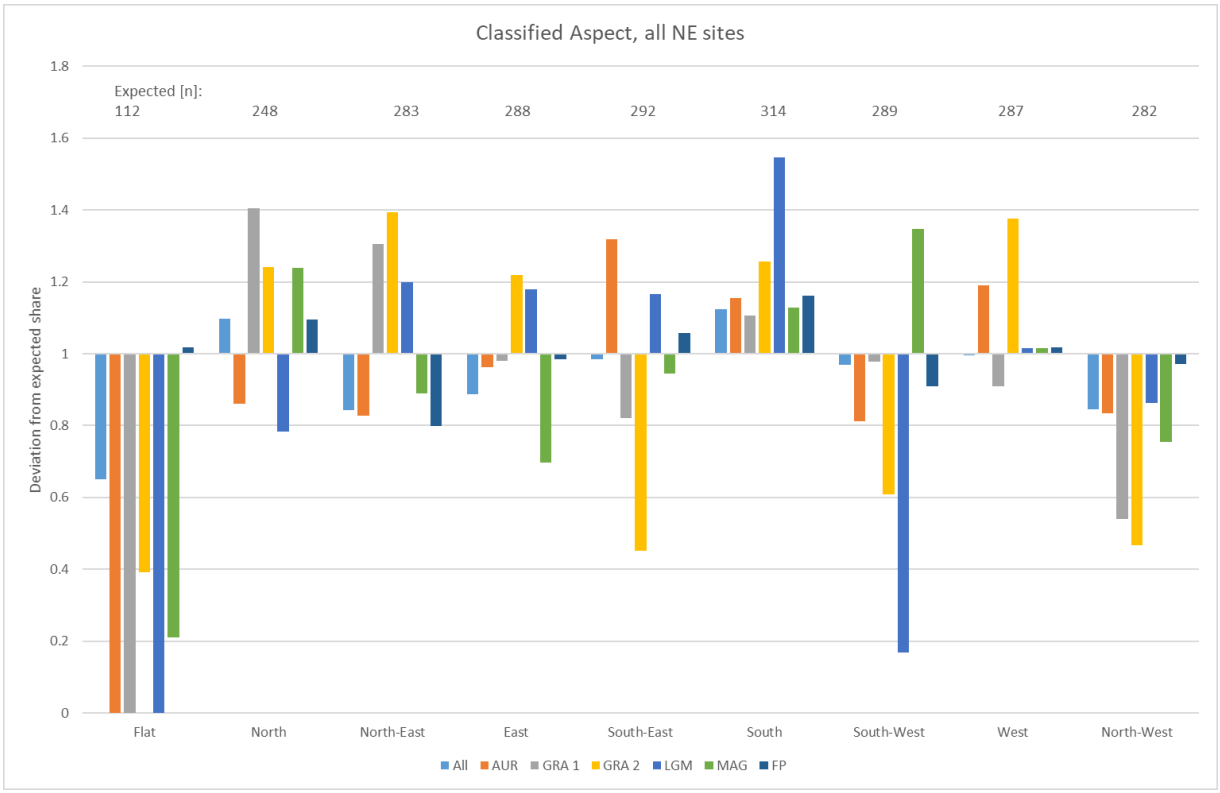



*Figure A75: Bar chart of the over- and underrepresentation. The expected values represent the share of all sites that would equal the area share of each respective surface. The overrepresentation is displayed as a factor of the expected value. Sites: NE, Variable: Classified aspect*

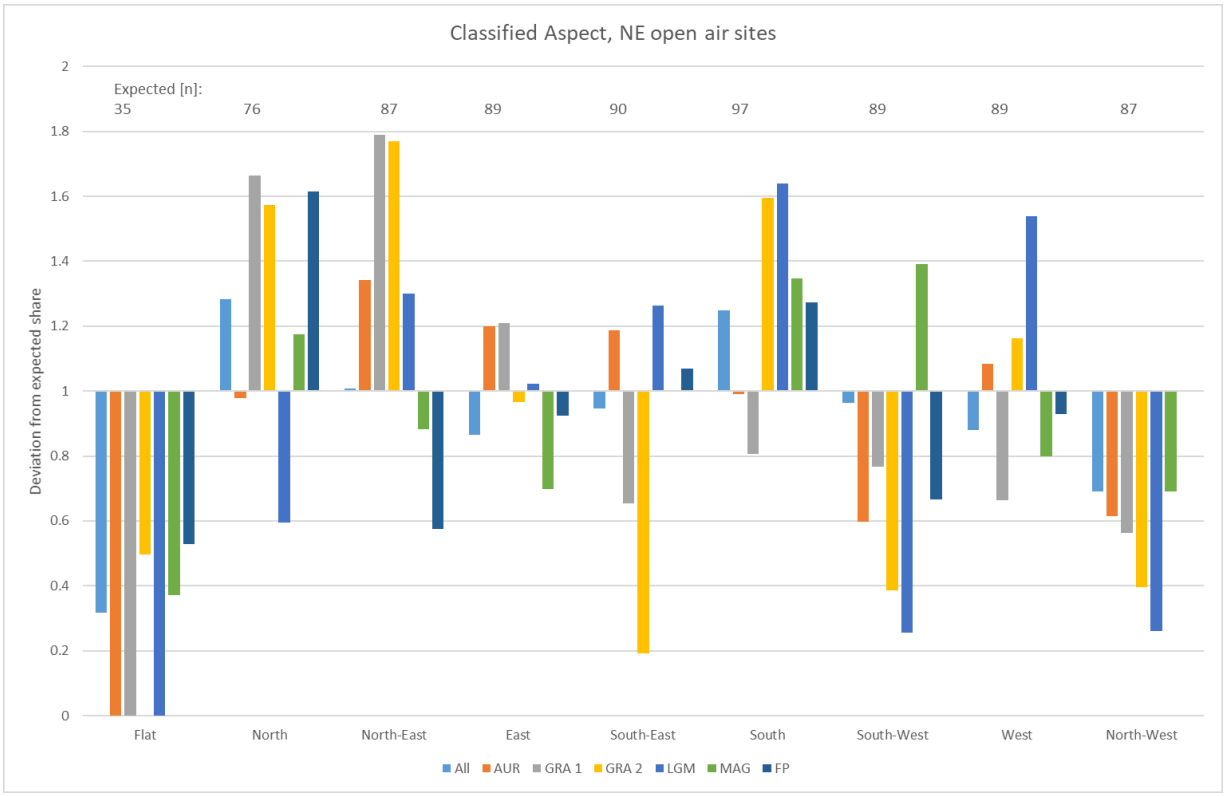



*Figure A76: Bar chart of the over- and underrepresentation. The expected values represent the share of all sites that would equal the area share of each respective surface. The overrepresentation is displayed as a factor of the expected value. Sites: NE open air, Variable: Classified aspect*

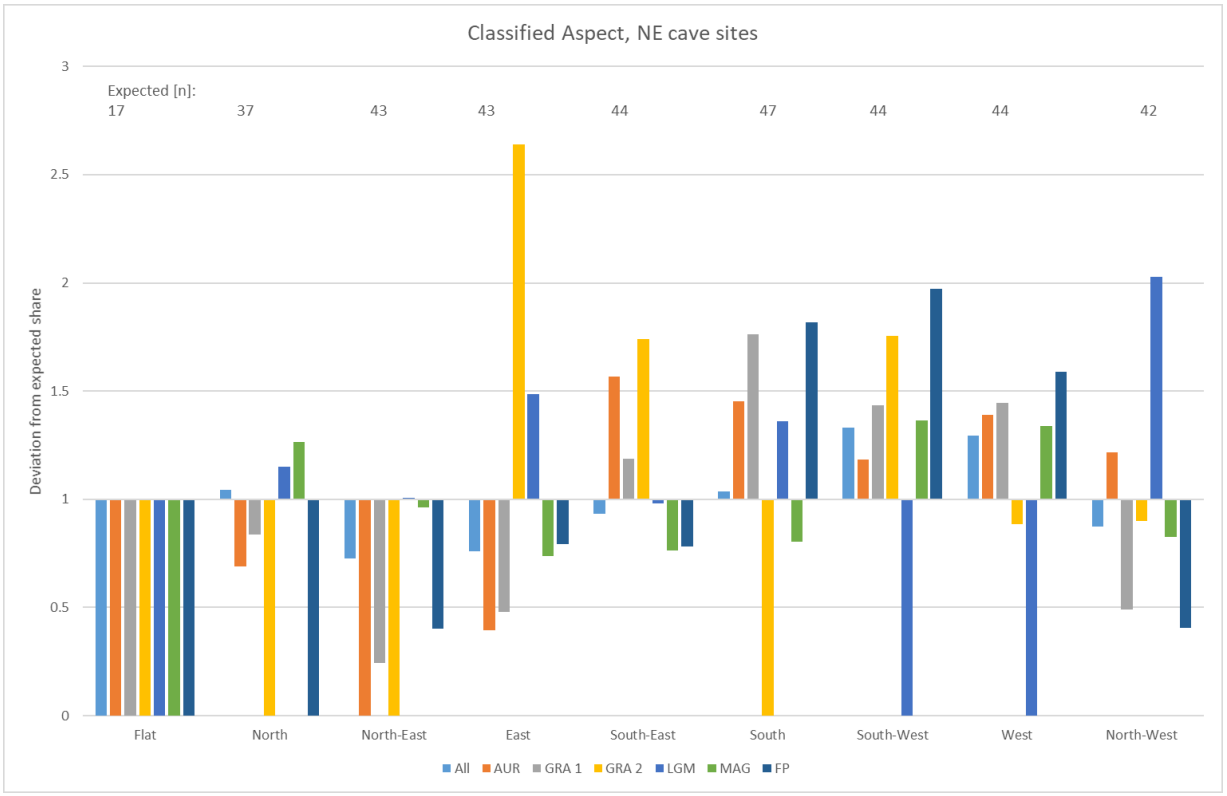



*Figure A77: Bar chart of the over- and underrepresentation. The expected values represent the share of all sites that would equal the area share of each respective surface. The overrepresentation is displayed as a factor of the expected value. Sites: NE cave, Variable: Classified aspect*

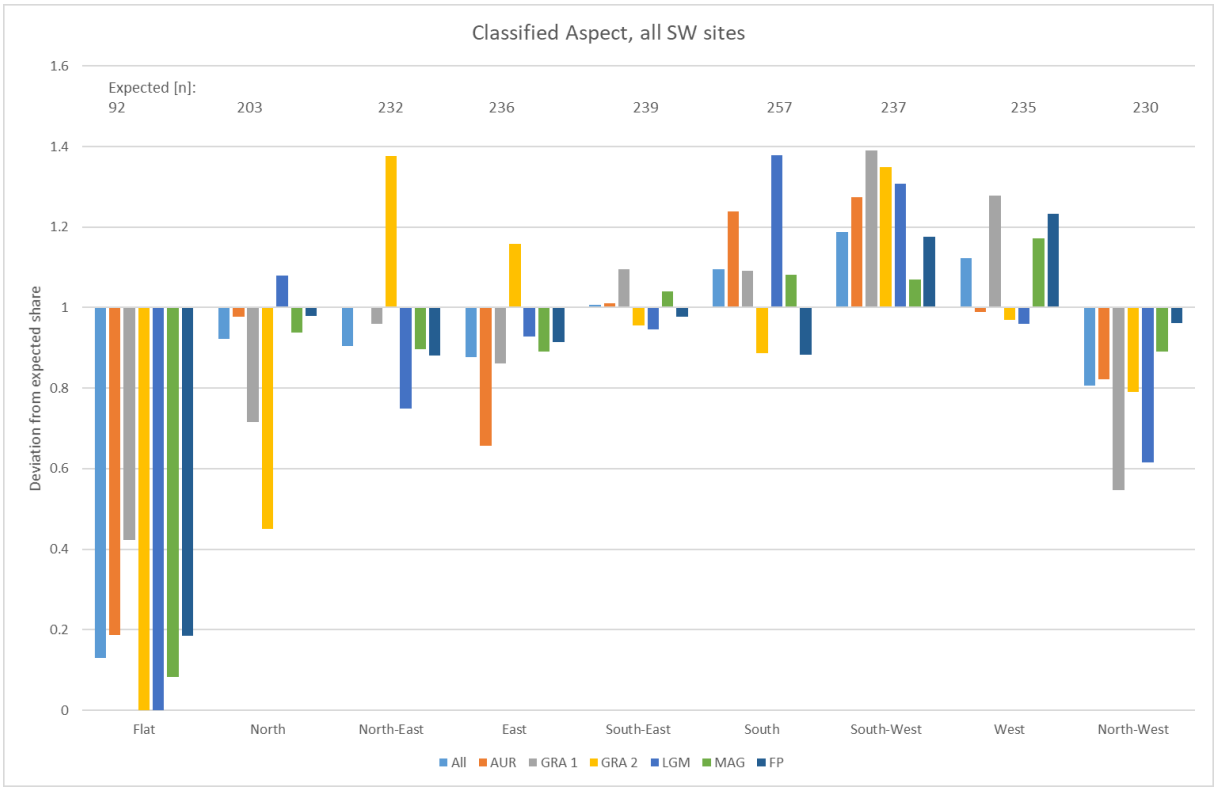



*Figure A78: Bar chart of the over- and underrepresentation. The expected values represent the share of all sites that would equal the area share of each respective surface. The overrepresentation is displayed as a factor of the expected value. Sites: SW, Variable: Classified aspect*

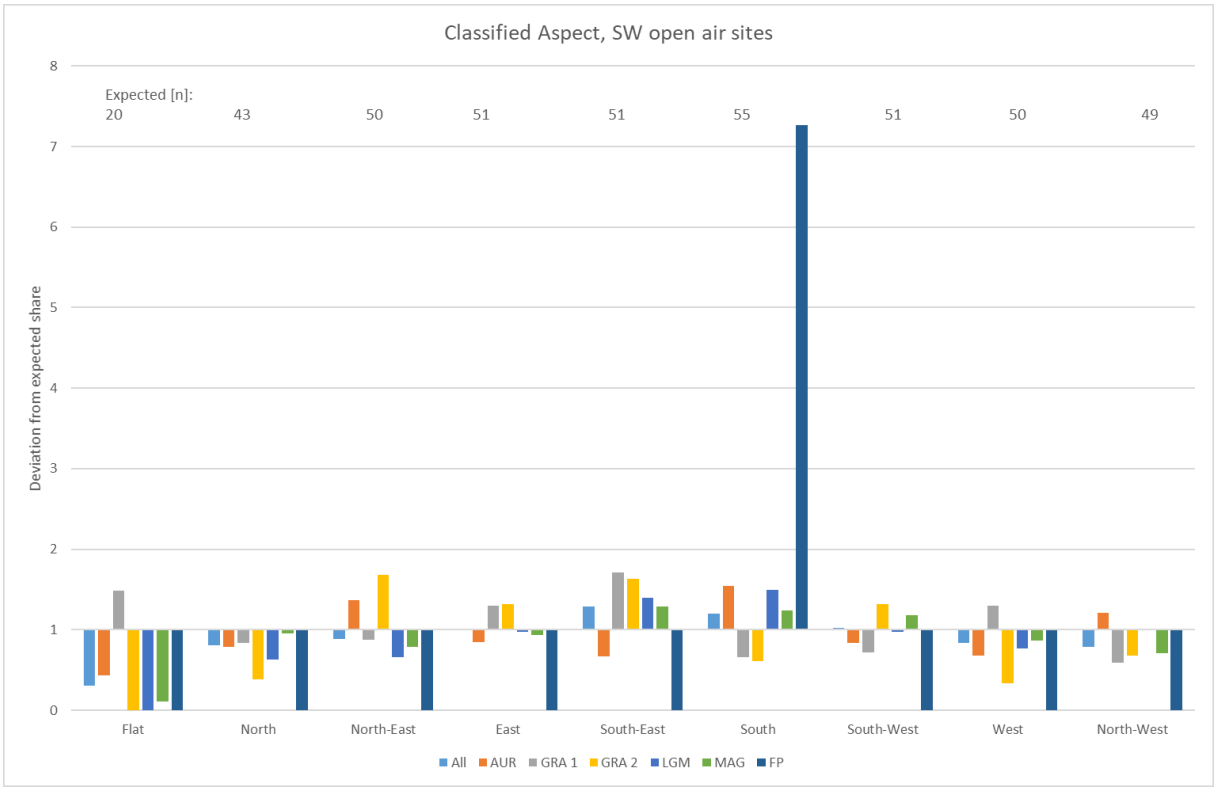



*Figure A79: Bar chart of the over- and underrepresentation. The expected values represent the share of all sites that would equal the area share of each respective surface. The overrepresentation is displayed as a factor of the expected value. Sites: SW open air, Variable: Classified aspect*

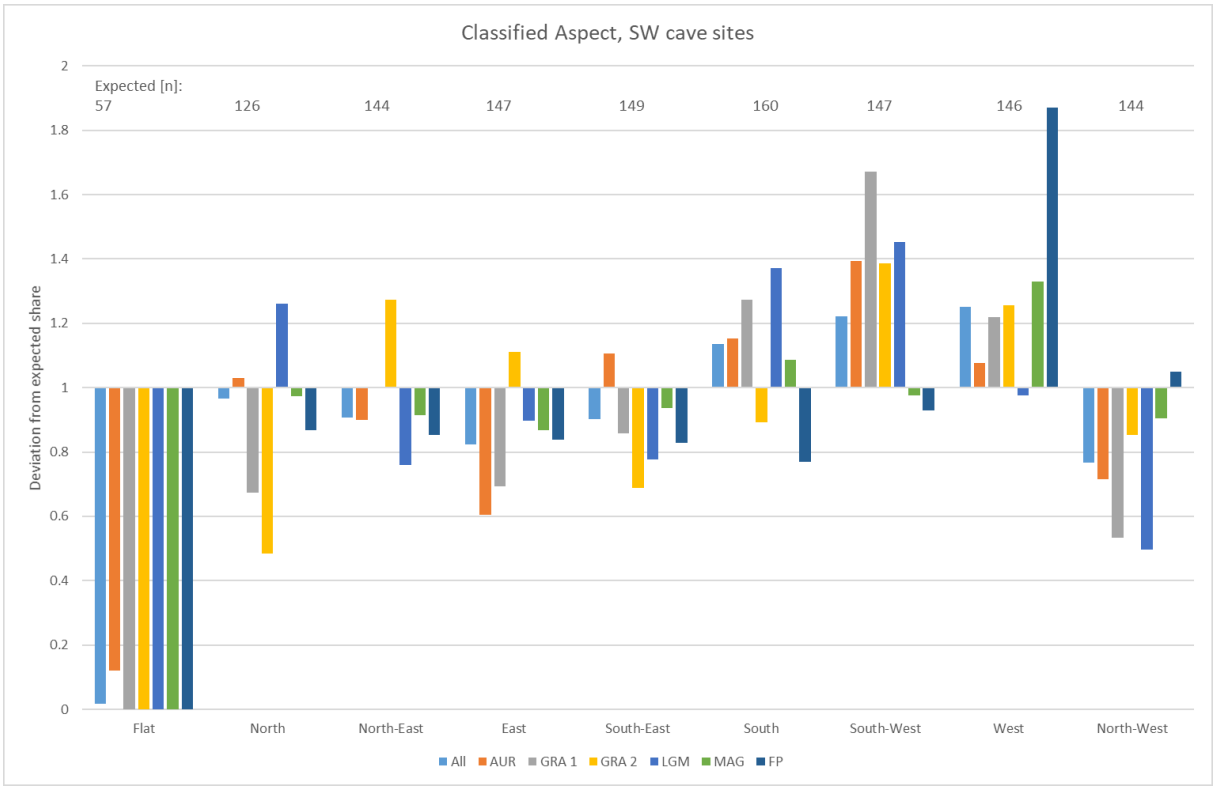



*Figure A80: Bar chart of the over- and underrepresentation. The expected values represent the share of all sites that would equal the area share of each respective surface. The overrepresentation is displayed as a factor of the expected value. Sites: SW cave, Variable: Classified aspect*

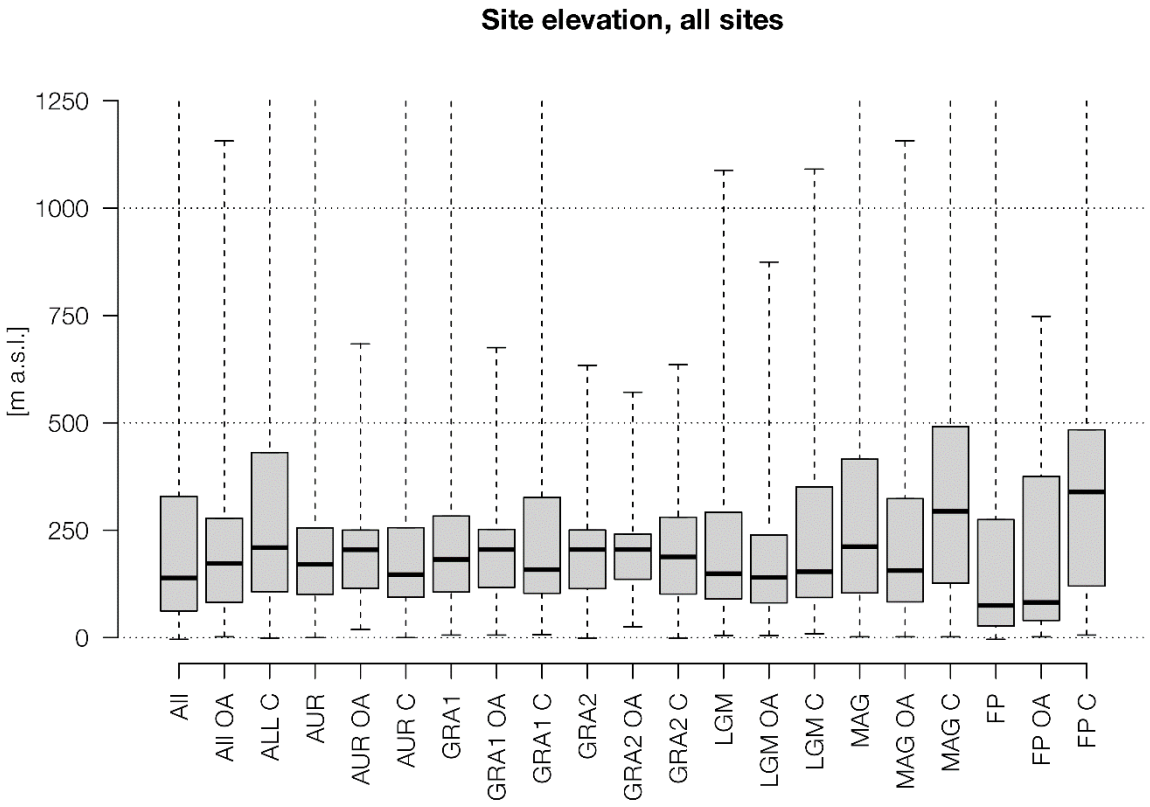



*Figure A81: Box plot showing the median and percentiles of site elevation for all sites*

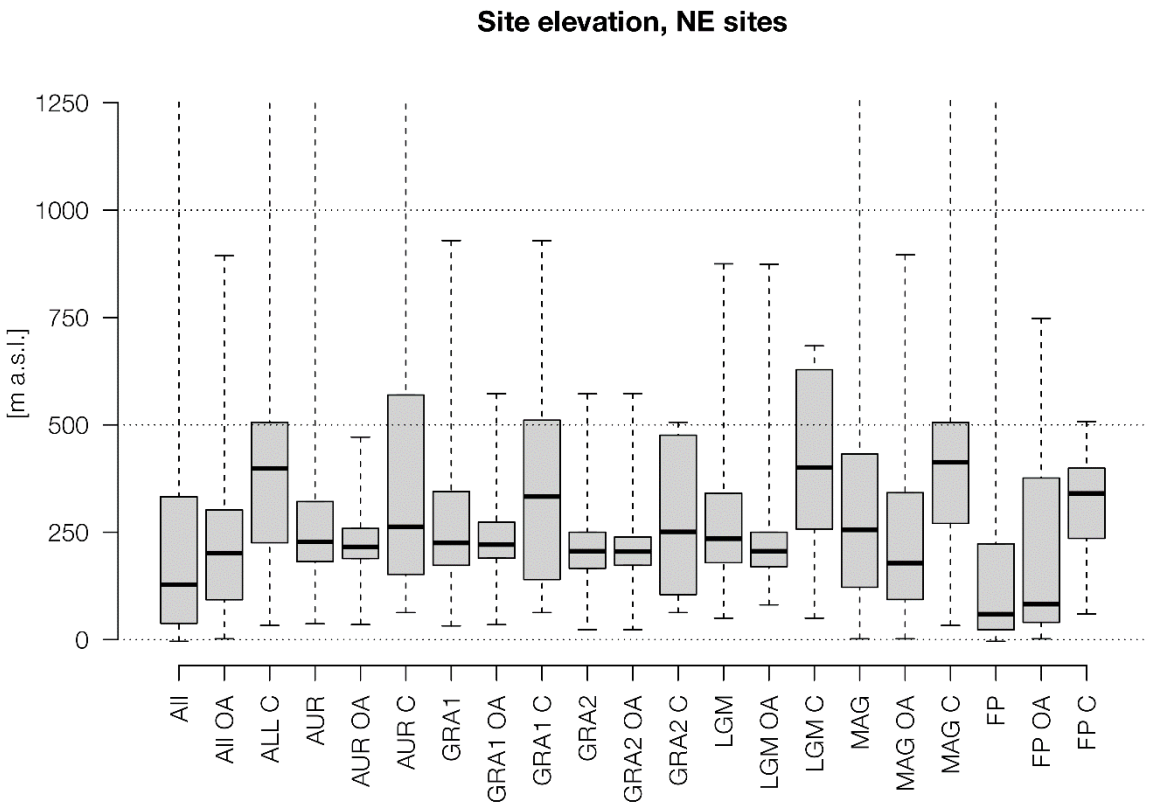



*Figure A82: Box plot showing the median and percentiles of site elevation for NE sites*

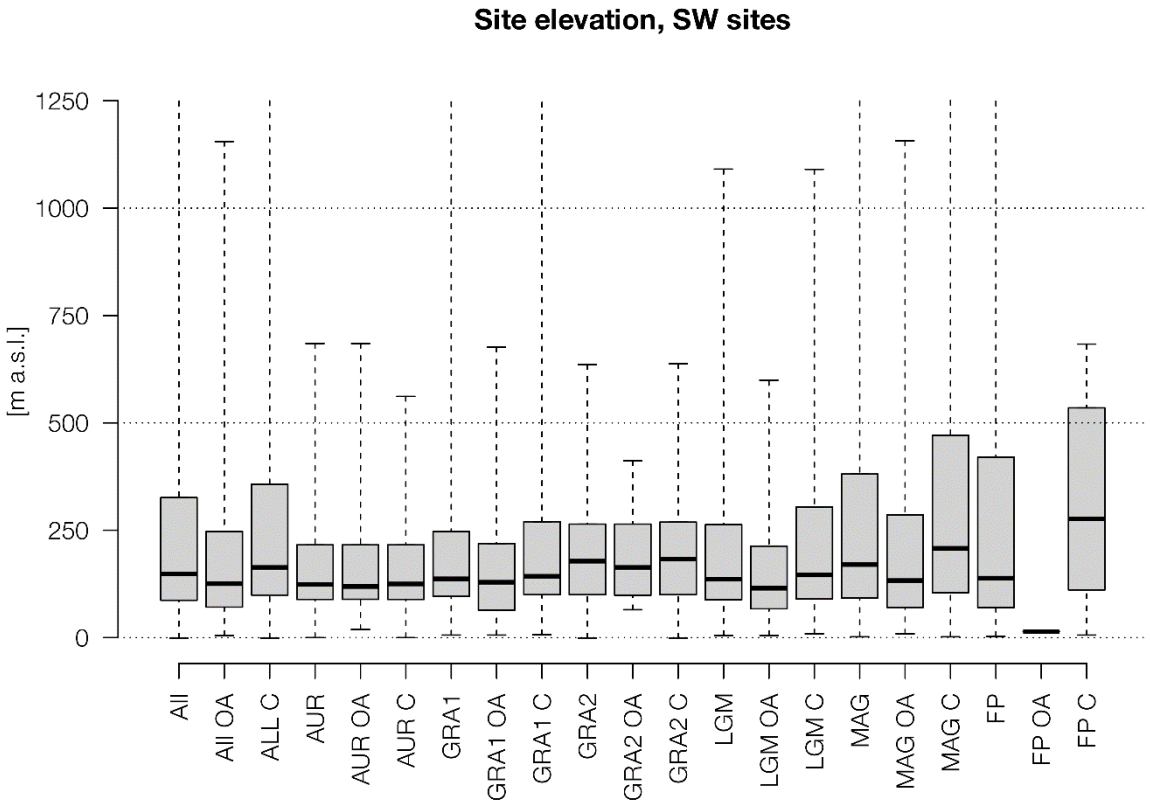



*Figure A83: Box plot showing the median and percentiles of site elevation for SW sites*

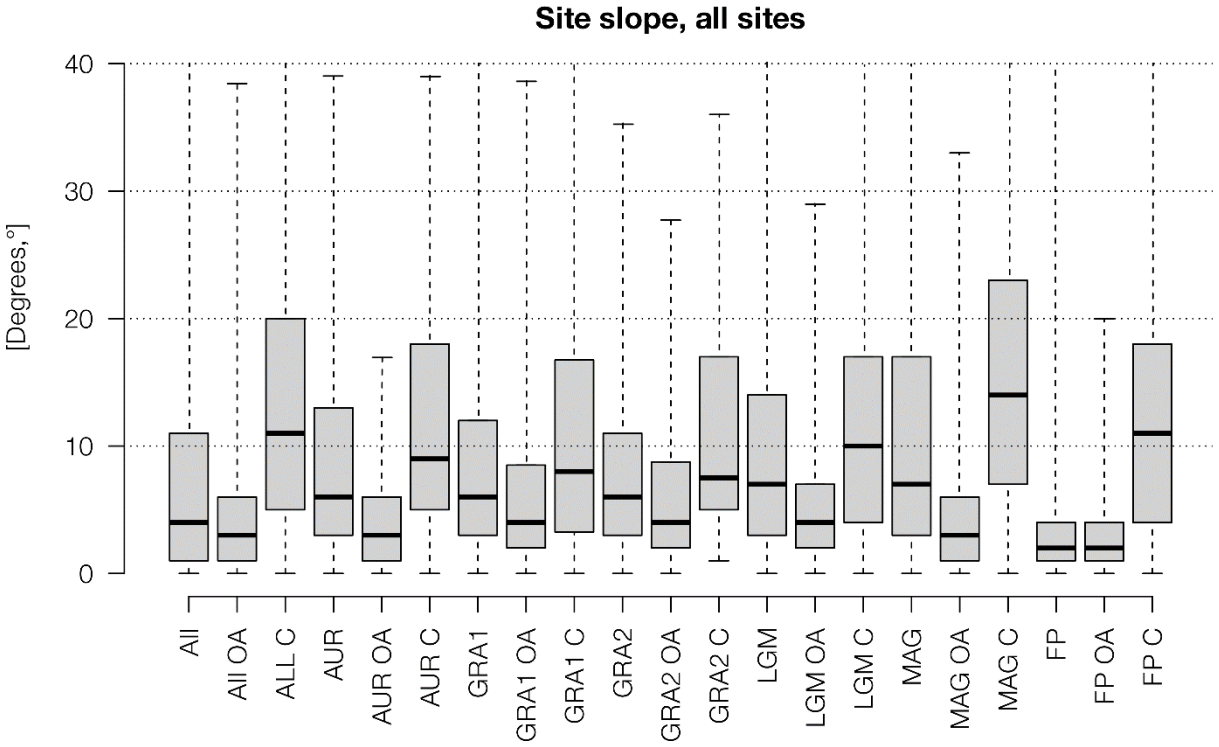



*Figure A84: Box plot showing the median and percentiles of site slope for all sites*

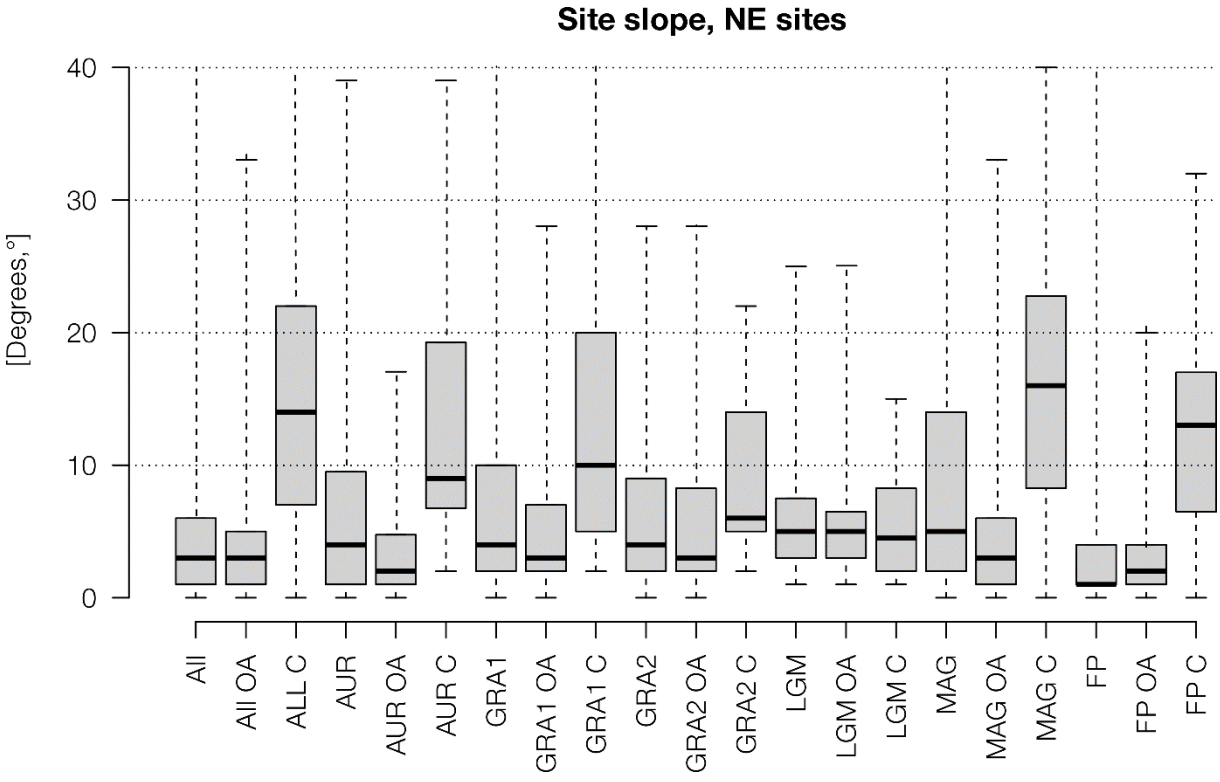



*Figure A85: Box plot showing the median and percentiles of site slope for NE sites*

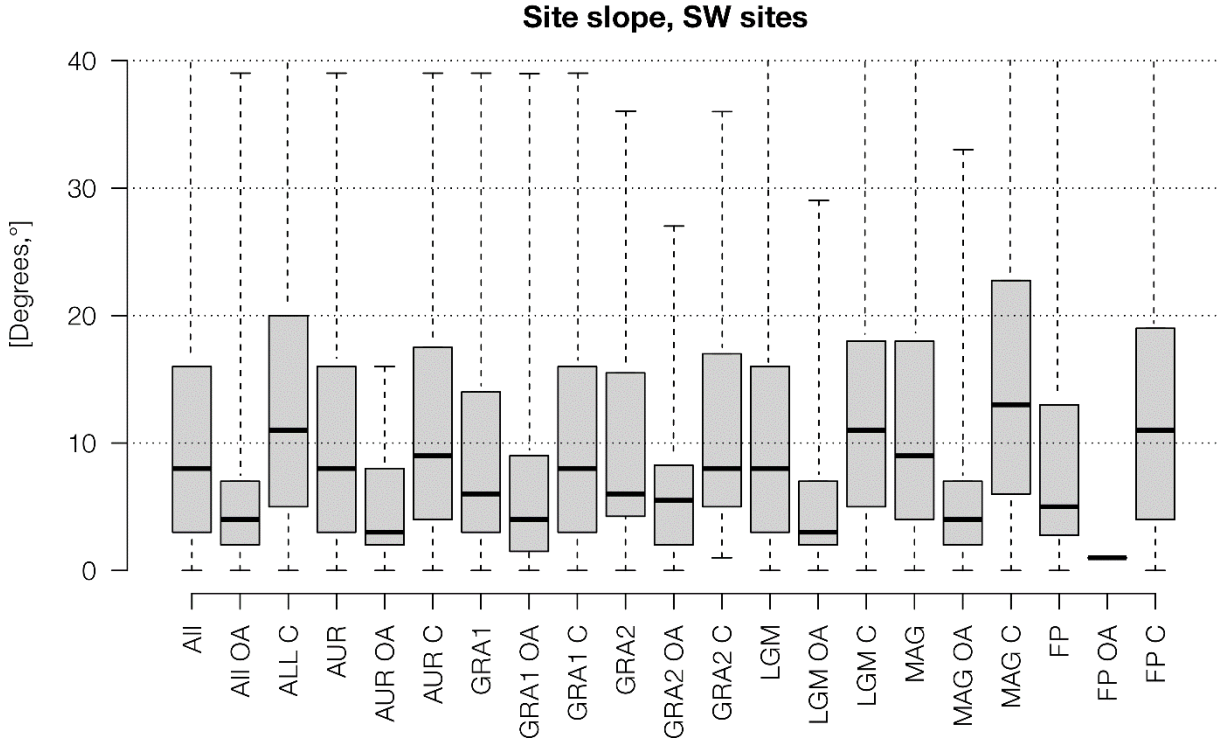



*Figure A86: Box plot showing the median and percentiles of site slope for SW sites*

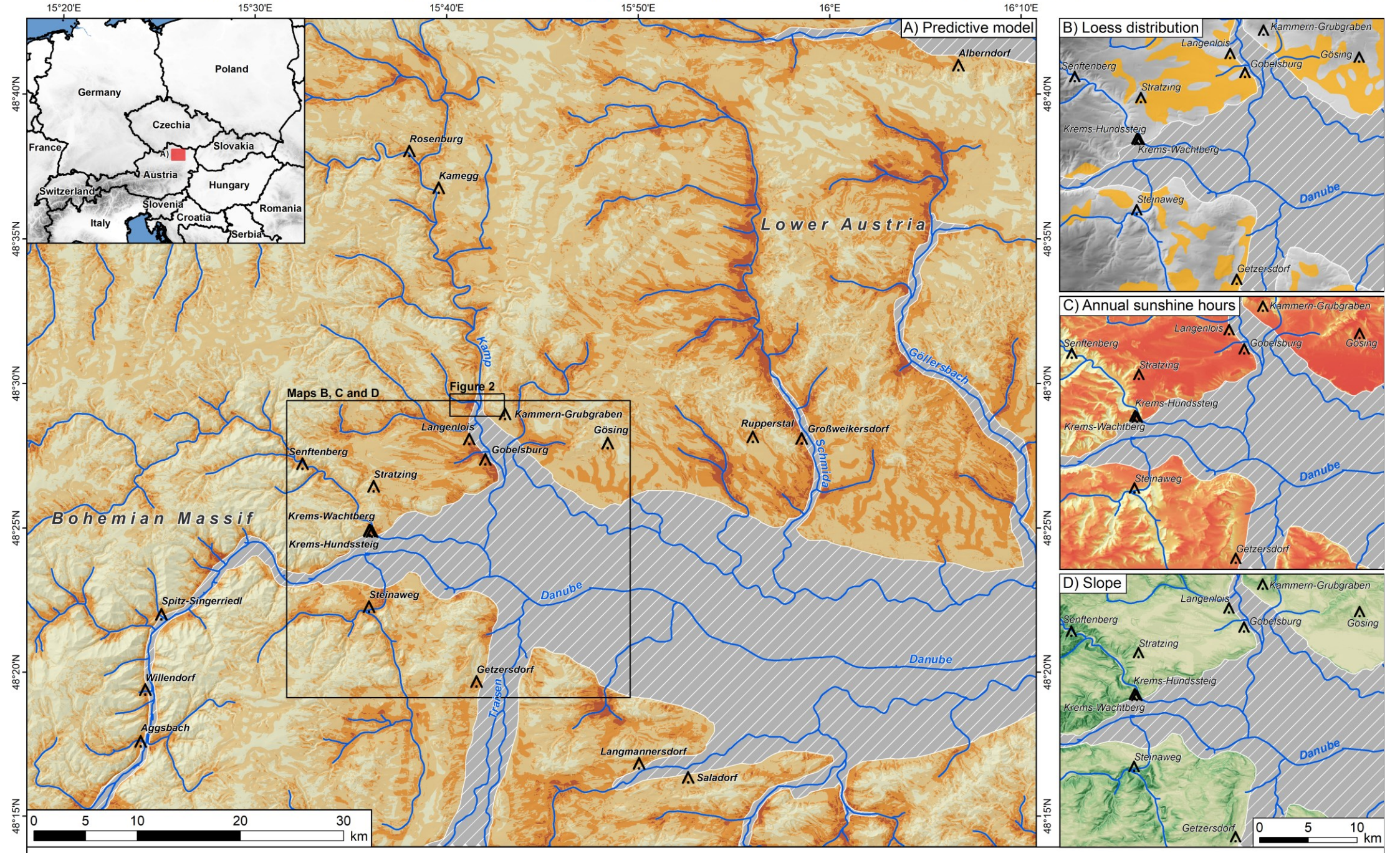
# Appendix B

## Supplementary material for chapter 3: Upper Palaeolithic site probability in Lower Austria – a geoarchaeological multi-factor approach

Main map showing the results of the archaeological predictive modelling on the left and three selected environmental predictors on the right.

# Upper Palaeolithic site probability in Lower Austria – a geoarchaeological multi-factor approach

Bruno Boemke[1], Thomas Einwögerer[2], Marc Händel[2], Frank Lehmkuhl[1] [1]Department of Geography, RWTH Aachen University, Aachen, Germany, [2]Austrian Archaeological Institute, Austrian Academy of Sciences, Vienna, Austria




GCS: GCS_WGS_1984
Datum: D_WGS_1994
Projection: Mercator_Auxiliary_Sphere
EPSG: 3857
Cartography: Bruno Johannes Boemke