From Radiance to Geometry: Identifying European Forest Clearings with Potential Heritage Value

Maximilian Schob[1], Luis Callejas[2,3]

[1]Norwegian University of Life Sciences (NMBU)/Norway · maximilian.schob@nmbu.no
[2]The Oslo School of Architecture and Design (AHO)/Norway · luis.callejas@aho.no
[3]Harvard University GSD/Cambridge · callejas@gsd.harvard.edu

Abstract: Forest clear-cuts, as both a forestry technique and a direct spatial consequence of forestry, are currently being phased out in Europe (European Commission, 2020, 2021). However, clearings resulting from this long-standing landscape practice will remain across various forested regions of Europe for the foreseeable future. The physical traces derived from forest clear-cutting may hold potential landscape heritage value, as they are the remnants of an important cultural landscape undergoing drastic change due to its environmental impacts. Despite their ubiquitous presence, valuable clearings are hidden in plain sight. Sufficient attention is not given yet to their spatial qualities (Callejas, 2021b; Callejas & Hansson, 2022): they are not cataloged geographically and have yet to be incorporated into landscape surveys documenting the digital and physical heritage of European cultural landscapes.

The techniques explored in this article are twofold: first is to develop a method for identifying the clearings, and second is to provide sufficient geometric information to value and categorize the shapes of selected clearings. Among various potential uses that such visualizations can illustrate, the most promising appears to be determining previously underrepresented landscape configurations based on perceived formal parameters (Callejas, 2021b) applicable to landscape heritage and other cases wherein geometric information might be needed.

The outcome is a novel model for the localization, identification, and geometric processing of forest clearings within the geographical scope of Europe. This model is executed through connecting the cloud computing capabilities and extensive data catalogs of Google Earth Engine (Gorelick et al., 2017; Wu, 2020) with effective image segmentation based on the Meta Segment Anything Model (Kirillov et al., 2023; Wu & Osco, 2023) and several open-source geospatial data analysis tools developed for geometric processing. Hansen Global Forest Change data (Hansen et al., 2013) is employed to detect disturbances in forest patterns on a global scale, serving as a preliminary filter to identify forest clearings in multiple national, high-resolution, color infrared (CIR), orthoimagery acquisitions across Finland, Latvia, and Germany. Expanding on the geographical scope of our previous study (Schob & Callejas, 2023), the present methodology allows us to efficiently identify, filter, select, and process a rich set of sample clearings. For each clearing, a series of infrared image tiles, precise outlines, and geometric synthetizations are produced. Furthermore, outlines and abstract geometries in plan view are categorized in relation to their spatial location.

While the collection of clearings could be expanded further, including other European countries, we offer a systematic, robust, and innovative approach that integrates different methods and data through a single accessible and centralized interface. Classification and segmentation models commonly used in the geosciences serve to determine landscape typologies and observe their transformations. However, they may be inherently limited in detecting geometrically fuzzy or temporarily unstable objects and surfaces. This research not only challenges the limitations of some of these tools and methods but also provides a clear workflow for identifying, collecting, categorizing, comparing, and ultimately valuing the complex shapes of forest clearings on a European scale.

Keywords: Forest clearing, Hansen global forest change, color infrared orthoimagery, image segmentation, geometric synthetization

Journal of Digital Landscape Architecture, 9-2024, pp. xxx-xxx. © Wichmann Verlag, VDE VERLAG GMBH · Berlin · Offenbach. ISBN 978-3-87907-xxx-x, ISSN 2367-4253, doi:10.14627/537xxxxxx.

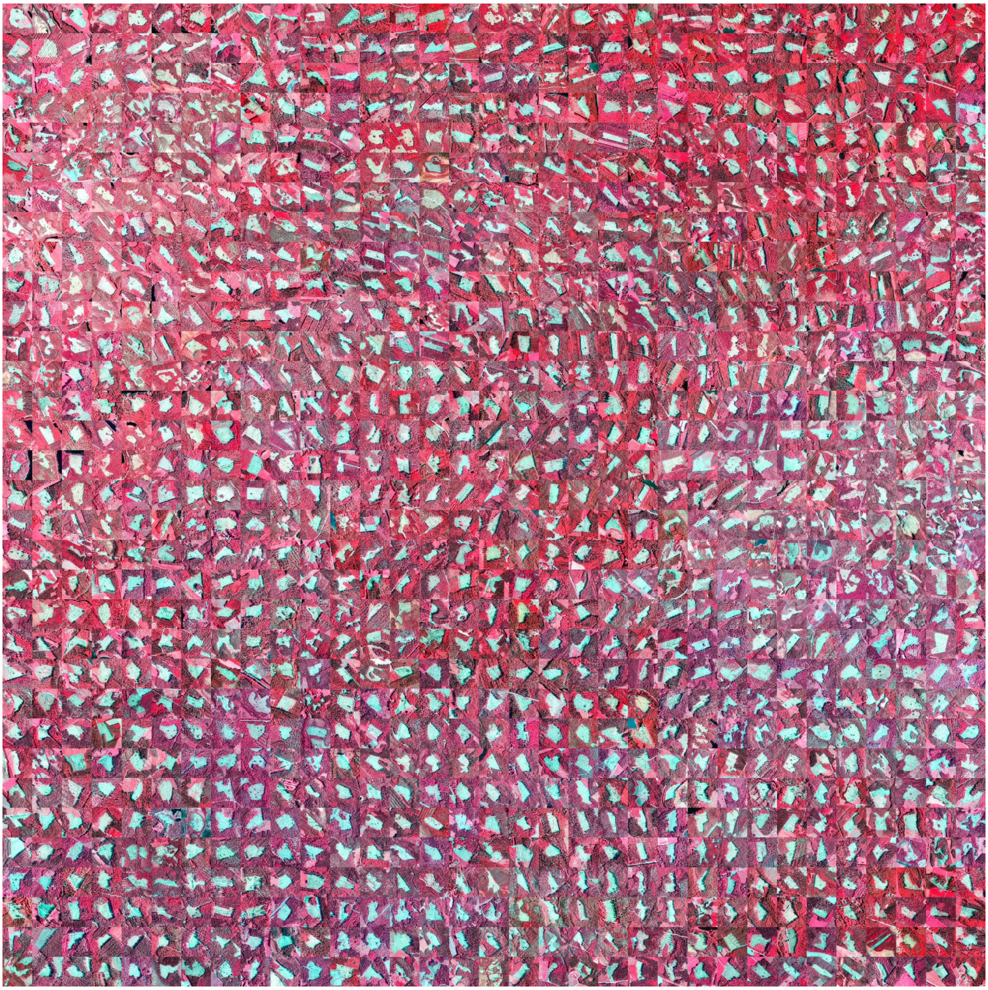

Fig. 1: 1089 clearing image tiles extracted from the Finland CIR dataset for our initial study on clearing typologies, visualization processed using a custom t-SNE script.

1 Preserving Europe's Forest Clearings: Documenting Landscape Heritage through Contemporary Survey Techniques

As forest clear-cuts, as both a technique and spatial consequence of forestry, are being phased out in Europe (European Commission, 2020, 2021), clearings resulting from this long-standing landscape practice will persist across various forested regions of Europe for the conceivable future. The physical traces left by forestry practices, particularly those resulting from clear-cutting, hold potential landscape heritage value as remnants of a practice that will undergo a drastic change over the years to come due to environmental impacts. Even if changes proposed by the European Parliament are generally positive and well-intentioned in terms of addressing their environmental effects, the physical traces left in European forests over centuries are among the most tangible manifestations of a practice that has shaped Europe's forests, its material culture, and the evolving relationship of Europeans' with forested land.

These clearings are not adequately discussed or sufficient attention is given to their description: they are not cataloged geographically and have yet to be incorporated into landscape surveys that document the digital and physical heritage of European landscapes. This article deploys techniques to begin documenting these fleeting landscape entities. As "the clearing is a landscape entity defined by the absence of that which materializes its perimeter" (Callejas, 2021a), it lacks the definition that other material traces with heritage value have in Europe, particularly those with architectural value or more tangible and measurable presence, such as ancient farming plots and mines.

Through utilizing different contemporary survey techniques, ranging from lidar scanning to aerial and satellite optical imaging, the article aims to define a methodology for detecting, classifying, and documenting forest clearings across a range of European forest conditions. The intended goal is to propose that a number of these clearings should be included within heritage categorizations, eventually preserving them in a similar manner to other remnants of landscape practices (such as farming and mining), which have been extensively documented for digital and physical preservation purposes in Europe.

Previous projects, positioned at the intersection of the arts, culture, and geosciences, have successfully employed remote sensing techniques that operate beyond the visible spectrum for spatial explorations of the forest (LCLA Office et al., 2021, 2023, 2024) and (Berrizbeitia & Hildebrand, 2024; Formafantasma, 2021; Mosse, 2022; Tavares, 2010); however, these examples focus on regions of the world where the data is more or less consistent in terms of the employed technique, temporal resolution, and spatial resolution. In contrast to our approach, which focused on utilizing remote sensing imagery produced as a part of large-scale automated remote sensing systems, the process of data making, the creation of project, site, and purpose-specific data through the author, is at the core of many such projects.

Previous authors have discussed the limitations of lidar data for addressing the dynamic nature of forest clearings as landscape entities in a constant state of change due to natural succession. Starting from the limitations described in our previous study (Schob & Callejas, 2023) and the fact that forested land in Europe is surveyed through disparate techniques with uneven levels of spatial and temporal resolution, the article lays out a methodology for the initial capture and documentation of the most valuable clearings as expressed and valued by their geometry and dimensions. Furthermore, as a clear, standardized classification for forest clearing does not exist, nor does an attempt to document this fleeting landscape entity in Europe, the outcome of this research will serve as a precursor to start discussing European clearings in both the emerging field of digital landscape heritage as well as cultural landscape heritage.

2 Methodology: Integrating Earth Engine, Segment Anything Model & Geometric Processing with Python in Google Colab

Forest clearings are fleeting and ambiguous landscape figures, emerging as and maintained by a threshold state: a void inside the dense fabric of a forest and a space surrounded by trees, the absence of chlorophyll detectable through aerial infrared imaging techniques and the exposure of the ground this forest arises from. To frame the forest clearing in a larger European context, we propose an integrated model, combining a software and data stack that synergizes the capabilities of multiple data, processing tools, and software environments. We provide one coherent model that can be divided into three functional parts:

Data Collection & Processing with Earth Engine and Geemap: We use Hansen Global Forest Change (Hansen GFC) data to detect disturbances in forest patterns on a global scale. Hansen GFC serves as the preliminary filter for identifying clearings in national, high-resolution, color infrared (CIR), orthoimagery acquisitions. We generate square CIR image tiles for each clearing, coupled with a set of sample points for forest and clearing regions, which are later used as input for the image segmentation process.

Image Segmentation with SAM and SamGeo: We use Segment Anything Model (SAM) with SamGeo to delineate clear-cut areas from CIR image tiles. SamGeo utilizes an automatic, prompt-based, or input-based method for segmenting images. We use the input-based method with sample points to produce high-resolution alpha masks and precise outlines for forest clearings.

Geometric Processing with Python: Using a selection of Python libraries, we create abstract, generalized geometric descriptions for each of the clearings, which allow us to assess, categorize, and document their shape.

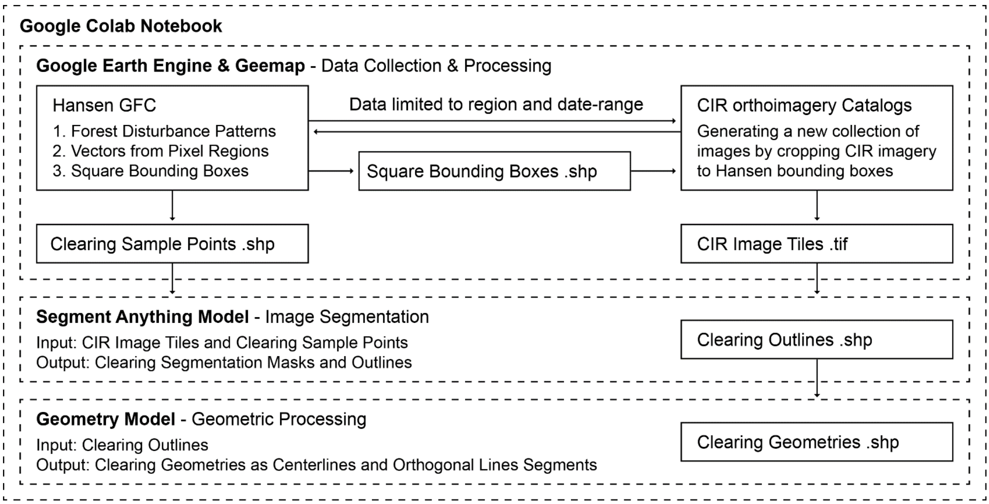

Fig. 2: Processing diagram of our model showing the sequential processing steps and relevant data products.

The cloud-based Python development environment Google Colab (GC) serves as a central interface in our project to combine multiple data sources, processing methods, and tools (Google LLC, 2024a). This allows us to create, edit, and share interactive notebooks in a web-based Jupyter environment hosted by GC and stored in Google Drive. We use GC notebooks to combine Google Earth Engine, Segment Anything Model, and several Python libraries for geometric processing.

Each part of our script generates distinct georeferenced data products in .shp or .tif format, which are saved as temporary files in a specified GC cloud directory. This sequential processing approach removes thresholds of closed environments and enables significant flexibility in working with different data products. The limitations of a specific environment are overcome by exporting data products from one environment to a GC cloud directory, which then serves as input for subsequent processing steps. For instance, processing methods that cannot be called in the geemap environment can be used on the products that are exported from geemap and saved to a GC cloud directory. This allows for the inclusion of other web- or cloud-based datasets, such as WMS layers, in the processing workflow.

In the following three parts, 2.1 Data Collection and Processing in Earth Engine, 2.2 Segmentation using Segment Anything Model, and 2.3 Geometric Processing, we are presenting the sequential processing steps of our methodology. The aim is to suggest a functional model and methodology, demonstrating the potential use and application of each part as a functional component of a sequence of processing steps. The complexity of the larger model may impose some limitations on the depth of information we are able to provide for some of the mentioned Python libraries, processing steps, or data products. However, for each step, we offer the most relevant references and visualizations of data products to enhance the accessibility of our model and methodology.

2.1 Data Collection and Processing with Google Earth Engine: Hansen Global Forest Change and National CIR Imagery

The first part of our model is dedicated to data collection and processing. We introduce Google Earth Engine and Geemap as operational tools for working with a selection of datasets from a global to a local scale. We characterize the main datasets used in our study, such as Hansen GFC and several national CIR imagery datasets, and describe the first segment of processing steps and data products.

Google Earth Engine (GEE), or Earth Engine, is a cloud-based service for geospatial data processing on a planetary scale, which combines a multi-petabyte-scale data catalog with extensive cloud processing capabilities and an accessible API implemented in Python and JavaScript (Gorelick et al., 2017). The GEE data catalog aggregates a large amount of publicly available geospatial datasets, usually only accessible through a vast and growing array of institutions, services, and platforms. The centralization of datasets in one accessible repository eliminates significant constraints in data interoperability. Typical processing steps that need to be considered when working with geospatial data, such as coordinate reference system transformations, are rendered invisible to the user. As part of a rapid expansion of remote sensing environments, "big data deals with various issues, such as capturing data, searching, sharing, storing, transferring, visualizing, querying, and updating information" (Tamiminia et al., 2020). GEE has perhaps gone the furthest in developing a centralized environment that has moved geospatial data processing to the cloud, eliminating accessibility thresholds, such as "data acquisition and storage; parsing obscure file formats; managing databases, machine allocations, jobs and job queues, CPUs, GPUs, and networking; and using any of the multitudes of geospatial data processing frameworks" (Gorelick et al., 2017). This is a timely reflection of Google's mission to "organize the world's information and make it universally accessible and useful" (Google LLC, 2024b).

Geemap is a Python library developed for processing, analyzing, and visualizing geospatial data with GEE (Wu, 2020). Geemap enhances the capabilities of the GEE Python API, enabling the interactive exploration and processing of large geospatial datasets in a web-based Jupyter environment (Wu, 2024). While early-stage testing of our scripts has been executed in the GEE JavaScript API code editor, integrating geemap into a GC environment, as presented in this paper, offers a more streamlined approach to data processing and the visualization of results.

Fig. 3: Global forest disturbance patterns, elaborated by the Hansen Global Forest Change dataset, between 2000 (red) and 2022 (black).

Definitions of forest clearings pertain to a pronounced threshold in datasets derived from remote sensing observations. Understood as a forest/non-forest variable, these clearings may be described in land cover classification datasets, most commonly used for detecting and monitoring change in a set of predefined classes. Land cover classification considers individual classes as generalized, continuous areas based on the classification method and resolution of the underlying observation dataset. While there are numerous classification datasets available from a local to a global level, the highest resolution, continuous, global datasets still maintain a spatial resolution of tens of meters.

More precise datasets, at centimeter resolution, that allow us to consider forest clearings at the level of the individual trees are largely managed by and often uniquely available through regional or national institutions. Based on the previously described trade-offs between spatial continuity and resolution, we employ two different types of datasets to frame clearings on a European level: Hansen GFC and national CIR orthoimagery datasets.

We use the Hansen Global Forest Change v1.10 2000-2022 (Hansen GFC) dataset (Hansen et al., 2013) to detect forest cover disturbances on a global scale and by year of gross forest cover loss event. The global scale analysis of change in forest patterns in the Hansen dataset consists of a "time-series analysis of Landsat images in characterizing global forest extent and change" (Hansen et al., 2023). A forest cover loss event, described as an annual variable in the lossyear attribute, is defined as a "stand-replacement disturbance, or a change from a forest to non-forest state" (Hansen et al., 2023). We utilize Hansen GFC for the preliminary site selection process.

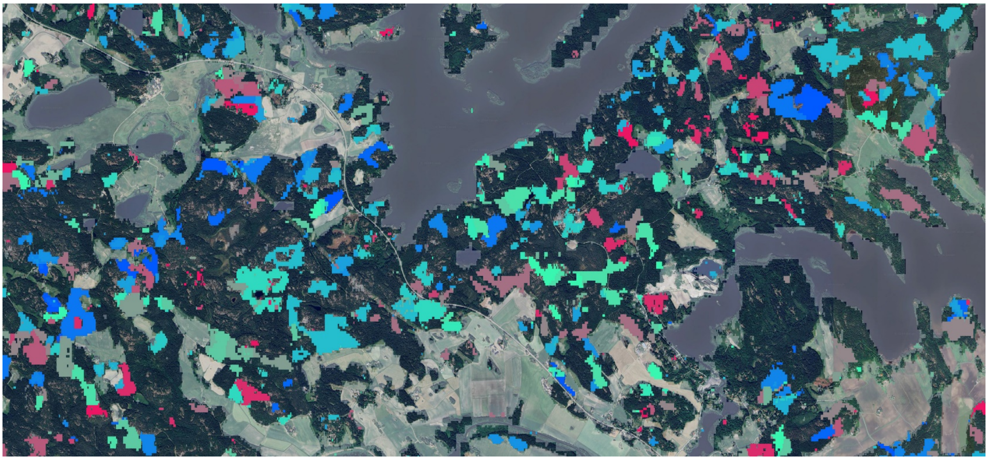

Fig. 4: Hansen Global Forest Change as a gradient of pixel areas from 2000 (red), 2010 (green), to 2022 (blue) over aerial imagery.

While Hansen GFC is useful for identifying disturbances in forest patterns on a larger scale, the spatial resolution of approximately 30.92 meters is unsuitable for developing a precise, sensible, geometric description of the forest clearing. The above image shows Hansen GFC pixel regions colored by the lossyear attribute, as a gradient from 2000 (red), 2010 (green), to 2022 (blue). Thresholds of the spatial resolution of a dataset for the identification of forest clearings were discussed in our previous study (Schob & Callejas, 2023).

We use color infrared (CIR) composite orthoimages with three bands: near-infrared (N), red (R), and green (G) mapped to RGB as a false-color composite. An orthophoto is a digitally corrected orthorectified aerial image. The use and applications of infrared optical imagery for remote sensing of vegetation have been extensively demonstrated in the field (Jensen, 2014). Plant canopies with healthy leaves reflect a particularly high amount of near-infrared energy, resulting from factors, such as plant physiology or leaf additive reflectance (Jensen, 2014). The specific reflectance, transmittance, and absorption characteristics of vegetation in the near-infrared and visible spectrum allow to accurately sense and describe different types of vegetation and land cover characteristics.

For our study, we used a selection of CIR imagery datasets readily available through the GEE data catalog. The three sample regions focus on datasets in **Finland**, **Latvia**, and **Germany**.

While there is clear potential for expansion of the geographic scope and selection of datasets, we prove that our model is not limited to a specific region but rather to the extent and availability of CIR datasets. Moreover, other datasets available through cloud- or web-based services, such as WMS layers, could potentially be added to the data selection and processed in the same model.

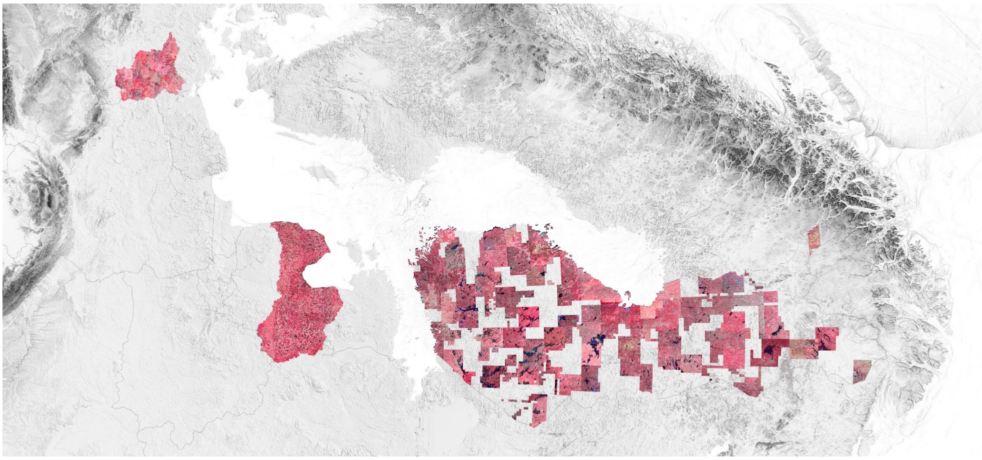

Fig. 5: Overview of CIR imagery collections from Finland, Latvia, and Germany used for our study, north to the right.

Finland NRG NLS orthophotos (National Land Survey of Finland, 2023b) are provided by The Energy Agency (formerly SMK) and accessible through the National Land Survey of Finland (NLS). NLS orthophotos provide complete coverage of Finland and are captured in a three-year cycle following the regional map sheet division system. Digital false color (GBN) images are produced by SMK as GeoTIFF files at a spatial resolution of 50 cm, in 8-bit, and a native projection of EPSG:3067 (National Land Survey of Finland, 2023a). The GEE data catalog hosts images of this collection between 01.01.2015 and 01.01.2022.

Latvia Color InfraRed (CIR) orthophotos (Latvijas Geospatial Information Agency, 2023a) are accessible through Latvijas Geospatial Information Agency. The dataset contains digital false color (GBN) images for the whole of Latvia, captured in a regional map sheet division at a spatial resolution of 20 cm, in 8-bit, and a native projection of EPSG:3059 (Latvijas Geospatial Information Agency, 2023b). In the GEE data catalog, images are available from 01.01.2007 to 01.01.2018.

Brandenburg (Germany) RGBN orthophotos (Landesvermessung und Geobasisinformation Brandenburg (LGB), 2023b) is an orthophoto dataset provided by the state government of Brandenburg (LGB) and covers the regions of Brandenburg and Berlin, the only regions in Germany that have made their data accessible through the earth engine data catalog at the time of surveying. The data is captured in a three-year cycle. The dataset contains digital RGBN imagery at 20 cm spatial resolution, 8-bit, and a native projection of EPSG:7837 (Landesvermessung und Geobasisinformation Brandenburg (LGB), 2023a). Images are available through the GEE data catalog between 23.08.2021 and 20.01.2023.

For spatial and temporal congruence, we filter the region and date range of Hansen GFC to the extents and image acquisition dates of the CIR collections, limiting Hasen GFC to the extent of the CIR collections and remapping lossyear values in Hansen GFC to include only values from one year prior to the CIR imagery acquisition dates. By aligning the parameters of the two datasets, we ensure the inclusion of the most recent data while reducing the search area to a minimal extent. Since Hansen GFC data composites are produced at annual intervals, however, the timing of a forest disturbance pattern can only be dated to a year, not to an exact date. Consequently, there may be a temporal offset of several months between what is registered by the Hansen dataset and the image acquisition date of the CIR collections.

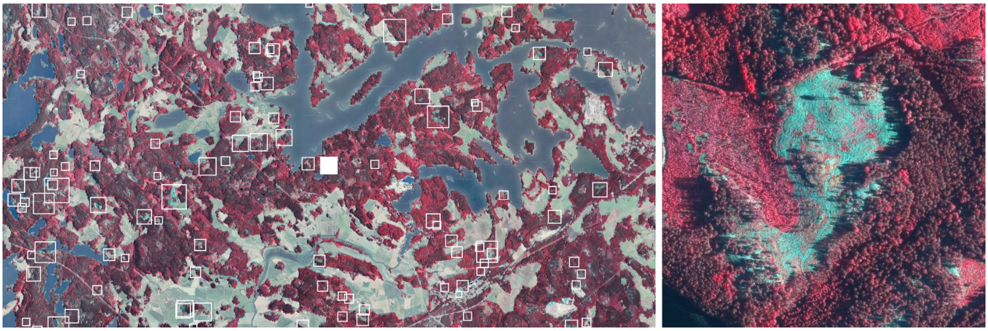

Fig. 6: Collection of clearings found in the CIR imagery landscapes of Finland (left), with a white square for the position of the resulting square infrared image tile (right).

We reduce Hansen GFC raster pixel regions to vector polygons, calculate their area, and filter them, with reference to the size of clearings in our previous study (Schob & Callejas, 2023), keeping clearings of an area between 10000 m² and 200000 m². We create offset square buffer regions around each polygon, which serve as clip regions to extract individual clearing images from the CIR imagery dataset. Each bounding box produces an image that is added to a new image collection. We export square CIR image tiles as .tif files for each clearing region.

We were able to extract 33,850 clearing image tiles from the 2021 Finland CIR imagery acquisitions using Hansen data between 2019 and 2021, 4169 tiles from the German CIR acquisitions between 2021 and 2023 using Hansen data between 2019 and 2021, and 6012 tiles from the 2016 CIR acquisitions in Latvia using Hansen data between 2015 and 2016. We used a custom t-SNE script to organize our image tiles into a grid layout and obtain an initial overview of potential patterns in our CIR image tile collections. Our current selection comprises more than 44,000 clearing image tiles that are primarily constrained by the parameters specified as part of our process rather than the number of forest disturbance pattern areas, and therefore potential clearings, found throughout the European forests.

To streamline the subsequent segmentation process, we create a set of sample points for clearings and surrounding forest areas by sampling the pixel values of the Hansen GFC raster within our bounding boxes. In addition, we construct a buffer region from the Hansen GFC vector polygons, as shown in Fig. 7, to exclude points near the edge between forest and clearing.

Given the sparse resolution of Hansen GFC, we expect that the pixel regions representing forest disturbance patterns may not precisely align with the actual ground in clearing areas as visibly encountered in the infrared imagery. Although SamGeo provides a robust algorithm capable of detecting outliers in sample points, we choose to implement a buffer region to prioritize accuracy in the samples we provide, even if that might mean fewer samples in quantity. The sample points are presented as a formatted list of coordinates in the form of "Clearing Coordinates: [[x, y], [x,y], [x,y…"

The collection of national IR image tiles is constrained by the selection of national datasets that are readily available through the GEE data catalog. However, while we provide the most streamlined workflow, our method also accommodates the use of datasets outside of the GEE data catalog, such as local GeoTIFF files sourced individually from other national inventories or WMS image layers.

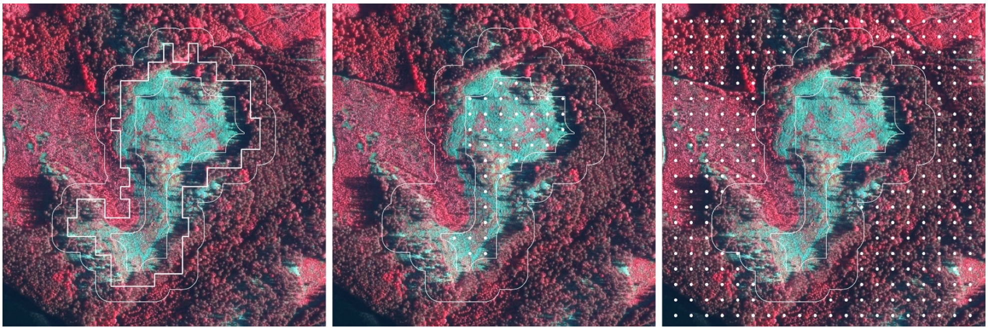

Fig. 7: CIR sample clearing with Hansen vector region and buffer zone (left), clearing sample points (center), forest sample points (right).

2.2 Image Segmentation with Segment Anything Model: from CIR Image Tiles to Images Masks and Vector Outlines

Image segmentation is a common technique in geospatial data analysis, employed to subdivide an image into distinct regions based on a set of specified objects or landscape features (Kotaridis & Lazaridou, 2021). While traditional segmentation techniques require extensive data-specific training and human intervention to produce adequate results, recent advancements in foundational models for computer vision show a promising development for automatic image segmentation tasks.

We use the Segment Anything Model (SAM) (Kirillov et al., 2023) to identify clear-cut areas in our CIR image tiles and generate precise image masks and outlines for the clearings. SAM is described as "a promptable segmentation system with zero-shot generalization to unfamiliar objects and images, without the need for additional training" (Meta AI, 2023). Zero-shot learning is a machine learning approach wherein a model can learn to recognize classes it has not explicitly encountered during training. SAM has not been specifically trained for the segmentation of remote sensing images for land cover classification, nor for the identification of forest clearings. The model has also not been trained to account for the different image calibrations of individual national aerial imagery datasets.

SamGeo, or Segment-Geospatial, is a Python library that advances the capabilities of SAM, specifically for working with geospatial data (Wu & Osco, 2023). The integration of SAM with remote sensing data processing holds significant promise in simplifying previously complex segmentation workflows, as well as making them more accessible. While there is potential for further task-specific improvements and addressing certain limitations, the model delivers notable performance for remote sensing applications (Osco et al., 2023).

SamGeo offers three pre-trained image encoder models, with the most advanced being vit_h or vit-huge—vit being an abbreviation for "vision transformer," a relatively recent but promising development in deep learning for image segmentation tasks (Kotaridis & Lazaridou, 2021). SamGeo employs automatic, prompt-based, or input (point) -based methods to segment images. We use previously generated infrared .tif image tiles and a corresponding list of point coordinates for the clearing as input. The segmentation process is fully automated, requiring minimal to no human input.

Fig. 8: Outputs derived from Meta Segment Anything Model (SAM), image mask (left), masked forest with clearing (center), masked clearing surrounded by forest (right).

We produce image masks as .tif files at the spatial resolution of the input image and subsequently convert the masks to polygon .shp files, outlining the edges of each clearing.

In our CIR image tiles, forest clearings can typically be described through idealized metrics of land cover classification as a clear distinction between two main classes: closed forest cover (forest) and bare ground or soil (clearing). Other observable objects or surfaces encountered in some images include buildings, roads, water surfaces, swamps, or croplands. However, the presence of or proximity to each of these classes could introduce additional variables to further delimit the selection of clearings through prevalent land cover classification or feature vector datasets.

The performance of the model and the accuracy of the output are, in any case, contingent on the input data. The substantial number of image tiles and reference points we produced (Fig. 1) exhibit a considerable variation in both data quality and observed conditions. Data quality may be influenced by various sensor types, calibration standards, and spatial resolution, as well as the kind of surfaces observed, which may influence the general detectability of features and surfaces in the image. On the ground, the process of forest clear-cutting, identified by Hansen GFC as a disturbance in forest patterns, may result in areas that are entirely cleared or display a more subtle removal of individual trees, diminishing the legibility of the figure-ground relationship between forest and clearing.

With this article, we present foundational research on the operational use and application of SAM. While there is potential for fine-tuning SAM for specific tasks and data (Osco et al., 2023), we have not yet conducted additional refining to further align the model with our specific data.

2.3 Processing of Clearing Outlines to Geometric Synthetisations

The segmentation masks generated with SAM are the precursor to a series of geometric operations aimed at deconstructing the clearings. The initial two stages of our model focus on the identification and segmentation of forest clearings using GEE in conjunction with

SAM. This involves the production of infrared image tiles as .tif files and clearing outlines as .shp files. The final step translates formerly unsorted pixel values to vector conclusions that can be further described, categorized, and documented.

In processing a larger volume of clearing outlines, we are faced with a richly diverse array of complex shapes. The final phase of our model is dedicated to working with these clearing outlines using a selection of Python libraries to process simplified, generalized, geometric descriptions for clearings that allow us to collect, categorize, and compare them.

There are numerous Python libraries available for working with .shp geometries in a geospatial data environment, and while GEE and SAM each provide a closed environment with rather unique functionalities, in this final step, we combine a range of independent Python libraries to process the geometries of our clearings. The selection of libraries for this stage may reliably facilitate the outlined purpose, yet the process and products are more experimental compared to the previous two parts, and there is potential for adjustments or additions to the selection of libraries.

Among the selection of Python libraries we use, the most important are Centerline (Todić, 2023), GeoPandas (Van den Bossche et al., 2023), Shapely (Gillies et al., 2024), Rasterio (Gillies & Others, 2013), and Matplotlib (Hunter, 2007).

We employ the Centerline library (Todić, 2023) to process an internal central network of lines from simplified polygon clearing outlines. We then create a set of orthogonal lines by seeding points at predetermined distances along the centerline network. Finally, we sort the orthogonal lines by length and limit their number, first to minimum and maximum lengths, then to keeping them to a maximum of fifteen lines most representative of an evenly spaced distribution of lengths using an iloc indexer and slice notation.

For the visualization and styling of our data products, we utilize Rasterio (Gillies & Others, 2013) and Matplotlib (Hunter, 2007). We plot composite images for each clearing, consisting of infrared .tif tiles as backgrounds overlaid with .shp files for SAM-derived clearing outlines, forest and clearing sample points, and the centerline network, including orthogonal line segments.

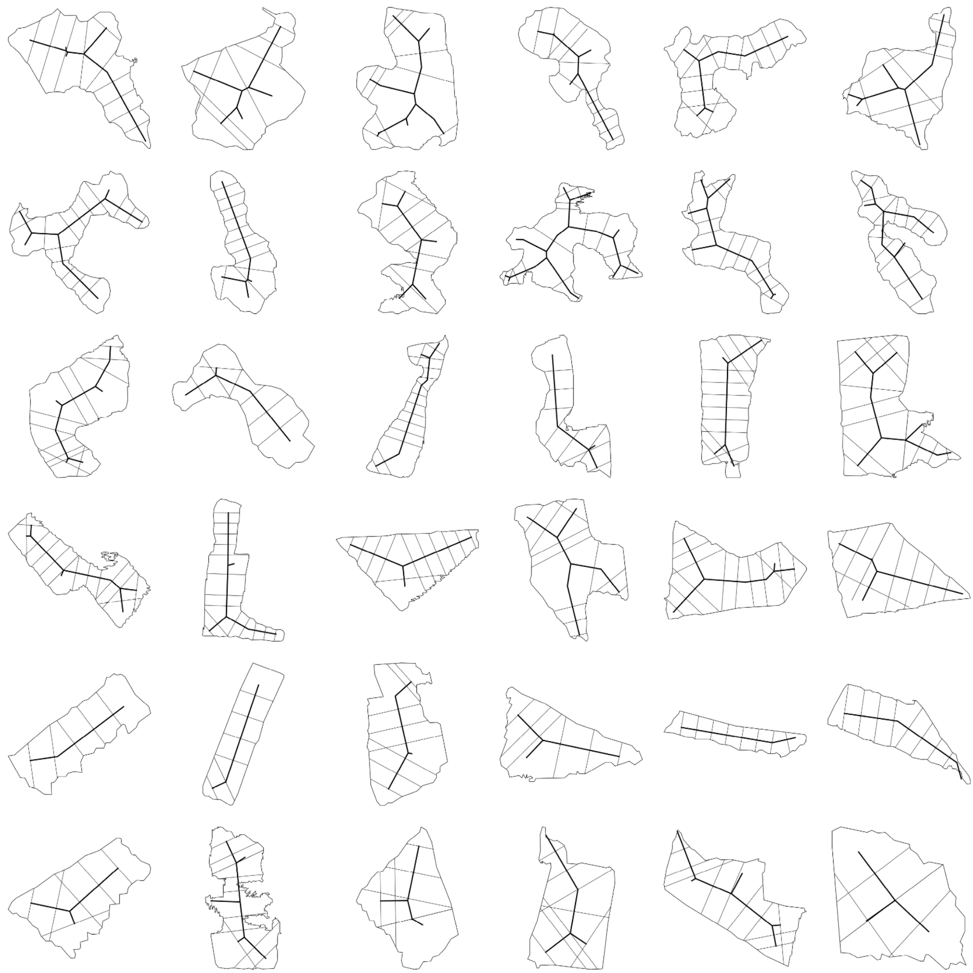

Fig. 9: Derived data products: clearing images, outlines, and geometries; a total of twelve samples in two rows per country, from the top: Finland, Latvia, and Germany.

3 Shapes of Clearings, Value of Shapes & Collection

The heterogeneity of geospatial data environments is an inherent part of the diversification and convergence of technological sensing systems. Assembling data sources, methods, approaches, definitions, and terms may remain an important task for qualifying operative models for the identification of particular landscape features. We provide a model that aims for convergence, overcoming several limitations of datasets, processing environments, or class variables. There is clear potential for further refinement, exchange or addition of individual segments, or expanding the capabilities of the model to alter the target or outcome entirely.

The methodology enables automation of clearing detection across multiple datasets. Our model uses Google Earth Engine to harness the global-scale analysis capabilities of the Hansen GFC dataset and advance its resolution-based limitations by employing several national high-resolution aerial imagery datasets. The intention is to translate temporal patterns of forest disturbance detected through Hansen data to concrete, spatially defined forest clearing sites.

The rapid processing offered by our model, using pre-existing datasets, demonstrates that the geometric descriptions of forest clearings as landscape entities can be perceived as versatile, as data is agnostic to cultural references, which might come from different sources depending on which country we are looking at, rather than a closed dataset. Since the data are stored on various external servers, the model becomes the only point of contact with the data; the images serve as the interface to the model. No data, apart from the final data composition images, have been downloaded as part of the process, and there may be no need to create a copy of the original data in one central location as long as the individual input data remains online and the model is able to produce images that capture and describe the clearings.

The large-scale identification and collection of clearings have revealed the differences between forest clear-cutting and the forest clearing. In literature, these differences may not be as clearly pronounced as we conceive them through the multitude and vast array of images we have produced. In addition to our previous observations, the idealized clearing can be further described through the following:

1. an open space surrounded not only by a row of trees but also by a buffer region of dense forest cover,
2. the clearing area is cleared at once and not part of multiple years of clear-cutting in the same area, resulting in the growth of new trees in previously cleared areas, and the same applies to the density of the surrounding forest areas—both are flat,
3. there is a certain uniqueness to clearings in the way that they are usually not portrayed as a network or a group of clearings that are nearby one another, but rather perceived as the singular transition from forest to open space,
4. a clearing may partially overlap with other spatial typologies, such as roads, agricultural fields, lakes, or rocks, but should not be dominated by them as the notion of enclosure is still important in preserving the character of the clearing, allowing for the possibility to enter the clearing by crossing a threshold between inside and outside,
5. the clearing is independent of the type of trees it is surrounded by or the type of ground that is revealed as part of its making, which is a condition true for the specific case of our study areas.

Classification and segmentation models commonly used in the geosciences serve to determine landscape typologies and observe transformations. However, they may be inherently limited in detecting fuzzy or rather unstable objects and surfaces, especially when discussed only within specific fields. Recent advancements in LLMs having achieved to remove thresholds in accessing scripting and software programming may suggest that SAM could likewise eliminate constraints for building custom classification models. This development would move from a definite set of classes to nearly unlimited input variables. Land cover classifications may be extended from the boundary of a particular classification system to the boundaries of language to define and describe spatial variables. This may provide the opportunity to respond to previous critiques on land cover classification systems (M'-Closkey & VanDerSys, 2022) and allow for a more pronounced integration of complex descriptions of landscapes as part of technological systems that sense, analyze, and describe landscapes.

Despite our systematic approach, aimed at designing a robust model for the described task, some level of processing noise is expected in each outcome. Incongruencies between ground truth and model outcomes may arise from false positive points resulting from the coarse spatial resolution of the Hansen dataset, the temporal difference between the one-year product of the Hansen dataset and the acquisition date of the aerial imagery, sparse forests, or thick undergrowth beneath trees, leading to a less clear distinction between forest and clearing.

Fig. 10: Large-scale t-SNE grid of 33,124 clearing image tiles extracted from the Finland CIR dataset (left), zoom-in of image tiles with clearings bordering lakes (center), zoom-in of image tiles with clearings fully enclosed by forest (right).

4 Conclusion

The research develops a novel model for the localization, identification, and geometric valuation of previously undescribed and undocumented landscape features on a European scale. The study expands upon the spatial and spectral scope of our previous research on forest clearings (Schob & Callejas, 2023). This earlier work was limited to a specific region of Northern Europe with relatively homogeneous data sets. By expanding the geographical area of study to encompass different countries, it is possible to discuss the value of forest clearings as landscape heritage at a European level while including disparate data sets with different levels of spatial and temporal resolution. We not only broaden the research scope through introducing more diverse data sets, but also trace the distinct shapes of clearings and their underlying formal and temporal dynamics.

The cloud-based Python development environment Google Colab (GC) serves as a central interface to integrate the cloud computing capabilities and extensive data catalog of Google Earth Engine (GEE) with efficient image segmentation based on Segment Anything Model (SAM) and a selection of Python libraries for geometric processing. Furthermore, the combination of multiple data sources, methods, and tools grants us the opportunity to overcome initial thresholds of separate data collections, processing environments, APIs, constraints in resolution, and issues of data interoperability. Our method is based entirely on open access—publicly available geospatial datasets, open-source software packages and libraries, and free cloud-based environments and services.

Hansen Global Forest Change (Hansen GFC) data is utilized to detect forest cover disturbances on a global scale and by the year of the gross forest cover loss event. Hansen GFC thus becomes the preliminary filter for identifying clearings in national, high-resolution, color infrared (CIR), orthoimagery acquisitions. For each clearing, we produce square CIR image tiles, accompanied by a list of sample points for forest and clearing regions. Segment Anything Model and SamGeo are employed to delineate clear-cut areas from our image tiles by using a point-based method based on our reference coordinates to produce high-resolution alpha masks and precise shape outlines. To collect, assess, categorize, and document the complex shapes of these outlines, we engage with a selection of Python libraries to create abstract, generalized geometric descriptions for each clearing. The resulting composite data images consist of the infrared .tif tiles as backgrounds overlaid with .shp files for SAM-derived clearing outlines, forest and clearing sample points, and centerline networks, including orthogonal line segments.

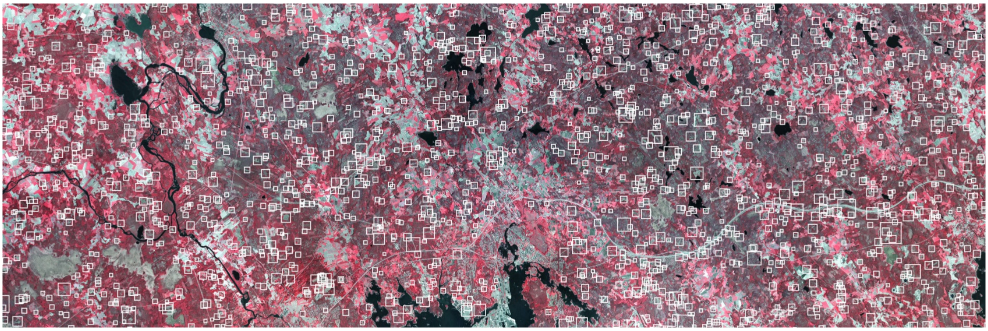

Fig. 11: Identification of clearings throughout the CIR imagery landscapes of Finland.

Our model facilitates the rapid accumulation of a vast number of clearing image tiles and derivative data products. Although our model is largely automated, human pattern recognition skills still play a role at the end of the processing cycle, particularly for the curation of specific sample clearings, as presented in this paper. While we propose a functional model, fine-tuning each individual part may certainly improve the precision of the model outputs. The analysis of tens of thousands of sample image tiles has already enhanced our understanding of the particular definitions of forest clearing and the kind of metrics and variables that may be employed to identify forest clearings on a larger scale and with greater precision.

The valuation of some clearings over others, or the assessment of their geometry in terms of potential heritage or aesthetic value, is still a process that cannot be automated. Human pattern recognition continues to be essential in identifying geometric value. In the future, this study can serve as a precursor in laying the groundwork for defining and detecting certain morphological characteristics automatically. There is potential for both digital heritage and the updating of current physical landscape heritage lists. Digital heritage application is clearer and requires less certainty, with no need for transnational European agreements. In short, today's work is on capturing, valuing and storing entities as images or models, determining their value will be a task to undertake in the future—by the field of digital archeology, for instance. On the other hand, real landscape heritage must adhere to certain European conventions and charters, and this research might serve as an introduction to proposing the clearings as heritage sites to the pertinent authorities.

The identification of clearings is data-dependent but not data-specific. By proposing a multi-resolution data stack comprising different types of inputs and products, we build upon our previous study, which has used lidar data to identify clearings (Schob & Callejas, 2023). Even though there is a possibility that contemporary analysis techniques may soon be rendered obsolete, state-of-the-art techniques are in parallel with the moment in which forest clearings must be understood as potential landscape entities with heritage value. We believe this research is timely regarding how the study aligns with the propositions and demands of the European Parliament on forestry and forested land.

We demonstrate that clearings, as remnants of an important historical practice of forest management, are not only identifiable as individual entities in specific regions with particular definitions of datasets but also traceable across the complex configurations of European forests, which are among the many possible applications of this knowledge, allowing us to recognize them as a tangible part of European cultural landscape heritage. The images of selected clearings themselves may become objects of digital heritage value, as actual clearings will inevitably change or disappear over time. Similar to other cultural landscape or archaeological practices, digital heritage could be another emerging field in which these techniques are deployed to capture the formal attributes of human impact on the surface of the earth—in this case, the subtractive impacts over vegetated masses dominated by the presence of trees.

Professionals involved in developing novel descriptions for (parts of) landscapes may increasingly engage with an array of techniques, tools, and data (Girot, 2013, 2019), as outlined in our methodology. The goal is to create custom models and definitions for particular landscape features, surfaces, or objects. Future models for remote sensing data may be oriented less toward task-specific specializations and offer a broader range of processing tasks in a more accessible manner instead, enabling more articulate and seamless transformations of sensor data from radiance to geometry.

References

BERRIZBEITIA, A. (2024). Forest Futures. Harvard University GSD, Druker Gallery, Cambridge, MA, USA.

CALLEJAS, L. (2021a). The Forest Clearing Archetype. Oculs Reseach Center Web. http://oculs.no/projects/object-as-ground-as-landscape/ (24.02.2024)

CALLEJAS, L. (2021b). Pan Scroll Zoom 16: Luis Callejas, The Forest Clearing as Archetype. https://drawingmatter.org/pan-scroll-zoom-16-luis-callejas (24.02.2024)

CALLEJAS, L., & HANSSON, C. (2022). The archetypal clearing, or the form of an idea. In Reimagining the Civic (Print). Yale School of Architecture.

EUROPEAN COMMISSION. (2020). EU Biodiversity Strategy for 2030 (52020DC0380).

EUROPEAN COMMISSION. (2021). New EU Forest Strategy for 2030 (52021DC0572).

FORMAFANTASMA. (2021). Formafantasma Cambio (R. Badano, R. Lewin, & N. Grabowska, Eds.). König, Walther.

GILLIES, S. & OTHERS. (2013). Rasterio (1.3.9) [Computer software]. Mapbox. https://github.com/rasterio/rasterio (24.02.2024)

GILLIES, S., VAN DER WEL, C., VAN DEN BOSSCHE, J., TAVES, M. W., ARNOTT, J., WARD, B. C., & OTHERS. (2024). Shapely (2.0.3) [Computer software]. https://doi.org/10.5281/ZENODO.5597138

GIROT, C. (2013). The Elegance of Topology. In C. Girot, A. Freytag, & A. Kirchengast (Eds.), Landscript 3: Topology. Jovis.

GIROT, C. (2019). "Cloudism": Towards a new culture of making landscapes. In Routledge research companion to landscape architecture (pp. 113–123). Routledge.

GOOGLE LLC. (2024a). Google Colaboratory. https://colab.research.google.com (24.02.2024)

GOOGLE LLC. (2024b). Google—About Google. https://about.google (24.02.2024)

GORELICK, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment, 202, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

HANSEN, M. C., POTAPOV, P. V., MOORE, R., HANCHER, M., TURUBANOVA, S. A., TYUKAVINA, A., THAU, D., STEHMAN, S. V., GOETZ, S. J., LOVELAND, T. R., KOMMAREDDY, A., EGOROV, A., CHINI, L., JUSTICE, C. O., & TOWNSHEND, J. R. G. (2013). High-Resolution Global Maps of 21st-Century Forest Cover Change. Science, 342(6160), 850–853. https://doi.org/10.1126/science.1244693

HANSEN, M. C., POTAPOV, P. V., MOORE, R., HANCHER, M., TURUBANOVA, S. A., TYUKAVINA, A., THAU, D., STEHMAN, S. V., GOETZ, S. J., LOVELAND, T. R., KOMMAREDDY, A., EGOROV, A., CHINI, L., JUSTICE, C. O., & TOWNSHEND, J. R. G. (2023). Hansen Global Forest Change v1.10. https://developers.google.com/earth-engine/datasets/catalog/UMD_hansen_global_forest_change_2022_v1_10 (24.02.2024)

HUNTER, J. D. (2007). Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering, 9(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

JENSEN, J. R. (2014). Remote sensing of the environment: An earth resource perspective (Second edition). Pearson.

KIRILLOV, A., MINTUN, E., RAVI, N., MAO, H., ROLLAND, C., GUSTAFSON, L., XIAO, T., WHITE-HEAD, S., BERG, A. C., LO, W.-Y., DOLLÁR, P., & GIRSHICK, R. (2023). Segment Anything (arXiv:2304.02643). arXiv. http://arxiv.org/abs/2304.02643

KOTARIDIS, I., & LAZARIDOU, M. (2021). Remote sensing image segmentation advances: A meta-analysis. ISPRS Journal of Photogrammetry and Remote Sensing, 173, 309–322. https://doi.org/10.1016/j.isprsjprs.2021.01.020

LANDESVERMESSUNG UND GEOBASISINFORMATION BRANDENBURG. (2023a). Brandenburg (Germany) RGBN orthophotos 20 cm. Earth Engine Data Catalog. https://developers.-google.com/earth-engine/datasets/catalog/Germany_Brandenburg_orthos_20cm (24.02.2024)

LANDESVERMESSUNG UND GEOBASISINFORMATION BRANDENBURG. (2023b). Luftbilder aktuell | LGB_Startseite. https://geobasis-bb.de/lgb/de/geodaten/luftbilder (24.02.2024)

Latvijas Geospatial Information Agency. (2023a). Orthophoto maps | Latvian Geospatial Information Agency. https://www.lgia.gov.lv/lv/ortofotokartes-1 (24.02.2024)

Latvijas Geospatial Information Agency. (2023b, December 22). Latvia Color InfraRed (CIR) orthophotos. Earth Engine Data Catalog. https://developers.google.com/earth-engine/datasets/catalog/Latvia_Maamet_orthos_cir (24.02.2024)

LCLA OFFICE, CALLEJAS, L., & SCHOB, M. (2021). The Forest Clearing Archetype. arc en rêve centre d'architecture, Bordeaux, France.

LCLA OFFICE, CALLEJAS, L., & SCHOB, M. (2023). The Forest Clearing Archetype. Bergen School of Architecture, Bergen, Norway.

LCLA OFFICE, CALLEJAS, L., & SCHOB, M. (2024). The Forest Clearing Archetype. Harvard University GSD, Druker Gallery, Cambridge, MA, USA.

M'CLOSKEY, K., & VANDERSYS, K. (2022). Behind-the-Scenes: Multispectral imagery and land cover classification. Journal of Landscape Architecture, 17(1), 22–37. https://doi.org/10.1080/18626033.2022.2110417

META AI. (2023). Segment Anything | Meta AI. https://segment-anything.com/ (24.02.2024)

MOSSE, R. (2022). Broken Spectre. Loose joints.

NATIONAL LAND SURVEY OF FINLAND. (2023a). Finland RGB NLS orthophotos 50 cm by SMK. Earth Engine Data Catalog. https://developers.google.com/earth-engine/datasets/catalog/Finland_SMK_V_50cm (24.02.2024)

NATIONAL LAND SURVEY OF FINLAND. (2023b). NLS Orthophotos Documentation. https://www.maanmittauslaitos.fi/en/maps-and-spatial-data/datasets-and-interfaces/product-descriptions/orthophotos (24.02.2024)

OSCO, L. P., WU, Q., DE LEMOS, E. L., GONÇALVES, W. N., RAMOS, A. P. M., LI, J., & MAR-CATO, J. (2023). The Segment Anything Model (SAM) for remote sensing applications: From zero to one shot. International Journal of Applied Earth Observation and Geoinformation, 124, 103540. https://doi.org/10.1016/j.jag.2023.103540

SCHOB, M., & CALLEJAS, L. (2023). Landscapes between Signal and Data: Formal Identification and Analysis of Forest Clearings in Oslo through Lidar Data. Journal of Digital Landscape Architecture. https://doi.org/10.14627/537740024

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. ISPRS Journal of Photogrammetry and Remote Sensing, 164, 152–170. https://doi.org/10.1016/j.isprsjprs.2020.04.001

TAVARES, P. (2010). The Political Nature of the Forest: A Botanic Archaeology of Genocide. In U. K. Le Guin (Ed.), The word for world is still forest. TOR.

TODIĆ, F. (2023). Centerline (1.0.1) [Computer software]. https://centerline.readthedocs.io/en/latest/ (24.02.2024)

VAN DEN BOSSCHE, J., JORDAHL, K., FLEISCHMANN, M., MCBRIDE, J., WASSERMAN, J., RICHARDS, M., GARCIA BADARACCO, A., SNOW, A. D., TRATNER, J., GERARD, J., ´WARD, J., PERRY, M., FARMER, C., HJELLE, G. A., TAVES, M., TER HOEVEN, E., COCHRAN, M., RRAY-MONDGH, GILLIES, S., … BILOGUR, A. (2023). GeoPandas (v0.14.1) [Computer software]. Zenodo. https://doi.org/10.5281/ZENODO.2650956

WU, Q. (2020). geemap: A Python package for interactive mapping with Google Earth Engine. Journal of Open Source Software, 5(51), 2305. https://doi.org/10.21105/joss.02305

WU, Q. (2024). Sharing Work in Earth Engine: Basic UI and Apps. In J. A. Cardille, M. A. Crowley, D. Saah, & N. E. Clinton (Eds.), Cloud-Based Remote Sensing with Google Earth Engine (pp. 603–627). Springer International Publishing. https://doi.org/10.1007/978-3-031-26588-4_30

WU, Q., & Osco, L. P. (2023). samgeo: A Python package for segmenting geospatial data with the Segment Anything Model (SAM). Journal of Open Source Software, 8(89), 5663. https://doi.org/10.21105/joss.05663