

Big Earth Data

ISSN: 2096-4471 (Print) 2574-5417 (Online) Journal homepage: www.tandfonline.com/journals/tbed20

Harnessing multi-source geospatial big data to construct a pan-Arctic building footprint atlas: integrating geometric, morphological, temporal, and functional attributes

Shimeng Zhao , Jiayi Wang , Xiao Cheng , Yize Guo , Qi Zhang , Peijie Wang , Weijia Li , Da He & Chong Liu

To cite this article: Shimeng Zhao , Jiayi Wang , Xiao Cheng , Yize Guo , Qi Zhang , Peijie Wang , Weijia Li , Da He & Chong Liu (16 Nov 2026): Harnessing multi-source geospatial big data to construct a pan-Arctic building footprint atlas: integrating geometric, morphological, temporal, and functional attributes, Big Earth Data, DOI: 10.1080/20964471.2026.2759385

To link to this article: https://doi.org/10.1080/20964471.2026.2759385

© 2026 The Author(s). Published by Taylor & Francis Group and Science Press on behalf of the International Society for Digital Earth, supported by the International Research Center of Big Data for Sustainable Development Goals.

View supplementary material

Published online: 16 Nov 2026.

Submit your article to this journal

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=tbed20

BIG EARTH DATA
https://doi.org/10.1080/20964471.2026.2759385

RESEARCH ARTICLE

OPEN ACCESS Check for updates

Harnessing multi-source geospatial big data to construct a pan-Arctic building footprint atlas: integrating geometric, morphological, temporal, and functional attributes

Shimeng Zhao[a], Jiayi Wang[a], Xiao Cheng[a,b], Yize Guo[a], Qi Zhang[c], Peijie Wang[a], Weijia Li[d], Da He[e] and Chong Liu[a,b]

[a]School of Geospatial Engineering and Science, Sun Yat-sen University, Guangzhou, China; [b]Ministry of Education, Key Laboratory of Comprehensive Observation of Polar Environment (Sun Yat-sen University), Zhuhai, China; [c]Department of Geography and Environment, University of North Carolina at Chapel Hill, Chapel Hill, USA; [d]Shenzhen International Graduate School, Tsinghua University, Shenzhen, China; [e]School of Geography and Planning, Sun Yat-sen University, Guangzhou, China

ABSTRACT

Arctic infrastructure is increasingly exposed to permafrost degradation and thaw-related geohazards, yet building-resolved information remains insufficient for translating regional climate risk into object-level exposure assessments. Existing settlement, impervious-surface, and land-cover products delineate broad built-up extent but lack the object-level detail needed to characterize individual Arctic buildings. Here, we construct a harmonized multi-attribute pan-Arctic building footprint atlas by integrating complementary building-footprint products with information from multiple geospatial sources. The three input products were evaluated against 39,685 manually interpreted reference buildings across 12 Arctic benchmark sites, then integrated through hierarchical geometric conflation, redundancy removal, and multi-source attribute enrichment. Independent validation against regional reference building footprints yielded an overall area-based *F1-score* of 77.3% for the final fused atlas. The resulting atlas identifies approximately 15.08 million buildings north of 60°N, covering about 2,004 km^2, and provides object-level geometry together with available morphological, temporal, and functional attributes. The atlas reveals a highly uneven Arctic built environment, with dense concentrations along the southern Arctic margin, particularly in northern Europe, western Russia, southern Alaska, and Iceland, whereas the high Arctic remains sparsely built. Linking the atlas with environmental layers shows that, among footprints with valid permafrost assignments, 31.7% of footprint area and 21.9% of buildings occur in locations with permafrost probabilities of ≥50%. Under Representative Concentration Pathway 8.5, high-hazard-potential zones contain 51% of the footprint area with valid hazard assignments by 2061–2080. The Pan-Arctic Building Footprint Atlas v1 is publicly available via https://doi.org/10.5281/zenodo.20796872.

ARTICLE HISTORY
Received 5 July 2026
Accepted 13 October 2026

KEYWORDS
Building footprint; pan-Arctic; multi-source data fusion; object- level atlas

CONTACT Chong Liu liuc@mail.sysu.edu.cn School of Geospatial Engineering and Science, Sun Yat-sen University, Guangzhou, China

Supplemental data for this article can be accessed online at https://doi.org/10.1080/20964471.2026.2759385

© 2026 The Author(s). Published by Taylor & Francis Group and Science Press on behalf of the International Society for Digital Earth, supported by the International Research Center of Big Data for Sustainable Development Goals.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

1. Introduction

The Arctic is warming nearly four times faster than the global average (Rantanen et al., 2022), triggering cascading cryospheric transformations with far-reaching socio-economic consequences (Li et al., 2025; Moon et al., 2019; Myers-Smith et al., 2020). Foremost among these is the widespread degradation of ice-rich permafrost, which underlies roughly a quarter of the Northern Hemisphere's landmass and serves as the structural foundation for pan-Arctic human settlements and strategic industrial networks (Chen et al., 2023; Hjort et al., 2022; Obu, 2021). As ground ice melts, previously stable terrain is increasingly subjected to thermokarst subsidence, active layer thickening, and the formation of retrogressive thaw slumps. These processes are often amplified by localized anthropogenic disturbances such as vegetation removal and infrastructure-induced heat fluxes, shifting the built environment into a regime of escalating geohazard risk (Freitas et al., 2025; Hjort et al., 2022; Liu et al., 2026). Recent projections indicate that up to 70% of current Arctic infrastructure is situated in areas with high potential for near-surface thaw by 2050, threatening tens of billions of dollars in structural damage across regions such as Alaska and Russian Arctic (Hjort et al., 2018; Liu et al., 2024; Manos et al., 2025; Schneider von Deimling et al., 2021). Therefore, a consistent, building-resolved infrastructure inventory is needed to link individual structures with regional hazard projections and support spatially explicit assessments of exposure across the circumpolar North.

As the most granular physical units of human activity, buildings constitute the foundational elements of the pan-Arctic anthropogenic landscape and the primary loci of geotechnical vulnerability (Esch et al., 2022; Li et al., 2022; Tanguy et al., 2024). To date, satellite remote sensing has enabled the monitoring of built-up environments predominantly through moderate-resolution land-cover classifications (Li & Gong, 2016; Zhou & Weng, 2024) and multi-temporal impervious surface records (Gong et al., 2020; Liu et al., 2019; Ren et al., 2025). While these efforts provide invaluable insights into the broad distribution and historical expansion of human activities, they remain inherently limited in resolving the fine-scale spatial configuration and morphology of individual buildings. Such object-level granularity is essential for infrastructure-specific risk quantification and supports progress toward the Sustainable Development Goals in rapidly changing Arctic systems (Bartsch et al., 2021; Liu et al., 2022). Moreover, the limited availability of consistent building-resolved information constrains the integration of geophysical process models with socio-economic exposure datasets, limiting our ability to couple environmental change with infrastructure-level impacts (Bergmann et al., 2022). Recent advances in very-high-resolution (VHR) satellite imagery, coupled with rapid progress in deep learning techniques and petabyte-scale cloud computing capabilities, have made large-scale extraction of building footprints increasingly feasible. These developments have enabled the emergence of global datasets, such as Google Open Buildings, Microsoft's machine-learning-derived global building footprints, and the crowdsourced OpenStreetMap (OSM), that map the built environment with unprecedented geometric detail. Collectively, these datasets mark a transition from pixel-based representations of human activity to object-based geospatial intelligence, opening new avenues for fine-scale environmental risk analysis.

However, applying existing global building-mapping efforts to high-latitude environments exposes critical limitations, leaving the pan-Arctic substantially underrepresented (Raynolds et al., 2014). Optical approaches are challenged by persistent cloud cover, prolonged polar night, low solar illumination, and the weak spectral contrast between buildings and snow- or ice-covered surfaces, resulting in systematic omissions and geometric inconsistencies (Bartsch et al., 2016; Liu et al., 2023). Meanwhile, crowdsourced datasets such as OSM exhibit pronounced socio-geographic biases and variable data quality, including inconsistencies in positional accuracy, completeness, and mapping standards (Oostwegel et al., 2025). These biases are particularly acute in sparsely populated Arctic regions, where mapping activity is limited and often concentrated around a few accessible settlements. Importantly, because existing building footprint products are derived from different data sources and methodological designs, their spatial gaps are often complementary. This heterogeneity, while problematic for standalone applications, also creates an opportunity for synergistic integration through multi-source geospatial data fusion (Cui et al., 2024; Li et al., 2026; Liu et al., 2024). Moreover, assessing building-level exposure to permafrost thaw hazards requires information beyond geometric completeness, including structural morphology, functional context, and temporal persistence. Building height characterizes the structural scale and vertical morphology of exposed assets, while temporal attributes provide information on the timing of infrastructure development. Functional information further distinguishes residential, industrial, public-service, and other building types that differ in societal role and adaptation priority. These complementary attributes enable a more differentiated interpretation of permafrost-related exposure across the Arctic built environment. Although several recent studies have incorporated such attributes at regional or global scales (Che et al., 2024; Ren et al., 2025; Zhu et al., 2025), available products differ in geographic coverage, source data, spatial representation, and attribute availability. These differences motivate an integration that combines complementary information while retaining the provenance of individual building records.

Here, we construct a harmonized multi-attribute pan-Arctic building footprint atlas by integrating existing building-footprint products as the geometric foundation with complementary geospatial datasets. Our objective is to characterize the pan-Arctic built environment at the object level by resolving building geometry and integrating available information on structural morphology, functional composition, and temporal development. By addressing these fundamental spatial data deficits, this circumpolar inventory provides an object-level empirical basis for assessing infrastructure exposure to permafrost-related hazards and supporting localized climate-adaptation planning in Earth's most rapidly warming frontier.

2. Study area and data

2.1. Study area and input-product evaluation sites

The study domain encompasses the terrestrial pan-Arctic region, defined here as all northern landmasses situated above 60°N latitude (Figure 1). Permafrost occurrence

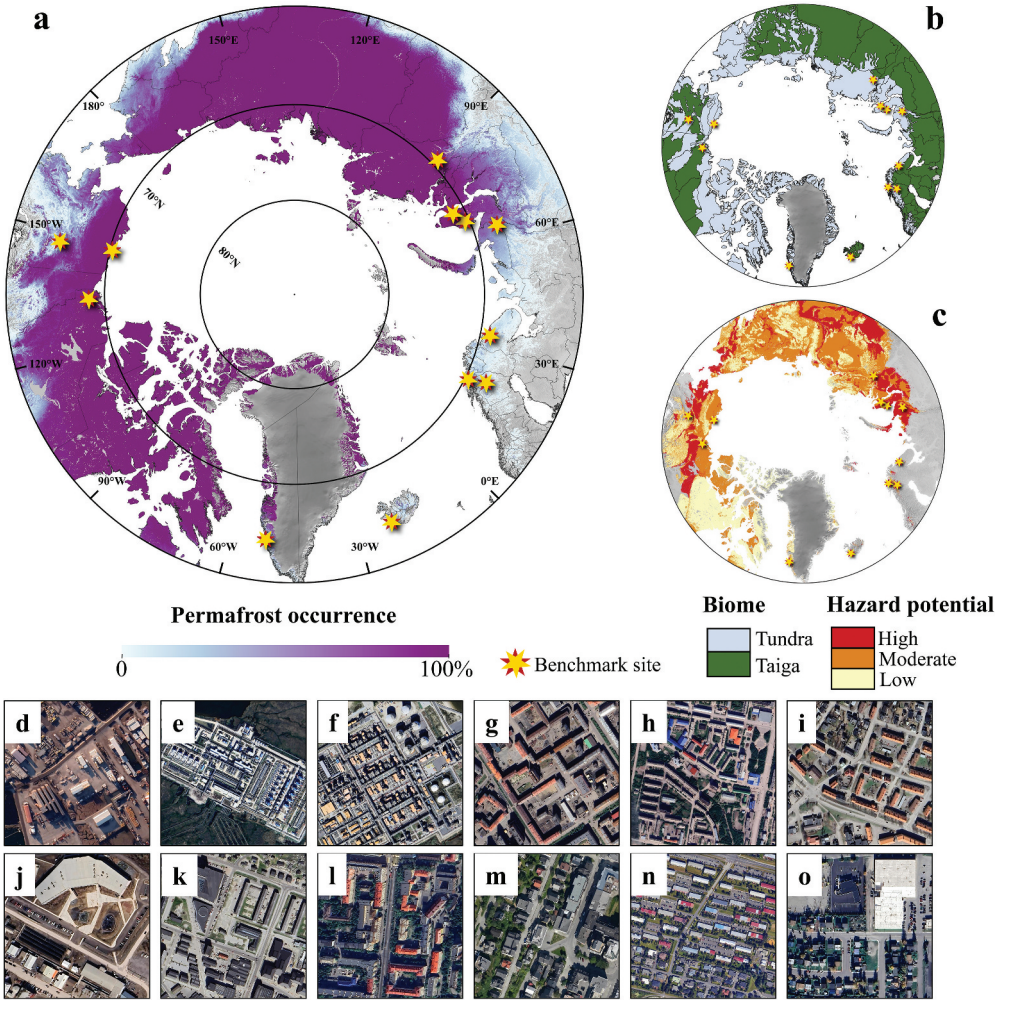


Figure 1. Geo-environmental context of the pan-Arctic study area and input-product evaluation sites. a) Spatial distribution of permafrost occurrence, showing the circumpolar gradient from 0% to 100% areal coverage (Obu et al., 2019). Yellow markers indicate the 12 benchmark sites used for input-product evaluation. b) Distribution of major ecological biomes, highlighting tundra and boreal forest/taiga regions based on the World Wildlife Fund (WWF) global classification (Dinerstein et al., 2017). c) Thaw-related infrastructure hazard potential by 2050 under the representative concentration Pathway 8.5 (RCP8.5) scenario, classified from low to high (Hjort et al., 2018). d–o) Examples of very-high-resolution (VHR) satellite imagery illustrating representative Arctic built-environment types and architectural morphologies: Prudhoe Bay (d), Bovanenkovo (e), Sabetta (f), Norilsk (g), Vorkuta (h), Kiruna (i), Inuvik (j), Nuuk (k), Murmansk (l), Tromsø (m), Reykjavík (n), and Fairbanks (o). The thematic layers in panels a–c are shown on the Natural Earth basemap (https://www.naturalearthdata.com/).

varies markedly across this vast circumpolar domain, from near-continuous coverage in the high Arctic to increasingly sparse and fragmented occurrence toward lower-latitude margins (Figure 1(a)). The study area also spans diverse ecological settings, predominantly characterized by tundra and boreal forest/taiga biomes as delineated by the World Wildlife Fund (WWF) global ecoregion classification

(Figure 1(b); Dinerstein et al., 2017). Projected thaw-related geohazard potential is likewise spatially heterogeneous, with areas of high hazard potential by 2050 distributed across parts of the circumpolar domain (Figure 1(c); Hjort et al., 2018). Some of these projected hazard-prone areas also contain existing Arctic settlements and infrastructure, including those located along coastal regions and within permafrost-affected landscapes. These cryospheric, ecological, and projected geohazard gradients provide the broader environmental context within which the pan-Arctic built environment is distributed.

To evaluate the performance and complementarity of the input building-footprint products, we established a benchmark comprising 12 sites across the circumpolar North (Figure 1(d–o)). Rather than constituting a statistically representative random sample of the entire pan-Arctic domain, these sites were selected to span major built-environment types, building densities, architectural morphologies, and background conditions encountered in Arctic regions. The benchmark sites encompass four broad built-environment contexts: isolated oil and gas extraction fields, represented by Prudhoe Bay, Bovanenkovo, and Sabetta; resource-dependent mining and industrial cities, represented by Norilsk, Vorkuta, and Kiruna; remote administrative outposts and coastal communities, represented by Inuvik and Nuuk; and modern Arctic metropolitan hubs, represented by Murmansk, Tromsø, Reykjavík, and Fairbanks. Together, these sites provide a diverse benchmark for examining product behavior across sparse industrial facilities, compact urban fabrics, remote settlements, and complex Arctic environments. Their purpose is to characterize source-product accuracy and complementarity under contrasting conditions rather than to provide a probabilistic estimate of pan-Arctic accuracy.

2.2. Data sources

A diverse set of geospatial datasets was assembled to support building-footprint integration, attribute enrichment, post-processing, validation, and subsequent characterization of the Arctic built environment. These datasets can be grouped into four categories: 1) contemporary building-footprint and settlement-coverage products, 2) datasets used for attribute enrichment and post-processing, 3) land-cover, permafrost, and geohazard datasets used for contextual analyses, and 4) VHR imagery and independent building references used for validation. Table S1 summarizes the spatial and temporal coverage, spatial resolution, key information, and specific role of each dataset in this study.

2.2.1. Building footprint and settlement-coverage datasets

Three contemporary vector building-footprint products were used as the foundational geometric inputs: OpenBuildingMap (OBM; Oostwegel et al., 2025), GlobalBuildingAtlas (GBA; Zhu et al., 2025), and the High-resolution Arctic Built Infrastructure and Terrain Analysis Tool (HABITAT; Manos, 2025). These products differ in their source data, mapping strategies, spatial coverage, and associated building attributes, providing complementary information for constructing the pan-Arctic atlas. OBM is a global building-footprint database containing approximately 2.7 billion building polygons generated through the conflation of multiple artificial intelligence (AI)-derived building datasets with OSM. In addition to building geometry, OBM incorporates occupancy and operational information

harmonized using the Global Earthquake Model taxonomy, which provides the principal source of functional-context information used in this study (Oostwegel et al., 2025). GBA provides global building polygons together with building-height estimates and Level of Detail 1 (LoD1) three-dimensional models (Zhu et al., 2025). HABITAT was included as an Arctic-oriented supplementary source to provide additional candidate footprints in areas not represented by the two global products. Because the three products were developed from different upstream datasets and mapping strategies, their relative performance and spatial complementarity were explicitly evaluated before integration, as described in Section 3.1.

Two built-environment products were further used to examine broad-scale spatial coverage. The World Settlement Footprint 2019 (WSF2019; Marconcini et al., 2020, 2021) is a global 10-m binary settlement product derived from the synergistic use of optical and synthetic aperture radar (SAR) observations. It provides continuous delineation of settlement extent across the pan-Arctic domain, although it does not resolve individual building geometries. The Sentinel-1/2-derived Arctic Coastal Human Impact dataset version 2 (SACHI-v2; Bartsch et al., 2023; Tanguy et al., 2024) maps Arctic settlements and infrastructure using Sentinel-1 and Sentinel-2 observations acquired during 2016–2020. SACHI-v2 was developed primarily for Arctic coastal regions and extends beyond the original 100-km coastal buffer in selected areas, providing an Arctic-specific representation of built infrastructure. In this study, WSF2019 was used as the primary full-domain benchmark for evaluating settlement-scale spatial coverage across the terrestrial pan-Arctic region, whereas SACHI-v2 provided a supplementary Arctic-specific benchmark within its available spatial extent. Neither dataset was treated as an object-level ground-truth reference or as a measure of absolute building completeness. Instead, they were used to assess broad-scale coverage consistency and regional omission patterns of the contemporary building footprint products. Detailed implementation of these assessments is described in Section 3.1.3.

2.2.2. Attribute-enrichment and post-processing datasets

Several complementary remote sensing datasets were used to enrich the morphological and temporal information associated with the fused building footprints and to support post-processing of potentially spurious detections. Building-height information was supplemented using the 10-m High-resolution Global Building Height Map (GBHM; Chen et al., 2024). GBHM provides globally consistent building-height estimates derived from Sentinel-1 SAR and Sentinel-2 optical observations using a Random Forest regression framework. Its spatially continuous raster representation provides complementary height information for fused building footprints without valid height attributes from GBA. Accordingly, GBHM was used in this study to fill these height-data gaps. The resulting building-height attribute was used to characterize broad-scale vertical building morphology and should be interpreted as satellite-derived height information rather than engineering-grade measurements. Temporal information was derived from the improved 30-m Global Impervious Surface Area (GISA) dataset (Huang et al., 2022; Ren et al., 2025), which provides a multi-temporal record spanning 1972–2021, including observations for 1972 and 1978 and annual observations from 1985 onward. The temporal sequence was used to identify the approximate timing at which impervious surfaces associated with individual building footprints first emerged. We therefore refer to this attribute as the

GISA-derived impervious-surface emergence year, from which an approximate building age in 2021 was calculated. This temporal attribute provides a remote sensing proxy for the timing of building development and should not be interpreted as a cadastral construction year. Sentinel-2 surface reflectance imagery and the Joint Research Centre (JRC) Global Surface Water v1.4 dataset (Pekel et al., 2016) were additionally used during post-processing. The normalized difference vegetation index (NDVI) derived from Sentinel-2 imagery was used to construct a growing-season maximum NDVI metric ($NDVI_{max}$) for identifying highly vegetated detections unlikely to represent buildings, while the JRC Global Surface Water dataset provided information on persistent surface water for screening water-related false detections. The temporal compositing, polygon-level statistics, and specific screening criteria applied to these datasets are described in Section 3.2.1.

2.2.3. Environmental and geohazard datasets

Three geospatial datasets were used to characterize the broader environmental and cryospheric settings associated with Arctic buildings. Land-cover information was obtained from the 2021 Moderate Resolution Imaging Spectroradiometer (MODIS) MCD12Q1 Land Cover Type 1 product (Friedl & Sulla-Menashe, 2022), which provides annual global land-cover classifications at 500-m spatial resolution following the International Geosphere–Biosphere Programme (IGBP) scheme. In this study, the MODIS land-cover classes were used as a spatially consistent contextual layer for characterizing the broader landscape setting of the pan-Arctic built environment. Permafrost conditions were characterized using the Northern Hemisphere permafrost map developed by Obu et al. (2019). The dataset represents modeled permafrost conditions for the 2000–2016 period at approximately 1-km spatial resolution and provides spatially distributed estimates of permafrost probability within its mapped domain. It was used to characterize the distribution of Arctic buildings along the permafrost-occurrence gradient. Projected thaw-related geohazard conditions were derived from the infrastructure hazard-potential maps of Hjort et al. (2018), which classify projected hazard potential associated with future near-surface permafrost degradation under different Representative Concentration Pathway (RCP) scenarios and time periods. The published low-, moderate-, and high-hazard-potential classes were spatially linked with the building footprints under the corresponding projections. These classes were adopted directly from the consensus geohazard index developed by Hjort et al. (2018), without reclassification in the present study. The existing geohazard layers were used to contextualize the building atlas, and no new permafrost-hazard model or building-damage function was developed in this study. Detailed overlay and summarization procedures are described in Section 3.4.

2.2.4. Validation imagery and reference building data

VHR optical imagery available through Google Earth was used to generate manually interpreted reference building footprints for evaluating the three input building-footprint products. For the 12 benchmark sites described in Section 2.1, we selected recent cloud-free images with the highest available spatial resolution, typically approximately 0.25–0.5 m. Image acquisitions were restricted to 2022–2025 to improve temporal consistency with the contemporary building-footprint products, and growing-season imagery was preferred where available to minimize snow and

ice obscuration. Individual building roof outlines were subsequently delineated as reference polygons following the annotation procedure described in Section 3.1.1. To provide an additional evaluation of the final fused atlas independent of both the input products and the manually interpreted benchmark, we assembled regional reference building-footprint datasets for Anchorage, Sodankylä, Sisimiut, and selected settlements in Yukon and the Northwest Territories. These datasets were obtained from the Municipality of Anchorage Building Footprints, the National Land Survey of Finland Topographic Database, the NunaGIS/Asiaq Technical Base Map, and Statistics Canada's Open Database of Buildings. None was used in atlas construction or attribute enrichment. Because their acquisition dates, mapping procedures, building definitions, and spatial coverage differ among regions, they were treated as independent regional reference datasets rather than as a uniform, error-free pan-Arctic ground truth. Detailed preparation and validation procedures are described in Section 3.3.

3. Methods

The overall methodological framework is summarized in Figure 2 and comprises five interconnected components. First, the three contemporary building footprint products were evaluated using site-based reference footprints and pan-Arctic spatial coverage benchmarks to characterize their relative accuracy, coverage, and complementarity.

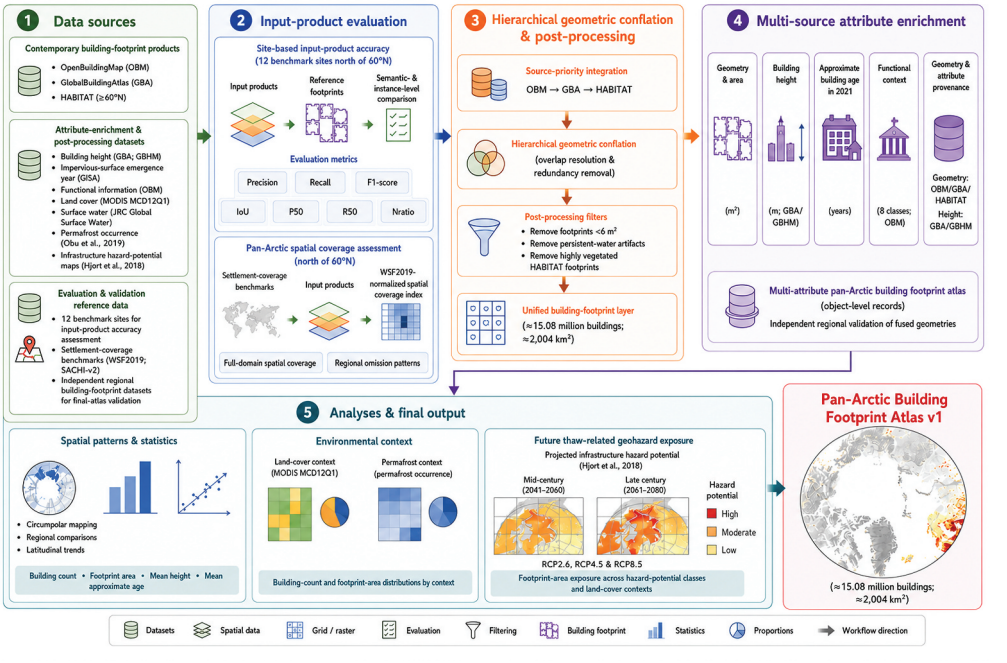


Figure 2. Workflow for constructing, evaluating, enriching, and analyzing the pan-Arctic building footprint atlas. This workflow illustration was prepared with assistance from GPT-6.0. The authors reviewed and verified its content against the study methods and take full responsibility for its integrity.

Second, the input products were integrated through hierarchical geometric conflation, followed by post-processing to remove potentially spurious detections. Third, the fused footprints were enriched with geometric, morphological, temporal, and functional attributes using complementary geospatial datasets. Fourth, the resulting atlas was independently evaluated using reference building datasets from local or national mapping authorities in selected Arctic regions. Finally, the atlas was used to characterize circumpolar spatial patterns and regional and latitudinal variations, and was further linked with land-cover, permafrost, and projected thaw-related geohazard datasets to assess environmental context and future exposure. Detailed procedures for each component are described in the following subsections.

3.1. Evaluation of three contemporary building-footprint datasets

3.1.1. Benchmark sample generation

To generate reference samples for site-level accuracy assessment, we manually delineated building footprints across the 12 benchmark sites using sub-meter Google Earth imagery. Image tiles of $1{,}024 \times 1{,}024$ pixels were generated for each site, with a sliding-window step of 256 pixels to reduce edge truncation and preserve sufficient overlap for spatial reconstruction. Tiles containing no visible buildings, including areas dominated by water, bare ground, vegetation, or snow/ice, were manually excluded before annotation. After this screening, 4,999 image tiles were retained for expert interpretation. Building roof outlines were then manually labeled as per-instance polygons using the Computer Vision Annotation Tool (CVAT, https://www.cvat.ai/). A flexible vertex-based digitization strategy was adopted to capture both simple residential structures and complex industrial facilities. To improve consistency, all annotations were checked through secondary manual correction and independent review, with particular attention to edge-truncated buildings, attached structures, shadows, and visually ambiguous objects. The final benchmark dataset contains 39,685 reference building instances, with a mean density of 7.94 buildings per tile. These samples cover three major Arctic built-environment contexts: settlements (32,029 instances), mining areas (6,816 instances), and oil and gas fields (840 instances). This benchmark dataset provides the reference basis for evaluating the geometric and instance-level performance of the three contemporary building footprint products.

3.1.2. Site-based accuracy assessment

To evaluate the performance of the contemporary building footprint products, we implemented a dual-level accuracy assessment across the 12 benchmark sites, combining semantic-level pixel agreement with instance-level object detection. Here, semantic-level assessment refers to the binary classification of each pixel as building or non-building. At this level, each predicted footprint product was rasterized and compared with the reference building labels generated from VHR imagery. For each site, true positive (*TP*), false positive (*FP*), and false negative (*FN*) pixels were aggregated across all validation tiles before metric calculation. We then computed four standard semantic-level accuracy metrics: *Precision*, *Recall*, *F1-score*, and Intersection over Union (*IoU*), which quantify the spatial agreement between predicted and reference building areas.

At the instance level, each discrete building polygon was treated as an independent analytical unit. Predicted and reference polygons were matched on a one-to-one basis according to their *IoU*, consistent with established object-detection and building-footprint evaluation practices (Everingham et al., 2010; Shermeyer et al., 2020). For a given *IoU* threshold t (with $t = 0.50$ for the primary instance-level assessment), a predicted polygon and a reference polygon were considered a valid match when their *IoU* was equal to or greater than t. Each predicted and reference polygon was allowed to participate in only one match. Successfully matched pairs were counted as true-positive instances (TP_t), unmatched predicted polygons as false-positive instances (FP_t), and unmatched reference polygons as false-negative instances (FN_t). Based on these classifications, we calculated the following instance-level metrics:

(1) Instance *Precision* at $IoU = 0.50$ ($P50$): Calculated as $TP_{0.50} / N_{pred}$, where N_{pred} is the total number of predicted building instances. This metric quantifies the proportion of predicted buildings that can be successfully matched to reference buildings at an *IoU* threshold of 0.50, with higher values indicating fewer commission errors.
(2) Instance *Recall* at $IoU = 0.50$ ($R50$): Calculated as $TP_{0.50} / N_{GT}$, where N_{GT} is the total number of reference building instances. This metric measures the proportion of reference buildings successfully detected at an *IoU* threshold of 0.50, with higher values indicating fewer omission errors.
(3) Instance Number Ratio (N_{ratio}): Calculated as N_{pred} / N_{GT}, where N_{pred} and N_{GT} are the numbers of predicted and reference building instances, respectively. This metric describes the relative number of mapped building objects but does not distinguish among different causes of count disagreement. A value greater than 1 indicates more predicted than reference instances and may reflect over-segmentation and/or commission errors, whereas a value below 1 indicates fewer predicted instances and may reflect under-segmentation and/or omission errors.

3.1.3. Pan-Arctic spatial coverage assessment

To assess broad-scale spatial coverage of the three contemporary building footprint products across the full terrestrial pan-Arctic domain, we used WSF2019 as a consistent settlement-scale baseline. WSF2019 was selected primarily because it provides spatially continuous 10-m coverage across the study domain and integrates optical and SAR observations, offering a full-domain representation of settlement extent under the cloud and illumination constraints common at high latitudes. However, WSF2019 represents settlement extent rather than individual building footprints and may itself contain omission and commission errors. In addition, its 2019 reference year predates some of the contemporary building footprint products evaluated here. We therefore did not treat WSF2019 as an object-level ground truth or as a measure of absolute building completeness. Instead, it was used as a full-domain benchmark for comparing macro-scale settlement-level coverage and identifying broad regional omission patterns among the three input products.

The assessment was implemented using a standardized 10×10 km grid covering the study domain. This scale was selected as a practical compromise between retaining settlement-scale spatial variability and obtaining stable area-based summaries across the sparse and unevenly distributed Arctic built environment. For each grid cell, we

calculated the total footprint area mapped by each building footprint product and compared it with the settlement area delineated by WSF2019. The resulting area ratio was interpreted as a WSF2019-normalized spatial coverage index. A value of 0 indicates that no building footprints were mapped within a WSF2019-delineated settlement grid and therefore suggests potential omission relative to the WSF2019 baseline. Values approaching 1 indicate closer agreement in aggregated mapped area, whereas values greater than 1 do not necessarily represent commission errors because they may also arise from differences between settlement-extent and building-footprint definitions, temporal change after 2019, or the detection of isolated infrastructure not fully represented by WSF2019. Accordingly, this metric was used for relative coverage comparison rather than as a direct measure of object-level accuracy or absolute completeness.

To examine the sensitivity of the WSF2019-based results to the choice of settlement-scale benchmark, we conducted a supplementary comparison using SACHI-v2 within its available geographic extent. Only the buildings-and-other-constructions class (class 20) was retained. All other SACHI-v2 classes, including transport infrastructure, other impacted areas, airstrips, and water-related features, were excluded. The same standardized 10×10 km grid was used, but a grid cell was included only when its entire area fell within the valid SACHI-v2 coverage extent. For each retained grid cell, the SACHI-v2-normalized spatial coverage index was calculated as the building-footprint area mapped by each input product divided by the corresponding SACHI-v2 class-20 area. The same product-area numerators were also compared with WSF2019 within the identical grid subset. Area-weighted indices were calculated as the summed product footprint area divided by the summed reference area across the retained grids. After applying these criteria, 777 grid cells with positive SACHI-v2 class-20 and WSF2019 reference areas were included in the paired comparison.

3.2. Multi-source data fusion and attribute enrichment

3.2.1. Hierarchical geometric conflation

To establish the geometric foundation of the pan-Arctic building atlas, we implemented a hierarchical geometric conflation strategy to integrate OBM, GBA, and HABITAT. The priority order for resolving overlapping geometries was determined primarily from the aggregate site-level geometric-accuracy assessment, while the pan-Arctic spatial-coverage assessment was used to evaluate source complementarity and the value of retaining non-overlapping footprints (Section 3.1; corresponding results in Section 4.1, Table 1, and Figures 3 and 5). Accordingly, overlapping geometries were resolved in the order OBM,

Table 1. Semantic- and instance-level accuracy of OBM, GBA, and HABITAT across the 12 benchmark sites. Abbreviations: OBM, OpenBuildingMap; GBA, GlobalBuildingAtlas; HABITAT, High-resolution Arctic Built Infrastructure and Terrain Analysis Tool; *IoU*, Intersection over Union; *P*50 and *R*50, instance precision and recall at an *IoU* threshold of 0.50, respectively; N_{ratio}, ratio of the number of predicted building instances to the number of reference building instances.

Product	*Precision* (%)	*Recall* (%)	*F1-score* (%)	*IoU* (%)	*P*50(%)	*R*50(%)	N_{ratio}
OBM	73.9	71.6	72.8	57.2	36.9	54.1	1.47
GBA	72.0	70.9	71.5	55.6	35.2	50.3	1.43
HABITAT	68.9	28.9	40.7	25.6	25.0	12.7	0.508

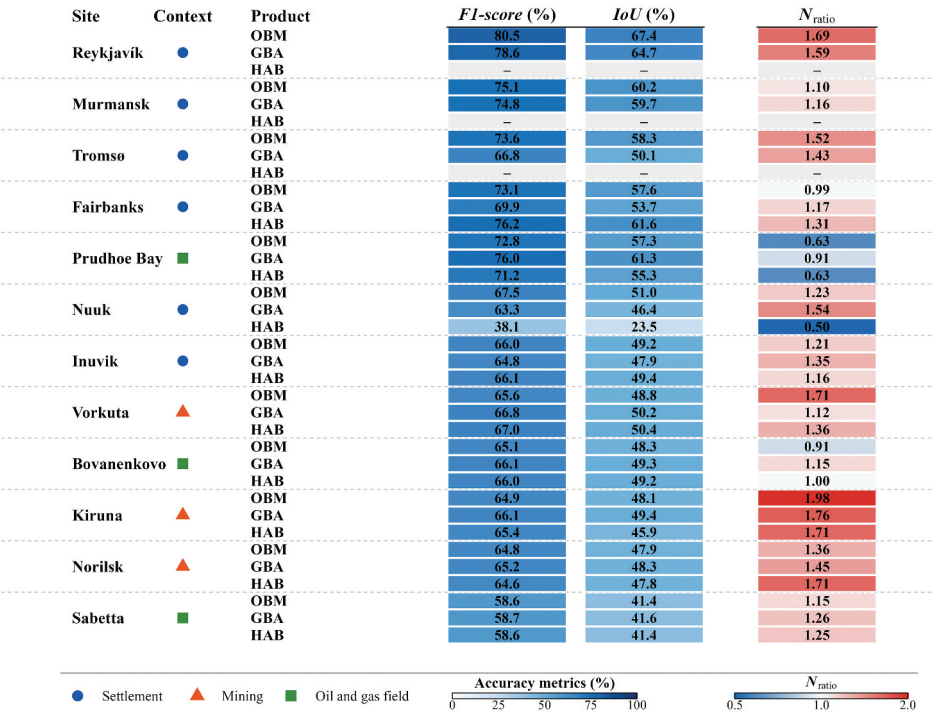


Figure 3. Site-level accuracy of three contemporary building-footprint products across 12 Arctic benchmark sites. Each row represents one product–site combination, with sites ordered by the *F1-score* of OpenBuildingMap. OBM and GBA denote OpenBuildingMap and GlobalBuildingAtlas, respectively. HAB is used only as a shortened figure label for HABITAT because of space constraints. Symbol colors and shapes distinguish three broad built-environment contexts: settlements, mining areas, and oil and gas fields.

GBA, and HABITAT. This ordering served only as a transparent and reproducible duplicate-resolution rule and was not intended to imply that OBM was universally superior across the pan-Arctic domain. During conflation, OBM geometries were retained where products overlapped. GBA geometries not overlapping OBM were then added, followed by HABITAT geometries not overlapping either global product. This procedure reduced redundant detections while preserving complementary footprints from the lower-priority sources, consistent with precedence-based integration strategies adopted in previous large-scale building-footprint studies (Manos et al., 2026; Oostwegel et al., 2025).

After hierarchical integration, we applied a series of post-processing filters to remove morphologically or environmentally implausible detections. First, polygons smaller than 6 m^2 were removed to suppress noise and widespread false positives such as bright geological outcrops, rocks, and mobile objects (Oostwegel et al., 2025). Persistent-water artifacts were then screened using the seasonality layer of the JRC Global Surface Water v1.4 dataset (Pekel et al., 2016). For each candidate footprint, the modal (most frequent) seasonality value of the intersecting JRC Global Surface Water pixels was determined. Because the seasonality layer records the number of months in which surface water is present, ranging from 0 to 12, footprints dominated by the

maximum value of 12 months were removed as persistent-water artifacts. Vegetation-related false positives were screened using Sentinel-2 surface reflectance imagery from 2019 to 2021. For each year, NDVI was calculated from the near-infrared (B8) and red (B4) bands after cloud, cloud-shadow, cirrus, and snow/ice masking, and the maximum NDVI ($NDVI_{max}$) during June–September was retained. The final $NDVI_{max}$ metric was calculated as the median of the three annual growing-season maxima and sampled at each building centroid at 10-m resolution. Because site-level inspection showed that strongly vegetated false detections occurred primarily in HABITAT, the $NDVI_{max}$ filter was applied only to HABITAT-derived footprints, with polygons exceeding 0.75 removed. The threshold was selected empirically as a conservative cutoff for strongly vegetated false detections while minimizing the removal of plausible buildings. Together, this hierarchical conflation and filtering procedure produced a geometrically consolidated building footprint layer for subsequent attribute integration and pan-Arctic analysis.

3.2.2. Integration of geometric, morphological, temporal, and functional attributes

Following hierarchical geometric conflation, we enriched the unified building footprint layer with available object-level attributes from the source products and complementary datasets. To maximize attribute availability, we exploited the spatial overlap between the fused polygons and the original source products. When a unified polygon intersected features from multiple input datasets, complementary information was extracted and harmonized across sources rather than relying only on the geometry source retained during conflation. Nevertheless, not every building record contains complete height, temporal, and functional information, because these attributes were derived from heterogeneous source datasets with different spatial coverage and attribute availability. Geometric attributes were derived directly from the fused polygon geometry, including footprint area and related planar measurements.

Morphological attributes focused primarily on building height. For footprints with valid GBA height information, we retained the native satellite-derived height estimates as the primary source. For polygons without valid GBA height attributes, we supplemented height information using the GBHM raster. Specifically, each unified polygon was overlaid with the GBHM layer, and valid GBHM pixels intersecting the building footprint were extracted. Non-positive or otherwise invalid values were excluded, and the median of the remaining pixel heights was assigned as the representative GBHM-based height for the building. Where neither source provided reliable height information, the height attribute was retained as missing. GBA height was retained when available because it is natively associated with individual building polygons, providing a direct object-level height attribute consistent with the spatial units of the atlas. GBHM, by comparison, provides height as a spatially continuous 10-m raster and was therefore used to fill gaps where the native GBA height attribute was unavailable. This source hierarchy was designed primarily to preserve the object-level information already associated with GBA footprints while exploiting the complementary spatial coverage of GBHM. To characterize cross-product correspondence and systematic differences between the two height sources, we compared GBA and GBHM estimates for building footprints with valid positive values from both datasets. GBHM heights were extracted using the same polygon-based procedure adopted for gap filling. The

paired estimates were summarized using Pearson's correlation coefficient (*r*), mean absolute error (*MAE*), root mean square error (*RMSE*), and mean signed difference (GBHM – GBA). Variation in the cross-product relationship was further examined across five GBA-defined height classes (<3, 3–<6, 6–<12, 12–<20, and ≥20 m) and across six Arctic regions (Alaska, Canada, Greenland, Iceland, northern Europe, and Russia). GBA was used to define height-class membership because it constitutes the primary object-associated height source in the final atlas. Because neither product served as an independent height reference, these statistics characterize cross-product correspondence rather than absolute height accuracy.

Temporal attributes were estimated using the GISA multi-temporal impervious-surface record. Each unified polygon was spatially overlaid with the multi-temporal GISA layer, and valid impervious-surface emergence-year values within the footprint were extracted. The modal year, representing the most spatially prevalent valid year within the polygon, was assigned as the GISA-derived impervious-surface emergence year. The approximate building age in 2021 was then calculated as the difference between 2021 and this emergence year. The year 2021 was selected as the common reference year because it represents the final year of the GISA temporal record used in this study. Accordingly, this attribute indicates the elapsed time since the first detected impervious-surface signal associated with a building footprint and should be interpreted as an approximate temporal proxy rather than the actual construction age of an individual building. As with height and function, temporal information was retained as missing where no reliable GISA-based emergence year could be assigned. Functional attributes were derived exclusively from the occupancy and operational information available in OBM. These metadata were harmonized into broad functional categories, including residential, commercial, industrial, governmental, educational, agricultural, assembly, and mixed-use buildings. Because functional information was unavailable for footprints not linked to OBM or lacking valid source metadata, missing functional labels were preserved rather than inferred.

To preserve source provenance in the released atlas, each building record includes a source field identifying the input product from which the retained footprint geometry originated (OBM, GBA, or HABITAT). Because building height may be obtained from two alternative sources, a separate height source field records whether the final height estimate was derived from GBA or GBHM. Provenance for attributes with a single derivation pathway is defined directly by their source datasets: the GISA-derived impervious-surface emergence year is derived from GISA, and approximate building age in 2021 is calculated from this emergence year, whereas functional information is inherited from OBM. These provenance fields and source relationships allow users to trace the origin of the principal geometry and attributes associated with each building record.

3.3. Independent validation of the fused atlas

To provide an evaluation independent of the 12 benchmark sites and the datasets used in atlas construction, the final fused atlas was compared with regional reference building-footprint datasets. All datasets were transformed to EPSG:3413 before spatial comparison. Because formal source-coverage polygons were not consistently available for all reference datasets, common evaluation domains were defined using complete, EPSG:3413-aligned

1×1 km cells containing at least one reference-building centroid. All atlas and reference footprints whose centroids fell within these cells were retained. This procedure avoided selecting atlas footprints according to their proximity to individual reference buildings while restricting the analysis to cells with demonstrated reference-data support. For the area-based assessment, footprint geometries were clipped to the union of the included cells. Area-based metrics (*Precision*, *Recall*, *F1-score*, and *IoU*) were calculated from the intersection and union of atlas and reference footprint areas within each regional evaluation domain.

At the instance level, atlas and reference polygons were matched one-to-one according to *IoU* using the procedure described in Section 3.1.2. *P*50, *R*50, and N_{ratio} were used as the primary instance-level metrics. To examine sensitivity to increasingly stringent geometric-overlap requirements, the matching procedure was additionally repeated at ten *IoU* thresholds from 0.50 to 0.95 in increments of 0.05. *P*50–95 and *R*50–95 were calculated as the arithmetic means of the corresponding instance-precision and instance-recall values across these thresholds. Because the atlas footprints do not have a confidence-ranking framework, these quantities are not equivalent to COCO-style average precision or average recall and were treated as supplementary measures of cross-source geometric agreement. The complete threshold-dependent results are reported in Figure S1. Overall results were micro-aggregated across the five regions by summing intersecting areas, atlas and reference areas, numbers of atlas and reference objects, and true-positive matches before calculating the corresponding metrics. To quantify the realized geometric contribution of the supplementary sources, the same evaluation was repeated for three nested subsets of the final atlas defined by retained geometry source: OBM only, OBM + GBA, and OBM + GBA + HABITAT (the final atlas).

3.4. Analysis methods

To characterize the pan-Arctic built environment using the unified multi-attribute building footprint atlas, we implemented a multi-scale analytical framework linking object-level building attributes with regional spatial patterns and environmental exposure. First, we quantified the overall building stock across all terrestrial regions north of 60°N, including total building count, cumulative/mean footprint area, building height, approximate building age in 2021, and functional composition. These statistics provided a circumpolar baseline for assessing the magnitude and structure of Arctic building development. Second, we analyzed the spatial organization of building footprints and attributes across multiple scales. At the pan-Arctic scale, building footprints were aggregated into standardized 10×10 km grid cells. This spatial scale provides a practical balance between retaining regional detail and limiting excessive spatial generalization in large-area analyses (Siedentop & Fina, 2010), while maintaining a consistent spatial unit across the two components of the study. The gridded summaries were used to identify broad regional clusters and attribute gradients across the circumpolar North. To quantify latitudinal structure, we further aggregated building attributes within 0.25° latitude belts from 60°N to 85°N and fitted locally estimated scatterplot smoothing (LOESS) trends to characterize south-to-north changes in building abundance, footprint area, height, and approximate age. Regional heterogeneity was assessed by summarizing grid-level

building attributes across major Arctic regions, including North America, Europe, Russia/Asia, and Iceland/Greenland.

We further assessed the environmental context and future geohazard exposure of Arctic buildings by spatially linking the unified footprint atlas with land-cover, permafrost, and projected thaw-related hazard datasets. Land-cover context was derived from 2021 MODIS IGBP classes, which were used to stratify the pan-Arctic built environment according to its broader landscape setting. To further characterize differences in building morphology among these landscape contexts, we calculated an enrichment index as the ratio between footprint-area share and building-count share. Permafrost context was evaluated using the permafrost occurrence dataset of Obu et al. (2019). Each building footprint was assigned the permafrost value extracted at its centroid and grouped into six probability classes (0–10%, 10–25%, 25–50%, 50–75%, 75–90%, and 90–100%). Footprints returning NoData or values outside the valid 0–100% range were excluded, and building-count and footprint-area shares were normalized over the remaining valid subset. Future geohazard exposure was quantified using the projected infrastructure hazard potential maps developed by Hjort et al. (2018), which classify thaw-related damage potential under different climate scenarios and time periods. The published low-, moderate-, and high-hazard-potential classes were derived from a consensus index combining three geohazard indices: a settlement index, a risk zonation index, and an analytic hierarchy process-based index. The final class was determined by majority agreement among the three indices; where all three indices assigned different levels, the moderate class was used (Hjort et al., 2018). We adopted these published consensus classes directly in the present analysis. For each scenario, we calculated the proportion of building footprint area assigned to the published low-, moderate-, and high-hazard-potential classes, using footprints with valid hazard-class assignments as the scenario-specific denominator. Within each land-cover context, we then calculated the share of validly assigned footprint area occurring in the high-hazard-potential class. This integrated framework links object-level infrastructure attributes with ecological setting, permafrost conditions, and future thaw-related hazard potential.

4. Results

4.1. Performance comparison of the input building footprint datasets

4.1.1. Site-based accuracy

Using the complete benchmark sample dataset, we evaluated the performances of OBM, GBA, and HABITAT at both semantic and instance levels (Table 1). Overall, OBM achieved the best performance, with a *Precision* of 73.9%, *Recall* of 71.6%, *F1-score* of 72.8%, and *IoU* of 57.2%. GBA performed comparably, with only slightly lower *F1-score* (71.5%) and *IoU* (55.6%). In contrast, HABITAT showed substantially lower detection recall, with *Recall* decreasing to 28.9% and *F1-score* to 40.7%, indicating pronounced omission errors across the complete benchmark sample. Instance-level metrics showed a similar ranking. OBM achieved the highest $P50$ (36.9%) and $R50$ (54.1%), followed closely by GBA ($P50 = 35.2\%$; $R50 = 50.3\%$). HABITAT had a much lower $R50$ of 12.7%, confirming its limited ability to detect individual buildings consistently. The instance number ratio further revealed contrasting object-count

tendencies. OBM and GBA contained more predicted than reference instances, with N_{ratio} values of 1.47 and 1.43, respectively, consistent with over-segmentation and/or commission errors. By contrast, HABITAT had substantially fewer predicted instances than the reference dataset ($N_{ratio} = 0.508$), consistent with under-segmentation and/or omission errors.

Site-level results revealed substantial heterogeneity in product performance across Arctic settlement and industrial contexts (Figure 3). OBM generally performed best in dense or well-mapped urban environments, achieving the highest *F1-scores* in Reykjavík (80.5%), Murmansk (75.1%), Tromsø (73.6%), and Nuuk (67.5%), and maintaining strong performance in Fairbanks (73.1%). This pattern indicates strong OBM performance in relatively complete urban mapping environments. GBA showed comparable or superior performance in several structured industrial and resource-dependent sites, including Prudhoe Bay (76.0%), Bovanenkovo (66.1%), Kiruna (66.1%), Norilsk (65.2%), and Sabetta (58.7%), indicating robust geometric extraction in industrial landscapes. HABITAT performed competitively in selected sites, such as Fairbanks (76.2%), Vorkuta (67.0%), Inuvik (66.1%), and Bovanenkovo (66.0%), but was unavailable or not evaluated in several major urban sites and performed poorly in Nuuk (*F1-score* = 38.1%; *IoU* = 23.5%), highlighting inconsistent spatial availability and generalization capacity.

The N_{ratio} patterns further clarified product-specific object-count differences. OBM and GBA frequently contained more predicted than reference instances in compact or morphologically complex sites such as Reykjavík, Tromsø, Kiruna, and Norilsk, consistent with fragmentation and/or commission errors. The oil and gas sites showed more variable behavior. In Prudhoe Bay, OBM and HABITAT contained fewer predicted than reference instances, whereas GBA was closer to parity. These site-level differences demonstrate that no single contemporary product consistently dominates across all Arctic built-environment types and highlight the value of integrating complementary information from multiple source products.

Product performance also varied systematically with building size (Figure 4). Both OBM and GBA showed a clear increase in accuracy from small to large building footprints. Small buildings were the most challenging class, with low *F1-scores* of 33.4% and 31.8% for OBM and GBA, respectively, and *IoU* values below 20.1%. Their N_{ratio} values were also markedly greater than one (2.80 and 2.57), indicating substantially more predicted than reference instances and suggesting pronounced fragmentation and/or commission errors for small structures. Performance improved for medium buildings, where both products reached more balanced N_{ratio} values close to 1.0 and higher *F1-scores*. The highest accuracies were obtained for large buildings, with *F1-scores* above 72% and *IoU* values around 57% for both OBM and GBA. HABITAT lacked small-building results and remained weaker for medium and large buildings, particularly in recall-related metrics, although its N_{ratio} values suggested less fragmentation than the small-building cases of OBM and GBA.

Together, these multi-level evaluations indicate that OBM provides the most robust general-purpose geometric baseline for pan-Arctic atlas construction, combining the highest overall accuracy, good site-level adaptability, and consistent size-dependent performance. GBA performs comparably and offers complementary strength in industrial and resource-extraction landscapes. HABITAT contributes useful regional information in selected sites but is constrained by lower detection recall, limited availability, and less consistent performance across site and size categories. These site-level results informed

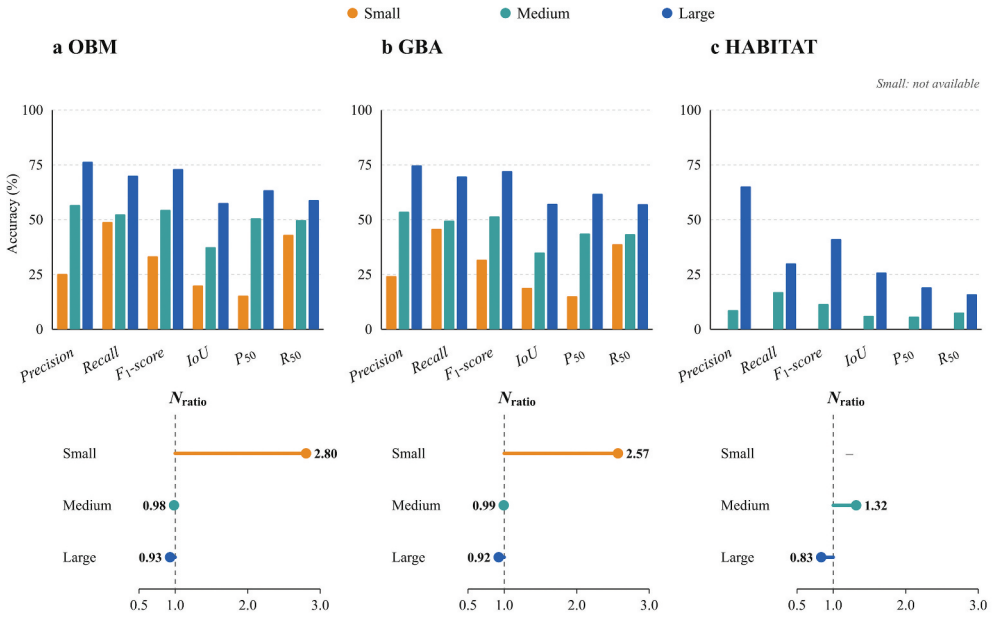


Figure 4. Size-dependent accuracy of three contemporary building-footprint products. a–c) Grouped bar charts compare four semantic-level metrics (*Precision, Recall, F1-score*, and *IoU*) and two instance-level metrics (*P50* and *R50*) across small, medium, and large building footprints for OBM, GBA, and HABITAT, respectively. All six accuracy metrics are expressed as percentages and share the same 0–100% scale. The lower plots show N_{ratio} for each available product–size combination. The dashed vertical line at $N_{ratio} = 1$ indicates equal numbers of predicted and reference instances. Building-size classes were defined using footprint-area thresholds of 1,024 and 9,216 image pixels. Small-building results were unavailable for HABITAT.

the priority order used to resolve overlapping geometries, with OBM assigned the highest priority, GBA used as a complementary secondary source, and HABITAT retained as an auxiliary Arctic-specific input.

4.1.2. Pan-Arctic spatial coverage

The WSF2019-normalized spatial coverage assessment revealed substantial spatial heterogeneity across the pan-Arctic domain, while also showing broad agreement between the two global building-footprint products (Figure 5(a–c)). OBM and GBA exhibited similar circumpolar patterns, with comparatively extensive mapped coverage across major built-up regions in northern Europe, Iceland, southern Alaska, and parts of western Russia, and more fragmented coverage across sparsely developed high-Arctic interiors. Their inset distributions were also closely aligned. Among the 13,795 WSF2019-positive grid cells evaluated, zero values accounted for 4.1% for OBM and 3.0% for GBA, while 13.0% and 13.2%, respectively, exceeded the displayed upper limit of 5. Within the displayed positive range, the WSF2019-area-weighted mean index values were 0.68 for OBM and 0.70 for GBA. These conditional means and the corresponding distributions indicate similar broad-scale area coverage, although values greater than 1 cannot be interpreted directly as commission because they may also reflect differences in data definitions, acquisition periods, geometric representation, or infrastructure not captured by WSF2019. HABITAT

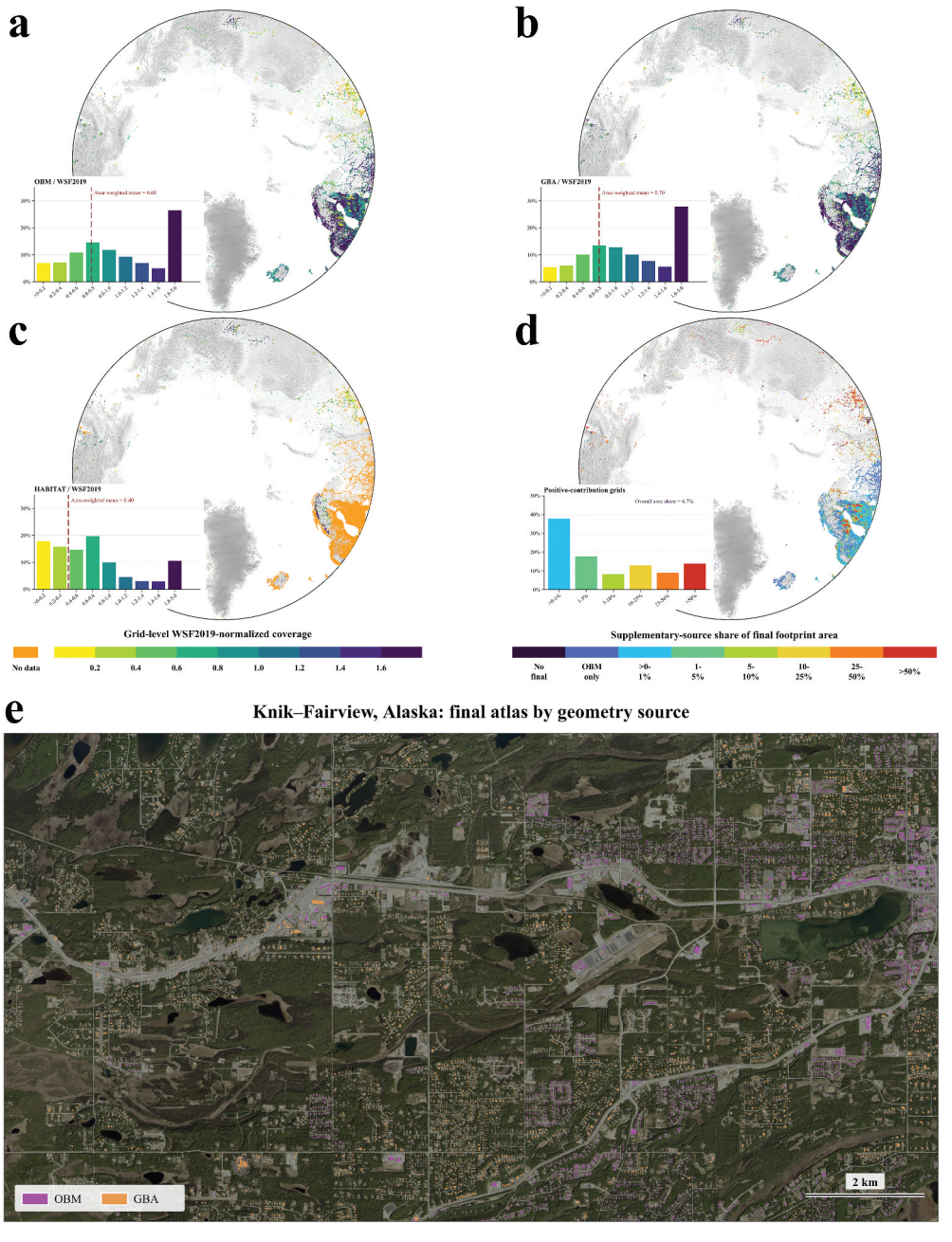


Figure 5. Pan-Arctic spatial coverage and complementary source contributions. a–c) Grid-level coverage of OBM, GBA, and HABITAT relative to WSF2019, calculated as building-footprint area divided by WSF2019 settlement area within each 10 × 10 km grid cell. Insets show the percentage distributions of the corresponding WSF2019-normalized spatial coverage indices within the interval (0, 5]. Zero values, values greater than 5, and NoData are excluded from the inset distributions and corresponding means. Dashed lines indicate WSF2019-area-weighted means calculated for the same displayed subsets. d) Share of final footprint area contributed by supplementary sources, calculated as retained GBA and habitat area divided by total final-atlas area in each grid cell. The inset shows the distribution among

displayed a markedly different full-domain pattern, consistent with its more restricted regional coverage (Figure 5(c)). Approximately 79.0% of WSF2019-positive grid cells contained no corresponding HABITAT footprint area, while 19.3% fell within the displayed positive range and 1.7% exceeded 5. The WSF2019-area-weighted mean for the displayed positive subset was 0.40. This value characterizes areas in which HABITAT contributes mapped footprints and should not be interpreted as a pan-Arctic coverage mean. The spatial results nevertheless show localized HABITAT contributions in parts of the Arctic, supporting its role as a regional supplementary source rather than a full-domain baseline.

The similarity between the pan-Arctic OBM and GBA distributions does not imply that their local footprint coverage is interchangeable. After hierarchical geometric conflation, geometries retained from GBA and HABITAT contributed to 65.3% of the grid cells included in the assessment (Figure 5(d)). These supplementary geometries represented 6.7% of total final-atlas footprint area. Among grids with positive supplementary contributions, 55.7% had shares no greater than 5%. However, the contribution was locally substantial, with 35.9% assigned to the 10–25%, 25–50%, or >50% classes and 13.9% receiving more than half of their final footprint area from supplementary sources. The combination of a modest overall area share with widespread and locally high contributions indicates that lower-priority sources primarily fill spatially uneven gaps rather than uniformly increasing mapped area across the domain. The Knik–Fairview example further illustrates this local complementarity at the object level (Figure 5(e)). Within the displayed extent, retained OBM and GBA geometries form spatially distinct clusters, and both contribute substantially to the combined atlas. OBM accounts for 50.9% and GBA for 49.1% of retained footprint area in this selected window. Thus, similar pan-Arctic coverage statistics can coexist with pronounced local differences in which buildings are represented by each source. This example describes retained geometry provenance after conflation rather than providing an independent estimate of raw-product completeness.

Because WSF2019 represents settlement extent rather than individual buildings and differs temporally from the input products, we additionally examined the sensitivity of these spatial-coverage patterns using the SACHI-v2 buildings-and-other-constructions class (class 20) within its available geographic extent (Figure S2a–c). Across the 777 fully covered paired grids, the area-weighted indices followed the same regional ordering under both references: GBA had the highest coverage, followed by HABITAT and OBM, with values of 0.520, 0.467, and 0.456 relative to WSF2019 and 0.699, 0.628, and 0.613 relative to SACHI-v2, respectively (Figure S2d). This ordering applies only to the SACHI-v2 geographic subset and should not be interpreted as the corresponding pan-Arctic ranking. The systematically higher SACHI-v2-normalized values reflect its smaller aggregate reference area, which was 74.4% of the corresponding WSF2019 settlement area (Figure S2d), although appreciable grid-level differences remained, as further illustrated by the local comparison in Pangnirtung (Figure S2e). This regionally constrained comparison complements the full-domain WSF2019 assessment.

grids with positive supplementary contributions, while the annotation reports the overall supplementary-source share of final-atlas area. “OBM only” and “no final” indicate grids without supplementary contributions and without final footprints, respectively. e) Object-level example from Knik–Fairview, Alaska, showing the combined atlas classified by geometry source.

4.2. The unified pan-Arctic building footprint atlas and its spatial-attribute patterns

The final fused atlas showed generally robust, although regionally variable, agreement with the independent reference datasets. Across all five regions, the micro-aggregated area-based *Precision*, *Recall*, *F1-score*, and *IoU* were 80.8%, 74.1%, 77.3%, and 63.0%, respectively, while the instance-level *P*50 and *R*50 were 71.4% and 53.6% (Table 2). The overall N_{ratio} was 0.750, indicating that the atlas contained fewer building objects than the reference datasets. Regional area-based *F1-scores* ranged from 59.9% in Yukon to 80.4% in the Northwest Territories, while *P*50 ranged from 32.6% in Yukon to 80.7% in Anchorage and N_{ratio} ranged from 0.616 to 0.923. These regional differences indicate that agreement depends not only on residual atlas omissions and geometric errors, but also on heterogeneity among the reference datasets in mapping source, building delineation, and the treatment of adjoining or subdivided structures. Instance-level precision and recall declined progressively as the *IoU* threshold increased (Figure S1). This pattern indicates that most matched buildings had broadly consistent locations and extents, whereas exact polygon correspondence was more sensitive to boundary delineation and split–merge differences between datasets. Within the same evaluation domains, the retained-source ablation showed that adding final-atlas geometries retained from GBA and HABITAT increased N_{ratio} from 0.721 to 0.750 and yielded 1,265 additional one-to-one matches at IoU ≥ 0.50 (Figure S3). These additions were associated with a completeness–precision trade-off: *Recall* and *R50* increased by 0.66 and 0.77 percentage points, whereas *Precision* and *P50* decreased by 1.33 and 1.87 percentage points; *F1-score* and *IoU* remained nearly unchanged (−0.24 and −0.32 percentage points, respectively)

Following these assessments, we characterized the spatial organization and multi-attribute structure of the resulting atlas across the terrestrial pan-Arctic region. The atlas identifies approximately 15.08 million individual building footprints, covering a total footprint area of approximately 2,004 km^2, and provides object-level geometry together with available morphological, temporal, and functional attributes. Across the atlas, valid height, GISA-derived temporal-proxy, and non-unknown functional-type attributes were available for 90.4%, 34.8%, and 59.1% of buildings, respectively, while all three enriched attributes were jointly available for 25.7% (Figure S4). Figure 6 presents a circumpolar overview of the fused atlas and summarizes its large-scale spatial organization across 10 × 10 km grids, latitude belts, and major Arctic regions. The pan-Arctic maps reveal a highly uneven geography of buildings across the circumpolar North (Figure 6(a–e)). Building count and footprint area show broadly similar spatial patterns, with dense concentrations in northern Europe, western Russia, southern Alaska, and Iceland, whereas most high-Arctic interiors

Table 2. Independent validation of the final fused atlas against regional reference building-footprint datasets.

Region	*Precision* (%)	*Recall* (%)	*F1-score* (%)	*IoU* (%)	*P*50 (%)	*R*50 (%)	N_{ratio}
Anchorage	84.6	75.3	79.7	66.2	80.7	59.9	0.742
Northwest Territories	81.4	79.4	80.4	67.2	79.7	49.1	0.616
Yukon	52.9	69.2	59.9	42.8	32.6	30.1	0.923
Sodankylä	70.4	60.0	64.7	47.9	47.0	36.7	0.781
Sisimiut	66.9	62.0	64.4	47.5	52.3	32.3	0.617
Overall	80.8	74.1	77.3	63.0	71.4	53.6	0.750

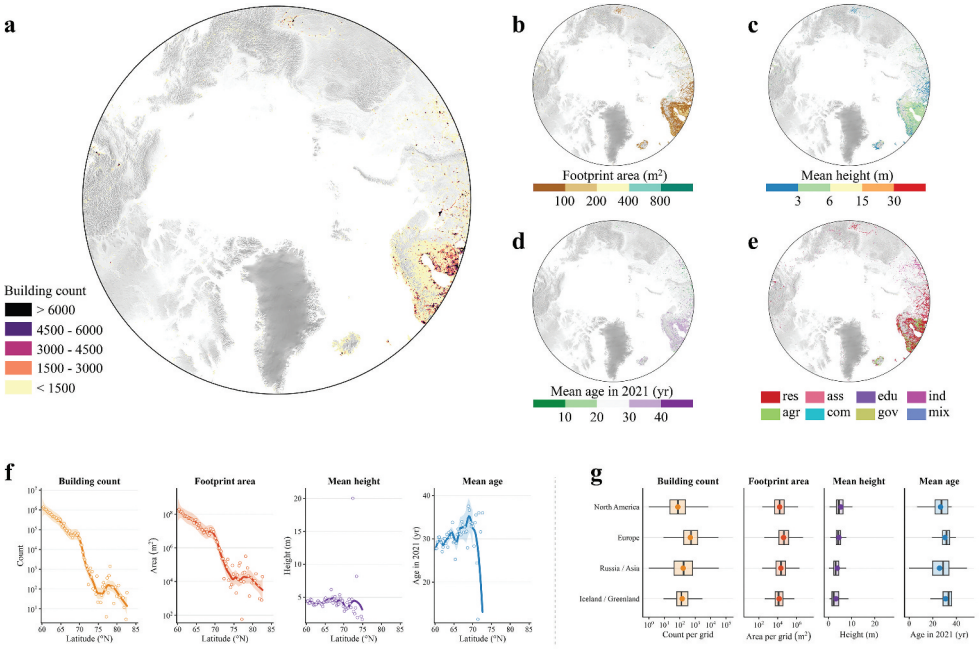


Figure 6. The unified pan-Arctic building footprint atlas and its circumpolar-scale attribute patterns. Maps show the spatial distribution of the resulting atlas north of 60°N and its associated object-level attributes aggregated within 10 × 10 km grid cells: building count (a), mean footprint area (b), mean height (c), mean approximate building age in 2021 (d), and dominant functional type (e). Functional abbreviations denote residential (res), agricultural (agr), assembly (ass), commercial (com), educational (edu), governmental (gov), industrial (ind), and mixed-use (mix) buildings. f) Latitudinal profiles of building footprint attributes aggregated within 0.25° latitude belts from 60°N to 85°N. Circles indicate observed values, solid lines show locally estimated scatterplot smoothing (loess) trends, and shaded bands indicate approximate 95% confidence intervals. For building height and approximate building age, confidence bands are shown only up to 70°N because valid samples become sparse farther north. g) Regional distributions of building footprint attributes across North America, Europe, Russia/Asia, and Iceland/Greenland. Boxplots show grid-level distributions, with boxes indicating interquartile ranges, center lines indicating medians, whiskers extending to 1.5× the interquartile range, and colored dots indicating regional means. Height, approximate-age, and functional-type summaries are calculated only from buildings with valid values for the corresponding attribute.

remain sparsely built. This correspondence indicates that regions with high building density also account for a large share of total built-up footprint area. However, local differences between count and area reveal important variation in building morphology: compact settlements tend to contain many small to medium-sized buildings, whereas industrial and resource-extraction landscapes can contribute disproportionately large footprint areas through fewer but larger facilities. Across most of the pan-Arctic, buildings are generally low-rise, reflecting the dominance of small settlements, dispersed communities, and low-density industrial infrastructure. Taller structures are more localized around larger urban centers and selected industrial nodes, indicating that vertical development remains spatially limited in the Arctic built environment. Where available, approximate building age and functional type further characterize the historical and operational structure of the pan-Arctic building stock. Among records with available temporal attributes, the approximate

building age map indicates that much of the mapped building stock is relatively mature, with older clusters concentrated in long-established settlement regions of northern Europe, western Russia, Iceland, and Greenland. Younger or more mixed-age patterns are more localized and are often associated with recent infrastructure development, industrial expansion, or resource-extraction facilities. Among buildings with non-unknown functional labels, their patterns show a strong contrast between settlement-dominated regions and specialized industrial landscapes. Residential and mixed-use categories are concentrated around towns and urban corridors, whereas industrial and related functional types are more evident in oilfield complexes, mining regions, ports, and logistical facilities.

Latitudinal profiles further quantify the strong south-to-north gradient in Arctic infrastructure (Figure 6(f)). Building count and cumulative footprint area decline rapidly with increasing latitude, with the overwhelming majority of buildings located between 60°N and 65°N and only a small fraction occurring above 70°N. This pattern reflects the concentration of settlements and transport-accessible infrastructure along the southern Arctic margin. In contrast, high-latitude buildings are sparse and more often associated with specialized facilities, remote communities, and logistical or extraction-related infrastructure. Among records with valid attributes, building height and approximate building age profiles show weaker but still informative gradients, suggesting that structural morphology and development timing vary less continuously with latitude than building abundance, but remain shaped by the geographic distribution of major settlements and industrial nodes. Regional summaries emphasize the continental heterogeneity of the pan-Arctic built environment (Figure 6(g)). Europe contains the largest share of mapped buildings and footprint area, reflecting its relatively dense settlement network across northern Nordic and sub-Arctic regions. Russia/Asia forms the second major concentration, but with a more uneven spatial pattern shaped by industrial cities, mining centers, and resource-dependent settlements embedded within vast sparsely populated territories. North America contributes fewer buildings overall but includes localized clusters with relatively large or taller structures, particularly around Alaskan and Canadian Arctic settlements and industrial facilities. Iceland/Greenland contains smaller and more dispersed building clusters, consistent with its coastal settlement structure and limited inland development.

Figure 7 further illustrates the local-scale representational capacity of our product across three contrasting Arctic built environments. In the compact urban core of Anchorage, the dataset resolves dense and heterogeneous building arrangements, including small residential structures, larger commercial or public buildings, and local variation in building height and approximate building age. This demonstrates the ability of the product to preserve fine-scale urban morphology in relatively complex settlement environments. In Prudhoe Bay, the product captures a markedly different spatial configuration: sparse but highly organized industrial infrastructure, characterized by isolated operational buildings, larger footprint areas, low-rise structures, and a dominant industrial functional context. This pattern highlights the value of the fused product for representing resource-extraction landscapes where infrastructure is dispersed but functionally specialized. In Vorkuta, the dataset delineates a mining and industrial city with clustered residential blocks, industrial facilities, and corridor-like development patterns, reflecting

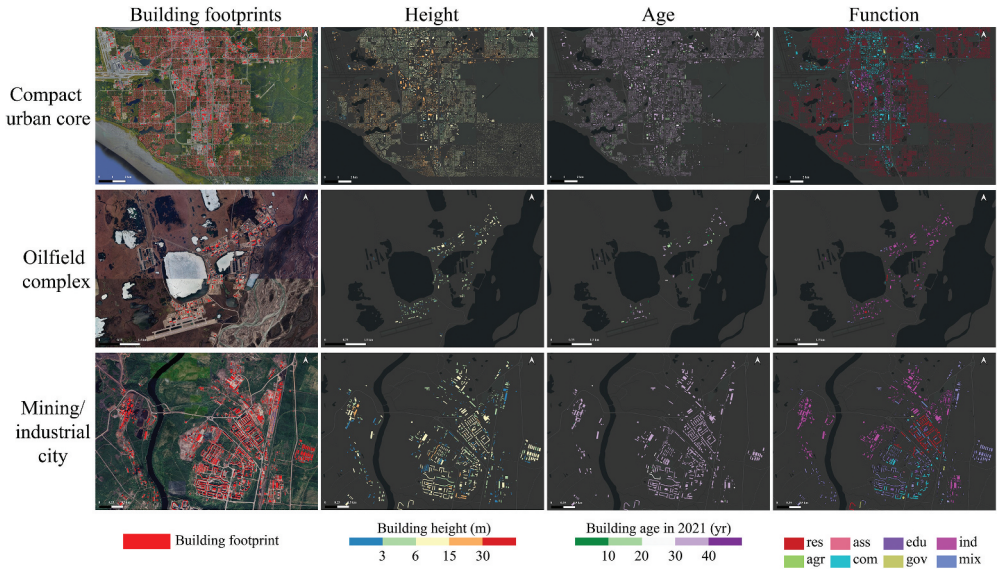


Figure 7. Representative zoom-in examples of the multi-attribute pan-Arctic building atlas in Arctic built environments. The three cases illustrate a compact urban core (Anchorage), an oilfield complex (Prudhoe Bay), and a mining/industrial city (Vorkuta). For each case, the fused product is shown with building footprint geometry and available attributes, including building height, approximate building age in 2021, and functional type.

the spatial imprint of extraction-oriented urbanization. Across these cases, the mapped footprint geometry, height, approximate age, and functional-type attributes reveal substantial local heterogeneity that cannot be represented by conventional built-up area or settlement-extent products. These zoom-in examples therefore demonstrate that the unified dataset bridges circumpolar-scale mapping and object-level characterization, providing a detailed spatial basis for subsequent analyses of infrastructure and geohazard exposure.

4.3. Environmental context and future geohazard exposure of Arctic buildings

The unified building footprint atlas enables a spatially explicit assessment of how Arctic buildings are embedded within ecological and cryospheric environments, and how these environmental settings translate into future geohazard exposure. Arctic buildings are concentrated primarily in open, low-stature landscapes rather than being confined to conventional urban/built-up land-cover classes (Figure 8(a)). Within this contextual stratification, grassland/open-vegetation landscapes represent the dominant environmental setting of the mapped Arctic built environment, accounting for 72.0% of mapped buildings and 70.3% of total footprint area. The corresponding shares for forested landscapes are 16.3% and 9.9%, respectively, suggesting smaller or more dispersed structures. By contrast, urban/built-up landscape contexts account for 5.8% and 13.8%, respectively, yielding a strong enrichment index of 2.38. Barren/sparse-vegetation landscapes show the highest enrichment index (3.25), despite their limited overall representation, suggesting a disproportionate contribution from large-footprint industrial, logistical, or

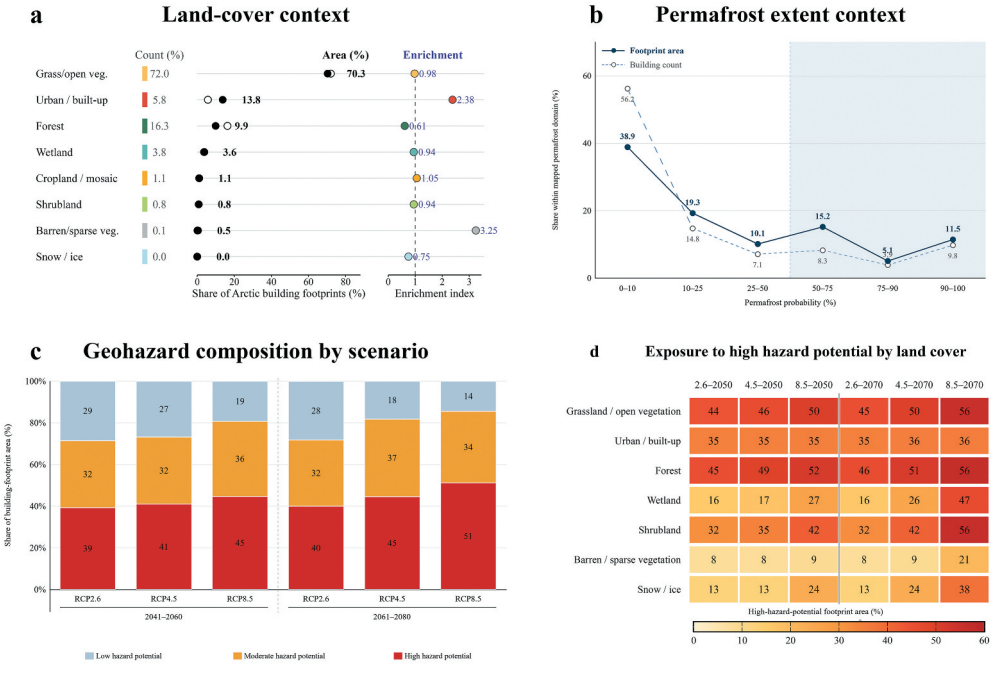


Figure 8. Environmental context and projected thaw-related geohazard exposure of pan-Arctic building footprints. a) Distribution of building count and footprint area across major land-cover contexts derived from the 2021 MODIS MCD12Q1 land cover type 1 product. The enrichment index is the ratio of footprint-area share to building-count share; the dashed line marks parity (enrichment = 1). b) Distribution of building count and footprint area across permafrost extent/probability classes. Percentages are calculated among footprints assigned valid values within the mapped permafrost domain; the shaded region indicates permafrost probabilities of ≥50%. c) Scenario-dependent composition of building-footprint area across low-, moderate-, and high-hazard-potential classes under representative concentration Pathway (RCP) 2.6, 4.5, and 8.5 for 2041–2060 and 2061–2080. Percentages are calculated among footprints with valid hazard-class assignments in each scenario. d) Share of building-footprint area assigned to the high-hazard-potential class within each land-cover context and climate scenario.

extraction-related structures. The permafrost analysis, which is restricted to footprints assigned valid values within the mapped permafrost domain, reveals a distribution across the full probability gradient (Figure 8(b)). The 0–10% probability class contains the largest shares, accounting for 38.9% of footprint area and 56.2% of building count within this valid subset. By comparison, the three classes with permafrost probabilities of ≥50% collectively contain 31.7% of footprint area and 21.9% of building count. Footprint-area shares exceed the corresponding building-count shares in each of these higher-probability classes, indicating a relatively greater contribution from larger building footprints.

Projected geohazard exposure increases under stronger warming and later-century scenarios (Figure 8(c)). Under RCP2.6 for 2041–2060, 39% of building-footprint area with valid hazard assignments is located in high-hazard-potential zones, whereas 29% occurs in low-hazard-potential zones. Under RCP8.5 for 2061–2080, the high-hazard-potential share increases to 51%, while the low-hazard-potential share declines to

14%. This change indicates a marked shift in the composition of projected exposure toward higher hazard potential under stronger warming. Even under the lower-emission mid-century projection, the high- and moderate-hazard-potential classes together account for most of the validly assigned footprint area. Exposure to high hazard potential also varies substantially among land-cover contexts (Figure 8(d)). Forest and grassland/open-vegetation contexts reach high-hazard-potential shares of 56% under RCP8.5 for 2061–2080, while shrubland reaches the same value. Urban/built-up contexts maintain a comparatively stable but substantial share of approximately 35–36% across the six scenarios. Wetland and snow/ice contexts show particularly pronounced increases, reaching 47% and 38%, respectively, under the late-century RCP8.5 scenario. Barren/sparse-vegetation contexts remain less exposed overall, although their high-hazard-potential share increases to 21% under the same scenario. These patterns indicate that projected thaw-related geohazard exposure is jointly structured by climate trajectory and the broader landscape setting of the built environment.

5. Discussion

5.1. Advancing a multi-attribute pan-Arctic building atlas

Existing built-environment products provide complementary information at different spatial scales and for different analytical purposes (Bartsch et al., 2024; Gong et al., 2020; Liu et al., 2023). Settlement-scale products such as WSF2019 and SACHI-v2 offer spatially consistent representations of settlement or infrastructure extent and are well suited to broad regional analyses, but they do not resolve individual buildings or their attributes. By contrast, contemporary building-footprint products such as OBM, GBA, and HABITAT provide object-level geometries and selected attributes, yet differ in spatial coverage, provenance, and attribute availability. The added value of the present atlas therefore lies not in any single data layer, but in bringing these complementary sources into a common building-level framework across terrestrial regions north of 60°N. Individual footprints are linked with available morphological, temporal, functional, and provenance information, allowing building geometry and associated attributes to be interpreted within the same spatial unit. This common structure also provides a consistent basis for comparing Arctic settlements and infrastructure across regions that differ markedly in building density, morphology, and development history.

At the analytical level, the atlas bridges broad built-up mapping and building-resolved characterization. It enables individual buildings to be counted, measured, grouped by available attributes, and directly linked with environmental and geohazard information. These capabilities support cross-regional assessment of the Arctic building stock, characterization of structural and temporal patterns, and spatially explicit analysis of infrastructure exposure. Importantly, the value of multi-source integration does not require each source product to perform uniformly well across all Arctic environments. It may also arise when a source contributes otherwise missing information in specific geographic settings. The retained-source ablation supports this interpretation by showing that supplementary geometries recovered additional reference-matched buildings while leaving the overall *F1-score* and *IoU* nearly unchanged (Figure S3). HABITAT illustrates this spatially selective contribution.

Although its aggregate performance is constrained by lower recall and more limited spatial availability, it can make substantial local contributions where the global products provide a less complete representation. In Noatak, Alaska, for example, 152 of the 161 buildings retained within the corresponding 10×10 km grid of the final atlas were retained from the HABITAT input product, demonstrating its locally important role as a supplementary source (Figure S5). More recently, the updated HABITAT–OSM product described by Manos et al. (2026) similarly expands building-resolved information by providing predicted occupancy, height, and story counts within the circumpolar permafrost region. In this evolving data landscape, the present atlas provides a harmonized, source-aware framework spanning terrestrial regions north of 60°N. It links building geometries and enriched attributes with environmental and geohazard information while retaining source provenance.

5.2. Spatial organization and heterogeneity of Arctic buildings

The spatial organization revealed by the atlas shows that the Arctic built environment is not a uniform northern extension of global urbanization, but a highly selective infrastructure geography shaped by environmental constraints, accessibility, historical settlement, and resource development (Figure 6). The strong concentration of buildings near the southern Arctic margin and in northern Europe reflects relatively favorable climatic conditions, denser transport networks, longer settlement histories, and closer integration with national urban systems. By contrast, the sparse building distribution across much of the high Arctic illustrates how extreme cold, remoteness, terrain constraints, sea-ice seasonality, and continuous permafrost have limited permanent settlement and large-scale construction. The more uneven spatial patterns in Russia/Asia and North America further indicate the imprint of resource-oriented development, where mining cities, oil and gas fields, ports, and logistical corridors create localized infrastructure clusters within otherwise sparsely populated landscapes. These contrasts are also reflected in the mapped attributes. The predominance of low-rise structures is consistent with the dispersed settlement forms and engineering constraints typical of Arctic environments (Liu et al., 2022), whereas localized taller or larger-footprint buildings mark urban centers, industrial facilities, and strategic infrastructure nodes. Similarly, older building clusters in long-established settlement regions contrast with younger or mixed-age infrastructure associated with recent industrial expansion and logistical development. Thus, the pan-Arctic building geography revealed here reflects the interaction between natural environmental filtering and human development priorities, linking settlement history, resource economies, and national infrastructure strategies within a single object-level spatial framework. These circumpolar patterns are summarized partly using the standardized 10×10 km grid, and their local expression may vary with alternative cell sizes or grid alignments. The gridded results are therefore most useful for characterizing broad regional structure, while the released object-level footprints allow the same patterns to be examined at finer or application-specific spatial scales.

5.3. Building-resolved exposure and implications for Arctic adaptation

Previous circumpolar studies have characterized infrastructure exposure and estimated potential damage from permafrost degradation using hazard indices, infrastructure

inventories, geotechnical models, and economic valuation (Hjort et al., 2018; Manos et al., 2026; Streletskiy et al., 2023; Suter et al., 2019). In this context, the present study uses the published geohazard layers of Hjort et al. (2018) to examine exposure through a harmonized building inventory rather than developing a new hazard or damage model. The inventory resolves individual buildings across terrestrial regions north of 60° N while retaining geometry, available attributes, and source provenance. The retained-source ablation shows that data integration recovers additional reference-matched buildings relative to OBM alone, with only minor changes in overall *F1-score* and *IoU* (Figure S3). Our building-level analysis shows that footprints within the mapped permafrost domain span the full probability gradient, with 31.7% of the validly assigned footprint area occurring in locations with modeled permafrost probabilities of ≥50%. It also shows that the share of footprint area exposed to high hazard potential increases markedly under stronger warming and later-century scenarios (Figure 8). These results characterize the distribution of building-level exposure, whereas realized risk also depends on asset vulnerability and potential consequences (Streletskiy et al., 2023). The landscape-context patterns further suggest that future geohazard exposure is not evenly distributed across the Arctic built environment. Distinct exposure trajectories are observed among areas characterized by open vegetation, forests, shrublands, wetlands, and urban land cover, reflecting broad differences in ground conditions, settlement history, and infrastructure function. These associations should be interpreted at the 500-m MODIS landscape scale. A single MCD12Q1 pixel may encompass heterogeneous local land-cover types and multiple built structures, particularly along settlement margins and in mixed urban–vegetation landscapes, so the immediate environmental setting of individual buildings may not be fully represented. The results are therefore intended as a broad circumpolar landscape-context characterization. A temporal mismatch also remains because the 2021 MCD12Q1 layer does not correspond exactly to the acquisition or update periods of all source building-footprint products. This mismatch may be more relevant in areas affected by recent construction, settlement expansion, or land-cover change and should therefore be considered when interpreting local landscape-context relationships. Future building-scale environmental analyses would benefit from the development of land-cover products that combine substantially finer spatial resolution with a classification scheme tailored to the heterogeneous ecological and built-environment conditions of the Arctic. These products would enable more reliable characterization of the immediate environmental setting surrounding individual buildings and provide a stronger basis for evaluating and reducing the scale-related uncertainty identified here. Such information is directly relevant for prioritizing adaptation actions, because the same projected thaw hazard may imply different management needs for residential communities, oilfield infrastructure, mining settlements, ports, and logistical facilities. By linking individual buildings with permafrost and geohazard layers, the atlas provides a spatially explicit basis for screening locations that may warrant more detailed monitoring, engineering assessment, or adaptation planning.

Our findings underscore that an object-level Arctic building inventory is not merely a scientific geospatial data product, but a practical foundation for infrastructure governance, community adaptation, and risk-informed decision-making across the circumpolar North. Many Arctic settlements and industrial systems are small and logistically constrained, yet they support essential housing, energy production, transportation, public

services, and regional economic activity (Carioli et al., 2026; Kumpula et al., 2012). In these settings, even localized ground instability can produce disproportionate consequences because maintenance costs are high, alternative infrastructure is limited, and emergency response capacity is often constrained (Manos et al., 2025). By identifying where individual buildings are located, which functional contexts they occupy, and how they intersect with permafrost and projected geohazard zones, the atlas provides a scalable basis for screening locations that may warrant monitoring, engineering inspection, maintenance investment, or adaptation planning (Li et al., 2026). This is particularly relevant for local governments and Arctic stakeholders that must allocate limited resources across dispersed settlements, industrial facilities, and transport-linked infrastructure. More broadly, the atlas can support assessments of resilient infrastructure, sustainable settlements, and climate adaptation in high-latitude regions by providing a consistent spatial denominator for comparing exposure across communities, regions, and functional building types. Rather than treating Arctic infrastructure as a homogeneous exposed asset class, the atlas enables exposure to be compared in relation to building morphology, function, settlement structure, and environmental setting. This building-resolved exposure layer can be combined with independent vulnerability and consequence information in future risk assessments.

5.4. Uncertainties and future directions

Several limitations and uncertainties remain and should be considered when interpreting and applying the current atlas. First, although the hierarchical geometric conflation integrates complementary contemporary footprint datasets, the final atlas inevitably inherits residual errors from its source products, including omission of small or low-visibility buildings, commission errors over complex natural surfaces, and segmentation inconsistencies in dense or morphologically complex settlements. The fixed source-priority rule used to resolve overlapping geometries represents an additional limitation because product performance varies among built-environment types, building sizes, and local conditions, and a lower-priority product may occasionally provide a more accurate geometry in a specific local context. The adopted hierarchy should therefore be interpreted as a transparent and reproducible duplicate-resolution strategy rather than as evidence of universal superiority of one source product. Future versions could incorporate locally adaptive or object-level quality-aware conflation as sufficiently representative reference data become available across the pan-Arctic domain. These uncertainties are particularly difficult to avoid in the Arctic because building extraction is affected by persistent cloud cover, snow and ice backgrounds, low solar illumination, and limited reference data. The WSF2019-based spatial coverage assessment introduces an additional source of uncertainty. WSF2019 represents settlement extent rather than individual building geometry, and differences in feature definition and spatial aggregation can therefore affect the resulting coverage ratios. WSF2019 may also contain regional omission and commission errors, while its 2019 reference year introduces a temporal mismatch with some of the more recent building footprint products. Consequently, the WSF2019-normalized ratios should be interpreted as relative settlement-scale coverage indicators rather than measures of absolute building completeness or object-level accuracy. The supplementary SACHI-v2 comparison reduces reliance on a single settlement-scale

benchmark but does not eliminate reference-data uncertainty. SACHI-v2 has spatially selective regional coverage, its class-20 definition includes buildings and other constructions, and its 2016–2020 observation period does not correspond exactly to all input products. The 12 benchmark sites were purposively selected to span contrasting Arctic built-environment conditions rather than through probabilistic sampling and therefore should not be interpreted as a statistically representative sample of the full pan-Arctic domain. Their site-level evaluation is complemented by the full-domain spatial coverage assessment and, in the evaluation framework, by independent regional validation of the final fused atlas.

Second, the enriched attributes differ in completeness and reliability because they were derived from heterogeneous data sources and inference strategies. Although valid height estimates were available for most buildings, temporal and functional information had substantially lower and spatially heterogeneous availability (Figure S4). The corresponding footprint-area availability was 95.9% for height, 53.9% for the GISA-derived temporal proxy, and 68.5% for functional type, while buildings containing all three attributes represented 43.0% of total footprint area. The consistently higher area-based than count-based percentages indicate greater attribute availability among larger-footprint buildings. Height and temporal-proxy availability generally decreased toward higher latitudes, whereas functional-type availability was more spatially variable because it depended on non-unknown OBM metadata. These patterns delimit the subsets suitable for attribute-based analysis and should be interpreted as non-missing attribute coverage rather than as evidence of attribute accuracy. Building height was compiled primarily from GBA and supplemented with GBHM where valid GBA height attributes were unavailable. A supplementary cross-product comparison showed limited building-level correspondence, with differences varying across height classes and Arctic regions (Figure S6). The two height products should therefore not be treated as interchangeable. Nevertheless, the comparison supports the adopted source hierarchy: object-associated GBA estimates were retained as the primary height source, whereas the spatially continuous 10-m GBHM was used only for gap filling. GBA and GBHM account for 97.3% and 2.7%, respectively, of valid final height attributes, and their provenance is recorded for each building. The resulting height attribute is therefore most appropriate for characterizing broad-scale vertical building morphology, while individual-building estimates should be interpreted with caution. Further improvements could draw on independent elevation information and Arctic-specific height-retrieval approaches. For example, Manos et al. (2026) estimated building heights from 2-m ArcticDEM digital surface models using footprint-level roof and surrounding-ground elevations. Their comparison with airborne light detection and ranging (LiDAR) demonstrated the potential of this approach but also revealed systematic height underestimation, highlighting the importance of local co-registration and calibration. Airborne LiDAR, where available, could additionally provide high-quality regional reference heights for calibrating or training future Arctic-specific height models. In addition, approximate building age was derived from long-term impervious-surface records rather than direct construction records, and functional type was mainly inherited from occupancy and operational metadata available in OBM. These attributes therefore provide a useful first-order characterization, but they should not be interpreted as equivalent to cadastral or engineering-grade

building records. Future versions would benefit from more attribute-specific reference datasets, particularly for evaluating building height and functional type.

Third, the geohazard analysis quantifies spatial exposure rather than realized structural damage. Overlaying footprints with permafrost occurrence and projected hazard layers identifies where buildings coincide with potentially unstable terrain, but actual vulnerability will also depend on foundation design, building materials, ground-ice content, maintenance history, drainage conditions, and local adaptation measures. Future work should therefore move from product fusion toward model-assisted updating, using existing footprint datasets as prior knowledge while incorporating VHR optical imagery, SAR, digital elevation models, local cadastral data, and field observations within deep learning or active learning frameworks. Such approaches could improve geometric accuracy, harmonize building attributes, and enable dynamic monitoring of Arctic infrastructure change. Despite these limitations, the atlas provides a coherent, object-level representation of pan-Arctic buildings that can support future product updates and risk-oriented infrastructure analyses.

6. Conclusions

This study developed a unified, multi-attribute pan-Arctic building footprint atlas, providing an object-level inventory of building geometry together with available morphological, temporal, and functional information across terrestrial regions north of 60°N. By integrating complementary global and Arctic-specific footprint products through hierarchical geometric conflation and multi-source attribute enrichment, the atlas identifies approximately 15.08 million buildings covering about 2,004 km^2, revealing a highly uneven Arctic built environment. The resulting building-resolved perspective shows that Arctic infrastructure is not only spatially fragmented and predominantly low-rise, but also distributed across a broad range of permafrost conditions and exposed to spatially heterogeneous future thaw-related hazard potential. By linking individual buildings with environmental and hazard layers, the atlas supports spatially explicit assessments of infrastructure exposure and adaptation priorities across the circumpolar North. The Pan-Arctic Building Footprint Atlas v1 has been publicly released through Zenodo (Zhao et al., 2026) as an open resource for research on Arctic infrastructure, permafrost-related exposure, and climate adaptation.

Author contributions

CRediT: **Shimeng Zhao:** Methodology, Validation, Writing – original draft; **Jiayi Wang:** Data curation, Formal analysis, Writing – original draft; **Xiao Cheng:** Funding acquisition, Supervision, Writing – review & editing; **Yize Guo:** Data curation, Formal analysis, Software; **Qi Zhang:** Conceptualization, Writing – review & editing; **Peijie Wang:** Data curation, Software, Validation; **Weijia Li:** Resources, Supervision, Writing – review & editing; **Da He:** Conceptualization, Supervision, Writing – review & editing; **Chong Liu:** Conceptualization, Supervision, Writing – review & editing.

Disclosure statement

No potential conflict of interest was reported by the authors.

During the preparation of this manuscript, GPT-5.5 and GPT-6.0 were used only to improve language readability, grammar, clarity, and overall expression. Generative AI was also used to assist in preparing the workflow illustration in Figure 2. After using these tools, all authors carefully reviewed, edited, and verified the manuscript content as required, and take full responsibility for the integrity and accuracy of the published work.

Funding

This work was supported by the National Key Research and Development Program of China under Grant [2025YFE0219100]; the National Natural Science Foundation of China under Grant [42371344]; and Discipline Breakthrough Precursor Project of the Ministry of Education of China under Grant [JYB2025XDXM803].

Notes on contributors



Shimeng Zhao is an M.S. student at the School of Geospatial Engineering and Science, Sun Yat-sen University. Her research focuses on individual building footprint extraction from high-resolution remote sensing imagery and the development of high-resolution infrastructure datasets for Arctic and other challenging environments.



Jiayi Wang is an M.S. student at the School of Geospatial Engineering and Science, Sun Yat-sen University. Her research focuses on mapping and interpretation of sparse human infrastructure in high-latitude environments.



Xiao Cheng is a Professor at Sun Yat-sen University, where he serves as Dean of the School of Geospatial Engineering and Science, Director of the Polar Research Center, and Director of the Key Laboratory of Comprehensive Observation of Polar Environment, Ministry of Education. His research focuses on polar remote sensing, atmosphere-ice-ocean-land interactions, ice shelf and glacier dynamics, sea-ice parameter retrieval, and polar climate change. He received the Award for Advanced Individual of China's Polar Exploration in 2017.



Yize Guo is an M.S. student at the School of Geospatial Engineering and Science, Sun Yat-sen University. His research focuses on the remote sensing characterization of Arctic surface properties and environmental conditions.



Qi Zhang is a Research Assistant Professor in the Department of Geography and Environment at the University of North Carolina at Chapel Hill. His research focuses on coupled human and natural systems, land-use change, forest conservation, and household livelihoods. His work contributes to understanding the spatial processes linking land systems, infrastructure, and socio-ecological change.



Peijie Wang is an M.S. student at the School of Geospatial Engineering and Science, Sun Yat-sen University. His research focuses on multimodal remote sensing and multi-source data fusion.



Weijia Li is an Associate Professor at the Shenzhen International Graduate School, Tsinghua University. She received her bachelor's degree from Sun Yat-sen University and her Ph.D. from Tsinghua University. Her research interests include computer vision, large multimodal models, and urban remote sensing, with a focus on extracting and interpreting built-environment information from multi-source geospatial data.



Da He is an Associate Professor at the School of Geography and Planning, Sun Yat-sen University. His research interests include multi- and hyperspectral remote sensing image classification, deep learning, data fusion, and change detection.



Chong Liu is an Associate Professor at the School of Geospatial Engineering and Science, Sun Yat-sen University. He received the B.S. degree in remote sensing science and technology and his Ph.D. degree in photogrammetry and remote sensing from Wuhan University in 2010 and 2015, respectively. His research focuses on satellite remote sensing, land-cover mapping and monitoring, Arctic and permafrost environments, and geospatial big data analytics.

ORCID

Xiao Cheng http://orcid.org/0000-0001-6910-6565

Data availability statement

The Pan-Arctic Building Footprint Atlas v1 generated in this study is publicly available from Zenodo: Zhao, S., Wang, J., Guo, Y., & Liu, C. (2026). Pan-Arctic Building Footprint Atlas v1 (Version v1) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.20796872. Additional intermediate data and processing scripts are available from the corresponding author upon reasonable request.

References

Bartsch, A., Efimova, A., Widhalm, B., Muri, X., von Baeckmann, C., Bergstedt, H., Ermokhina, K., Hugelius, G., Heim, B., & Leibman, M. (2024). Circumarctic land cover diversity considering wetness gradients. *Hydrology and Earth System Sciences*, *28*(11), 2421–2481. https://doi.org/10.5194/hess-28-2421-2024

Bartsch, A., Höfler, A., Kroisleitner, C., & Trofaier, A. M. (2016). Land cover mapping in northern high latitude permafrost regions with satellite data: Achievements and remaining challenges. *Remote Sensing*, *8*(12), 979. https://doi.org/10.3390/rs8120979

Bartsch, A., Pointner, G., Nitze, I., Efimova, A., Jakober, D., Ley, S., Högström, E., Grosse, G., & Schweitzer, P. (2021). Expanding infrastructure and growing anthropogenic impacts along Arctic coasts. *Environmental Research Letters*, *16*(11), 115013. https://doi.org/10.1088/1748-9326/ac3176

Bartsch, A., Widhalm, B., von Baeckmann, C., Efimova, A., Tanguy, R., & Pointner, G. (2023). *Sentinel-1/2 derived arctic coastal human impact dataset (SACHI)* (Version v2.0) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.10160636

Bergmann, M., Collard, F., Fabres, J., Gabrielsen, G. W., Provencher, J. F., Rochman, C. M., van Sebille, E., & Tekman, M. B. (2022). Plastic pollution in the Arctic. *Nature Reviews Earth and Environment*, *3*(5), 323–337. https://doi.org/10.1038/s43017-022-00279-8

Carioli, A., Uhl, J. H., Maffenini, L., Schiavina, M., Krasnodębska, K., Pesaresi, M., Florio, P., Melchiorri, M., & Kemper, T. (2026). Human settlement dynamics of the Arctic: Insights from remote sensing and census observations. *Applied Geography*, *189*, 103922. https://doi.org/10.1016/j.apgeog.2026.103922

Che, Y., Li, X., Liu, X., Wang, Y., Liao, W., Zheng, X., Zhang, X., Xu, X., Shi, Q., Zhu, J., Zhang, H., Yuan, H., & Dai, Y. (2024). 3D-GloBFP: The first global three-dimensional building footprint dataset. *Earth System Science Data*, *16*(11), 5357–5374. https://doi.org/10.5194/essd-16-5357-2024

Chen, X., Xiong, J., Yuan, Y., Liu, Y., Hou, X., Wu, Q., & Zhang, Z. (2024). *High-resolution global building height map (GBHM, 10 m) developed based on Sentinel imagery* [Data set]. Data set. https://doi.org/10.5281/zenodo.13285155

Chen, Y., Cheng, X., Liu, A., Chen, Q., & Wang, C. (2023). Tracking lake drainage events and drained lake basin vegetation dynamics across the Arctic. *Nature Communications*, *14*(1), 7359. https://doi.org/10.1038/s41467-023-43207-0

Cui, Y., Dong, J., Zhang, C., Yang, J., Chen, N., Guo, P., Di, Y., Chen, M., Li, A., & Liu, R. (2024). Validation and refinement of cropland map in southwestern China by harnessing ten contemporary datasets. *Scientific Data*, *11*(1), 671. https://doi.org/10.1038/s41597-024-03508-5

Dinerstein, E., Olson, D., Joshi, A., Vynne, C., Burgess, N. D., Wikramanayake, E., Hahn, N., Palminteri, S., Hedao, P., Noss, R., Hansen, M., Locke, H., Ellis, E. C., Jones, B., Barber, C. V., Hayes, R., Kormos, C., Martin, V., Crist, E., … Saleem, M. (2017). An ecoregion-based approach to protecting half the terrestrial realm. *BioScience*, *67*(6), 534–545. https://doi.org/10.1093/biosci/bix014

Esch, T., Brzoska, E., Dech, S., Leutner, B., Palacios-Lopez, D., Metz-Marconcini, A., Marconcini, M., Roth, A., & Zeidler, J. (2022). World settlement footprint 3D - A first three-dimensional survey of the global building stock. *Remote Sensing of Environment*, *270*, 112877. https://doi.org/10.1016/j.rse.2021.112877

Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., & Zisserman, A. (2010). The PASCAL visual object classes (VOC) challenge. *International Journal of Computer Vision*, *88*(2), 303–338. https://doi.org/10.1007/s11263-009-0275-4

Freitas, N. L., Walter Anthony, K., Lenz, J., Porras, R. C., & Torn, M. S. (2025). Substantial and overlooked greenhouse gas emissions from deep Arctic lake sediment. *Nature Geoscience*, *18* (1), 65–71. https://doi.org/10.1038/s41561-024-01614-y

Friedl, M., & Sulla-Menashe, D. (2022). *MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500m SIN Grid V061* [Data set]. NASA Land Processes Distributed Active Archive Center. https://doi.org/10.5067/MODIS/MCD12Q1.061

Gong, P., Li, X., Wang, J., Bai, Y., Chen, B., Hu, T., Liu, X., Xu, B., Yang, J., Zhang, W., & Zhou, Y. (2020). Annual maps of global artificial impervious area (GAIA) between 1985 and 2018. *Remote Sensing of Environment*, *236*, 111510. https://doi.org/10.1016/j.rse.2019.111510

Hjort, J., Karjalainen, O., Aalto, J., Westermann, S., Romanovsky, V. E., Nelson, F. E., Etzelmüller, B., & Luoto, M. (2018). Degrading permafrost puts Arctic infrastructure at risk by mid-century. *Nature Communications*, *9*(1), 5147. https://doi.org/10.1038/s41467-018-07557-4

Hjort, J., Streletskiy, D., Doré, G., Wu, Q., Bjella, K., & Luoto, M. (2022). Impacts of permafrost degradation on infrastructure. *Nature Reviews Earth and Environment*, *3*(1), 24–38. https://doi.org/10.1038/s43017-021-00247-8

Huang, X., Yang, J., Wang, W., & Liu, Z. (2022). Mapping 10 m global impervious surface area (GISA-10m) using multi-source geospatial data. *Earth System Science Data*, *14*(8), 3649–3672. https://doi.org/10.5194/essd-14-3649-2022

Kumpula, T., Forbes, B. C., Stammler, F., & Meschtyb, N. (2012). Dynamics of a coupled system: Multi-resolution remote sensing in assessing social-ecological responses during 25 years of gas field development in Arctic Russia. *Remote Sensing*, *4*(4), 1046–1068. https://doi.org/10.3390/rs4041046

Li, B., Liu, X., Zhuang, H., Shi, Q., Zeng, L., Cai, Y., Zhang, H., Cai, Y., Wu, C., & Xu, X. (2026). ALCC: Temporally consistent annual land cover maps over China from 1985 to 2022 based on an ensemble change detection method. *Journal of Remote Sensing*, *6*, 1029. https://doi.org/10.34133/remotesensing.1029

Li, J., Huang, X., Tu, L., Zhang, T., & Wang, L. (2022). A review of building detection from very high resolution optical remote sensing images. *GIScience and Remote Sensing*, *59*(1), 1199–1225. https://doi.org/10.1080/15481603.2022.2101727

Li, X., & Gong, P. (2016). An "exclusion-inclusion" framework for extracting human settlements in rapidly developing regions of China from landsat images. *Remote Sensing of Environment*, *186*, 286–296. https://doi.org/10.1016/j.rse.2016.08.029

Li, X., Guo, H., Cheng, G., Song, X., Ran, Y., Feng, M., Che, T., Li, X., Wang, L., Duan, A., Shangguan, D., Chen, D., Jin, R., Deng, J., Su, J., & Cao, B. (2025). Polar regions are critical in achieving global sustainable development goals. *Nature Communications*, *16*(1), 3879. https://doi.org/10.1038/s41467-025-59178-3

Li, Z., Li, L., Hu, T., Cheng, M., He, W., Qiu, T., Zhang, L., & Zhang, H. (2026). Satellite mapping of every building's function in urban China reveals deep built environment disparities. *Nature Communications*, *17*(1), 2827. https://doi.org/10.1038/s41467-026-69589-5

Liu, C., Huang, H., Zhang, Q., Chen, X., Xu, X., Xu, H., & Cheng, X. (2022). Arctic's man-made impervious surfaces expanded by over two-thirds in the 21st century. *Science Bulletin*, *67*(14), 1425–1429. https://doi.org/10.1016/j.scib.2022.06.001

Liu, C., Huang, H., Zhang, Q., Feng, X., Hou, X., Liu, C., Xu, H., & Cheng, X. (2024). Fine-resolution mapping and assessment of artificial surfaces in the northern hemisphere permafrost environments. *International Journal of Digital Earth*, *17*(1), 2302579. https://doi.org/10.1080/17538947.2024.2302579

Liu, C., Wang, J., Zhang, Y., Zhang, Q., Dong, N., Liu, C., Li, T., & Cheng, X. (2026). Anthropogenic disturbance dampens climate-driven greening in the Arctic. *Journal of Remote Sensing*, *6*, 1062. https://doi.org/10.34133/remotesensing.1062

Liu, C., Xu, X., Feng, X., Cheng, X., Liu, C., & Huang, H. (2023). CALC-2020: A new baseline land cover map at 10 m resolution for the circumpolar Arctic. *Earth System Science Data*, *15*(1), 133–153. https://doi.org/10.5194/essd-15-133-2023

Liu, C., Zhang, Q., Luo, H., Qi, S., Tao, S., Xu, H., & Yao, Y. (2019). An efficient approach to capture continuous impervious surface dynamics using spatial-temporal rules and dense Landsat time series stacks. *Remote Sensing of Environment*, *229*, 114–132. https://doi.org/10.1016/j.rse.2019.04.025

Manos, E. (2025). *Arctic circumpolar permafrost region building footprints from <1 meter resolution Maxar satellite imagery and OpenStreetMap (2018–2023)* [Data set]. NSF Arctic Data Center. https://doi.org/10.18739/A2R49GB8X

Manos, E., Streletskiy, D. A., Witharana, C., Haggerty, E., & Liljedahl, A. K. (2026). Refined modeling of Arctic circumpolar building stock increases estimated mid-century permafrost degradation damages. *Earth's Future*, *14*(7), e2026EF008578. https://doi.org/10.1029/2026EF008578

Manos, E., Witharana, C., & Liljedahl, A. K. (2025). Permafrost thaw-related infrastructure damage costs in Alaska are projected to double under medium and high emission scenarios. *Communications Earth & Environment*, *6*(1), 221. https://doi.org/10.1038/s43247-025-02191-7

Marconcini, M., Metz-Marconcini, A., Esch, T., & Gorelick, N. (2021). Understanding current trends in global urbanisation-the world settlement footprint suite. *GI_Forum*, *1*, 33–38. https://doi.org/10.1553/giscience2021_01_s33

Marconcini, M., Metz-Marconcini, A., Üreyen, S., Palacios-Lopez, D., Hanke, W., Bachofer, F., Zeidler, J., Esch, T., Gorelick, N., Kakarla, A., Paganini, M., & Strano, E. (2020). Outlining where humans live, the world settlement footprint 2015. *Scientific Data*, *7*(1), 242. https://doi.org/10.1038/s41597-020-00580-5

Moon, T. A., Overeem, I., Druckenmiller, M., Holland, M., Huntington, H., Kling, G., Lovecraft, A. L., Miller, G., Scambos, T., Schädel, C., Schuur, E. A. G., Trochim, E., Wiese, F., Williams, D., & Wong, G. (2019). The expanding footprint of rapid Arctic change. *Earth's Future*, *7*(3), 212–218. https://doi.org/10.1029/2018EF001088

Myers-Smith, I. H., Kerby, J. T., Phoenix, G. K., Bjerke, J. W., Epstein, H. E., Assmann, J. J., John, C., Andreu-Hayles, L., Angers-Blondin, S., Beck, P. S. A., Berner, L. T., Bhatt, U. S., Bjorkman, A. D., Blok, D., Bryn, A., Christiansen, C. T., Cornelissen, J. H. C., Cunliffe, A. M., Elmendorf, S. C., . . . Wipf, S. (2020). Complexity revealed in the greening of the Arctic. *Nature Climate Change*, *10*(2), 106–117. https://doi.org/10.1038/s41558-019-0688-1

Obu, J. (2021). How much of the Earth's surface is underlain by permafrost? *Journal of Geophysical Research: Earth Surface*, *126*(5), e2021JF006123. https://doi.org/10.1029/2021JF006123

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., & Zou, D. (2019). Northern hemisphere permafrost map based on TTOP modelling for 2000–2016 at 1 km2 scale. *Earth-Science Reviews*, *193*, 299–316. https://doi.org/10.1016/j.earscirev.2019.04.023

Oostwegel, L. J. N., Schorlemmer, D., & Guéguen, P. (2025). From footprints to functions: A comprehensive global and semantic building footprint dataset. *Scientific Data*, *12*(1), 1699. https://doi.org/10.1038/s41597-025-06132-z

Pekel, J.-F., Cottam, A., Gorelick, N., & Belward, A. S. (2016). High-resolution mapping of global surface water and its long-term changes. *Nature*, *540*(7633), 418–422. https://doi.org/10.1038/nature20584

Rantanen, M., Karpechko, A. Y., Lipponen, A., Nordling, K., Hyvärinen, O., Ruosteenoja, K., Vihma, T., & Laaksonen, A. (2022). The Arctic has warmed nearly four times faster than the globe since 1979. *Communications Earth & Environment*, *3*(1), 168. https://doi.org/10.1038/s43247-022-00498-3

Raynolds, M. K., Walker, D. A., Ambrosius, K. J., Brown, J., Everett, K. R., Kanevskiy, M., Kofinas, G. P., Romanovsky, V. E., Shur, Y., & Webber, P. J. (2014). Cumulative geoecological effects of 62 years of

infrastructure and climate change in ice-rich permafrost landscapes, Prudhoe Bay Oilfield, Alaska. *Global Change Biology*, *20*(4), 1211–1224. https://doi.org/10.1111/gcb.12500

Ren, H., Huang, X., Yang, J., & Zhou, G. (2025). Improving 30-meter global impervious surface area (GISA) mapping: New method and dataset. *ISPRS Journal of Photogrammetry & Remote Sensing*, *220*, 354–376. https://doi.org/10.1016/j.isprsjprs.2024.12.023

Schneider von Deimling, T., Lee, H., Ingeman-Nielsen, T., Westermann, S., Romanovsky, V., Lamoureux, S., Walker, D. A., Chadburn, S., Trochim, E., Cai, L., Nitzbon, J., Jacobi, S., & Langer, M. (2021). Consequences of permafrost degradation for Arctic infrastructure - bridging the model gap between regional and engineering scales. *The Cryosphere*, *15*(5), 2451–2471. https://doi.org/10.5194/tc-15-2451-2021

Shermeyer, J., Hogan, D., Brown, J., Van Etten, A., Weir, N., Pacifici, F., Hänsch, R., Bastidas, A., Soenen, S., Bacastow, T., & Lewis, R. (2020). SpaceNet 6: Multi-sensor all weather mapping dataset. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 768–777. https://doi.org/10.1109/CVPRW50498.2020.00106

Siedentop, S., & Fina, S. (2010). Monitoring urban sprawl in Germany: Towards a GIS-based measurement and assessment approach. *Journal of Land Use Science*, *5*(2), 73–104. https://doi.org/10.1080/1747423X.2010.481075

Streletskiy, D. A., Clemens, S., Lanckman, J.-P., & Shiklomanov, N. I. (2023). The costs of arctic infrastructure damages due to permafrost degradation. *Environmental Research Letters*, *18*(1), 015006. https://doi.org/10.1088/1748-9326/acab18

Suter, L., Streletskiy, D., & Shiklomanov, N. (2019). Assessment of the cost of climate change impacts on critical infrastructure in the circumpolar arctic. *Polar Geography*, *42*(4), 267–286. https://doi.org/10.1080/1088937X.2019.1686082

Tanguy, R., Bartsch, A., Nitze, I., Irrgang, A., Petzold, P., Widhalm, B., von Baeckmann, C., Boike, J., Martin, J., Efimova, A., Vieira, G., Whalen, D., Heim, B., Wieczorek, M., & Grosse, G. (2024). Pan-Arctic assessment of coastal settlements and infrastructure vulnerable to coastal erosion, sea-level rise, and permafrost thaw. *Earth's Future*, *12*(12), e2024EF005013. https://doi.org/10.1029/2024EF005013

Zhao, S., Wang, J., Guo, Y., & Liu, C. (2026). *Pan-Arctic Building Footprint Atlas v1* (Version v1) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.20796872

Zhou, Y., & Weng, Q. (2024). Building up a data engine for global urban mapping. *Remote Sensing of Environment*, *311*, 114242. https://doi.org/10.1016/j.rse.2024.114242

Zhu, X. X., Chen, S., Zhang, F., Shi, Y., & Wang, Y. (2025). GlobalBuildingAtlas: An open global and complete dataset of building polygons, heights and LoD1 3D models. *Earth System Science Data*, *17*(12), 6647–6668. https://doi.org/10.5194/essd-17-6647-2025