PROJECT NARRATIVE

The Alaska Soil Data Bank (ASDB): A database for delivering non-NRCS legacy data for digital soil mapping initiatives in Alaska

i. Identification of which research priority the proposal is being written to address: This proposal addresses the following research priority: ***Novel techniques to leverage big data for applications in soil survey research.***

ii. Problem statement/Need and justification for project: Identification of which research priority the proposal is being written to address.

Approximately 300 million acres (> 70%) of the NOTCOM areas in the United States are in Alaska. Little data is available in unmapped areas of Alaska in the NASIS database (Figure 1A, Figure 1B). Harnessing disparate non-NRCS legacy datasets for digital soil mapping initiatives in Alaska will therefore be critical for completing the Soils2026 initiative. In order to properly harness big data in soil survey in Alaska, where a significant amount of non-NRCS soil data exists in critical areas, we need to develop a platform that allows public/private data, provides attribution, is flexible and nimble enough to accept ongoing contributions without a large burden on the submitter, can be easily scaled, and can provide export workflows to ingest data into NASIS. The utility of such a database (what we will refer to as the Alaska Soil Data Bank – ASDB) will be demonstrated through the production of digital soil mapping products at multiple scales and using multiple methods. We will use the harmonized ASDB dataset and NASIS together to produce the first digital statewide predictive map of Alaska soil orders and suborders at 30m resolution, and apply a novel approach to digital soil mapping (using landscape segmentation analysis) for Katmai National Park and Preserve.

iii. Expected project deliverables that will contribute to achieving the Soil Survey mission.

This project will have 3 major deliverables, each of which contribute to achieving the Soil Survey mission:

1. *The Alaska Soil Data Bank (ASDB).* We will acquire, curate, and centralize non-NRCS legacy data for the state of Alaska and stand up a database for continuing contributions on the G.E.M.S. platform of the University of Minnesota Supercomputing Institute. The ASDB will result in a harmonized dataset for use in digital soil mapping in Alaska and scripts will be written to export the data into a NASIS-readable format for ingestion.
2. *AK Statewide Predictive Soil Order/Suborder Map AK.* We will generate digital predictive soil order and soil suborder maps of the state of Alaska at a resolution of 30m.
3. *Segmentation Analysis Approach to DSM for Katmai National Park and Preserve.* We will utilize a novel landscape segmentation analysis approach to digital soil mapping for Katmai National Park and Preserve.

iv. A brief review of current literature.

Data compilation efforts in Alaska for digital soil mapping purposes. Over the past 10 years, several data compilation efforts have been attempted for the purposes of mapping soil organic carbon stocks (Bliss and Maursetter, 2010; Mishra and Riley, 2012; Mishra et al., 2017, Mishra et al., 2021), active layer thickness (Pastick et al., 2015; Mishra et al., 2017), and organic material thickness (Wylie et al., 2016) in Alaska. Currently, no attempts to produce a statewide map of Alaska soil classes from a data compilation have been made, with the exception of

SoilsGrids 250m, which utilizes the WoSIS database (Hengl et al., 2017), and performs poorly with soil class prediction in Alaska due to data paucity. To our knowledge, only one digital soil mapping product for soil classes has been produced for a portion of Alaska (Paul, 2018) using data in NASIS, in the Dalton Highway corridor. Until now, the NRCS data (and the NCSS Soil Characterization database subset) has provided the core of all of these efforts and there is significant overlap between existing non-NRCS databases. For example, the World Soil Information Service (WoSIS) soil profiles database contains 2,325 points in Alaska, with 555 points from NCSS soil characterization database and ~1,800 redundant with the International Soil Carbon Network (ISCN) database – Nave et al., 2017).

Practical considerations for data harmonization. Data harmonization efforts must accommodate the needs of human researchers and the computer systems that are increasingly used to search, organize, and analyze large datasets (Wilkinson et al., 2016). Thus, an effective data harmonization effort must be flexible enough to incorporate heterogeneous datasets and also facilitate standardized labelling, attribution, and storage for future discovery and reuse. One effective way to achieve this goal is through the use of controlled vocabularies and domain-specific ontologies to facilitate the linking and integration of diverse datasets using common, standardized terms for metadata on a field by field basis (L'Abate et al., 2015). Using these techniques to create detailed, consistent field metadata has been recognized as the crucial "last mile" in scientific data sharing and interoperability (Shaw et al., 2020).

v. Research objective(s). The objectives of this research are to harness non-NRCS soil data in Alaska to build a large database of point data for digital soil mapping initiatives. From this data compilation effort, in combination with NASIS point data, we will produce a digital soil order and suborder predictive class map of the state at 30m resolution and utilize segmentation analysis to produce a novel delineated raster product for Katmai National Park and Preserve.

vi. Hypothesis to be tested. This research is not explicitly hypothesis driven. However, we explore the following themes: non-NRCS soil point data can be harnessed to address major gaps in soil survey to assist the Soils2026 initiative. That data can be utilized at multiple scales (statewide and regional) to inform digital soil mapping applications using novel techniques.

vii. Research methods and procedures:

Deliverable 1: The Alaska Soil Data Bank (ASDB): A scalable harmonization approach using controlled vocabularies and field metadata.

Why construct another database? **We would like to be very clear: the ASDB is not meant to be a stand-alone replacement for NASIS as the primary storage for non-NRCS Alaska legacy data. Rather, the ASDB would utilize the Minnesota Supercomputing Institute's powerful GEMS platform (described below) to support data acquisition and harmonization of non-NRCS legacy data in Alaska *for eventual automated input into NASIS*. There are numerous advantages to this approach, including post-entry harmonization scripts which allow flexible format data submission and lower barriers to data acquisition both now and in the future.** There are several databases (International Soil Carbon Network (ISCN), World Soil Information Services Soil (WoSIS) Soil Profiles Database, Northern Circumpolar Soil Carbon Database (NCSCD) – Hugelius et al., 2013) and Alaska-relevant data repositories (NSF Arctic Data Center) that contain non-NRCS soil data in Alaska. However, these are not at all comprehensive, cannot deal with both private and public data

(ISCN, WoSIS, NCSCD), do not have the capacity to deal with soil morphological data which is critical for building a comprehensive dataset of point data in Alaska (ISCN, NCSCD), or lack data harmonization tools and the functionality to construct project specific controlled vocabularies for field metadata (All). Separating public and private data is critical for the ASDB, because some of the point data for critical areas comes from private companies who own the data and therefore, the entire dataset (or collection of datasets) cannot be completely open. Therefore, a novel data platform for the ASDB is necessary. *Crucially, the ASDB will be highly complementary to and not duplicative of ongoing NRCS Initial Mapping and Database Focus team efforts related to point data curation. We will work closely with NRCS staff, including Cory Cole (AK State Soil Scientist), Jessica Jobe (AK Regional Soil Survey Director), and Dave White (DSM Specialist and Alaska 2026 Taskforce Lead - see attached letter of collaboration in Background Information). These efforts will significantly expand point observations available for DSM in Alaska, particularly in areas which are data-sparse and unlikely to be visited by NRCS field staff due to high costs and logistical limitations (Figure 1).*

Database platform selection: MSI-GEMS. A database platform to house non-NRCS legacy soil data for digital soil mapping initiatives in Alaska must be able to separate a) both public and private data, b) provide a stable, continuously maintained platform, and c) host a user interface, d) support controlled vocabularies for field metadata, data harmonization and cleaning tools, and e) support back end export scripting. The University of Minnesota Supercomputing Institute (MSI) Genetic, Environmental, Management, and Socioeconomic (G.E.M.S®) data platform meets all of these characteristics (see Facilities and Equipment and Data Management Plan and Data Sharing and Usage Policy for more details). The GEMS platform is stable and receives long-term sustained funding from the Minnesota State Government.

GEMS approach to data harmonization. The challenge to all data harmonization approaches is to take source datasets in different formats and produce target data in a common format (Vellino, 2015). There are essentially two different approaches to data harmonization. The first is the template approach, which is time intensive, has a high barrier to submission, and requires the user or compiler to translate all data sources into a common template. Each data source must be faithfully transcribed into the template and field names in the original data often discarded and replaced with field names in the template. While very effective and easy to use on the output side, this represents a significant burden to the data contributor. For these reasons, the template approach leads to lower levels of sustained submission, and makes it very difficult to scale the database. An alternative approach to data harmonization utilizes controlled vocabularies for field metadata (L'Abate et al., 2015). Essentially, each field is tagged with controlled vocabulary terms, which can be referenced by custom scripts to generate data tables and harmonize the dataset. The GEMS platform supports large controlled vocabularies, has excellent functionality for field metadata tags during the data contribution process, and additionally uses fuzzy logic to assist contributors in identifying the most likely metadata tags for field metadata.

Ongoing Legacy Data Acquisition and Centralization. The acquisition of novel soil point data not contained in established non-NRCS databases (WoSIS, NASIS, ISCN) represents a major opportunity to expand the available data for digital soil mapping efforts in Alaska. This legacy data ranges in scope from point observations of permafrost depth or organic layer thickness to full morphological descriptions and a wide range of soil physical and chemical properties. There are a large number of sources that this data can come from for Alaska. A non-exhaustive list of example datasets includes academic sources (University of Alaska Fairbanks; University of Minnesota; Woods Hole Research Center), the National Park Service Arctic

Research Coordination Network (NPS-ARCN), the U.S. Army Corps of Engineers Cold Regions Research and Engineering Laboratory (CRREL) and the BLM-AIM database. In addition, Alaska Biological Research (ABR, Inc., Anchorage, AK) is an important partner on this proposal and will compile legacy soils data collected by ABR in Alaska from 1995–2021 into a PostgreSQL database for delivery and ingestion into the Alaska Soil Data Bank. The ABR data represent a significant portion of the non-NRCS soil data acquired to date for Alaska, particularly in key remote areas that have no coverage from other available datasets, which is the reason for ABR's inclusion in this proposal as a contract for professional services from the lead institution (UMN) (Figure 1C).

Additional Legacy Data Acquisition Opportunities. Additional data acquisition opportunities include data from the peer-reviewed and grey literature, US Geological Survey (USGS), US Department of Energy (DOE), Bureau of Land Management (BLM), and the US Forest Service (USFS). PI Jelinski has initiated conversations with representatives from several of these agencies to begin discussions of data acquisition. As a part of this project we will also undertake an extensive search of the grey literature and peer-reviewed literature, as well as mining descriptive pedon data from unpublished legacy soil surveys throughout the state. Thus, we believe that there is a significant potential to expand the data already acquired beyond the example datasets shown in Figure 1C.

Data Centralization and Ingestion. During the data ingestion process a dataset can be categorized as private (where the geographic scope of the dataset and metadata is visible to public users, but the data itself is available only by request) or into GEMS-Open, which allows the data to be explored and downloaded by public users. All datasets contributed to GEMS-Open receive a unique digital object identifier (DOI). See Data Management Plan and Data Sharing and Usage Policy section for specifics on data sharing philosophy, term definitions, user roles, and licenses.

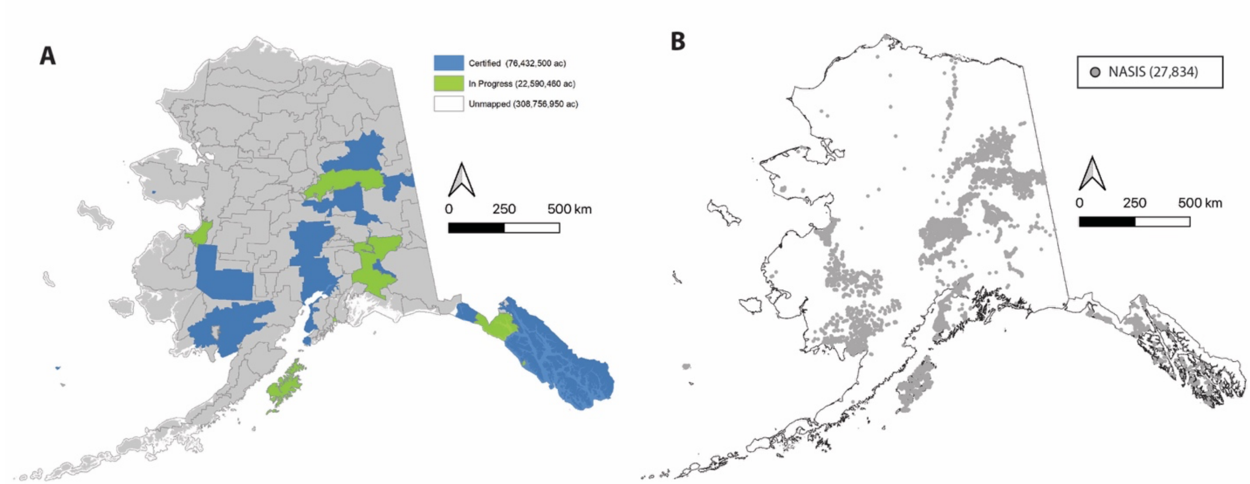
The GEMS platform supports controlled vocabularies, which can be customized and project specific. Controlled vocabularies for field metadata tags will be built from standard USDA-NRCS vocabularies (Field Manual for Describing and Sampling Soils, Keys to Soil Taxonomy 12th Ed., KSSL Laboratory Methods Manual) and crosswalks of terms (schema matching) with WoSIS and ISCN will be developed to facilitate later data export into these databases. NRCS standardized terminologies serve as the primary source of controlled vocabularies because NASIS is the main target of export and integration of the harmonized dataset. Novel methodology codes and controlled vocabularies will need to be developed for site level data of importance to Alaskan soils, such as organic layer thickness (Wylie et al., 2016), depth to permafrost/active layer thickness (Pastick et al., 2015), and for harmonizing different eras of morphological codes and classification schemes.

Data Quality and Enhancement During Staging. During the data staging process, flags are applied to fields that do not use controlled vocabularies, which will allow our team to expand the scope of our controlled vocabulary where necessary. Additionally, we will develop and document a minimum of four metrics of data quality and completeness rankings:

- *Geospatial Data Quality*. The quality of geospatial locations will be identified by the approximate accuracy of the profile location (following a numerical code of decimal place accuracy from Batjes et al. (2019)). The method of geospatial data collection will also be identified (aerial photo mark-up, GPS (with associated uncertainty), or inferred from text location description) as described in previous projects by co-PI Grunwald for the southeastern U.S. (Florida Soil Characterization Dataset – metadata).

Analytical Data Robustness. For datasets including measured values (e.g., pH, carbon, particle size), measurement units (including unknown) will be identified using coding conventions from the controlled vocabulary. These will be attached to each dataset allowing users to filter data based on the method used to measure each soil property.

Data Completeness Rankings. We will develop standardized data completeness rankings which will be to allow users to filter data based on completeness. Point observations will be tagged with these completeness rankings.



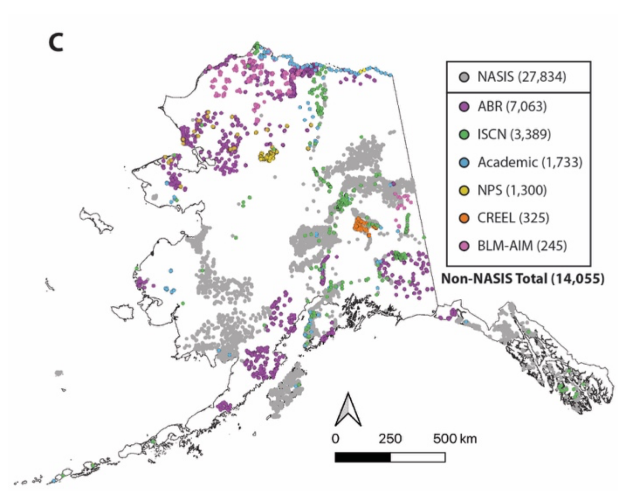


Figure 1. Figure 1. a) Status of Soil Survey in Alaska (Completed, In Progress, and NOTCOM) - June, 2020; b) NASIS data coverage in Alaska - April, 2019; c) Spatial distribution of potential non-NRCS point data acquisition.

Soil Classification Rules and Classification Uncertainty. Because one of the deliverables of this project is the development of a digital map of predictive soil classes across the state of Alaska, we will attempt to gap fill soil classification where possible through a set of documented rules. For example, if a datapoint does not have complete data, but has organic materials (or an SOC concentration of > 18%) to a depth of > 40cm and permafrost within 1m of the soil surface, the soil would unambiguously be classified as a Histel. The level to which point observations without native classification data can be classified will depend on the completeness of the data and the level of classification. If datapoints already contain classification information, the date/year of collection will be used to flag datapoints for which classification may need to be reviewed based on outdated classification rules. Finally, a classification uncertainty ranking will be developed and applied, to allow users to filter datapoints with a high degree of certainty in classification (where morphological and laboratory data available), from those with lower degrees of classification uncertainty (full or partial morphological data only).

Data Harmonization. Scripts to generate novel, harmonized data tables (study, site, horizon) from controlled vocabularies in field metadata will be constructed and made publicly available. The structure of these fully harmonized data tables will utilize the NASIS import data structure as a foundation to ensure the harmonized data can be ingested into NASIS. We will not be developing or applying gap filling or pedotransfer functions in the harmonization process.

Data export and interoperability with other databases. The PI and Co-PIs, in consultation with NRCS staff (Soil Survey Database Team, Initial Mapping Team), will construct export scripts to enable delivery of all harmonized ASDB data to NASIS. GEMS-Open data in the ASDB will be available to all users, agencies and organizations to download as they wish. Metadata for each ASDB dataset will be provided to the NRCS NOTCOM data catalog.

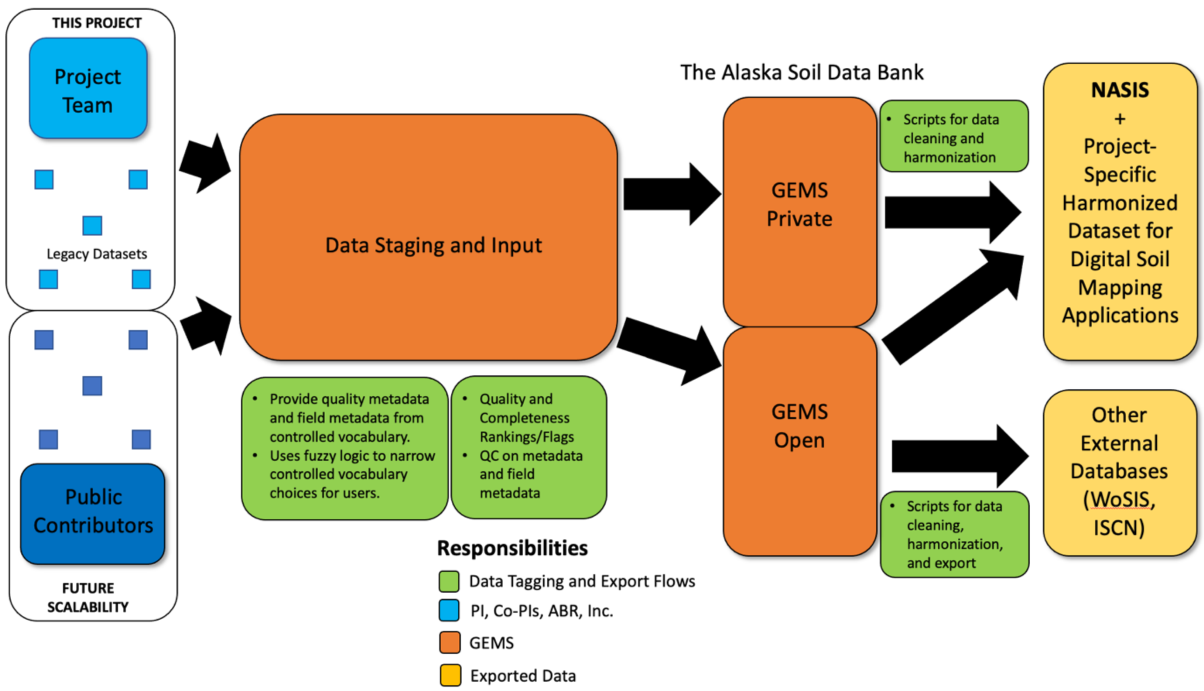


Figure 2. Scheme and Responsibilities for Alaska Soil Data Bank Data Acquisition, Centralization, Staging, Storage and Export

Scalability. Because developing the ASDB on the GEMS platform will involve the construction of controlled vocabularies and crosswalks with existing databases, the bank is readily scalable. Co-PI Grunwald will test platform scalability and export capabilities early in Project Year 1 using existing pre-consolidated legacy data (from the southeastern U.S.) as an example dataset.

<u>Deliverable 2:</u> Produce a 30m Alaska state-wide digital soil class map (Soil order and suborder).

Our approach to generating a 30m statewide predictive map of soil classes for AK based on the harmonized Alaska Soil Data Bank and NASIS point data will follow that of Ramcharan et al. (2018). Briefly, we will utilize a random forest method (Breiman, 2001) in the ranger package for R. The PI (Jelinski) has curated a stack of covariates and their derivatives at a resolution of 30m for the state of Alaska. This covariate stack includes 1) a 30m DEM and ten derived terrain attributes, 2) PRISM climate covariates (30-year normal) and 19 derived bioclimate indices, 3) Landsat 30m (cloud free mosaic) and 13 derivatives. Optimal covariate sets will be selected by recursive feature elimination (Brungard et al, 2015) or the Boruta algorithm to filter relevant variables before running Random Forest models (Xiong et al., 2014). The overall density of the statewide data already acquired is ~1 observation per 41 km^2. However, given that classification may not be possible for all point observations, and spatial

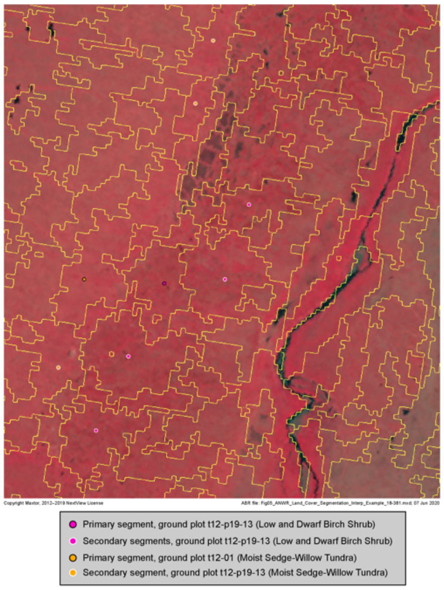


Figure 3. Example of image segmentation results with field plots, and primary and secondary segment interpretation points. Figure by ABR, Inc.--Environmental Research and Services from the land cover mapping project for the North Slope of the Arctic National Wildlife Refuge, Alaska, 2019. Approximate map scale 1:3,000.

distribution is highly uneven (Figure 1), it is likely that large spatial differences in uncertainties will exist. Out of bag (OOB) accuracy for a random forest approach to soil class mapping along the Dalton Highway Corridor in northern Alaska (Paul, 2018), showed the OOB accuracy was greatest for the Soil Order (69%) and Suborder (60%) levels, with significantly lower Great Group OOB accuracy (47%). Given that it is likely that few non-NRCS data points will be able to be classified to Great Group, and the OOB accuracies are likely to be < 50% (as data distribution is expected to be even lower than in Paul (2018)), only predictive class maps at the order and suborder levels will be attempted at a statewide extent.

Deliverable 3: Produce an innovative raster-based, delineated soil map by segmentation analysis for Katmai National Park and Preserve.

Modern high-resolution (≤ 2-m pixel) satellite imagery mosaics, combined with the computing power and analytical tools available in Google Earth Engine (GEE) (Gorelick et al. 2017), and a team of skilled satellite imagery interpreters, make it possible to classify and map broad areas rapidly and at a relatively fine scale (1:10,000–1:24,000) previously limited to hand-digitized maps. We will use an existing normalized, high-resolution (2-m pixel) satellite imagery mosaic prepared by ABR for Katmai National Park and Preserve (KATM) to prepare a soil map using existing field data. To prepare the map, we will perform a segmentation analysis on the high-resolution imagery using Simple Non-Iterative Clustering (SNIC) (Achanta and Susstrunk, 2017) in GEE and develop soil map-unit concepts by classifying existing field data using multivariate statistical analysis techniques in R. ABR's team of skilled imagery interpreters will (1) select and annotate training segments representative of the existing field plots and (2) assign soil map-unit and component classes to a stratified, random sample of segments.

The training data from KATM curated and prepared by ABR in the first project year as part of the legacy data compilation effort (Deliverable 1) includes approximately 407 points along 67 transects in KATM. This data will be used to classify the segments from imagery analysis to soil map-unit components using random forest models (Breiman 2001) utilizing spectral metrics from high-resolution imagery and Landsat data, as well as ancillary data such as elevation and climate metrics. The end product will be a soil map equivalent to a traditional hand-delineated NRCS vector-based soil map, at the scale of an order 2 soil survey (1:24,000) but prepared using a digital soil mapping approach for an approximately 4 million acre study area. The final map product will have attributes of both vector- and raster-based soil maps. The map product resulting from this pilot study will represent a significant advance in the rate of completion and spatial resolution of soil mapping for an area of this size and complexity. ABR will work with UMN, NMSU, and UFL to co-author a manuscript for peer-review.

viii. Project timetable. See Table 1.

ix. Information transfer plan. The results of this work will be distributed in several ways: 1) project reporting to NRCS, 2) a minimum of three peer-reviewed publications (1 on each deliverable, 3) a minimum of three presentations at scientific meetings (Soil Science Society of America and AGU) and three webinars (1 on each deliverable). All project documentation will be archived using a GitHub page as a repository. All harmonized data will be transferred to NASIS.

x. Documentation of the involvement of and training provided to graduate and/or undergraduate students and to NRCS field scientists. This project will involve a graduate student (1 full-time UMN). This student will be trained on data management, harmonization scripting, and digital soil mapping procedures by the PI and Co-PIs in consultation with NRCS staff. This project will involve NRCS field soil scientists in several steps: 1) subjective evaluation of digital soil mapping products following the procedures of Paul (2018) will occur with field soil scientists in Alaska, coordinated with Jessica Jobe (Regional Director of Soil Survey for Alaska); 2) harmonization scripts for export of data into NASIS will be closely coordinated with the Soil Survey Database Team and Alaska DSM Team; 3) Additionally, 3 webinars will be delivered in various working groups (Initial Mapping Focus Group – Deliverable 1), Digital Soil Mapping Focus Group, Practitioners Discussion (Deliverables 2 and 3).

xi. References

Achanta, R. and Susstrunk, S., 2017. Superpixels and polygons using simple non-iterative clustering. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 4651-4660).

Batjes, N.H., Ribeiro, E., van Oostrum, A., 2019. Standardised soil profile data for the world (WoSIS Snapshot) https://dx.doi.org/10.17027/isric-wdcsoils.20190901

Bliss, N.B. and Maursetter, J. (2010), Soil Organic Carbon Stocks in Alaska Estimated with Spatial and Pedon Data. Soil Sci. Soc. Am. J., 74: 565-579. doi:10.2136/sssaj2008.0404

Breiman, L. 2001. Random forests. Machine Learning, 45:5–32.

Brungard, C.W., Boettinger, J.L., Duniway, M.C., Wills, S.A., Edwards Jr., T.C. 2015. Machine learning for predicting soil classes in three arid landscapes. Geoderma, 239-240, 68-83.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D. and Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote sensing of Environment, 202, pp.18-27.

Hengl T, Mendes de Jesus J, Heuvelink GBM, Ruiperez Gonzalez M, Kilibarda M, et al. (2017) SoilGrids250m: Global gridded soil information based on machine learning. PLOS ONE 12(2): e0169748. https://doi.org/10.1371/journal.pone.0169748

Hugelius G., Bockheim J.G., Camill P., Elberling B., Grosse G., Harden J.W., Johnson K., Jorgenson T., Koven C.D., Kuhry P., Michaelson G., Mishra U., Palmtag J., Ping C.-L., O'Donnell J., Schirrmeister L., Schuur E.A.G., Sheng Y., Smith L.C., Strauss J. and Yu Z. *Earth System Science Data*, Vol 5, pp. 393–402 (2013). DOI:10.5194/essd-5-393-2013

L'Abate, G., Caracciolo, C., Pesce, V., Geser, G., Protonotarios, V., Costantini, E.A.C. 2015. Exposing vocabularies for soil as Linked Open Data. Information Processing in Agriculture, 2 (3-4): 208-216

Michaelson, G., Ping, C.-L., Clark, M. 2013. Soil Pedon Carbon and Nitrogen Data for Alaska: An Analysis and Update. *Open Journal of Soil Science*, Vol. 3 No. 2, 2013, pp. 132-142. doi: 10.4236/ojss.2013.32015.

Mishra, U. and Riley, W. J. 2012. Alaskan soil carbon stocks: spatial variability and dependence on environmental factors, Biogeosciences, 9, 3637–3645, https://doi.org/10.5194/bg-9-3637-2012.

Mishra, U., Drewniak, B., Jastrow, J.D., Matamala, R.M., Vitharana, U.W.A. 2017. Spatial representation of organic carbon and active-layer thickness of high latitude soils in CMIP5 earth system models. Geoderma, 300, 55-63, https://doi.org/10.1016/j.geoderma.2016.04.017.

Mishra U, Hugelius G, Shelef E, et al. 2021. Spatial heterogeneity and environmental predictors of permafrost region soil organic carbon stocks. Science Advances 7(9):eaaz5236. Published 2021 Feb 24. doi:10.1126/sciadv.aaz5236

Nave L, Johnson K, van Ingen C, Agarwal D, Humphrey M, Beekwilder N. (2017). International Soil Carbon Network (ISCN) Database, Version 3. DOI: 10.17040/ISCN/1305039. Custom download from Map-based Data Access Interface. Accessed 5 June 2020

Pastick, N. J., M. T. Jorgenson, B. K. Wylie, S. J. Nield, K. D. Johnson, and A. O. Finley (2015), Distribution of near-surface permafrost in Alaska: Estimates of present and future conditions, Remote Sens. Environ., 168, 301–315, doi:10.1016/j.rse.2015.07.019.

Paul, J. 2018. Multiresolution digital soil mapping of permafrost soils using a random forest classifier: An investigation along the Dalton Highway corridor, Alaska, M.S. Thesis, University of Alaska-Fairbanks.

Ramcharan, A., Hengl, T., Nauman, T., Brungard, C., Waltman, S., Wills, S. and Thompson, J. (2018), Soil Property and Class Maps of the Conterminous United States at 100-Meter Spatial Resolution. Soil Science Society of America Journal, 82: 186-201. doi:10.2136/sssaj2017.04.0122

Ribeiro E, Batjes NH and van Oostrum AJM, 2018. World Soil Information Service (WoSIS)-Towards the standardization and harmonization of world soil data. Procedures manual 2018, Report 2018/01, ISRIC - World Soil Information, Wageningen. doi:10.17027/isric-wdcsoils.20180001.

Shaw, F., Etuk, A., Minotto, A., Gonzalez-Beltran, A., Johnson, D., Rocca-Serra, P., … Davey, R. P. 2020. COPO: a metadata platform for brokering FAIR data in the life sciences [version 1; peer review: awaiting peer review]. F1000Research, 9(495).

Xiong X., S. Grunwald, D.B. Myers, J. Kim, W.G. Harris and N.B. Comerford. 2014. Holistic environmental soil-landscape modeling of soil organic carbon. Environmental Modeling and Software J. 57: 202-215.

Vellino, A. 2015. Harmonizing the metadata among diverse climate change datasets. International Journal of Data Curation 10(1): 268-279.

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3(1), 160018. https://doi.org/10.1038/sdata.2016.18

Wylie, B.K., N.J. Pastick, K.D. Johnson, N. Bliss, H. Genet. 2016. Soil carbon and permafrost estimates and susceptibility to climate change in Alaska. Chapter 3(pp. 53-76) In: Zhu, Z., and McGuire, A.D. (Eds.). Baseline and Projected Future Carbon Storage and Greenhouse-Gas Fluxes in Ecosystems of Alaska. US Geological Survey Professional Paper 1826, U.S. Geological Survey, Reston, Virginia.

Table 1. Timetable of Project Milestones and Deliverables		Year 1 (OCT 2022- SEP 2023)		Year 2 (OCT 2023 – SEP2024)		
Activity	Responsible Project Personnel	QTR 1-2	QTR 3-4	QTR 1-2	QTR 3-4	Deliverables (D)
0.0 Review of Project Milestones and Deliverables	Lead: Jelinski					
Task 1: Alaska Soil Data Bank						
1.1 Data Acquisition	Lead: Jelinski, Graduate Student Support: Brungard, Grunwald, ABR					(1)Construction of Alaska Soil Data Bank on UMN MSI GEMS data platform.
1.2. Data Centralization	Lead: Jelinski, Graduate Student					
1.3. Data Quality and Enhancement	Lead: Jelinski, Graduate Student					
1.4. Data Harmonization	Lead: Jelinski, Graduate Student					
1.5. Data Export and Integration with Other Databases	Lead: Jelinski, Graduate Student					
Task 2: Generate Statewide 30m Predictive Digital Soil Mapping Product						
2.1. Covariate Curation & Script Development	Lead: Jelinski Support: Brungard					(2) Statewide 30m digital soil class (order and suborder) map
2.2. Predictive Model Runs	Lead: Brungard, Grunwald					
Task 3: Generate Digital Soil Mapping Product for Katmai National Park and Preserve						
3.1. Segmentation Analysis	Lead: ABR					(3) Digital Soil Map of Katmai Natl Park & Preserve
3.2. Predictive Model Runs and Segment Classification	Lead: ABR					
Task 4: Reporting and Project Closeout						
4.1. Scientific Communication of Results	Lead: Jelinski, Graduate Student Support: Grunwald, Brungard, ABR					
4.2. Final Reporting and Project Closeout	Lead: Jelinski Support: Grunwald, Brungard, ABR					