

Highlights from RMBL Data Products Community Survey



Ian Breckheimer
Research Scientist
Rocky Mountain Biological Laboratory
9 January 2020

Context

In October 2019, research scientist Ian Breckheimer sent an online survey to the PI listserv soliciting feedback on proposed spatial data products that could be developed as part of RMBL's push to improve resources available to scientists. The survey asked respondents to assess the importance of 38 proposed data products to their future research program, as well as their overall importance, and barriers to utilizing these products. This document summarizes responses to that survey in an attempt to understand which data products would have the highest utility to the research community, and what barriers might stand in the way of their use.

Key Results

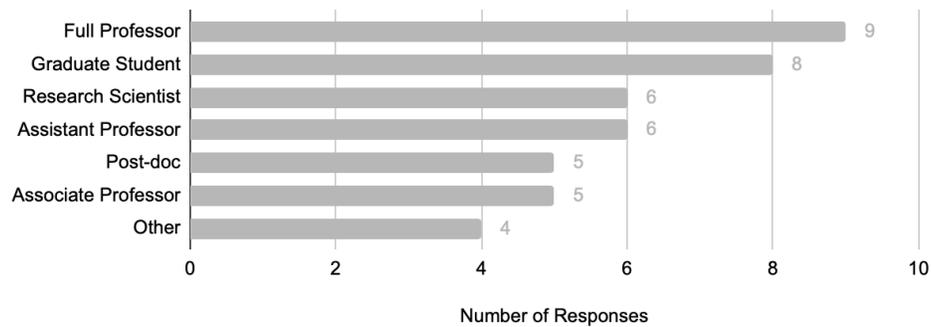
We present detailed summaries of survey participation and results in the following sections, but highlight a few key results here:

1. 43 scientists from a wide variety of disciplines and at many different career stages responded to the survey.
2. We have broad support for the spatial data initiative among survey respondents. 84% of them indicated that the data products would be "Very Important" or "Critical" to their research.
3. Products dealing with land surface dynamics (snow cover, phenology), near-surface microclimate (temperature, snowpack, precipitation), and ecosystem processes (GPP, NPP) have the highest perceived utility.
4. Key barriers to using these datasets include lack of availability at non-RMBL field sites, lack of skill in manipulating spatial data, and mismatches in scale between field measurements and spatial datasets.
5. Scientists want datasets to be accessible and citable, suggesting investment in making the platform easy to use, and publishing methods / data papers along with the products.

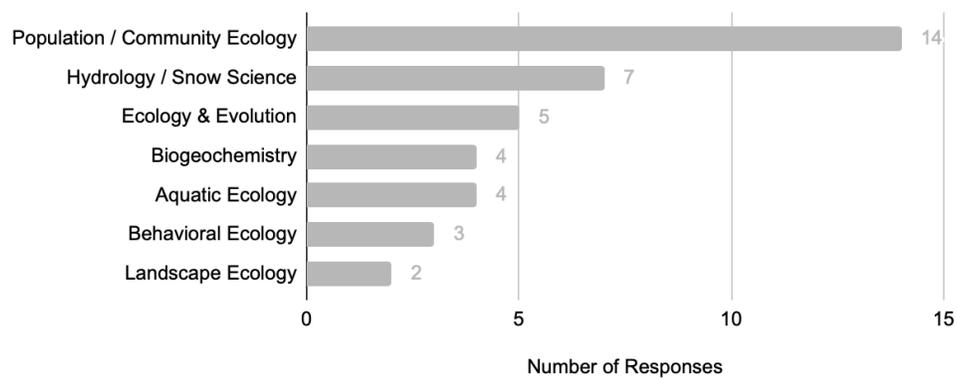
Responses

From October to December 2019, we received 43 responses to the survey from RMBL scientists, students, technicians, and other stakeholders, representing a broad (though not necessarily representative) sample of the RMBL-affiliated scientific community. Figure 1 shows the distribution of respondents by career stage, broad discipline, and study system. Note that discipline and study system categories were manually generated from open-ended responses to the question “How you would describe your primary research area?”

What best describes your career stage?



Discipline Category



Primary Study System Category

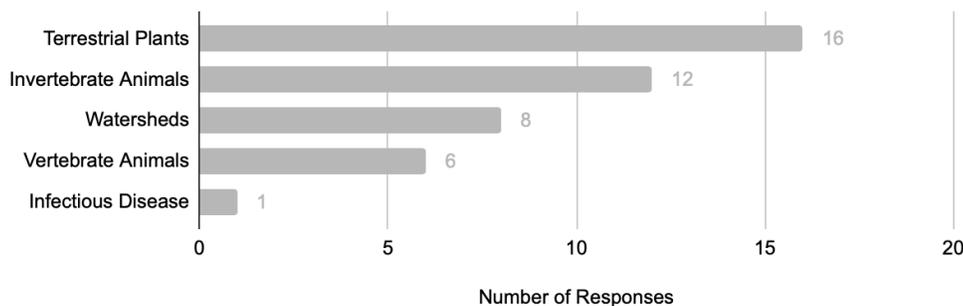


Figure 1: Survey respondents represented a broad cross-section of RMBL scientists in a variety of disciplines and career stages.

Perceived Utility of Individual Data Products

We asked respondents to evaluate the usefulness of 38 potential spatial data products in supporting their current or future research program. Respondents placed each product in one of four ranked categories: 1 – Not Useful, 2—Somewhat Useful, 3—Very Useful, 4—Mission-critical. Scores for each product were summarized (Table 1) by calculating the percentage of respondents that gave each product a score of 3 or 4 (“>=Very Useful “ in Table 1), the percentage that gave each product a score of 4 (“Mission Critical” in Table 1), as well as by computing the arithmetic average of the numeric scores. These measures were strongly correlated with one another, but each represents a slightly different way of evaluating how the scientist community views each product.

Table 1: Perceived utility of 38 potential data products, ranked by the proportion of respondents indicating the product would be “Very Useful” or “Mission Critical” in their current or future research.

Data Product	Type	>= Very Useful (%)	Mission Critical (%)	Mean Score
[Snowpack accumulation (10-30m)]	Microclimate	90.7%	44.2%	3.35
[Snow onset / departure dates (10-30m)]	Microclimate	86.0%	55.8%	3.40
[Accumulated precipitation (90-180m)]	Microclimate	83.7%	39.5%	3.23
[Snow-free Growing Degree Days (10 - 30m)]	Microclimate	72.1%	25.6%	2.91
[Near-surface (10cm, 0cm) air temperature (10-30m res.)]	Microclimate	69.8%	23.3%	2.91
[Timing of vegetation green-up and brown-down (3-10m)]	Phenology	69.8%	25.6%	2.88
[Meadow vegetation community type (1-3m)]	Vegetation	65.1%	14.0%	2.65
[Near-surface (0cm, -10cm) soil temperature (10-30m)]	Microclimate	65.1%	20.9%	2.74
[Digital Elevation Model (0.5 - 1m)]	Geophysical	62.8%	23.3%	2.79
[Near-surface (0-10cm) soil moisture (10-30m)]	Microclimate	62.8%	20.9%	2.72
[Timing of first and last freeze (30 - 90m)]	Microclimate	62.8%	20.9%	2.77
[Incident solar radiation (30 - 90m)]	Microclimate	58.1%	18.6%	2.63
[DEM Derivatives (slope, aspect, TMI, 0.5 - 1m)]	Geophysical	53.5%	11.6%	2.47
[Above-surface (2m) air temperature (10-30m res.)]	Microclimate	51.2%	25.6%	2.70
[Gross primary productivity (10 - 30m)]	Eco. Process	51.2%	14.0%	2.37
[Net primary productivity (10 - 30m)]	Eco. Process	51.2%	14.0%	2.40
[Vegetation height growth rate (1 - 3m)]	Eco. Process	51.2%	11.6%	2.44
[Peak-season Aboveground biomass (5-10m)]	Vegetation	48.8%	14.0%	2.35
[Timing of peak leaf area (10 - 30m)]	Phenology	48.8%	14.0%	2.37
[Deeper (10cm - 1m) soil moisture (30-90m)]	Microclimate	46.5%	14.0%	2.35
[Peak vegetation cover (1-3m)]	Vegetation	44.2%	7.0%	2.30
[Peak-season cover of dominant plant species (1-3m)]	Vegetation	44.2%	7.0%	2.37
[Soil sand, silt, clay proportions (1 - 3m)]	Geophysical	44.2%	7.0%	2.35
[Vegetation height (1-3m)]	Vegetation	41.9%	4.7%	2.28
[Hydrologically Corrected DEM (0.5 - 1m)]	Geophysical	39.5%	14.0%	2.23
[Soil chemical properties (N, P, K, 1 - 3m)]	Geophysical	39.5%	14.0%	2.23
[Soil hydrologic conductivity (1 - 3m)]	Geophysical	39.5%	16.3%	2.19
[Reference evapotranspiration (30-90m)]	Microclimate	37.2%	4.7%	2.16
[Vapor pressure deficit (30-90m)]	Microclimate	37.2%	9.3%	2.23

[Overstory (>1m height) Density (1-3m)]	Vegetation	32.6%	4.7%	2.05
[Peak Leaf Area Index (5-10m)]	Vegetation	32.6%	4.7%	2.02
[Wind speed (90-180m)]	Microclimate	32.6%	11.6%	2.23
[Climatic water deficit (30-90m)]	Microclimate	30.2%	4.7%	2.21
[Geologic Parent Material (1 - 3m)]	Geophysical	30.2%	9.3%	2.19
[Leaf canopy chemical properties (N, P, water, 1-3m)]	Vegetation	30.2%	16.3%	2.02
[Understory (<=1m height) Density (1-3m)]	Vegetation	27.9%	4.7%	1.95
[Soil microbial biomass (1 - 3m)]	Eco. Process	25.6%	4.7%	1.84
[Soil microbial functional composition (1 - 3m)]	Eco. Process	20.9%	4.7%	1.84

We also asked respondents to suggest additional products in each category, and products that were suggested by more than one respondent are listed in Table 2.

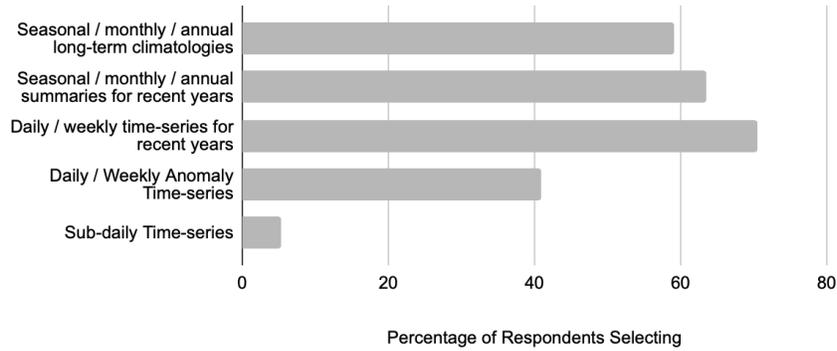
Table 2: Additional data products suggested by the research community that were mentioned by more than one survey respondent.

Product Type	Format	Times Mentioned
Flower / Floral Resource Abundance	Weekly Time-series	4
Downwelling (SW / LW) Radiation	Daily Time-series	3
Wind Direction	Daily Time-series	2
NDVI / Greenness	Daily Time-series	2
Wetland / Riparian Vegetation Type	Static Map	2
Wetland Hydroperiod / Stream Flow	Daily / Weekly Time-series	2
Vegetation Disturbance	Annual Time-series	2

Product Frequency and Format

For products dealing with microclimate and ecosystem processes, we asked respondents to indicate what format and frequency of data products would be useful in their research. Respondents could select more than one option. Figure 2 shows the percentage of respondents selecting each option. Responses to this question highlight the diversity of different data types that the research community would find useful, with a large proportion of respondents desiring long-term climatologies, and shorter-term summaries (annual, seasonal, monthly), as well as daily or weekly time-series.

Desired Product Types - Microclimate



Desired Product Types - Ecosystem Process

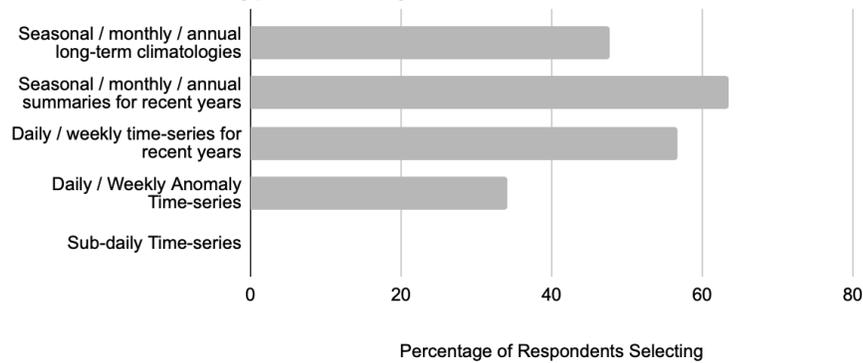


Figure 2: The research community would find many different forms of data products useful in their research.

Desired Uses of Data Products

The research community would likely use the research products for a wide variety of purposes, from boosting understanding of how the environment affects their study system, to fitting spatially explicit models. Figure 3 shows the percentage of respondents that selected each potential use of the data products. Respondents could select more than one option.

What would you use the data products for?

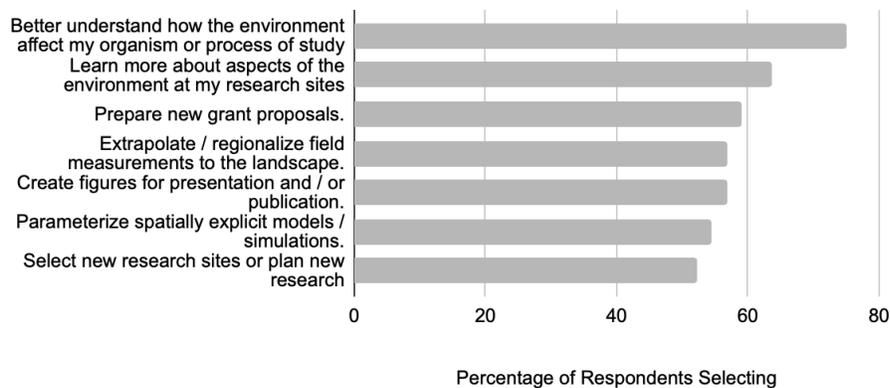


Figure 3: Percentage of respondents selecting different possible uses for the spatial data products. Respondents could select more than one option.

Overall Importance

A large proportion (84%) of respondents indicated that the data products would be “Very Important” or “Critical” to their research program. No respondents selected the lowest category of importance “Incidental — I’m unlikely to use these datasets for any of my work”.

Overall, how important would these data products be to your research program?

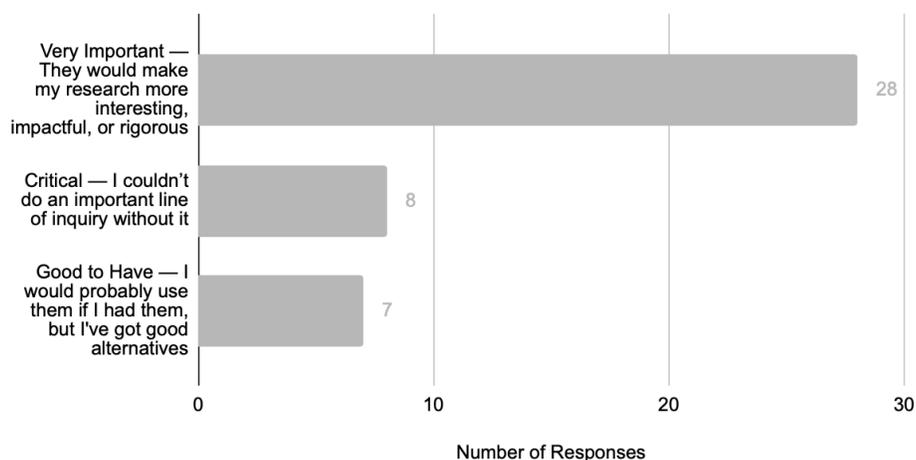


Figure 3: A large proportion of respondents indicated that the data products would be very useful or critical to their research program.

Timelines for Incorporation

Most respondents (79%) indicated that they would be able to incorporate the data products into scientific publications within 24 months of receiving them. Few respondents indicated that this timeline was shorter than 6 months or longer than 24 months (Figure 4).

How quickly do you think you would incorporate these data products into scientific publications after receiving them?

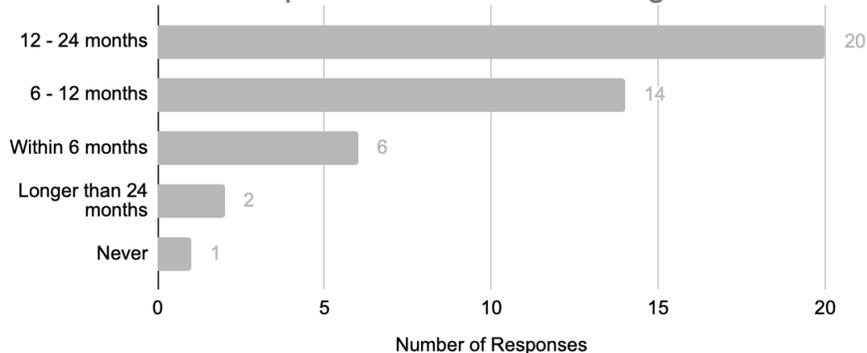


Figure 4: A large proportion of respondents indicated that the data products would be very useful or critical to their research program.

Barriers to Use

We asked respondents which barriers they anticipated could prevent them from utilizing the spatial data products. Specifically, we asked respondents to rank each potential barrier on an ordinal scale: 1- no barrier / not limiting, 2 - a nuisance / small barrier, 3 – significant barrier that can be overcome with effort, 4 – critical barrier very difficult to overcome. The results are summarized in Table 3, which ranks barriers by the percentage of respondents indicating that the barrier was larger than “a nuisance”. We also calculated the percentage of respondents that indicated that the barrier was ranked 4 (“very difficult to overcome”), as well as an arithmetic average of the ordinal scores.

Table 3: Anticipated barriers to use of spatial data products.

Barrier to Use	>=Nuisance (%)	Critical (%)	Mean Score
Lack of availability of the data at non-RMBL field sites	48.5%	21.2%	2.30
Lack of skills / experience in manipulating spatial data (GIS)	45.2%	7.1%	2.24
Mismatches in scale between field data and gridded data products	44.4%	13.9%	2.50
Limited awareness of measurement types in my research specialty	23.3%	0.0%	1.70
Potential reviewer skepticism about using extrapolated measurements	17.1%	2.9%	1.89

We also asked respondents to suggest other barriers that might prevent them from utilizing the products. A selection of those responses is below:

“...we need DOI numbers for datasets, metadata and methods. Perhaps you can publish a methods paper we can all cite?”

“Infrequent updating of the datasets could impair my research plans.”

“I would need help getting data extracted so that I could fit models...”

Unstructured Feedback

At the end of the survey, we asked respondents “Do you have any other feedback for us?”. Here is a selection of those responses:

“perhaps some of your datasets could have a finer scale for intensive sites such as around Gothic, the pumphouse, research meadow, etc”

“Please keep the community updated on this project. It's a new model for RMBL, and there are concerns about the process.”

"I hope that the finished products are available to local land managers and resource professionals."

"High-res climate data is badly needed at RMBL, so well done for making it happen."

"having the ability to go back and forward in time with these measurements would be GREAT"

"One of the major topics of our meetings about modeling is how to scale from plots to landscape to region to globe."

"I assume these are all terrestrial - what about the aquatic systems?"