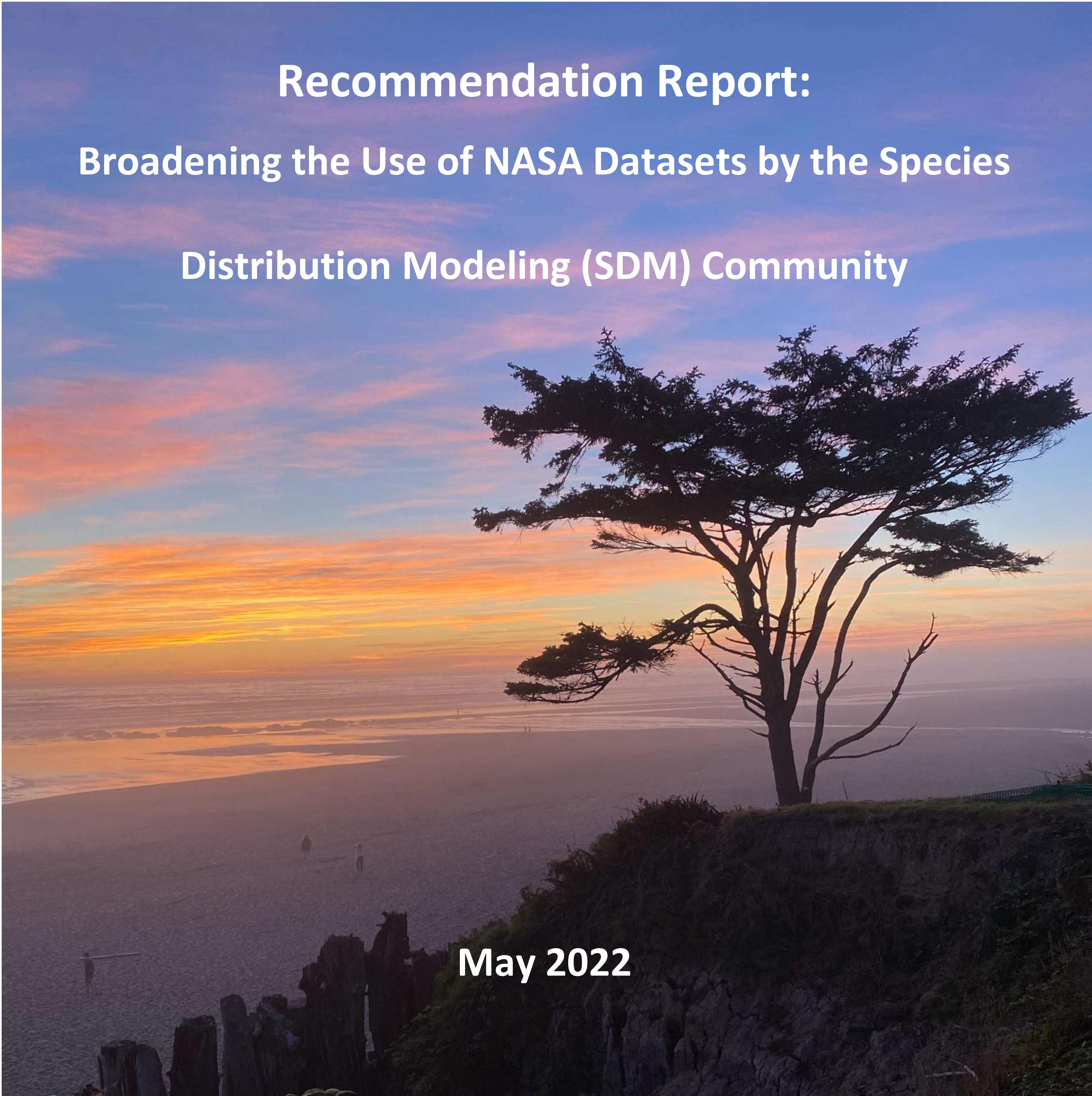


Recommendation Report:

Broadening the Use of NASA Datasets by the Species Distribution Modeling (SDM) Community

May 2022

Prepared by the Spatial Analysis Lab at the University of Montana and SDM Workshop Participants

Workshop made possible by the NASA Understanding User Needs to Broaden Outside Use of NASA Data (UNBOUND) Program under a proposal submitted to the Topical Workshops, Symposia, and Conferences (TWSC) element of the Research Opportunities in Space and Earth Sciences (ROSES) Solicitation (21-TWSC21-0005; Grant No. 80NSSC21K1984)

Recommendation Report: Broadening the use of NASA datasets by the Species Distribution Modeling (SDM) community

May 18th 2022

Prepared by the Spatial Analysis Lab at the University of Montana and SDM workshop participants

Spatial Analysis Lab
Melissa Hart (melissa.hart@mso.umt.edu)
Marie Johnson (marie3.johnson@umconnect.umt.edu)
Jessica Mitchell (jessica.mitchell@mso.umt.edu)
Ryan Rock (ryan.rock@mso.umt.edu)
Ethan Shafron (ethan.shafron@mso.umt.edu)
Claudine Tobalske (claudine.tobalske@mso.umt.edu)

SDM Workshop Contributors
Sudeep Banad, Indian Institute of Technology Delhi (qiz188545@iitd.ac.in)
Mary Blair, American Museum of Natural History (mblair1@amnh.org)
Israel Borokini, University of California, Berkeley (iborokini@berkeley.edu)
Natalie Burgos, Chicago Region Trees Initiative - Morton Arboretum (nburgos@luc.edu)
Patrick Burns, Northern Arizona University (pb463@nau.edu)
Nathalie Chardon, University of British Columbia (nathalie.chardon@gmail.com)
Paige Copenhaver-Parry, Wyoming Natural Diversity Database (pcopenha@uwyo.edu)
Kelly Easterday, The Nature Conservancy (kelly.easterday@tnc.org)
Xiao Feng, Florida State University (fengxiao.sci@gmail.com)
Matt Fitzpatrick, University of Maryland (mfitzpatrick@umces.edu)
Beth Gerstner, Michigan State University (gerstn11@msu.edu)
Reza Goljani-Amirkhiz, University of South Dakota (Reza.GoljaniAmirkhiz@coyotes.usd.edu)
Chris Hakkenberg, Northern Arizona University (chris.hakkenberg@nau.edu)
Tim Howard, New York Natural Heritage Program (tim.howard@dec.ny.gov)
Patrick Jantz, Northern Arizona University (Patrick.Jantz@nau.edu)
Abhishek Kumar, Panjab University, Chandigarh (abhikumar.pu@gmail.com)
Zhihua Liu, University of Montana (zhihua.liu@umontana.edu)
Ellen C. Martin, Norwegian University of Science and Technology (ecmartin33@gmail.com)
Cory Merow, University of Connecticut (cory.merow@gmail.com)
Eric Nielsen, Oregon State University, Institute for Natural Resources (emn2@pdx.edu)
Giovanni Rapacciuolo, NatureServe (gio@natureserve.org)
Sangeeta Roy, Visva-Bharati University (sangeetaroy.rs@visva-bharati.ac.in)
Scott Rush, Mississippi State University (scott.rush@msstate.edu)
Leah Schwizer, Booz Allen Hamilton (leah.schwizer@nasa.gov)
Adam B. Smith, Missouri Botanical Garden (adam.smith@mobot.org)
Ho Yi Wan, California State Polytechnic University Humboldt (hw83@humboldt.edu)
Lu Zhai, Oklahoma State University (zhailulu@gmail.com)

Contents

List of figures

List of tables

EXECUTIVE SUMMARY

In January and February of 2022, University of Montana's Spatial Analysis Lab hosted a virtual workshop on broadening use of NASA datasets by the Species Distribution Modeling (SDM) community. The workshop was made possible by the NASA Understanding User Needs to Broaden Outside Use of NASA Data (UNBOUND) Program under a proposal submitted to the Topical Workshops, Symposia, and Conferences (TWSC) element of the Research Opportunities in Space and Earth Sciences (ROSES) Solicitation (21-TWSC21-0005; Grant No. 80NSSC21K1984).

As part of the SDM workshop effort, five subgroups (Vegetation, Surface Water/Marine, Edaphic, Climate, and Human Disturbance) made general recommendations for improving access and usability by SDM practitioners. In many cases, workshop participants and subgroup teams explored NASA datasets in detail and provided product-specific recommendations detailed in this report (see Table 1 for a selected summary of recommendations). Four participants indicated interest in testing and reporting the sensitivity of Global Ecosystem Dynamics Investigation (GEDI) Lidar products on SDM model performance for a single species; the sensitivity of coarse and downscaled (if feasible) nighttime light data on SDM model performance for one to several species; and the sensitivity of SDMs for niche plants to microtopographic variables available through the Commercial Satellite Data Acquisition Program (2-meter Digital Surface Models from EarthDEM).

In general the data in Earthdata need another layer of comprehensive and systematic curation, not just the extensive but disparate metadata exposure. A major workshop outcome is the recommendation that NASA continue working with the SDM community to prepare a technical guidebook that helps users navigate through the troves of data that are available, identify and locate relevant layers, and interpret which datasets are appropriate for different types of SDMs (e.g., terrestrial, aquatic, intertidal, marine, niche species). A user's guidebook should provide high-level information on what the datasets represent; address tradeoffs between datasets to help users select appropriate inputs, and provide options for a tutorial when accessing a product. One option for supporting an initiative to standardize, centralize, and improve reproducibility of SDM-relevant datasets is expanding the Species Distribution Modeling Pathfinder to include product descriptions with information on the applicability of use in SDMs. Additional considerations include centralized access to relevant datasets in AppEEARS! or a similar system, and adding tags and/or additional metadata to relevant data products (see Edaphic Variables section of the report). This recommendation evolved from an earlier suggestion to provide a raster stack of available datasets in R-compatible format such as TIFF, with user-defined layer download selection or to otherwise generate a use case where species of interest have been targeted for modeling (e.g. coldwater fish). While it is critical to provide SDM practitioners with mechanisms for identifying relevant datasets and easy access to these layers, this "stack" recommendation was dismissed on account of maintenance burdens and potential for a stack to generate bias in SDM models.

General recommendations that emerged from the workshop are summarized below and loosely prioritized based on level of undertaking:

- NASA data products would be more widely used if available in a centralized location, with standardization along a few common dimensions, such as plain language descriptions and references to appropriate uses in different Species Distribution Models and applications (i.e. a user's guidebook with a glossary of remote sensing terms).
- Consider making changes to NASA datasets that a large percent of practitioners would benefit from by establishing Worldclim/ CHELSA as a baseline to which supplemental NASA data could be added.

    - Enable customized spatial and temporal subsetting and aggregation before downloading or otherwise promote the availability of wall-to-wall continuous datasets

- Datasets are too frequent and large in volume. Use a descriptive, easy-to-follow file and data set naming convention, with easy-to-digest metadata/standard descriptions in a format consistent across products.
    - Enable EarthData searches that can be filtered by spatial resolution and tags in order to help find or discover the best data for a given study area.
    - A meta-literature search can be conducted by a few PhD students to determine different tag categories (see Zellweger et al. 2019).

- It would be useful to have suggestions provided within the dataset for sensible pixel quality/QAQC thresholds to grab the best data for SDM applications.

- If any resources are available to generate additional higher level products that would support the SDM community, the workshop group expressed interest in additional GEDI L3 gridded products (i.e., Plant Area Index, canopy cover, and foliage height diversity) in addition to canopy height and ground elevation. At minimum, the community of practice could establish standardized guidelines and code for generating additional gridded products. Participants also expressed need for annual canopy cover maps (similar to MODIS Vegetation Continuous Field (VCF) products) at 30 m resolution.

- Citizen science data are expanding but are frequently disconnected from Earth observation datasets. Link citizen science data with NASA products in data portals, with a possible note or filter to indicate what data were collected through citizen science so that a user may be able to select to include that data or not.

- Publish datasets in consistent and common file formats, such as TIFF and when this is not feasible, provide dataset specific resources for converting data into a more useable format.

- Provide access to datasets through commonly used software such as R, ESRI Basemaps, or Google Earth Engine. Environmental datasets gain in popularity if they can directly be downloaded and processed from R or if compatibility with R is somehow facilitated.

- Publish large datasets in a tiled format to speed up downloading time.

- In cases where tutorials are available for Python users, but not published with instruction using R, translation would broaden use by SDM practitioners, including use in the classroom as lab materials.

- In some cases, modeled products (e.g., MODIS EVI) are discontinued and methods cannot easily be replicated using more recent data. Jarnevich et al. (2021) discusses this as one of many challenges in seamlessly iterating SDMs that require ‘original’ models be continually refit for model comparisons.

- Improve ease of access and reduce confusion by creating clean websites that are easy to use (e.g., worldclim.org), both for downloading data and accessing specifications; promote datasets with Earthdata links on sites outside of NASA.

- Improve product awareness by continuing to clearly feature links to specific NASA datasets on major sites and in newsletters; consider additional exposure through links on external software websites that are willing to do the same and create a link or page specifically for NASA datasets.

Table 1 Selected product-specific recommendations made by workshop groups.

DATASET	ACCESS	FORMAT	CLARITY	TEMPORAL SCALE	ACCESS
Vegetation - GEDI	Centralize GEDI data so they can be accessed from the same DAAC	Aggregate point and polygon data (where feasible) to raster grids of manageable size (e.g., 5 by 5 degree), store on a server allowing for fast downloads			GEE, ORNL, AppEARS!
Surface water/ Marine - MUR SST	R packages needed for direct access			Compilation of daily datasets into monthly, seasonal, or yearly averages	PODAAC, Earthdata
Climate - Ecostress	Add a download link to Earthdata	Package in a more user friendly format (not H5)	Add a description page including a summary of dataset attributes, definitions of acronyms, specific applications for each dataset	Compilation of daily datasets into monthly, seasonal, or yearly averages	AppEARS, EarthExplorer, Earthdata
Climate - MERRA2 precipitation and monthly extremes detection indices	Simplify download process (no redirecting to Earthdata login; provide a unique step to download data directly onto computer)	Package in a more user friendly format (not NetCDF)	List the data as a separate column, e.g., Monthly Mean, Monthly Percentiles, Monthly Aggregated, Hourly Time-Averaged, etc. Standardize initial file naming to be comprehensible, clean, and concise		Earthdata, Giovanni
Climate - MODIS snow cover			Investigate the cause of differences in the NDSI values between the Terra and the Aqua layers for data even from the same day, along with some potential spatial misalignments		GEE, AppEARS!
Human Dimensions - Black Marble		Package in a more user-friendly format (not H5) or link to post-processing options available through LAADS DAAC.	Provide a quick description of each product in the table on the landing page (vs a moving text)	Compilation of daily datasets into monthly, seasonal, or yearly averages	Earthdata, WorldView

INTRODUCTION

This Recommendation Report is the outcome of a virtual workshop convened by the Spatial Analysis Lab (SAL) at the University of Montana in January and February 2022 on broadening use of NASA datasets by the Species Distribution Modeling (SDM) community, including agency land managers, academic and governmental scientists, and non-governmental agencies. Throughout the workshop sessions, practitioners with expertise in ecology, remote sensing, and modeling worked together to identify current and upcoming NASA missions with high potential for the emergence of novel predictor variables and improved SDM performance (He et al., 2015). The workshop focused on disseminating science and resultant data products relevant to SDM practitioners, including how to make products more accessible and useful to end users. We define here the scope of SDMs as models that quantify relationships between environmental factors (e.g., continuous raster datasets) and the spatial distribution of plant and animal species (e.g., point observations of occurrence). Model results include empirically derived measures of the importance of environmental factors in predicting species' distributions across unsampled areas. A key SDM outcome is improved understanding of ecological consequences of interactions between organisms and changing environmental conditions (e.g., Elith and Leathwick, 2009; Miller, 2010), conditions which are represented in the diverse array of NASA data products. The recommendations in this report are thus highly relevant to related modeling efforts and applications, such as habitat suitability, species diversity, ecological integrity, disease and invasive species potential, and conservation planning (e.g., Bradley et al., 2012; Elith, 2017; Jetz and Rahbek, 2002; Porfirio et al., 2014; Rogers, 2006; and Schmitt et al., 2017).

Key attributes of environmental datasets widely used in SDM research

To help understand the needs of the SDM community with respect to environmental data and, thus the perspective of many workshop participants, it is useful to highlight several key characteristics of environmental datasets widely used in species distribution modeling research (reviewed by Peterson et al. 2011. Ecological Niches and Geographic Distributions. Monographs in Population Biology, Princeton University Press):

- Format: The vast majority of SDM software require environmental predictor data in raster GIS format (i.e. cell- or pixel-based files where each cell is assigned a value for the corresponding continuous or categorical environmental variable). Ideally, all cells have a value but NA cells are tolerated. Data in other formats such as points, polygons or lines must be converted to raster grids by the user, with potential loss of information.
- Cartographic system: Environmental predictor data must match the cartographic system (e.g., geographic projection, datum, coordinate system, units) of the corresponding occurrence data. Widely used environmental datasets generally do not require much additional processing to be matched with widely used species occurrence datasets, which frequently are provided as geodetic datum WGS84.
- Spatial extent: Widely used environmental datasets have a broad spatial extent, ideally global or continental, unconstrained by jurisdictional boundaries.
- Spatial resolution: Widely used environmental datasets have high spatial resolutions in proportion to the spatial extent.
- Temporal extent: Widely used environmental datasets have a temporal extent that matches with widely used species occurrence datasets, often stretching back to 1970-1980.
- Temporal resolution: Widely used environmental datasets do not necessarily have a high temporal resolution. Monthly, seasonal, or yearly summaries often suffice or are more desirable than daily or even finer time increments.
- Metadata: Comprehensive metadata is important to enable modelers to assess the biases and uncertainties in environmental datasets and thus the potential consequences on SDMs.
- Cohesion: The vast majority of SDM software require that all environmental predictor data have matching attributes, including spatial extent, resolution, and cartographic system.

Additionally, because R is such a key component of contemporary analytics in the field of ecology, with 58% of ecological studies reporting using R as their primary tool in data analysis in 2017 (Lai et al 2017), environmental datasets gain in popularity if they can directly be downloaded and processed from R or if compatibility with R is somehow facilitated.

A number of environmental datasets have become extremely popular in SDM research because they have many of the characteristics listed above. These include WorldClim and WorldClim2 (Fick and Hijmans 2017), Climatologies at high resolution for the earth's land surface areas (CHELSA; Karger et al. 2021), and Ocean Climate Layers for Marine Spatial Ecology (MARSPEC; Sbrocco and Barber 2013). Due to their popularity in ecological research, these datasets and their corresponding characteristics have become central components of SDM software (e.g. Wallace; Kass et al. 2018) and research (e.g., some of the most cited SDM publications in Elith et al. 2006 and Wisz et al 2008). Several recommendations at the subgroup level are related to establishing Worldclim/ CHELSA as a baseline to which supplemental NASA data could be added.

The connection between SDM and biodiversity mapping

The current framework has been to develop SDM maps for individual species, and then to aggregate predicted distributions across species and cumulatively arrive at biodiversity distribution mapping (e.g., Cord et al., 2013). This approach may overestimate diversity totals by conflating aggregate fundamental niches (the environmental space capable of supporting a given species) with realized niche (where species actually are) due to unaccounted-for effects from inter-species competition and other factors (Moullec et al. 2022). Approaches such as the Generalized Joint Attribute Modeling for biodiversity analysis attempt to explicitly account for this inter-species interaction though joint-modeling based on composition (e.g., Clark et al. 2017). Alternatively, models may be parameterized to predict diversity values as a continuous response variable across spatial scales (e.g., Hakkenberg et al. 2018).

We now have species richness measurements across individual SDMs in some weighted fashion (e.g., imperiled species, rare species, species with restricted ranges) and spectrally based measurements of alpha (species richness) and beta diversity to define biodiversity patterns. A large number of additional biodiversity indicators that are derived from imaging spectroscopy are anticipated, particularly with the upcoming Surface Biology Geology mission and other upcoming missions that include imaging spectrometers (e.g.,The Environmental Mapping and Analysis Program (EnMAP), Copernicus Hyperspectral Imaging Mission for the Environment (CHIME)). If both environmental and biodiversity indicators can be generated at the same scale, how would they be compared? ***Participants recognized a paradigm shift in biodiversity modeling and a compelling need (though outside the workshop scope) to address the matter of how SDMs can integrate new remote sensing variables that are more directly sensed indicators of biodiversity.***

Network of networks

It should be noted that some workshop hosts and participants have been funded under National Science Foundation (NSF) Macrosystems Biology (MSB) programs, where predicting connections between fine scale processes and broad-scale spatiotemporal patterns requires integrating large ecological and remote sensing datasets generated by individuals and ecological research networks (e.g., Environmental Data Initiative, National Ecological Observation Network). Science organizations are investing in infrastructure to store and host data, yet demand for data storage far exceeds data use. A Network of Networks (NoN) concept is emerging that includes shared collection protocols, processing standards, best practices, training and Findable, Accessible, Interoperable, and Reusable (FAIR) data principles, and resources (Wilkinson et al., 2016; Goldman et al., 2021). There is a measurable overlap between the SDM and MSB communities and opportunity for NASA to reach ecological research networks and audiences broader than the SDM community when considering recommendations presented in this report that involve data tracking and integration tools (e.g., Servilla et al., 2016).

WORKSHOP DESIGN AND OBJECTIVES

The workshop was structured around a series of six sessions (January 6th, 11th, 13th, 20th, and February 3rd and 4th, 2022), where each session was 2.5 hours in length, with a two-week pause near the end to allow participants time to explore datasets each breakout group identified as relevant. A TWSC Workshop Announcement and Call for Participants (Appendix A) was broadly circulated through social media, professional listservs, (e.g., ecology, Ecological Society of America), and an invitation list that was carefully developed with input from expert practitioners, NASA, and NASA Principal Investigators (previously or currently funded). A survey of user needs was created in Qualtrics, distributed as part of the announcement, and used in lieu of an application. We received a total of 87 completed surveys from across the world (Figure 1), the majority of which were interested applicants (Appendix B). Survey results (Appendix C) were summarized and a rubric for ranking applicants was developed based on attaining the technical aim of the workshop, which was to generate a recommendation list to broaden the use NASA datasets that capture important predictor variables, but are under-used because of accessibility, interoperability, or other technical challenge. Participant support was provided in the amount of $2500 stipends for 12 of the applicants. A team of six reviewers independently ranked application responses. We prioritized top scoring applications in terms of stipend eligibility, (e.g., need based, soft-funded positions), and unique qualifications. From this shortlist, we then ensured that applicants with a range of career levels, themes, and backgrounds (organizational and geographical) were offered stipends. The stipend amount was based on a commitment to attend all workshop sessions and provide detailed feedback and recommendations as requested, or the equivalent of approximately five workdays. Workshop attendance varied at any given time and ranged from 30 to over 50 participants (Figure 2). Workshop sessions were recorded and made available to participant collaborators and all applicants, including postdoctoral researchers and students of all levels.

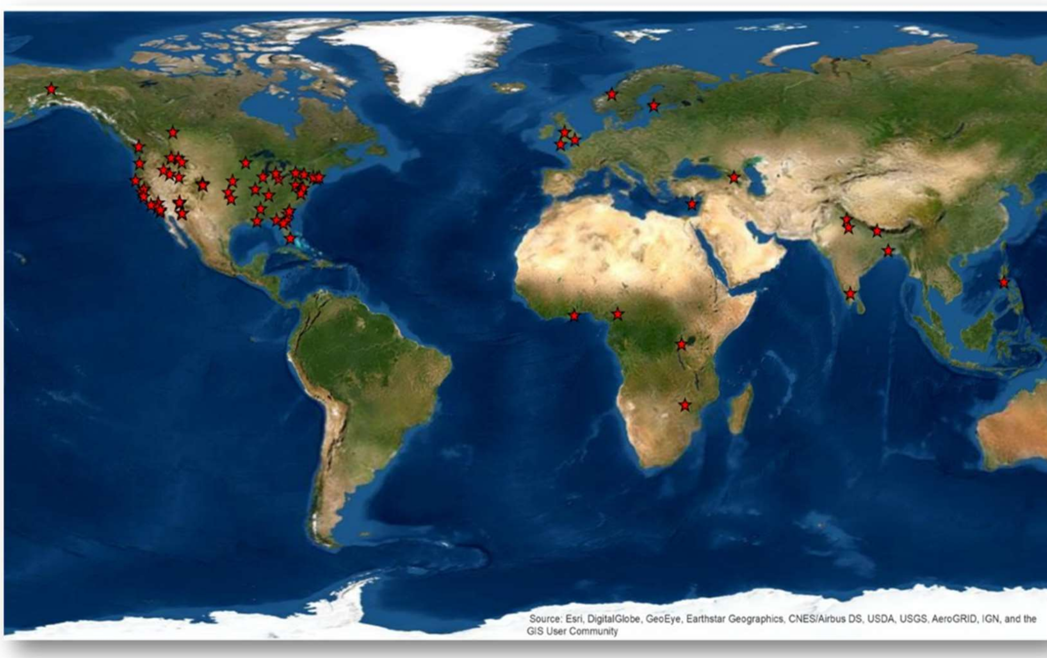


Figure 1. Geographic distribution of applicants who participated in the workshop on broadening the use of NASA datasets by the Species Distribution Modeling (SDM) community.

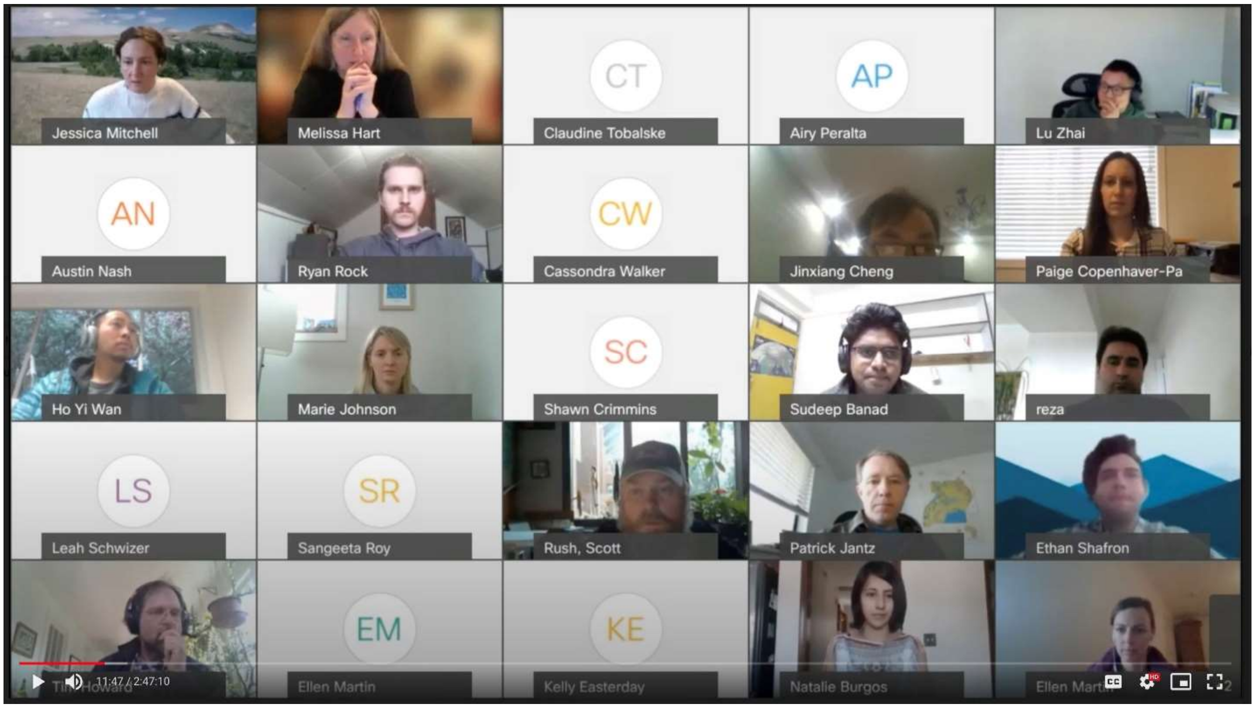


Figure 2. Example (screenshot) of participants attending one of the virtual workshop sessions on broadening use of NASA datasets by the Species Distribution Modeling (SDM) community.

The workshop agenda loosely followed open data science concepts, where data investigation and analysis were team contributions (Lowndes et al., 2019). The workshop agenda (Appendix D) was designed to **identify and address shared needs** (e.g.,lightning talks) and **normalize data discussion** among SDM practitioners with a range of backgrounds and user levels through invited talks and introductions to NASA datasets, organized by theme (climate, edaphic, vegetation, human disturbance, surface water). Workshop hosts introduced NASA EOSDIS Distributed Active Archive Centers (DAACs); data interfaces (access tools such as NASA AppEEARS!, Giovanni, and Earthdata); new resources for discovering, accessing, and using data that are transitioning to the cloud (e.g., Harmony API and Earthdata Cloud Cookbook); and skill-building and training opportunities such as ARSET webinars for remote sensing background, and the DAAC Mentor Cohort program. During lightning talks and discussions in the first session, a general recommendation emerged to address the need for data standardization, with software and workflows that can be used for reproducibility, analysis, and storage.

Workshop objectives

Workshop objectives (Figure 3) were developed in close coordination with NASA and defined as follows:

1. Identify NASA data products that capture important predictive distribution variables, but are under-used by the species distribution modeling community because of usability, accessibility, interoperability, or other technological challenges.
2. Conduct a detailed evaluation of technologies, methods, and capabilities for accessing and processing static and time-series data products relevant to species distribution modeling.
3. Prioritize candidate variables based on level of effort to modify products or technologies for dissemination and use, and potential for significantly boosting accuracy results for a wide-range of species distribution models.

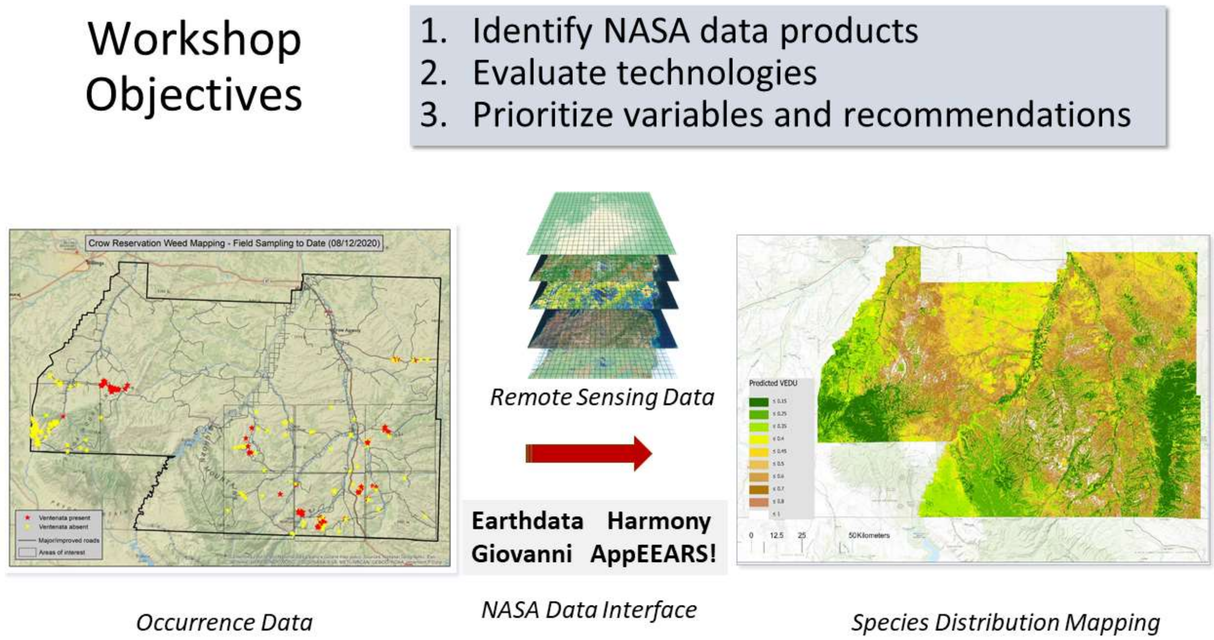


Figure 3. A conceptual overview of workshop objectives, where remote sensing datasets and NASA repositories can play a pivotal role in generative predictive distribution maps for terrestrial and aquatic plant and animal species.

IDENTIFICATION OF RELEVANT DATASETS

To address the first workshop objective to identify NASA data products that capture important predictive distribution variables, the workshop hosts began by generating an exhaustive list of likely relevant datasets, organized by theme. The team conducted preliminary evaluations for a number of datasets based on expert input and survey feedback. The preliminary investigations were developed as lines of inquiry and to solicit detailed input from workshop participants. Background material was presented on relevant variables to normalize data discussion, along with presentations by Gerald “Stinger” Guala, Ryan Pavlich, Ryan Rock, Zhihua Liu, Laura Duncason, Marie Johnson, Ethan Shafron, Kerry Cawse-Nicholson, Patrick Burns, Mary Blair, Beth Gerstner, and Jeremy Werdell.

After polling in workshop sessions, participants were assigned to breakout groups based on thematic interest to address Workshop Objective 2 (conduct a detailed evaluation of technologies, methods and capabilities for accessing and processing static and time-series data products relevant to species distribution modeling). Participants were prompted to (1) identify what NASA products they would like to evaluate in detail by downloading a sample dataset (single tile or large dataset) and providing feedback, and (2) decide on criteria the group will use for evaluating datasets and providing feedback (e.g., spatial resolution, repeat frequency, data standards, mission continuity, technological challenges, testing in SDMs with published results). Participants were encouraged to use Earthdata search tools and referred to the [Species Distribution Modeling Data Pathfinder](#) where raster-based datasets available for model input are organized thematically, along with tools for data access and visualization. NASA datasets that the workshop identified as relevant to SDMs did not include composite response variables such as ecological integrity indicators (Hansen et al., 2021).

FINDINGS

Working group notes for vegetation, surface water/marine, edaphic, climate, and human disturbance variables are included below. General recommendations made by each group are provided below and have been summarized in the executive summary.

Vegetation variables

Working group: Patrick Burns, Paige Copenhaver-Parry, Matt Fitzpatrick, Chris Hakkenberg, Tim Howard, Marie Johnson, Abhishek Kumar, Zhihua Liu, Cory Merow, Eric Nielsen, Lu Zhai

Relevant products: The vegetation subgroup identified Global Forest Cover Change (GRCC) Products, GEDI products, Harmonized Landsat Sentinel-2, and Landsat 9 reflectance bands as the primary sources of relevant NASA datasets.

Global Forest Cover Change (GRCC) feedback

The advantageous use of Global Forest Cover Change (GRCC) products (Sexton et al., 2013) was demonstrated in research presented during the workshop. Depending on the product, temporal extents of these products are roughly 1999/2000 through 2014/2015, which limits their use in some SDM applications that backcast suitable habitat based on more recent species occurrence datasets (e.g., Gavrutenko et al., 2021).

Table 2 . Global Forest Cover Change products relevant to the Species Distribution Modeling community.

Global Forest Cover Change (GFCC) Products	DAAC Link	Short Name
Forest Cover Change	LP_DAAC	GFCC30FCC
Surface Reflectance Estimates	LP_DAAC	GFCC30SR
Tree Cover *	LP_DAAC	GFCC30TC
Water Cover	LP DAAC	GFCC30WC

* *Version 4 of Tree Cover has been created but is not available via Earthdata or DAAC.*

Recommendation

One of the recommendations agreed upon is to extend the temporal extent of this dataset to present, and if feasible, pre-1999.

GEDI feedback

Helpful tools

- Level 3 gridded land surface metrics
  - No coding necessary using the Spatial Data Access Tool (SDAT), spatial subsetting is easy if user has general GIS skills
  - https://webmap.ornl.gov/ogc
- The webinar on using GEDI Level 1B and Level 2 products, which include geolocated full waveform measurements (L1B), elevation and height metrics (L2A), and canopy cover and vertical profile metrics (L2B) footprints
  - https://youtu.be/UlrCC1Xp-wk

- The webinar session that summarizes where and how GEDI lidar data can be accessed, followed by a demo showing participants how to open, interpret, and analyze GEDI data for assessments of vegetation structure, followed by a Q&A session
    - https://lpdaac.usgs.gov/resources/e-learning/accessing-and-analyzing-gedi-lidar-data-for-vegetation-studies/
- The Spatial Data Access Tool (SDAT) for accessing Level 3 and Level 4B products

Challenges
GEDI products and the challenges encountered include scattered tutorial and resources, broken code, installation issues, and multiple user guides for one process. A description of product applications and limitations would be helpful, particularly in the context of SDMs. A PhD student evaluating the LP DAAC GEDI Finder Service and the GEDI Spatial Querying and Subsetting Quick Guide noted an inability to spatially or temporally subset data (https://git.earthdata.nasa.gov/projects/LPDUR/repos/gedi-subsetter/browse). A similar evaluation was conducted for RGEDI and issues with installation were reported (summer 2022). The package was removed from CRAN. It was determined during the workshop that some of the issues noted were due to GEDI data being split between two DAACS. The information was communicated to NASA on 2/15/2022.

Table 3. GEDI products relevant to the Species Distribution Modeling community.

GEDI Products	DAAC Link	Short Name
Geolocated Waveform L1B	LP DAAC	GEDI01_B
Elevation and height metrics L2A	LP DAAC	GEDI02_A
Canopy covery and vertical profile metrics L2B	LP DAAC	GEDI02_B
Land surface metrics (V2) L3	ORNL DAAC	GEDI_L3_LandSurface_Metrics_V2_1952
Aboveground biomass density (V2) L4A	ORNL DAAC	GEDI_L4A_AGB_Density_V2_1_2056
Aboveground biomass density (gridded) L4B v2	ORNL DAAC	GEDI_L4B_Gridded_Biomass_2017

* *Only V2 products are included as all V1 products are flagged as out of date now on the DAACs.*

Recommendations
Combine and update the GEDI Quick Guide and GEDI finder user guides. Consider providing flowcharts and short videos. There is opportunity to increase the usability of the RGEDI code if it is well maintained and updated. Overall, providing access to GEDI products in a more streamlined process is the top recommendation. For example, a user may start going through code through the Earthdata bitbucket (https://git.earthdata.nasa.gov/projects/LPDUR/repos/gedi-tutorials/browse), but later find that a less confusing point of entry would have been through Earthdata E-Learning resources (https://lpdaac.usgs.gov/resources/e-learning/). We suggest deleting code and other resources that are out of date to decrease confusion (Appendix E). Consider generating new resources, for example a short (3-minute) video on GEDI finder. Consider more R resources and a shapefile of GEDI transects.

Harmonized Landsat Sentinel-2 (HLS-2) feedback

Harmonized Landsat Sentinel-2 data are geometrically corrected, solar angle-adjusted, BRDF corrected 30m resolution data available every 2-3 days. This processing pipeline allows for interoperability between Sentinel 2 and Landsat 8 data and are available from 2013 to present. These data are available on EarthExplorer, but not on Earth Engine or other central repositories. Given its high temporal frequency, HLS-2 data have been used in phenological analyses for invasive grass species mapping across large areas (Dahal, et al, 2022).

Table 4. Harmonized Landsat Sentinel-2 products relevant to the Species Distribution Modeling community.

Harmonized Landsat Sentinel-2 Products	DAAC Link	Short Name
HLS Operational Land Imagery Surface Reflectance and TOA Brightness Daily Global 30m	LP DAAC	HLSL30
HLS Sentinel-2 Multi-spectral Instrument Surface Reflectance Daily Global 30m	LP DAAC	HLSS30
MuSLI Multi-Source Land Surface Phenology Yearly North America 30m	LP DAAC	MSLSP30NA

Recommendations
A detailed evaluation of the HLS dataset was not performed as part of this workshop; however, in addition to identifying specific issues and challenges, we suggest testing the sensitivity of topographic corrections in relation to SDM model performance. To our knowledge, the HLS products are corrected to surface reflectance but have not been topographically corrected. If generating a user's guidebook, we also suggest identifying ways in which HLS and other time-series datasets could support dynamic SDMs for conservation planning and decision support applications.

Landsat 9 reflectance bands

Landsat 9 images are currently only available for download through USGS channels, not yet through NASA Earthdata. Landsat 9 data were not widely available during the workshop (released in February 2022) so specific issues and recommendations around using the data in SDM will take some time to emerge. Given the similarity between Landsat 8 and 9, these data offer opportunities for use in SDM. Continuous, remotely-sensed variables have been shown to be better model predictors than land-cover variables, which tend to overpredict habitat suitability for plants (Cord et. al, 2013). Additionally, derived vegetation indices and surface temperature data, when combined with other biologic and topographic variables, can improve the accuracy of SDM (Schwager and Berg, 2021).

Table 5. Landsat 9 spectral bands relevant to the Species Distribution Modeling community.

Spectral Band	Wavelength in micrometers	Resolution in meters
Operational Land Imager-2 (OLI-2)		
Band 1 - Ultra blue (coastal/aerosol)	0.435-0.451	30
Band 2 - Blue	0.452-0.512	30
Band 3 - Green	0.533-0.590	30
Band 4 - Red	0.636-0.673	30
Band 5 - Near infrared	0.851-0.879	30
Band 6 - Shortwave infrared 1	1.566-1.651	30
Band 7 - Shortwave infrared 2	2.107-2.294	30
Panchromatic	0.503-0.676	30
Cirrus	1.363-1.384	
Thermal Infrared Sensor-2 (TIRS-2)		
Band 10 - Thermal infrared 1	10.60-11.19	100*
Band 11 - Thermal infrared 2	11.50-12.51	100*

** Thermal bands are acquired at 100m resolution and resampled to 30m.*

Recommendations
As with similar remote sensing products, ease of access and processing difficulties are likely the biggest barriers to entry for using reflectance datasets in SDMs. Making endpoints for users to programmatically work with these data or access via GIS software without excessive manual searching are key to increasing use. As such, building out APIs for making data requests, storing these datasets on google earth engine (and having tutorials for working with them), and making spatial/temporal querying as streamlined as possible are logical ways forward.

General comments and recommendations

The data do not align with the skill sets and needs of the user community. All of the identified products were of interest to the subgroup and suggested they would be more widely used if available in a centralized location, with standardization along a few common dimensions, such as plain language descriptions and references to appropriate uses in different Species Distribution Models and applications.

Certain datasets, like GEDI, are problematic to access because of large file sizes and slow download rates. The group would recommend aggregating point and polygon data (where feasible) to raster grids, chunking these data into even smaller pieces (e.g. 5x5 degree grids), and storing on a server that allows for faster downloads. Typical GEDI download speeds were on the order of 1-5 mbps, for the full ~60 TB that would take 100+ days to download.

For basic users:

- Some common spatial grains (e.g. 1km) that are compatible with WorldClim and Chelsa would provide a gateway to introduce people to the raw products.
- A common repository for derived products in a common res/extent, with a way for all users to be aware of assumptions underlying the common res/extent
- Temporal aggregation, because temporal precision is < important that spatial precision in many applications

For intermediate users:

- A tool that allows for customized spatial and temporal subsetting and aggregation before downloading. NASA AppEEARS has some functionality in this regard.
- One interface to access all the data sets would be ideal. GEE would be easy. Again something like AppEEARS could work. People started using the Species Distribution Modeling Data Pathfinder website more frequently as discussions continued and participants became more aware of the resource; however, the platform does not provide direct data access.

For advanced users:

- Standardized descriptions that could be read in ~5 min. And automate detection of citations that can be included in a ref list. Perhaps vignettes with usage examples.
- Primarily working with raw data and their own custom scripts. Ease/speed of access and documentation are biggest considerations.

Surface water/marine variables

Working group: Sudeep Banad, Natalie Burgos, Sangeeta Roy

Relevant products: MEaSUREs - Multi-scale Ultra-high Resolution (MUR) Sea Surface Temperature (MEaSUREs-MUR)

MEaSUREs feedback

Sea surface temperature (SST) is one of the main environmental factors regulating marine species distribution. There is literature available where SST was used as predictor for determining distribution and anthropogenic risk exposure for highly mobile species, like the blue whale (Abrahms et al., 2019;

Bedriñana□Romano et al., 2018, 2021; Barlow and Torres, 2021), Atlantic herring, Atlantic mackerel, butterfish (Wang et al., 2018), hammerhead sharks (Mason et al., 2019), Cuvier's beaked whales, sperm whale, and Risso's dolphin (Frasier et al., 2021).

Issues/challenges

Datasets are too frequent and large in volume. For example, from the MUR SST (v4.1, JPL NASA; 11.03.2015) data direct download link, users can access a total 7,207 temporal files to date (02/25/22). For MUR SST v4.2 (JPL NASA; 18.10.2019), the available number of files is 7,115. Compilation of daily datasets into monthly or yearly averages is needed for use in SDM.

MEaSUREs-MUR is a part of the NASA Making Earth System Data Records for Use in Research Environments (MEaSUREs) Program. The mission objective is to develop a coherent and consistent daily map of SST at the highest spatial (horizontal) resolution possible at global scale. The datasets are gridded, gap-free, and easy to download.

Table 6. MEaSUREs - Multi-scale Ultra-high Resolution (MUR) Sea Surface Temperature (MEaSUREs-MUR) products relevant to the Species Distribution Modeling community.

Data products Abbreviation	Spatial resolution	Time span	Temporal resolution	Format
MUR SST (v4.1)	0.01 degrees (Latitude) *x* 0.01 degrees (Longitude)	2002-May-31 to Present	Hourly - < Daily	netCDF-4
MUR SST (v4.2)	0.25 degrees (Latitude) x 0.25 degrees (Longitude)	2002-Aug-31 to Present	Hourly - < Daily	netCDF-4

Link: https://podaac.jpl.nasa.gov/MEaSUREs-MUR?tab=background&sections=about%2Bdata

Data access and other information

Data recipe:

https://github.com/podaac/data-readers

https://github.com/nasa/podaac_tools_and_services/tree/master/subset_opendap

MUR SST (v4.1)

- Direct Access

https://cmr.earthdata.nasa.gov/virtual-directory/collections/C1996881146-POCLOUD

- Earthdata Search

https://search.earthdata.nasa.gov/search/granules?p=C1996881146-POCLOUD

MUR SST (v4.2)

- Direct Access

https://cmr.earthdata.nasa.gov/virtual-directory/collections/C2036880657-POCLOUD

- Earthdata Search

https://search.earthdata.nasa.gov/search/granules?p=C2036880657-POCLOUD

Recommendations
Datasets with a spatial resolution less than 1 km are necessary for accurate prediction of some species distribution, like range- restricted or unique species. Tutorials with screenshots of stepwise processes (for using datasets in SDM) are needed. NASA MUR SST data layers should be directly accessible by R packages.

Edaphic Variables

Working group: Israel Borokini, Nathalie Chardon, Catherine Jarnevich, Scott Rush

Relevant products: MERRA-2 soil moisture, MERRA-2 precipitation, MERRA-2 monthly extremes detection indices, AMSR surface soil moisture, SMAP radar soil moisture

The NASA edaphic datasets available are at coarse spatial resolutions, which is problematic for organisms that rely on microhabitat variation (e.g., small plants), but useful for organisms that respond to broad landscape patterns in edaphic conditions (e.g., locusts; Gómez et al. 2019). In particular, predictor variables from remote sensing could be superior to downscaled products (He et al., 2015) and advances in remote sensing products are yielding better estimates of the microclimate important for many organisms (Zellweger et al., 2019). The SMAP products have been used in some SDM studies (e.g., Lauer et al. 2018), whereas other products containing edaphic variables (e.g., MERRA and AMSR) have not, to our knowledge, been used in SDM studies. Rather, MERRA-2 products and AMSR are widely used for for drought monitoring in agricultural landscapes, atmospheric pollution assessment, and hydrological assessment (e.g., Ines et al., 2013; Zaussinger et al., 2019; Rizza et al., 2019; Forgotson et al., 2020; Modanesi et al., 2020; Ukhov et al., 2020; Das et al., 2021, Xu et al., 2022). These products have also been assessed in various systems for their reliability in validation studies (Brocca et al. 2011, Draper et al 2008). Moreover, the study where MERRA-2 products were officially published did not mention SDMs as a potential use of these products, but focused on their use for aerosol observations and understanding stratospheric and cryospheric processes (Gelaro et al., 2017).

Current issues and potential solutions

Many of the datasets that the Edaphic Subgroup tried to download had similar issues, which primarily had to do with accessibility and useability.

Issue: The subgroup found it difficult to differentiate how each different dataset could be used in SDMs (resolution, temporal coverage, parameters available).

Suggestions: Use descriptive filenames, provide easy-to-digest metadata, and examples of what kinds of scenarios each dataset would be appropriate to use in and caveats of use. Along with many others in this workshop, we recommend a metadata table for each dataset that contains:

- Layer name: e.g., SMAP Soil Moisture
- Keywords: searchable terms for each layer (e.g., terrestrial, vegetation)
- Temporal extent: e.g., 2015-2022
- Temporal resolution: e.g., 60 min
- Spatial extent: e.g., global
- Spatial resolution_x: e.g., 9 km
- Spatial resolution_y: e.g., 9 km
- Units
- CRS
- Calculation: e.g., statistically weighted regression from raw product
- Description: 3 sentence description and link to short 5 min talk (as seen in SDM workshop) explaining dataset
- Common uses: e.g., covariate in vegetation SDMs

- Examples of use: link to papers
- Recommendations
- User comments: Link to forum where users discuss this one dataset
- Dependencies
- File size/unit
- Data license: e.g., CC BY
- Version: e.g., 1.4

Issue: Coarse spatial resolutions in edaphic variables are not meaningful predictors for many of the organisms (e.g., smaller plants) that SDMs are built for. Edaphic variables are closely linked with the topography of the terrain they describe, and these values can greatly vary within small distances.

Suggestion: Generate predictors at different spatial scales. For example, edaphic products at 2 m and 30 m resolution could be made at state or regional levels, while coarse resolution products can be made at national, continental, and global levels. From literature and end-user perspective, continental and global studies often use low resolution predictors for analysis, given the computation limitations. However, species-specific SDMs often use predictors at resolutions that match the scale at which species-environment signals can be detected. For edaphic predictors, this would be at high resolution, especially for endemic and at-risk species that exhibit edaphic specializations. Continuous products at high spatial resolutions are most valuable to use in SDMs. Furthermore, products at high temporal resolution would be most useful.

Issue: Most datasets are very large in size, making it difficult (or even impossible) for users to download data onto their personal computer.

Suggestion: We recommend allowing users to customize their download area and optionally choose their temporal and spatial extent.

Issue: Point products are not useful predictors to use in SDMs, which need continuous spatial variables to generate prediction maps.

Suggestion: Generate wall-to-wall data.

Issue: Data exists for short time periods instead of for the traditional 30-year climatology lengths. This creates a mismatch between covariate temporal periods, and potentially requires the user to process individual datasets to get matching temporal periods.

Suggestion: Creating 30-year averages that are the same temporal extent as bioclimatic variables calculated at 30-year scales (e.g., WorldClim, CHELSA) would be ideal. In many cases NASA missions don't align well with WorldClim or CHELSA timespans (30-year averages); therefore, we suggest generating 1-year or 10-yr averages, where users create their own multi-decadal averages through a selection interface.

Issue: Many data products are only available for the finite length of the mission, yielding discontinuous data products. This poses problems for scientific reproducibility.

Suggestion: If a product is made available for SDM use, keep it available.

Issue: There is a substantial access barrier to download the data, such as needing an account and needing to click through many different links to finally download the dataset.

Suggestion: Create clean websites that are easy to use (e.g., see https://chelsa-climate.org/downloads/) both for downloading data and accessing the data specifications.

Issue: Many datasets are available in uncommon file formats (e.g., h5, nc4), making them difficult to import into R or even requiring separate software to read them.

Suggestion: Publish datasets in consistent and common file formats, such as TIFF.

Issue: Datasets have to be individually downloaded, forcing the user to process all data.

Suggestion: Produce a raster stack of all available datasets (i.e. layers in R-compatible format such as TIFF), with user-defined layers to download. As many of these variables are likely to be correlated and thus should not be used together as covariates, a correlation analysis between variables should be available to help users make decisions on which layers to use.

General comments and recommendations

Environmental variables produced from remote sensing have increased in recent years, providing better and more reliable applications in SDM studies than existing datasets. For example, Waltari et al. (2014) compared surface temperature products from MERRA and ASMR-E with those generated from WorldClim for fitting SDMs for 20 vertebrate species in the Amazonia, the Atlantic Forest, the Cerrado, and Patagonia; four distinct biomes in South America. They reported that MERRA products performed equally well or better than WorldClim, while the latter performed better than AMSR-E products in three of the four biomes (Waltari et al., 2014). Furthermore, MERRAclim data were released recently, which includes the Antarctic continent that is missing from WorldClim and other related datasets (Vega et al., 2017). These comparative studies underscore the superiority of remote sensing data compared to products derived from the interpolation of in situ climatic data.

However, despite these geospatial advancements, spatial resolution remains a major limitation. The ambition of providing bioclimatic variables of global extent comes at the expense of spatial resolution, which limits their use for SDMs. Most SDM studies are species-specific, covering the known ranges of a species, which are often at regional and sub-national levels, thus NASA products need to match these spatial grains for increased usability. MERRAclim products, for example, are only available at 2.5, 5, and 10 arc-minutes (Vega et al., 2017). Moreover, these coarse scaled products are not adequate to capture heterogeneous and geodiverse soil properties which occur at very fine spatial scales.

In order for NASA data to become more useful in the SDM community, we urge NASA and users alike to think about what NASA datasets were originally designed to measure, and how, for example, changes in daily maximum surface temperature differs substantially from what SDM users need as covariates (e.g., 30-year averages). The most frequently used ecological variable datasets were developed with a minimum of 30-year averages, with updates almost every year. This includes the 1 km^2 WorldClim data (1970-2000; Fick and Hijmans, 2017), as well as the Parameter-elevation Regressions on Independent Slopes Model (PRISM) data (1980 - 2020; Daly et al., 2008). Under the circumstances of mismatched temporal resolutions, end users would be forced to choose either NASA or other products for their SDM studies. We therefore recommended that NASA products should be made available in a flexible manner that would allow end users to generate decadal averages (e.g., 30-year averages) of predictor variables.

On the other hand, there is an increasing demand for predictor variables at finer scale spatial resolutions to capture landscape heterogeneity and microclimatic conditions, which are suitable for detecting reliable species-environment relationships. There are current efforts to downscale the 1 km2 resolution of WorldClim data (1970-2000; Hijmans et al., 2017) to 100 m resolution for increased use in SDM studies (Poggio et al., 2018). High spatial and temporal resolution is very key for the many organisms that rely on microhabitat variations, and thus products need to be at high resolutions to be useful in some SDMs (see Maclean et al. 2018 & Maclean 2019 for downscaling). This is particularly important for accurately capturing edaphic heterogeneity on the landscape. Moreover, SDMs are increasingly being used as a conservation tool for many endemic and range-restricted species. Many of these range-restricted species also exhibit edaphic specializations, and their habitats can be reliably detected in high spatial resolution. Therefore, edaphic variables should be available in fine scales to enhance their applicability for SDMs. It should be noted that these fine-data are not meant to be mutually-exclusive from having 30-year time scales and both parameters are preferred.

In regards to edaphic conditions, soil temperature can be more important than soil moisture in certain systems, such as deserts, meaning that soil moisture might be an inadequate predictor in these systems and soil temperature data would be preferred (e.g., Lembrechts et al., 2020). Other variables that can be important as edaphic covariates, but currently do not exist through remote sensing products, are soil density, soil chemistry, and microbial community structure. Some of these are already available in POLARIS, a 30 m probabilistic soil series map of the United States (Chaney et al., 2016).

Finally, we emphasize to NASA that many end-users of these products are not computer scientists (e.g., NGOs) and will be discouraged if it is too complicated to access and download data.

Climate variables

Working group: Reza Goljani-Amirkhiz, Ellen Martin, Ho Yi Wan

Products evaluated: Ecostress, Merra-2, MODIS snow cover, Global Precipitation Measurements (GPMs), and NLDAS-2

Species distribution models look at the relationship between species' distributions and their environment, often quantified using spatial climate variables (Stoklosa et al. 2014). NASA provides a variety of climate datasets such as Ecostress, Merra-2, MODIS snow cover, Global Precipitation Measurements (GPMs), and NLDAS-2, which were evaluated by workshop participants. In terms of specific examples, Anand et al. (2021) used MODIS and Ecostress, among other variables, to show the superiority of Convolutional Neural Network models over BioClim for predicting the regional distribution of *Rhododendron arboreum*; Zhu and Guo (2022) developed a suite of winter habitat indices by fusing MODIS snow products and NASA MEaSUREs daily freeze/thaw records and tested how these indices could improve the explanation of species richness patterns across China; and Deblauwe et al. (2016) showed that datasets derived from the Tropical Rainfall Measuring Mission (NASA/JAXA) and MODIS provided better models than those from BioClim for a large percent of plant species in the tropics.

ECOSTRESS

The land surface temperature data from Ecostress were downloaded and evaluated as an alternative to MODIS land surface temperature for use in modeling the distribution and abundance of pheasants in South Dakota. ECOSTRESS data were never tested in the models due to access and usability challenges, summarized below.

Issues/challenges

- Thirteen matching collections (Figure 4); difficulties due to confusion on differences, acronyms and which one(s) is/are best for the user's needs.
- Datasets are only available in .h5 format (.h5 is not a common format for SDM analysis).
- Only daily data are available, whereas SDMs mostly need monthly and seasonal data. For example, for a pheasant habitat suitability model, monthly data would be needed. (Goljani-Amirkhiz, pers. comm.). Daytime and nighttime LST are not available separately.
- Limited temporal coverage (2018- present), whereas many SDM studies use past occurrence records to evaluate trends in distribution and abundance. Pheasant observation records, for example, were obtained from 2015-2019 (Goljani-Amirkhiz, pers. comm.).

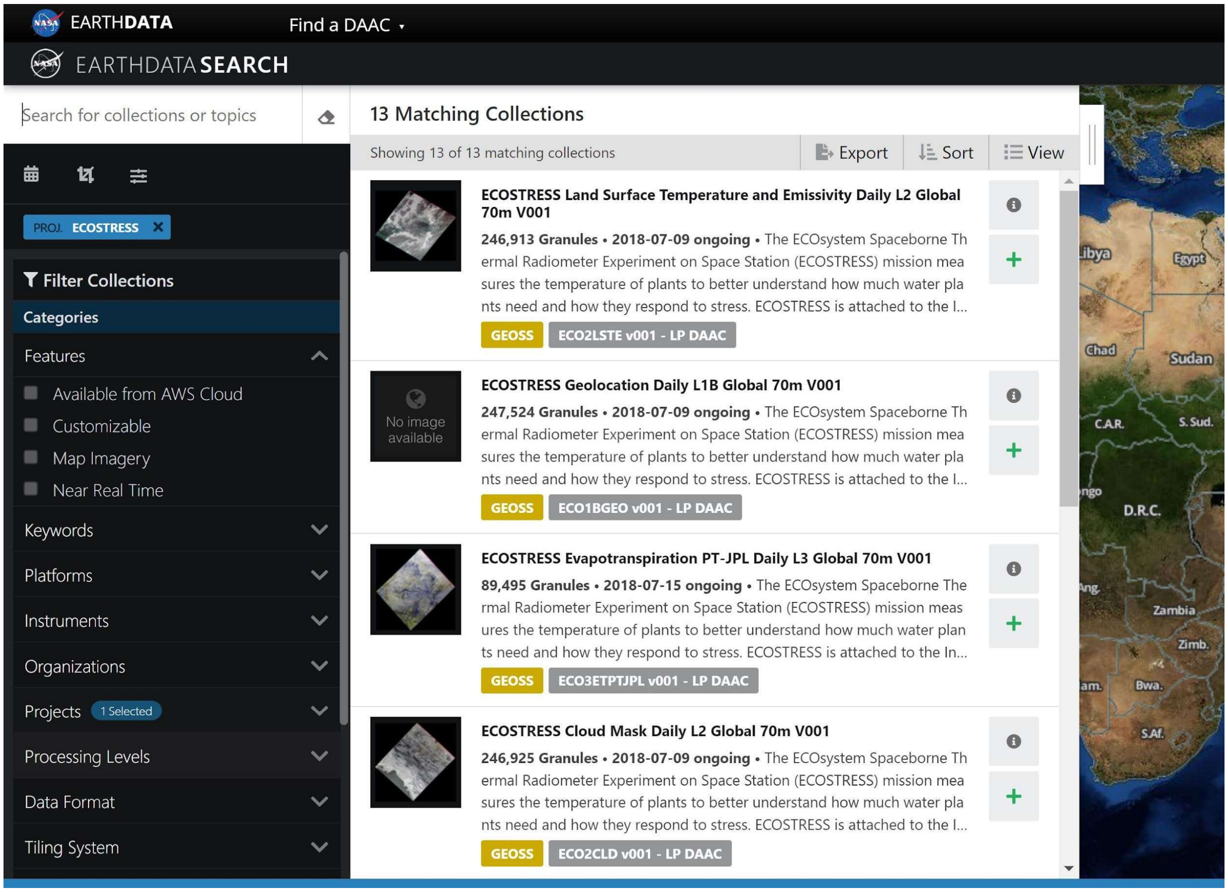


Figure 4. Example of the large number of matching collections and difficulty selecting relevant datasets without plain language product descriptions or knowledge of applicability in Species Distribution Models.

Recommendations

Possible improvements could include adding a description page that includes:

- A summary of the project and datasets' attributes
- Definitions of acronyms (e.g., L2, L1B, V001)
- Differences and specific applications
- Links to references

MERRA-2 precipitation & monthly extremes detection indices

MERRA-2 data products offer a long time-series (1980-present) of coarse resolution (0.5 degree by 0.625 degree) global climate data at hourly temporal resolution. The climate working group acknowledged that the global coverage of MERRA-2 products is advantageous, as opposed to working with discreet weather station data. After the workshop ended, NASA announced the addition of MERRA-2 products to the THREDDS Data server, which will enable more efficient access to the time-series and may alleviate some of the access and processing issues outlined in this section, although issues around naming conventions and technical language will likely continue to discourage many users.

Issues and recommendations

- Standardize the initial file naming to be comprehensible, clean, and concise. Can keep this long, acronym-filled name for the user once we have identified it as the dataset we are looking for.

- Seemingly no reason to select one versus another. There are many different versions of MERRA-2 products. List the data as a separate column (e.g., Monthly Mean, Monthly Percentiles, Monthly Aggregated, Hourly Time-Averaged) so that it is easy to identify. This is buried in the file name (Figures 5 and 6).

- Clearly delineated workflow which helps you decide which product format to download from the bevy of options.

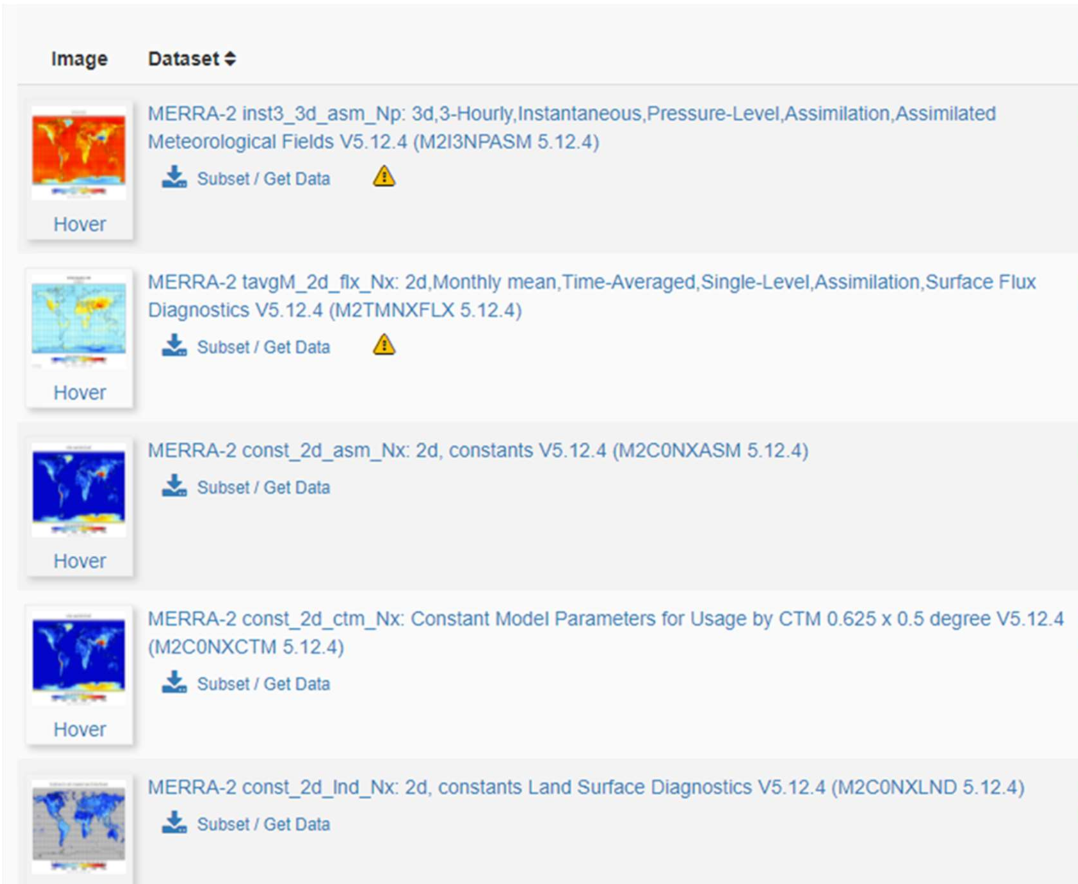


Figure 5. Example of multiple Merra-2 product versions and subsequent difficulty selecting datasets most appropriate for use in Species Distribution Models.

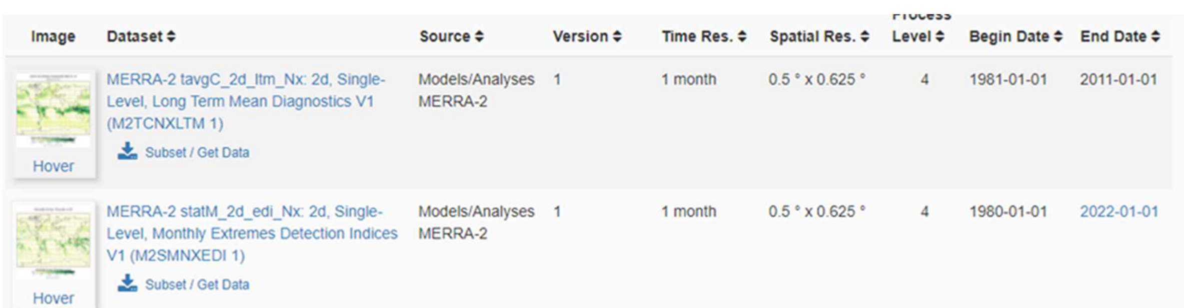


Image	Dataset	Source	Version	Time Res.	Spatial Res.	Process Level	Begin Date	End Date
Hover	MERRA-2 tavgC_2d_ltm_Nx: 2d, Single-Level, Long Term Mean Diagnostics V1 (M2TCNXLTM 1) Subset / Get Data	Models/Analyses MERRA-2	1	1 month	0.5 ° x 0.625 °	4	1981-01-01	2011-01-01
Hover	MERRA-2 statM_2d_edi_Nx: 2d, Single-Level, Monthly Extremes Detection Indices V1 (M2SMNXEDI 1) Subset / Get Data	Models/Analyses MERRA-2	1	1 month	0.5 ° x 0.625 °	4	1980-01-01	2022-01-01

Figure 6. Example of multiple Merra-2 product versions and subsequent difficulty selecting datasets most appropriate for use in Species Distribution Models.

- There are two steps necessary to actually download the data. Hitting "Download Data" does not download the data onto your computer locally. Another final, but small step is needed. You must hit the "Save" button to specifically bump the files to your local drive. This caused much lost time searching locally for the files which never actually made it onto the computer (Figure 7).

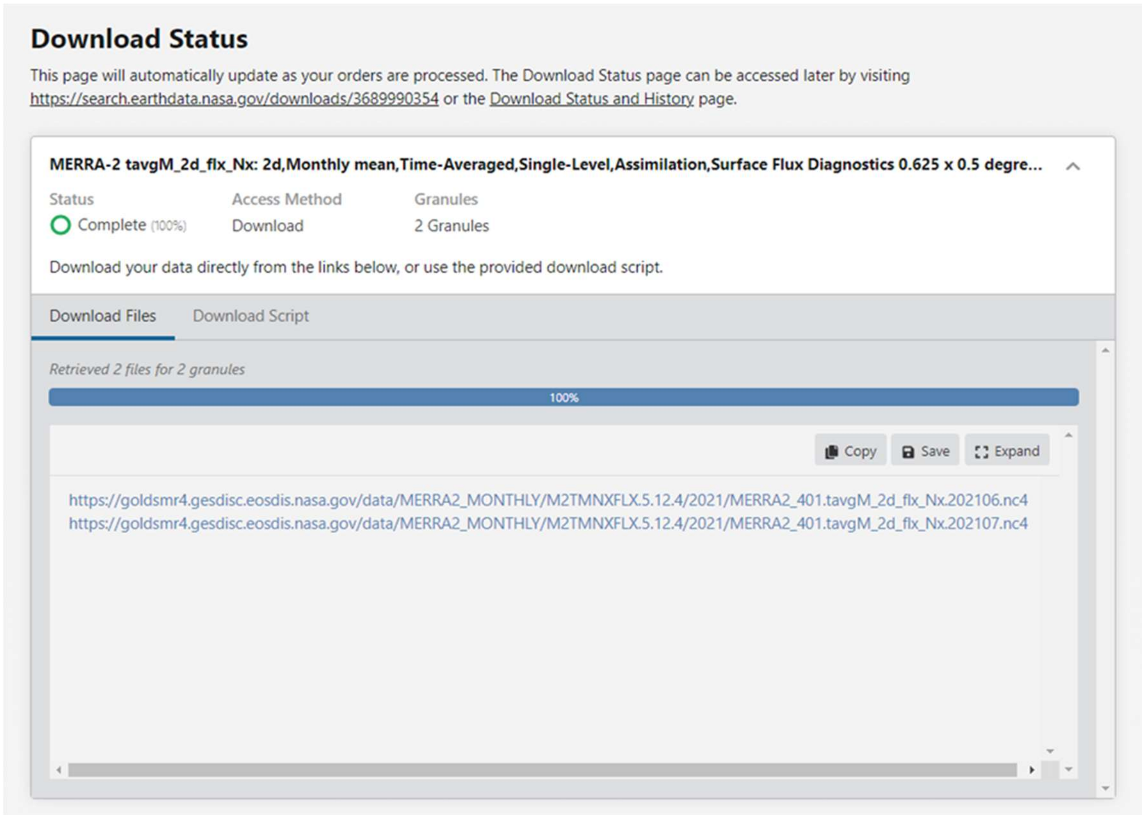


Figure 7. Example of a technical issue related to the Merra-2 download steps.

- Unnecessary redirecting to the webpage which requires Earthdata Login for Data Access. Happened three times after initial log-in before I downloaded the data successfully.
- There is no obvious way to download the data as a raster. Additional steps are necessary to convert the data from Nc4 to something more user-friendly.

Global Precipitation Measurements (GPMs)

The GPM Global Microwave Imager (GMI) sensor is a passive microwave radiometer with 13 bands that measure microwave energy from precipitation. The resolution is 10 km every 30 minutes and is available with data from 2000 to the present.

Variables include precipitation rates, probability of liquid precipitation, precipitation accumulation. The following datasets were identified as being potentially useful but were not evaluated in detail as to the specific benefits they may provide to SDM.

- "Early run" NRT gridded precipitation data available for every 30 minutes within hours of acquisition (for applications in weather forecasting, disaster management, etc.).

- "Late run" NRT gridded precipitation data available for every 30 minutes within 15 hours of acquisition with quasi-Lagrangian time interpolation.

- “Final run” gridded precipitation data available for every 30 minutes within 3.5 months of acquisition with quasi-Lagrangian time interpolation, gauge data, climatological adjustments.

- Data access via FTPS, HTTPS, GES DISC, Giovanni, and STORM.

- GPM IMERG Image Service: https://www.arcgis.com/home/item.html?id=039c16e86b254bca80cdf198f054dce6.

MODIS snow cover

Snow is an important variable that can strongly affect species distribution. For example, moisture content from snow packs can affect the growth of many plant species. The presence of snow also affects habitat selection and movement patterns of many wildlife animals. Therefore, tracking the spatiotemporal variation of snow cover conditions is useful in predicting species distributions, especially in their seasonal changes. Further, snow regimes are projected to be altered across many ecosystems under climate change, which can have considerable impact on species distribution.

NASA offers two products - Terra and Aqua MODIS snow cover collections - that detect reflective signals from snow using the Normalized Difference Snow Index (NDSI). We downloaded and tested these two products to explore areas for recommendations. First, we conducted a search using a rectangle polygon around the state of Washington as the spatial extent boundary. Second, because of the broad temporal range that the collections cover (i.e., data coverage since February 2000 and July 2002 for Terra and Aqua, respectively), we limited our search to 1st January, 2021 to 15th January, 2021. Multiple collections of different temporal resolutions (i.e., 5-minute, daily, 8-day, monthly, etc.) are available under the umbrella of the Terra and Aqua snow cover products. Overall, we downloaded one grid from each of these collections for testing. These collections included:

Table 7. MODIS snow cover products relevant to the Species Distribution Modeling community.

MODIS Snow Cover Products	DAAC Link	Short Name
Aqua Snow Cover Daily L3 Global 500m SIN Grid V061	NSIDC	MYD10A1
Aqua Snow Cover Daily L3 Global 0.05Deg CMG V061	NSIDC	MYD10C1
Aqua Snow Cover 5-Min L2 Swath 500m V061	NSIDC	MYD10_L2
Aqua Snow Cover 8-Day L3 Global 500m SIN Grid V061	NSIDC	MYD10A2
Aqua Snow Cover Monthly L3 Global 0.05Deg CMG V061	NSIDC	MYD10CM
Aqua CGF Snow Cover Daily L3 Global 500m SIN Grid V061	NSIDC	MYD10A1F
Aqua Snow Cover 8-Day L3 Global 0.05Deg CMG	NSIDC	MYD10C2
Terra Snow Cover Daily L3	NSIDC	MOD10A1

Global 500m SIN Grid V061		
Terra Snow Cover 8-Day L3 Global 500m SIN Grid V061	NSIDC	MOD10A2
Terra Snow Cover 5-Min L2 Swath 500m V061	NSIDC	MOD10_L2
Terra Snow Cover Monthly L3 Global 0.05Deg CMG V061	NSIDC	MOD10CM
Terra CGF Snow Cover Daily L3 Global 500m SIN Grid V061	NSIDC	MOD10A1F
Terra Snow Cover Daily L3 Global 0.05Deg CMG V061	NSIDC	MOD10C1
Terra Snow Cover 8-Day L3 Global 0.05Deg CMG V061	NSIDC	MOD10C2

Recommendations
Because of the fine temporal resolution and broad temporal coverage that the Terra and Aqua snow cover products provide, the two products can be especially useful in developing seasonal distribution models as well as in detecting changes in their distribution for species whose movement patterns and occupancy change with the presence of snow. Spatially, the two products provide global coverage, making it superior to similar products that estimate snow cover. For example, USGS offers a LANDSAT-based snow cover product (i.e., Landsat Collection 1 Level-3 Fractional Snow Covered Area Science Product), but spatially it only covers the western U.S. Therefore, researchers who study locations beyond the western U.S. will find the Terra and Aqua products particularly useful. Also, the LANDSAT snow cover collection only produces an image about every two weeks, making the Terra and Aqua products more useful to researchers who demand a finer temporal resolution for their projects.

However, the USGS's snow cover collection offers a finer spatial resolution at 30m x 30m pixel, which is a huge improvement over Terra and Aqua's 500m x 500m resolution. Therefore, researchers who study species in the western U.S. will likely be using the USGS product for assessing snow cover. On the other hand, the Terra and Aqua products might be better in assessing species distribution at a broader scale (e.g., regional or global) for faster computational time.

Finally, we noticed some obvious differences in the NDSI values between the Terra and the Aqua layers even for data from the same day, along with some potential spatial misalignments. We know that Terra passes the equator in the morning, while Aqua passes the equator in the afternoon. However, it is not very likely that snow cover changes that rapidly and dramatically within such a short period of time across a broad landscape. We think perhaps the abundance, thickness, and spatial variation of clouds might be causing some of the differences. We recommend developers of these products to investigate the causes of these differences so that users can have greater confidence in using these products.

NLDAS-2 feedback
Differences between the numerous datasets with the same primary name (NLDAS-2) are unclear. There are 26 matching collections for NLDAS-2 on Earthdata because of different temporal aggregation methods or the type of land surface model used for simulation.

Helpful tools

- GES DISC subsetter
  - allows users to subset data easily and export in multiple formats without any scripting
- GIOVANNI
  - Useful for downloading time averaged maps of variables
- Google Earth Engine
  - NLDAS-2 Forcing data is hosted in GEE which allows for easy manipulation and export of the data for users with GEE/scripting experience

Challenges

- Virtually no tutorials available/easily accessible for working with this dataset
- Data is most often in GRIB or NetCDF format without documentation on extracting variables of interest
- Coarse scale may be a limiting factor for some practitioners
- Large number of similar datasets and technical language in the dataset names may discourage some practitioners

Human disturbance variables

Working group: Natalie Burgos, Kelly Easterday, Xiao Feng, Patrick Jantz, Giovanni Rapacciuolo, Adam B. Smith

A brief literature review on relationships between remote sensing human disturbance variables and habitat/species ranges was conducted for this report. Schirmer et al. (2019) combined Black Marble imagery with ground-level illumination measurements throughout Chicago to look at nighttime light influence on animals' activity levels in the city. Ditmer et al. (2021) used data from the VIIRS suite to estimate loss and fragmentation of dark environments within the ranges of 351 mammals in the United States. Wilson et al. (2021) developed models showing that light pollution (based on monthly averages from the VIIRS Day/Night Band) negatively impacts bird behavior across North America.

In general, the breakout group found connections between NASA products, human disturbance variables, and species distribution modeling to be limited. For example, the [Data Pathfinder]() for Species Distribution modeling does not list human disturbance as a data category in the main area on the webpage where raster-based datasets available for model input are provided; however, a link to "Find Human Impact Data" is called out in a separate callbox elsewhere on the webpage (Figure 8). This link redirects to surface reflectance, nighttime lights, and socioeconomic data. The subgroup overlooked this link and recommends including "Human Impacts" as a data category under the list of raster-based datasets available for model input. Currently, the closest category listed would likely be "Land Cover Type/Dynamics". Another note is that, in the absence of guidelines, it would be difficult for a beginner to intermediate practitioner to determine how to incorporate reflectance data into an SDM in an informative way.

The subgroup notes that many relevant products related to human disturbance may be encompassed by other high-level categories and suggests organizing potential products according to the [IUCN threats classification]() (Salafsky et al. 2008):

Figure 8. Access to Human Disturbance variables on the Pathfinder Data webpage for Species Distribution Modeling.

- Development (e.g., nightlights, several urban cover data products)
- Agriculture
- Energy/mining
- Transportation
- Biological resource use
- Human intrusion/disturbance
- Natural systems modification (fire, dams, etc.)
- Invasives, non-native genes, etc. (examples using airborne sensors exist)
- Pollution (e.g., MODIS particulate matter)
- Natural events (covered by other themes but in cases may be precipitated by humans; e.g., landslides)
- Climate (not in this theme)

Relevant datasets identified:

Nighttime lights

- DMSP/OLS (1989, but especially 1992-2013, 3 km)
- Landsat 8 OLI sensor (2013 onward)
- Black Marble (VIIRS instrument on the Suomi National Polar-orbiting Partnership (NPP) satellite since 2011/2012, 30 m, daily to annual)

Urban-specific

- Global Human Settlement Layer (Sentinel-2, 10 m)
- Global Urban Footprint (TerraSAR-X and TanDEM-X, ~12 m)

General

- GlobeLand30 (Landsat 4 and 7 and HJ-1, 30 m): “Artificial surfaces” land class type
- Anthropogenic Biomes (MODIS IGBP land cover, Landscan population density, ~10 km)
- Human Footprint (MODIS UMD cropland, GlobCover croplands, Gridded population of the world, gROADS, 1 km)
- Global Human Modification of Terrestrial Ecosystems (Global Human Settlement, OSM roads, 1 km)
- Global accessibility indicators (https://www.nature.com/articles/s41597-019-0265-5, Global Human Settlement, MODIS MCD12Q1 land cover, OSM, ~1km)

Pollution

- MODIS aerosol optical depth (for night pollution) and PACE

Black Marble feedback

Issues/challenges

1. The landing page provides a great overview, but differences among products (near the bottom) are not explained; for example, the difference between VNP46A1 and VNP46A2, which have the same spatial/temporal resolution, operational time, etc. Likewise, how is Black Marble NRT different from VNP46A2?
2. Provide SDM user-ready versions of Black Marble that aggregate annual/seasonal/monthly values across all available years (like WorldClim).
3. Data format can be a barrier. Black Marble comes in HDF format, the script to convert to TIFF is not straightforward.

Table 8. List of Black Marble products relevant to the Species Distribution Modeling community.

Black Marble Products	DAAC Link	Short Name
Black Marble VNP46A1	LAADS DAAC	VNP46A1
Black Marble VNP46A2	LAADS DAAC	VNP46A2
Black Marble NRT (daily)	GSFC MODAPS	VNP46A1_NRT
Black Marble NRT (hourly)	GSFC MODAPS	VNP46A1G_NRT

Recommendations
This dataset would benefit greatly from better access and aggregation tools, similar to other optical, high temporal resolution products. While HDF formats are convenient for many applications, they are not ideal for usability in the SDM community. Converting to GeoTIFF via a Python script works for some users, but with R being the lingua franca of ecologists, this is likely a barrier to use. Adding onto a platform like AppEARS or Google Earth Engine would help alleviate some of these issues. Additionally, further technical details on how to interpret each dataset would be useful so that ecologists could identify the product that suits their needs.

Common issues across products and potential solutions

Issue

Users may have limited technical understanding of products due to a lack of tutorials on product usage.

Suggestions

1. Provide high-level information on what the data represent

2. Provide assessment of strengths and weaknesses of different datasets

3. Provide options for a tutorial when accessing a product

Issue

There are a large number of products, many of which are closely related with little to no specification on the differences between them and the data they represent or how they are best used.

Suggestions

1. Provide a list of related datasets that might be relevant for certain applications.

2. Develop a dichotomous key-like set of filters for dataset filtering.

3. Identify datasets that currently might be disconnected but are very related in terms of information content.

4. Refine the set of filters on the DAAC e.g. spatial resolution, temporal resolution, dataset characteristics.

Issue

Citizen science data are expanding but are frequently disconnected from Earth observation datasets.

Suggestions

Link citizen science data with NASA products in data portals, with a possible note or filter to indicate what data were collected through citizen science so that a user may be able to select to include that data or not.

Other issues

- Data may exist [in Earthdata] but variables are sometimes separated/at incompatible resolutions and cannot be easily blended for immediate/rapid use.
- Data format can be a barrier (e.g. Black Marble comes in HDF format, the script to convert to TIFF is not straightforward).
- It is not always clear what different versions of the same dataset comprise.
- Human disturbance is a very broad theme. Sometimes it is unclear what aspect of human disturbance the data capture and if they will be useful in SDM.
- Time period or resolution may not be relevant for a given SDM application (e.g. standardized nighttime lights from 1992-present would expand applicability).
- Products may require additional processing/aggregation to be relevant (e.g. users may desire summary across all years, seasonal across all years, monthly averages across all years akin to WorldClim).
- Product comparison: it would be useful to further test differences in using nighttime lights vs NLCD impervious surface.
- There is a bulk archive download feature for Active Fires, but limited control downloading the Monitoring Trends in Burn Severity datasets.

General comments and recommendations

- Consider building a use case where species of interest have been targeted for modeling (e.g. coldwater fish); provide mechanisms for identifying relevant datasets and easy access to these layers.
- It would be ideal to have a “one stop shop” for data access where data are processed in a common format. As would being able to access those datasets through R, GEE, or other software that are frequently used
- It would be useful to have suggestions provided within the dataset for sensible pixel quality/QAQC thresholds to grab the best data for SDM applications.
- A guidebook for SDM users selecting NASA datasets could be useful. For example, which datasets are best for which types of applications?
- Enable Earthdata searches that can be filtered by spatial resolution and tags in order to help find or discover the best data for a given study area.
- A meta-literature search can be conducted by a few PhD students to determine different tag categories (see Zellweger et al. 2019).

REFERENCES

Abrahms, B., Welch, H., Brodie, S., et al., 2019. Dynamic ensemble models to predict distributions and anthropogenic risk exposure for highly mobile species. *Diversity and Distributions* 25(8):1182–1193. https://doi.org/10.1111/ddi.12940.

Anand, A., Pandey, M.K., Srivastava, P.K., Gupta, A., and Latif Khan, M., 2021. Integrating multi-sensors data for species distribution mapping using deep learning and envelope models. *Remote Sens*. 13(16): 3284. https://doi.org/10.3390/rs13163284

Barlow, D.R., and Torres, L.G., 2021. Planning ahead: dynamic models forecast blue whale distribution with applications for spatial management. *Journal of Applied Ecology* 58:2493–2504. https://doi.org/10.1111/1365-2664.13992.

Bedriñana-Romano, L., Hucke-Gaete, R., Viddi, F.A., et al., 2018. Integrating multiple data sources for assessing blue whale abundance and distribution in Chilean Northern Patagonia. *Diversity and Distributions* 24:991–1004. https://doi.org/10.1111/ddi.12739.

Bedriñana-Romano, L., Hucke-Gaete, R., Viddi, F.A., et al., 2021. Defining priority areas for blue whale conservation and investigating overlap with vessel traffic in Chilean Patagonia, using a fast-fitting movement model. *Scientific Reports* 11:2709. https://doi.org/10.1038/s41598-021-82220-5.

Bradley, B.A., Olsson, A.D., Wang, O., Dickson, B.G., Pelech, L., Sesnie, S.E. and Zachmann, L.J., 2012. Species detection vs. habitat suitability: are we biasing habitat suitability models with remotely sensed data? *Ecological Modelling* 244:57-64. https://doi.org/10.1016/j.ecolmodel.2012.06.019.

Brocca, L., Hasenauer, S, Lacava, T., et al., 2011. Soil moisture estimation through ASCAT and AMSR-E sensors: an intercomparison and validation study across Europe. *Remote Sensing of Environment* 115:3390-3408. http://dx.doi.org/10.1016/j.rse.2011.08.003.

Burns, P., Clark, M., Salas, L., Hancock, S., Leland, D., Jantz, P., Dubayah, Rl., and Goetz, S.J., 2020. Incorporating canopy structure from simulated GEDI lidar into bird species distribution models. Environmental Research Letters 15:095002. https://doi.org/10.1088/1748-9326/ab80ee.

Chaney, N.W., Wood, E.F., McBratney, A.B., Hempel, J.W., Nauman, T.W., Brungard, C.W., and Odgers, N.P., 2016. POLARIS: A 30-meter probabilistic soil series map of the contiguous United States. *Geoderma* 274:54–67. https://doi.org/10.1016/j.geoderma.2016.03.025.

Clark, J.S., Nemergut, D., Seyednasrollah, B.,Turner, P.J., and Zhang, S.,, 2017. Generalized joint attribute modeling for biodiversity analysis: median-zero, multivariate, multifarious data. *Ecological Monographs* 87(1): 34-56. https://doi.org/10.1002/ecm.1241.

Cord, A., Klein, D., Mora, F., Dech, S. 2014. Comparing the suitability of classified land cover data and remote sensing variables for modeling distribution patterns of plants. *Ecological Modelling*: 272(2014) 129-140. https://doi.org/10.1016/j.ecolmodel.2013.09.011

Dahal D, Pastick NJ, Boyte SP, Parajuli S, Oimoen MJ, Megard LJ. Multi-Species Inference of Exotic Annual and Native Perennial Grasses in Rangelands of the Western United States Using Harmonized Landsat and Sentinel-2 Data. *Remote Sensing*. 2022; 14(4):807. https://doi.org/10.3390/rs14040807.

Daly, C., Halbleib, M., Smith, J.I., Gibson, W.P., Doggett, M.K., Taylor, G.H., Curtis, J., and Pasteris, P.A., 2008. Physiographically sensitive mapping of temperature and precipitation across the conterminous United States. *International Journal of Climatology* 28:2031-2064. https://doi.org/10.1002/joc.1688.

Das, P.K., Mohinuddin, S.K., Midya, S.K., Das, D.K., Sharma, R., and Bandyopadhyay, S., 2021. Can multiscalar meteorological drought indices detect soil moisture droughts? A study of Indian regions. *Hydrological Sciences Journal* 66(9):1475-1487. https://doi.org/10.1080/02626667.2021.1942475.

Deblauwe, V., Droissart, V., Bose, R., Sonke, B., Blach-Overgaard, A., Svenning, J.-C., Wieringa, J.J., Ramesh, B.R., Stewart, T, and Couvreur, L.P., 2016. Remotely sensed temperature and precipitation data improve species distribution modelling in the tropics. *Global Ecology and Biogeography* 25(4):443-454. https://doi.org/10.1111/geb.12426

Ditmer, M.A., Stoner, D.C., and Carter, N.H., 2021. Estimating the loss and fragmentation of dark environments in mammal ranges from light pollution. *Biological Conservation* 257:109135. https://doi.org/10.1016/j.biocon.2021.109135.

Draper, C.S., Walker, J.P., Steinle, P.J., de Jeu, R.A.M., Holmes, T.R.H., 2008. An evaluation of AMSR-E derived soil moisture over Australia. *Remote Sensing of the Environment* 113:703-710. http://dx.doi.org/10.1016/j.rse.2008.11.011.

Elith, J., 2017. Predicting distributions of invasive species. *Invasive species: Risk assessment and management*, pp. 93-129. https://doi.org/10.1017/9781139019606.006.

Elith, J. and Leathwick, J., 2009. Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics* 40:677–697. http://dx.doi.org/10.1146/annurev.ecolsys.110308.120159.

Elith, J., H. Graham*, C., P. Anderson, R., Dudík, M., Ferrier, S., Guisan, A., J. Hijmans, R., Huettmann, F., R. Leathwick, J., Lehmann, A. and Li, J., 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29(2): 129-151. https://doi.org/10.1111/j.2006.0906-7590.04596.x

Fick, S. E. and Hijmans, R. J. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* 37(12): 4302-4315. https://doi.org/10.1002/joc.5086

Forgotson, C., O'Neill, P.E., Carrera, M.L., et al., 2020. How satellite soil moisture data can help to monitor the impacts of climate change: SMAP case studies. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 12:1590-1596. https://doi.org/10.1109/JSTARS.2020.2982608.

Frasier, K.E., Garrison, L.P., Soldevilla, M.S., Wiggins, S.M., and Hildebrand, J.A., 2021. Cetacean distribution models based on visual and passive acoustic data. *Scientific Reports* 11:8240. https://doi.org/10.1038/s41598-021-87577-1

Gavrutenko, M., Gerstner, B.E., Kass, J.M., Goodman, S.M. and Anderson, R.P., 2021. Temporal matching of occurrence localities and forest cover data helps improve range estimates and predict climate change vulnerabilities. *Global Ecology and Conservation* 27:e01569. https://doi.org/10.1016/j.gecco.2021.e01569

Gaston, K.J., 2003. *The structure and dynamics of geographic ranges*. Oxford University Press on Demand.

Gelaro, R., McCarthy, W., Suárez, M.J., Todling, R., Molod, A., Takacs, L., Randles, C.A., Darmenov, A., Bosilovich, M.G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A.M., Gu, W., Kim, G.-K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J.E., Partyka, G., Pawson, S., Putman, W., Rienecker. M., Schubert, S.D., Sienkiewicz, M., and Zhao, B., 2017. The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *Journal of Climate* 30(14):5419-5454. https://doi.org/10.1175/JCLI-D-16-0758.1.

Goldman, A. E.,Emani, S. R.,Pérez-Angel, L. C.,Rodríguez-Ramos, J. A.,Stegen, J. C., and Fox, P., 2021, Special collection on open collaboration across geosciences. *Eos* 102, published on 06 January 2021. https://doi.org/10.1029/2021EO153180.

Gómez, D, Salvador, P., Sanz, J., Casanova, C., Taratiel, D., and Casanova, J.L,. 2018. Machine learning approach to locate desert locust breeding areas based on ESA CCI soil moisture. *Journal of Applied Remote Sensing* 12(3):036011. https://doi.org/10.1117/1.JRS.12.036011.

Gómez, D., Salvador, P., Sanz, J., Casanova, C., Taratiel, D., and Casanova, J.L, 2019. Desert locust detection using Earth observation satellite data in Mauritania. *Journal of Arid Environments* 164:29-37. https://doi.org/10.1016/j.jaridenv.2019.02.005.

Hakkenberg, C.R., K. Zhu, R.K. Peet and C. Song. (2018). Mapping multi-scale vascular plant richness in a forest landscape with integrated LiDAR and hyperspectral remote-sensing. *Ecology* 99(2):474-487. https://doi.org/10.1002/ecy.2109.

Hansen, A.J., Noble, B.P., Veneros, J., East, A., Goetz, S.J., Supples, C., Watson, J.E., Jantz, P.A., Pillay, R., Jetz, W. and Ferrier, S., 2021. Toward monitoring forest ecosystem integrity within the post□ 2020 Global Biodiversity Framework. *Conservation Letters* 14(4):e12822. https://doi.org/10.1111/conl.12822

He, K.S., Bradley, B.A., Cord, A.F., Rocchini, D., Tuanmu, M.N., Schmidtlein, S., Turner, W., Wegmann, M. and Pettorelli, N., 2015. Will remote sensing shape the next generation of species distribution models?. *Remote Sensing in Ecology and Conservation* 1(1): 4-18. https://doi.org/10.1002/rse2.7.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., and Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25(15):1965-1978. https://doi.org/10.1002/joc.1276.

Fick, S.E. and Hijmans, R.J., 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* 37(12):4302-4315. https://doi.org/10.1002/joc.5086.

Ines, A.V.M., Das, N.N., Hansen, J.W., and Njoku, E.G., 2013. Assimilation of remotely sensed soil moisture and vegetation with a crop simulation model for maize yield prediction. *Remote Sensing of Environment* 138:149-164. https://doi.org/10.1016/j.rse.2013.07.018.

Jarnevich, C.S., Belamaric, P.N., Fricke, K., Houts, M., Rossi, L., Beauprez, G., Cooper, B. and Martin, R., 2021. Challenges in updating habitat suitability models: An example with the lesser prairie-chicken. *PloS one* 16(9):e0256633. https://doi.org/10.1371/journal.pone.0256633.

Jetz, W. and Rahbek, C., 2002. Geographic range size and determinants of avian species richness. *Science* 297(5586):1548-1551. https://doi.org/10.1126/science.1072779.

Karger, D.N., Conrad, O., Bohner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., Zimmerman, N.E., Linder, P. Kessler, M. 2017. Climatologies at high resolution for the Earth land surface areas. *Scientific Data* 4 (170122). https://doi.org/10.1038/sdata.2017.122

Kass JM, Vilela B, Aiello□Lammens ME, Muscarella R, Merow C, Anderson RP. 2018. Wallace: A flexible platform for reproducible modeling of species niches and distributions built for community expansion. *Methods in Ecology and Evolution* 9:1151–1156. https://doi.org/10.1111/2041-210X.12945

Lai, J., Lortie, C.J., Muenchen, R.A., Yang, J., Ma, K. Evaluating the popularity of R in ecology. *Ecosphere* 10(1):e02567. https://doi.org/10.1002/ecs2.2567

Lauer, I., Coats, D., Kucera, L., Sforzo, Z., and Broddle, M., 2018. Estimating soil moisture in semiarid sagebrush steppe utilizing NASA satellite imagery [Technical Report]. Idaho Water Resources, Pocatello.

Lembrechts, J.J., Aalto, J., Ashcroft, M.B., et al., 2020. SoilTemp: A global database of near-surface temperature. *Global Change Biology* 26(11):6616-6629. https://doi.org/10.1111/gcb.15123.

Lowndes, J.S.S., Froehlich, H.E., Horst, A., Jayasundara, N., Pinsky, M.L., Stier, A.C., Therkildsen, N.O. and Wood, C.L., 2019. Supercharge your research: a ten-week plan for open data science. *Nature*. https://doi.org/10.1038/d41586-019-03335-4.

Maclean, I.M.D., Mosedale, J.R., and Bennie, J.J., 2018. Microclima: An R package for modeling meso- and microclimate. *Methods in Ecology and Evolution* 10(2):280-290. https://doi.org/10.1111/2041-210X.13093.

Maclean, I.M.D., 2019. Predicting future climate at high spatial and temporal resolution. *Global Change Biology* 26(2):1003-1011. https://doi.org/10.1111/gcb.14876.

Mason, J.G., Alfaro-Shigueto, J., Mangel, J.C., et al., 2019. Convergence of fishers' knowledge with a species distribution model in a Peruvian shark fishery. *Conservation Science and Practice* 1(4):e13. https://doi.org/10.1111/csp2.13.

Miller, J., 2010. Species distribution modeling. *Geography Compass* 4(6):490-509. https://doi.org/10.1111/j.1749-8198.2010.00351.x.

Modanesi, S., Massari, C., Camici, S., Brocca, L., and Amarnath, G., 2020. Do satellite surface soil moisture observations better retain information about crop□yield variability in drought conditions? *Water Resources Research* 56:e2019WR025855. https://doi.org/10.1029/2019WR025855.

Moullec, F., Barrier, N., Drira, S., Guilhaumon, F., Hattab, T., Peck, M.A. and Shin, Y.J., 2022. Using species distribution models only may underestimate climate change impacts on future marine biodiversity. *Ecological Modelling* 464:109826. https://doi.org/10.1016/j.ecolmodel.2021.109826

Pinto-Ledezma, J.N., and Cavender-Bares, J., 2021. Predicting species distributions and community composition using satellite remote sensing predictors. *Scientific Reports* 11:16448. https://doi.org/10.1038/s41598-021-96047-7.

Poggio, L., Simonetti, E., and Gimona, A., 2018. Enhancing the WorldClim data set for national and regional applications. *Science of the Total Environment* 625:1628-1643. https://doi.org/10.1016/j.scitotenv.2017.12.258.

Porfirio, L.L., Harris, R.M., Lefroy, E.C., Hugh, S., Gould, S.F., Lee, G., Bindoff, N.L. and Mackey, B., 2014. Improving the use of species distribution models in conservation planning and management under climate change. *PLoS One* 9(11):e113749. https://doi.org/10.1371/journal.pone.0113749.

Rizza, U., Mancinelli, E., Morichetti, M., Passerini, G.,and Virgili, S., 2019. Aerosol optical depth of the main aerosol species over Italian cities based on the NASA/MERRA-2 model reanalysis. *Atmosphere* 10:709. https://doi.org/10.3390/atmos10110709.

Rogers, D.J., 2006. Models for vectors and vector-borne diseases. *Advances in parasitology* 62:1-35. https://doi.org/10.1016/S0065-308X(05)62001-5.

Salafsky, N., Salzer, D., Stattersfield, A.J., Hilton-Taylor, C., Neugarten, R., Butchart, S.H.M., Collen, B., Cox, N., Master, L.L., O'Connor, S., and Wilkie, D., 2008. A standard lexicon for biodiversity conservation: unified classifications of threats and actions. *Conservation Biology* 22(4):897-911. https://doi.org/10.1111/j.1523-1739.2008.00937.x.

Sbrocco, E.J. and Barber, P.H., 2013. MARSPEC: ocean climate layers for marine spatial ecology: Ecological Archives E094□086. *Ecology* 94(4):979-979. https://doi.org/10.1890/12-1358.1.

Schirmer, A.E., Gallemore, C., Liu, T., Magle, S., DiNello, E., Ahmed, H., and Gilday, T., 2019. Mapping behaviorally relevant light pollution levels to improve urban habitat planning. *Scientific Reports* 9:11925. https://doi.org/10.1038/s41598-019-48118-z.

Schmitt, S., Pouteau, R., Justeau, D., De Boissieu, F. and Birnbaum, P., 2017. ssdm: An r package to predict distribution of species richness and composition based on stacked species distribution models. *Methods in Ecology and Evolution* 8(12):1795-1803. https://doi.org/10.1111/2041-210X.12841.

Schwager, P., and Berg, C., 2021. Remote sensing variables improve species distribution models for alpine plant species. Basic and Applied Ecology 54:1-13. https://doi.org/10.1016/j.baae.2021.04.002.

Servilla, M., Brunt, J., Costa, D., McGann, J. and Waide, R., 2016. The contribution and reuse of LTER data in the Provenance Aware Synthesis Tracking Architecture (PASTA) data repository. *Ecological Informatics* 36:247-258. https://doi.org/10.1016/j.ecoinf.2016.07.003.

Sexton, J.O., Song, X.P., Feng, M., Noojipady, P., Anand, A., Huang, C., Kim, D.H., Collins, K.M., Channan, S., DiMiceli, C. and Townshend, J.R., 2013. Global, 30-m resolution continuous fields of tree cover: Landsat-based rescaling of MODIS vegetation continuous fields with lidar-based estimates of error. *International Journal of Digital Earth*, 6(5):427-448. https://doi.org/10.1080/17538947.2013.786146.

Stoklosa, J., Daly, C., Foster, S.D., Ashcroft, M.B, and Warton, D.I., 2015. A climate of uncertainty: accounting for error in climate variables for species distribution models. *Methods in Ecology and Evolution* 6(4):412-423. https://doi.org/10.1111/2041-210X.12217

Ukhov, A., Mostamandi, S., Krotkov, N., Flemming, J., da Silva, A., Li, C., Fioletov, V., McLinden, C., Anisimov, A., Alshehri, Y.M., and Stenchikov, G., 2020. Study of SO_2 pollution in the Middle East using MERRA-2, CAMS data assimilation products, and high-resolution WRF-chem simulations. *JGR Atmospheres* 125(6):e2019JD031993. https://doi.org/10.1029/2019JD031993.

Vega, G.C., Pertierra, L.R., and Olalla-Tárraga, M.A., 2017. MERRAclim, a high-resolution global dataset of remotely sensed bioclimatic variables for ecological modelling. *Scientific Data* 4:170078. https://doi.org/10.1038/sdata.2017.78.

Waltari, E., Schroeder, R., McDonald, K., Anderson, R.P., and Carnaval, A., 2014. Bioclimatic variables derived from remote sensing: assessment and application for species distribution modelling. *Methods in Ecology and Evolution* 5:1033–1042. https://doi.org/10.1111/2041-210X.12264.

Wang, L., Kerr, L.A., Record, N.R., et al., 2018. Modeling marine pelagic fish species spatiotemporal distributions utilizing a maximum entropy approach. *Fisheries Oceanography* 27(6):571-586. https://doi.org/10.1111/fog.12279.

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al., 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3:160018. https://doi.org/10.1038/sdata.2016.18.

Wilson, A.A., Ditmer, M.A., Barber, J.R., Carter, N.H., Miller, E.T., Tyrell, L.P., and Francis, C.D., 2021. Artificial night light and anthropogenic noise interact to influence bird abundance over a continental scale. *Global Change Biology* 27(17):3987-4004. https://doi.org/10.1111/gcb.15663.

Wisz, M.S., Hijmans, R.J., Li, J., Peterson, A.T., Graham, C.H., Guisan, A. and NCEAS Predicting Species Distributions Working Group, 2008. Effects of sample size on the performance of species distribution models. *Diversity and distributions* 14(5):763-773. https://doi.org/10.1111/j.1472-4642.2008.00482.x.

Xu, X., Frey, S.,and Ma, D., 2022. Hydrological performance of ERA5 and MERRA-2 precipitation products over the Great Lakes Basin. *Journal of Hydrology: Regional Studies* 39:100982. https://doi.org/10.1016/j.ejrh.2021.100982.

Zaussinger, F., Dorigo, W., Gruber, A., Tarpanelli, A., Filippucci, P., and Brocca, L., 2019. Estimating irrigation water use over the contiguous United States by combining satellite and reanalysis soil moisture data. *Hydrology and Earth System Sciences* 23:897–923, https://doi.org/10.5194/hess-23-897-2019.

Zellweger, F., De Frenne, P., Lenoir, J., Rocchini, D., and Coomes, D., 2019. Advances in microclimate ecology arising from remote sensing. *Trends in Ecology & Evolution* 34(4):327-341. https://doi.org/10.1016/j.tree.2018.12.012.

Zhu, L., and Guo, Y., 2022. Remotely sensed winter habitat indices improve the explanation of broad-scale patterns of mammal and bird species richness in China. *Remote Sens*. 14(3):794. https://doi.org/10.3390/rs14030794

APPENDIX A: Workshop announcement

NASA Topical Workshop:
Broadening the use of NASA datasets by the Species Distribution Modeling community

Partner

Call for Workshop Participants

The material contained in this document is based upon work supported by a National Aeronautics and Space Administration (NASA) grant or cooperative agreement. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of NASA.

The University of Montana Spatial Analysis Lab is hosting a *practitioner-driven* workshop to identify gaps in NASA data usage by species distribution modelers. We are recruiting applicants and survey respondents across a range of taxonomic groups, career levels, and backgrounds.

Workshop Goals

Participants will meet and work together virtually to identify NASA data products that capture important predictive distribution variables, but are under-used because of accessibility, interoperability, or other technological challenges.

To achieve this outcome, we are evaluating capabilities for accessing and processing static and time-series data products relevant to species distribution modeling.

Candidate variables will be prioritized based on spatiotemporal scales of observation, level of effort to modify products or technologies for dissemination and use, and potential for significantly boosting accuracy results for a wide range of species distribution models.

Workshop Details

- Session Dates (11:00 am – 1:30 pm MST): **January 6th, 11th, 13th, 20th, February 3rd and 4th**
- Fully virtual; $2500 stipends available for 12 participants
- Commitment to development of a Workshop Recommendation Report

Application Guidelines

Apply by responding to a brief online survey by November 24th, 2021.

Please consider sharing this notice and encouraging *practitioners* to apply or complete our short survey to identify important predictor variables. All applications and survey responses will be used to help inform the workshop and identify data needs in the species distribution modeling community.

The application and questionnaire is available at: https://bit.ly/nasa-workshop-survey

Email spatial.analysis@umontana.edu with any questions.

APPENDIX B: List of applicants (n = 87)

Seventy-seven people applied to participate in the workshop, and 10 (listed last) responded to provide information on their use of NASA datasets for Species Distribution Modeling. Some survey respondents selected “Other” as their organizational type and provided specific descriptions, e.g., “Other: prospective graduate student”.

Name	Organizational type
Jason Alstad	Non-profit organization
Alex Amoakoh	Academic institution (faculty, staff, student)
Ashley Anderson	Private sector commercial business
Caitlin Annear	Other: prospective graduate student
Sudeep Banad	Academic institution (faculty, staff, student)
Namrata Bhandari	Other: master’s graduate student
Rachel Bonafilia	State government (e.g., state agency)
Israel Borokini	Academic institution (faculty, staff, student)
Ericka Buckreis	National/Federal government
Natalie Burgos	Non-profit organization
Patrick Burns	Academic institution (faculty, staff, student)
Alexis Cambridge	Academic institution (faculty, staff, student)
Josh Carrell	Academic institution (faculty, staff, student)
Sean Carter	Academic institution (faculty, staff, student)
Rafael Castillo	Academic institution (faculty, staff, student)
Nathalie Chardon	Academic institution (faculty, staff, student)
Vimbai Chizarura	Private sector commercial business
MARIA CHRISTOU	Other: Research Institute
Angelito Cinco	Academic institution (faculty, staff, student)
Patrick Comer	Non-profit organization
Jackson Connors	Academic institution (faculty, staff, student)
Paige Copenhaver-Parry	Other: State Natural Heritage Program (WYNDD)

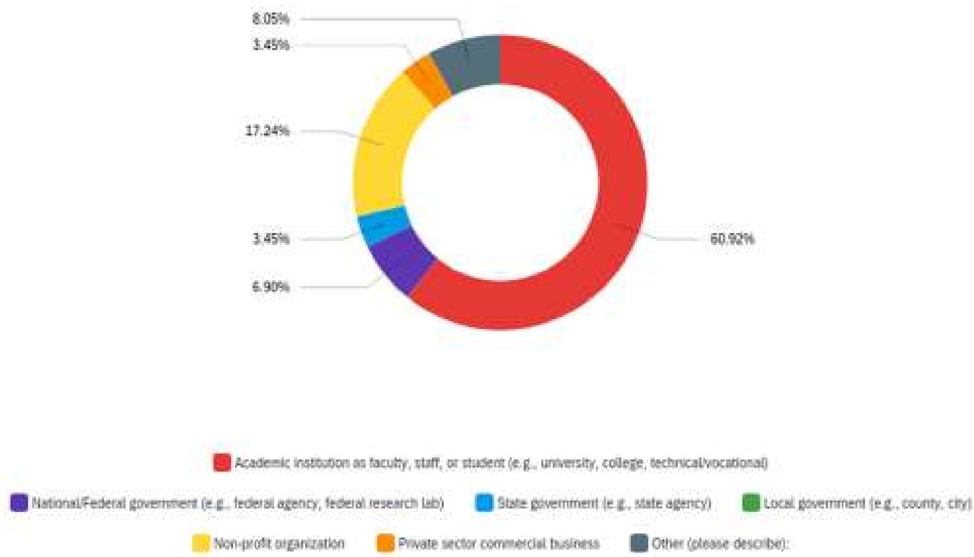
Name	Organizational type
Shawn Crimmins	National/Federal government
Andrea De las casas	Academic institution (faculty, staff, student)
Alexandra De Sousa	Academic institution (faculty, staff, student)
Kelly Easterday	Non-profit organization
Shira Ellenson	Academic institution (faculty, staff, student)
Helen Enander	Non-profit organization
Helen Enander	Non-profit organization
Xiao Feng	Academic institution (faculty, staff, student)
Matt Fitzpatrick	Academic institution (faculty, staff, student)
Allison Foster	Academic institution (faculty, staff, student)
Beth Gerstner	Academic institution (faculty, staff, student)
Joshua Goldberg	Non-profit organization
Reza Goljani Amirkhiz	Academic institution (faculty, staff, student)
Airy Gonzalez Peralta	Academic institution (faculty, staff, student)
Ericka Griggs	Academic institution (faculty, staff, student)
Robert Guralnick	Academic institution (faculty, staff, student)
Chris Hakkenberg	Academic institution (faculty, staff, student)
Paul Hillbrand	Non-profit organization
Tim Howard	Non-profit organization
Hyla Howe	Academic institution (faculty, staff, student)
Clarisse Ishimwe	Academic institution (faculty, staff, student)
Patrick Jantz	Academic institution (faculty, staff, student)
Catherine Jarnevich	National/Federal government
Cheng Jinxiang	Academic institution (faculty, staff, student)
Kate Kallenbach	Academic institution (faculty, staff, student)
Isaac Kissiedu	Non-profit organization
Abhishek Kumar	Academic institution (faculty, staff, student)

Name	Organizational type
Peter Lesica	Non-profit organization
Zhihua Liu	Academic institution (faculty, staff, student)
Covir Maju	Other: Individual
Sparkle Malone	Academic institution (faculty, staff, student)
Ellen Martin	Academic institution (faculty, staff, student)
Anne Mauro	Non-profit organization
Bryce Maxell	State government (e.g., state agency)
Bryan McLean	Academic institution (faculty, staff, student)
Cory Merow	Academic institution (faculty, staff, student)
Adam Miles	Academic institution (faculty, staff, student)
Austin Nash	Academic institution (faculty, staff, student)
Eric Nielsen	Academic institution (faculty, staff, student)
Fiona Noonan	Academic institution (faculty, staff, student)
Alexiana Pou	Academic institution (faculty, staff, student)
Yi Qi	Academic institution (faculty, staff, student)
Giovanni Rapacciuolo	Non-profit organization
Sangeeta Roy	Academic institution (faculty, staff, student)
Scott Rush	Academic institution (faculty, staff, student)
Skye Salganek	National/Federal government
Adam Smith	Non-profit organization
Ashish Suman	Academic institution (faculty, staff, student)
Erana Taylor	Academic institution (faculty, staff, student)
Romaric Tegang Pagning	Academic institution (faculty, staff, student)
Cavalcante Thiago	Academic institution (faculty, staff, student)
Cassondra Walker	National/Federal government
Ho Yi Wan	Academic institution (faculty, staff, student)
Xian Wang	Academic institution (faculty, staff, student)

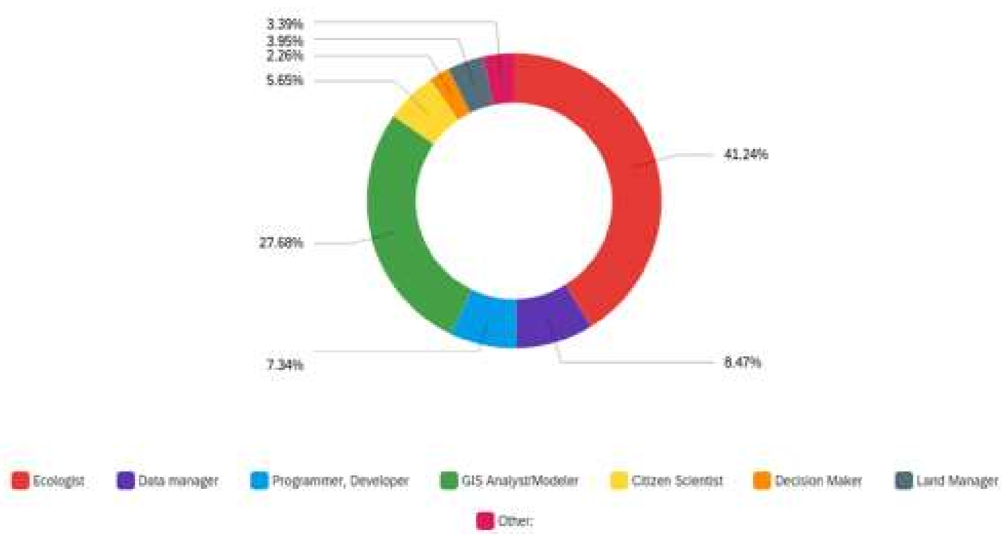
Name	Organizational type
Lu Zhai	Academic institution (faculty, staff, student)
[only email provided]	Non-profit organization
stephanie brodie	Academic institution (faculty, staff, student)
David Bucklin	State government (e.g., state agency)
Andy Cameron	Non-profit organization
Patrick Donnelly	National/Federal government
Scott Goetz	Academic institution (faculty, staff, student)
Mamuka Gvilava	Private sector commercial business
Chenzherui Liu	Academic institution (faculty, staff, student)
Raja Natarajan	Academic institution (faculty, staff, student)
Jaris Emmanuel Veneros Guevara	Academic institution (faculty, staff, student)

APPENDIX C: Selected response summaries from a survey on the use of NASA datasets in Species Distribution Modeling

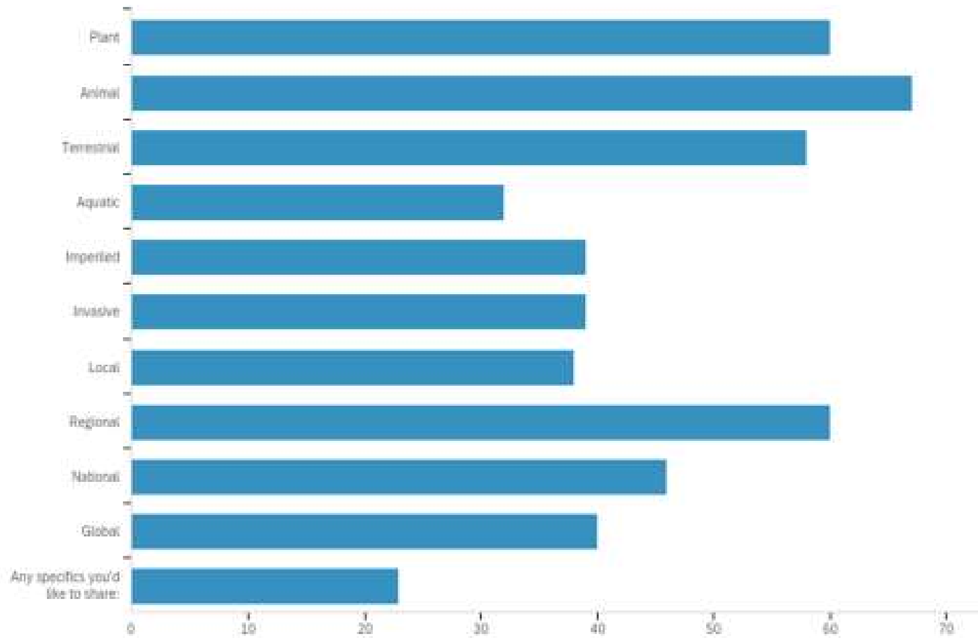
Survey responses:
Which of the following best describes your organization?



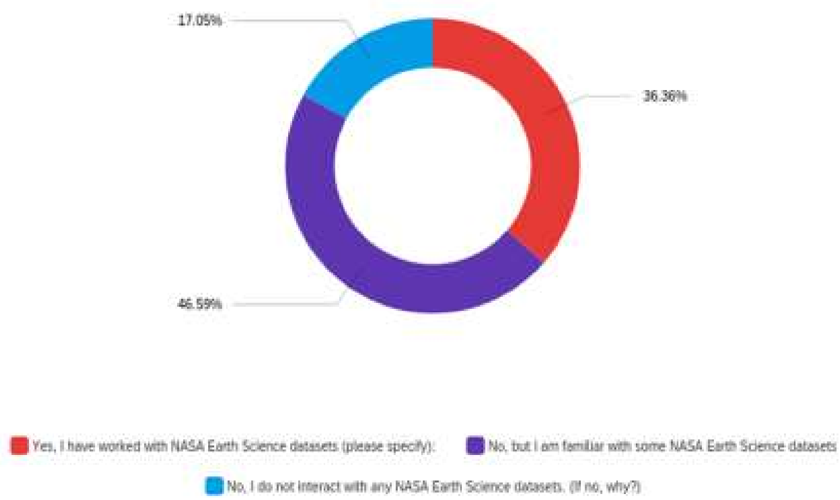
Survey responses:
In what capacity do you interact with Species Distribution Modeling datasets?



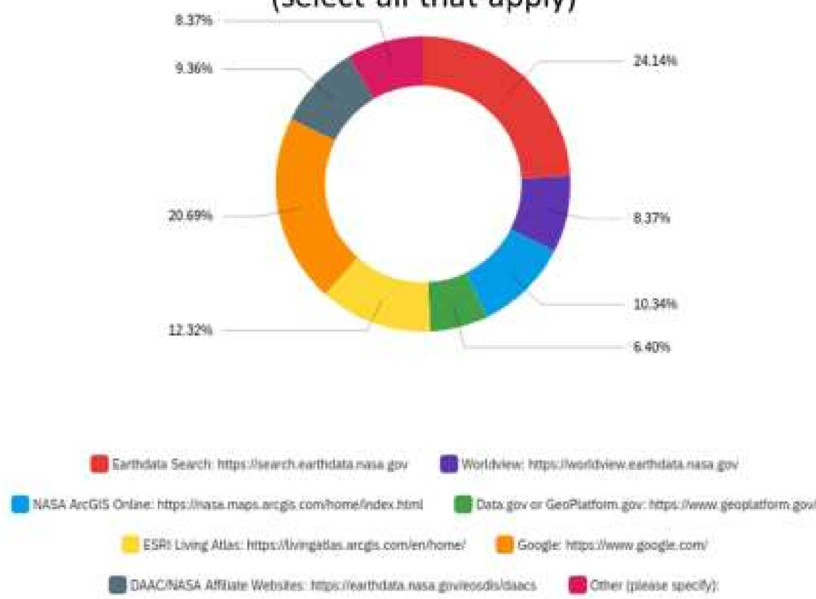
Survey responses: Please tell us a little bit about your Species Distribution Modeling interests (select all that apply)



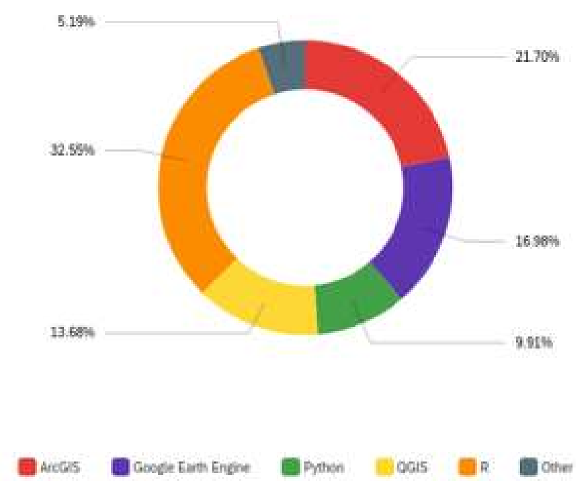
Survey responses:
Have you worked with NASA Earth Science datasets for Species Distribution Modeling?



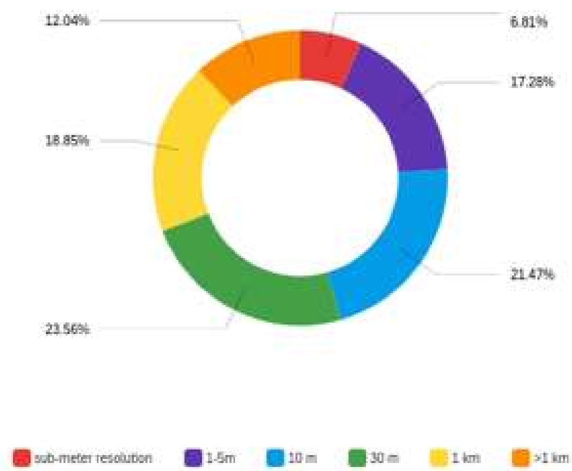
Survey responses: Where do you discover/access NASA data to use in a GIS tool? (select all that apply)



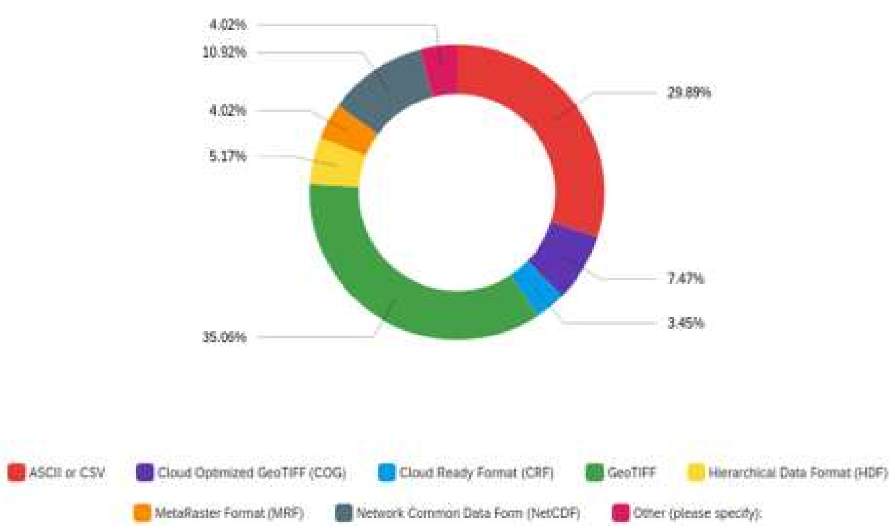
Survey responses:
How are you analyzing your data? (select all that apply)



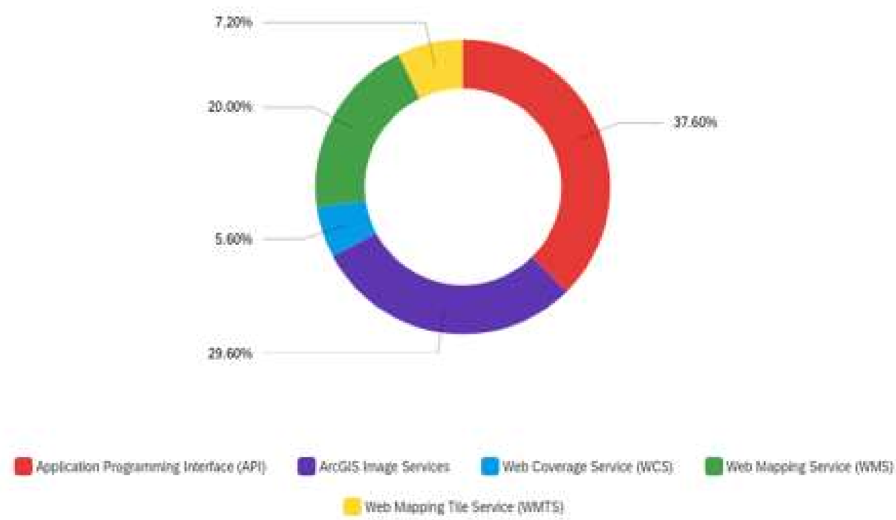
Survey responses:
What spatial resolution is most useful to you in your Species Distribution Modeling work? (select all that apply)



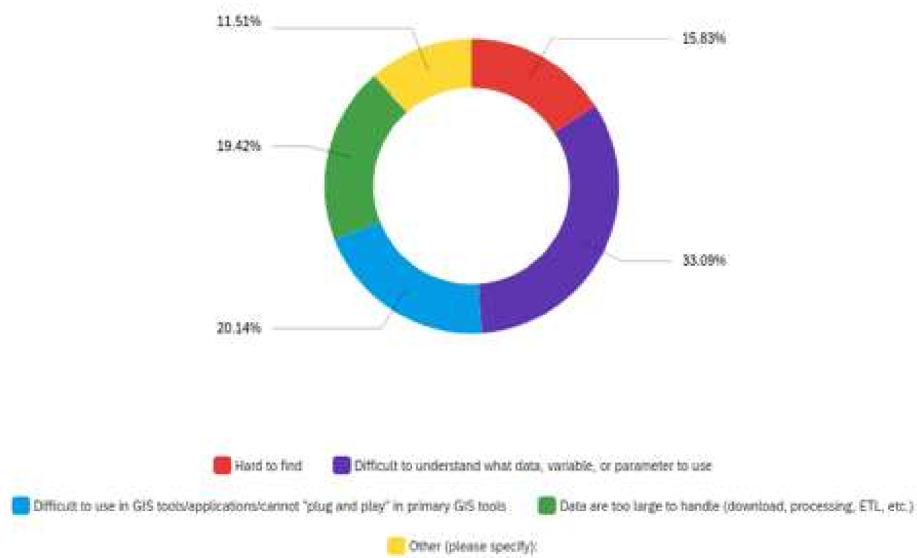
Survey responses:
What data formats do you prefer to work with? (select all that apply)



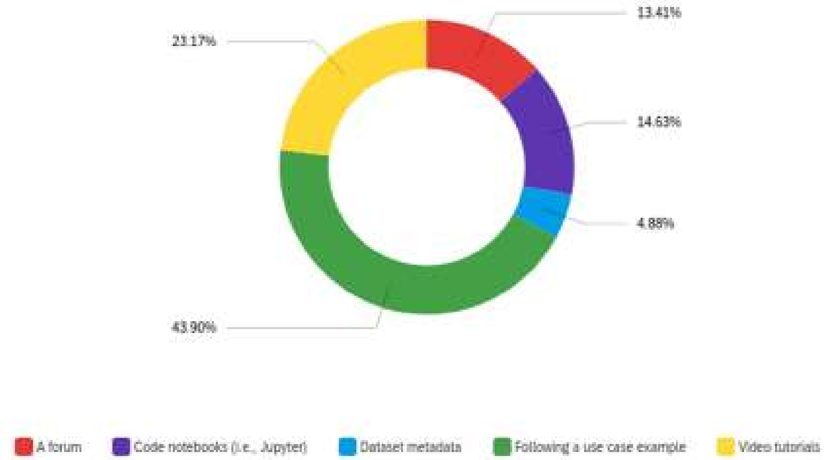
Survey responses:
Which web services types/formats do you prefer? (select all that apply)



Survey responses:
What challenges do you have in accessing and using NASA data? (select all)



Survey responses:
How do you prefer to learn about using NASA data in Species Distribution Modeling?



Important Variables

Survey question:
In your work with Species Distribution Modeling, what variables have you frequently identified as important predictors across multiple taxa or ecological boundaries?

- 73 people responded.
- Most people listed more than one variable, but many of those fell into the same group.
    - For example, if elevation and topographic position index were both listed, that counted as one for topographic variables.

FREQUENTLY IMPORTANT VARIABLES	CITED BY
Climate/weather	56
Topographic variables	28
Categorical land cover	22
Soils/geology	17
Anthropogenic variables, urban mapping, land use	16
Spectral indices (NDVI, EVI, NDMI...)	15
Vegetation characteristics (structure, % canopy...)	14
Remotely sensed data	9
Habitat	6
Disturbance	5
Hydrologic variables/stream morphology	5
Species occurrence data	4
Species ecology	4
Water quality	4
Land surface temperature	4
Surface water occurrence	3
Marine variables	3
Landscape metrics	3
Invasive species	2
Abiotic variables	2
Protected areas	2
Water availability	2
Restoration efforts	1
Nutrient availability	1
Field size	1
Genetic maps	1
Hardiness zones	1
Latitude	1

Wishlist Variables

Survey question:
In your work with Species Distribution Modeling, have you identified variables that you suspect would be important predictors, but are currently not included in models due to unavailability, complexity, binning, etc.?

- 59 people responded.
- Most people listed more than one variable, but many of those fell into the same group.
    - For example, if snow free dates and snow depth were both listed, that counted as one for climate/weather.
- Many people noted the need for finer-resolution data than may be currently available.

WISHLIST VARIABLE	CITED BY
Climate/weather	16
Anthropogenic variables, urban mapping, land use	11
Vegetation characteristics (structure, % canopy...)	10
Soils/geology	9
Disturbance	8
Microclimates and microhabitat variables	6
Species ecology	5
Remotely sensed data	5
Categorical land cover	4
General scale/resolution issues (incl. temporal)	4
Surface water occurrence	3
Ecosystem variables	3
Topographic variables	2
Spectral indices	2
Species occurrence data	1
Hydrologic variables/stream morphology	1
Marine variables	1
Landscape metrics	1
Water availability	1
Restoration efforts	1
pH levels	1
Pest and controller data (agriculture)	1
Models often oversimplified	1

APPENDIX D: Workshop agenda

Broadening the use of NASA Datasets by the Species Distribution Modeling Community (21-TWSC21-0005; Grant No. 80NSSC21K1984) January – February 2022

** Our workshop agenda is adaptive and subject to change. Please check back periodically for updates.
All sessions run 11:00 am - 1:30 pm Mountain Time
Session 1 Thursday January 6th

- **11:00 am - 11:40 am** Workshop overview, survey responses and SDM variables
- **11:40 am - 11:50 am** Break
- **11:50 am - 12:50 pm** 2 min Lightning Talks (background, data needs) Participants
- **1:00 pm - 1:10 pm** Break
- **1:10 pm - 1:30 pm**; Data access landscape; Q & A

Session 2 - Tues January 11th

- **11:00 am -11:15 am** Opening remarks Gerald "Stinger" Guala
- **11:15 am -12:00 pm** Surface Biology Geology (SBG) Study Ryan Pavlich
- **12:00 pm - 12:10 pm** Break
- **12:10 pm - 12:40 pm** Climate datasets Ryan Rock
- **12:40 pm - 12:50 pm** Break
- **12:50 pm - 1:10 pm** Edaphic variables (soil moisture) Zhihua Liu
- **1:10 pm - 1:30 pm** Q&A

Session 3 - Thurs Jan. 13th

- **11:00 am - 11:15 am** Workshop recap
- **11:15 am - 12:00 pm** Vegetation Structure (GEDI, IceSAT-2) Laura Duncanson
- **12:00 pm - 12:10 pm** Break
- **12:10 pm - 12:30 pm** Data demo/ user experience Marie Johnson; Q&A
- **12:30 pm - 12:50 pm** Human disturbance; AppEEARS Ethan Shafron
- **12:50 pm - 1:00 pm** Break
- **1:00 pm - 1:20 pm** Warmup Breakout Session
- **1:20 - 1:30 pm** Q&A

Session 4 - Thurs Jan 20

- **11:00 am - 11:10 am** Workshop recap
- **11:10 am - 11:40 pm** ECOSTRESS Kerry Cawse-Nicholson
- **11:40 am - 12:10 pm** Using GEDI in SDMs Patrick Burns; GEODE Lab
- **12:10 pm - 12:20 pm** Break
- **12:20 pm - 12:50 pm** Creating better biodiversity maps Mary Blair, Beth Gerstner
- **12:50 pm - 1:00 pm** Break
- **1:00 pm - 1: 30 pm** Breakout groups meet

Session 5 - Feb 3

- **11:00 am - 11:05 am** Recap; Q & A
- **11:05 am - 11:45pm** Breakout groups meet
- **11:45 pm - 11:55 pm** Break
- **11:55 pm - 12:50 pm** Presentations by theme (20 min. each for 5 themes (10 min. presentation & 10 min. discussion): climate, vegetation, edaphic, surface water, anthropogenic)
- **12:50 pm - 1:00 pm** Break
- **1:00 pm - 1:40 pm** Presentations by theme (20 min. each for 5 themes (10 min. presentation & 10 min. discussion): climate, vegetation, edaphic, surface water, anthropogenic)

Session 6 - Feb 4

- **11:00 - 11:10 am** Recap
- **11:10 am - 11:55 am** PACE presentation, Jeremy Werdell
- **11:55 - 12:05 pm** Break
- **12:05 am - 12:35 pm** Human Dimensions reporting
- **12:35 pm - 1:00 pm** Final breakout session
- **1:00 pm - 1:10 pm** Break
- **1:05 pm - 1:30 pm** Final recommendations are reported (5 min. each breakout group)