Multi-Source Synthesis, Harmonization, and Inventory of Critical Infrastructure and Human-Impacted Areas in Permafrost Regions of Alaska (SIRIUS)

Soraya Kaiser[1,2], Julia Boike[1,2], Guido Grosse[1,3], and Moritz Langer[1,4]

[1]Permafrost Research Section, Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, Telegrafenberg A45, 14473 Potsdam, Germany

[2]Geography Department, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany

[3]Institute of Geosciences, University of Potsdam, Karl-Liebknecht-Str. 24-25, 14476 Potsdam, Germany

[4]Department of Earth Sciences, Vrije Universiteit Amsterdam, De Boelelaan 1085, 1081 HV Amsterdam, Netherlands

Correspondence: Soraya Kaiser (soraya.kaiser@awi.de)

Abstract. The Arctic region has undergone warming at a rate more than three times higher than the global average. This warming has led to the degradation of near-surface permafrost, resulting in decreased ground stability. This instability not only poses a primary ~~threat~~ hazard to Arctic infrastructure and human-impacted areas, but can also lead to secondary ecological hazards from infrastructure failure associated with hazardous materials. This development underscores the need for a comprehensive inventory of critical infrastructure and human-impacted areas~~, that is~~. The inventory should be linked to environmental data to assess their susceptibility to permafrost degradation as well as the ecological consequences that may arise from infrastructure failure. ~~In this study~~Here, we provide such an inventory for Alaska, a vast state covering approximately 1.7 million km^2, with a population of over $733,000$ people and a history of industrial development on permafrost. Our ~~SIRIUS inventory~~ Synthesized Inventory of CRitical Infrastructure and HUman-Impacted Areas in AlasSka (SIRIUS) integrates data from (i) the Sentinel-1/2 derived Arctic Coastal Human Impact dataset (SACHI), (ii) OpenStreetMap, (iii) the pan-Arctic Catchment Database (ARCADE), (iv) a dataset of permafrost extent, probability and mean annual ground temperatures, and (v) a contaminated sites database and reports to create a unified new dataset of critical infrastructure and human-impacted areas as well as permafrost and watershed information for Alaska. The integration ~~steps involved~~ process included harmonizing spatial references, extents, and geometries ~~, the usage of~~ across all datasets, as well as incorporating a uniform usage type classification scheme for the infrastructure data. Additionally, we employed text mining techniques to generate ~~additional geospatial data~~ supplementary geospatial data from textual reports on contaminated sites~~ – including~~ , including details on contaminants, cleanup duration, and affected ~~medium – from textual reports, and the incorporation of a uniform usage type classification scheme for infrastructure~~media. The combination of SACHI and OSM enhanced the detail of the usage type classification for infrastructure from 5 to 13 categories, ~~which allows for~~ allowing the identification of elements critical to Arctic communities beyond industrial sites. Further, the new inventory ~~unites~~ integrates the high spatial ~~accuracy from~~ detail of OSM with the ~~high level of completeness from SACHI~~unbiased infrastructure detection capability of SACHI, accurately representing 94 % of polygonal infrastructure and 78 % of linear infrastructure, respectively. The SIRIUS dataset is presented as a GeoPack-

age, enabling spatial analysis and queries of its components, either in dependence or combination with one another. The dataset is available on Zenodo under DOI 10.5281/zenodo.8311242.

1 Introduction

In the past decades, the Arctic has experienced a pronounced warming, entailing an increase in air temperature that is more than three times as high as the global average (Rantanen et al., 2022), referred to as Arctic Amplification (Cohen et al., 2014). These increasing air temperatures led to a warming and thawing of permafrost since the 1980s, as borehole measurements across the Arctic demonstrate (Biskaborn et al., 2019; Smith et al., 2022). Modeling studies indicate that the initiation of permafrost warming can be traced back to as early as 1900 (Langer et al., 2024). As 15 % of the exposed land surface of the Northern Hemisphere are underlain by permafrost (Obu et al., 2019), this warming trend affects a vast area and has major implications for ecosystems and livelihoods in the Arctic and subarctic. With permafrost degrading, we not only expect the mobilization of one of the largest soil carbon pools (Schuur et al., 2015, 2022), but also substantial land surface changes that result from ground subsidence and thermal erosion (Kokelj and Jorgenson, 2013). Permafrost warming trends can also be observed in mountain regions worldwide (Biskaborn et al., 2019), leading to the destabilization of slopes and increased movement of rock glaciers (Haeberli, 2013; Haeberli et al., 2024). Numerous studies demonstrate intensifying land surface changes in the permafrost region which encompass for example processes such as thaw slumping (e.g. Runge et al., 2022; Ramage et al., 2017; Leibman et al., 2021), the development of thermokarst ponds and lakes (e.g. Muster et al., 2017; Jones et al., 2011), thermo-erosional gullying (e.g. Fortier et al., 2007; Godin et al., 2012), ice wedge degradation (e.g. Liljedahl et al., 2016; Jorgenson et al., 2006) and ~~rock avalanches~~ mass movement processes such as rock avalanches and falls in mountainous regions ~~(e.g. Bessette-Kirton and Coe, 2020; Smith et al., 2023)~~ (e.g. Bessette-Kirton and Coe, 2020; Smith et al., 2023; Stoffel et al., 2024) all pointing to an increasing loss in ground stability. Some of these processes such as thaw slumps have impacts not just locally but even far away in downstream areas as sediments, solubles, and organic matter are eroded from thaw features and may follow different trajectories of transport, biogeochemical processing, and sedimentation depending on environmental conditions (Lamhonwah et al., 2016; Keskitalo et al., 2021; Kokelj et al., 2013) and can also impact ecosystems in these downstream areas (Levenstein et al., 2020).

For Arctic settlements, the destabilization of the ground can cause severe infrastructure failure. Damage to housing units, transport networks (roads and airstrips), and water supply and sewage systems are frequently reported (Liew et al., 2022). Degradation of permafrost also ~~threatens~~ poses a hazard to industrial infrastructure, including sites relevant for ~~e.g.~~ natural resource extraction ~~, energy production , and further processing~~ and energy production whose failure can result in environmental contamination (Rajendran et al., 2021; Langer et al., 2023). With the expansion of human activities and infrastructure development in the Arctic (Bartsch et al., 2021), increasing human-induced effects on snow and vegetation, as well as permafrost degradation, are observed in their vicinity, which further accelerates the destabilization of the ground (Walker et al., 2022; Bergstedt et al., 2022; Raynolds et al., 2014; Hammar et al., 2023). Model projections focusing on RCP 4.5 (Representative Concentration Pathways) (van Vuuren et al., 2011) indicate that approximately 69 % of Arctic infrastructure will face impacts of near-surface permafrost degradation by 2050 (Hjort et al., 2018). This will influence the lives of about 5 million people living in ~~over~~ more than 1000 settlements across the Arctic permafrost region (Ramage et al., 2021) (see Figure 1a). Given the potential impact of near future permafrost degradation, it becomes imperative to generate comprehensive inventories of critical Arctic infrastructure and areas of human activity, allowing the assessment of their specific usage types, potential to failure, and relevance to local and regional livelihoods. Such an inventory is a prerequisite for determining exposure to natural hazards such as thaw induced ground destabilization, coastal erosion, and flooding which is pivotal to risk assessments.

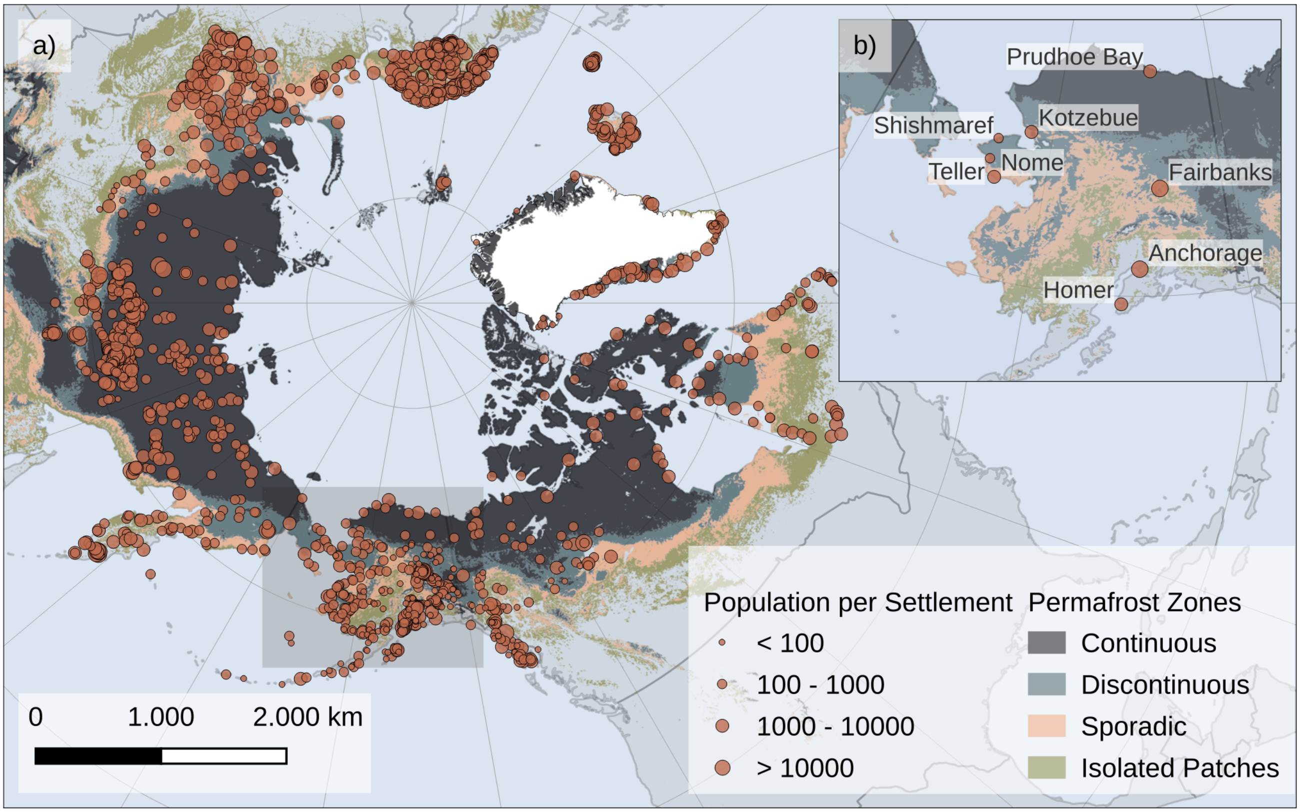


Figure 1. ~~Pan-Arctic~~ Figure a) shows pan-Arctic permafrost extent as modeled by Obu et al. (2019) together with population numbers of settlements in the Arctic Circumpolar Permafrost Region (ACPR) (Wang et al., 2021). The different sizes of the circles represent logarithmic scaling of the population numbers. Our study focuses on the state Alaska as shown in inset map b). Basemap was made with Natural Earth: Free vector and raster map data @www.naturalearthdata.com.

Therefore, substantial efforts are being made to map settlements, areas of human activity, and industrial sites throughout the Arctic. Extensive databases have ~~already~~ been compiled regarding population numbers (Wang et al., 2021; Ramage et al., 2021), the occurrence and development of infrastructure along coastlines (Bartsch et al., 2020, 2021), and the distribution of industrial sites in the Arctic (Langer et al., 2023). The datasets focusing on Arctic infrastructure in particular and areas of human activities in general, however, are limited in spatial coverage (coastal areas, north of treeline (e.g. Bartsch et al., 2021; Xu et al., 2022)), spatial resolution and lack specific detail regarding usage type. Furthermore, because of their diverse research approaches, these datasets ~~do not exhibit consistency in~~ are inconsistent with respect to spatial references and geometry types (vector/ raster). To date, there is no comprehensive inventory that synthesizes various information about infrastructure and areas of human activity in the Arctic and combines these information with essential environmental data such as permafrost occurrence and watersheds. In addition, for Canada and the U.S. there is a substantial volume of state and federal data on contaminated sites available (Langer et al., 2023). However, the geospatial data provided by government agencies is highly heterogeneous, offering the full range of detailed site chronologies (e.g. affected containment structures, mandated cleanup measures) as well as data about the polluting substances to sometimes only basic information about location, cleanup status, and responsible personnel. Additional details can then be found in written reports (Langer et al., 2023; State of Alaska Department of Environmental Conservation, 2023a) and each have to be extracted first, before they can be put into a spatial context. However, this detailed information is urgently required in a geospatial data format over large regions, not only to estimate the vulnerability of critical infrastructure and human-affected areas to permafrost degradation but also to assess the ecological consequences of contamination resulting from industrial infrastructure site failures.

Focusing on Alaska, we thus (i) harmonised existing multi-source data on infrastructure and human-impacted areas into a coherent usage type classification scheme, (ii) created a statewide inventory of these elements and enriched it with data on permafrost characteristics (extent, probability, and ground temperatures), watersheds, and sites of contamination, for which we extracted information on contaminants, cleanup duration, and the affected medium from available text reports, and (iii) enabled the spatial analysis and queries of the inventory together with ecological information in a database-like structure.

Following the CIIP manual (Critical Information Infrastructure Protection, CIIP2008) (Brunner and Suter, 2008), we define critical infrastructure as those sectors essential for the reliable functioning of communities. These core categories include among others food and water supply, and health and sanitation. To better align with the modern and traditional ways of life in the Arctic and subarctic region, we have adjusted the internationally recognized core categories and extended them~~, as elaborated in Section 2.2.1~~. Please refer to Section 2.2.1 (Infrastructure Usage Types) and Table 1 for a full list of categories.

2 Material & Methods

2.1 Study Site

Alaska is the largest and northernmost state ~~in the~~ of the United States of America (U.S.). With a population of over 733,000 inhabitants and a land area of approx. 1.7 million km^2 (The Information Architects of Encyclopaedia Britannica, 2023), it is also the least densely populated state in the U.S., with a population density of 0.5 people per square kilometer (1.3 people per square mile), compared to the rest of the U.S. with a density of 35.9 per square kilometer (93 people per square mile) (Department of Labor and Workforce Development, 2020; World Bank, 2024). Alaska is home to over 300 communities, with Anchorage, Juneau and Fairbanks City being the biggest municipalities, housing 49 % of the overall population. The other near half of the population (44 %) resides in smaller settlements with fewer than 10,000 people (Department of Labor and

Workforce Development, 2020), dispersed across the entire state. Many of these smaller settlements are only reachable by air or barge (Hamilton et al., 2016).

~~Naturally, with its vast expanse, it includes~~ Alaska encompasses a range of different landscapes, from glaciers in the Brooks Range to tundra in the North Slope and boreal forests in the Alaska-Yukon region (Raynolds et al., 2019; Jorgensen and Meidlinger, 2015). There are also substantial variations in meteorological and permafrost characteristics, following a North-South gradient. In the North, a cold polar tundra climate (Beck et al., 2018) prevails, with mean annual air temperatures (MAAT) of $-10.4°$ C (Climate Normals 1991-2010 of Deadhorse, see NCEI, 2023a) and a continuous permafrost extent (see Figure 1). The South on the other hand, is still characterized by a cold climate (Beck et al., 2018), but with much higher temperatures ($4.5°$ C MAAT for Homer, see NCEI, 2023b) and a permafrost extent transitioning to a sporadically underlain land surface and isolated patches.

It is important to note that approx. 80 % of the state's area – accounting for nearly 200 settlements (refer to Figure 1) – fall within the permafrost region (Jorgenson et al., 2008; Ramage et al., 2021), which is projected to undergo massive changes in the upcoming decades (Chadburn et al., 2017; McGuire et al., 2018). Challenges such as ground subsidence across the region and coastal erosion along the extensive and ~~highly~~ populated coastline (occupied by 83% of the population (NOAA Office for Coastal Management, 2023)), will pose a high risk to the Alaskan population and economy (Melvin et al., 2016; Liew et al., 2022; Wang et al., 2023).

~~Apart from its value to the global fishing industry (Markon et al., 2018), Alaska has many other industries highly contributing to the economy: transportation and warehousing (including cargo, passengers but also tourism), finance, insurance, real estate, and government and government enterprises (including community services such as military, postal service, etc.) (Bureau of Economic Ana . However, the~~ The most important contribution to Alaska's economy stems from the mining, quarrying, and oil and gas extraction industry (Bureau of Economic Analysis, 2023a). Notably, the oil exploration units in the North Slope and Cook Inlet play a vital role in Alaska's revenue, having contributed 38 % of the general funds in the 2019 fiscal year (Alaska Oil and Gas Association, 2020, 2021). ~~Nevertheless, the~~ In addition to the significant impact of oil and gas, Alaska's fishing industry also plays a crucial role in the economy. The Alaska Seafood Marketing Institute (Alaska Seafood Marketing Institute, 2024) reports that, in 2021/22, the fishing industry employed 17,000 Alaskans (from a total of 48,000 workers) from more than 142 communities, making it the top employer in the Alaskan manufacturing sector. Moreover, more than 60 % of the total U.S. seafood harvest comes from Alaska's fisheries (Alaska Seafood Marketing Institute, 2024). Further industries contributing to the economy are: transportation and warehousing (including cargo, passengers but also tourism), finance, insurance, real estate, and government and government enterprises (including community services such as military, postal service, etc.) (Bureau of Economic Analysis, 2023a, b). However, the economic growth comes with environmental consequence. The continued development of infrastructure, expansion of human-impacted areas and oil exploration sites in the North, along with the associated transportation and infrastructure networks, have already led to an increase in thermokarst occurrence (Raynolds et al., 2014; Walker et al., 2022). Furthermore, given the extensive oil and gas production operations, there is an inherent risk of environmental contamination resulting from infrastructure failures. This, in conjunction with both natural and human-induced degradation processes, underscores the need for a comprehensive and freely accessible database encompassing critical infrastructure and human-impacted areas on one hand and environmental information concerning watersheds and permafrost on the other.

2.2 Data Harmonization & Mining

The SIRIUS (**S**ynthesized **I**nventory of C**R**itical **I**nfrastructure and H**U**~~uman-Impacted~~ man-Impacted Areas in Alas**S**ka) dataset synthesizes data from five different sources: ~~(i) the~~

(i) Sentinel-1/2 derived Arctic Coastal Human Impact dataset (SACHI) (Bartsch et al., 2021) ~~, (ii)~~(acquired June 11, 2021),

(ii) OpenStreetMap dataset for the infrastructure and land use information (OpenStreetMap Contributors and Geofabrik GmbH, 2018) ~~, (iii)the pan-Arctic~~ (acquired January 20, 2023),

(iii) Pan-Arctic Catchments Database (ARCADE) for the watersheds (Speetjens et al., 2022) ~~, (iv)the modeled~~ (acquired January 17, 2023),

(iv) Modeled Northern Hemisphere permafrost map by Obu et al. (2018) ~~, and (v)the contaminated~~ (acquired August 31, 2023), and

(v) Contaminated sites database and reports by the State of Alaska Department of Environmental Conservation (2023a) (DEC) ~~. After acquiring the latest updates (see Table ?? in the Appendix) of these individual spatial datasets, the~~ (acquired March 2, 2023).

The primary task was to harmonize them to create a semantically and geometrically coherent and uniform data product (see Figure 2). Initially a thorough homogenization of the spatial reference was required. All datasets were reprojected to the the World Geodetic System 1984 with an Alaska polar stereographic map projection (EPSG Code 5936). Subsequently, we clipped every dataset's spatial extent to the state boundary of Alaska as provided by the National Weather Service (2023). Each dataset had to undergo further geometric harmonization processes such as merging individual vector files, creating buffer zones along linear features, and clipping to layer spatial extents. Thereafter, we performed spatial analyses such as spatial overlays and joins to determine overlapping features and retrieve their information. ~~Detailed information on the dataset's content and applied processing steps are explained in detail in the following sections.~~ All data processing was done using Python with its geospatial data processing libraries geopandas, pandas, numpy, ~~osgeo~~gdal, rasterio, and rioxarray (Jordahl et al., 2022; pandas development team, 2023; Harris et al., 2020; Rouault et al., 2023; Gillies et al., 2013–; Rio, 2024) . The data processing scripts are downloadable from our Zenodo repository (Kaiser et al., 2023).

2.2.1 Infrastructure and Human-Impacted Areas

Sentinel-1/2 derived Arctic Coastal Human Impact

The Sentinel-1/2 derived Arctic Coastal Human Impact (hereafter SACHI) dataset contains buildings, road and railway networks and other human-impacted areas in the Arctic coastal regions up to 100 km inland (Bartsch et al., 2020). The infrastruc-

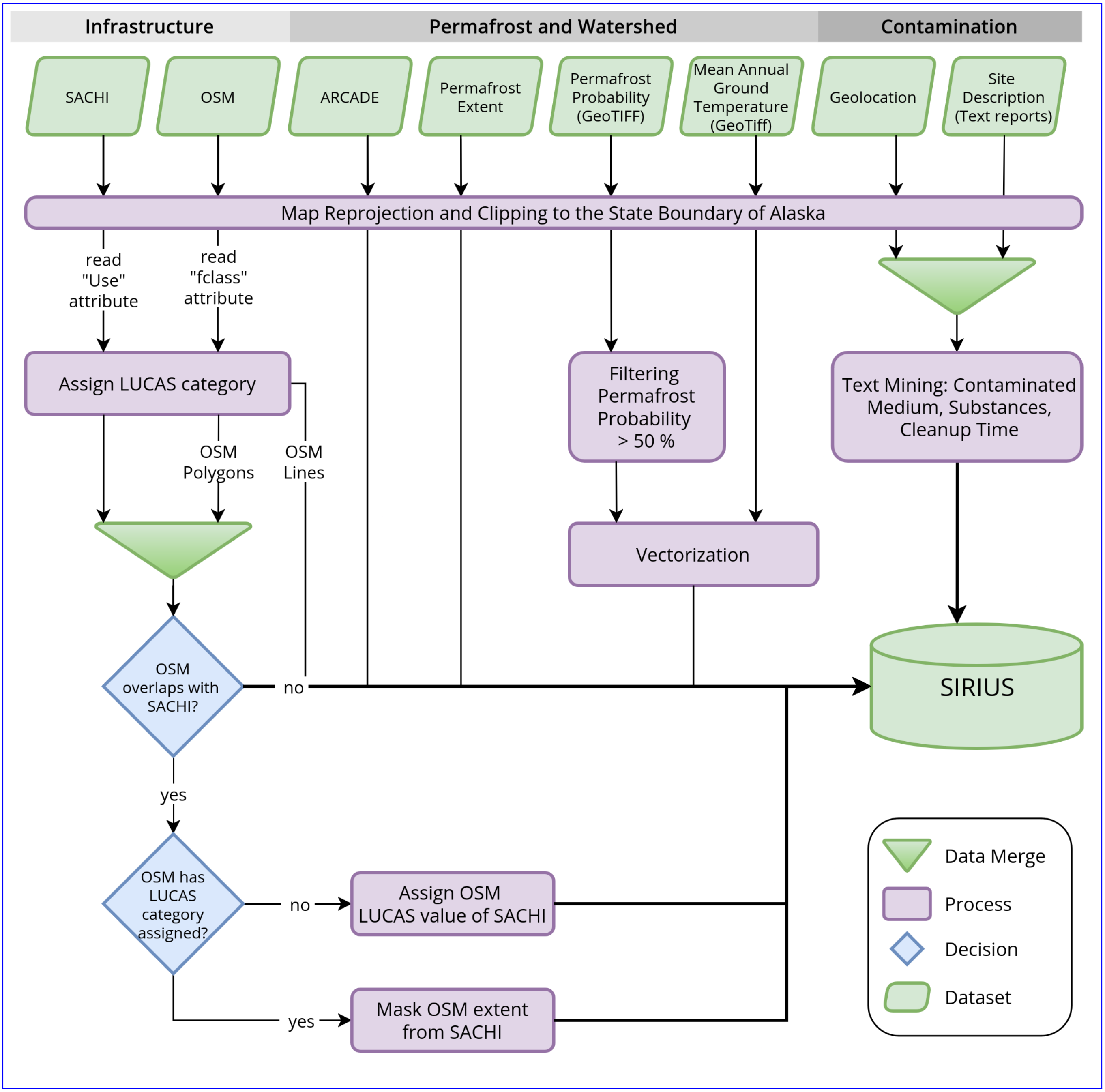


Figure 2. Flowchart of Harmonization Process. If not indicated otherwise, all input datasets are of ESRI Shapefile format.

ture features in SACHI were derived from Sentinel satellite imagery using machine learning and were blended with auxiliary information from other datasets (Bartsch et al., 2021). Each infrastructure feature holds among others information on the settlement name, the feature's class, the primary economic activity (attribute "Use") and the general economic activity (attribute "Use main") (Bartsch et al., 2021). The value of the attribute "settlement name" was assigned on the basis of the settlement dataset by ~~(Wang et al., 2021)~~Wang et al. (2021), with a 40 km buffer applied to also incorporate surrounding infrastructure. Features outside this buffer were labeled following the Google hybrid data layer (Bartsch et al., 2021). Each ~~settlements~~ settlement (and surrounding) was then assigned one economic activity category. This procedure resulted in a rather coarse definition of use categories. For example, the settlement of Nome is assigned the general use category "Mining", with no further distinction, and for the Nome-Teller highway connecting ~~both settlements~~ the settlements Nome and Teller, the southern part (Nome) is assigned "Mining", while the northern part counts towards the "Fishing" industry in Teller City. This generalization does not allow the differentiation of use categories within settlements and beyond. As the SACHI dataset was derived using a pixel-based approach, linear infrastructure is also represented as polygons. The "class" attribute specifies whether a feature corresponds to linear transport infrastructure (class = 1), a building (class = 2), or another human-impacted area (class = 3). When ~~examining~~ we visually examined the linear transport infrastructure, we observed some gaps in the data, particularly in settlements: extracting narrow paths or distinguishing between a linear gravel road and other human-impacted areas, such as driveways or exploration pads, were difficult with the limited spatial resolution of the Sentinel sensors (10 m). In addition, the "road" class showed a particularly low mapping accuracy compared to the "building" class (Bartsch et al., 2021). As OSM on the other hand is estimated to represent 83 % of the global road network (Barrington-Leigh and Millard-Ball, 2017; Hjort et al., 2018), we decided to use OpenStreetMap data to represent the linear transport infrastructure.

OpenStreetMap

The OpenStreetMap (hereafter OSM) project is a collaborative initiative involving mappers from around the globe, aiming to provide highly detailed and comprehensive map data (OpenStreetMap Foundation, 2023). It offers a wide range of geographic features, encompassing various categories such as settlement types (e.g., cities, hamlets, villages), road classifications (e.g., motorways, footways, primary and secondary roads), railway networks, amenities, man-made structures, and more (OpenStreetMap Wiki, 2023). Notably, the road and railway networks in OSM are represented as line features~~, which enables the execution of spatial queries. For instance, it~~. This trait facilitates queries about the total length of road network sections situated on different types of permafrost or within specific catchment areas, as well as the identification of potential contamination along transportation routes. Another advantage of OSM is its data availability for the entire region of Alaska. ~~We acquired the latest OpenStreetMap dataset for Alaska from 20 January 2023 (OpenStreetMap Contributors and Geofabrik GmbH, 2018) (Figure A1). Our focus lay~~ Our focus is on areas (farmland, commercial areas, etc.) and elements (small-scale features such as hunting stands, memorials, etc.) that are directly influenced by human activities and are shaped by practical land use. Therefore, we excluded OSM files which contained information about water bodies and natural features: "waterways" for the linear infrastructure files and "natural" and "water" for the polygonal and point infrastructure files. We also excluded information on the orientation (Buddhist, Jewish, etc.) of religious sites: "pofw" (places of worship). Buildings such as churches, chapels, and burial grounds (cemeteries) were retained. Subsequently, we merged the linear OSM infrastructure files into one dataset.

To assess how the linear OSM infrastructure dataset compares to the pixel-based SACHI dataset, we compared their polygonal representations. For this, we converted the linear OSM infrastructure to polygons by applying a buffer around each linear feature: major highways and roads (OpenStreetMap Wiki, 2023) were assigned a width of 20 m to account for possible embankments, sliproads ~~, ramps, etc.~~or ramps. For the rest of the road network and the railway lines, we assumed a width of 10 m. Subsequently, we clipped the polygonal OSM dataset - representing the linear infrastructure features - to the spatial extent of the SACHI dataset and compared their respective areas to each other.

After merging the linear ~~rail~~ railway and road network OSM data, we combined the polygonal OSM infrastructure data into a single GeoDataFrame. The attribute "fclass" of the polygonal OSM GeoDataFrame contains the tag, which people use to describe the mapped feature. In the OSM Wiki (OpenStreetMap Wiki, 2023), these tags are listed following a certain key and value combination, a mapping standard most members of the community follow. As a first step, we derived the unique values of the attribute "fclass" and compared them to the OSM values defined in the Wiki (OpenStreetMap Wiki, 2023). Generally, the tags under "fclass" were in agreement with the OSM values of the Wiki. Some mismatches originated from different expressions, e.g. „town_hall“ instead of „townhall“, „archaeological“ instead of „archaeological_site“ or „mobile_phone_shop“ instead of „mobile_phone“. Some tags were ~~actual additions from~~ unofficial additions created individually by the OSM mapping community, e.g. parking_multistorey, recycling_paper. Further, we removed any occurring tags describing natural features (waterfalls, etc.) and places (island, heath, village, etc.), which portray localities and their population in which multiple usage types are possible. Table A1 shows the retrieved values of "fclass" and their corresponding OSM keys and values, which we assigned manually following the above mentioned Wiki. The predominant tag under "fclass" was "building“. This tag ~~represents~~ represented 81% of the polygonal OSM dataset. To determine the usage type for these buildings, we analyzed their attribute "osm_type" of the dataset and once again compared the tags under "osm_type" to the OSM keys and values of the OSM Wiki. Having identified all of the tags under "fclass" and "osm_type" and assigned them an OSM key and value, we had gathered information on the features's main usage and purpose and could categorize them into usage categories.

Infrastructure Usage Types

For this, we followed the Land Use / Cover Area frame statistical Survey (LUCAS) of Eurostat (E4.LUCAS (ESTAT), 2018), which provides a framework for a consistent classification and harmonization of land use/ land cover data (see Table 1).

This categorization allows us to incorporate the aspect of sectors critical to the functioning of Arctic communities. ~~While our~~ Our core categories of critical infrastructure align with internationally defined sectors (Brunner and Suter, 2008), which include food and water supply, banking and finance, government services and institutions, transport and mobility, information and communication, energy production, health and sanitation~~, we also introduce two additional~~. In addition, we introduce two supplementary categories: ecological & traditional sustainability, and environmental protection. The latter category refers to any infrastructure that may pose environmental ~~threats~~ hazards in the event of failure. This category is particularly significant for traditional lifestyles, such as hunting and fishing, which we consider within the ecological & traditional sustainability category, as they rely on intact terrestrial and aquatic ecosystems. In this category, we also include sites of cultural heritage (cemeteries, tents, yert, etc., see e.g. Irrgang et al. (2019)).

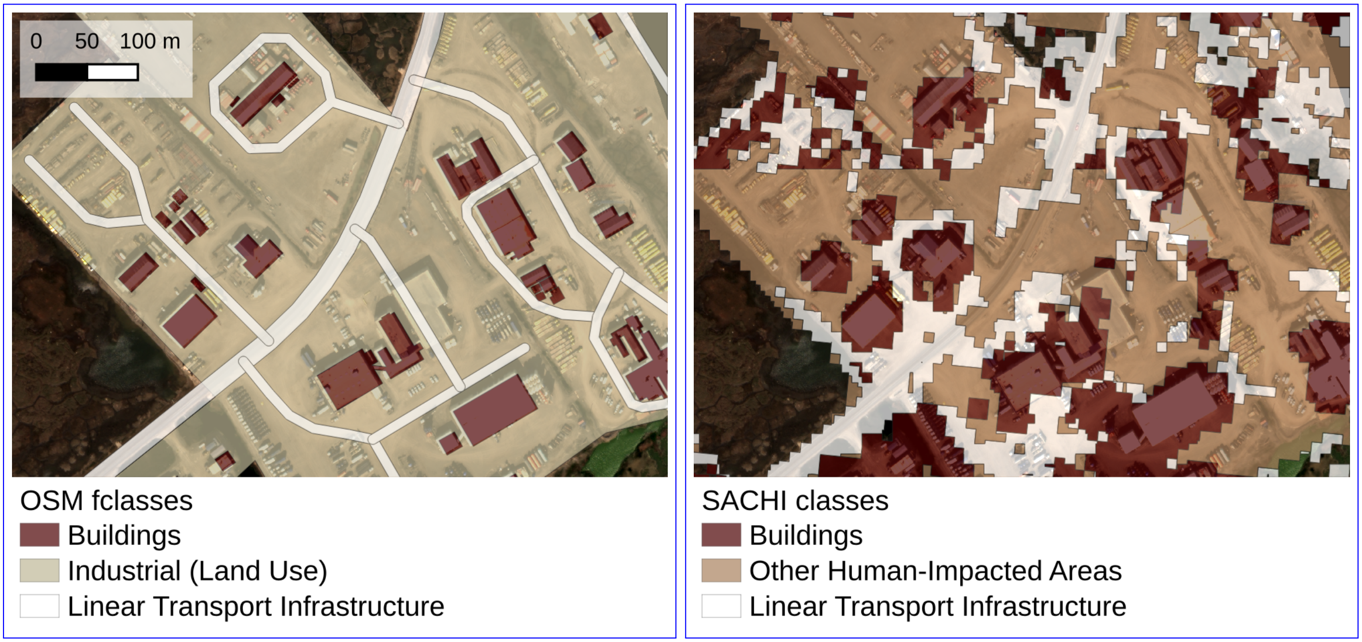
Table 1. LUCAS categories with their respective sectors critical to Arctic and subarctic communities.

Category Nr.	LUCAS	Critical Sector
01	Agriculture	Food Supply
02	Commerce, finance and business	Banking & Finance
03	Community services	Health & Sanitation, Government services, Ecological & Traditional Sustainability
04	Construction	–
05	Energy production	Energy Production
06	Fishing	Ecological & Traditional Sustainability
07	Forestry	Ecological & Traditional Sustainability
08	Hunting	Ecological & Traditional Sustainability
09	Industry and manufacturing	Environmental Protection
10	Mining and quarrying	Environmental Protection
11	Recreational, leisure and sport	–
12	Residential	–
13	Transport, communication networks, storage and protective works	Transport & Mobility, Information & Communication
14	Unused	–
15	Water and waste treatment	Water Supply, Health & Sanitation

Table A~~2~~1 shows the assigned LUCAS category for each OSM tag. As the linear OSM data only consists of railway and road network data, no further classification was needed.

After implementing the initial assignment based on the given scheme, we noticed that all of the tags under "fclass" were effectively categorized except for one: the "building" tag posed a challenge as the corresponding "osm_type" attribute lacked detailed information on the usage type for 86 % out of $144,000$ building features. To address this, we ~~internally overlaid these unknown usage type building with the known usage type~~ sub-sampled the features with the "fclass" building that hadn't been assigned a usage type yet and "internally" overlaid them with features of any other "fclass" (other than building) that already had a usage type assigned. We then assigned the usage type of the non-building ~~features and assigned their tag for~~ feature to the building feature in the overlapping areas. This analysis revealed that the ~~buildings tag frequently features various~~ features with the tag building (e.g. a shopping mall) frequently contain various smaller features and, thus, usage types, such as shops, offices, parking areas, and more. To harmonize this, we aggregated these diverse usage types and assigned the predominant usage type.

We processed the point OSM infrastructure data files in the same way: generating one GeoDataFrame containing all point features and assigning them a LUCAS category based on their tag under "fclass". Eventually, we repeated the LUCAS category assignment for the SACHI dataset: each usage value was assigned a LUCAS category, see table A2.



(a) (b)

Figure 3. Comparison of level of detail of original (a) OSM and (b) SACHI dataset. OSM shows a higher detail in mapping buildings, land use boundaries and linear transport infrastructure in contrast to SACHI, where the delineation is done with a pixel-based classifier (Bartsch et al., 2021). Background RGB high-resolution imagery of Deadhorse is from WorldView-3 (Copyright: DigitalGlobe, 2016). OSM data copyrighted by OpenStreetMap contributors, licensed under the Open Data Commons Open Database License (ODbL).

~~When comparing~~ **Combining SACHI and OSM**

When visually examining subsets of the SACHI and OSM datasets, we again observed ~~,~~ that the OSM data had a higher level of detail. The buildings' boundaries of the OSM dataset were delineated accurately (see Fig. 3a), while the buildings' outlines of the SACHI dataset were coarse and contained adjacent non-building areas due to the pixel-based approach (Fig. 3b). However, the SACHI approach detected more building area. Therefore, we implemented a decision tree structure for the last harmonization step of the infrastructure and usage type datasets. As a first step, we retrieved all overlapping features of the OSM and SACHI dataset with a spatial join (see Figure 2). When the OSM feature already had a LUCAS category assigned, we stored it in the final infrastructure and usage type dataset. If not, we assigned it the LUCAS category of the overlapping SACHI feature. All other non-overlappping SACHI and OSM features were also stored in the final infrastructure and usage type dataset.

2.2.2 Accuracy Assessment

To assess and quantify the accuracy of our data integration of infrastructure and human-impacted areas, we sub-sampled an area of 0.3 km^2 of the coastal settlement Shishmaref for which very high-resolution imagery was available. We built a reference dataset by manually digitizing all presumably permanent infrastructure elements using multi-spectral (RGB+NIR) orthophotos with a spatial resolution of 10 cm acquired in 2021 with the Modular Aerial Camera System (MACS) by Rettelbach et al. (2023). Buildings and other polygonal infrastructure features, such as repurposed shipping containers, small sheds, and coastal protection structures were mapped at a scale of 1:500. An infrastructure feature was considered permanent when it exhibited characteristics indicating a fixed location, such as supply pipes for shipping containers ~~, fixed roofing, etc.~~or fixed roofing. Roads were mapped at a scale of 1:2500 and solely if they exhibited an approximate width of 10 m or more to comply with the spatial resolution of the Sentinel sensors of SACHI. Subsequently, we created a grid layer spanning the mapped area with a size of 10 by 10 m for each grid cell. Each grid cell was assigned the corresponding values of the i) reference dataset and ii) the SIRIUS infrastructure and human-impacted area dataset: the OSM keys and values, "fclass", and the binary information if an infrastructure feature intersected with the grid cell (yes/ no). This allowed the calculation of a confusion matrix for the linear and polygonal infrastructure to determine the performance of the SIRIUS dataset.

In a confusion matrix, the classified dataset – in our case the SIRIUS infrastructure and human-impacted areas data – is compared with the reference dataset to determine the performance of the classification (Maxwell et al., 2021). The matrix provides information on correctly classified pixels (true positives: a "true" infrastructure feature of the reference dataset is also represented in the SIRIUS inventory; true negatives: a grid cell of the reference dataset does not show an infrastructure feature, neither does the SIRIUS inventory) and missclassifications (false positives and false negatives). A common metric derived from a confusion matrix is the overall accuracy (OA), the ratio of correctly classified pixels (true positive and true negative) to the total number of pixels (true or false) (Albertini et al., 2022).

2.2.3 Contaminated Sites of Alaska

The Contaminated Sites Program (CSP) of the Alaskan Department of Environmental Conservation (DEC) provides statewide information about the contamination by hazardous substances and manages their cleanup (State of Alaska Department of Environmental Conservation, 2023a). The DEC dataset entails information on the site name, address, geographic coordinates, cleanup status, responsible staff, contact person and the URL to a detailed site report. This site report contains complementary information on the contaminated medium (soil, groundwater, etc.), the substances (diesel, petroleum, etc.), and the date and type of cleanup measurements. ~~For our purpose of providing~~ We downloaded the detailed site report for each location to provide a harmonized dataset on contamination and infrastructure and human-impacted ~~elements~~ areas which allows users to assess their interrelation with permafrost degradation and hydrological watersheds in Alaska~~, we downloaded the detailed site report for each location~~. With basic text mining tasks (regular expressions, filtering for words in uppercase, etc.), we firstly derived all abbreviations of the site report. We compared the abbreviations to the DEC glossary (State of Alaska Department of Environmental Conservation, 2023b) and saved the ones indicating a substance or containment structure associated with contamination (e.g. LUST - Leaking Underground Storage Tank, PCBs - Polychlorinated Biphenyls, etc.) to a new attribute "contaminants" of the dataset. Subsequently, we deemed the dates followed by the expressions "Site Added to Database" and "Site Closure Approved" or "Cleanup Complete" (after 2008, (State of Alaska Department of Environmental Conservation, 2023c)) as the start and end date of the cleanup and saved them to the attributes "first_date" and "last_date", which allowed us to calculate the total cleanup time (attribute "cleanup_days"). If these expressions didn't appear in the site chronology report, we assumed the first and last mentioned date to be the start and finish of the cleanup. From this, we calculated the total cleanup time in days and saved it as an additional attribute. These simple text mining analyses were sufficient for deriving dates and abbreviations in uppercase letters as well as for comparing our list of toxic substances and containment related keywords against the full-text reports. However, we also wanted to provide information on the predominantly contaminated medium, so whether the groundwater, soil, or adjacent waterbodies were impacted. Here, we had to deal with a high heterogeneity in the structure of each report. Some reports listed the contaminated medium under the section „Contaminant Information". By comparing a set of medium keywords (soil, groundwater, river, etc.) against this section, we retrieved the contaminated ~~medium~~media.

2.2.4 Permafrost Data

As described for the infrastructure and contamination datasets, we assigned the joint spatial reference to the permafrost datasets and clipped their extent to the state boundary of Alaska. We derived the permafrost information from the modeled Northern Hemisphere permafrost map for 2000-2016 by Obu et al. (2018). The dataset comprises three GeoTIFF raster files containing the mean annual ground temperature (MAGT), the MAGT standard deviation, the permafrost probability fraction, and one vector file (ESRI Shapefile) giving information on the permafrost extent. The dataset is an estimation based on the TTOP (temperatures at the top of permafrost) model, which uses the mean annual air temperatures (MAAT) to model the MAGT and subsequently the permafrost probability and zonation (Obu et al., 2019). It has a resolution of 1 km^2 and was validated by borehole data (Obu et al., 2019). Within our study, we integrated the data on permafrost probability fraction and filtered for raster values where the probability of permafrost occurrence was greater than 50%, complying with the definition of the permafrost model domain (Langer et al., 2023). The filtering step ~~enabled us to concentrate on regions where permafrost is most likely to exist. Following that~~provides users with an additional filtering option for relevant permafrost information, as it allows the integration of mean annual ground temperatures. Subsequently, we vectorized the raster data to ensure compatibility with the other vector datasets. Given that each pixel value in the MAGT raster file was provided with precision to five decimal places, our initial step involved rounding these values to a single decimal place before proceeding with the vectorization process. We also included the vector data on the permafrost extent (zones) to allow the user to query data in dependence of permafrost zone, e.g. continuous, sporadic, etc..

2.2.5 ARCADE Watershed Database

The pan-Arctic Catchments Database, referred to as ARCADE, comprises a comprehensive collection of over 40 000 catchments draining into the Arctic Ocean down to a Strahler order of five (Speetjens et al., 2022). The geometries of the watersheds were derived from the Copernicus Digital Elevation Model with a spatial resolution of 30 arc seconds (approximately 1 km). Additional information regarding the catchments' characteristics (elevation, slope, etc.), climatology (precipitation, snowfall, runoff, etc.) and physiography (soil characteristics, permafrost parameters and extent, land surface data, etc.) were already incorporated to enrich the dataset (Speetjens et al., 2022). However, the permafrost extent and information on the MAGT were averaged over the extent of each watershed, which reach sizes of up to 3.1×10^6 km^2 (Speetjens et al., 2022). Therefore, we chose to include the information on every 1 km^2 grid cell of the permafrost MAGT dataset by Obu et al. (2019), see section 2.2.4.

2.3 Data Usability

To enhance spatial queries involving different usage types, contaminated sites, watersheds, and permafrost information, it was necessary to consolidate the individual pre-processed files into a single container. For this, we chose the GeoPackage format, as specified by the Open Geospatial Consortium (OGC). The GeoPackage format facilitates the exchange of geospatial data across different platforms, is open-source (Open Geospatial Consortium, 2023), and eliminates the need to handle multifile data formats like ESRI Shapefiles. Thus, it is highly suitable for accommodating the diverse data handling preferences of potential users. As GeoPackage uses a SQLite database container, the user is able to conduct their analyses within established geographic information systems such as ArcGIS, QGIS or spatial databases ~~(Geopackage Contributors, 2020; Rouault et al., 2023).~~

~~We demonstrate the usability of our data product by presenting and discussing various contexts in which the data can be used, see Section 3.2.1. These contexts may encompass risk assessments related to public health and potential infrastructure failures, statistical analyses, as well as basic cost calculations for clean up measurements at contaminated sites~~(Geopackage Contributors, 2020; War

.

3 Results

3.1 Data Harmonization & Mining

In this section, we outline the enhancements made to the infrastructure and human-impacted ~~elements~~ areas dataset of Alaska, as well as the information on contaminated sites. To showcase the advancements achieved by combining the SACHI and OSM data, we focused on two coastal regions, Nome and Prudhoe Bay, by sub-sampling their respective datasets. Furthermore, we investigated the performance of simple text mining tasks for the contaminated sites. For this, we randomly selected ten sites from the dataset and verified the accuracy of the derived start and end dates, cleanup duration, and information regarding the substances and contaminated medium. Subsequently, we analyzed in which cases the simple text mining approach performed well and identified its limitations in other instances.

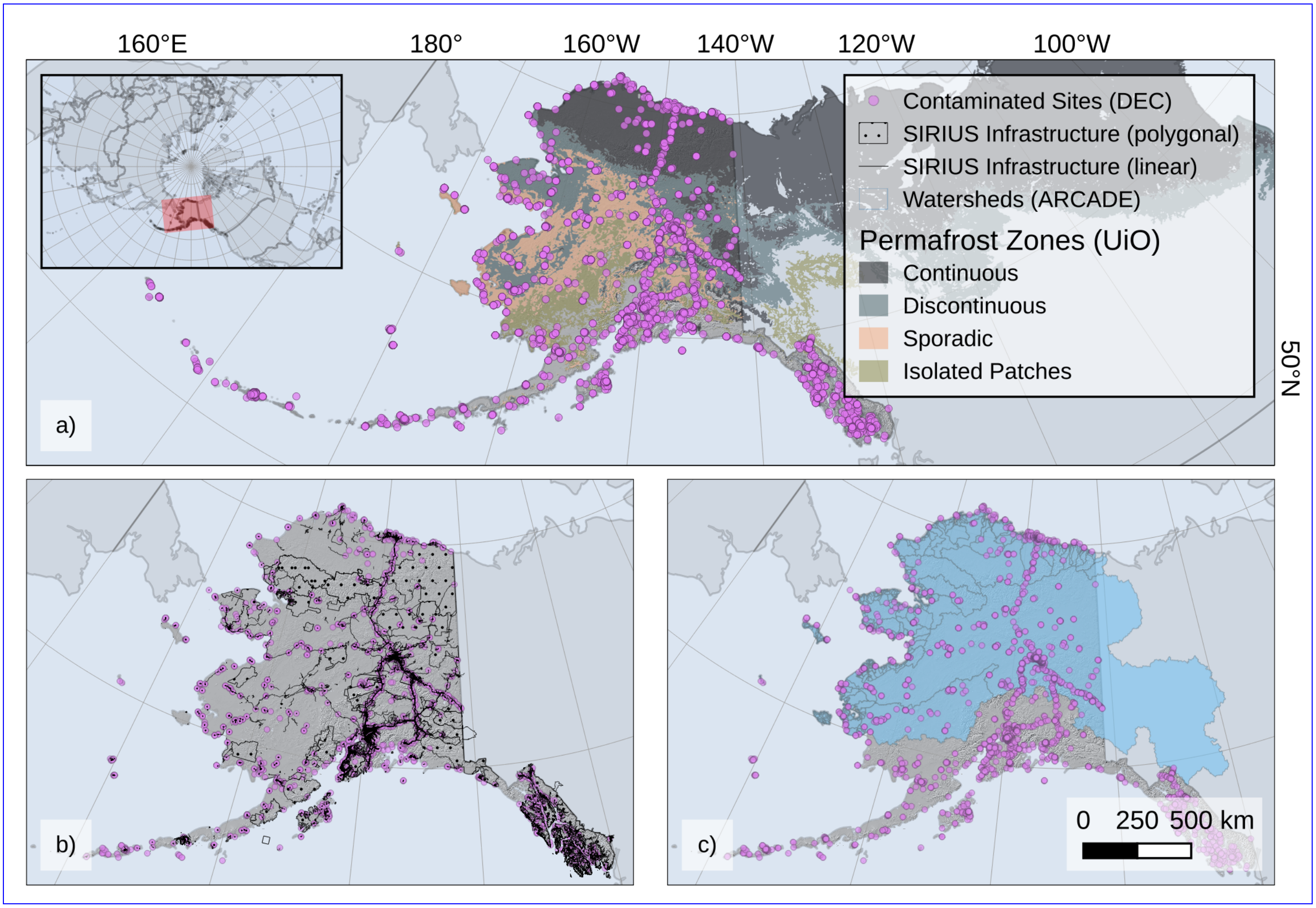


Figure 4. Overview of synthesized data: Contaminated sites and a) modeled permafrost zones, b) combined SACHI and OSM infrastructure and human-impacted areas, and c) watersheds draining into the Arctic Ocean and Bering Sea. OSM data copyrighted by OpenStreetMap contributors, licensed under ODbL. Basemap was made with Natural Earth: Free vector and raster map data @www.naturalearthdata.com..

3.1.1 Infrastructure and Human-Impacted Areas

The data fusion of the OSM and SACHI datasets resulted in an infrastructure and human-impacted areas map with a higher spatial detail and coverage than the original data sets. ~~Through the incorporation of OSM data, we successfully extended the coverage of the SACHI coastal area data from~~ While SACHI only covered the coastal region with an area of $62\ \text{km}^2$~~to span~~ , the incorporation of OSM data has extended the infrastructure map to encompass the entire state, now ~~encompassing~~ covering an expansive $640,593\ \text{km}^2$.

Furthermore, ~~this~~ the integration allowed us to enhance the level of detail regarding the usage categories for various infrastructure features. While we could initially assign five LUCAS categories to the SACHI data, including Fishing, Mining and Quarrying, Energy Production, Community Services, and Recreational, leisure, and sport, the inclusion of OSM data expanded this categorization to include an additional eight categories: Agriculture, Commerce, finance and business, Construction, Forestry, Industry and Manufacturing, Residential, Transport and communication networks, and Waste and Water Treatment ~~. (Refer~~ (refer to Table A3 and Figure 6 for a detailed breakdown~~.)~~).

This comprehensive categorization enhancement enabled us to refine the generalized approach. For example, we discovered that energy production sites, initially thought as dominant with an area of 28 km^2 in coastal regions, were, in ~~reality~~fact, less extensive, covering only 17 km^2 across the entire state (see Table A3 and Figure 6).

However, by incorporating the SACHI dataset, the map now also encompasses small and isolated elements like gravel pads and small paths, which ~~weren't~~ were not mapped by the OSM community but successfully derived from the satellites (refer to section 2.2). On the other hand, the integration of OSM data provided a heightened level of detail, enabling clear identification and differentiation of roads and single buildings (see Figure ~~??~~5c).

Looking at the settlement Nome under SACHI, we identified "Mining and Quarrying" as the primary land use category, aside from the transport network. These categories were determined by applying a buffer around each settlement (refer to section 2.2.1) and assigning it one predominant value (see Figure ~~??~~5b). Combining the SACHI with the OSM data ~~not only~~ enhanced the quality of the transport network, where streets are clearly defined even within areas with a high density of buildings and other human impacted areas~~, but~~. It also improved the detail of these usage type categories (Figure ~~??~~5c). We learned that the majority of the settlement's area is actually residential, characterized by houses and recreational areas such as pitches and parks (see Figures ~~?? and ??~~5a and 5c). The OSM data also added detail, where the spatial resolution of the SACHI product derived from Sentinel satellites fell short. For example, the pier in the Western area was not captured by the Sentinel satellites, but digitized by the OSM community. However, comparing the resulting human-impacted areas and infrastructure map with aerial imagery from Bing (as accessed via the QGIS Plugin OpenLayers (Sourcepole AG, 2024)) revealed that there is a second pier, which did not appear in the OSM nor in the SACHI dataset. Nonetheless, the true added value of the SACHI dataset ~~lay~~ is in its information on small features such as extraction pads and others, which only occasionally appear in the OSM data.

~~OSM SACHI~~

A closer examination of the Prudhoe Bay area confirmed this observation. Once again, the SACHI dataset showed more human-impacted ~~elements~~areas, probably from expanding exploration sites, while OSM offered more spatial detail. Furthermore, at both sites, we found that OSM exhibited higher quality in terms of linear infrastructure objects such as ~~roads and rail~~ road and railway lines. As mentioned in section 2.2, we compared the areas of the linear transport network between SACHI and OSM to evaluate the potential limitations of using OSM data. However, we discovered that the difference in area was only 5 km^2 (or 6 % of the total SACHI linear infrastructure area), as shown in table A3.

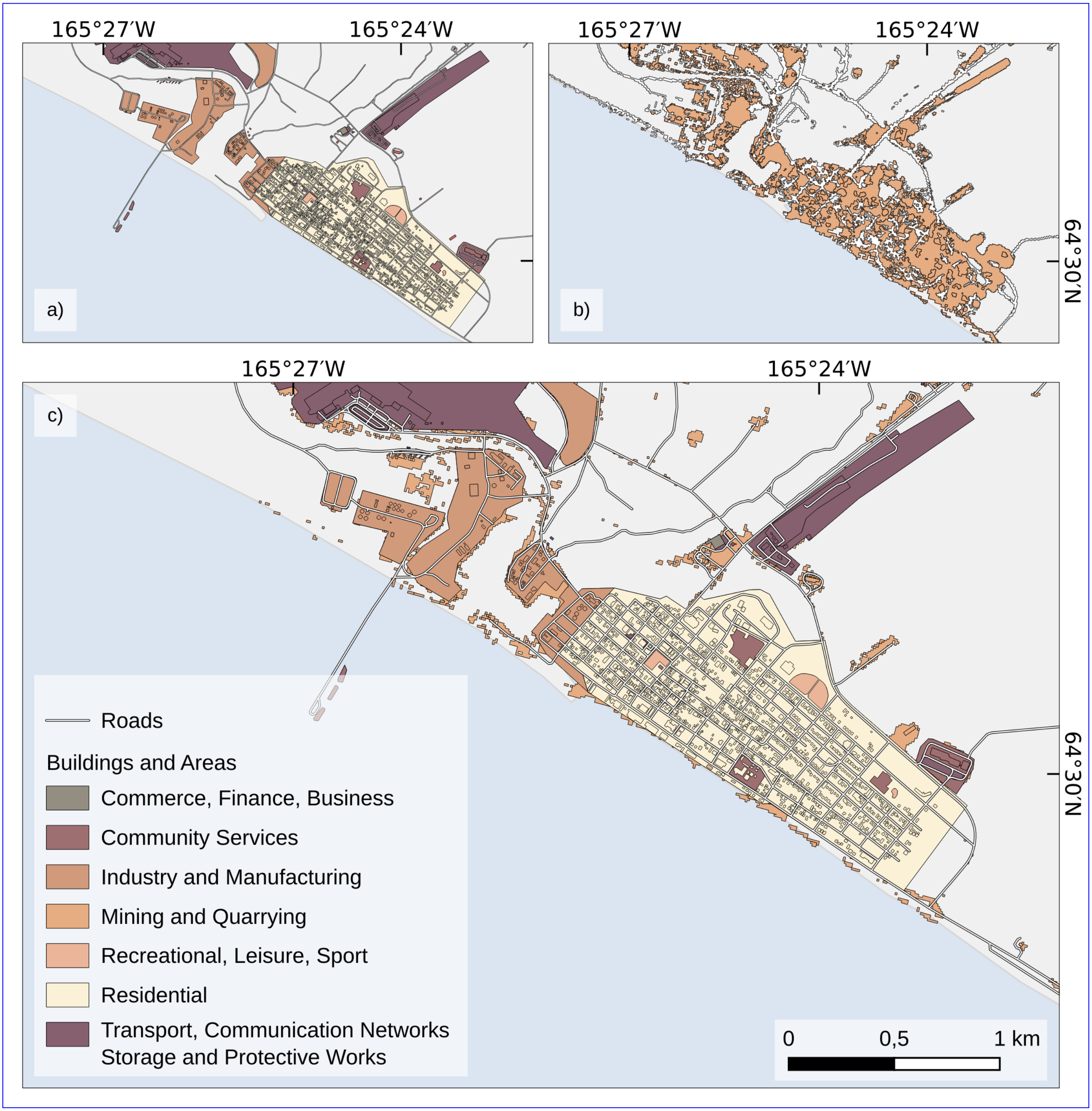


Figure 5. ~~Resulting~~ Input data from a) OSM and b) SACHI assigned to LUCAS categories for the example of the settlement Nome located along the Bering Sea coast. Map c) shows the harmonized data on infrastructure and human-impacted areas ~~map after fusion~~ of ~~SACHI and SIRIUS.~~ OSM data copyrighted by OpenStreetMap contributors, licensed under ODbL. Basemap was made with Natural Earth: Free vector and raster map data @www.naturalearthdata.com.

~~Input data from a) OSM and b) SACHI assigned to LUCAS categories for the example of the settlement Nome located along the Bering Sea coast. Map c) shows the harmonized data on infrastructure and human-impacted areas. OSM data copyrighted by OpenStreetMap contributors, licensed under ODbL. Basemap was made with Natural Earth: Free vector and raster map data @.~~

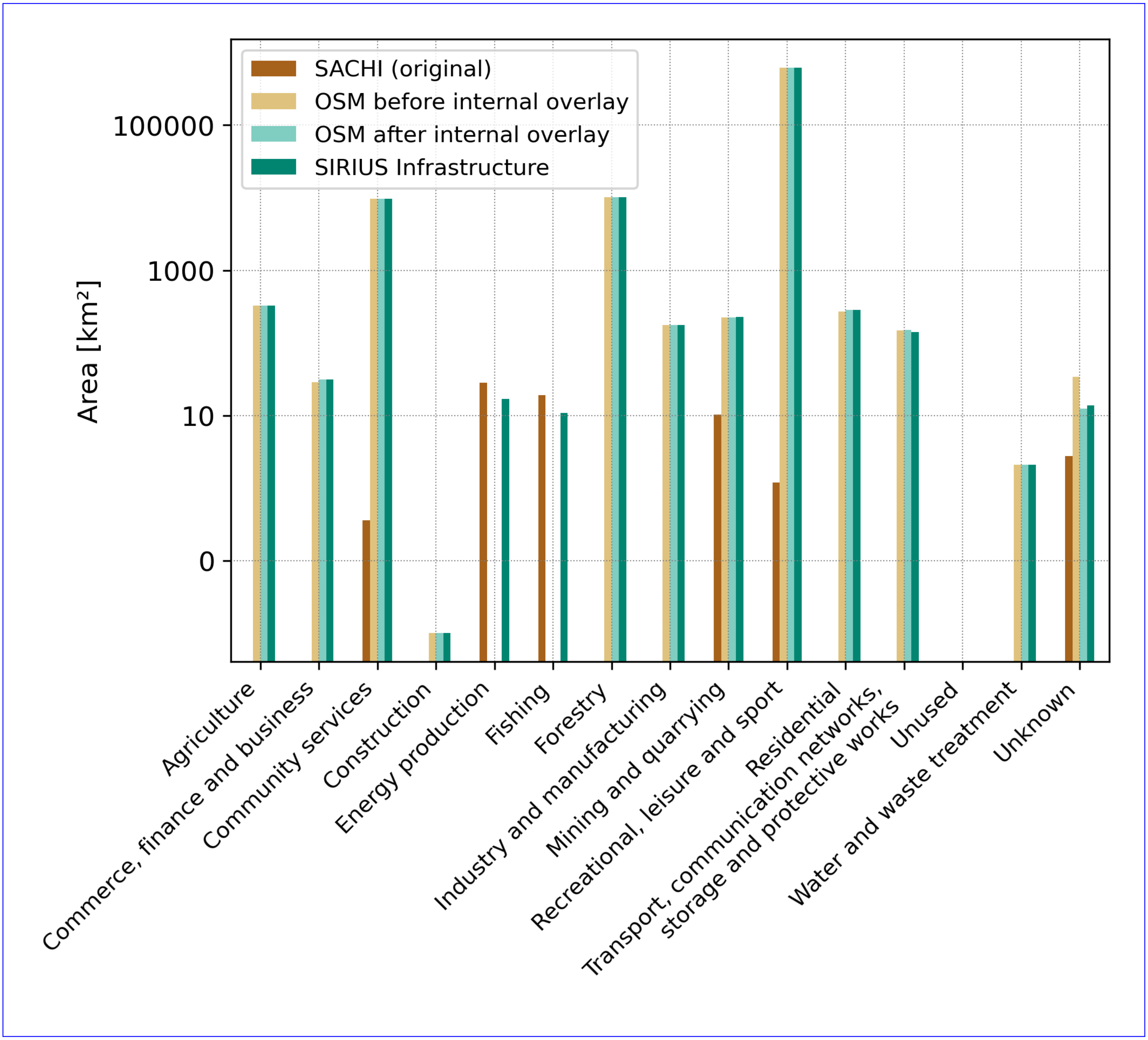


Figure 6. Improvement of spatial coverage and usage type categorization. Area [km^2] per LUCAS category for (i) the original SACHI dataset (only coastal areas), (ii) OSM before and (iii) after the internal overlay (complete extent of Alaska), and (iv) after combining both datasets within our SIRIUS inventory of infrastructure and human-impacted areas. For detailed values refer to Table A3.~~LUCAS category numbers are defined in Table 1.~~

390 The resulting ~~infrastructure and land use layer~~ SIRIUS infrastructure and human-impacted areas inventory not only represents economic activities but also incorporates ~~the population's requirements~~fundamental functions for living (Maier et al., 1977)

, including agricultural areas, commercial and residential zones, recreational spaces, ~~and more~~waste and water treatment, and community services. We also observed a significant decrease in the number of features with unknown land use types by internally overlaying OSM buildings with non-building OSM information, see Figure 6 and Table A3. Prior to the internal overlay, the area with unknown land use was $34\ \mathrm{km}^2$, whereas after the overlay, it reduced to only $13\ \mathrm{km}^2$ (refer to Table A3). This enhanced level of usage type detail allows for various applications, such as risk assessments for energy production facilities and transportation networks, as well as evaluations of contaminated sites close to recreational or agricultural areas (refer to section 3.2.1).

3.1.2 Accuracy Assessment

The overall accuracy (OA) of the confusion matrix represents the ratio of correctly classified pixels to the total number of all pixels (positive and negative~~pixel values~~, true and false~~. For~~). The OA of the linear infrastructure data of SIRIUS ~~the OA value~~ is 0.5. While this value seems relatively low, we need to zoom in on a specific detail: Of all 310 true road grid cells of the reference dataset showing a road infrastructure, 241 grid cells, thus 78 %, were accurately represented in the SIRIUS dataset, see Figure ~~??~~7a. A visual examination further reveals, that of the remaining 69 true road grid cells supposedly not represented by SIRIUS, 45 (65 %) were captured but with a slight spatial offset (see Figure 8~~A~~a), leading to a "false negative" when indeed it was only a positional inaccuracy. Taking into account these offset grid cells, the overall accuracy of the SIRIUS dataset improves to 0.69 and the true positive value increases from 0.78 to 0.92, indicating that 92 % of the road infrastructure is mapped in the SIRIUS inventory. All of the SIRIUS road grid cells, which were not mapped in the reference dataset (false positives) were either small tracks, footways or narrow residential roads, with a width of less than 10 m and thus not mapped, see Figure 8~~B~~b.

The overall accuracy of the polygonal infrastructure and human-impacted areas of the SIRIUS dataset shows a similarly low value of 0.53. However, the true positive value, representing the ratio of correctly classified values in SIRIUS per actual positive values, is 94 % (686 of 731 true polygonal infrastructure grid cells) (Figure ~~??~~7b). Of the remaining 45 false negative grid cells, 13 % were indeed missing, another 18 % occurred again because of a spatial offset, and 69 % appeared along the breakwater, protecting the shore (see Figure 9~~A~~a). OSM did not capture this structure and due to the relatively coarse spatial resolution of the Sentinel sensors, the representation of the breakwater was sparse and patchy in SACHI, leading to an underestimation and high number of false negatives.

However, substantially distorting the overall accuracy is the high number of false positives: 568 grid cells showed an intersection with polygonal infrastructure in the SIRIUS dataset (Figure ~~??~~7b), which was not captured in the reference dataset. 23 % of these false positives stem from an overestimation of the airport area in the SACHI and an altogether more generous mapping of the area in the OSM data. The Eastern part of the runway, for instance, appears re-vegetated and allows the conclusion that it is no longer in use, despite being still represented in the OSM data (refer to Figure 9~~B~~b). Yet, the highest number of false positives originates from areas affected by human activities represented in the SIRIUS dataset. These human-impacted areas posed a challenge in accurately mapping them for the reference dataset on the basis of the orthophotos alone. Some features, for example a playground, were either not visible or difficult to delineate accurately. Figure 9~~C~~ c shows an example of a human-impacted area mapped as industrial landuse by the OSM community. While the single storage structures are represented in the reference dataset, there was no indication of an enclosed area visible.

In summary, the low overall accuracy of the polygonal infrastructure data is distorted by a high number of false positives, that originate from either an overestimation of areas (e.g. airport) or a (conceptual) definition of landuse (e.g. playground, industrial usage, etc.) difficult to reproduce with orthophotos alone. However, it is important to note, that SIRIUS achieved a representation of 78 % for linear infrastructure and 94 % for polygonal infrastructure, respectively, of the true infrastructure values.

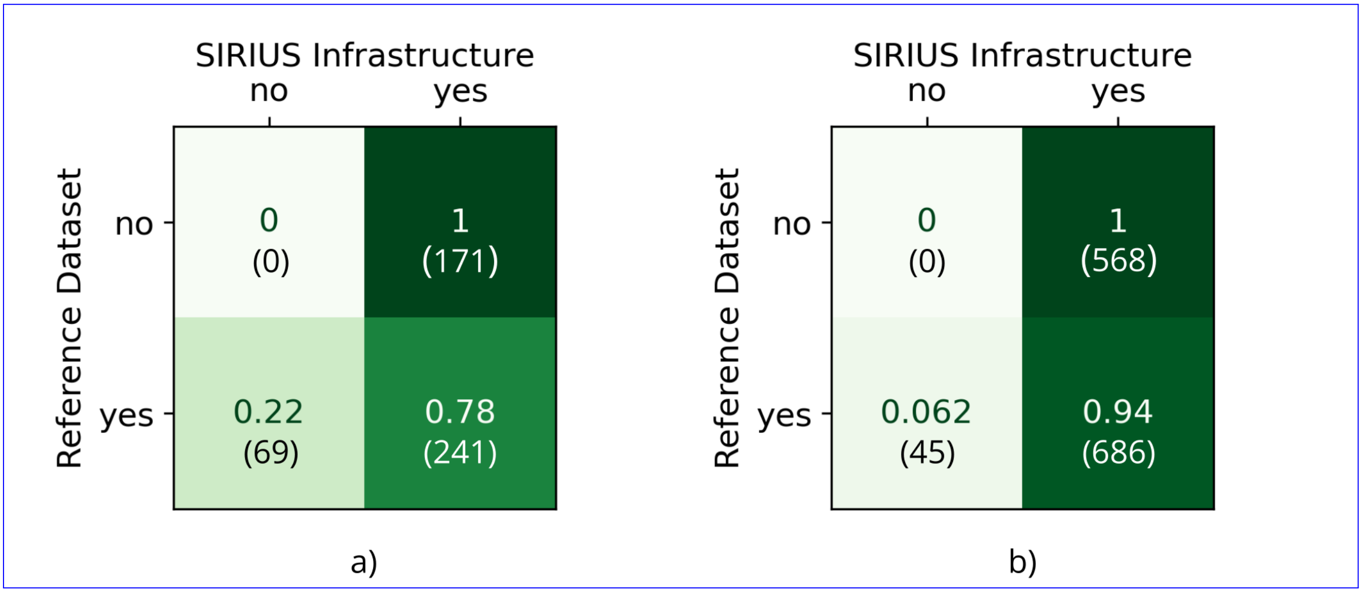


Figure 7. ~~Linear Infrastructure~~Confusion matrices were used to evaluate the accuracy of the SIRIUS dataset. The integrated SIRIUS inventory is compared with the reference data, which was mapped on the basis of orthophotos acquired in 2021. The matrices were normalized to the 'true' value, representing the ratio of correctly and incorrectly (SIRIUS-mapped) features for each true class label (values [0-1]). Figure (a) shows the accuracy of the linear infrastructure features with a true positive value of 0.78 and a false negative value of 0.22. For the polygonal infrastructure the true positive value is 0.94, deeming the SIRIUS inventory highly thorough.

~~Polygonal Infrastructure. Confusion Matrices were used to evaluate the accuracy of the SIRIUS dataset. The integrated SIRIUS inventory is compared with the reference data, which was mapped on the basis of orthophotos acquired in 2021. The values of the matrices were normalized to the 'true' value, representing the ratio of SIRIUS-mapped features to true features (values 0-1). Figure (a) shows the accuracy of the linear infrastructure features with a true positive value of 0.78 and a false negative value of 0.22. For the polygonal infrastructure the true positive value is 0.94, deeming the SIRIUS inventory highly thorough.~~

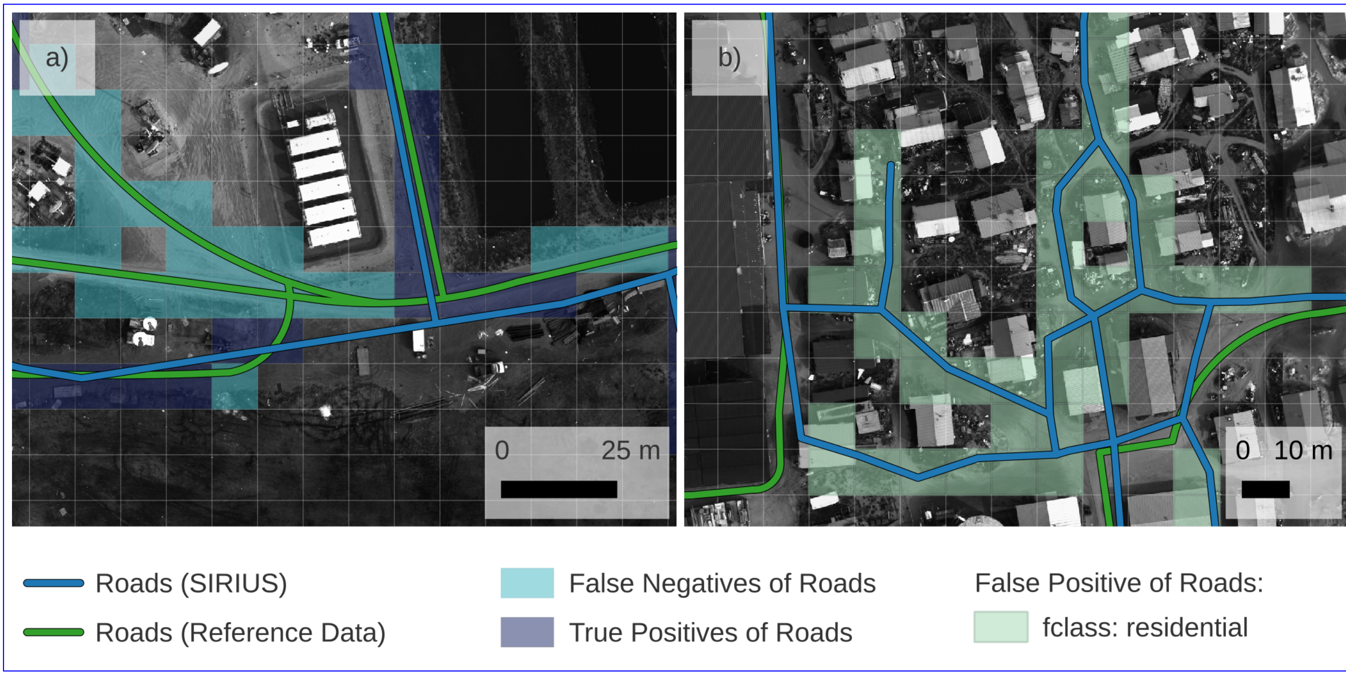


Figure 8. Comparison of the road network as represented in the SIRIUS inventory (integrated from OSM and SACHI from 2023 and 2021, respectively) and the reference data, which was mapped on the basis of multi-spectral (RGB + NIR) very high-resolution orthophotos from 2021. Subfigure ~~A~~a) showcases the presumably false negative values (0.22) of the SIRIUS road network, revealing that the roads are indeed present but exhibit a slight offset. Subfigure ~~B~~b) shows a section of the SIRIUS road network, which was deemed a false positive (1.0). However, the roads in SIRIUS are clearly visible on the imagery, yet they were not mapped due to their width being less than 10 m. Background imagery: orthophotos of Shishmaref, used to build the reference dataset (Rettelbach et al., 2023).

3.1.3 Contaminated Sites of Alaska

With the text mining approach, we successfully extracted additional information from the site reports of the contaminated sites program. The use of regular expressions allowed us to identify dates, abbreviations, and references to substances from the DEC glossary or any contaminated medium mentioned in the text. Consequently, we were able to calculate the total cleanup time at inactive sites and provide a comprehensive list of substances mentioned in the site reports. ~~By sub-sampling the data, we confirmed~~ To assess the accuracy, we retrieved a sample of 10 data entries from the dataset, see Table A4. We confirm the successful extraction of dates, following the pattern described in Section 2.2.3. The expressions "Sites Added to Database" and "Sites Closure Approved/ Cleanup Complete" were considered as the first and last action dates, respectively. In cases where no specific expressions were present, the first and last mentioned action dates were used instead. However, ~~we observed entries where~~ in 491 (6 %) entries the cleanup duration was recorded as 0 days (~~see~~ for an example see Hazard ID 361 of the sample in Table A5), and in ~~some instances~~214 (2 %) cases, even negative values were reported. This again points to a heterogeneous approach or methodology used by the agency to input data into the database. In these cases, "Site Closure Approved/ Cleanup

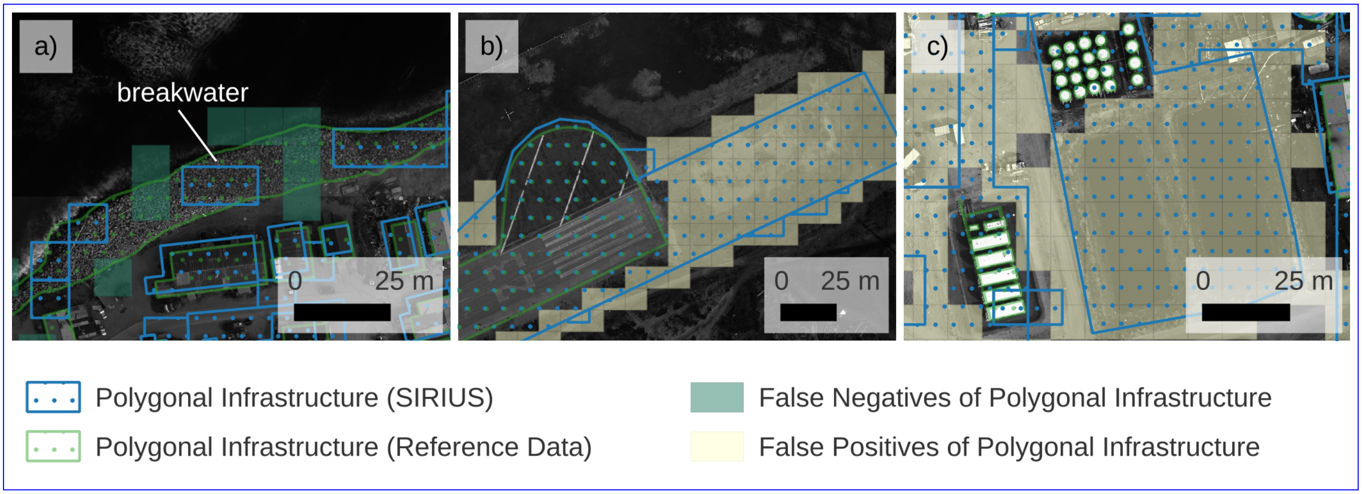


Figure 9. Comparison of the polygonal infrastructure and human-impacted areas as represented in the SIRIUS inventory (integrated from OSM and SACHI from 2023 and 2021, respectively) and the reference data, which was mapped on the basis of multi-spectral (RGB + NIR) very high-resolution orthophotos from 2021. Subfigure ~~A~~a) shows a subset of false negatives (0.06 in total) along the breakwater as a consequence of the patchy representation of this feature in SACHI. Subfigure ~~B~~b) displays the overestimation of the airport's runway in the SIRIUS dataset by including a re-vegetated area seemingly no longer in use. In subfigure ~~C~~c) the area close to the storage features is represented as industrial landuse in SIRIUS, which could not be identified on the basis of the orthophotos alone and is thus considered a false positive. Background imagery: orthophotos of Shishmaref, used to build the reference dataset (Rettelbach et al., 2023).

Complete" was entered on the same date or even before "Sites Added to Database." ~~The~~ For the sample, the retrieval of contaminants was highly successful, as all substances and containment structures listed in the DEC glossary (see Table A6) were found. However, any substances not appearing in the glossary won't be retrieved with our approach. Also, the information regarding the contaminated medium was limited as the DEC rarely provides details in the "Contaminant Information" section of the reports. Consequently, we were only able to derive the contaminated medium for 3321 (39 %) out of 8533 sites.

3.2 Data Usability

The resulting GeoPackage with our pre-processed spatial data layers contains all the input data on watersheds, permafrost probability, zones, and MAGT within the geographic extent of Alaska, projected to a joint spatial reference (EPSG code 5936). Additionally, it includes information on the contaminated sites, infrastructure features, and other human-impacted ~~elements~~areas. These datasets have undergone harmonization and enrichment, specifically focusing on the retrieval of detailed land usage information and the types of contamination, duration of cleanup measures, etc., as outlined in section 2.2.1 and 2.2.3. These datasets are now stored as separate layers (see Figure A2), eliminating the need for managing multiple ESRI Shapefiles and their auxiliary files. While retaining their original fields such as id, geometry, watershed names, etc., the files have been enriched with new information recorded in additional fields.

We deployed two GeoPackages with the same data. However, in PermaRisk_RRNetworkLine.gpkg the ~~rail~~ railway and road network are represented as line geometries, in PermaRisk_RRNetworkPolygonal.gpkg as polygons, based on the geometry buffers we defined in section 2.2.1. This allows more detailed spatial queries, such as deriving the length of a road or ~~rail~~ railway line within a specific research domain (see section 3.2.1). Considering the different user's requirements, the GeoPackage can be imported into a spatially enabled database, such as PostgreSQL/PostGIS, loaded into a Geographic Information System (GIS) or used within geospatial processing libraries, such as Python's GeoPandas. In this section, we will showcase the use of our GeoPackage within QGIS, perform SQL queries and access it via GeoPandas to generate exemplary statistics and explore potential application scenarios.

3.2.1 Application

As a first application scenario, we wanted to retrieve the total length of the road and ~~rail~~ railway lines within Alaska's continuous permafrost zone. As GeoPackage uses a SQLite database container, we could easily query spatial information by using the "Execute SQL" command in QGIS:

```
SELECT SUM(ST_Length(RRnetwork.geom))
FROM SACHI_OSM_InfrastructureHIElements_RRNetwork AS RRnetwork
JOIN UiO_PermafrostZones AS permafrost
ON ST_Intersects(RRnetwork.geom, permafrost.geom)
WHERE permafrost.EXTENT = 'Cont';
```

This query provided us with a length of 8456 km for the ~~rail~~ railway and road network intersecting with the continuous permafrost zone.

Another possible application is to determine the number of contaminated sites per watershed. To achieve this, the user can for example use the QGIS tool "count points in polygon". We tested this and discovered that the Yukon watershed, which is also Alaska's largest watershed draining into the Arctic Ocean, contained the highest number of contaminated sites, totaling 2256. However, to account for the huge differences in watershed sizes and normalize the number of sites per area, we further calculated the number of contaminated sites per square kilometer per watershed, showing that the watersheds along the coast of the Beaufort Sea (Figure 10 A and C) and Kotzebue (10 B) depict the highest density of contaminated sites per square kilometer (see Figure 10).

We further derived which land use category or infrastructure type shows the most contamination. For this analysis, we showcase the use of GeoPandas as a third processing option for our GeoPackage. By creating a spatial join between the SACHI_OSM_InfrastructureHIElements and SACHI_OSM_InfrastructureHIElements_RRNetwork (as polygonal representation) layers along with the DEC_ContaminatedSitesAK, we first derived all infrastructure and human-impacted areas and elements intersecting with a contaminated site. Next, we dissolved these intersecting elements based on their LUCAS attribute. Subsequently, we counted the number of contaminated sites by examining the points within these dissolved polygons, representing the aggregated LUCAS attribute. For the Python code see Appendix B.

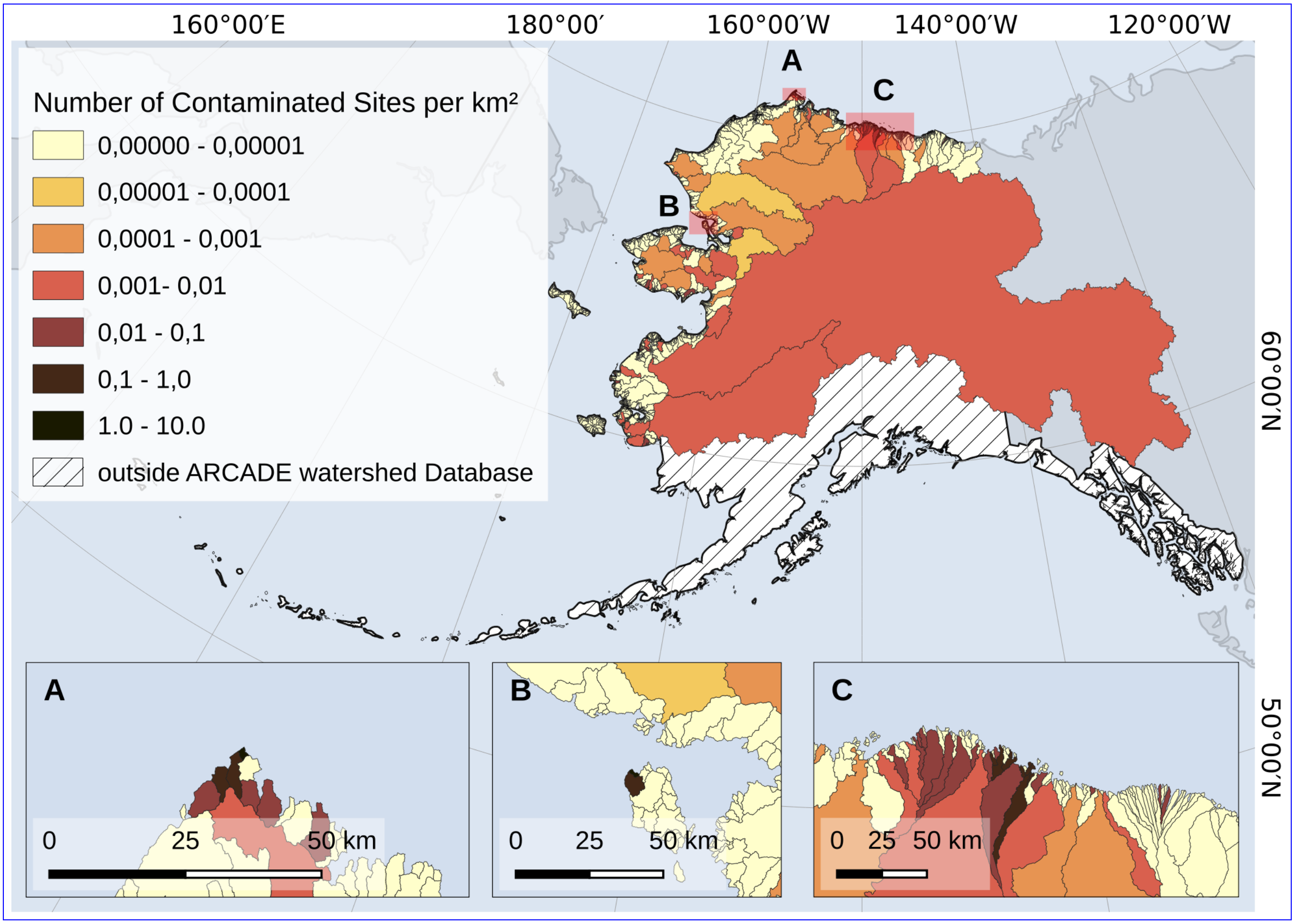


Figure 10. Number of contaminated sites per ARCADE watershed per square kilometer. Inset map A) shows a watershed along the coast of the Beaufort Sea with the highest value of 1.76 contaminated sites per square kilometer. Other watersheds exceeding more than one contamination per square kilometer were located in Kotzebue (inset map B) and on St. Lawrence Island. Inset map C) shows a range of watersheds of the Prudhoe Bay area. Basemap was made with Natural Earth: Free vector and raster map data @www.naturalearthdata.com.

This application example showed, that most of the contamination occurs in the land use categories "Community Services" (under which among others fall military installations, see table A2), "Transport, communication networks, storage and protective works", "Industry and Manufacturing", and "Recreational, leisure and sport" (see Table 2).

Table 2. Number of contaminated sites per land use category.

LUCAS	Nr. of Contaminated Sites
Agriculture	6
Commerce, finance and business	654
Community services	1989
Energy production	37
Fishing	79
Forestry	144
Industry and manufacturing	840
Mining and quarrying	32
Recreational, leisure and sport	755
Residential	531
Transport, communication networks, storage and protective works	1978
Unknown	210
Water and waste treatment	11

4 Discussion

4.1 Data Harmonization & Mining

4.1.1 Infrastructure and Human-Impacted Areas

The resulting inventory on infrastructure and human-impacted areas and elements in Alaska provides a detailed and comprehensive overview of various human activities, encompassing not only economic functions, but also ~~recreational purposes, agricultural and commercial components, etc.~~fundamental functions for living. Compared to the original SACHI dataset, we have achieved higher spatial detail and coverage throughout the entire state by incorporating OSM data (see Table A3 and Figure 6). On the other hand, the SACHI dataset has made a substantial contribution by capturing small elements that had been missed by the mapping efforts of the OSM community. This limitation may be attributed to the peripheral status of Arctic environments within the global OSM mapping network, primarily due to their sparse population. Hjort et al. (2018) report such a limitation for isolated, smaller communities and with regional variability (e.g. better coverage in North America and Eurasia compared to Asia). This deficit in mapped regions underscores the necessity for infrastructure products derived from remote sensing images, such as SACHI, as the underlying algorithms used to retrieve these features remain unbiased in terms of area selection. However, as described in Section 2.2.1, the algorithms fall short in densely populated areas, which makes distinguishing between adjacent features of different classes – e.g. buildings, roads, ~~extraction pads, etc.~~ or extraction pads – challenging. To meet the spatial detail of the OSM additions, the retrieval of infrastructure and human-impacted features could be enhanced by analysing remote sensing data with sub-meter spatial resolution. However, this improvement would come at a significant cost as most of these satellite images are commercial. On a pan-Arctic scale, this approach is nearly impossible due to the large spatial coverage necessary and associated high-resolution imagery costs (Manos et al., 2022). ~~Consequently, this~~ A compromise could involve using satellite imagery from providers that offer educational programs or discounted rates for researchers, such as Planet's Planetscope with a spatial resolution of 4 m (Sentinel Hub, 2024). Alternatively, deep learning models could be leveraged to generate high-resolution images from lower-resolution sources like Sentinel-2 (Wang et al., 2018) . Considering these challenges, emphasizes the need to rely on ~~crowed-sourced~~ crowd-sourced map data. These map data can also be generated remotely, using accessible Web Map Servers or GIS Plugins (e.g. Bing). Using OpenStreetMap as this data source serves as a gateway for this purpose. It establishes a low threshold for non-researchers, including citizen scientists, who can not only map various elements but eventually also incorporate valuable information on contamination, that has not been captured by official environmental agencies, highlighting the unique potential of OSM in this context. In addition, this approach allows the continuous development of suitable tags (attribute "fclass" in our data). However, based on own field visits, we have identified instances where certain areas and elements that contribute to the critical sector of "health and sanitation" are not accurately represented in OSM. For example, the Middle Salt Lagoon in Barrow, which is used for sewage purposes, is labeled as "water" in OSM and is thus not included in our SIRIUS dataset. This underlines the need for a comprehensive review of the mapping tags, before basing future inventories of critical infrastructure and human-impacted areas on OSM. Fortunately, due to OSM's open design and accessibility, these revisions can be easily implemented. Given that OSM undergoes daily updates through user contributions, the integration of OSM data also facilitates periodic updates within our inventory.

4.1.2 Accuracy Assessment

While the linear infrastructure data exhibited low overall accuracy in our Shishmaref test area, about two-thirds of the false negatives resulted from a spatial offset (see Figure 8). ~~Thus, the information of a road's presence is indeed given,~~ This indicates the presence of roads but with reduced positional accuracy~~. This is likely the result of~~ , likely due to an image offset between the MACS data and the ~~imagery used for mapping the road network in OSM . All the false positive values correspond to~~ OSM imagery. All false positives were narrow residential roads or small paths ~~of the SIRIUS dataset. Although clearly visible in the orthophotos , they were not digitized for~~ visible in orthophotos but not digitized in the reference dataset~~because the mapping adhered to the Sentinel spatial resolution of~~ , to comply with the Sentinel 10 ~~m. Including the narrow residential roads and small footways~~ meter resolution. Including these narrow features in the reference dataset would have ~~improved the accuracysubstantially.~~ substantially improved accuracy.

Nonetheless, 78 % of the true road grid cells were accurately represented in the SIRIUS dataset~~. When accounting for the offset grid cells, this value increases~~ , increasing to 92 % ~~. This value underlines~~ when accounting for offset grid cells. This highlights the effectiveness of OSM ~~for~~ in representing linear infrastructure ~~opposed~~ compared to SACHI. OSM ~~allows not only a clear distinction~~ distinguishes between roads and adjacent infrastructure areas ~~but also the inclusion of~~ and includes narrow roads and footways. ~~Looking ahead~~To improve the accuracy, it could ~~prove~~ be beneficial to integrate official data from local or federal agencies (e.g. Alaska Department of Transportation) to evaluate the comprehensiveness of the OSM linear infrastructure data. Further, incorporating the Trans-Alaskan-Pipeline would provide a spatial context for contamination, oil exploration and transportation data.

In the case of the polygonal infrastructure for the Shishmaref test area, the SIRIUS dataset achieves a representation of 94 % of all true values. Distorting the overall accuracy are the false positives, approximately a quarter of which belong to the section of Shishmaref's airport runway no longer in use. ~~Arguably, in~~ It is important to note, that OSM encourages users to regularly update features. If a user finds a feature no longer physically exists, they should delete it or tag it as "nonexistent" (OpenStreetMap Wiki, 2024b). If a feature still physically exists but is no longer in use, users are encouraged to tag it as "disused" (OpenStreetMap Wiki, 2024a). In this specific context and considering the potentials of contamination, it could be seen as an asset to have former land usage and industrial legacies represented in the SIRIUS dataset. An interesting approach might thus be to specifically filter for the OSM tags "nonexistent" and "disused" – in the regularly updated and historical OSM database – to highlight potential contamination sites.

The same applies to the human-impacted areas, such as playgrounds and industrial landuse. While these features are important infrastructures critical to Arctic communities they largely can not be mapped on the basis of orthophotos alone. Accordingly, the polygonal infrastructure lacks this level of detail when derived from SACHI. As discussed in Section 4.1.1, the coarse spatial resolution of the Sentinel sensors poses a challenge in densely populated areas. In such regions, buildings and human-impacted areas become difficult to separate from adjacent roads. This challenge contributes to the high number of false positives, where roads are misclassified as buildings, and areas of human activities are overestimated. However, this issue could be addressed using imagery with a higher spatial resolution.

It is important to note that the high mapping accuracy of 78 % (92 %) for linear and 94 % for polygonal infrastructure in our test area of Shishmaref can likely be expected for most of the coastal regions (until 100 km inland). For inland areas (beyond the extent of SACHI), the infrastructure data relies solely on OSM, which may show the above mentioned limitations. Once again, integrating further official data sources could improve its quality.

4.1.3 Contaminated Sites of Alaska

We could successfully enhance the DEC contaminated sites dataset with complementary information regarding substances, the affected medium, and the duration of cleanup measurements. However, the text mining approach, using regular expressions to compare site reports against the DEC glossary for retrieving the contaminant and affected medium, encountered limitations where data was entered heterogeneously into the database (see Section 3.1.3). For instance, only 39 % of the site reports included information about the contaminated medium in the designated section "Contaminant Information". In addition, in some cases, comparing the medium keywords (soil, groundwater, etc.), against this section led to false positives as these terms are frequently used to describe the hazard level of substances. The first entry (Hazard ID 26994) of our validation sample (refer to Table A4), is one of these false positives. The site report actually lists "soil" as contaminated medium, but the level description for the substances "Benzo(a)anthracene" and "Benzo(a)pyrene" is "Between Method 2 Migration to Groundwater and Human Health/Ingestion/Inhalation". Consequently, our approach also lists groundwater as a contaminated medium, which is not accurate. If we were to compare the full report against these keywords, it would result in even more incorrect classifications, as these terms are also employed to describe a suspicion of contamination. Furthermore, using regular expressions for the retrieval of the polluting substances, does not differentiate between the presence and absence of a contaminant, e.g. "PCB was found" vs. "PCB was not found". Although, we did not encounter statements of absent contaminants in the reports of our sample, we can not rule out the possibility of false positives of this kind.

These shortfalls could be addressed by implementing advanced text classification approaches from natural language processing and text mining. This could provide a more comprehensive understanding of toxic substances, including those not mentioned in the DEC glossary. Furthermore, these methods would extract and classify information about the contaminated medium from the entire report, rather than solely relying on the ~~sub-sampled~~ sampled section labeled "Contaminant Information." Another viable alternative would be the integration of Large Language Models (LLM). We tested our particular false positive case (Hazard ID 26994) with the LLM chatbot ChatGPT Version 3.5. by copying the full report into the prompt and requesting: "reading this text, tell me what medium (soil, groundwater, river, lake, etc.) was contaminated:" and it correctly classified the affected medium:

> *"Based on the provided text, the medium that was contaminated is "Soil." The text mentions that soil samples collected during site assessment activities showed elevated concentrations of contaminants, specifically "benzo(a)pyrene" and "benzo(a)anthracene," which exceeded certain cleanup levels. Therefore, the contamination occurred in the soil medium."*

This way, inconsistencies in data entries and false classifications could be easily addressed.

4.2 Data Usability

4.2.1 Application

All resulting datasets have been organized as individual layers within a single GeoPackage, which is available for download from our Zenodo repository (see Section 6). The GeoPackage does not have to be extracted (e.g. like a .zip archive) nor does it rely on the handling of multifile data formats such as ~~Shapefiles. You~~ ESRI Shapefiles. The user can seamlessly integrate it by either opening it in a GIS application or importing it into a spatially-enabled database like PostgreSQL/PostGIS. This way, each layer can be analyzed independently or in conjunction with the others, facilitating easy querying of critical infrastructure and human-impacted areas, and their interrelation with environmental parameters.

To achieve a more comprehensive understanding of the socio-economic implications of permafrost degradation, we advocate to incorporate additional environmental data, such as soil and waterbody databases, which are important for assessing the contamination severity and the significance of waterbodies as water resources. Additionally, incorporating demographic factors like age distribution, education, employment, and income numbers can provide valuable insights into the impacts of permafrost degradation on the population’s well-being.

5 Conclusions

The SIRIUS dataset offers a comprehensive inventory of critical infrastructure and human-impacted areas in Alaska. It enables researchers and local communities to explore data in a spatial context, providing valuable information on permafrost extent, permafrost probability, mean annual ground temperatures, and watersheds, allowing for an in-depth analysis of their interdependencies.

By combining the OSM and SACHI datasets, the information content regarding the type of infrastructure usage was greatly improved, increasing the number of usage categories from five (in SACHI) to a total of 13. The new usage categories now go beyond industrial and other economically important infrastructure by distinguishing elements of health care, food and water supply, sanitation, and areas of cultural heritage that are crucial to the well-being of local communities. Leveraging the OSM data and internally overlaying building features with non-building-features, we were also able to reduce the number of buildings with unknown usage type by 63 % (from $34.15\ \mathrm{km}^2$ to $12.58\ \mathrm{km}^2$).

As we move forward, ~~further enhancements of text classification methods and infrastructure data detail, will solidify the SIRIUS dataset as a foundational resource for pan-Arctic multi-source synthesis and data integration initiatives~~we have identified several steps to enhance the SIRIUS dataset further. Future updates will incorporate the new version of the SACHI dataset, which was released during the review period of this manuscript. Version 2.0 encompasses i) a refinement of the linear infrastructure features, now distinguishing between asphalt and gravel transport infrastructure, ii) airstrips, iii) humanly influenced waterbodies and reservoirs, and iv) additional regions further inland (Bartsch et al., 2023). The inclusion of water reservoirs affected by human activity is expected to improve the "health and sanitation" category by providing information on water and waste treatment facilities.

Further, an improvement of the text mining approach could be achieved by implementing transformer-based Large Language Models such as GPT or BERT. This could enhance information accuracy, density, and open up new pathways to incorporate contamination-related data from heterogeneous text sources, including online reports, historical documents, and analogue text data.

Researchers and volunteers can contribute to improving the dataset by providing feedback, additional data, or participating in (community) collaborative mapping efforts. The integration of OpenStreetMap into the Land Use / Cover Area frame statistical Survey (LUCAS) framework not only promotes harmonization across international boundaries, but also opens avenues for automated and regularly updated data retrieval through Python libraries like OMSnx (Boeing, 2017). Leveraging crowd-sourced data can encourage future mapping endeavors, including the identification of previously unregistered contamination sources. ~~This approach also allows the continuous and unrestricted expansion of~~

We aim to establish the SIRIUS dataset as a foundation for multi-source synthesis and data integration initiatives, consolidating infrastructure, environmental, and health-related information to facilitate the analysis of spatial trends and patterns, with the ~~capacity to assist decision-makers in effectively managing risks associated with permafrost degradation~~potential to be upscaled to the pan-Arctic region.

6 Code and data availability

The GeoPackage and Python code are available from https://doi.org/10.5281/zenodo.8311243 (Kaiser et al., 2023) (accessed on September 25, 2023).

Appendix A: Figures

```
OSM
├── 01_polygon
│   ├── gis_osm_buildings_a_free_1-5936.shp
│   ├── gis_osm_landuse_a_free_1-5936.shp
│   ├── gis_osm_natural_a_free_1-5936.shp
│   ├── gis_osm_places_a_free_1-5936.shp
│   ├── gis_osm_pofw_a_free_1-5936.shp
│   ├── gis_osm_pois_a_free_1-5936.shp
│   ├── gis_osm_traffic_a_free_1-5936.shp
│   ├── gis_osm_transport_a_free_1-5936.shp
│   └── gis_osm_water_a_free_1-5936.shp
├── 02_line
│   ├── gis_osm_railways_free_1-5936.shp
│   ├── gis_osm_roads_free_1-5936.shp
│   └── gis_osm_waterways_free_1-5936.shp
└── 03_point
    ├── gis_osm_natural_free_1-5936.shp
    ├── gis_osm_places_free_1-5936.shp
    ├── gis_osm_pofw_free_1-5936.shp
    ├── gis_osm_pois_free_1-5936.shp
    ├── gis_osm_traffic_free_1-5936.shp
    └── gis_osm_transport_free_1-5936.shp
```

Figure A1. Tree structure of OSM input data folder. OSM data was retrieved from OpenStreetMap Contributors and Geofabrik GmbH (2018) on January 20, 2023.

~~Flowchart of Harmonization Process. If not indicated otherwise, all input datasets are of ESRI Shapefile format.~~

```
PermaRisk.gpkg
 |__ARCADE_WatershedsDB
 |__DEC_ContaminatedSitesAK
 |__OSM_Point_InfrastructureHIElements
 |__SACHI_OSM_InfrastructureHIElements
 |__SACHI_OSM_InfrastructureHIElements_RRNetwork
 |__UiO_MAGT
 |__UiO_PermafrostProbability
 |__UiO_PermafrostZones
```

Figure A2. Tree structure of GeoPackage.

Appendix B: Tables

~~Synthesized datasets and their date of acquisition. Dataset Date of Acquisition Sentinel-1/2 derived Arctic Coastal Human Impact June 11, 2021OpenStreetMap January 20, 2023Pan-Arctic Catchments Database January 17, 2023 Northern Hemisphere Permafrost Map August 31, 2023 Alaska Department of Environmental Conservation Contaminated Sites Program March 2, 2023~~

Table A1: Assigning OSM keys and values to the "fclass" and "osm_type" attributes of the OSM ESRI Shapefiles, followed by LUCAS categorization.

fclass	osm_type	OSM_key	OSM_value	LUCAS
airfield	NaN	military	airfield	Community services
airport	NaN	aeroway	aerodrome	Transport, communication networks, ...
allotments	NaN	landuse	allotments	Residential
allotments	NaN	place	allotments	Residential
alpine_hut	NaN	tourism	alpine_hut	Recreational, leisure and sport
apron	NaN	aeroway	apron	Transport, communication networks, ...
archaeological	NaN	Historic	archaeological_site	Community services
arts_centre	NaN	amenity (Entertainment, Arts & Culture)	arts_centre	Recreational, leisure and sport
artwork	NaN	tourism	artwork	Recreational, leisure and sport
atm	NaN	amenity (financial)	atm	Commerce, finance and business
attraction	NaN	tourism	attraction	Recreational, leisure and sport
bakery	NaN	shop (food & beverages)	bakery	Commerce, finance and business
bank	NaN	amenity (financial)	bank	Commerce, finance and business
bar	NaN	amenity (Sustenance)	bar	Recreational, leisure and sport
beauty_shop	NaN	shop (Health and beauty)	beauty	Commerce, finance and business
bench	NaN	amenity (facilities)	bench	Community services
beverages	NaN	shop (food & beverages)	beverages	Commerce, finance and business
bicycle_rental	NaN	amenity (transportation)	bicycle_rental	Transport, communication networks, ...
bicycle_shop	NaN	shop (Outdoors and sport, vehicles)	bicycle	Commerce, finance and business
biergarten	NaN	amenity (Sustenance)	biergarten	Recreational, leisure and sport
bookshop	NaN	shop (Stationery, gifts, books, newspapers)	books	Commerce, finance and business
building	NaN	NaN	NaN	NaN

Continuation of Table A2				
fclass	osm_type	OSM_key	OSM_value	LUCAS
bus_station	NaN	amenity (transportation)	bus_station	Transport, communication networks, ...
bus_stop	NaN	highway (other highway features)	bus_stop	Transport, communication networks, ...
butcher	NaN	shop (food & beverages)	butcher	Commerce, finance and business
cafe	NaN	amenity (Sustenance)	cafe	Recreational, leisure and sport
camera_surveillance	NaN	man_made	surveillance	Transport, communication networks, ...
camp_site	NaN	tourism	camp_site	Recreational, leisure and sport
car_dealership	NaN	shop (Outdoors and sport, vehicles)	car	Commerce, finance and business
car_rental	NaN	amenity (transportation)	car_rental	Transport, communication networks, ...
car_wash	NaN	amenity (transportation)	car_wash	Transport, communication networks, ...
caravan_site	NaN	tourism	caravan_site	Recreational, leisure and sport
cemetery	NaN	landuse	cemetery	Community services
chalet	NaN	tourism	chalet	Recreational, leisure and sport
chemist	NaN	shop (Health and beauty)	chemist	Commerce, finance and business
cinema	NaN	amenity (Entertainment, Arts & Culture)	cinema	Recreational, leisure and sport
city	NaN	NaN	NaN	(removed)
clinic	NaN	amenity (healthcare)	clinic	Community services
clothes	NaN	shop (Clothing, shoes, accessories)	clothes	Commerce, finance and business
college	NaN	amenity (education)	college	Community services
college	NaN	building	college	Community services
commercial	NaN	building	commercial	Commerce, finance and business
commercial	NaN	landuse	commercial	Commerce, finance and business
comms_tower	NaN	man_made	communications_tower	Transport, communication networks, ...

Continuation of Table A2				
fclass	osm_type	OSM_key	OSM_value	LUCAS
community_centre	NaN	amenity (Entertainment, Arts & Culture)	community_centre	Recreational, leisure and sport
computer_shop	NaN	shop (Electronics)	computer	Commerce, finance and business
convenience	NaN	shop (food & beverages)	convenience	Commerce, finance and business
county	NaN	NaN	NaN	(removed)
courthouse	NaN	amenity (Public Service)	courthouse	Community services
crossing	NaN	footway	crossing	Transport, communication networks, ...
crossing	NaN	highway (other highway features)	crossing	Transport, communication networks, ...
crossing	NaN	railway	crossing	Transport, communication networks, ...
dam	NaN	waterway (Barriers on waterways)	dam	Community services
dentist	NaN	amenity (healthcare)	dentist	Community services
department_store	NaN	shop (General store, department store, mall)	department_store	Commerce, finance and business
doctors	NaN	amenity (healthcare)	doctors	Community services
dog_park	NaN	leisure	dog_park	Recreational, leisure and sport
doityourself	NaN	shop (Do-it-yourself, household, building mater...	doityourself	Commerce, finance and business
drinking_water	NaN	amenity (facilities)	drinking_water	Community services
drinking_water	NaN	emergency	drinking_water	Community services
embassy	NaN	office	diplomatic	Community services
farmland	NaN	landuse	farmland	Agriculture
farmyard	NaN	landuse	farmyard	Agriculture
fast_food	NaN	amenity (Sustenance)	fast_food	Recreational, leisure and sport
ferry_terminal	NaN	amenity (transportation)	ferry_terminal	Transport, communication networks, ...

Continuation of Table A2				
fclass	osm_type	OSM_key	OSM_value	LUCAS
fire_station	NaN	amenity (Public Service)	fire_station	Community services
fire_station	NaN	building	fire_station	Community services
florist	NaN	shop (Do-it-yourself, household, building mater...	florist	Commerce, finance and business
food_court	NaN	amenity (Sustenance)	food_court	Recreational, leisure and sport
forest	NaN	boundary	forest	Forestry
forest	NaN	landuse	forest	Forestry
fort	NaN	Historic	fort	Community services
fountain	NaN	amenity (Entertainment, Arts & Culture)	fountain	Recreational, leisure and sport
fuel	NaN	amenity (transportation)	fuel	Transport, communication networks, ...
fuel	NaN	waterway	fuel	Transport, communication networks, ...
furniture_shop	NaN	shop (Furniture and interior)	furniture	Commerce, finance and business
garden_centre	NaN	shop (Do-it-yourself, household, building mater...	garden_centre	Commerce, finance and business
general	NaN	shop (General store, department store, mall)	general	Commerce, finance and business
gift_shop	NaN	shop (Stationery, gifts, books, newspapers)	gift	Commerce, finance and business
golf_course	NaN	NaN	none	Recreational, leisure and sport
grass	NaN	landuse	grass	Community services
graveyard	NaN	amenity (Others)	grave_yard	Community services
greengrocer	NaN	shop (food & beverages)	greengrocer	Commerce, finance and business
guesthouse	NaN	tourism	guest_house	Recreational, leisure and sport
hairdresser	NaN	shop (Health and beauty)	hairdresser	Commerce, finance and business
hamlet	NaN	NaN	NaN	(removed)

Continuation of Table A2				
fclass	osm_type	OSM_key	OSM_value	LUCAS
heath	NaN	NaN	NaN	(removed)
helipad	NaN	aeroway	helipad	Transport, communication networks, ...
hospital	NaN	amenity (healthcare)	hospital	Community services
hospital	NaN	building	hospital	Community services
hostel	NaN	tourism	hostel	Recreational, leisure and sport
hotel	NaN	building	hotel	Recreational, leisure and sport
hotel	NaN	tourism	hotel	Recreational, leisure and sport
hunting_stand	NaN	amenity (Others)	hunting_stand	Hunting
ice_rink	NaN	leisure	ice_rink	Recreational, leisure and sport
industrial	NaN	building	industrial	Industry and manufacturing
industrial	NaN	landuse	industrial	Industry and manufacturing
industrial	NaN	usage	industrial	Industry and manufacturing
island	NaN	NaN	NaN	(removed)
jeweller	NaN	shop (Clothing, shoes, accessories)	jeweller	Commerce, finance and business
jeweller	NaN	shop (Clothing, shoes, accessories)	jewelry	Commerce, finance and business
kindergarten	NaN	amenity (education)	kindergarten	Community services
kindergarten	NaN	building	kindergarten	Community services
kiosk	NaN	building	kiosk	Commerce, finance and business
laundry	NaN	shop (others)	laundry	Commerce, finance and business
library	NaN	amenity (education)	library	Community services
lighthouse	NaN	man_made	lighthouse	Community services
locality	NaN	NaN	NaN	(removed)
mall	NaN	shop (General store, department store, mall)	mall	Commerce, finance and business

Continuation of Table A2

fclass	osm_type	OSM_key	OSM_value	LUCAS
marina	NaN	leisure	marina	Recreational, leisure and sport
market_place	NaN	amenity (Others)	marketplace	Commerce, finance and business
meadow	NaN	landuse	meadow	Agriculture
memorial	NaN	Historic	memorial	Community services
military	NaN	building	military	Community services
military	NaN	landuse	military	Community services
military	NaN	usage	military	Community services
mini_roundabout	NaN	highway (other highway features)	mini_roundabout	Transport, communication networks, ...
mobile_phone_shop	NaN	shop (Electronics)	mobile_phone	Commerce, finance and business
monument	NaN	Historic	monument	Community services
motel	NaN	tourism	motel	Recreational, leisure and sport
motorway_junction	NaN	highway (other highway features)	motorway_junction	Transport, communication networks, ...
museum	NaN	tourism	museum	Recreational, leisure and sport
nature_reserve	NaN	leisure	nature_reserve	Recreational, leisure and sport
newsagent	NaN	shop (Stationery, gifts, books, newspapers)	newsagent	Commerce, finance and business
nightclub	NaN	amenity (Entertainment, Arts & Culture)	nightclub	Recreational, leisure and sport
observation_tower	NaN	NaN	none	Community services
optician	NaN	shop (Health and beauty)	optician	Commerce, finance and business
orchard	NaN	landuse	orchard	Agriculture
outdoor_shop	NaN	shop (Outdoors and sport, vehicles)	outdoor	Commerce, finance and business
park	NaN	leisure	park	Recreational, leisure and sport

Continuation of Table A2

fclass	osm_type	OSM_key	OSM_value	LUCAS
parking	NaN	amenity (transportation)	parking	Transport, communication networks, ...
parking	NaN	building	parking	Transport, communication networks, ...
parking_bicycle	NaN	amenity (transportation)	bicycle_parking	Transport, communication networks, ...
parking_multistorey	NaN	NaN	none	Transport, communication networks, ...
parking_underground	NaN	NaN	none	Transport, communication networks, ...
pharmacy	NaN	amenity (healthcare)	pharmacy	Community services
picnic_site	NaN	tourism	picnic_site	Recreational, leisure and sport
pier	NaN	man_made	pier	Community services
pitch	NaN	leisure	pitch	Recreational, leisure and sport
playground	NaN	leisure	playground	Recreational, leisure and sport
police	NaN	amenity (Public Service)	police	Community services
post_box	NaN	amenity (Public Service)	post_box	Community services
post_office	NaN	amenity (Public Service)	post_office	Community services
prison	NaN	amenity (Public Service)	prison	Community services
pub	NaN	amenity (Sustenance)	pub	Recreational, leisure and sport
public_building	NaN	man_made	public_building	Community services
quarry	NaN	landuse	quarry	Mining and quarrying
railway_halt	NaN	railway	halt	Transport, communication networks, ...
railway_station	NaN	railway	station	Transport, communication networks, ...
recreation_ground	NaN	landuse	recreation_ground	Recreational, leisure and sport
recycling	NaN	amenity (waste management)	recycling	Water and waste treatment
recycling_clothes	NaN	NaN	none	Water and waste treatment
recycling_glass	NaN	NaN	none	Water and waste treatment
recycling_metal	NaN	NaN	none	Water and waste treatment
recycling_paper	NaN	NaN	none	Water and waste treatment

Continuation of Table A2

fclass	osm_type	OSM_key	OSM_value	LUCAS
residential	NaN	building	residential	Residential
residential	NaN	highway	residential	Residential
residential	NaN	landuse	residential	Residential
restaurant	NaN	amenity (Sustenance)	restaurant	Recreational, leisure and sport
retail	NaN	building	retail	Commerce, finance and business
retail	NaN	landuse	retail	Commerce, finance and business
ruins	NaN	building	ruins	Community services
ruins	NaN	Historic	ruins	Community services
school	NaN	amenity (education)	school	Community services
school	NaN	building	school	Community services
school	NaN	military	school	Community services
scrub	NaN	NaN	NaN	(removed)
service	NaN	building (power/ technical buildings)	service	unknown
service	NaN	highway (Special road types)	service	unknown
shelter	NaN	amenity (facilities)	shelter	Community services
shoe_shop	NaN	shop (Clothing, shoes, accessories)	shoes	Commerce, finance and business
slipway	NaN	leisure	slipway	Recreational, leisure and sport
speed_camera	NaN	highway (other highway features)	speed_camera	Transport, communication networks, ...
sports_centre	NaN	leisure	sports_centre	Recreational, leisure and sport
sports_shop	NaN	shop (Outdoors and sport, vehicles)	sports	Commerce, finance and business
stadium	NaN	building	stadium	Recreational, leisure and sport

Continuation of Table A2				
fclass	osm_type	OSM_key	OSM_value	LUCAS
stadium	NaN	leisure	stadium	Recreational, leisure and sport
stationery	NaN	shop (Stationery, gifts, books, newspapers)	stationery	Commerce, finance and business
stop	NaN	highway (other highway features)	stop	Transport, communication networks, ...
street_lamp	NaN	highway (other highway features)	street_lamp	Transport, communication networks, ...
suburb	NaN	NaN	NaN	(removed)
supermarket	NaN	building	supermarket	Commerce, finance and business
supermarket	NaN	shop (General store, department store, mall)	supermarket	Commerce, finance and business
swimming_pool	NaN	leisure	swimming_pool	Recreational, leisure and sport
taxi	NaN	amenity (transportation)	taxi	Transport, communication networks, ...
telephone	NaN	amenity (facilities)	telephone	Community services
theatre	NaN	amenity (Entertainment, Arts & Culture)	theatre	Recreational, leisure and sport
theme_park	NaN	tourism	theme_park	Recreational, leisure and sport
toilet	NaN	amenity (facilities)	toilets	Community services
toilet	NaN	building	toilets	Community services
tourist_info	NaN	tourism	information	Recreational, leisure and sport
tower	NaN	Historic	tower	unknown
tower	NaN	lifeguard	tower	unknown
tower	NaN	man_made	tower	unknown
tower	NaN	power	tower	unknown
town	NaN	NaN	NaN	(removed)
town_hall	NaN	amenity (Public Service)	townhall	Community services

Continuation of Table A2

fclass	osm_type	OSM_key	OSM_value	LUCAS
toy_shop	NaN	shop (others)	toys	Commerce, finance and business
track	NaN	leisure	track	Recreational, leisure and sport
traffic_signals	NaN	highway (other highway features)	traffic_signals	Transport, communication networks, ...
travel_agent	NaN	office	travel_agent	Commerce, finance and business
turning_circle	NaN	highway (other highway features)	turning_circle	Transport, communication networks, ...
university	NaN	amenity (education)	university	Community services
university	NaN	building	university	Community services
vending_any	NaN	NaN	none	unknown
vending_machine	NaN	amenity (Others)	vending_machine	unknown
vending_parking	NaN	NaN	none	Transport, communication networks, ...
veterinary	NaN	amenity (healthcare)	veterinary	Community services
video_shop	NaN	shop (Art, music, hobbies)	video	Commerce, finance and business
viewpoint	NaN	tourism	viewpoint	Recreational, leisure and sport
village	NaN	NaN	NaN	(removed)
waste_basket	NaN	amenity (waste management)	waste_basket	Water and waste treatment
wastewater_plant	NaN	man_made	wastewater_plant	Water and waste treatment
water_tower	NaN	building	water_tower	Water and waste treatment
water_tower	NaN	man_made	water_tower	Water and waste treatment
water_well	NaN	man_made	water_well	Water and waste treatment
water_works	NaN	man_made	water_works	Water and waste treatment
waterfall	NaN	NaN	NaN	(removed)
wayside_cross	NaN	Historic	wayside_cross	Community services

Continuation of Table A2				
fclass	osm_type	OSM_key	OSM_value	LUCAS
weir	NaN	waterway (Barriers on waterways)	weir	Community services
windmill	NaN	man_made	windmill	Community services
zoo	NaN	tourism	zoo	Recreational, leisure and sport
NaN	Dump Station	NaN	none	Water and waste treatment
NaN	amphitheatre	NaN	none	Recreational, leisure and sport
NaN	apartments	building	apartments	Residential
NaN	barn	building	barn	Agriculture
NaN	boathouse	NaN	none	Transport, communication networks, ...
NaN	bridge	building	bridge	Transport, communication networks, ...
NaN	bungalow	building	bungalow	Residential
NaN	bunker	building	bunker	Community services
NaN	cabin	building	cabin	Residential
NaN	carport	building	carport	Residential
NaN	cathedral	building	cathedral	Community services
NaN	chapel	building	chapel	Community services
NaN	church	building	church	Community services
NaN	civic	building	civic	Community services
NaN	classrooms	NaN	none	Community services
NaN	commercial;apartment	NaN	none	unknown
NaN	construction	building	construction	Construction
NaN	construction	landuse	construction	Construction
NaN	container	NaN	none	Transport, communication networks, ...
NaN	cowshed	building	cowshed	Agriculture
NaN	detached	building	detached	Residential
NaN	disused	NaN	none	Unused

Continuation of Table A2

fclass	osm_type	OSM_key	OSM_value	LUCAS
NaN	dormitory	building	dormitory	Community services
NaN	farm	building	farm	Agriculture
NaN	farm_auxiliary	building	farm_auxiliary	Agriculture
NaN	fire station	building	fire_station	Community services
NaN	garage	building	garage	Residential
NaN	garages	building	garages	Transport, communication networks, ...
NaN	gazebo	NaN	none	Community services
NaN	government	building	government	Community services
NaN	grandstand	building	grandstand	Recreational, leisure and sport
NaN	greenhouse	building	greenhouse	Agriculture
NaN	hangar	building	hangar	Transport, communication networks, ...
NaN	historic	NaN	none	Community services
NaN	house	building	house	Residential
NaN	houseboat	building	houseboat	Residential
NaN	hut	building	hut	Transport, communication networks, ...
NaN	lodge	NaN	none	Recreational, leisure and sport
NaN	manufacture	NaN	none	Industry and manufacturing
NaN	mil	building	military	Community services
NaN	monastery	building	monastery	Community services
NaN	no	NaN	none	unknown
NaN	office	building	office	Commerce, finance and business
NaN	pavilion	building	pavilion	Recreational, leisure and sport
NaN	public	building	public	Community services
NaN	radio_station	NaN	none	Transport, communication networks, ...
NaN	railway_shed	NaN	none	Transport, communication networks, ...
NaN	recreation center	NaN	none	Recreational, leisure and sport

Continuation of Table A2

fclass	osm_type	OSM_key	OSM_value	LUCAS
NaN	religious	building	religious	Community services
NaN	roof	building	roof	unknown
NaN	roof;office	NaN	none	unknown
NaN	sauna	NaN	none	Recreational, leisure and sport
NaN	semidetached_house	building	semidetached_house	Residential
NaN	shed	building	shed	Transport, communication networks, ...
NaN	ship	NaN	none	Transport, communication networks, ...
NaN	sports_hall	building	sports_hall	Recreational, leisure and sport
NaN	stable	building	stable	Agriculture
NaN	static_caravan	building	static_caravan	Recreational, leisure and sport
NaN	storage	NaN	none	Transport, communication networks, ...
NaN	storage_tank	building	storage_tank	Transport, communication networks, ...
NaN	strip mall	NaN	none	Commerce, finance and business
NaN	tent	building	tent	Community services
NaN	terminal	aeroway	terminal	Transport, communication networks, ...
NaN	terrace	building	terrace	Residential
NaN	toilets	amenity (facilities)	toilets	Community services
NaN	toilets	building	toilets	Community services
NaN	tower_block	NaN	none	unknown
NaN	train_station	building	train_station	Transport, communication networks, ...
NaN	transmitter	NaN	none	Transport, communication networks, ...
NaN	transportation	building	transportation	Transport, communication networks, ...
NaN	wall	barrier	wall	unknown
NaN	warehouse	building	warehouse	Commerce, finance and business
NaN	yert	building	ger	Community services
NaN	NaN	NaN	none	unknown

Continuation of Table A2

fclass	osm_type	OSM_key	OSM_value	LUCAS

End of Table A2

Table A2. Assigning use categories of SACHI dataset to LUCAS classification

SACHI.Use_main	SACHI.Use	LUCAS
Fishing	Fishing	Fishing
Fishing	Fishing, Tourism	Fishing
Mining	Mining	Mining and quarrying
Mining	Quartz Mining	Mining and quarrying
Mining	Gold Mining	Mining and quarrying
other	NaN	NaN
Gas/Oil	Gas, Oil, Tourism	Energy production
Gas/Oil	Gas, Oil	Energy production
Military	Military	Community services
NaN	Historical	Community services
NaN	Tourism	Recreational, leisure and sport
NaN	NaN	NaN
NaN	unknown	NaN

Table A3. Improvement of spatial coverage and usage type categorization. Area [km^2] per LUCAS category for (i) the original SACHI dataset (only coastal areas), (ii) OSM before and (iii) after the internal overlay (complete extent of Alaska), and (iv) after combining both datasets within our SIRIUS inventory of critical infrastructure and human-impacted areas. For a visualization, see Figure 6.

	SACHI (original)	OSM before internal overlay	OSM after internal overlay	~~Joint IS & HI elements~~ SIRIUS
Spatial Extent	Coastal Areas of Alaska	Entire State of Alaska, including the Alaskan Peninsula and Aleutian Islands, and the Inside Passage		
Total Area [km^2]	62	641631	641631	640593
Area per category [km^2]				
Agriculture	NaN	328.33	328.35	328.34
Commerce, finance and business	NaN	28.90	31.32	31.29
Community services	0.36	9662.34	9665.45	9657.55
Construction	NaN	0.01	0.01	0.01
Energy production	28.21	NaN	NaN	16.72
Fishing	19.05	NaN	NaN	10.42
Forestry	NaN	10207.61	10207.88	10207.73
Industry and manufacturing	NaN	175.00	177.53	177.52
Mining and quarrying	10.35	224.77	224.81	227.18
Recreational, leisure and sport	1.19	620546.48	620547.67	619495.11
Residential	NaN	271.87	283.13	283.04
Transport, communication networks, storage and protective works	NaN	149.27	149.98	141.91
Unused	NaN	0.00	0.00	0.00
Water and waste treatment	NaN	2.08	2.11	2.11
Unknown	2.78	34.15	12.58	14.00
Linear Infrastructure clipped to SACHI extent	86.42	81.07	and after fusion:	826.21

Table A4. Part I: Sample of final contaminated sites dataset. Columns ~~hazard~~ Hazard ID, borough, and status (IC stands for "Institutional Controls") were given in the original file, whereas the first and last date, cleanup days, contaminated medium and contaminants information were derived using simple text mining tasks (refer to 2.2.3).

Hazard ID	Borough	Status	First Date	Last Date	Cleanup Days	Medium	Contaminants
26994	Anchorage	Cleanup Complete	2019-02-13	2019-06-28	135.0	soil, groundwater	Petroleum, Benzene, Toluene, Ethylbenzene and Xylene, Benzene, Toluene, Ethylbenzene, Diesel Range Organics, Diesel, Fuel, Ethylene Dibromide, Gasoline Range Organics, Gasoline, Leaking Underground Storage Tanks, Volatile Organic Compound
3100	Northwest Arctic	Active	1998-06-09	NaN	NaN	NaN	Aboveground Storage Tanks, Petroleum, Benzene, Toluene, Ethylbenzene and Xylene, Diesel Range Organics, Diesel, Gasoline Range Organics, Gasoline, Oil, Residual Range Organics
23929	Anchorage	Cleanup Complete	1995-06-05	2004-06-22	3305.0	NaN	Petroleum, Benzene, Toluene, Ethylbenzene and Xylene, Benzene, Toluene, Ethylbenzene, Xylene, Diesel Range Organics, Diesel, Fuel, Gasoline Range Organics, Gasoline, Leaking Underground Storage Tanks, Residual Range Organics, Total Petroleum Hydrocarbon
3606	Aleutian Islands	Cleanup Complete (IC)	2000-04-15	2006-04-05	2181.0	soil	Unexploded Ordnance

Table A5. Part II: Sample of final contaminated sites dataset. Columns ~~hazard~~ Hazard ID, borough, and status (IC stands for "Institutional Controls") were given in the original file, whereas the first and last date, cleanup days, contaminated medium and contaminants information were derived using simple text mining tasks (refer to 2.2.3).

Hazard ID	Borough	Status	First Date	Last Date	Cleanup Days	Medium	Contaminants
3774	Kodiak Island	Cleanup Complete	2001-06-21	2005-12-30	1653.0	NaN	Aboveground Storage Tanks, Diesel Range Organics, Diesel, Fuel, Gasoline Range Organics, Gasoline, Leaking Underground Storage Tanks, Oil
23690	Anchorage	Cleanup Complete	1994-11-10	2006-04-17	4176.0	NaN	Petroleum, Benzene, Toluene, Ethylbenzene and Xylene, Benzene, Toluene, Gasoline Range Organics, Leaking Underground Storage Tanks
24611	Aleutians East	Cleanup Complete	1998-09-02	2002-04-24	1330.0	NaN	Petroleum, Leaking Underground Storage Tanks
2361	North Slope	Cleanup Complete	1995-09-08	1995-09-08	0.0	NaN	Diesel Range Organics, Diesel, Fuel
24927	Juneau	Cleanup Complete	1995-10-27	1998-10-22	1091.0	NaN	Petroleum, Benzene, Toluene, Ethylbenzene and Xylene, Benzene, Gasoline Range Organics, Gasoline, Leaking Underground Storage Tanks
24867	Anchorage	Cleanup Complete	1996-11-20	1996-12-01	11.0	NaN	Petroleum, Diesel, Leaking Underground Storage Tanks, Petroleum, Oil and Lubricants, Oil, Lubricants

Table A6. Abbreviations indicating toxic substances and contaminant related containment structures. Source: ADEC Glossary (State of Alaska Department of Environmental Conservation, 2023b).

abbreviation	meaning
AST	Aboveground Storage Tanks
	Petroleum
BTEX	Benzene, Toluene, Ethylbenzene and Xylene
	Benzene
	Toluene
	Ethylbenzene
	Xylene
DNAPL	Dense Non-Aqueous Phase liquid
DRO	Diesel Range Organics
	Diesel
	Fuel
	Kerosin
	Dioxin
EDB	Ethylene Dibromide
GRO	Gasoline Range Organics
	Gasoline
HAZMAT	Hazardous Materials
LNAPL	Light Non-Aqueous Phase Liquid
LUST	Leaking Underground Storage Tanks
NAPL	Non-aqueous phase liquid
PAHs	Polycyclic Aromatic Hydrocarbons
PCB	Polychlorinated Biphenyls
PCE	Perchloroethylene
PCE	Tetrachloroethylene
PERC	Tetrachloroethylene
POL	Petroleum, Oil and Lubricants
RRO	Residual Range Organics
TCE	Trichloroethylene
TPH	Total Petroleum Hydrocarbon
UXO	Unexploded Ordnance
VOC	Volatile Organic Compound

Appendix B: Application Code Snippets

```
import geopandas as gpd

## load GPKG file
geopackage_path = "/path/to/geopackage/PermaRisk_RRNetworkPolygonal_v01_r00.gpkg"

## load layers
polygon_layer = gpd.read_file(geopackage_path,
                              layer = 'SACHI_OSM_InfrastructureHIElements')
line_area_layer = gpd.read_file(geopackage_path,
                              layer = 'SACHI_OSM_InfrastructureHIElements_RRNetwork')
points_layer = gpd.read_file(geopackage_path,
                              layer = 'DEC_ContaminatedSitesAK')

## join IS-HI polygon and line layer
polygon_layer = polygon_layer.append(line_area_layer)

## create query
subset = gpd.sjoin(polygon_layer, points_layer, how='inner', predicate='intersects')
dfcount = subset.groupby('LUCAS')['geometry'].count().rename('pointcount').reset_index()
```

Author contributions. Conceptualization, S.K., J.B. and G.G. and M.L.; methodology, S.K. and M.L.; software, S.K.; validation, S.K.; formal analysis, S.K., G.G., J.B. and M.L.; resources, M.L.; data curation, S.K. and M.L.; writing—original draft preparation, S.K.; writing—review and editing, S.K., G.G., J.B. and M.L.; visualization, S.K.; supervision, G.G., J.B. and M.L.; project administration, M.L.; funding acquisition, M.L. and S.K.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This work was conducted within the young investigator group PermaRisk, which is funded by the German Federal Ministry of Education and Research (BMBF) under the funding reference number 01LN1709A. We further acknowledge suppport by the BMBF funded projects UndercoverEisAgenten (ref. nr. 01BF2115A) and ThinIce (ref. nr. 03F0943A) and the EU project Illuq (grant agreement id 101133587). S.K. also received funding from the Caroline von Humboldt Scholarship Program of Humboldt-Universität Berlin. G.G. acknowledges support by EU Arctic Passion (grant agreement id 101003472). ChatGPT 3.5 was used to improve the readability of Python scripts and support the identification of errors in the code.

References

rioxarray, https://github.com/corteva/rioxarray, [Online; accessed 10. Jun. 2024], 2024.

Alaska Oil and Gas Association: The Role of the Oil and Gas Industry in Alaska's Economy, Tech. rep., https://www.aoga.org/wp-content/uploads/2021/01/Reports-2020.1.23-Economic-Impact-Report-McDowell-Group-CORRECTED-2020.12.3.pdf, [Online; accessed 15. Sep. 2023], 2020.

Alaska Oil and Gas Association: Alaska Oil & Gas Association - State Revenue, https://www.aoga.org/state-revenue, [Online; accessed 20. Sep. 2023], 2021.

Alaska Seafood Marketing Institute: The Economic Value of Alaska's Seafood Industry, https://www.alaskaseafood.org/resource/economic-value-report-april-2024, [Online; accessed 10. Jun. 2024], 2024.

Albertini, C., Gioia, A., Iacobellis, V., and Manfreda, S.: Detection of Surface Water and Floods with Multispectral Satellites, Remote Sensing, 14, 6005, https://doi.org/10.3390/rs14236005, 2022.

Barrington-Leigh, C. and Millard-Ball, A.: The world's user-generated road map is more than 80% complete, PLOS ONE, 12, e0180 698, https://doi.org/10.1371/journal.pone.0180698, 2017.

Bartsch, A., Pointner, G., Ingeman-Nielsen, T., and Lu, W.: Towards Circumpolar Mapping of Arctic Settlements and Infrastructure Based on Sentinel-1 and Sentinel-2, Remote Sensing, 12, 2368, https://doi.org/10.3390/rs12152368, 2020.

Bartsch, A., Pointner, G., Nitze, I., Efimova, A., Jakober, D., Ley, S., Högström, E., Grosse, G., and Schweitzer, P.: Expanding infrastructure and growing anthropogenic impacts along Arctic coasts, Environmental Research Letters, 16, 115 013, https://doi.org/10.1088/1748-9326/ac3176, 2021.

Bartsch, A., Widhalm, B., von Baeckmann, C., Efimova, A., Tanguy, R., and Pointner, G.: Sentinel-1/2 derived Arctic Coastal Human Impact dataset (SACHI), https://doi.org/10.5281/zenodo.10160636, 2023.

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., and Wood, E. F.: Present and future Köppen-Geiger climate classification maps at 1-km resolution, Scientific Data, 5, 1–12, https://doi.org/10.1038/sdata.2018.214, 2018.

Bergstedt, H., Jones, B. M., Walker, D., Peirce, J., Bartsch, A., Pointner, G., Kanevskiy, M., Raynolds, M., and Buchhorn, M.: The spatial and temporal influence of infrastructure and road dust on seasonal snowmelt, vegetation productivity, and early season surface water cover in the Prudhoe Bay Oilfield, Arctic Science, https://cdnsciencepub.com/doi/full/10.1139/as-2022-0013, 2022.

Bessette-Kirton, E. K. and Coe, J. A.: A 36-Year Record of Rock Avalanches in the Saint Elias Mountains of Alaska, With Implications for Future Hazards, Frontiers in Earth Science, 8, 557 922, https://doi.org/10.3389/feart.2020.00293, 2020.

Biskaborn, B. K., Smith, S. L., Noetzli, J., Matthes, H., Vieira, G., Streletskiy, D. A., Schoeneich, P., Romanovsky, V. E., Lewkowicz, A. G., Abramov, A., Allard, M., Boike, J., Cable, W. L., Christiansen, H. H., Delaloye, R., Diekmann, B., Drozdov, D., Etzelmüller, B., Grosse, G., Guglielmin, M., Ingeman-Nielsen, T., Isaksen, K., Ishikawa, M., Johansson, M., Johannsson, H., Joo, A., Kaverin, D., Kholodov, A., Konstantinov, P., Kröger, T., Lambiel, C., Lanckman, J.-P., Luo, D., Malkova, G., Meiklejohn, I., Moskalenko, N., Oliva, M., Phillips, M., Ramos, M., Sannel, A. B. K., Sergeev, D., Seybold, C., Skryabin, P., Vasiliev, A., Wu, Q., Yoshikawa, K., Zheleznyak, M., and Lantuit, H.: Permafrost is warming at a global scale, Nat. Commun., 10, 264, 2019.

Boeing, G.: OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks, Computers, Environment and Urban Systems, 65, 126–139, https://doi.org/10.1016/j.compenvurbsys.2017.05.004, 2017.

Brunner, E. M. and Suter, M.: International CIIP Handbook 2008/2009: An Inventory of 25 National and 7 International Critical Information Infrastructure Protection Policies, Center for Security Studies (CSS), ETH, Zürich, Switzerland, https://doi.org/10.3929/ethz-b-000009792, 2008.

Bureau of Economic Analysis: Real value added to the gross domestic product of Alaska in the United States in 2022, by industry (in billion chained 2012 U.S. dollars), in: Statista, Statista, https://www.statista.com/statistics/1064725/alaska-real-gdp-by-industry/, [Online; accessed 15. Sep. 2023], 2023a.

Bureau of Economic Analysis: Regional Economic Accounts: Regional Definitions, https://apps.bea.gov/regional/definitions, [Online; accessed 15. Sep. 2023], 2023b.

Chadburn, S. E., Burke, E. J., Cox, P. M., Friedlingstein, P., Hugelius, G., and Westermann, S.: An observation-based constraint on permafrost loss as a function of global warming, Nature Climate Change, 7, 340–344, https://doi.org/10.1038/nclimate3262, 2017.

Cohen, J., Screen, J. A., Furtado, J. C., Barlow, M., Whittleston, D., Coumou, D., Francis, J., Dethloff, K., Entekhabi, D., Overland, J., and Jones, J.: Recent Arctic amplification and extreme mid-latitude weather, Nature Geoscience, 7, 627–637, https://doi.org/10.1038/ngeo2234, 2014.

Department of Labor and Workforce Development: Alaska Population Overview. 2019 Estimates, https://live.laborstats.alaska.gov/pop/estimates/pub/19popover.pdf, [Online; accessed 14. Sep. 2023], 2020.

E4.LUCAS (ESTAT): LUCAS 2018 (Land Use / Cover Area Frame Survey). Technical reference document C3 Classification (Land cover Land use), Tech. rep., Eurostat Regional Statistics and Geographic Information, https://ec.europa.eu/eurostat/documents/205002/8072634/LUCAS2018-C3-Classification.pdf, [Online; accessed 18. Sep. 2023], 2018.

Fortier, D., Allard, M., and Shur, Y.: Observation of rapid drainage system development by thermal erosion of ice wedges on Bylot Island, Canadian Arctic Archipelago, Permafrost and Periglacial Processes, 18, 229–243, https://doi.org/10.1002/ppp.595, 2007.

Geopackage Contributors: geopackage/guidance/getting-started.md at gh-pages · opengeospatial/geopackage, https://github.com/opengeospatial/geopackage/blob/gh-pages/guidance/getting-started.md, 2020.

Gillies, S. et al.: Rasterio: geospatial raster I/O for Python programmers, https://github.com/rasterio/rasterio, 2013–.

Godin, E., Fortier, D., and Burn, C.: Geomorphology of a thermo-erosion gully, Bylot Island, Nunavut, Canada1,21This article is one of a series of papers published in this CJES Special Issue on the theme ofFundamental and applied research on permafrost in Canada.2Polar Continental Shelf Project Contribution 043-11., Canadian Journal of Earth Sciences, 49, 979–986, https://doi.org/10.1139/e2012-015, 2012.

Haeberli, W.: Mountain permafrost — research frontiers and a special long-term challenge, Cold Regions Science and Technology, 96, 71–76, https://doi.org/10.1016/j.coldregions.2013.02.004, 2013.

Haeberli, W., Arenson, L. U., Wee, J., Hauck, C., and Mölg, N.: Discriminating viscous-creep features (rock glaciers) in mountain permafrost from debris-covered glaciers – a commented test at the Gruben and Yerba Loca sites, Swiss Alps and Chilean Andes, Cryosphere, 18, 1669–1683, https://doi.org/10.5194/tc-18-1669-2024, 2024.

Hamilton, L. C., Saito, K., Loring, P. A., Lammers, R. B., and Huntington, H. P.: Climigration? Population and climate change in Arctic Alaska, Population and Environment, 38, 115–133, https://doi.org/10.1007/s11111-016-0259-6, 2016.

Hammar, J., Grünberg, I., Kokelj, S. V., van der Sluijs, J., and Boike, J.: Snow accumulation, albedo and melt patterns following road construction on permafrost, Inuvik–Tuktoyaktuk Highway, Canada, Cryosphere, 17, 5357–5372, https://doi.org/10.5194/tc-17-5357-2023, 2023.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., and Oliphant, T. E.: Array programming with NumPy, Nature, 585, 357–362, https://doi.org/10.1038/s41586-020-2649-2, 2020.

Hjort, J., Karjalainen, O., Aalto, J., Westermann, S., Romanovsky, V. E., Nelson, F. E., Etzelmüller, B., and Luoto, M.: Degrading permafrost puts Arctic infrastructure at risk by mid-century, Nature Communications, 9, 5147, https://doi.org/10.1038/s41467-018-07557-4, 2018.

Irrgang, A. M., Lantuit, H., Gordon, R. R., Piskor, A., and Manson, G. K.: Impacts of past and future coastal changes on the Yukon coast — threats for cultural sites, infrastructure, and travel routes, Arctic Science, https://cdnsciencepub.com/doi/full/10.1139/as-2017-0041, 2019.

Jones, B. M., Grosse, G., Arp, C. D., Jones, M. C., Anthony, K. M. W., and Romanovsky, V. E.: Modern thermokarst lake dynamics in the continuous permafrost zone, northern Seward Peninsula, Alaska, Journal of Geophysical Research: Biogeosciences, 116, https://doi.org/10.1029/2011JG001666, 2011.

Jordahl, K., den Bossche, J. V., Fleischmann, M., McBride, J., Wasserman, J., Richards, M., Badaracco, A. G., Snow, A. D., Gerard, J., Tratner, J., Perry, M., Ward, B., Farmer, C., Hjelle, G. A., Taves, M., ter Hoeven, E., Cochran, M., rraymondgh, Gillies, S., Caria, G., Culbertson, L., Bartos, M., Eubank, N., Bell, R., sangarshanan, Flavin, J., Rey, S., maxalbert, Bilogur, A., and Ren, C.: geopandas/geopandas: v0.12.2, https://doi.org/10.5281/zenodo.7422493, 2022.

Jorgensen, T. and Meidlinger, D.: The Alaska Yukon Region of the Circumboreal Vegetation map (CBVM)., https://oaarchive.arctic-council.org/items/0d744f89-1e18-4249-b6aa-d64bac1bcdf3, [Online; accessed 15. Sep. 2023], 2015.

Jorgenson, M., Yoshikawa, K., Kanevskiy, M., Shur, Y., Romanovsky, V., Marchenko, S., and Jones, B.: Permafrost Characteristics of Alaska + Map, Ninth International Conference on Permafrost, https://www.researchgate.net/publication/334524021_Permafrost_Characteristics_of_Alaska_Map, 2008.

Jorgenson, M. T., Shur, Y. L., and Pullman, E. R.: Abrupt increase in permafrost degradation in Arctic Alaska, Geophysical Research Letters, 33, https://doi.org/10.1029/2005GL024960, 2006.

Kaiser, S., Boike, J., Grosse, G., and Langer, M.: SIRIUS - Synthesized Inventory of CRitical Infrastructure and HUman-Impacted Areas in Permafrost Regions of AlaSka, https://doi.org/10.5281/zenodo.8311243, 2023.

Keskitalo, K. H., Bröder, L., Shakil, S., Zolkos, S., Tank, S. E., van Dongen, B. E., Tesi, T., Haghipour, N., Eglinton, T. I., Kokelj, S. V., and Vonk, J. E.: Downstream Evolution of Particulate Organic Matter Composition From Permafrost Thaw Slumps, Frontiers in Earth Science, 9, 642 675, https://doi.org/10.3389/feart.2021.642675, 2021.

Kokelj, S. V. and Jorgenson, M. T.: Advances in Thermokarst Research, Permafrost and Periglacial Processes, 24, 108–119, https://doi.org/10.1002/ppp.1779, 2013.

Kokelj, S. V., Lacelle, D., Lantz, T. C., Tunnicliffe, J., Malone, L., Clark, I. D., and Chin, K. S.: Thawing of massive ground ice in mega slumps drives increases in stream sediment and solute flux across a range of watershed scales, Journal of Geophysical Research: Earth Surface, 118, 681–692, https://doi.org/10.1002/jgrf.20063, 2013.

Lamhonwah, D., Lafrenière, M. J., Lamoureux, S. F., and Wolfe, B. B.: Multi-year impacts of permafrost disturbance and thermal perturbation on High Arctic stream chemistry1, Arctic Science, https://cdnsciencepub.com/doi/10.1139/as-2016-0024, 2016.

Langer, M., von Deimling, T. S., Westermann, S., Rolph, R., Rutte, R., Antonova, S., Rachold, V., Schultz, M., Oehme, A., and Grosse, G.: Thawing permafrost poses environmental threat to thousands of sites with legacy industrial contamination, Nature Communications, 14, 1–11, https://doi.org/10.1038/s41467-023-37276-4, 2023.

Langer, M., Nitzbon, J., Groenke, B., Assmann, L.-M., Schneider von Deimling, T., Stuenzi, S. M., and Westermann, S.: The evolution of Arctic permafrost over the last 3 centuries from ensemble simulations with the CryoGridLite permafrost model, Cryosphere, 18, 363–385, https://doi.org/10.5194/tc-18-363-2024, 2024.

Leibman, M., Kizyakov, A., Zhdanova, Y., Sonyushkin, A., and Zimin, M.: Coastal Retreat Due to Thermodenudation on the Yugorsky Peninsula, Russia during the Last Decade, Update since 2001–2010, Remote Sensing, 13, 4042, https://doi.org/10.3390/rs13204042, 2021.

Levenstein, B., Lento, J., and Culp, J.: Effects of prolonged sedimentation from permafrost degradation on macroinvertebrate drift in Arctic streams, Limnology and Oceanography, 66, S157–S168, https://doi.org/10.1002/lno.11657, 2020.

Liew, M., Xiao, M., Farquharson, L., Nicolsky, D., Jensen, A., Romanovsky, V., Peirce, J., Alessa, L., McComb, C., Zhang, X., and Jones, B.: Understanding Effects of Permafrost Degradation and Coastal Erosion on Civil Infrastructure in Arctic Coastal Villages: A Community Survey and Knowledge Co-Production, Journal of Marine Science and Engineering, 10, 422, https://doi.org/10.3390/jmse10030422, 2022.

Liljedahl, A. K., Boike, J., Daanen, R. P., Fedorov, A. N., Frost, G. V., Grosse, G., Hinzman, L. D., Iijma, Y., Jorgenson, J. C., Matveyeva, N., Necsoiu, M., Raynolds, M. K., Romanovsky, V. E., Schulla, J., Tape, K. D., Walker, D. A., Wilson, C. J., Yabuki, H., and Zona, D.: Pan-Arctic ice-wedge degradation in warming permafrost and its influence on tundra hydrology, Nature Geoscience, 9, 312–318, https://doi.org/10.1038/ngeo2674, 2016.

Maier, J., Paesler, R., Ruppert, K., Schaffer, F., and Wirth, E.: DIE DEUTSCHE SOZIALGEOGRAPHIE IN IHRER THEORETISCHEN KONZEPTION UND IN IHREM VERHÄLTNIS ZU SOZIOLOGIE UND GEOGRAPHIE DES MENSCHEN on JSTOR, https://www.jstor.org/stable/27817927, [Online; accessed 20. Jun. 2024], 1977.

Manos, E., Witharana, C., Udawalpola, M. R., Hasan, A., and Liljedahl, A. K.: Convolutional Neural Networks for Automated Built Infrastructure Detection in the Arctic Using Sub-Meter Spatial Resolution Satellite Imagery, Remote Sensing, 14, 2719, https://doi.org/10.3390/rs14112719, 2022.

Markon, C., Gray, S., Berman, M., Eerkes-Medrano, L., Hennessy, T., Huntington, H. P., Littell, J., McCammon, M., Thoman, R., and Trainor, S. F.: Chapter 26 : Alaska. Impacts, Risks, and Adaptation in the United States: The Fourth National Climate Assessment, Volume II, Tech. rep., https://doi.org/10.7930/nca4.2018.ch26, 2018.

Maxwell, A. E., Warner, T. A., and Guillén, L. A.: Accuracy Assessment in Convolutional Neural Network-Based Deep Learning Remote Sensing Studies—Part 1: Literature Review, Remote Sensing, 13, 2450, https://doi.org/10.3390/rs13132450, 2021.

McGuire, A. D., Lawrence, D. M., Koven, C., Clein, J. S., Burke, E., Chen, G., Jafarov, E., MacDougall, A. H., Marchenko, S., Nicolsky, D., Peng, S., Rinke, A., Ciais, P., Gouttevin, I., Hayes, D. J., Ji, D., Krinner, G., Moore, J. C., Romanovsky, V., Schädel, C., Schaefer, K., Schuur, E. A. G., and Zhuang, Q.: Dependence of the evolution of carbon dynamics in the northern permafrost region on the trajectory of climate change, Proceedings of the National Academy of Sciences, 115, 3882–3887, https://doi.org/10.1073/pnas.1719903115, 2018.

Melvin, A. M., Larsen, P., Boehlert, B., Neumann, J. E., Chinowsky, P., Espinet, X., Martinich, J., Baumann, M. S., Rennels, L., Bothner, A., Nicolsky, D. J., and Marchenko, S. S.: Climate change damages to Alaska public infrastructure and the economics of proactive adaptation, Proceedings of the National Academy of Sciences of the United States of America, 114, E122–E131, https://doi.org/10.1073/pnas.1611056113, 2016.

Muster, S., Roth, K., Langer, M., Lange, S., Cresto Aleina, F., Bartsch, A., Morgenstern, A., Grosse, G., Jones, B., Sannel, A. B. K., Sjöberg, Y., Günther, F., Andresen, C., Veremeeva, A., Lindgren, P. R., Bouchard, F., Lara, M. J., Fortier, D., Charbonneau, S., Virtanen, T. A., Hugelius, G., Palmtag, J., Siewert, M. B., Riley, W. J., Koven, C. D., and Boike, J.: PeRL: a circum-Arctic Permafrost Region Pond and Lake database, Earth System Science Data, 9, 317–348, https://doi.org/10.5194/essd-9-317-2017, 2017.

National Oceanic and Atmospheric Administration. National Centers for Environmental Information: NOAA NCEI U.S. Climate Normals Quick Access, https://www.ncei.noaa.gov/access/us-climate-normals/#dataset=normals-monthly&timeframe=30&station=USW00027406, [Online; accessed 20. Aug. 2023], 2023a.

National Oceanic and Atmospheric Administration. National Centers for Environmental Information: NOAA NCEI U.S. Climate Normals Quick Access, https://www.ncei.noaa.gov/access/us-climate-normals/#dataset=normals-monthly&timeframe=30&station=USW00025507, [Online; accessed 20. Aug. 2023], 2023b.

National Weather Service: U.S. States and Territories, https://www.weather.gov/gis/USStates, 2023.

NOAA Office for Coastal Management: Alaska, https://coast.noaa.gov/states/alaska.html, [Online; accessed 15. Sep. 2023], 2023.

Obu, J., Westermann, S., Kääb, A., and Bartsch, A.: Ground Temperature Map, 2000-2016, Northern Hemisphere Permafrost, PANGAEA, https://doi.org/10.1594/PANGAEA.888600, 2018.

Obu, J., Westermann, S., Bartsch, A., Berdnikov, N., Christiansen, H. H., Dashtseren, A., Delaloye, R., Elberling, B., Etzelmüller, B., Kholodov, A., Khomutov, A., Kääb, A., Leibman, M. O., Lewkowicz, A. G., Panda, S. K., Romanovsky, V., Way, R. G., Westergaard-Nielsen, A., Wu, T., Yamkhin, J., and Zou, D.: Northern Hemisphere permafrost map based on TTOP modelling for 2000–2016 at 1 km2 scale, Earth-Science Reviews, 193, 299–316, https://doi.org/https://doi.org/10.1016/j.earscirev.2019.04.023, 2019.

Open Geospatial Consortium: GeoPackage Encoding Standard - Open Geospatial Consortium, https://www.ogc.org/standard/geopackage/, 2023.

OpenStreetMap Contributors and Geofabrik GmbH: Geofabrik Download Server, http://download.geofabrik.de/, [Online; accessed 20. Jan. 2023], 2018.

OpenStreetMap Foundation: Main Page — OpenStreetMap Foundation,, https://osmfoundation.org/w/index.php?title=Main_Page&oldid=11226, [Online; accessed 20. Sep. 2023], 2023.

OpenStreetMap Wiki: Map features — OpenStreetMap Wiki, https://wiki.openstreetmap.org/w/index.php?title=Map_features&oldid=2488629, [Online; accessed 10. May 2023], 2023.

OpenStreetMap Wiki: Key:disused:∗ - OpenStreetMap Wiki, https://wiki.openstreetmap.org/wiki/Key:disused:%2A, [Online; accessed 21. Jun. 2024], 2024a.

OpenStreetMap Wiki: Nonexistent features - OpenStreetMap Wiki, https://wiki.openstreetmap.org/wiki/Nonexistent_features, [Online; accessed 21. Jun. 2024], 2024b.

pandas development team, T.: pandas-dev/pandas: Pandas, https://doi.org/10.5281/zenodo.7549438, 2023.

Rajendran, S., Sadooni, F. N., Al-Kuwari, H. A.-S., Oleg, A., Govil, H., Nasir, S., and Vethamony, P.: Monitoring oil spill in Norilsk, Russia using satellite data, Scientific Reports, 11, 1–20, https://doi.org/10.1038/s41598-021-83260-7, 2021.

Ramage, J., Jungsberg, L., Wang, S., Westermann, S., Lantuit, H., and Heleniak, T.: Population living on permafrost in the Arctic, Population and environment, 43, 22–38, 2021.

Ramage, J. L., Irrgang, A. M., Herzschuh, U., Morgenstern, A., Couture, N., and Lantuit, H.: Terrain controls on the occurrence of coastal retrogressive thaw slumps along the Yukon Coast, Canada, Journal of Geophysical Research: Earth Surface, 122, 1619–1634, https://doi.org/10.1002/2017JF004231, 2017.

Rantanen, M., Karpechko, A. Yu., Lipponen, A., Nordling, K., Hyvärinen, O., Ruosteenoja, K., Vihma, T., and Laaksonen, A.: The Arctic has warmed nearly four times faster than the globe since 1979, Communications Earth & Environment, 3, 1–10, https://doi.org/10.1038/s43247-022-00498-3, 2022.

Raynolds, M. K., Walker, D. A., Ambrosius, K. J., Brown, J., Everett, K. R., Kanevskiy, M., Kofinas, G. P., Romanovsky, V. E., Shur, Y., and Webber, P. J.: Cumulative geoecological effects of 62 years of infrastructure and climate change in ice-rich permafrost landscapes, Prudhoe Bay Oilfield, Alaska, Global Change Biology, 20, 1211–1224, https://doi.org/10.1111/gcb.12500, 2014.

Raynolds, M. K., Walker, D. A., Balser, A., Bay, C., Campbell, M., Cherosov, M. M., Daniëls, F. J. A., Eidesen, P. B., Ermokhina, K. A., Frost, G. V., Jedrzejek, B., Jorgenson, M. T., Kennedy, B. E., Kholod, S. S., Lavrinenko, I. A., Lavrinenko, O. V., Magnússon, B., Matveyeva, N. V., Metúsalemsson, S., Nilsen, L., Olthof, I., Pospelov, I. N., Pospelova, E. B., Pouliot, D., Razzhivin, V., Schaepman-Strub, G., Šibík, J., Telyatnikov, M. Yu., and Troeva, E.: A raster version of the Circumpolar Arctic Vegetation Map (CAVM), Remote Sensing of Environment, 232, 111 297, https://doi.org/10.1016/j.rse.2019.111297, 2019.

Rettelbach, T., Nitze, I., Grünberg, I., Hammar, J., Schäffler, S., Hein, D., Gessner, M., Bucher, T., Brauchle, J., Hartmann, J., Sachs, T., Boike, J., and Grosse, G.: Super-high-resolution aerial imagery, digital surface model and 3D point cloud of Shishmaref, Alaska, https://doi.pangaea.de/10.1594/PANGAEA.962678, [Online; accessed 20. Dec. 2023], 2023.

Rouault, E., Warmerdam, F., Schwehr, K., Kiselev, A., Butler, H., Łoskot, M., Szekeres, T., Tourigny, E., Landa, M., Miara, I., Elliston, B., Chaitanya, K., Plesea, L., Morissette, D., Jolma, A., and Dawson, N.: GDAL, https://doi.org/10.5281/zenodo.7986215, 2023.

Runge, A., Nitze, I., and Grosse, G.: Remote sensing annual dynamics of rapid permafrost thaw disturbances with LandTrendr, Remote Sensing of Environment, 268, 112 752, https://doi.org/10.1016/j.rse.2021.112752, 2022.

Schuur, E. A. G., McGuire, A. D., Schädel, C., Grosse, G., Harden, J. W., Hayes, D. J., Hugelius, G., Koven, C. D., Kuhry, P., Lawrence, D. M., Natali, S. M., Olefeldt, D., Romanovsky, V. E., Schaefer, K., Turetsky, M. R., Treat, C. C., and Vonk, J. E.: Climate change and the permafrost carbon feedback, Nature, 520, 171–179, https://doi.org/10.1038/nature14338, 2015.

Schuur, E. A. G., Abbott, B. W., Commane, R., Ernakovich, J., Euskirchen, E., Hugelius, G., Grosse, G., Jones, M., Koven, C., Leshyk, V., Lawrence, D., Loranty, M. M., Mauritz, M., Olefeldt, D., Natali, S., Rodenhizer, H., Salmon, V., Schädel, C., Strauss, J., Treat, C., and Turetsky, M.: Permafrost and Climate Change: Carbon Cycle Feedbacks From the Warming Arctic, Annual Review of Environment and Resources, 47, 343–371, https://doi.org/10.1146/annurev-environ-012220-011847, 2022.

Sentinel Hub: PlanetScope, https://docs.sentinel-hub.com/api/latest/data/planet/planet-scope, [Online; accessed 21. Jun. 2024], 2024.

Smith, S. L., O'Neill, H. B., Isaksen, K., Noetzli, J., and Romanovsky, V. E.: The changing thermal state of permafrost, Nature Reviews Earth & Environment, 3, 10–23, https://doi.org/10.1038/s43017-021-00240-1, 2022.

Smith, W. D., Dunning, S. A., Ross, N., Telling, J., Jensen, E. K., Shugar, D. H., Coe, J. A., and Geertsema, M.: Revising supraglacial rock avalanche magnitudes and frequencies in Glacier Bay National Park, Alaska, Geomorphology, 425, 108 591, https://doi.org/10.1016/j.geomorph.2023.108591, 2023.

Sourcepole AG: qgis-openlayers-plugin, https://github.com/sourcepole/qgis-openlayers-plugin, [Online; accessed 28. May 2024], 2024.

Speetjens, N. J., Hugelius, G., Gumbricht, T., Lantuit, H., Berghuijs, W., Pika, P., Poste, A., and Vonk, J.: The Pan-Arctic Catchment Database (ARCADE), Earth System Science Data Discussions, 2022, 1–25, https://doi.org/10.5194/essd-2022-269, 2022.

State of Alaska Department of Environmental Conservation: About the Contaminated Sites Program, https://dec.alaska.gov/spar/csp/about, 2023a.

State of Alaska Department of Environmental Conservation: Glossary, https://dec.alaska.gov/spar/glossary.htm, 2023b.

State of Alaska Department of Environmental Conservation: Glossary. Closure of a contaminated site, https://dec.alaska.gov/spar/glossary.htm#closure, 2023c.

Stoffel, M., Trappmann, D. G., Coullie, M. I., Ballesteros Cánovas, J. A., and Corona, C.: Rockfall from an increasingly unstable mountain slope driven by climate warming, Nature Geoscience, 17, 249–254, https://doi.org/10.1038/s41561-024-01390-9, 2024.

The Information Architects of Encyclopaedia Britannica: Alaska, https://www.britannica.com/facts/Alaska, [Online; accessed 14. Sep. 2023], 2023.

van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G. C., Kram, T., Krey, V., Lamarque, J.-F., Masui, T., Meinshausen, M., Nakicenovic, N., Smith, S. J., and Rose, S. K.: The representative concentration pathways: an overview, Climatic Change, 109, 5–31, https://doi.org/10.1007/s10584-011-0148-z, 2011.

Walker, D. A., Raynolds, M. K., Kanevskiy, M. Z., Shur, Y. S., Romanovsky, V. E., Jones, B. M., Buchhorn, M., Jorgenson, M. T., Šibík, J., Breen, A. L., Kade, A., Watson-Cook, E., Matyshak, G., Bergstedt, H., Liljedahl, A. K., Daanen, R. P., Connor, B., Nicolsky, D., and Peirce, J. L.: Cumulative impacts of a gravel road and climate change in an ice-wedge-polygon landscape, Prudhoe Bay, Alaska, Arctic Science, https://cdnsciencepub.com/doi/full/10.1139/as-2021-0014, 2022.

Wang, S., Ramage, J., Bartsch, A., and Efimova, A.: Population in the Arctic Circumpolar Permafrost Region at settlement level, https://doi.org/10.5281/zenodo.4529610, 2021.

Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Loy, C. C., Qiao, Y., and Tang, X.: ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks, ArXiv e-prints, https://doi.org/10.48550/arXiv.1809.00219, 2018.

Wang, Z., Xiao, M., Nicolsky, D., Romanovsky, V., McComb, C., and Farquharson, L.: Arctic coastal hazard assessment considering permafrost thaw subsidence, coastal erosion, and flooding, Environmental Research Letters, 18, 104 003, https://doi.org/10.1088/1748-9326/acf4ac, 2023.

Warmerdam, F., Rouault, E., , and contributors, G.: GPKG – GeoPackage vector, https://gdal.org/drivers/vector/gpkg.html#gpkg-geopackage-vector, 2023.

World Bank: Population density in the United States from 2002 to 2021 (inhabitants per square kilometer), https://www.statista.com/statistics/269965/population-density-in-the-united-states/, [Online; accessed 27. Mar. 2024], 2024.

Xu, X., Liu, C., Liu, C., Hui, F., Cheng, X., and Huang, H.: Fine-resolution mapping of the circumpolar Arctic Man-made impervious areas (CAMI) using sentinels, OpenStreetMap and ArcticDEM, Big Earth Data, 6, 196–218, https://doi.org/10.1080/20964471.2022.2025663, 2022.