scientific **data**

OPEN

DATA DESCRIPTOR

PAVC: The foundation for a Pan-Arctic Vegetation Cover database

Morgan R. Steckler[1 ✉], Jitendra Kumar[2], Amy L. Breen[3], Tianqi Zhang[2], Forrest M. Hoffman[1], William W. Hargrove[4], Donald A. Walker[5], Aaron F. Wells[6], Amanda Droghini[7], Timm W. Nawrocki[7], Stan D. Wullschleger[2], Matthew J. Macander[8], Gerald V. Frost[8], Verity G. Salmon[2], David T. Barnett[9] & Colleen M. Iversen[2]

Field-measured Arctic vegetation cover data is essential for creating accurate, high-quality vegetation structure and composition maps. Extrapolating field data into high-resolution cover maps provides detailed, function-specific information for use in Earth System Models, vegetation classifications, and monitoring vegetation change over time and space. However, field campaigns that collect plant cover vary substantially in scope, method, and purpose, which makes them difficult to unify across data stores, and they are often not designed to meet remote sensing needs. In this work, we synthesized and harmonized field-based fractional cover data from various data stores to create a high-quality, consistent repository schema for remote sensing-based vegetation cover mapping applications. We developed a reproducible workflow for synthesizing visual estimate and point-intercept fractional cover data. The resultant Pan-Arctic Vegetation Cover (PAVC) database contains synthesized fractional cover at both the species and plant functional type levels. The latter includes absolute foliar cover for deciduous shrubs and trees, evergreen shrubs and trees, forbs, graminoids, lichen, bryophytes, and "other" vegetation, as well as absolute cover for litter and top cover for water and bare ground.

Background & Summary

Complex interactions between climate, the environment, and vegetation make the impact of anthropogenic emissions-induced warming difficult to estimate and predict. Notably, Arctic amplification has advanced the spread of tall shrubs and narrowed the coverage of non-vascular plants across the arctic tundra—a composition change that directly affects permafrost freeze and thaw dynamics[1–4]. Rapid warming and its potential effects on permafrost stability can disrupt global carbon cycles, so researchers are closely monitoring ongoing vegetation recomposition[5–8]. Vegetation models within Earth System Models (ESMs) are designed to represent important structural and functional variables that together control land-surface energy, carbon, nutrient, and water budgets[9]. To represent this complex information, species are aggregated into plant functional types (PFTs) that have been carefully classified and parameterized for capturing dynamic climate responses. PFT classes are based on characteristic similarities like growth form, photosynthetic pathway, and response to climate. Proper representation of PFT classes and coverage can significantly improve ESM prediction uncertainties[3]. However, field-based vegetation cover data are temporally and spatially limited and are often collected at the species-level.

To improve the coverage of vegetation data, recent studies have extrapolated field-measured cover for PFTs into gridded estimations using optical remote sensing imagery at multimeter[10–14] and sub-meter resolutions[15–18]. For example, Macander et al.[10,16] used 30-meter multispectral Landsat imagery to map PFT cover in Alaska. In another study, Nelson et al.[11] used 5-meter hyperspectral AVIRIS-NG flights to map species-level traits. They suggested bridging high-resolution flight imagery with lower-resolution satellite data to create detailed circumpolar maps of vegetation. More recently, several studies have used unoccupied aerial vehicles to collect spectral data in the field, and to map centimeter-resolution cover[15–18]. The authors emphasized how extremely

[1]Oak Ridge National Laboratory Computational Science and Engineering Division, Oak Ridge, TN, USA. [2]Oak Ridge National Laboratory Environmental Sciences Division, Oak Ridge, TN, USA. [3]University of Alaska Fairbanks International Arctic Research Center, Fairbanks, AK, USA. [4]USDA Forest Service Southern Research Station, Asheville, NC, USA. [5]University of Alaska Fairbanks Institute of Arctic Biology, Fairbanks, AK, USA. [6]AECOM Technical Services, Anchorage, AK, USA. [7]University of Alaska Anchorage Alaska Center for Conservation Science, Anchorage, AK, USA. [8]Alaska Biological Research, Inc., Fairbanks, AK, USA. [9]National Ecological Observatory Network, Battelle, Boulder, CO, USA. ✉e-mail: stecklermr@ornl.gov

high-resolution (sub-meter) products could be connected with coarser (multi-meter) satellite images to produce detailed regional maps. These recent steps in Arctic cover mapping are paving the way toward panarctic-scale, high-quality cover products based on data collected in the field.

In order to create high-quality gridded models of vegetation coverage, synthesized cover data is required. Field campaigns that collect cover data vary substantially in scope, method, and purpose, which makes them difficult to unify across data stores. Regardless, several groups have collected and standardized disparate datasets. For example, Walker and Reynolds[19] initiated the Arctic Vegetation Archive (AVA), which is a repository of plot data synthesized for use in an Arctic vegetation classification. The AVA is an on-going project where data are stored within regional databases[20]. Specifically, the AVA-Alaska contains publicly accessible data from Alaska and select sites in the Canadian high Arctic. Up-to-date AVA-Russia data are available upon request from their respective website[21], but users may access a subset version on the Data Dryad publishing platform[22]. Similar plot data in Europe—including Svalbard, Iceland, Faroe Islands, and northern Scandinavia—are currently available upon request via the European Vegetation Archive[23].

The International Tundra Experiment (ITEX) has documented Arctic vegetation change internationally for thirty years[24]. They store their survey unit data in the Polar Data Catalog from the Canadian Cryospheric Information Network. This data store is an excellent source for locating Arctic-related data, but the cover datasets are mixed in among hundreds of other types of data. An additional source is the Alaska Vegetation (AKVEG) Database[25], which is a high-resolution vegetation, classification, and taxonomic data repository for the state of Alaska that is coordinated by the Alaska Geospatial Council Vegetation Working Group. However, the most recent version of their database is under development and was not publicly available at the time this manuscript was submitted. Overall, vegetation cover datasets can be stored in disparate formats and languages; they are spread across many electronic databases with various formatting and quality standards; and they contain data that were collected using different taxonomic structures, controlled vocabularies, methodologies, and scales. There also exist many datasets that are not publicly available at all.

With vegetation cover mapping applications in mind, the goal of this study was to design a standardized database and pipeline for synthesizing cover data from field campaigns and repositories across the globe. Specifically, we synthesized a subset of vegetation surveys from across Alaska, USA, to instantiate our database, PAVC. We (1) developed a pipeline for synthesizing cover data derived from survey units in Arctic Alaska. The workflow is reproducible and open-source in Python, and our dataset incorporates vegetation cover collected using both visual estimate and point-intercept methods described in the Usage Notes Section of this paper. The workflow (2) produced a database of survey unit-wise Arctic Alaska cover stored at both a standardized species level and standardized PFT level. We incorporated auxiliary information and flags for sorting and filtering survey units, such as collection date, field collection method, author, citation, the purpose of field work, GPS accuracy, cover type, etc. Lastly, we (3) summarized the statistics and patterns of PAVC data to emphasize not only how field campaigns differ greatly, but also how they can be used together in remote sensing-based studies to develop accurate cover estimates at a regional scale.

Methods

Figure 1 illustrates the data synthesis and harmonization workflow. We (1) gathered data from publicly available data sources, (2) preprocessed data values and tabular structures, (3) defined and extracted metadata, (4) standardized species names and PFTs, and (5) aggregated species-level cover to the PFT level. Throughout the workflow, we iteratively checked for errors and consulted Arctic vegetation experts for guidance. We exclusively used species-level cover data, and we applied our workflow consistently across each data source to produce two cover datasets and standardized plot information. The synthesized database contains cover at the species level, which includes 644 unique Arctic plant species from across 977 survey units (Fig. 2) derived from a selection of source datasets (Table 1). The database also contains cover as an aggregation of the species-level dataset to the PFT level. These PFTs are an expansion of the Nawrocki *et al.*[26] Checklist of Vascular Plants, Bryophytes, Lichens, and Lichenicolous Fungi of Alaska—henceforth the AKVEG Checklist—which contains species-to-PFT assignments. Leaf retention information for shrubs is derived from Macander *et al.*[2] Not all PFTs and non-vegetation classes may have been recorded as part of the survey design in source observations and studies, and thus we record the absence of a PFT in the survey by encoding the cover with a no-data value in the PAVC database.

Data sources.   All of the survey units we synthesized are located in Arctic Alaska (Fig. 2), which we delineated based on the Circumpolar Arctic Vegetation Map (CAVM). The CAVM defines the Arctic according to its unique climate and flora, with the treeline defining the southern limit[27]. In Alaska, this area is predominantly composed of the Arctic Coastal Plain, Arctic Foothills, Brooks Range, Seward Peninsula, and Yukon-Kuskokwim Delta ecoregions[28]. For the years 2010 to 2021, we extracted cover datasets from five US data stores: the AKVEG Database[25,29], the North Slope Science Catalog[10], the National Ecological Observatory Network[30], the Alaska Arctic Vegetation Archive[31–37], and vegetation survey units from the Next Generation Ecosystem Experiments in the Arctic (NGEE Arctic)[38].

For initial development purposes and time constraints, we selected only a small subset of available data—plots that used either visual estimation or point-intercept collection methods, had associated non-vegetation data, and that were collected on or after 2010. For future plans on PAVC development, refer to the Usage Notes section of this paper.

Table 1 summarizes information about the source datasets we synthesized. For each dataset, we developed a custom *jupyter* notebook used for data cleaning and standardization. Most steps were automated using functions from a custom Python module, but when unavoidable, some steps were performed manually in a spreadsheet software like Excel. Regardless, each step of the standardization process is clearly documented using Markdown
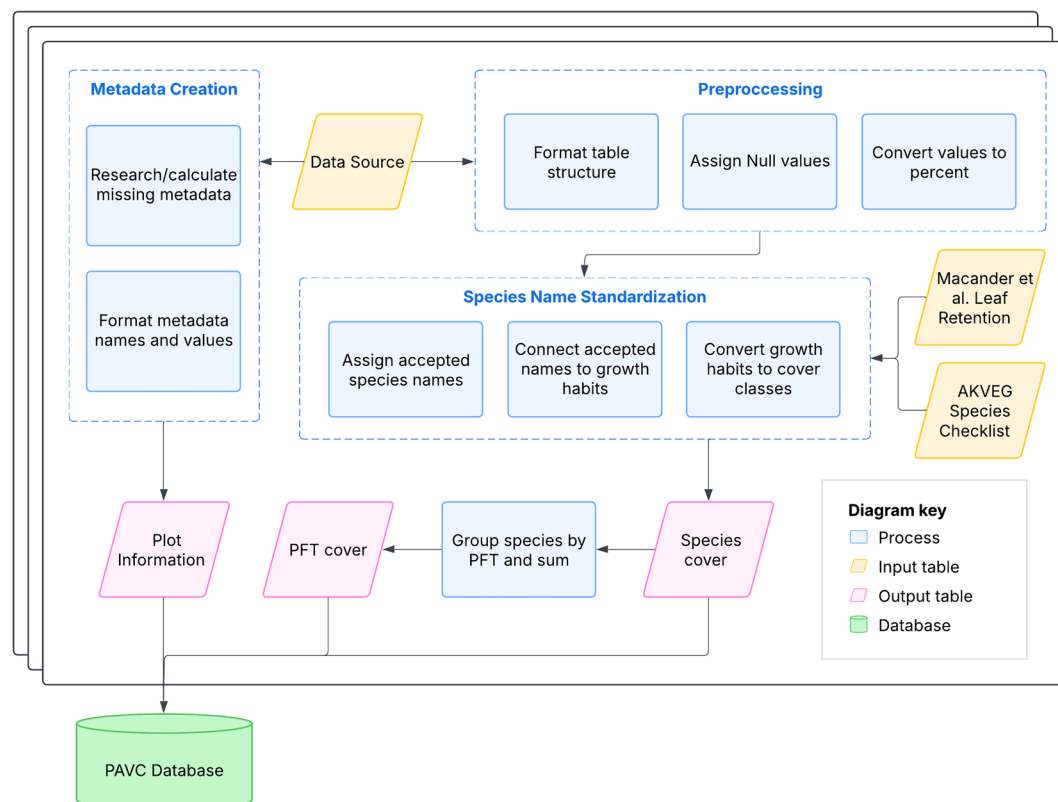
Fig. 1 General workflow that can be applied to all data sources synthesized into the PAVC database.
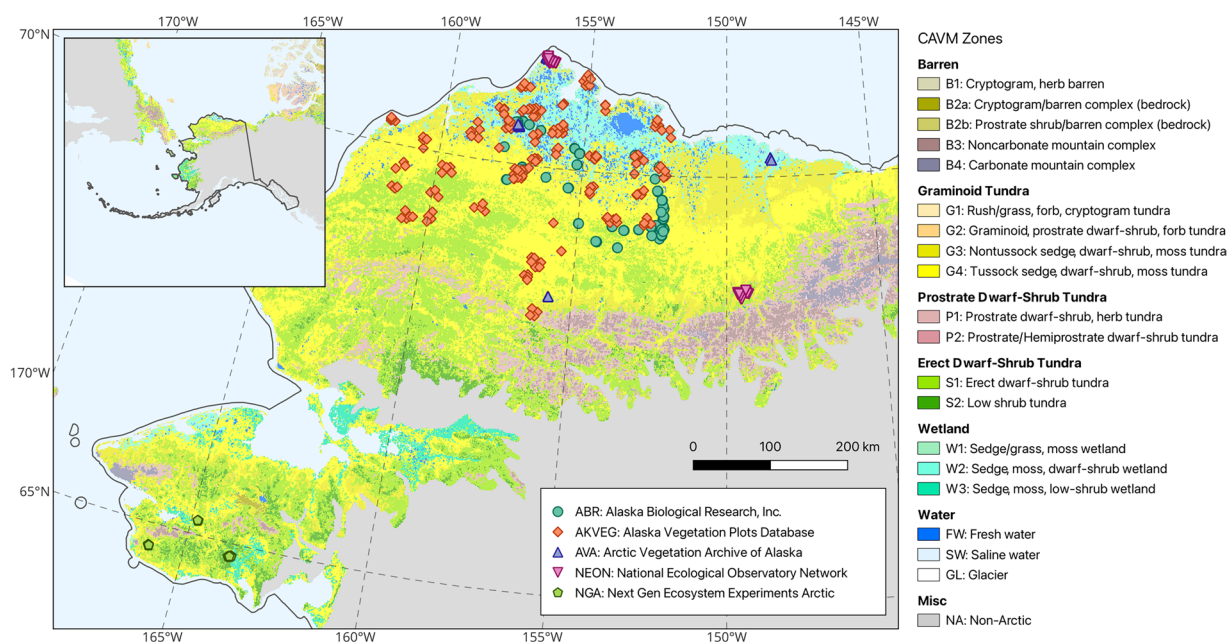

Fig. 2 Database coverage map underlain with Circumpolar Arctic Vegetation Map (CAVM) zones[27]. Most plots in ABR (62.6%), AVA (71.8%), and NEON (72.5%) are located on CAVM Graminoid Tundra. AKVEG plots (44.2%) are mostly located on CAVM Graminoid Tundra and secondarily on Wetlands (32.4%). Lastly, 68.4% of NGA plots are found on CAVM Erect Dwarf-Shrub Tundra.

cells in the data source-specific notebooks. A separate notebook was developed to further integrate the standardized datasets into a synthesized and harmonized database with four tables—species-level cover, PFT-level cover, survey unit auxiliary information, and a species to PFT checklist.

Data store	Source	Subsource	Dataset year(s)	n units	n species
Alaska Vegetation Plots Database	AKVEG & BLM	Nawrocki et al.[25]	2012–2017	185	460
North Slope Science Catalog	ABR	Macander et al.[10]	2012	107	294
National Ecological Observatory Network	NEON	NEON[30]	2021	280	93
Alaska Arctic Vegetation Archive	AVA	Villarreal et al.[31]	2010	31	59
		Villarreal[32]	2010	33	111
		Sloan et al.[33,34]	2012	48	37
		Davidson[35]	2014	140	62
		Walker et al.[36]	2014	29	52
		Walker et al.[37]	2015	27	43
NGEE Arctic	NGA	Breen et al.[38]	2016–2018	97	238
Total				977	644

Table 1. Summary of source datasets synthesized into PAVC.

Once the data sources were gathered, we implemented the generalized workflow visualized in Fig. 2. Details on the synthesis workflow steps are outlined in the following subsection.

Synthesis. We downloaded the cover data from each data store and inspected tables to correct for any lingering erroneous values, syntax errors, or formatting inconsistencies. We standardized the table format so that row indices represented unique adjudicated species names—carefully examined taxonomic names—and column headers contained unique, dataset-specific survey unit IDs. Sometimes, original datasets contained non-numeric values for identifying missing cover, which we automatically detected and converted to *numpy* null values. For the AVA data source in particular, surveys often used ordinal cover scales, such as the 7-step Braun-Blanquet scale, instead of exact percentages as the cover unit. The ordinal cover code is generally an integer representing a range of visually estimated cover. We therefore converted fractional cover scales into their associated midpoint percentage[39]. For survey units that were returned to at a later date—identified by duplicate IDs or coordinates—only the most recent observations were retained. Any missing metadata were manually identified from the associated manuscripts and recorded. If metadata were incorrectly formatted, we automatically corrected them where possible and manually re-assigned values where necessary.

Next, for each source dataset, we assigned all of the species names to a standardized accepted name. This way, we had one unified species checklist for all of the source datasets that did not contain duplicates in the form of misspelling, misapplications, or misnaming. We used standardized accepted names from the AKVEG Checklist. In the checklist, species names are formatted as *Genus species infraspecies-label infraspecies Author*. However, each source cover dataset used inconsistent synonyms, syntaxes, and spellings for the species names, so we could not always find a matching species name in the AKVEG Checklist on the first try. Thus, we implemented a method (visualized in Fig. 3) for matching a source dataset's species name to an AKVEG Checklist species name.

When a source dataset species name did not match any AKVEG Checklist species names, we then attempted to match only the *Genus species* part of the full names. If that failed again, we attempted to match only the *Genus*. If matched with a genus, we created a list of "potential" PFTs and "potential" accepted species names that we manually looked through to assign an accepted name. Finally, if no match was found, a specialist in botanical nomenclature manually identified the correct accepted species name and PFT. Often in these cases, the dataset genus was misspelled, misapplied, or it was not included in the AKVEG Checklist due to their ongoing lichen species updates.

Once we matched a source dataset's adjudicated species name to possible accepted names using a shared join key, we used the "habit" column from the AKVEG Checklist to assign potential PFTs. In this work, a PFT is defined as the potential mature growth habit (or growth form) of a species, which is mostly dependent on environmental limiting factors, but also genetic predispositions[40]. One species or genus may have many growth habits, or surveyors may have only recorded cover at the genus-level. So, we assigned the species or genus a list of unique "potential" habits. This was usually a list with only one associated habit. Additionally, shrub species have a wide range of growth forms, including dwarf shrubs, low shrubs, and tall shrubs, several of which are classified by the shrub's height. However, many surveys did not measure their shrub species, so we did not subcategorize shrubs based on height. Instead, shrubs were assigned a leaf retention boolean—deciduous or evergreen. This information was derived from Macander et al.[2].

After assigning PFTs, we exported the "potential" accepted name list for each source dataset's species name so we could manually standardize names across all datasets. Finally, we aggregated species-level cover to the PFT level. Our classification schema was based on those defined by the Energy Exascale Earth System Model (E3SM) Land Model (ELM)[41], being developed under the US Department of Energy's NGEE Arctic project whose goal is to advance the predictive understanding of the Arctic tundra ecosystem. Our final schema included 8 PFTs: non-vascular plants with lichen and bryophyte subcategories, trees with deciduous and evergreen subcategories, shrubs with deciduous and evergreen subcategories, graminoids (grasses), and forbs (herbaceous flowering plants). Trees are included in this Arctic vegetation schema because of boreal encroachment into the CAVM-delineated tree line, especially along the Seward Peninsula. We also aggregated non-vegetated areas into bare ground, water, and litter categories (standing dead vegetation, scat, and leaf litter). Any other cover that did not fit in our schema were classified as "other," which included fungi, algae, and cyanobacteria.

The scripts and notebooks used for synthesis are stored on Github at https://github.com/climatemodeling/pavc.
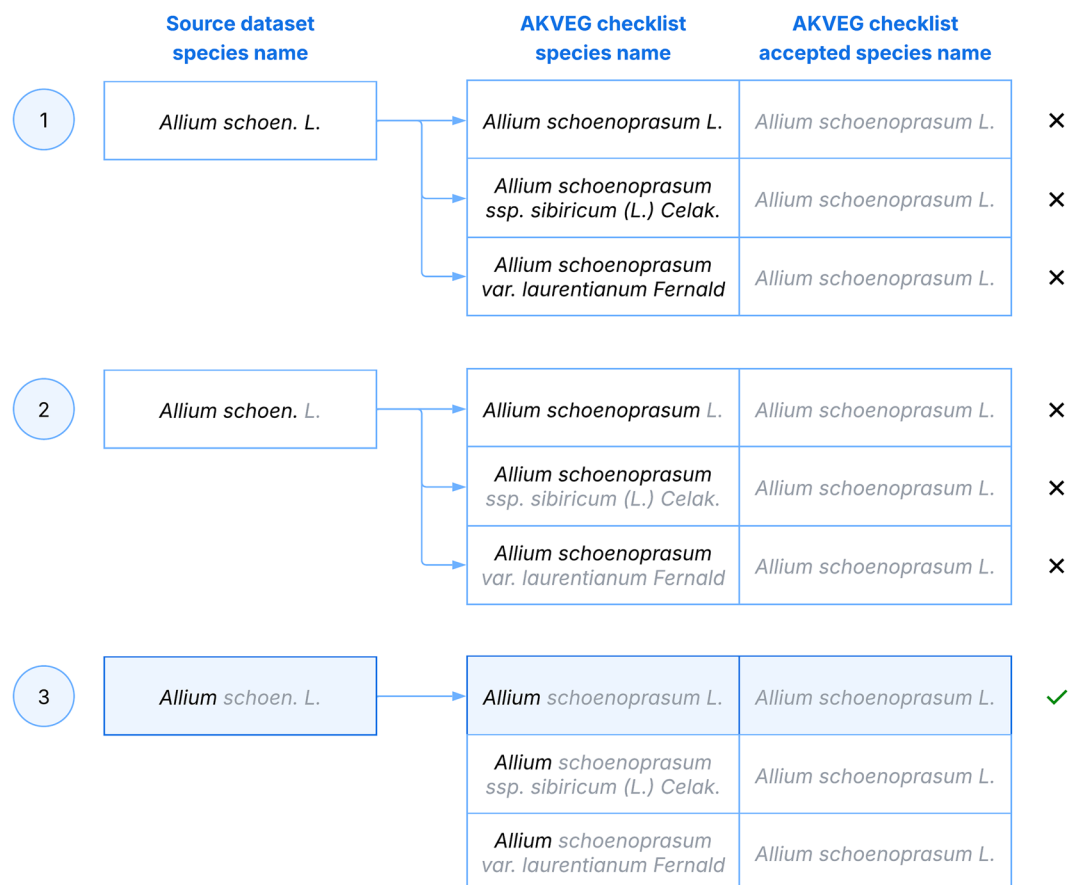
Fig. 3 Example workflow for identifying an accepted species name and PFT by (1) checking for a perfect match, (2) checking the first two words for a match, and (3) checking only the first word for a match.

Data Records

The PAVC database is archived on the Environmental System Science Data Infrastructure for a Virtual Ecosystem (ESS-DIVE) Repository[42] at https://doi.org/10.15485/2483557. The final synthesized database and meta-data were formatted according to Environmental System Science Data Infrastructure for a Virtual Ecosystem (ESS-DIVE) File-Level Meta-Data (FLMD) v1 standards[43]. ESS-DIVE is a data repository for Earth and environ-mental sciences data supported by the Department of Energy Biological and Environmental Research program. There are four datasets within the database: "species_pft_checklist," "synthesized_species_fcover," "synthesized_pft_fcover," and "survey_unit_information," all of which are stored in UTF-8-Sig encoded Comma Separated Value (CSV) format (Fig. 4). For each data file, there is an associated FLMD CSV that provides information about the data file itself, as well as a data dictionary (DD) CSV, which explains the data file's header information. This format ensures easy readability by people, scripting languages, and table-viewing software.

First, the "species_pft_checklist" is useful to users because it shows how data source species names were con-nected to a standardized accepted name and PFT. Users can modify this checklist to incorporate their own PFT schema and re-make the PFT files to match their needs. The "synthesized_species_fcover" and "synthesized_pft_fcover" datasets contain fractional cover (fcover) column(s) with values expressed as a percent. Percent cover values range from a minimum of 0% to a maximum of 163.71% and represent absolute (rather than top) foliar cover. For more information on the different types of cover, refer to the Usage Notes Section. To create the "synthesized_pft_fcover" dataset, the species-level cover values were summed to the PFT level so that each row is a survey visit_id, columns are PFTs, and cells were filled with summed fcover. Lastly, the "survey_unit_informa-tion" dataset contains rows of survey visit_ids and columns of auxiliary information on source data authorship, collection methods, collection dates, survey coordinates, spatial uncertainty, quality flags, and more. Figure 4 displays a complete layout of the database.

survey_unit_information.csv	
visit_id	str
plot_name	str
survey_year	int
survey_month	int
survey_day	int
plot_area	float
plot_shape	str
latitude_y	float
longitude_x	float
georef_source	str
georef_accuracy	float
coord_epsg	str
data_subsource	str
data_source	str
data_source_html	str
data_subsource_citation	str
survey_method	str
fcover_scale	str
survey_purpose	str
admin_unit	str
admin_country	str
fire_years	list
bioclim_subzone	str
duplicated_coords	list
duplicated_date	list
cavm_unit	str
cavm_description	str

synthesized_pft_fcover.csv	
visit_id	str
deciduous_shrub_cover	float
deciduous_tree_cover	float
evergreen_shrub_cover	float
evergreen_tree_cover	float
forb_cover	float
graminoid_cover	float
nonvascular_sum_cover	float
bryophyte_cover	float
lichen_cover	float
litter_cover	float
other_cover	float
water_cover	float
bareground_cover	float

species_pft_checklist.csv	
visit_id	list
dataset_species_name	str
accepted_species_name	str
accepted_species_name_author	str
data_source	list
data_subsource	list
taxon_rank	str
naming_authority	str
category	str
habit	str
pft	str
nonstandard_pft	str

synthesized_species_fcover.csv	
visit_id	str
accepted_species_name	str
fcover	float

legend	
join key	
data type	
column name	

Fig. 4 PAVC data files with their respective column names, data types, and join keys.

Currently, there are 978 survey units stored in PAVC, where 308 are sourced from AVA, 280 from NEON, 185 from AKVEG, 107 from ABR, and 98 from NGA. Between all 978 survey units, there were 630 unique Arctic plant species, not including species that were only identified to the genus level. Because survey unit data were recorded after 2010, the standard unit for cover observations was often percent, with a few older AVA units using Braun-Blanquet codes. Most survey units in our dataset (70.1%) were collected using visual estimate-based methods, where plots are on average 2.23 m area. Visual estimate-based sampling methods were used for three of the five input data sources—AVA, NEON, and NGA. In contrast, there are fewer point-intercept methods used in our database, partly because they are sourced from only two of the five datasets (ABR and AKVEG) and partly because these survey units are on average 5,319.5 m area, which covers over 2,385 times more area than the average visual estimate-based survey unit (2.23 m area). For detailed information on survey methodologies, refer to the Usage Notes section of this paper.

Technical Validation

To ensure the validity of our data values, we first performed iterative error checks throughout the synthesis processes. We visually assessed tables for structural errors like hanging columns or misplaced values. We corrected misformatted or erroneous values that raised data type errors. We standardized species names by only using accepted names to correct misspellings, misapplied synonyms, and unaccepted names. Then, we visualized cover distributions for each PFT (Fig. 5) to ensure that the cover values were within a reasonable range after aggregating to the PFT level. Shrubs, graminoids, bryophytes, and litter were most abundant across the survey units. Relatively fewer survey units recorded trees, forbs, lichen, bare ground, and water, though some surveys did not record lichen cover and used disparate non-vegetation definitions. Also, due to large survey unit sizes and immense vegetation heterogeneity across the Arctic, it is rare to have a single species or genus dominating the entire survey unit. "Other" vegetation types—e.g., algae, fungi, and cyanobacteria—only make up 4.3% of all non-zero cover values and are present in 30.09% of all survey units. The median cover of "other" vegetation is 8 percent.

Next, we compared the PFT fractional cover of our survey units to established CAVM map units, or ecoregions, to check for potential spatial discrepancies. The survey unit data are dispersed across 11 of the 20 CAVM map units (Fig. 6). These map units are defined by their dominant vegetation type's physiognomy, and are divided into 5 main categories—*B* barren, *G* graminoid-dominated tundra, *P* prostrate dwarf-shrub dominated tundra, *S* erect dwarf-shrub dominated tundra, *W* wetland, *FW* fresh water, and *SW* saline water, as well as their sub-classifications[27]. In the barren map unit, our survey units contain a mix of deciduous shrubs, graminoids, and bryophytes, with very few forbs, lichen, or evergreen shrubs. Forbs and lichen may be underrepresented or unmeasured in these areas given that these plants are expected to be of scattered distribution in the barren map unit.

In the graminoid-dominated tundra map unit, survey units contain high percentages of (especially) graminoids, litter, shrubs, and bryophytes, which aligns well with the map unit's description. However, CAVM accuracy is as low as 11% for graminoid tundra, and this confusion can be noted by the high spatial dispersal of
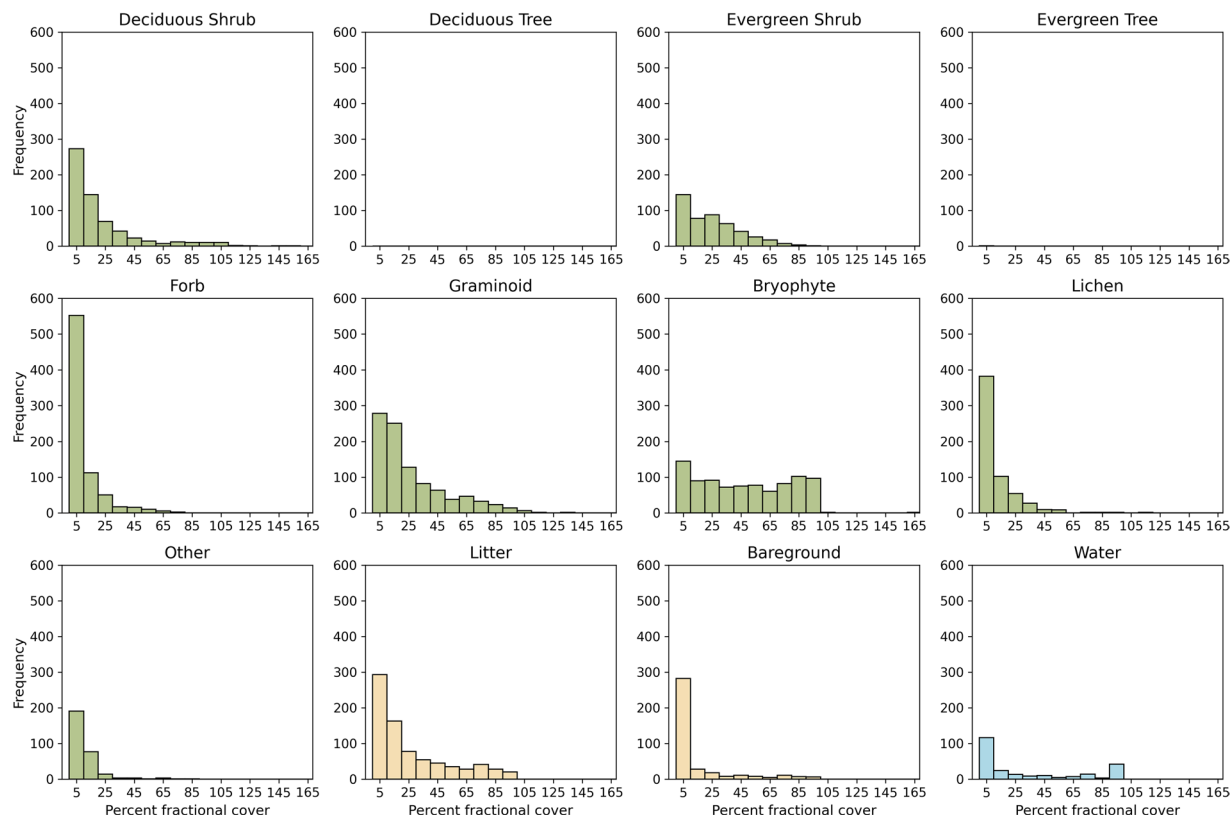
Fig. 5 Fractional cover distribution plots of each cover type in the PAVC database. Frequency counts indicated by bar height are green for vegetation, orange for non-vegetation, and blue for water. All of the x and y axis limits are shared, allowing for direct comparison between PFTs.

PAVC graminoids in Fig. 6. Prostrate dwarf-shrub dominated tundra are poorly represented in PAVC survey units, though P map units are sparse overall compared to other map units, and they are found near difficult terrain and mountainous regions. Deciduous shrubs and litter dominate the erect dwarf-shrub tundra, which has CAVM map unit accuracies between 43 and 62%. Deciduous shrubs produce seasonal litter, so this relationship is expected, though it is important to note that not all surveys measured litter cover. Bryophytes co-dominate the erect dwarf-shrub and moss tundra per the unit description, but CAVM erect dwarf shrubs were most often confused—usually with graminoids. Further, graminoids and bryophytes appear to dominate wetland complexes. Lastly, there is sizable overlap between fresh water, saline water, and PAVC survey unit cover measures of graminoids, litter, and bryophytes. Indeed, graminoids often grow in standing water, and bryophytes thrive in moist environments.

Usage Notes

Survey unit design. The five source datasets contain cover collected by a large number of researchers and projects over 11 years. Each project has their own unique study design that influences the selection of survey unit locations and protocols for vegetation surveys conducted in the field. These differences are important to consider and account for when data are used in remote sensing-based studies. For example, many of the AVA plots were 1 m area in size, while study areas from AKVEG often consist of circular plots with large, radial 25 m transects. Smaller survey units may not be spatially large enough to match the granularity of most publicly available satellite-based remote sensing platforms such as ESA Sentinel (10–20 m), Landsat (30 m), MODIS (250–500 m), VIIRS (370–740 m), etc. However, studies using extremely high resolution images from commercial satellites, airborne, or unmanned aerial system-based platforms may find 30 m AKVEG survey units to be too large in comparison, leading to mixed pixel effects. To manage discrepancies, users can filter out plots that do not meet their needs, or when possible, manipulate cover values to better match the resolution of satellite imagery. For example, Zhang *et al.*[44] successfully aggregated smaller plots into a larger area of representation by calculating the average coordinates and cover data of plots within *x* meters of each other.

Spatiotemporal context. It is also essential to consider the date, time of year, and location of survey unit data collection because some vegetation types may be more dynamic than others and change rapidly over time. Disturbance events, such as fires, may lead to drastic transitions in vegetation composition since data collection. In contrast to non-vascular plants, cover of fast growing graminoids can change quickly and shrubs can change significantly in their canopy height and crown size. Deciduous shrubs and forbs are phenologically dynamic,
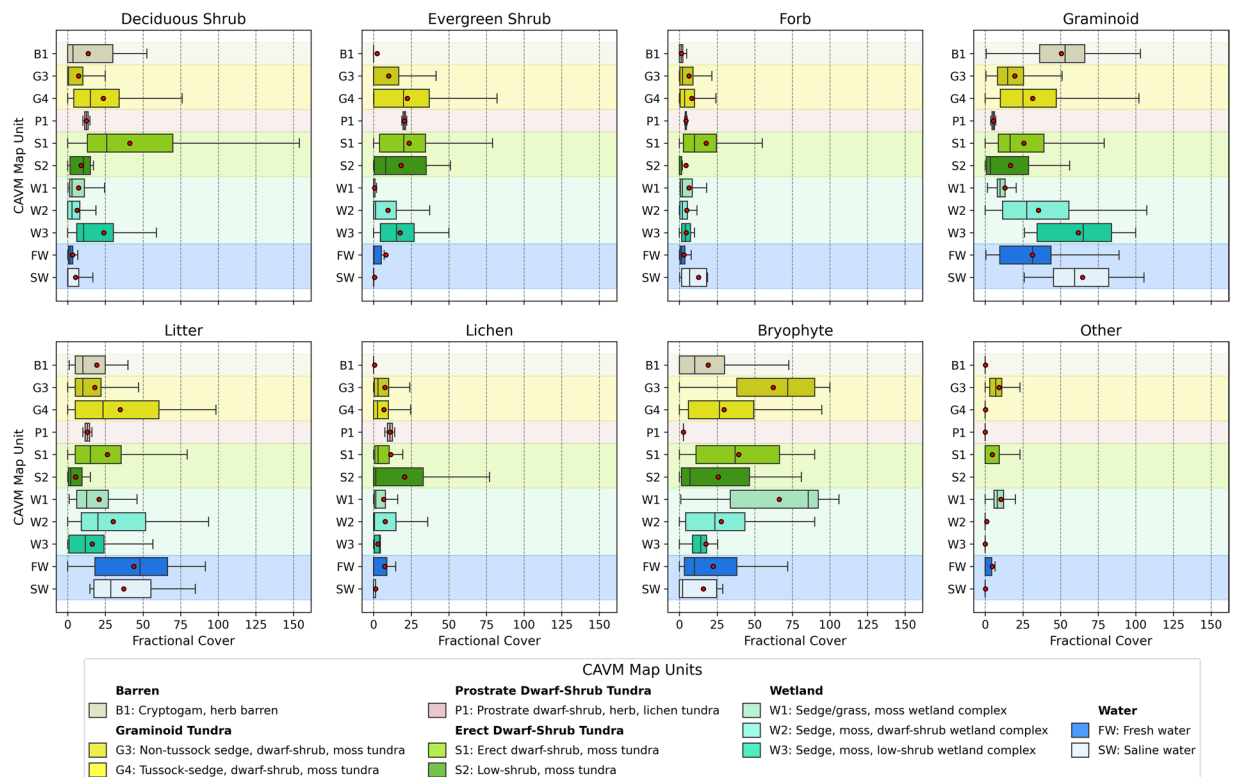
Fig. 6 Each panel shows the fractional cover distribution of a cover type grouped by the CAVM zone sampled via plot coordinates. Plot measurements of graminoids, litter, and bryophytes did not exhibit clear alignment with any particular CAVM zone, which is likely due to CAVM uncertainties being as low as 11% for G1 and G2 and as low as 43% for S2.

while graminoids and evergreen shrubs are less so. Thus, recent satellite imagery may have some discrepancy in comparison to conditions when the survey unit data were collected, especially for some PFTs. Further, geopositioning technology has improved over the years, and the recorded location coordinates for field survey units have geolocation errors that depend on the GPS equipment and methods, with older survey units tending to have poorer geolocation. This is also important when using data from very small survey units and imagery with small pixel sizes; a small GPS error in a small plot can lead to very skewed imagery comparisons.

Cover collection methods: visual estimation. Lastly, how fieldwork was designed, how cover was measured, and who performed these tasks can affect cover variance between survey units and study areas[45]. Among the synthesized datasets, we identified two broad categories for sampling design (Fig. 7). The first category is visual estimate sampling, where cover is visually estimated within a defined area. The size, shape, type, and spatial organization of visually sampled plots is dependent on the purpose of a study as well as the community surveyed. For example, shrublands require larger plots, while lichen communities do not[46]. Often, surveyors use visual estimation techniques to record the top or absolute cover of individual species. Absolute cover is the proportion of the plot's area covered by vegetation spread across all heights in the plant community, so it can sum to over 100 percent. In contrast, top cover represents the proportion of the plot's area covered by vegetation in only the top layer of the plant community, and it should always sum to 100[47]. Plots can be placed subjectively in a study area, or they can be placed stratified and/or randomly. The cover is recorded as visual percentage estimates, but historically, cover was estimated using ordinal cover classes. One such survey design, called a relevé, uses subjective plot placement and Braun-Blanquet estimations of cover, where a range of percentages is assigned to one cover class. For example, Braun-Blanquet class number 3 indicates cover between 1 and 5 percent, with a midpoint of 3 percent.
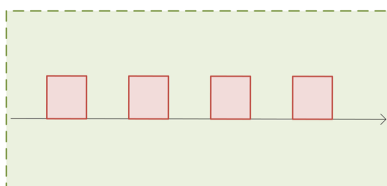
There are many ways to sub-classify visual estimation sampling techniques. Some common techniques include Braun-Blanquet relevés, the Daubenmire method, and the Whittaker multi-scale method, all of which may have various configurations and sub-variations[47,48]. We define 3 visual estimate methods: simple Plots (P), Plots along a Transect (PT), and Center-staked Plots along Transects (CPT). We found that 518 survey units employed the P method with plot sizes of 2.1 area. These small plots do not align with the spatial resolution of publicly available satellite imagery, e.g., 10–20 m Sentinel-2 or 30 m Landast images. Fewer survey units (168) used the PT method with 1 m area plots placed along transects. If plot measurements are averaged over the length of a transect, PTs could align well with the length or width of a Sentinel-2 pixel if the transect is long enough, wide enough, and oriented correctly. Lastly, none of the survey units included in the final database used the CPT method, but CPTs could represent the entire area of a Sentinel-2 pixel. Because these three sampling categories are generalized, the randomness of plot placement, number of plots along transects, and positioning
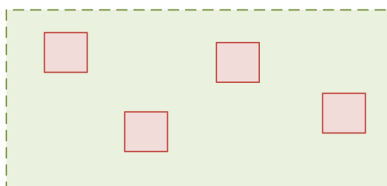
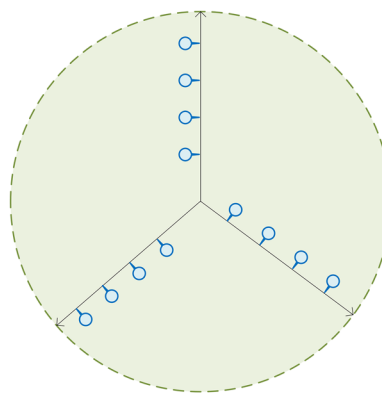**Visual estimate methods**
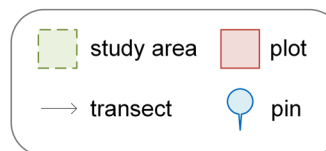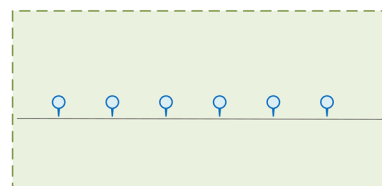
(CPT) Center-staked Plot along Transect

Point-Intercept methods

(CPIT) Center-staked Point-Intercept along Transect

(PT) Plot along Transect
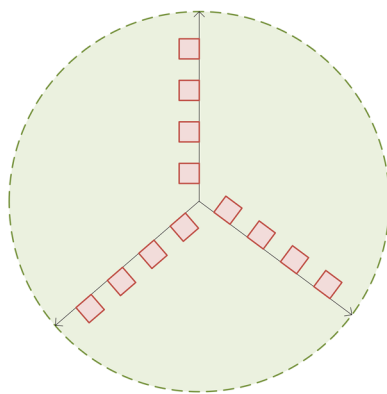
(PIT) Point-Intercept along Transect

(P) Simple Plot

| study area | plot |
| transect | pin |

Fig. 7 Aerial-view sketches of vegetation cover sampling methods.

of plots along transects varies by study. For more specific information on the collection method used for a survey unit, users should reference the citation metadata included in the database.

Cover collection methods: point-intercept. The second common sampling technique is the point-intercept method, or the "pin-dropping" method[48,49]. Surveyors use transects with equally-spaced intervals of pins, poles, lasers, or crosshairs. Surveyors identify and count the species that intersect, or "hit," each pin. If a species is a "first-hit," it will represent the uppermost canopy layer, or top cover. Any further hits of different species along the pin will elucidate lower canopy species presence, or absolute cover. The number of hits for an individual species is divided by the total number of pin points to get absolute cover. Note that there are sampling methods that record multiple hits of the same species, which reflects foliage density, and they are neither top nor absolute cover. Point-intercept methods often sample at the transect scale which makes them better suited for remote sensing applications. It can be overall more accurate and precise than visual estimation-based methods. However, point-intercept methods will not capture a complete picture of species diversity in a study area unless there is an effort to record trace species as well, which contributes to uncertainty in species level studies using these observations. In that regard, visual estimation methods are more likely to identify rarer species, while point-intercept methods alone can miss them. However, when aggregated to the PFT level, missing rare species will not undermine data completeness like it would in a species-level diversity study. In total, we identified two point-intercept subtypes: Point-Intercept along Transect (PIT) and Center-staked Point-Intercept along Transect (PITC). We found that 292 survey units used the PITC method, with an average unit size of 5,319.5 area. None of the observations in our synthesized dataset used the simpler PIT method.

Limitations and future work. Survey unit design, spatiotemporal context, and collection methods are some of the key ways to filter cover data in PAVC. However, future versions of the database will incorporate more filterable features such as the shape of a survey unit, geopositioning error relative to plot size, methodological sub-categories, species height, and boolean columns indicating whether trace species, lichen, bryophytes, litter, and/or dead vegetation were surveyed. Further, this database only includes visual estimate and point-intercept plots, but high-resolution aerial imagery has also been used to successfully extrapolate gridded map products.

The next version of PAVC will synthesize plots outside of Arctic Alaska and across the pan-Arctic: to Greenland, Canada, the European Union, and Russia. This also means that the current AKVEG Checklist will have to be strategically expanded into a pan-Arctic Vegetation Checklist with the guidance of international Arctic vegetation experts. For now, the PAVC is archived on ESS-DIVE[42], but we are also in the process of developing a standalone website for direct, server-side queries and downloads via a user interface and Python. This makes PAVC available and easily accessible to users and other data stores that wish to leverage the synthesis database.

In conclusion, the PAVC database contains vegetation cover at a standardized species level that was further aggregated into predefined cover classes. The creation of high-quality gridded products with coverage over immense remote landscapes are entirely dependent on the underlying plot data. The PAVC ensures that remote sensing scientists have publicly accessible, quality assessed, standardized, and filterable plot data in order to make informed decisions about which data is suitable for their analyses. Further, the database is immediately relevant and useful in remote sensing contexts. Many plots synthesized into the PAVC were used in Macander *et al.*[2,10] to map top fractional cover of PFTs for all of Alaska using Landsat imagery. The PAVC database was used by Zhang *et al.*[44] to produce absolute fractional coverage maps via Sentinel-2 multispectral, Sentinel-1 synthetic aperture radar, and ArcticDEM topographical products. Further, as a partial funder of the project, this database will be used by the NGEE Arctic team to meet their land modeling goals. In the future, full panarctic products are necessary for understanding Arctic amplification, change in vegetation resources, permafrost responses to increasing temperatures, and more. The remote sensing products derived from these plot data—as well as their role in ESMs—are necessary for making informed decisions on resource management, energy policy, infrastructure, and planning adaptive strategies in the face of ongoing climate shifts.

Code availability

Scripts and notebooks used to synthesize the data are publicly accessible on https://github.com/climatemodeling/pavc.

References

1. Wilcox, E. J., Bennett, K. E. & Boike, J. Bridging gaps in permafrost-shrub understanding. *PLOS Climate.* **3**, e0000360, https://doi.org/10.1371/journal.pclm.0000360 (2024).
2. Macander, M. J. *et al.* Time-series maps reveal widespread change in plant functional type cover across Arctic and boreal Alaska and Yukon. *Environ. Res. Lett.* **17**, 054042, https://doi.org/10.1088/1748-9326/ac6965 (2022).
3. Wullschleger, S. D. *et al.* Plant functional types in Earth system models: past experiences and future directions for application of dynamic vegetation models in high-latitude ecosystems. *Ann. Bot.* **114**, 1–16, https://doi.org/10.1093/aob/mcu077 (2014).
4. Elmendorf, S. C. *et al.* Plot-scale evidence of tundra vegetation change and links to recent summer warming. *Nature Clim. Change.* **2**, 453–457, https://doi.org/10.1038/nclimate1465 (2012).
5. Post, E. *et al.* The polar regions in a 2°C warmer world. *Sci. Adv.* **5**, eaaw9883, https://doi.org/10.1126/sciadv.aaw9883 (2019).
6. Overpeck, J. T. & Breshears, D. D. The growing challenge of vegetation change. *Science.* **372**, 786–787, https://doi.org/10.1126/science.abi9902 (2021).
7. Oehri, J. *et al.* Vegetation type is an important predictor of the arctic summer land surface energy budget. *Nat. Commun.* **13**, 6379, https://doi.org/10.1038/s41467-022-34049-3 (2022).
8. Schuur, E. A. G. *et al.* Ecosystem and soil respiration radiocarbon detects old carbon release as a fingerprint of warming and permafrost destabilization with climate change. *Philos. Trans. R. Soc. A.* **381**, 20220201, https://doi.org/10.1098/rsta.2022.0201 (2023).
9. Bonan, G. B. Forests and climate change: forcings, feedbacks, and the climate benefits of forests. *Science.* **320**, 1444–1449, https://doi.org/10.1126/science.1155121 (2008).
10. Macander, M. J., Frost, G. V., Nelson, P. R. & Swingley, C. S. Regional quantitative cover mapping of tundra plant functional types in arctic Alaska. *Remote Sens.* **9**, 1024, https://doi.org/10.3390/rs9101024 (2017).
11. Nelson, P. R. *et al.* Remote sensing of tundra ecosystems using high spectral resolution reflectance: opportunities and challenges. *J. Geophys. Res. Biogeosci.* **127**, e2021JG006697, https://doi.org/10.1029/2021JG006697 (2022).
12. Langford, Z. *et al.* Mapping Arctic plant functional type distributions in the Barrow Environmental Observatory using WorldView-2 and LiDAR datasets. *Remote Sens.* **8**, 733, https://doi.org/10.3390/rs8090733 (2016).
13. Langford, Z.L., Kumar, J. & Hoffman, F.M. Convolutional Neural Network approach for mapping Arctic vegetation using multi-sensor remote sensing fusion. *IEEE International Conference on Data Mining Workshops (ICDMW), New Orleans, LA, USA.* 322–331. https://doi.org/10.1109/ICDMW.2017.48 (2017).
14. Langford, Z. L., Kumar, J., Hoffman, F. M., Breen, A. L. & Iversen, C. M. Arctic vegetation mapping using unsupervised training datasets and Convolutional Neural Networks. *Remote Sens.* **11**, 69, https://doi.org/10.3390/rs11010069 (2019).
15. Orndahl, K. M. *et al.* Mapping tundra ecosystem plant functional type cover, height, and aboveground biomass in Alaska and northwest Canada using unmanned aerial vehicles. *Arctic Sci.* **8**, 1165–1180, https://doi.org/10.1139/as-2021-0044 (2022).
16. Macander, M. J. *et al.* Lichen cover mapping for caribou ranges in interior Alaska and Yukon. *Environ. Res. Lett.* **15**, 055001, https://doi.org/10.1088/1748-9326/ab6d38 (2020).
17. Thomson, E. R. *et al.* Multiscale mapping of plant functional groups and plant traits in the High Arctic using field spectroscopy, UAV imagery and Sentinel-2A data. *Environ. Res. Lett.* **16**, 055006, https://doi.org/10.1088/1748-9326/abf464 (2021).
18. Yang, D. *et al.* Remote sensing from unoccupied aerial systems: opportunities to enhance Arctic plant ecology in a changing climate. *J. Ecol.* **110**, 2812–2835, https://doi.org/10.1111/1365-2745.13976 (2022).
19. Walker, D.A. & Raynolds, M.K. *An International Arctic Vegetation Database: a foundation for panarctic biodiversity studies.* Report No. 5. (CAFF International Secretariat, 2011).
20. Walker, D. A. *et al.* Rescuing valuable arctic vegetation data for biodiversity models, ecosystem models and a panarctic vegetation classification. *Arctic.* **66**, 133–137, https://doi.org/10.14430/arctic4281 (2013).
21. Zemlianskii, V. *et al.* Russian Arctic Vegetation Archive—A new database of plant community composition and environmental conditions. *Glob. Ecol. Biogeogr.* **32**, 1699–1706, https://doi.org/10.1111/geb.13724 (2023).
22. Zemlianskii, V. *et al.* Russian Arctic Vegetation Archive – a new database of plant community composition and environmental conditions. *Dryad* https://doi.org/10.5061/dryad.5tb2rbp8d (2024).
23. Chytrý, M. *et al.* European Vegetation Archive (EVA): an integrated database of European vegetation plots. *Appl. Veg. Sci.* **19**, 173–180, https://doi.org/10.1111/avsc.12191 (2016).
24. Henry, G. H. R. *et al.* The International Tundra Experiment (ITEX): 30 years of research on tundra ecosystems. *Arct. Sci.* **8**, 550–571, https://doi.org/10.1139/as-2022-0041 (2022).

25. Nawrocki, T. W., Carlson, M. L., Osnas, J. L. D., Trammell, E. J. & Witmer, F. D. W. Regional mapping of species-level continuous foliar cover: beyond categorical vegetation mapping. *Ecol. Appl.* **30**, e02081, https://doi.org/10.1002/eap.2081 (2020).

26. Nawrocki, T. W. *et al.* Checklist of vascular plants, bryophytes, lichens, and lichenicolous fungi of Alaska. *Alaska Vegetation Plots Database (AKVEG)* https://akveg.uaa.alaska.edu/comprehensive-checklist/ (2020).

27. Raynolds, M. K. *et al.* A raster version of the Circumpolar Arctic Vegetation Map (CAVM. *Remote Sens. Environ.* **232**, 111297, https://doi.org/10.1016/j.rse.2019.111297 (2019).

28. Nowacki, G., Spencer, P., Fleming, M., Brock, T. & Jorgenson, T. *Unified Ecoregions of Alaska: 2001.* Open-File Report (U.S. Geological Survey, 2003).

29. U.S. Department of Interior Bureau of Land Management. Assessment, Inventory, and Monitoring (AIM) Terrestrial Indicators Raw Dataset. *BLM Natl AIM TerrADat Hub* https://www.arcgis.com/home/item.html?id=a00461ecbceb4056bc37d6c0dad66a41&sublayer=0 (2023).

30. NEON (National Ecological Observatory Network). Plant presence and percent cover (DP1.10058.001) RELEASE-2023. *NEON Data Portal* https://doi.org/10.48443/9579-a253 (2023).

31. Villarreal, S. *et al.* Tundra vegetation change near Barrow, Alaska (1972–2010). *Environ. Res. Lett.* **7**, 015508, https://doi.org/10.1088/1748-9326/7/1/015508 (2012).

32. Villarreal, S. International Polar Year (IPY) Back To The Future (BTF): Changes in Arctic ecosystem structure over decadal times scales. *Open Access Theses & Dissertations.* 1956 (2013).

33. Sloan, V.L. *et al.* Plant community composition and vegetation height, Barrow, Alaska, Ver. 1. *Next Generation Ecosystem Experiments Arctic Data Collection (NGEE Arctic)* https://doi.org/10.5440/1129476 (2014).

34. Sloan, V. L. *et al.* Soil temperature, soil moisture and thaw depth, Barrow, Alaska, Ver. 1. *Next Generation Ecosystem Experiments Arctic Data Collection (NGEE Arctic)* https://doi.org/10.5440/1121134 (2014).

35. Davidson, S. J. *et al.* Vegetation type dominates the spatial variability in CH_4 emissions across multiple Arctic tundra landscapes. *Ecosystems* **19**, 1116–1132, https://doi.org/10.1007/s10021-016-9991-0 (2016).

36. Walker, D.A. *et al. Infrastructure-thermokarst-soil-vegetation interactions at Lake Colleen Site A, Prudhoe Bay, Alaska.* Alaska Geobotany Center Data Report AGC 15-01 (Institute of Arctic Biology, University of Alaska Fairbanks, 2015).

37. Walker D.A., *et al. Road effects at airport study site Prudhoe Bay, Alaska.* Alaska Geobotany Center Data Report AGC 16-01 (Institute of Arctic Biology, University of Alaska Fairbanks, 2016).

38. Breen, A. *et al.* NGEE Arctic plant traits: plant community composition, Kougarok Road Mile Marker 64, Seward Peninsula, Alaska, 2016. *Next Generation Ecosystem Experiments Arctic Data Collection* https://doi.org/10.5440/1465967 (2020).

39. Wood, T., Essner, R. L. & Minchin, P. R. Effects of prescribed burning on grassland avifauna at riverlands migratory bird sanctuary. *Polymath: An Interdisciplinary Arts and Sciences Journal.* **3**, 19–38 (2013).

40. Cornelissen, J. *et al.* Handbook of protocols for standardised and easy measurement of plant functional traits worldwide. *Aust. J. Bot.* **51**, 335–380, https://doi.org/10.1071/BT02124 (2003).

41. Sulman, B. N. *et al.* Integrating arctic plant functional types in a land surface model using above- and belowground field observations. *J. Adv. Model. Earth Syst.* **13**, e2020MS002396, https://doi.org/10.1029/2020MS002396 (2021).

42. Steckler, M. R. *et al.* The Pan-Arctic Vegetation Cover (PAVC) database. *Next-Generation Ecosystem Experiments (NGEE) Arctic ESS-DIVE repository* https://doi.org/10.15485/2483557 (2024).

43. Velliquette, T. *et al.* ESS-DIVE Reporting Format for File-level Metadata. *Environmental Systems Science Data Infrastructure for a Virtual Ecosystem (ESS-DIVE) Repository* https://doi.org/10.15485/1734840 (2021).

44. Zhang, T., *et al* Mapping Wall-to-Wall Fractional Cover of Arctic Tundra Plant Functional Types in Alaska Using 20-M Spatial Resolution Satellite Imagery and Harmonized Plot Observations. Preprint at https://doi.org/10.2139/ssrn.5206100 (2025).

45. De Stefano, A., Fowers, B. & Mealor, B. A. Comparison of visual estimation and line-point intercept vegetation survey methods on annual grass–invaded rangelands of Wyoming. *Invasive Plant Sci. Manag.* **14**, 240–252, https://doi.org/10.1017/inp.2021.36 (2021).

46. Chytrý, M. & Otýpková, Z. Plot sizes used for phytosociological sampling of European vegetation. *J. Veg. Sci.* **14**, 563–570, https://doi.org/10.1111/j.1654-1103.2003.tb02183.x (2003).

47. Baer, K.C. *et al. Minimum standards for field observation of vegetation and related properties version 1.1.* (Vegetation Technical Working Group, Alaska Geospatial Council, 2022).

48. Coulloudon, B. *et al. Sampling vegetation attributes interagency technical reference.* Technical Reference 1734-4 (Bureau of Land Management's National Applied Resource Sciences Center, 1996).

49. Dong, S., Zhang, Y., Shen, H., Li, S. & Xu, Y. in *Grasslands on the Third Pole of the World.* https://doi.org/10.1007/978-3-031-39485-0 (Springer, 2023).

Acknowledgements

Author contributions

M.R.S., A.L.B., T.Z., J.K. developed the synthesis workflow, created the database and drafted the manuscript. T.W.N., V.G.S., A.L.B., A.D., provided input on the database structure. A.L.B., D.A.W., A.F.W., A.D., T.W.N., S.D.W., M.J.M., G.V.F., V.G.S., D.T.B., C.M.I. provided collected and curated input plot datasets. All authors provided feedback on the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.R.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.