Deck 7: All About Data!

File Paths, Data Creation and Management, Land Cover, Suitability Analysis

Intro to GIS – UMass Amherst – Michael F. Nelson

Overview

Data Management

File Paths

Your computer's file system is organized like a tree:

There is a root directory on your storage medium that contains all the other files and directories.

To find a file, your computer traverses the branches of the tree following the directions contained in the *path*

An *absolute* path:

gives your computer the map to find a file starting from the *root directory* of your storage medium.

- An *absolute path* is never ambiguous.

A *relative* path:

provides a map starting at the current *working directory*.

- A working directory is just the directory that a program is currently pointing to.

File Types

File extensions and associations

Filenames usually consist of two or more blocks of letters separated by periods:

file_01.txt

The extension is meant to communicate the *file type* or *file format*.

Archive files

Archive files can contain copies of a directory structure, which itself can contain files and other directories.

The most familiar archive format is .zip. Other common formats are .tar and .gz.

Archive files are typically (but not always) *compressed*

> Compressed files are in any of several specialized *binary* formats.

Archive files are a convenient way to share multiple files and directories as a single unit.

Those Pesky !

- The ! appear when you load a project into Arc, but it can't find your data.

- Think back to one of our definitions of GIS: Map documents are containers that contain pointers to the data, and information on how to display your data sources.

Data Management

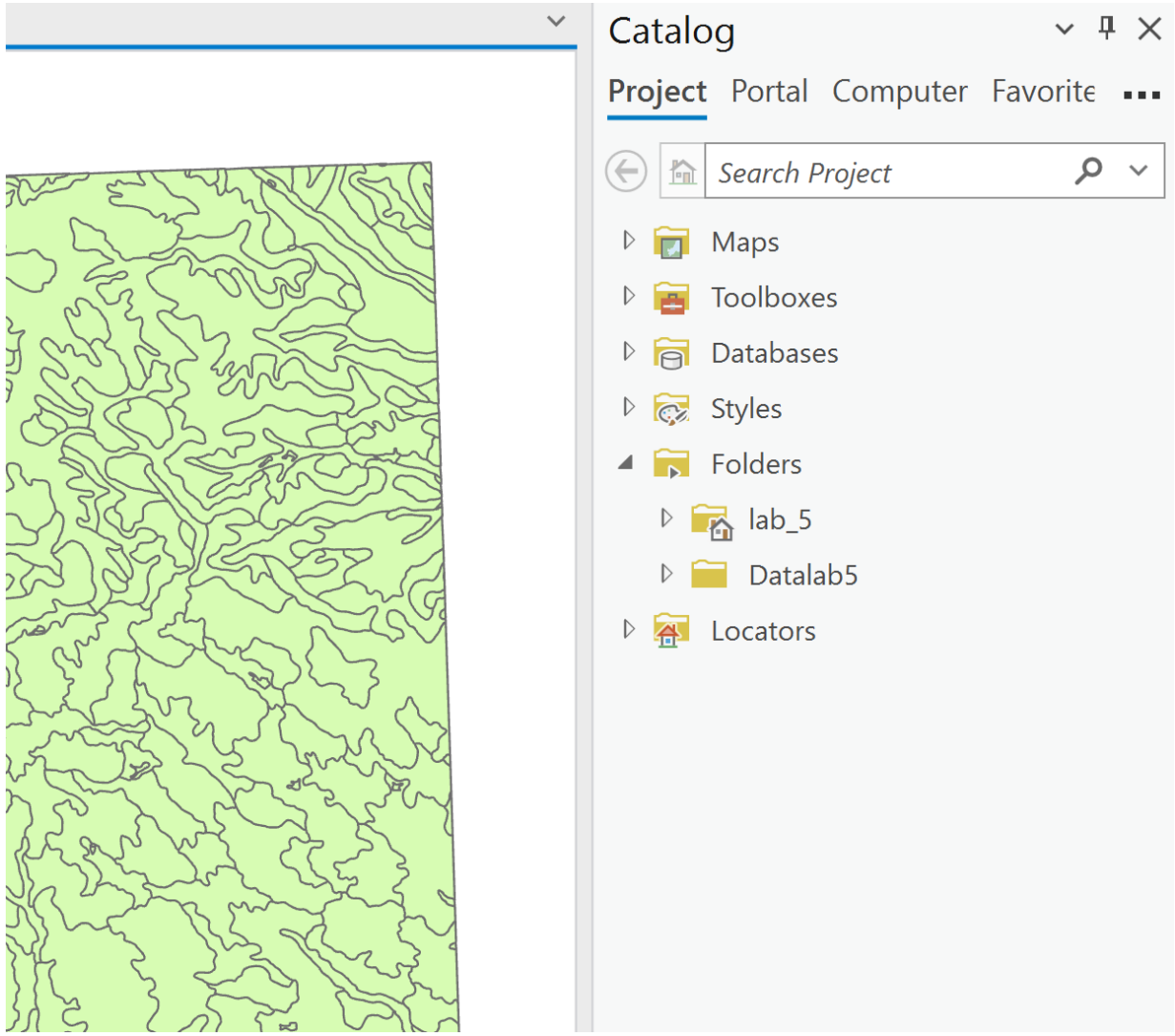
It’s easy to unintentionally lose your data/analysis in Arc

- Part of this we’ve discussed before organizing your data with sensible folders and filenames.
- Part of it is structural: Arc’s Geodatabases.

Saving, Processing, Defaults

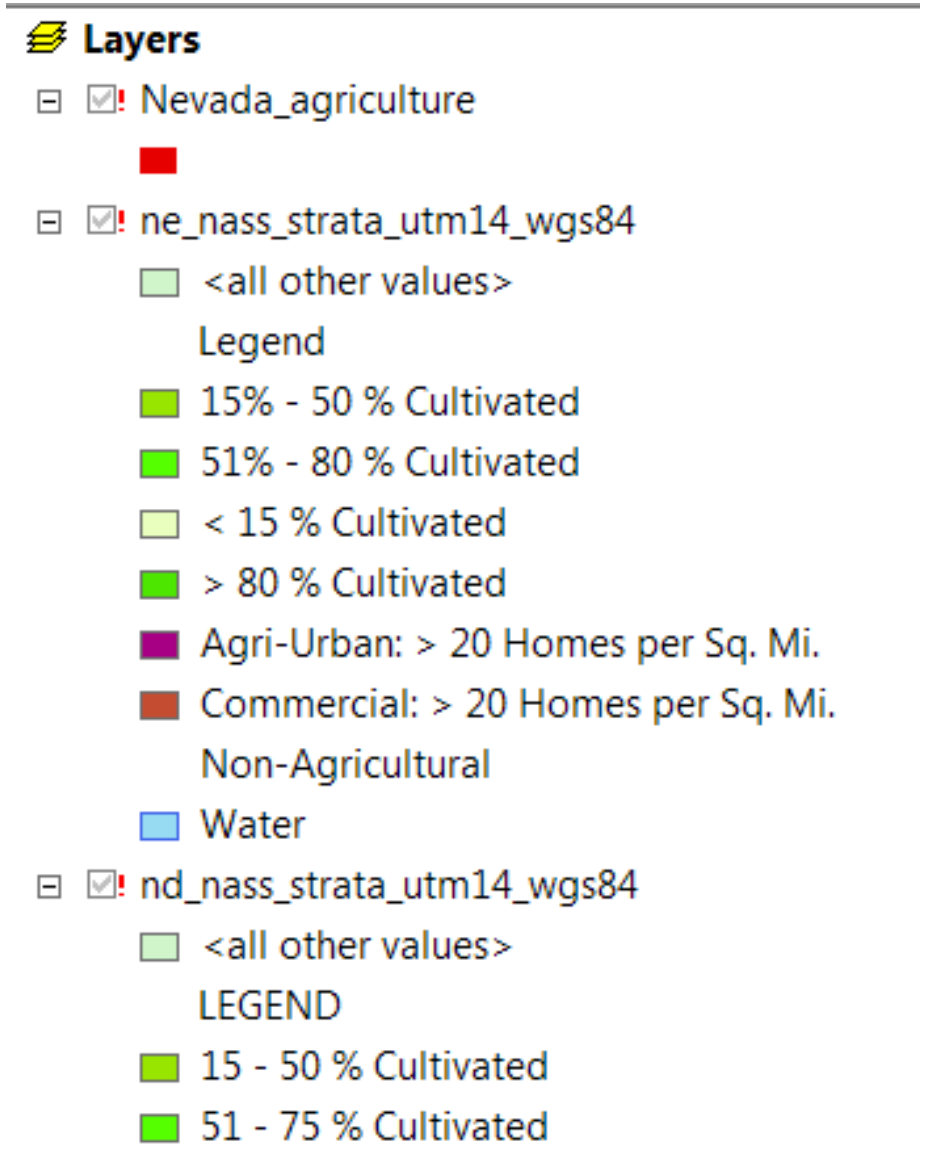
- Arc projects have default locations in which data and analyses live.
- They may not be the most convenient for you, especially if you want to share data.
- In order to remain in control, you need to know how Arc's data management works!

Arc Catalog


Catalog
Project
Portal
Computer
Favorite
Search Project
Maps
Toolboxes
Databases
Styles
Folders
lab_5
Datalab5
Locators

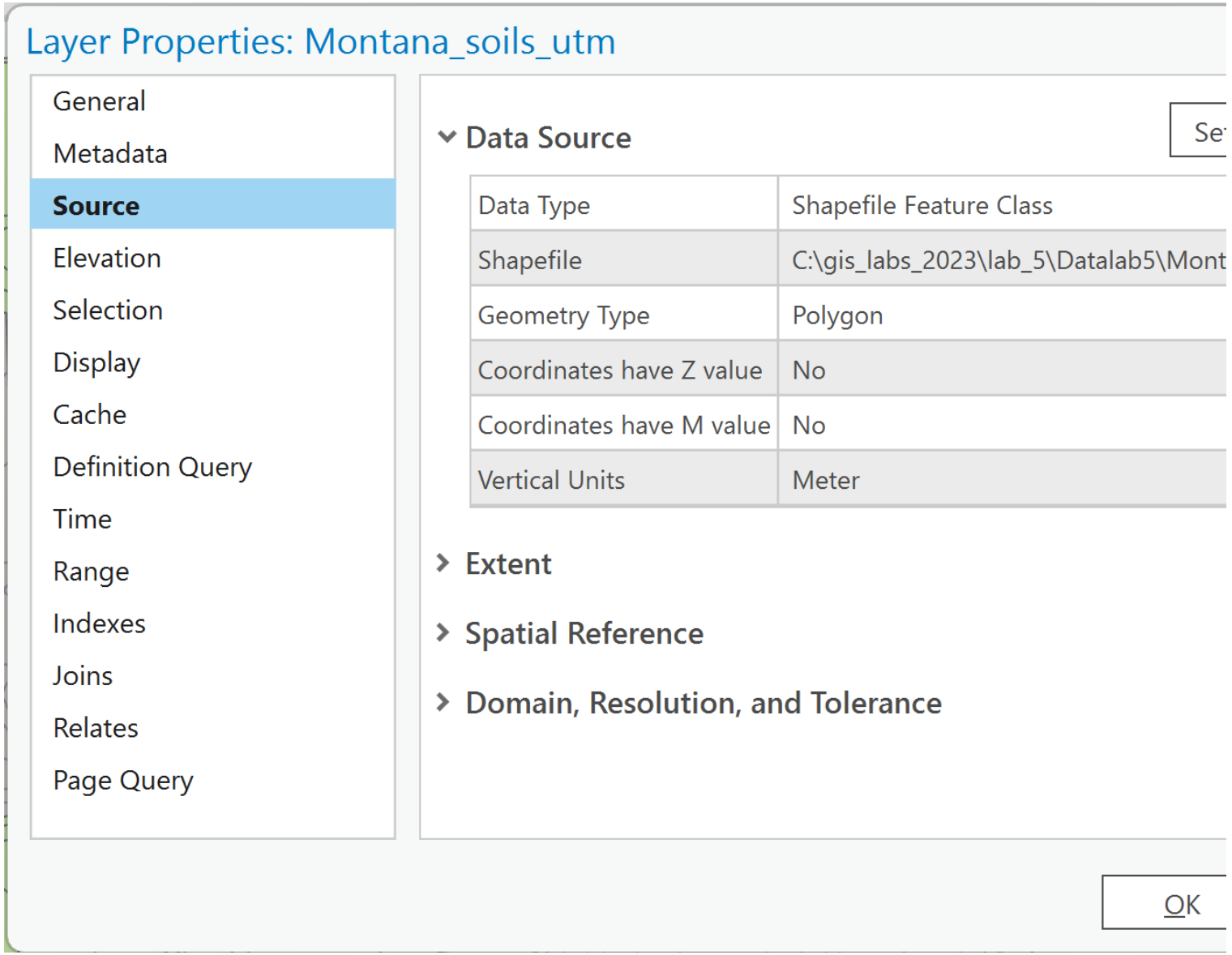
Changing computers or moving files can lead to lots of headaches!

And also the red exclamation marks...



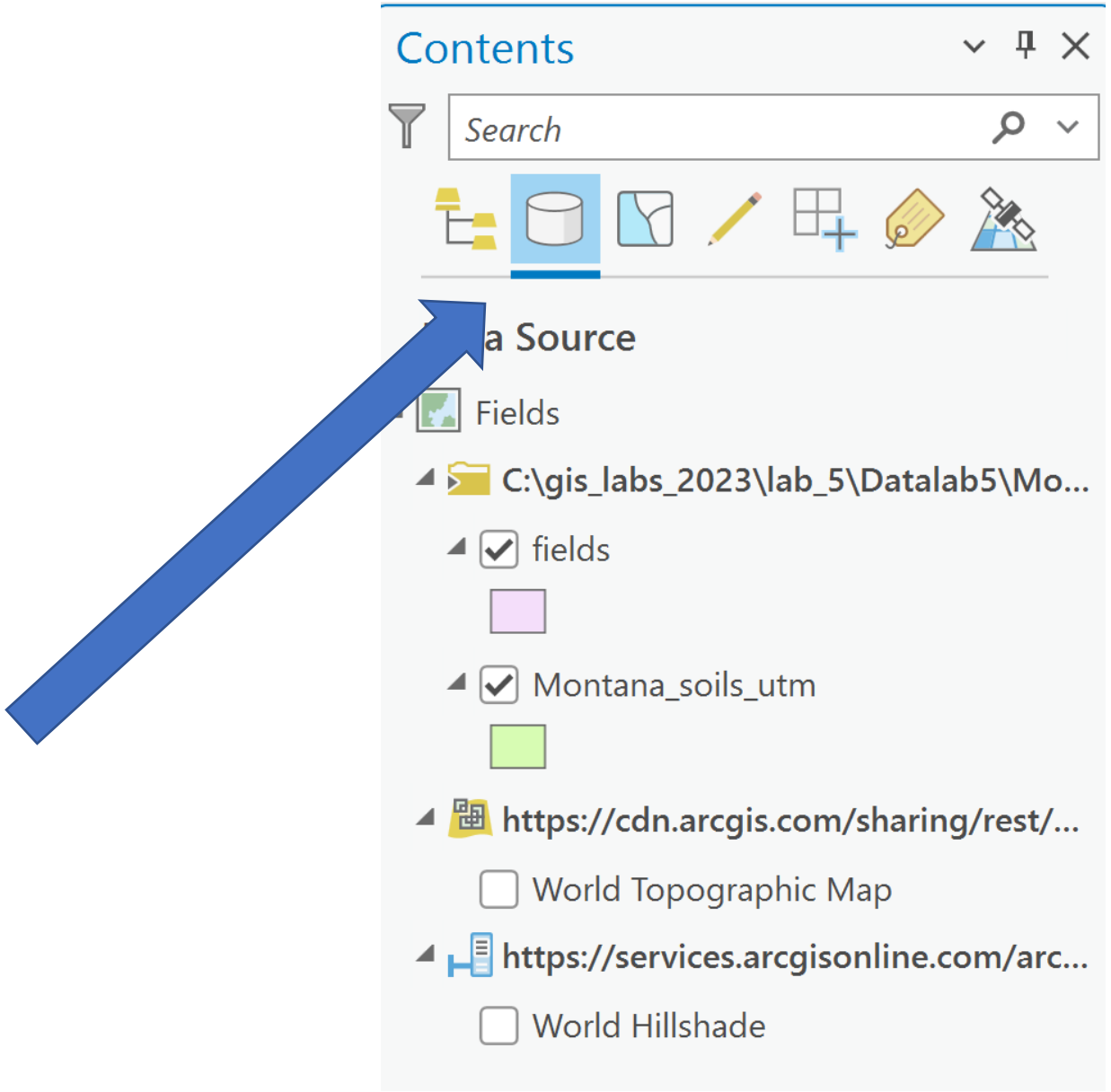
But, it's easy to fix incorrect paths (*provided you have the data*)

If the path is wrong: Set Data Source…



Use ‘list by source’ in ArcMap Table of Contents to view paths

- Avoid special-use directories:
  - Your Downloads folder
  - Your Desktop
  - Anything with ‘temp’ or ‘tmp’ in the name.



Those Pesky !

- The ! appear when you load a project into Arc but it can’t find your data.

- Think back to one of our definitions of GIS: The container of maps.

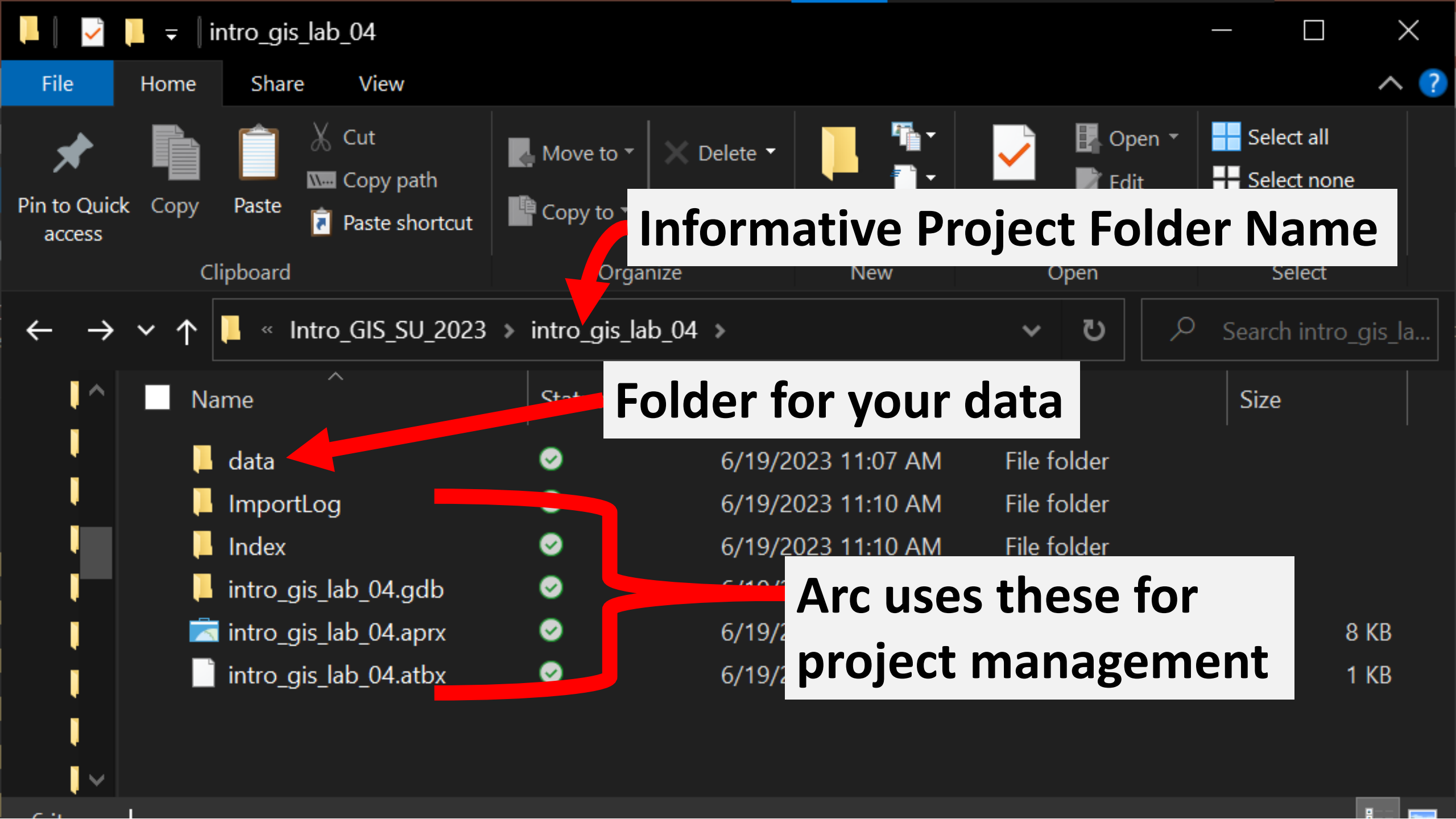
Arc Project Data – Suggested Practices

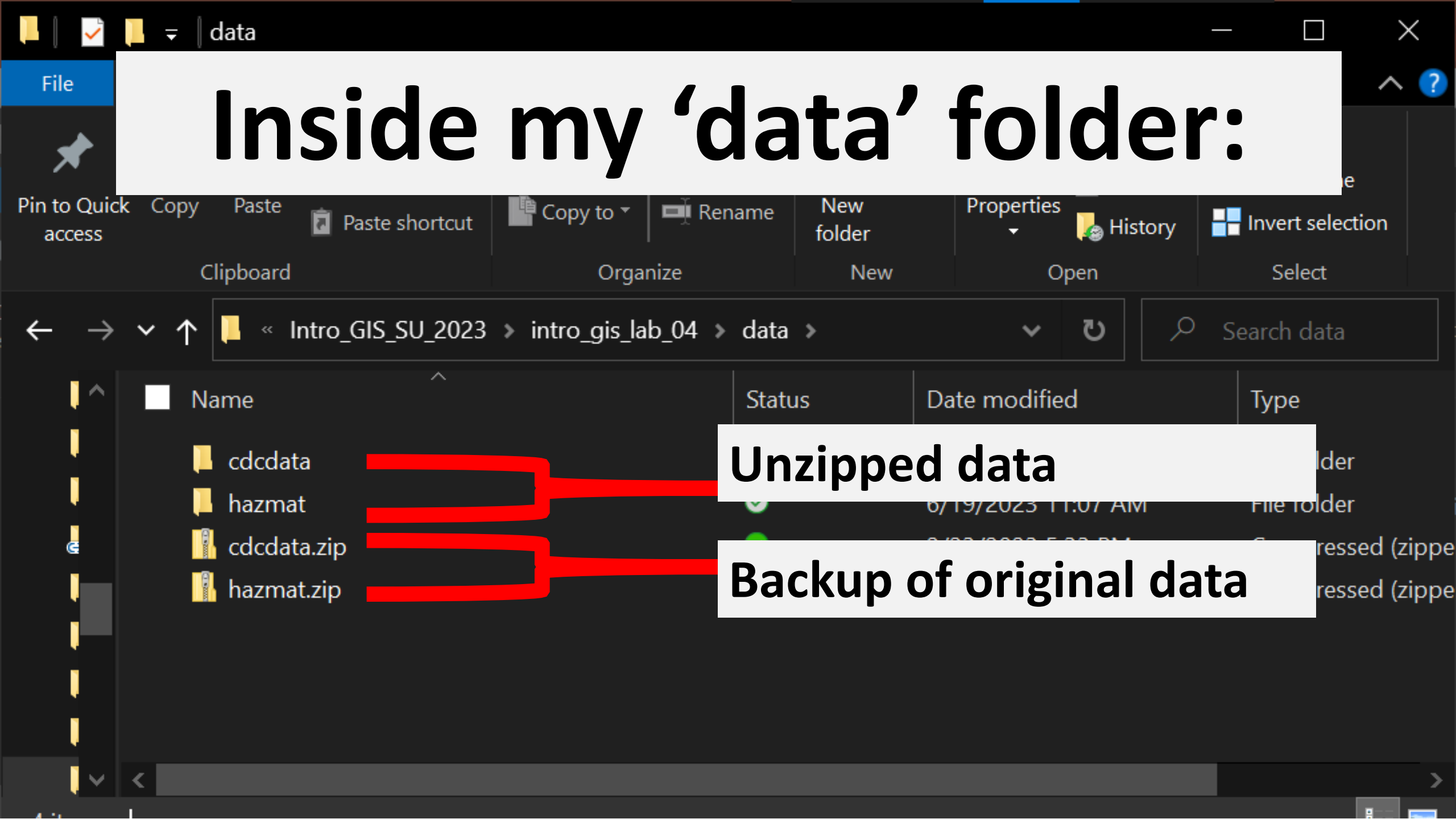
Create a new folder with an informative name for each Arc project. For example: lab_1, final_proj, etc.

Create the Arc project within your folder and give it an informative name. Don't use generic names like 'untitled project 37'.

Data files:

- Create a 'data' subdirectory in which you'll store all the data you import for the project. Keep backup copies of original data files as zip archives.
- Use the 'data' subdirectory to store any output data you might want to share or use for other projects.
- Use the project (default) database for intermediate datasets.


intro_gis_lab_04
File
Home
Share
View
Pin to Quick access
Copy
Paste
Cut
Copy path
Paste shortcut
Clipboard
Move to
Delete
Copy to
Organize
New
Open
Edit
Select all
Select none
Select
Informative Project Folder Name
Intro_GIS_SU_2023
intro_gis_lab_04
Search intro_gis_la...
Folder for your data
Name
Size
data
6/19/2023 11:07 AM
File folder
ImportLog
6/19/2023 11:10 AM
File folder
Index
6/19/2023 11:10 AM
File folder
intro_gis_lab_04.gdb
intro_gis_lab_04.aprx
8 KB
intro_gis_lab_04.atbx
1 KB
Arc uses these for project management

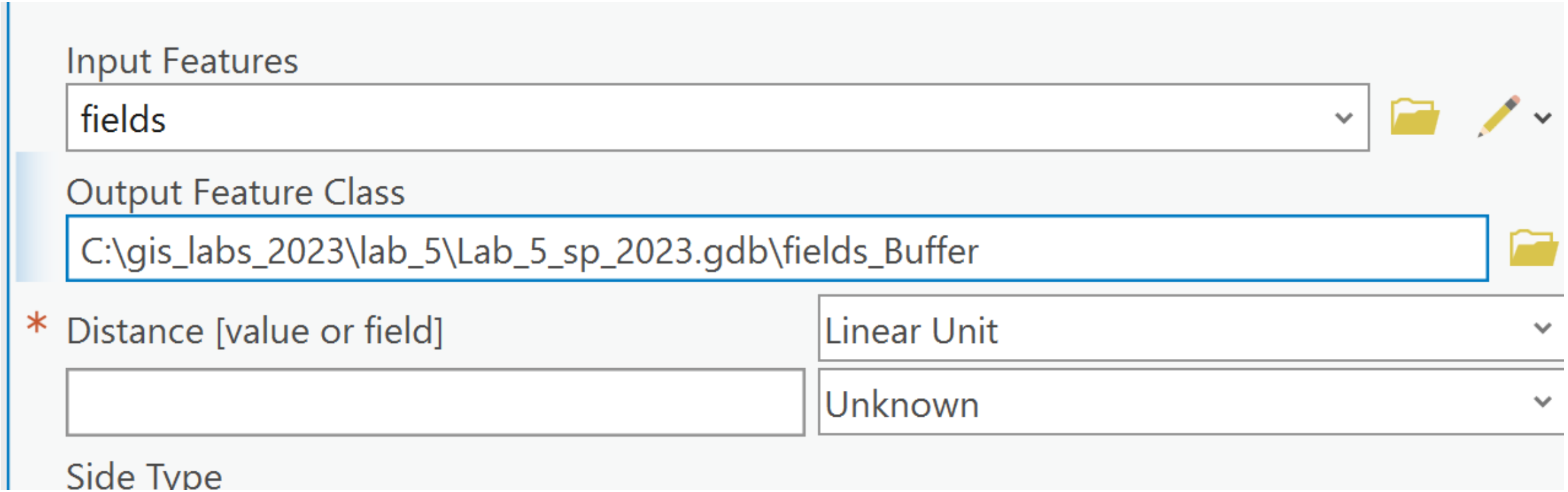

data
File
Inside my ‘data’ folder:
Pin to Quick access
Copy
Paste
Paste shortcut
Copy to
Rename
New folder
Properties
History
Invert selection
Clipboard
Organize
New
Open
Select
Intro_GIS_SU_2023
intro_gis_lab_04
data
Search data
Name
Status
Date modified
Type
cdcdata
hazmat
cdcdata.zip
hazmat.zip
Unzipped data
Backup of original data

Geodatabase?

- Via Arc Help:
  - ‘An ArcGIS geodatabase is a collection of geographic datasets of various types held in a common file system folder, or a multiuser relational DBMS (database management system)’.
- Your project geodatabase is a folder (that looks like a file) which contains all the files created by your arc project.

Default Geodatabase?

- By default, Arc puts stuff here.
- Located (unless you change it) in a subdirectory of your current project, e.g. C:\gis_labs_2023\lab_5\Lab_5_sp_2023.gdb



Where do spatial data come from? How reliable are they?

How to create your own data?

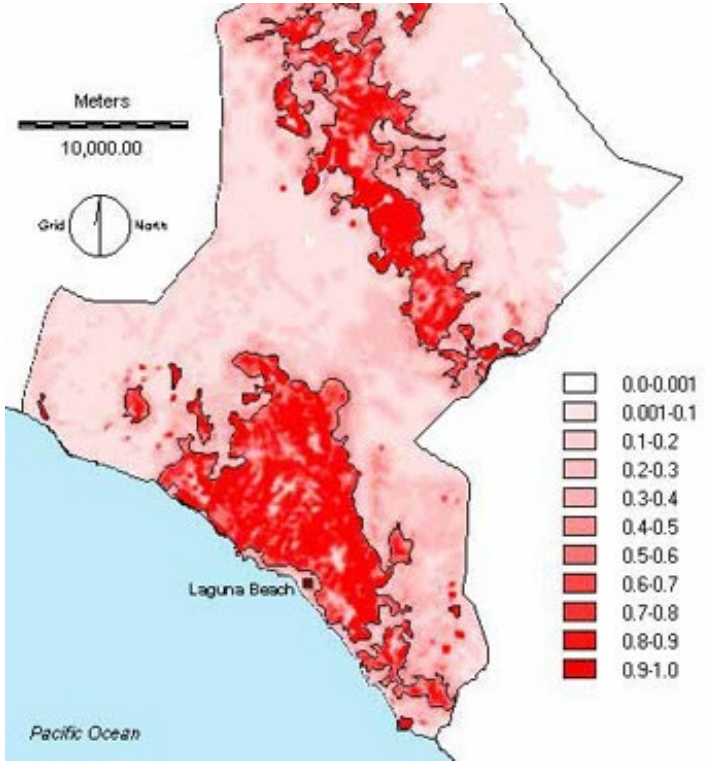
Data from Everywhere

Anything with spatial attributes can be mapped.

Even things without explicitly spatial attributes can be mapped!

If mappable, GIS can use it.

Habitat Suitability for CA Gnatcatcher



Habitat Suitability for Sudden Oak Death

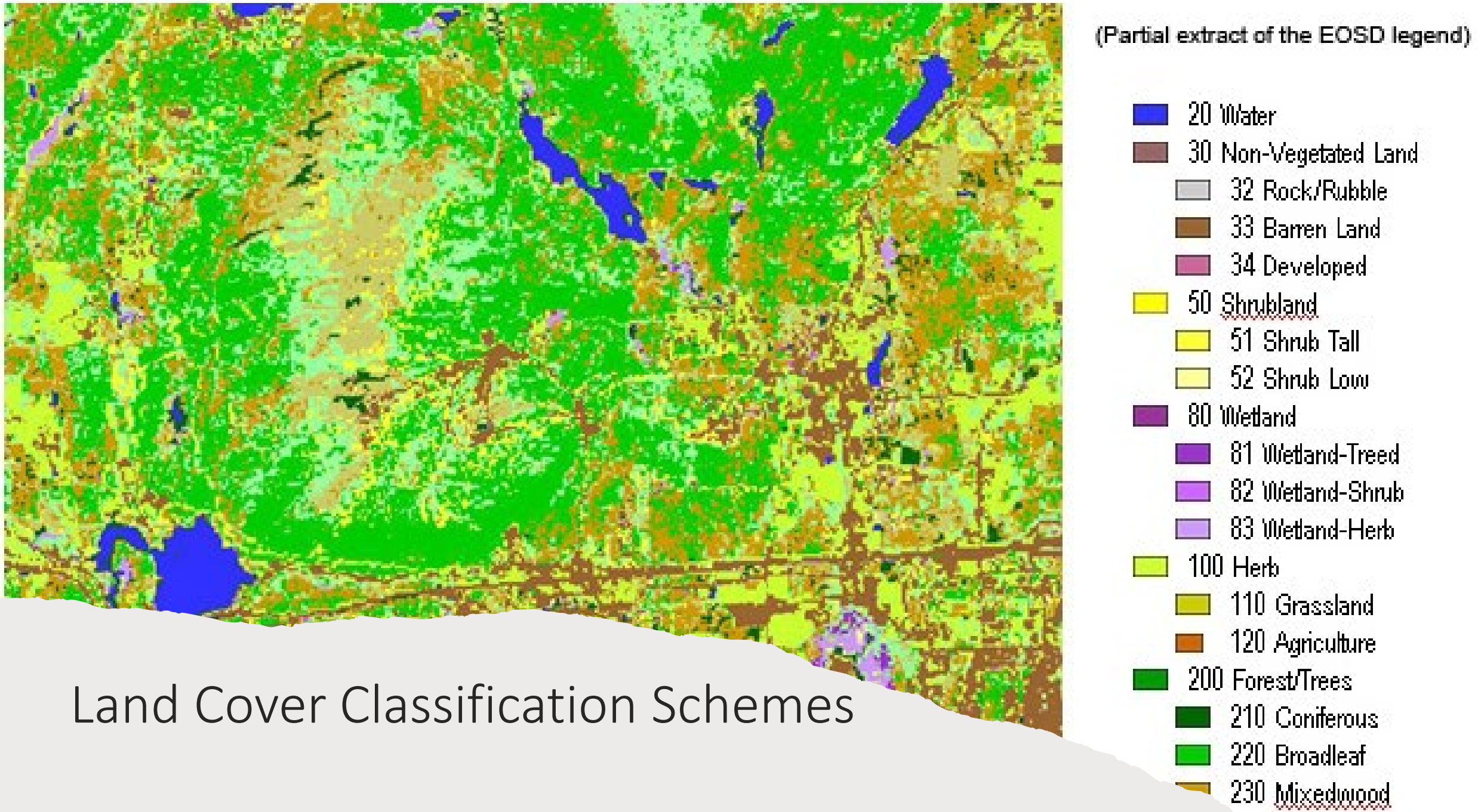
What kind of things are important to gnatcatchers, or fungus-like organisms?

Suitability analyses require DATA

DATA	SOURCE
Land cover and land use maps	Satellite image classification, aerial photo interpretation
Climate	Weather stations, interpolation
Census map	Mailed surveys
Topographic maps	Surveys, RADAR
LULC Change maps	Digitized paper maps, aerial photo interpretation
Occurrence/abundance	Surveys, collars

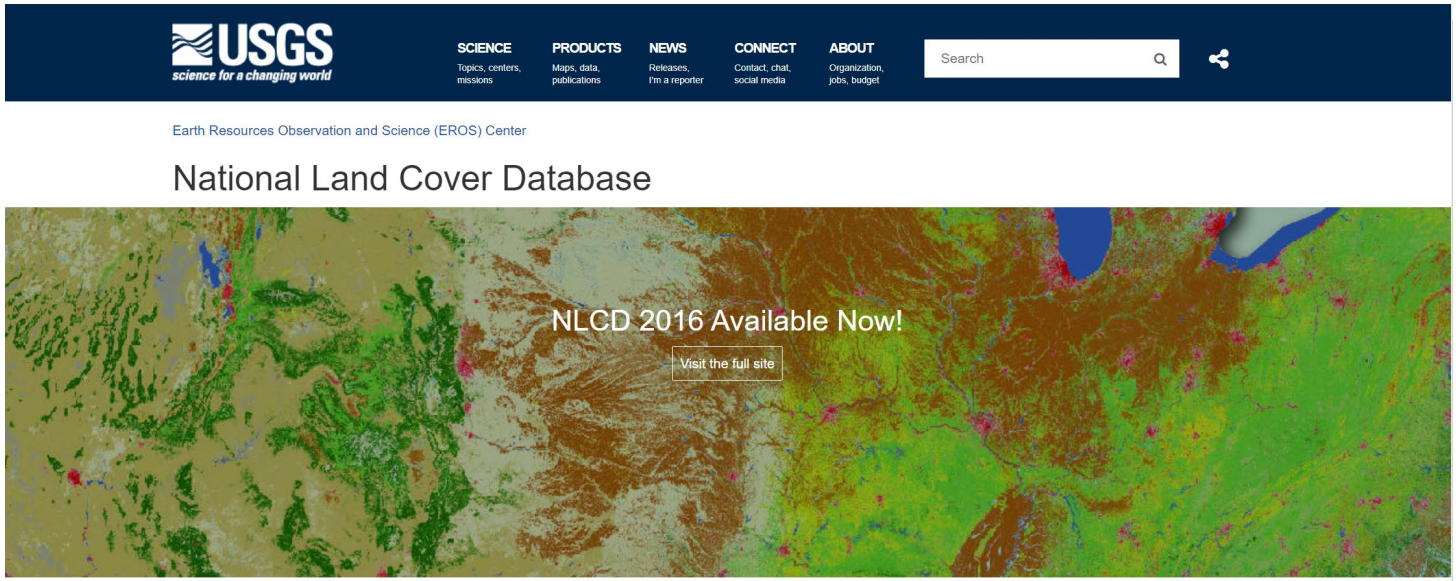
Land Cover Classification

- Landsat 7
- Satellite imagery



Land Cover Classification Schemes

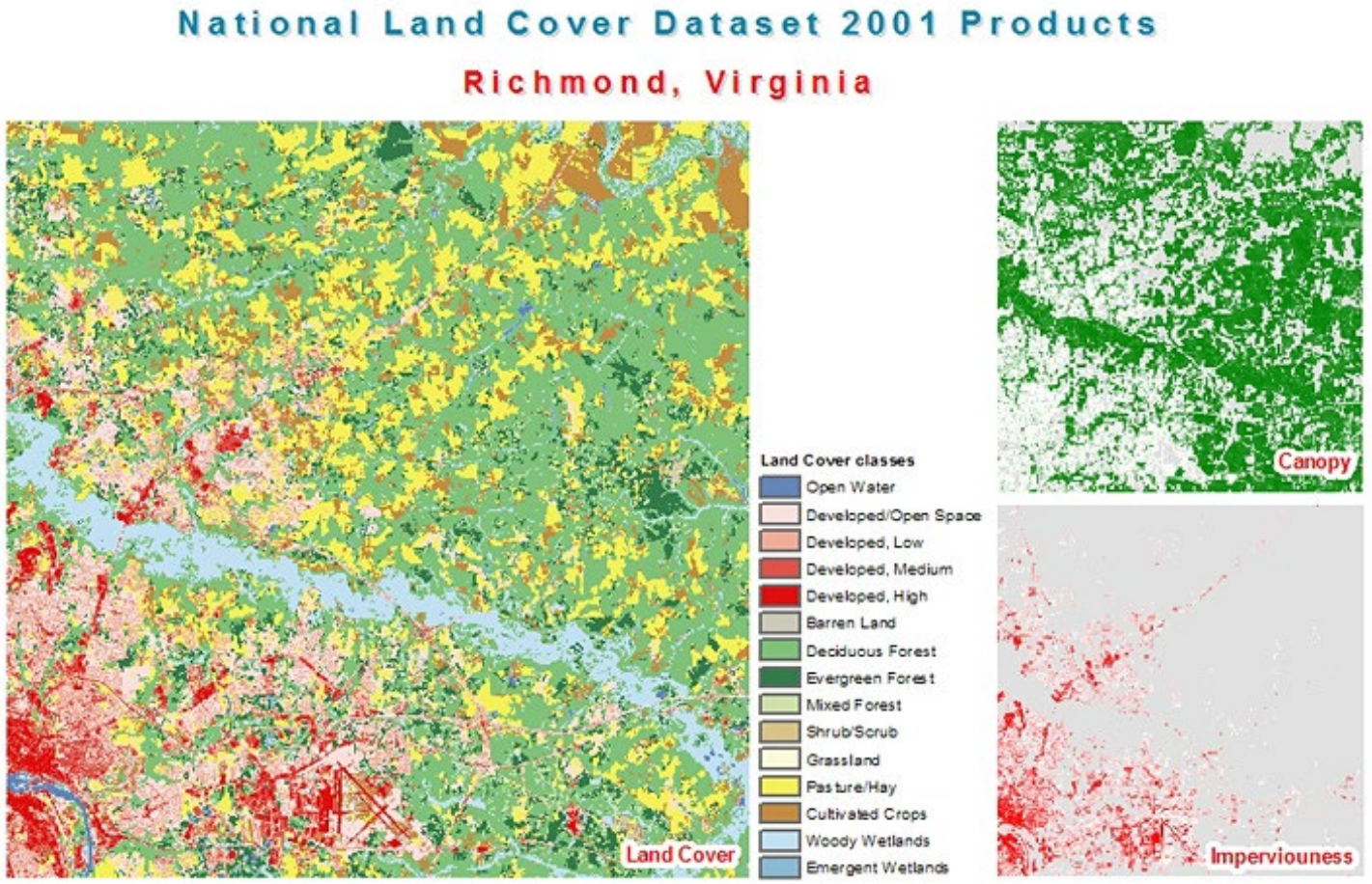
Land Use and Land Cover (LULC)



- **Land Cover:** How can we categorize the surface in physical, biological, ecological, etc. Terms?
- **Land Use:** What are people using the land for?

Composition and configuration

- Composition: How much of the area is forested?
- Configuration: How large are the forest patches?



Landscape Configuration

- Connectivity
- Fragmentation
- Habitiat, nonhabitat, matrix
- Characteristic scale?
- Configuration is usually harder to describe than composition.

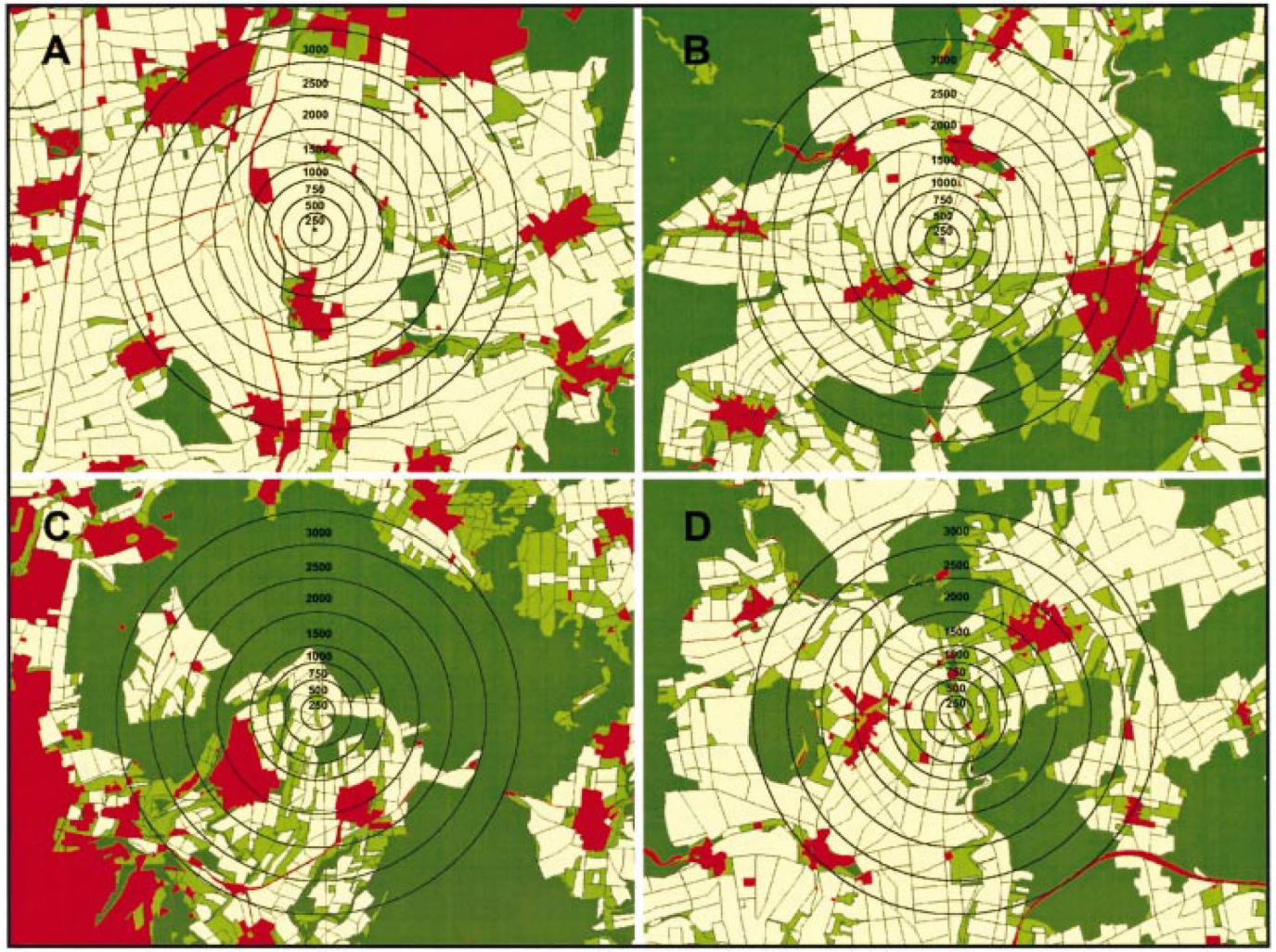
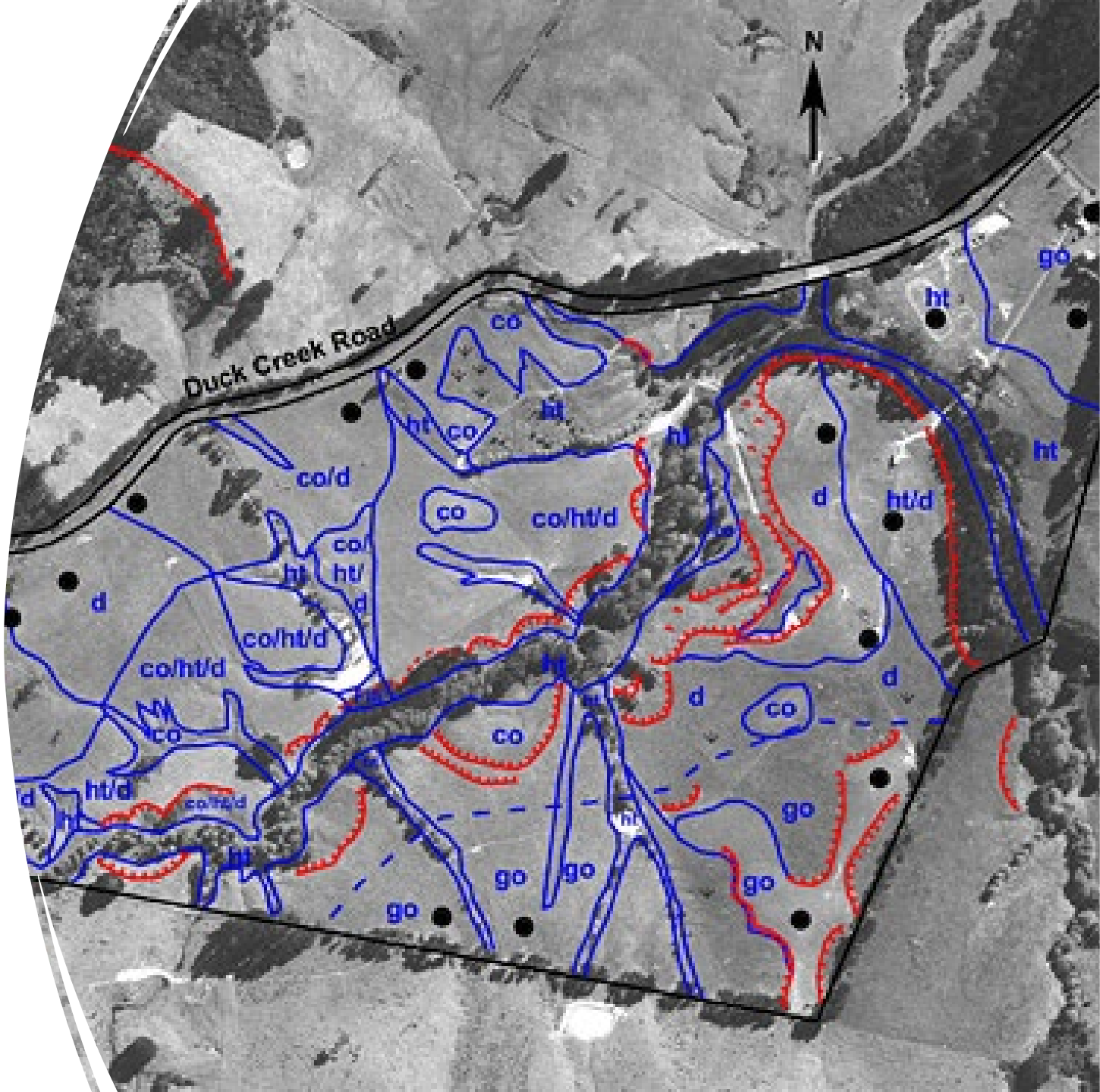
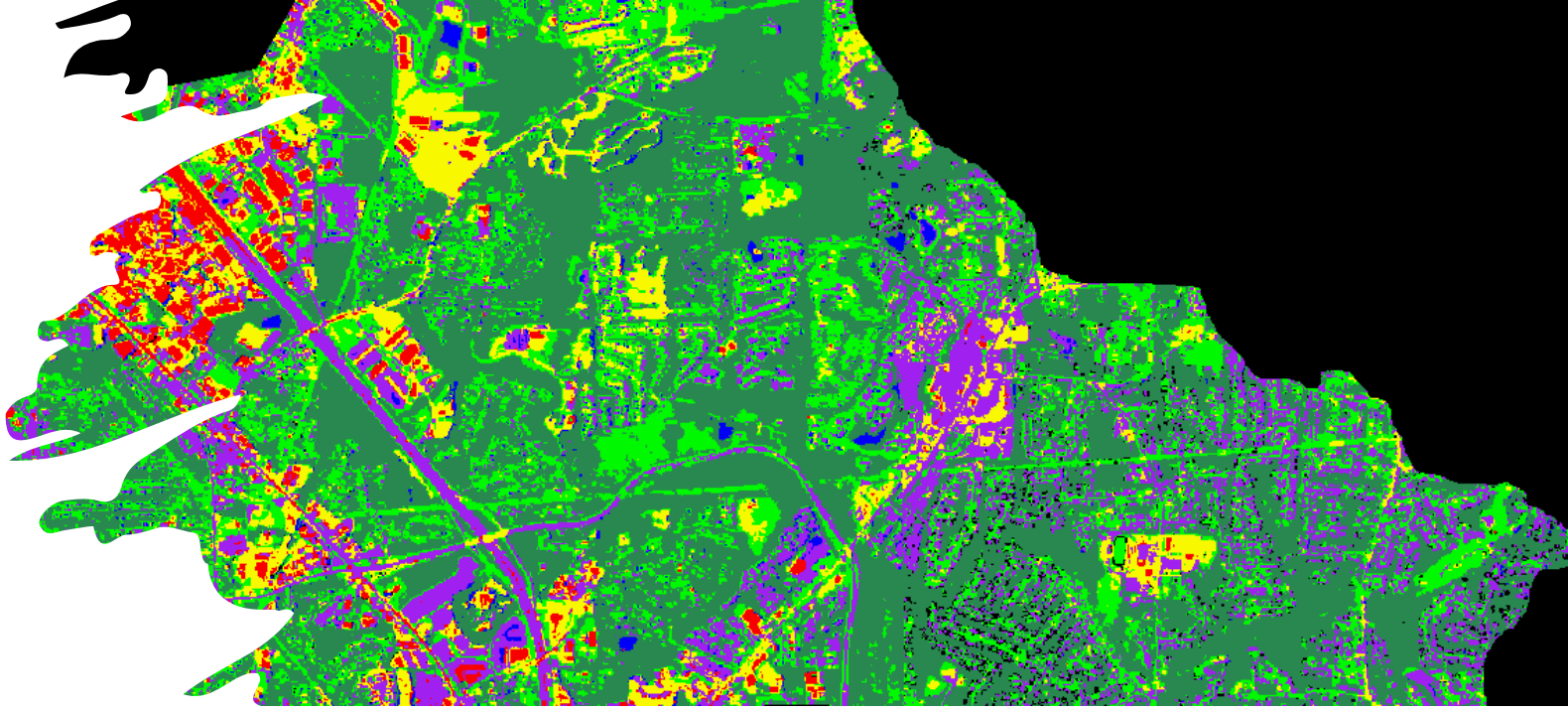


Figure 1 from Deventer et al. (2002) Scale-Dependent Effects of Landscape Context on Three Pollinator Guilds

Image interpretation



How do we assess data quality & accuracy?



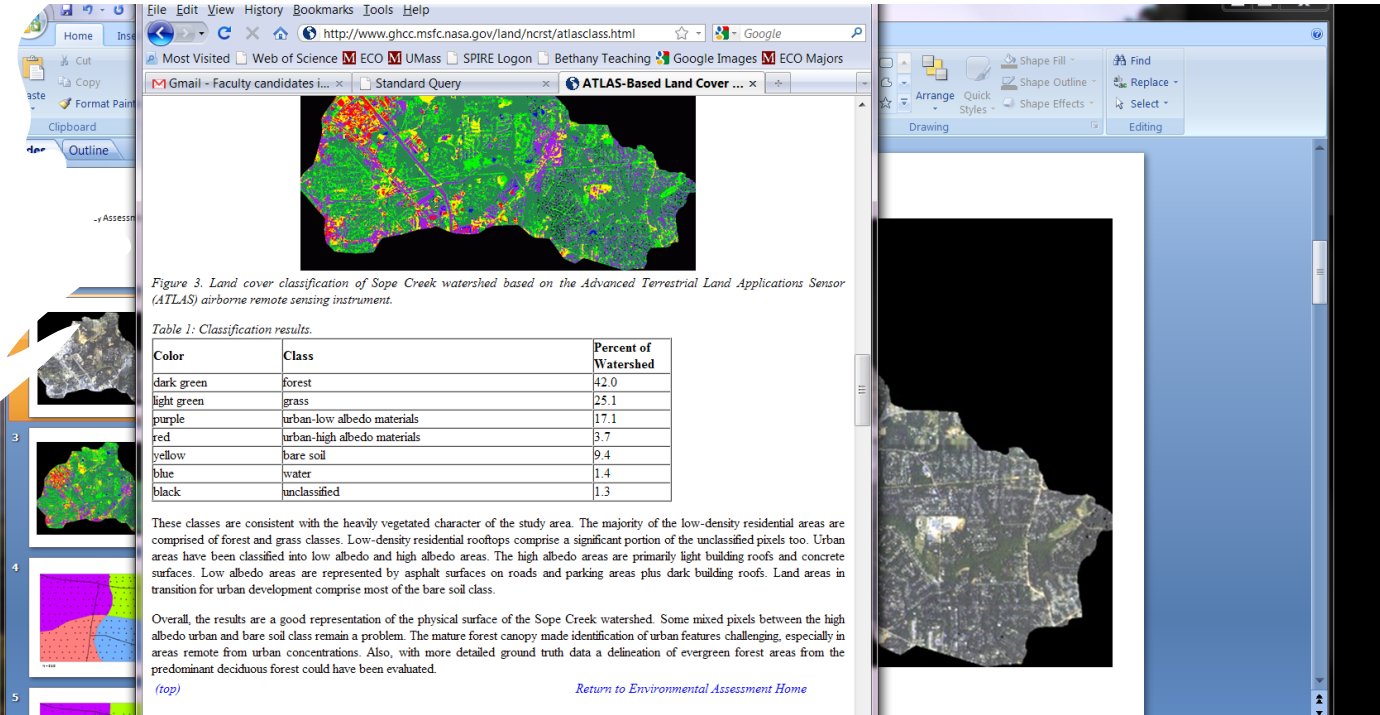


Figure 3. Land cover classification of Sope Creek watershed based on the Advanced Terrestrial Land Applications Sensor (ATLAS) airborne remote sensing instrument.

Table 1: Classification results.

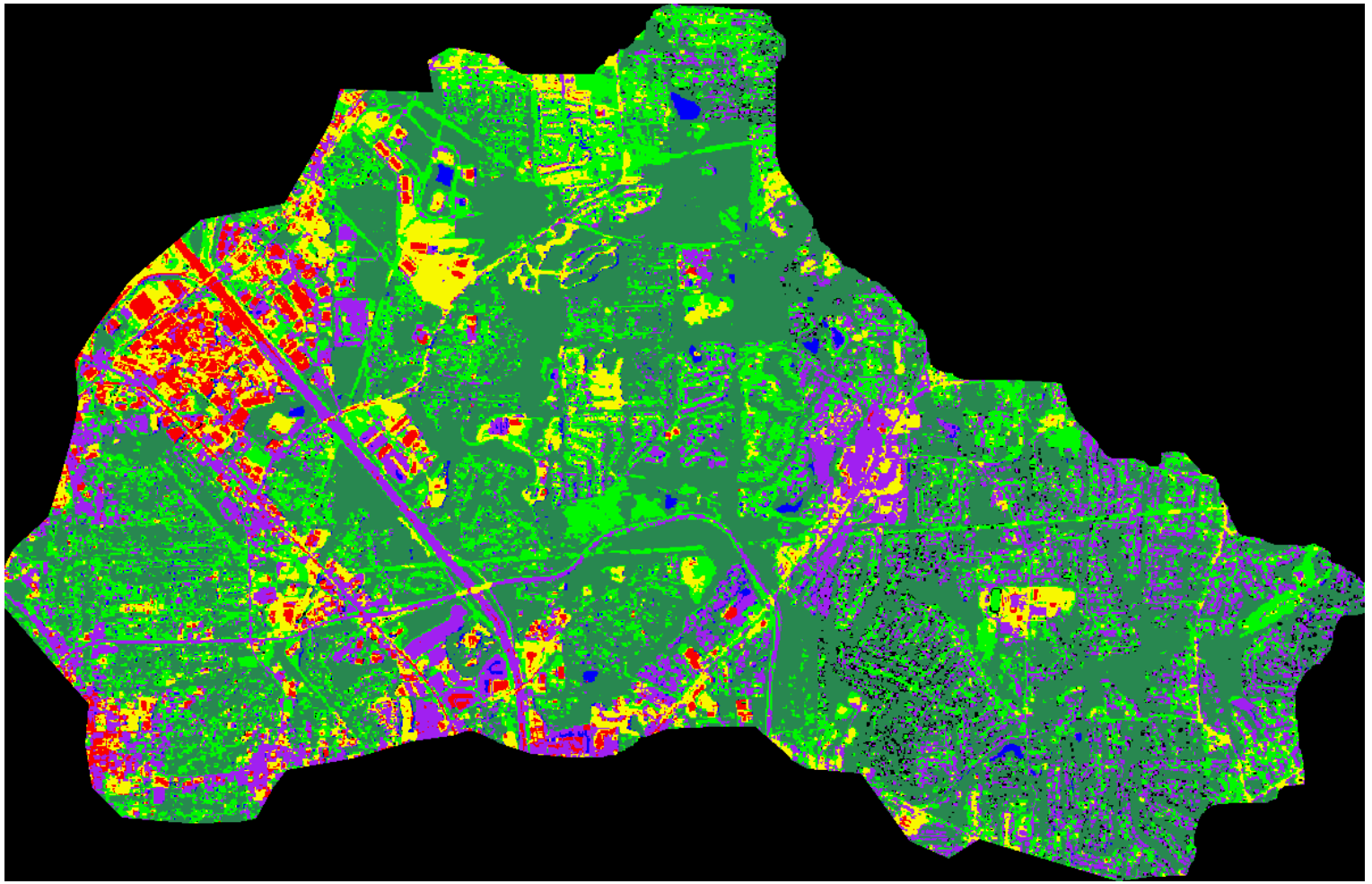
Color	Class	Percent of Watershed
dark green	forest	42.0
light green	grass	25.1
purple	urban-low albedo materials	17.1
red	urban-high albedo materials	3.7
yellow	bare soil	9.4
blue	water	1.4
black	unclassified	1.3

These classes are consistent with the heavily vegetated character of the study area. The majority of the low-density residential areas are comprised of forest and grass classes. Low-density residential rooftops comprise a significant portion of the unclassified pixels too. Urban areas have been classified into low albedo and high albedo areas. The high albedo areas are primarily light building roofs and concrete surfaces. Low albedo areas are represented by asphalt surfaces on roads and parking areas plus dark building roofs. Land areas in transition for urban development comprise most of the bare soil class.

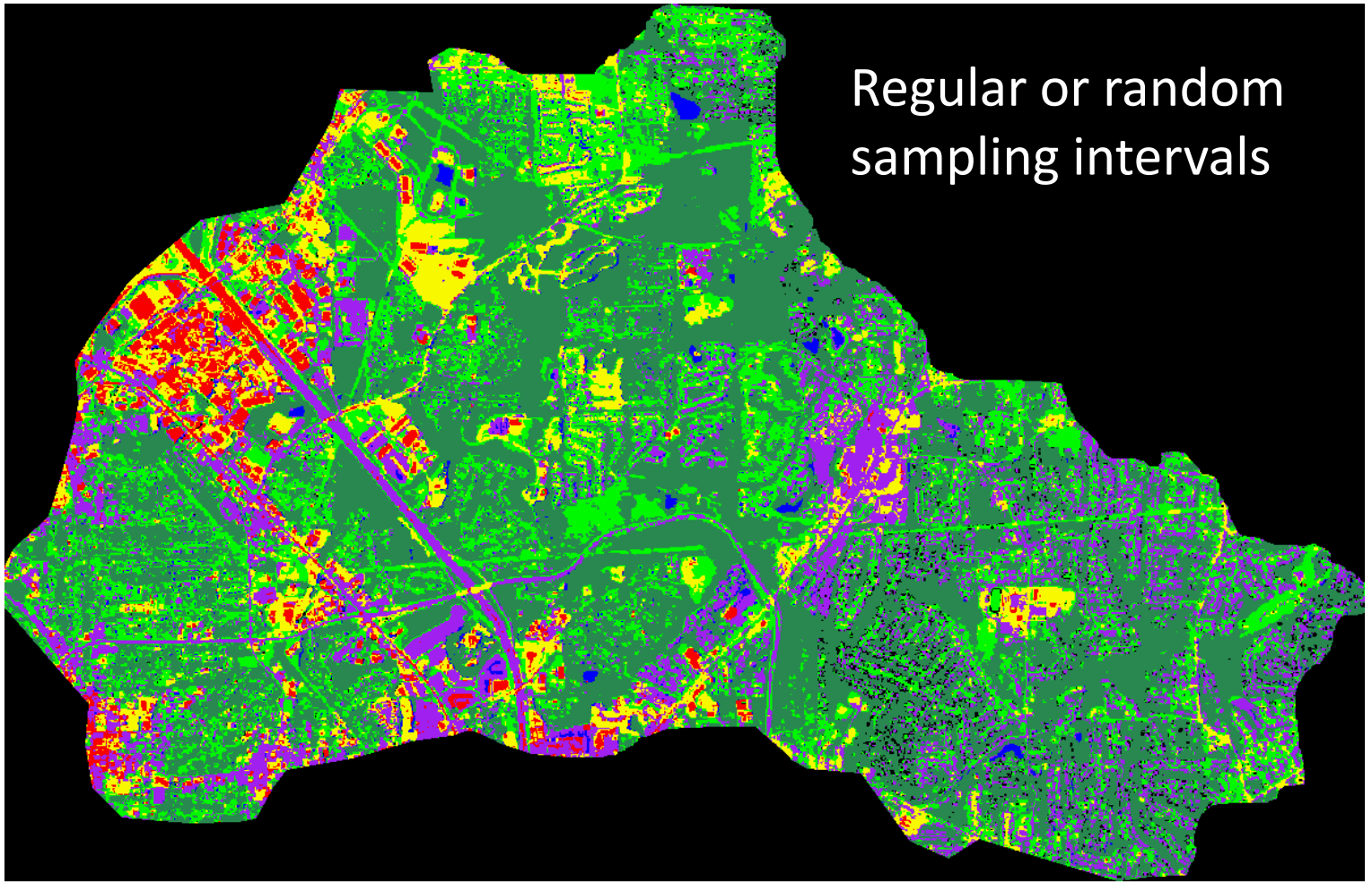
Overall, the results are a good representation of the physical surface of the Sope Creek watershed. Some mixed pixels between the high albedo urban and bare soil class remain a problem. The mature forest canopy made identification of urban features challenging, especially in areas remote from urban concentrations. Also, with more detailed ground truth data a delineation of evergreen forest areas from the predominant deciduous forest could have been evaluated.

(top) *Return to Environmental Assessment Home*

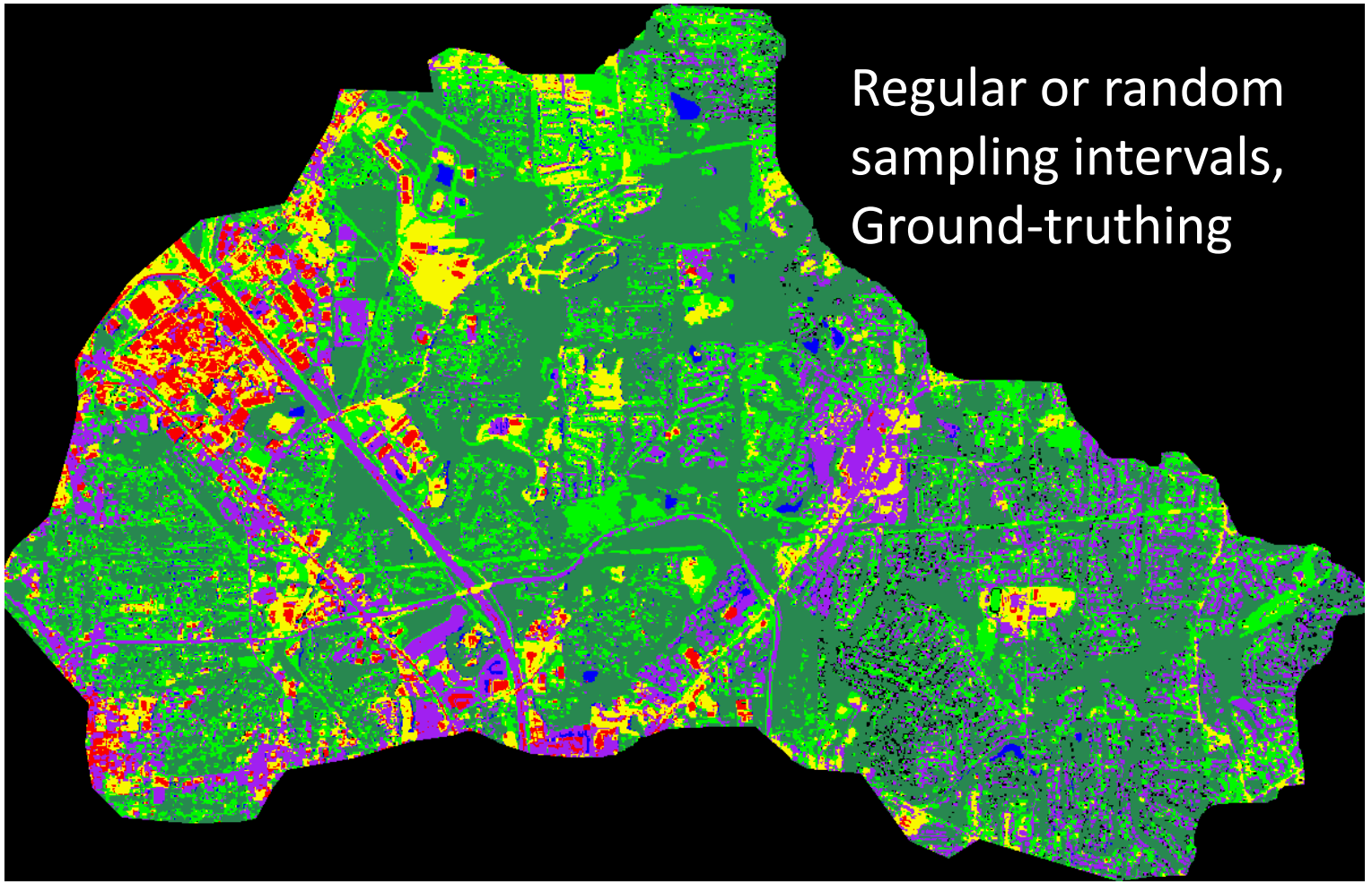
Collect a validation dataset



Collect a validation dataset



Collect a validation dataset



What kinds of errors are associated with land cover classification?

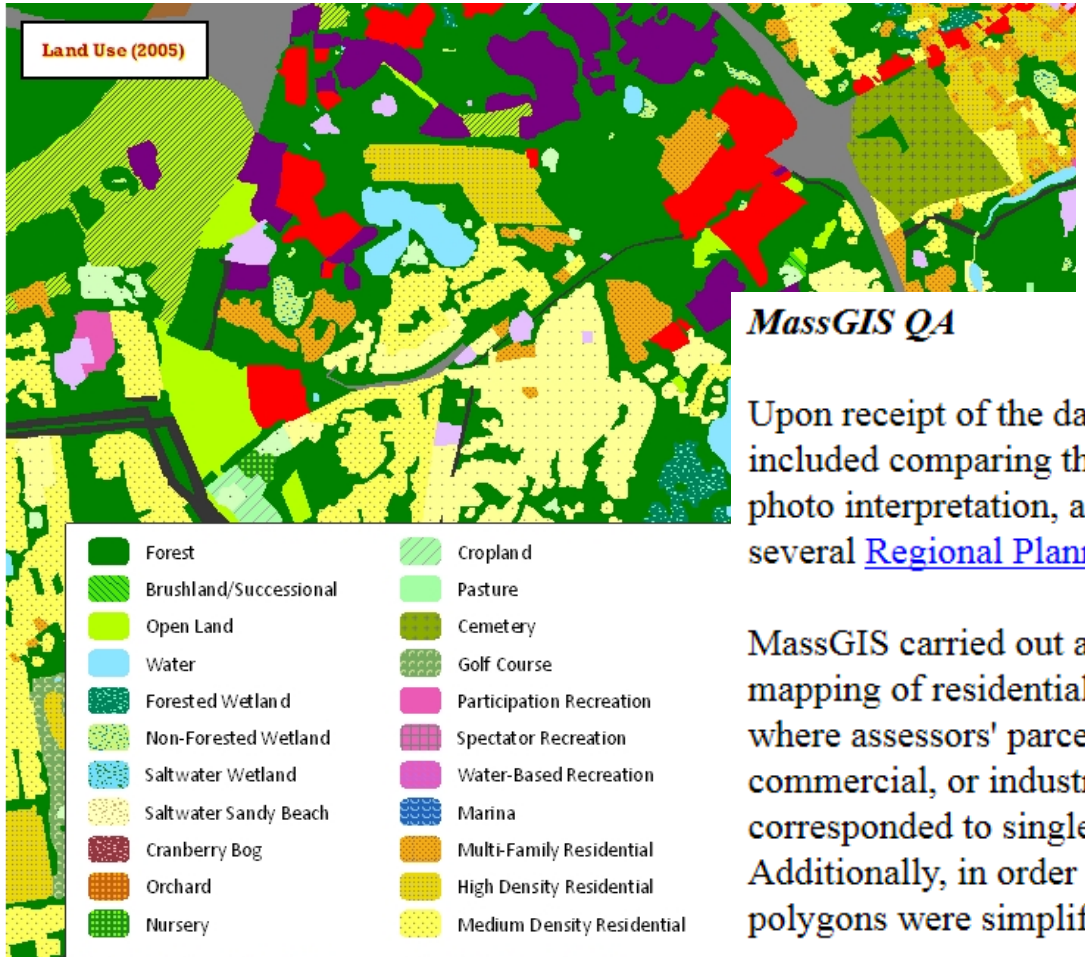
Over-estimation error *(mapping more of a particular land cover than actually exists)*

- *Commission Error*

Under-estimation error *(missing some of a particular land cover)*

- *Omission Error*

All land cover maps should have some quality assessment or accuracy assessment information



Sometimes that info is buried & vague

MassGIS QA

Upon receipt of the data, MassGIS performed extensive QA steps. These included comparing the draft land use with numerous ancillary datasets, manual photo interpretation, and field visits. Draft data was graciously reviewed by several Regional Planning Agencies across the state.

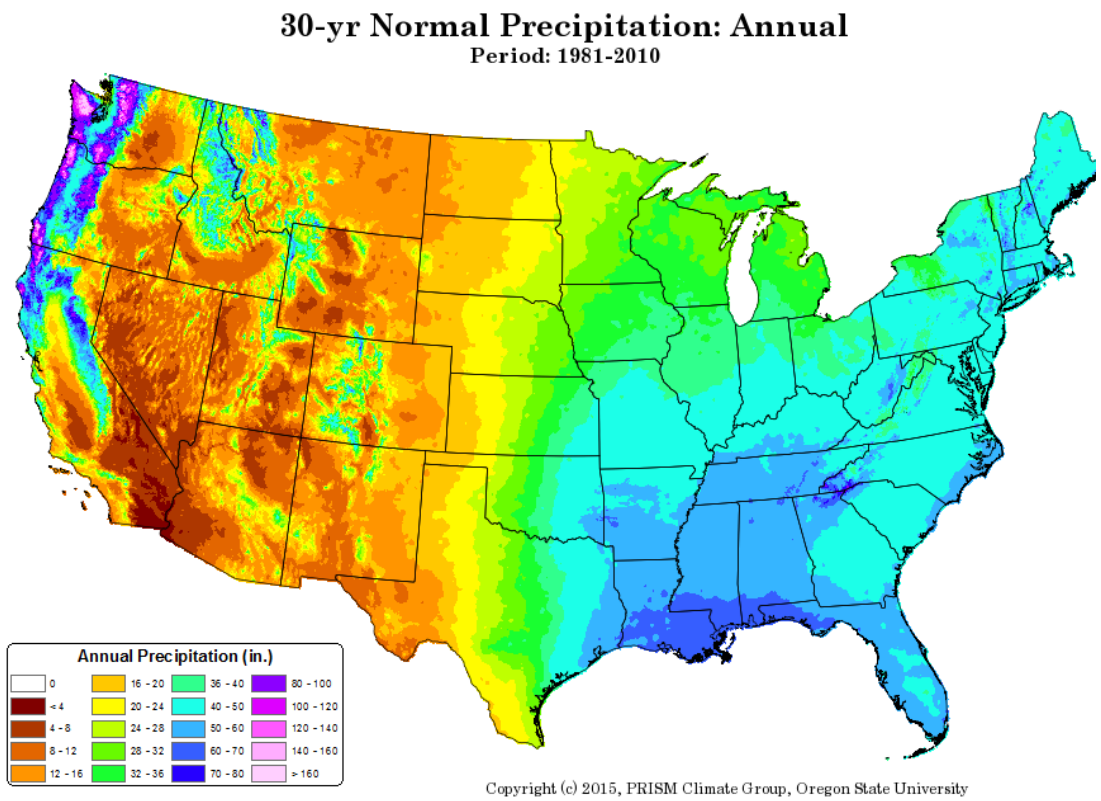
MassGIS carried out an additional post production step to further improve the mapping of residential, commercial, and industrial land use. In municipalities where assessors' parcels were available, the developed portions of residential, commercial, or industrial parcels with use codes that unambiguously corresponded to single land use codes were incorporated into the land use data. Additionally, in order to make the statewide dataset more manageable, polygons were simplified and slivers were removed.

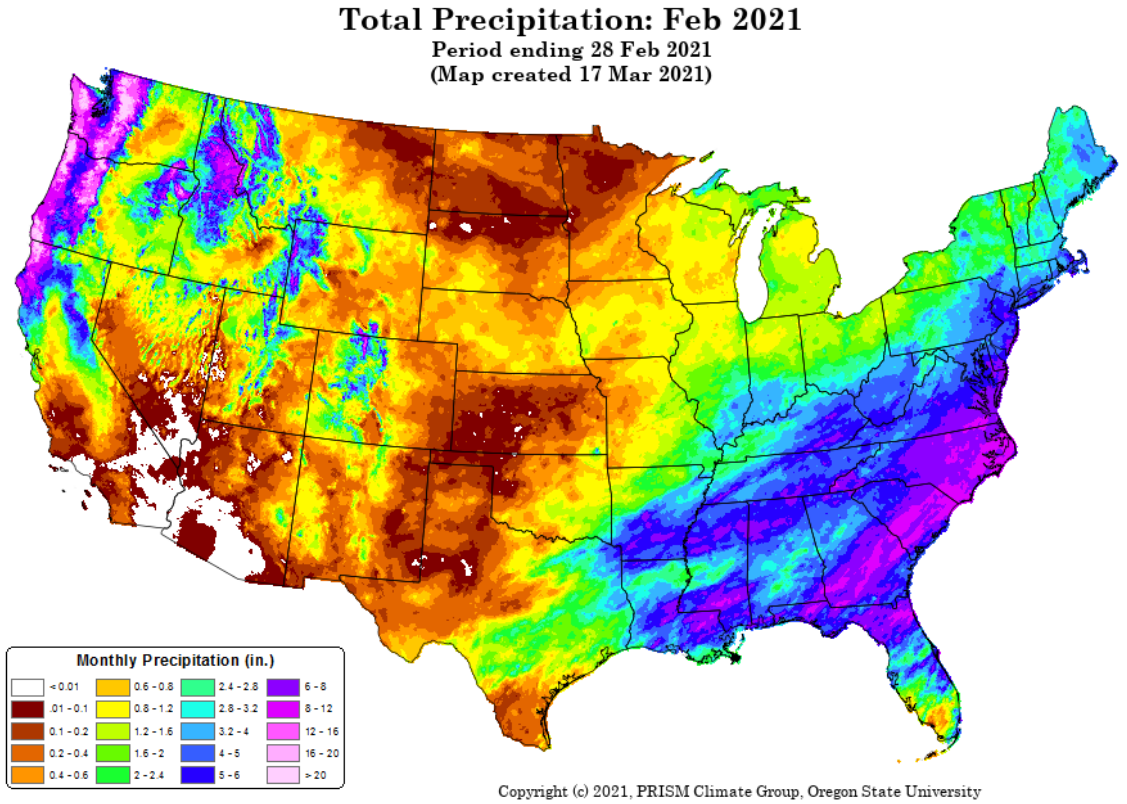
Suitability analyses require DATA

DATA	SOURCE
Land cover and land use maps	Satellite image classification, aerial photo interpretation
Climate	Weather stations, interpolation
Census map	Mailed surveys
Topographic maps	Surveys, RADAR
LULC Change maps	Digitized paper maps, aerial photo interpretation
Occurrence/abundance	Surveys, collars

Climate and weather data

- Historical records, trends
- Projections





Observed and Derived Variables

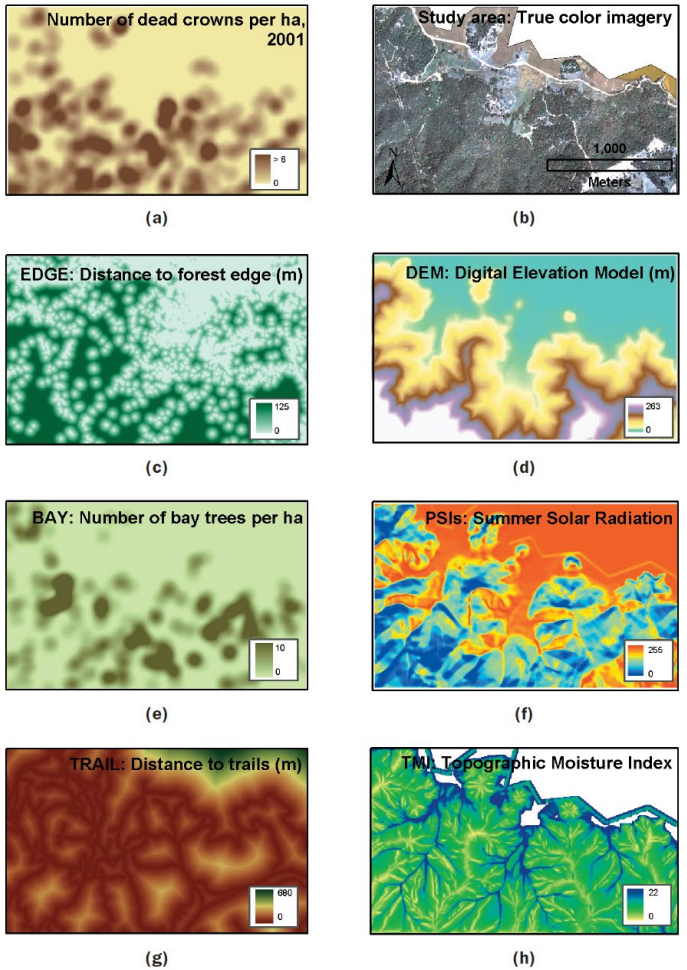


Plate 1. Derived environmental variables used in classification tree analysis. The clustering of dead crowns is shown in (a), and the study area from high-resolution imagery shown in true color (b) is presented for comparison. In (b) dark green areas are mixed oak woodlands affected with SOD; light areas indicate roads, paths, and gaps in forest cover. (c) and (e) are derived from high-resolution imagery (b). Potential Solar Insolation in the summer (f) and topographic moisture index (h) are derived from Digital Elevation Model (d). In (h) yellows indicate ridges, and blues indicate drainages.

Kelly and Meentemeyer (2002) Landscape Dynamics of the Spread of Sudden Oak Death

Interpolation

- Model-based approach to fill-in the areas in-between data points.
- Statistical and non-statistical approaches

Environment International 106 (2017) 234–247

Contents lists available at ScienceDirect

Environment International

journal homepage: www.elsevier.com/locate/envint

Mapping urban air quality in near real-time using observations from low-cost sensors and model information

Philipp Schneider*, Nuria Castell, Matthias Vogt, Franck R. Dauge, William A. Lahoz, Alena Bartonova

NILU - Norwegian Institute for Air Research, PO Box 100, Kjeller 2027, Norway

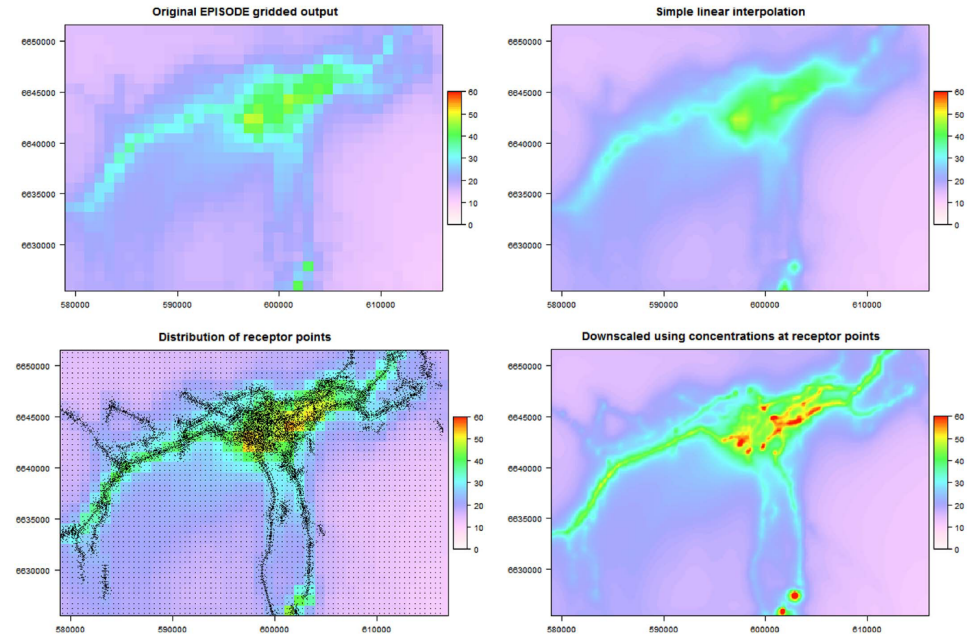


Fig. 2. Downscaling method for EPISODE, shown for the 2011 annual mean surface NO_2 field for Oslo, Norway, in units of μg m^{-3}. Top left panel: Original gridded output from the EPISODE model at 1000 m horizontal spatial resolution. Top right panel: The original field regridded to 100 m horizontal spatial resolution using simple linear interpolation as a reference. Bottom left panel: Original gridded EPISODE concentration with locations of receptor points overlaid in black. Bottom right panel: Downscaled concentration field at 100 m horizontal spatial resolution derived through spatial interpolation of receptor point concentrations. Coordinates in UTM32N/WGS84.

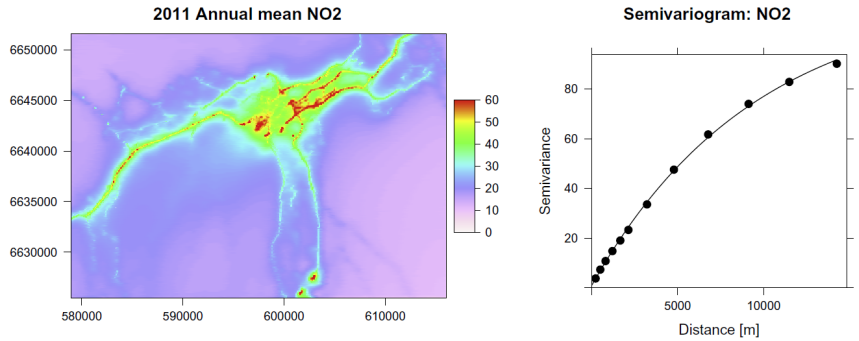


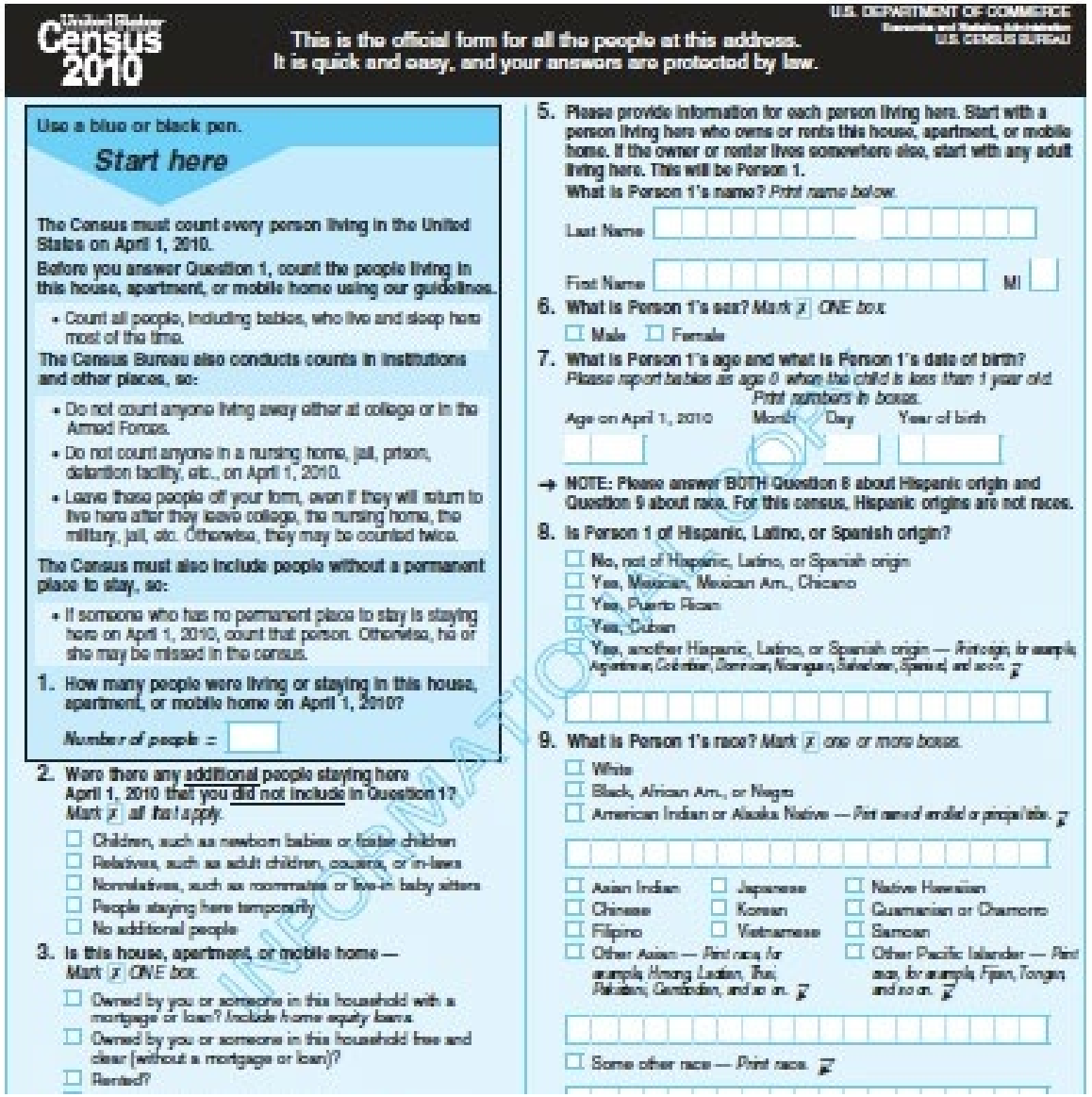
Fig. 3. Downscaled model-derived basemap (left panel) for NO_2 as produced by the EPISODE model and the corresponding semivariogram (right panel) for Oslo, Norway. The semivariogram was calculated over a total of 96,831 grid points. The concentrations are given in μg m^{-3}. Coordinates in UTM32N/WGS84.

Suitability analyses require DATA

DATA	SOURCE
Land cover and land use maps	Satellite image classification, aerial photo interpretation
Climate	Weather stations, interpolation
Census map	Mailed surveys
Topographic maps	Surveys, RADAR
LULC Change maps	Digitized paper maps, aerial photo interpretation
Occurrence/abundance	Surveys, collars

Demographic Data

- Number of people
- Age
- Sex
- Race
- Income
- Education



Census 2010

This is the official form for all the people at this address.
It is quick and easy, and your answers are protected by law.

U.S. DEPARTMENT OF COMMERCE
U.S. CENSUS BUREAU

Use a blue or black pen.

Start here

The Census must count every person living in the United States on April 1, 2010.

Before you answer Question 1, count the people living in this house, apartment, or mobile home using our guidelines.

- Count all people, including babies, who live and sleep here most of the time.

The Census Bureau also conducts counts in institutions and other places, so:

- Do not count anyone living away either at college or in the Armed Forces.
- Do not count anyone in a nursing home, jail, prison, detention facility, etc., on April 1, 2010.
- Leave these people off your form, even if they will return to live here after they leave college, the nursing home, the military, jail, etc. Otherwise, they may be counted twice.

The Census must also include people without a permanent place to stay, so:

- If someone who has no permanent place to stay is staying here on April 1, 2010, count that person. Otherwise, he or she may be missed in the census.

1. How many people were living or staying in this house, apartment, or mobile home on April 1, 2010?

   Number of people =

2. Were there any additional people staying here April 1, 2010 that you did not include in Question 1? Mark [x] all that apply.

   - ☐ Children, such as newborn babies or foster children
   - ☐ Relatives, such as adult children, cousins, or in-laws
   - ☐ Nonrelatives, such as roommates or live-in baby sitters
   - ☐ People staying here temporarily
   - ☐ No additional people

3. Is this house, apartment, or mobile home — Mark [x] ONE box.

   - ☐ Owned by you or someone in this household with a mortgage or loan? Include home equity loans.
   - ☐ Owned by you or someone in this household free and clear (without a mortgage or loan)?
   - ☐ Rented?

5. Please provide information for each person living here. Start with a person living here who owns or rents this house, apartment, or mobile home. If the owner or renter lives somewhere else, start with any adult living here. This will be Person 1.

   What is Person 1's name? Print name below.

   Last Name

   First Name MI

6. What is Person 1's sex? Mark [x] ONE box.

   ☐ Male ☐ Female

7. What is Person 1's age and what is Person 1's date of birth?
   Please report babies as age 0 when the child is less than 1 year old.
   Print numbers in boxes.

   Age on April 1, 2010 | Month | Day | Year of birth

→ NOTE: Please answer BOTH Question 8 about Hispanic origin and Question 9 about race. For this census, Hispanic origins are not races.

8. Is Person 1 of Hispanic, Latino, or Spanish origin?

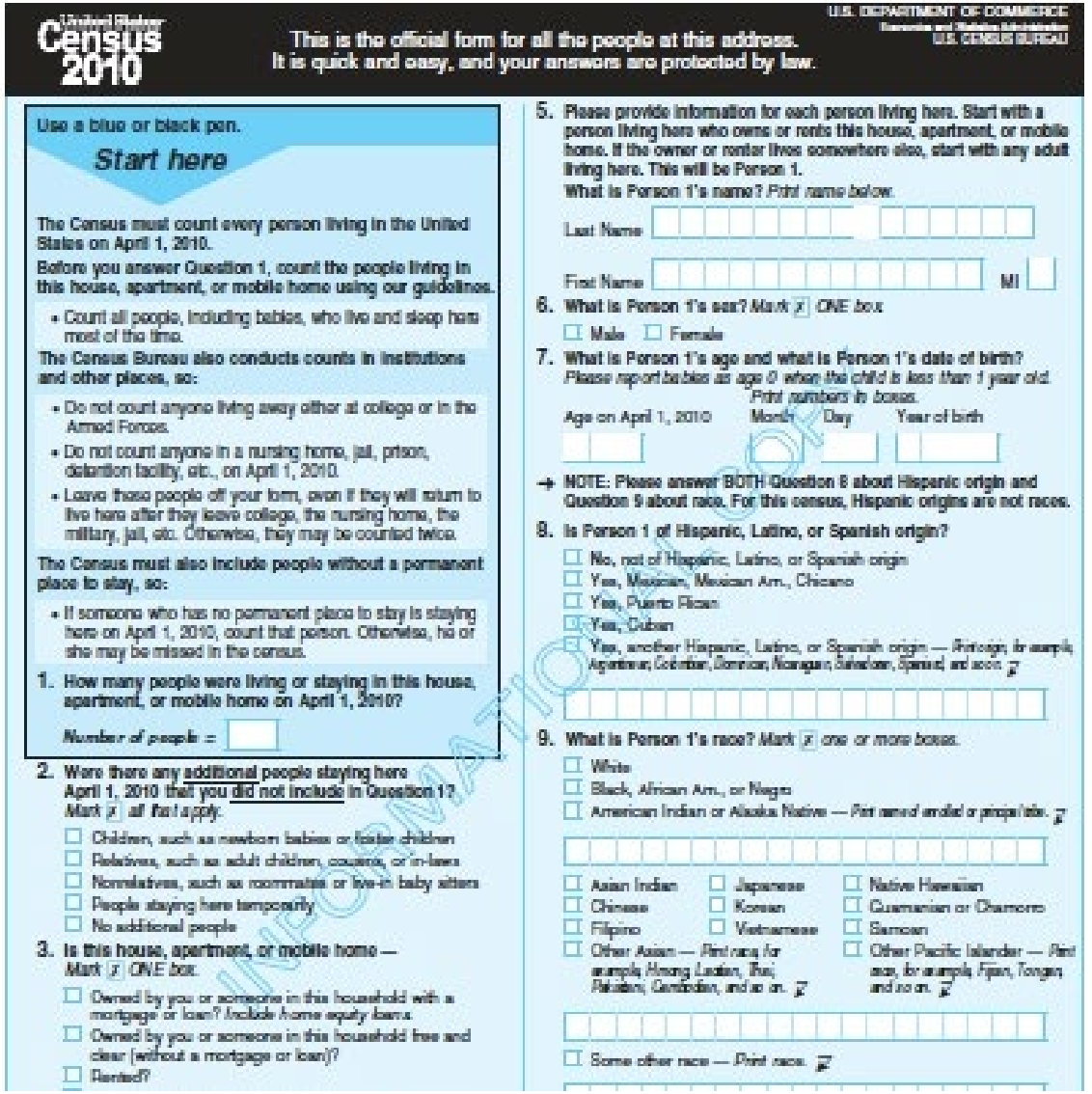
   - ☐ No, not of Hispanic, Latino, or Spanish origin
   - ☐ Yes, Mexican, Mexican Am., Chicano
   - ☐ Yes, Puerto Rican
   - ☐ Yes, Cuban
   - ☐ Yes, another Hispanic, Latino, or Spanish origin — Print origin, for example, Argentinean, Colombian, Dominican, Nicaraguan, Salvadoran, Spaniard, and so on.

9. What is Person 1's race? Mark [x] one or more boxes.

   - ☐ White
   - ☐ Black, African Am., or Negro
   - ☐ American Indian or Alaska Native — Print name of enrolled or principal tribe.
   - ☐ Asian Indian
   - ☐ Chinese
   - ☐ Filipino
   - ☐ Japanese
   - ☐ Korean
   - ☐ Vietnamese
   - ☐ Native Hawaiian
   - ☐ Guamanian or Chamorro
   - ☐ Samoan
   - ☐ Other Asian — Print race, for example, Hmong, Laotian, Thai, Pakistani, Cambodian, and so on.
   - ☐ Other Pacific Islander — Print race, for example, Fijian, Tongan, and so on.
   - ☐ Some other race — Print race.

Demographic Data

- **Address: city, state, county, etc.**
- **We can use database operations to join demographic and spatial data.**



Census 2010

This is the official form for all the people at this address.
It is quick and easy, and your answers are protected by law.

U.S. DEPARTMENT OF COMMERCE
U.S. CENSUS BUREAU

Use a blue or black pen.

Start here

The Census must count every person living in the United States on April 1, 2010.

Before you answer Question 1, count the people living in this house, apartment, or mobile home using our guidelines.

- Count all people, including babies, who live and sleep here most of the time.

The Census Bureau also conducts counts in institutions and other places, so:

- Do not count anyone living away either at college or in the Armed Forces.
- Do not count anyone in a nursing home, jail, prison, detention facility, etc., on April 1, 2010.
- Leave those people off your form, even if they will return to live here after they leave college, the nursing home, the military, jail, etc. Otherwise, they may be counted twice.

The Census must also include people without a permanent place to stay, so:

- If someone who has no permanent place to stay is staying here on April 1, 2010, count that person. Otherwise, he or she may be missed in the census.

1. How many people were living or staying in this house, apartment, or mobile home on April 1, 2010?

   Number of people =

2. Were there any additional people staying here April 1, 2010 that you did not include in Question 1? Mark [x] all that apply.
   - Children, such as newborn babies or foster children
   - Relatives, such as adult children, cousins, or in-laws
   - Nonrelatives, such as roommates or live-in baby sitters
   - People staying here temporarily
   - No additional people

3. Is this house, apartment, or mobile home — Mark [x] ONE box.
   - Owned by you or someone in this household with a mortgage or loan? Include home equity loans.
   - Owned by you or someone in this household free and clear (without a mortgage or loan)?
   - Rented?

5. Please provide information for each person living here. Start with a person living here who owns or rents this house, apartment, or mobile home. If the owner or renter lives somewhere else, start with any adult living here. This will be Person 1.
   What is Person 1's name? Print name below.

   Last Name

   First Name MI

6. What is Person 1's sex? Mark [x] ONE box.
   - Male
   - Female

7. What is Person 1's age and what is Person 1's date of birth?
   Please report babies as age 0 when the child is less than 1 year old.
   Print numbers in boxes.

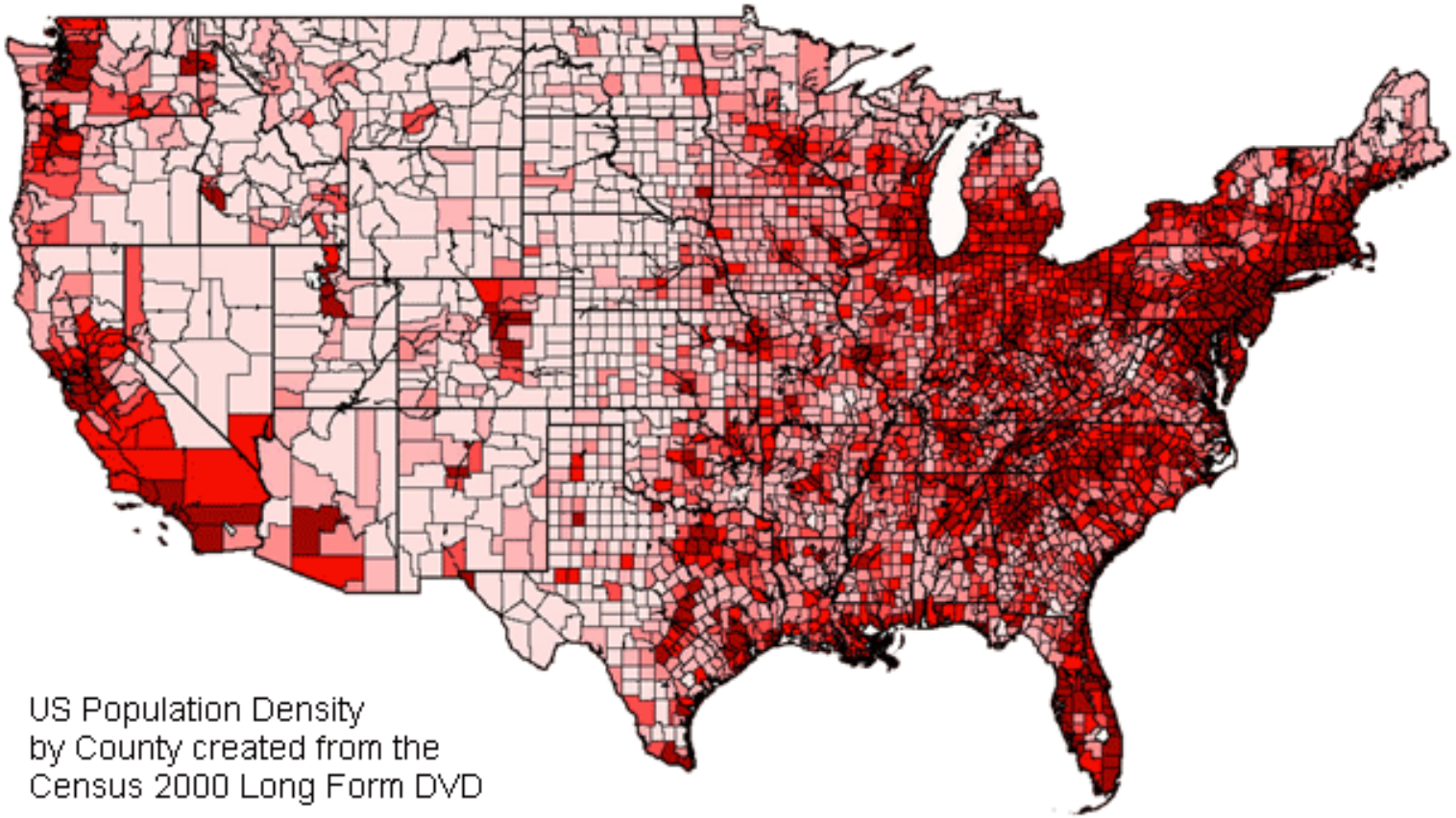
   Age on April 1, 2010 | Month | Day | Year of birth

→ NOTE: Please answer BOTH Question 8 about Hispanic origin and Question 9 about race. For this census, Hispanic origins are not races.

8. Is Person 1 of Hispanic, Latino, or Spanish origin?
   - No, not of Hispanic, Latino, or Spanish origin
   - Yes, Mexican, Mexican Am., Chicano
   - Yes, Puerto Rican
   - Yes, Cuban
   - Yes, another Hispanic, Latino, or Spanish origin — Print origin, for example, Argentinian, Colombian, Dominican, Nicaraguan, Salvadoran, Spaniard, and so on.

9. What is Person 1's race? Mark [x] one or more boxes.
   - White
   - Black, African Am., or Negro
   - American Indian or Alaska Native — Print name of enrolled or principal tribe.
   - Asian Indian
   - Japanese
   - Native Hawaiian
   - Chinese
   - Korean
   - Guamanian or Chamorro
   - Filipino
   - Vietnamese
   - Samoan
   - Other Asian — Print race, for example, Hmong, Laotian, Thai, Pakistani, Cambodian, and so on.
   - Other Pacific Islander — Print race, for example, Fijian, Tongan, and so on.
   - Some other race — Print race.

Population density by county

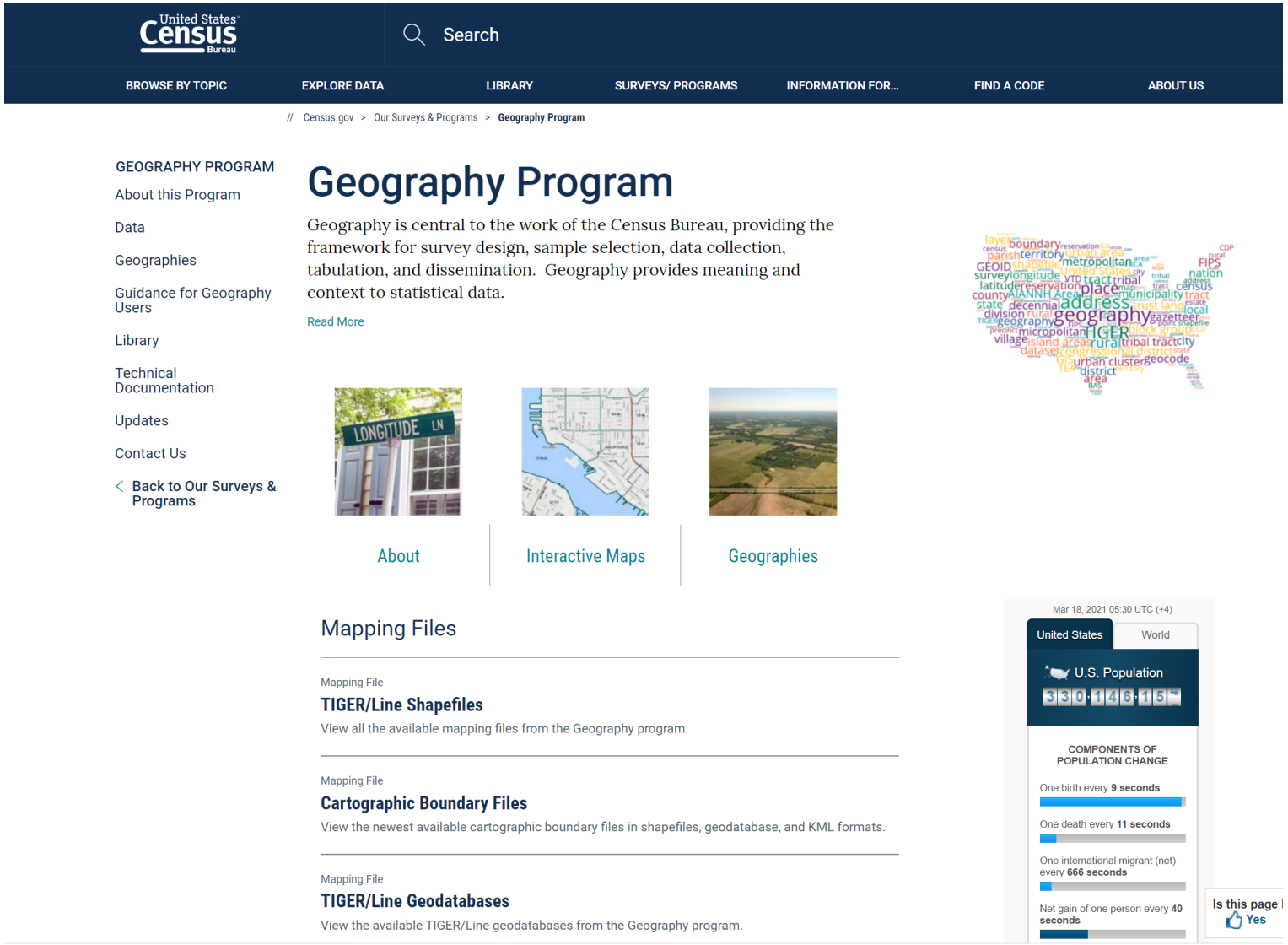


US Population Density
by County created from the
Census 2000 Long Form DVD

County-level census information

NAME	STATE_NAME	POP1999	MALES	FEMALES	WHITE	BLACK	AMERI_E	ASIAN_PI	HISPANIC	AGE_5_17	AGE_18_29	AGE_30_49	AGE_50_64
Phillips	Montana	4744	2537	2626	4741	2	390	14	44	1176	690	1399	716
Valley	Montana	8136	4110	4129	7423	9	770	19	62	1703	981	2309	1313
Daniels	Montana	1959	1115	1151	2242	0	6	18	12	475	189	608	389
Whatcom	Washington	159393	62848	64932	119229	650	4014	2363	3718	23201	26341	38357	14998
Bonner	Idaho	35901	13231	13391	26210	37	220	71	352	5721	2972	8332	3913
Ward	North Dakota	58560	28824	29097	54545	1411	962	594	857	11241	13497	15402	6259
Koochiching	Minnesota	15420	8440	7859	15633	45	451	50	185	3117	2513	4757	2443
Skagit	Washington	101320	39205	40350	74133	280	1712	782	4335	15167	11947	23329	11020
Williams	North Dakota	20025	10386	10743	20025	18	1010	43	110	4733	2883	6184	2728
McHenry	North Dakota	6024	3309	3219	6498	4	13	11	13	1377	676	1658	1064
St. Louis	Minnesota	192958	96435	101778	192053	1106	3682	1076	952	36136	33054	56233	27138
San Juan	Washington	12738	4967	5068	9811	23	79	86	121	1480	803	3316	1706
Roosevelt	Montana	10936	5374	5625	5569	17	5355	40	103	2684	1664	2982	1272
Mountrial	North Dakota	6590	3469	3552	5606	4	1395	14	25	1603	812	1861	951
Marshall	Minnesota	10190	5566	5427	10889	2	50	14	113	2474	1293	2866	1651
Ramsey	North Dakota	11973	6221	6460	12022	21	591	30	49	2462	1935	3350	1736
Walsh	North Dakota	13422	6890	6950	13453	17	97	59	441	2844	1826	3640	1952
Beltrami	Minnesota	39000	17011	17373	28409	100	5641	194	146	7356	7343	8949	3868
Pierce	North Dakota	4597	2498	2554	5011	2	23	15	1	1015	598	1142	802
Chelan	Washington	60730	25780	26470	48333	80	487	378	4786	9939	7760	15293	7026
Pondera	Montana	6424	3198	3235	5691	5	704	20	31	1448	792	1728	897
Clallam	Washington	64786	28084	28380	52509	321	2695	614	1150	10096	7312	15661	8274
Benson	North Dakota	6904	3631	3567	4417	0	2772	3	24	1787	996	1669	911
Chouteau	Montana	5125	2782	2670	5221	4	212	10	25	1153	622	1518	822
Snohomish	Washington	604856	232194	233448	434536	4767	6422	16467	10656	89762	83391	155542	53413
Island	Washington	71454	31340	28855	55034	1454	480	2553	2006	10554	12434	16696	7323
Sanders	Montana	10200	4377	4292	8135	12	471	37	104	1882	956	2509	1322

The US Census Bureau website is a rich source of spatial data.

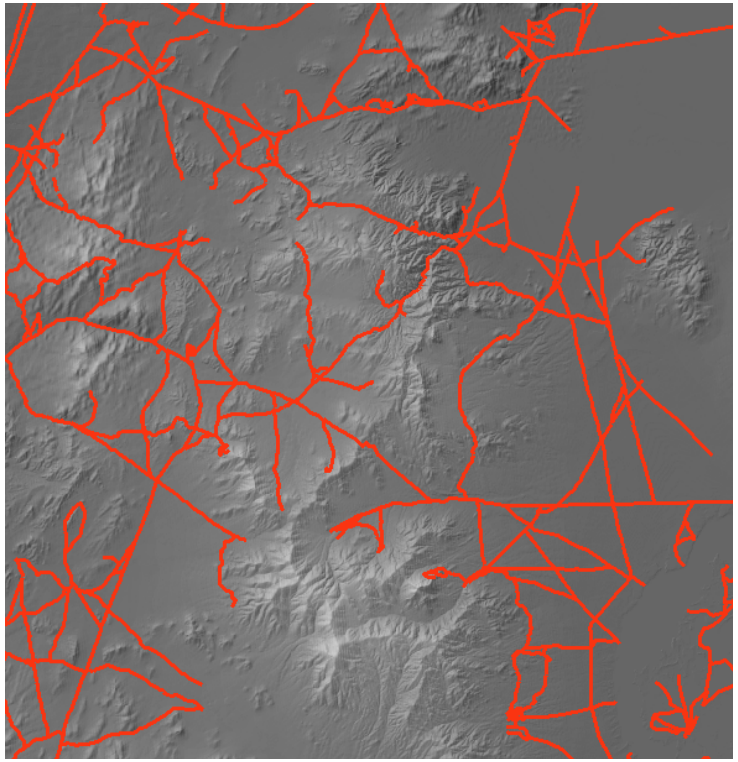


TIGER/line Shapefiles!

U.S. Census Bureau

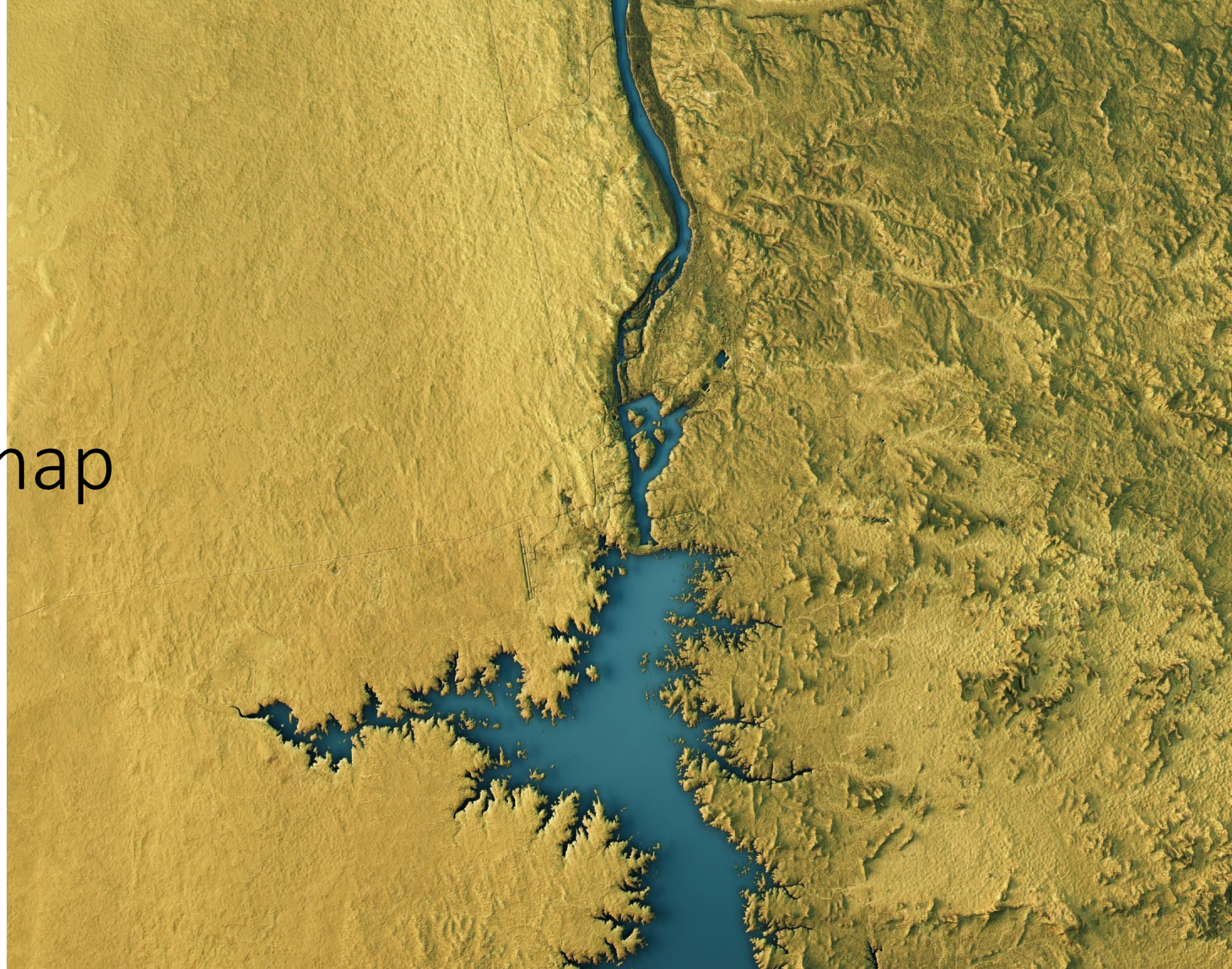
TIGER Navigation

- 2010 Census TIGER/Line Shapefiles Main
  - Download Shapefiles
  - Release Schedule
  - Technical Documentation
  - User Notes
- Previous Versions
  - 2009 TIGER/Line Shapefiles
  - 2008 TIGER/Line Shapefiles
  - 2007 TIGER/Line Shapefiles
  - TIGER/Line Files
- Other 2010 Census Geographic Products
- Geography Main Page



Other geocoded road information

But first, a map puzzler!



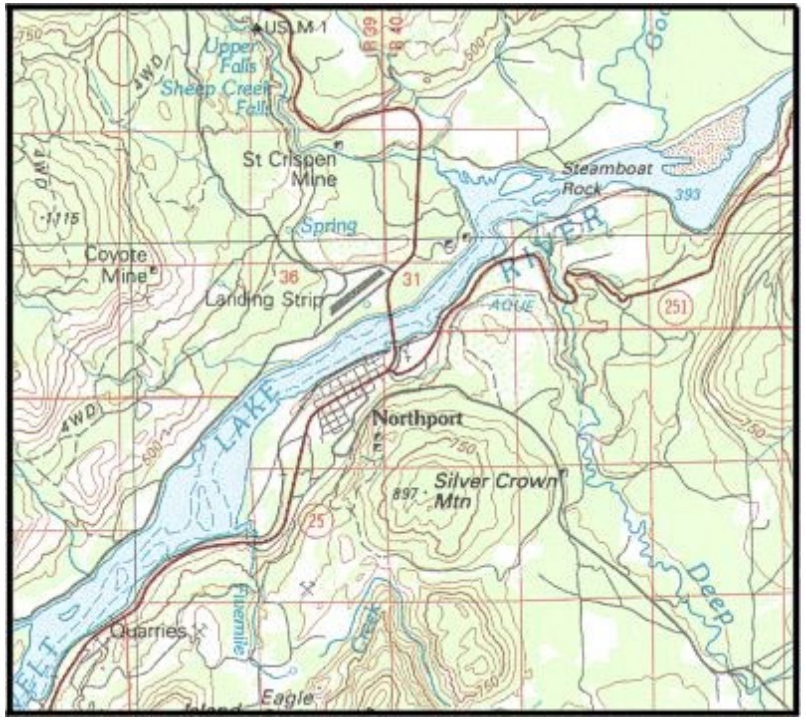

3D

Suitability analyses require DATA

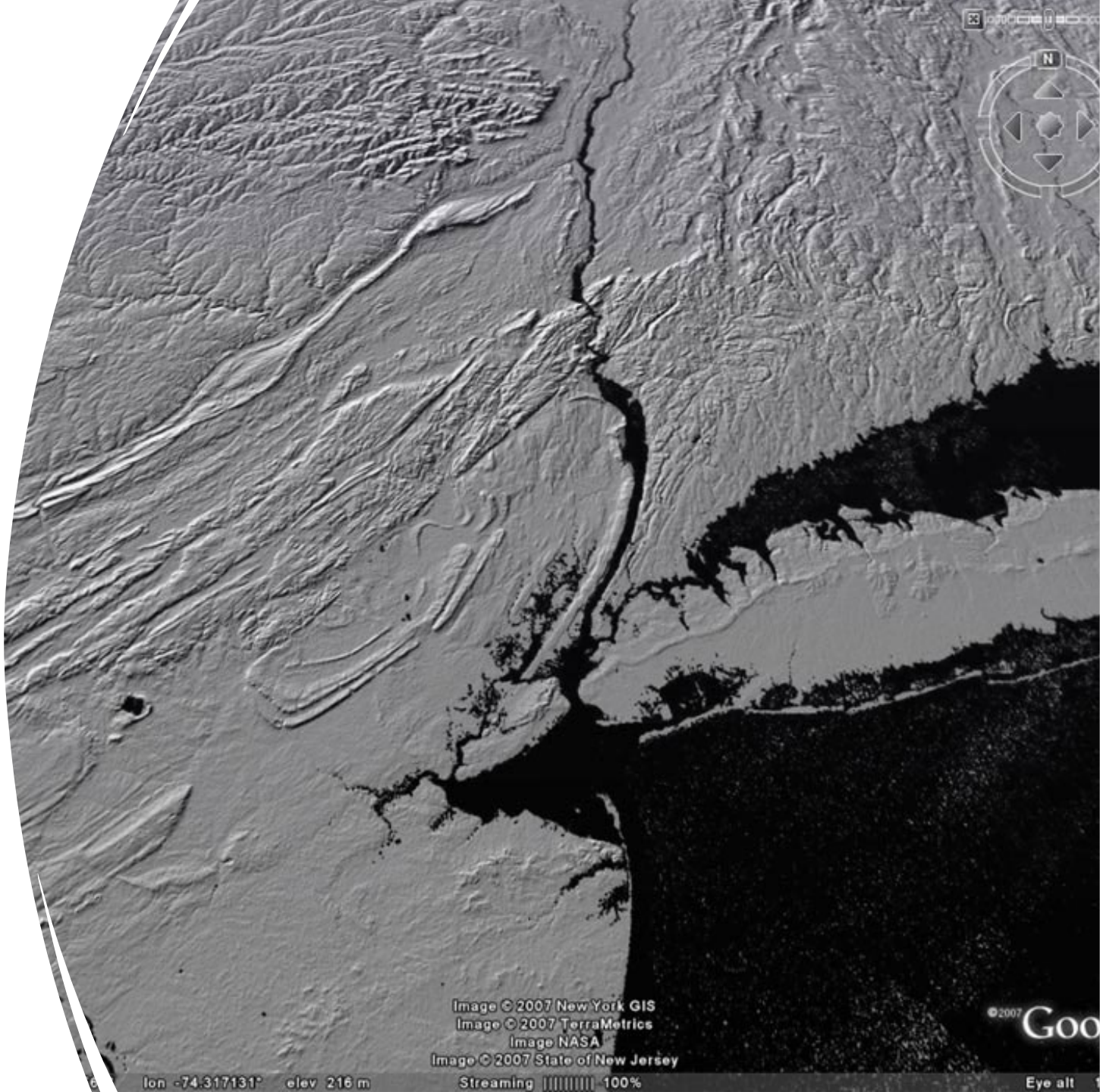
DATA	SOURCE
Land cover and land use maps	Satellite image classification, aerial photo interpretation
Climate	Weather stations, interpolation
Census map	Mailed surveys
Topography/Terrain maps	Surveys, RADAR
LULC Change maps	Digitized paper maps, aerial photo interpretation
Occurrence/abundance	Surveys, collars

Old-school topographic maps

- Example: USGS quadrangle topo lines defined based on widespread surveys

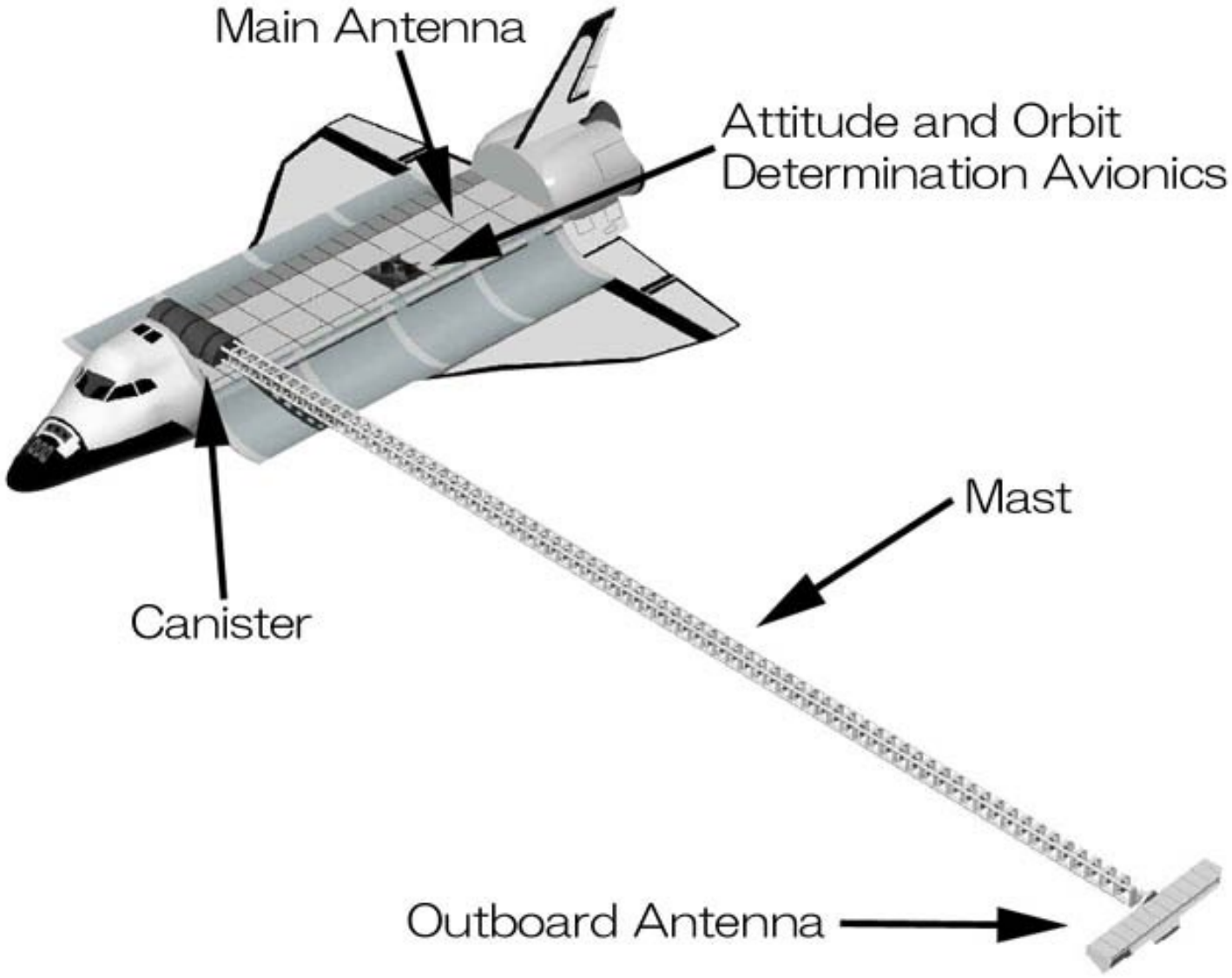


New-school topographic maps: SRTM

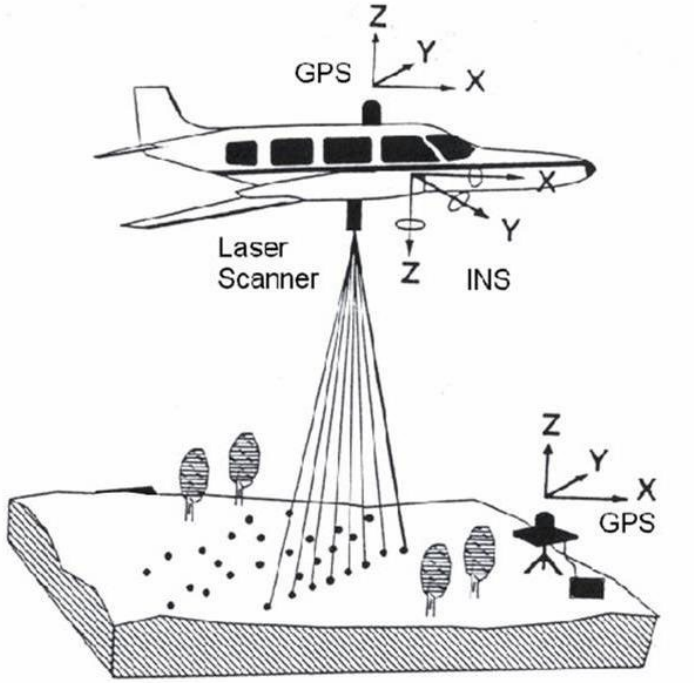


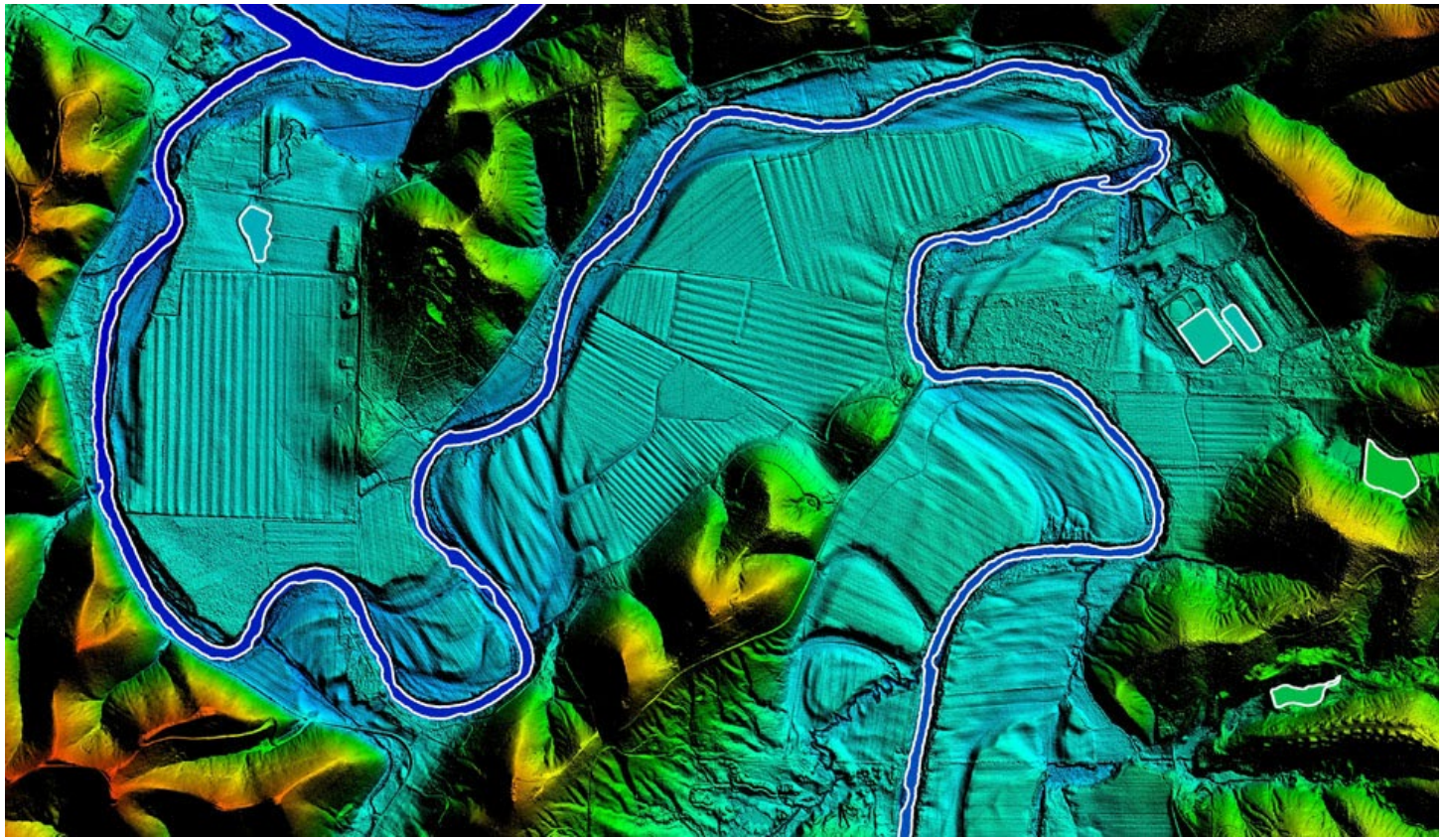
New-school topographic maps: SRTM

Shuttle Radar Topography Mission

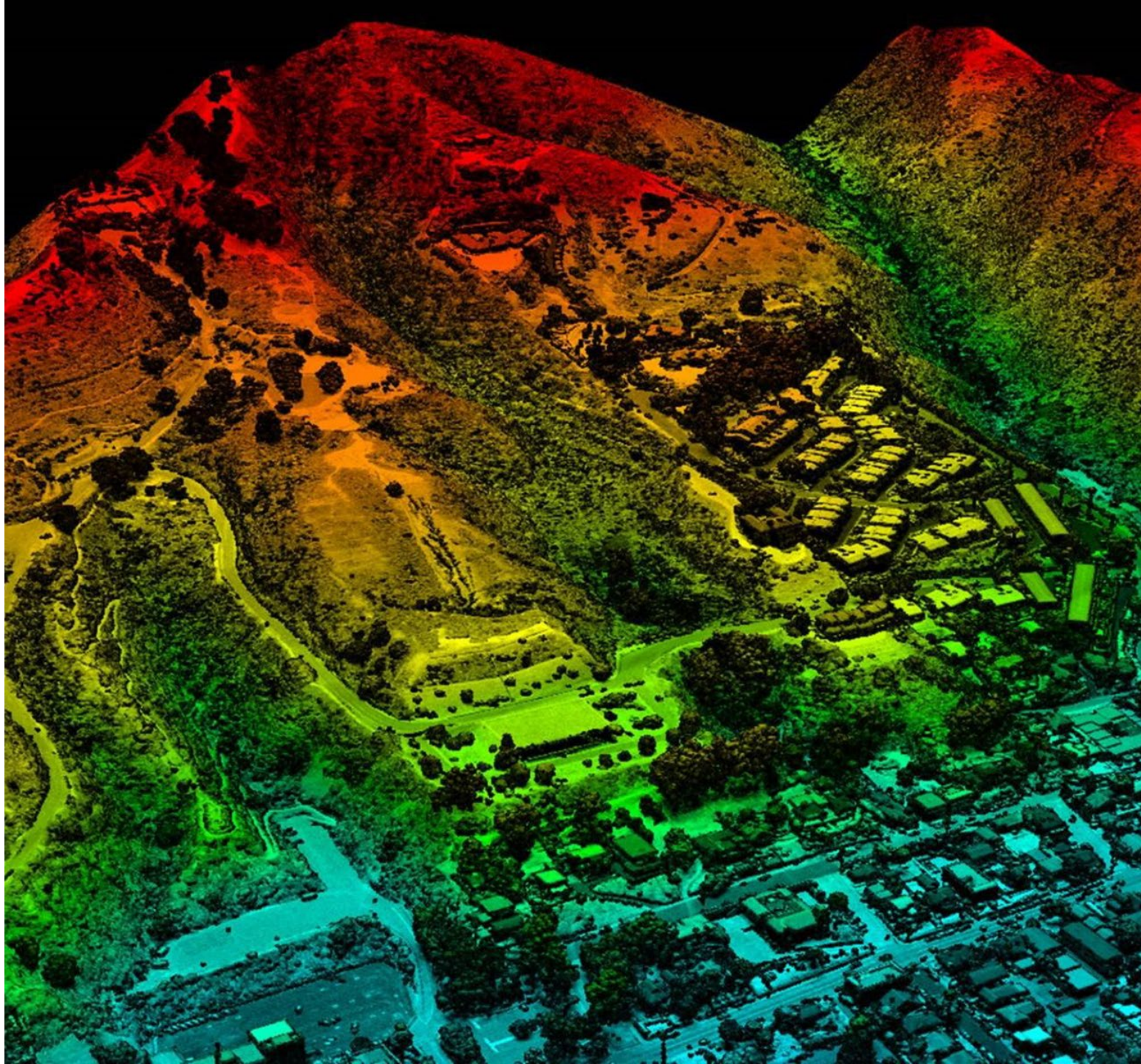


Topographic Maps from LASERS



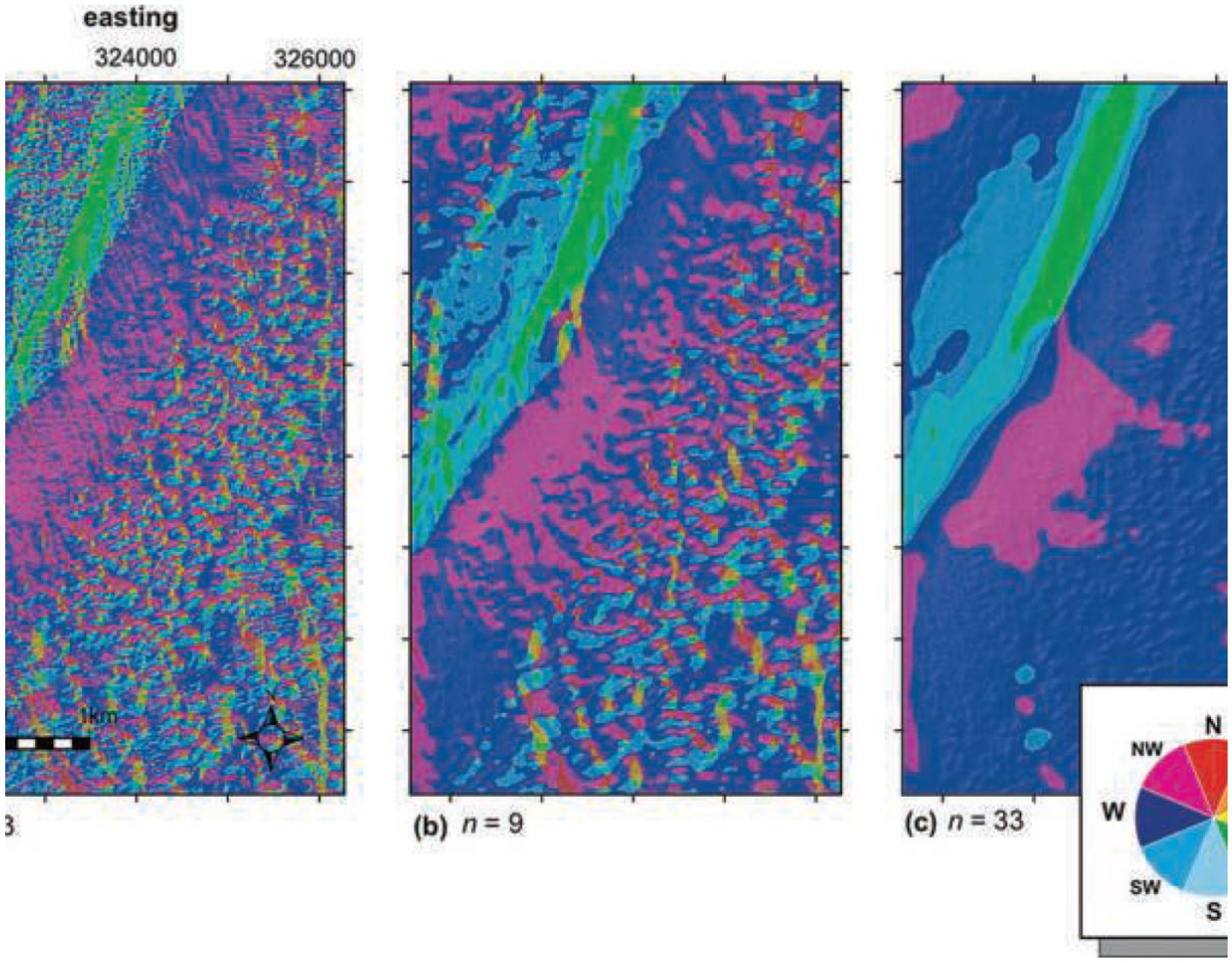


High-res topography from LIDAR



Terrain: Elevation, Slope, Aspect

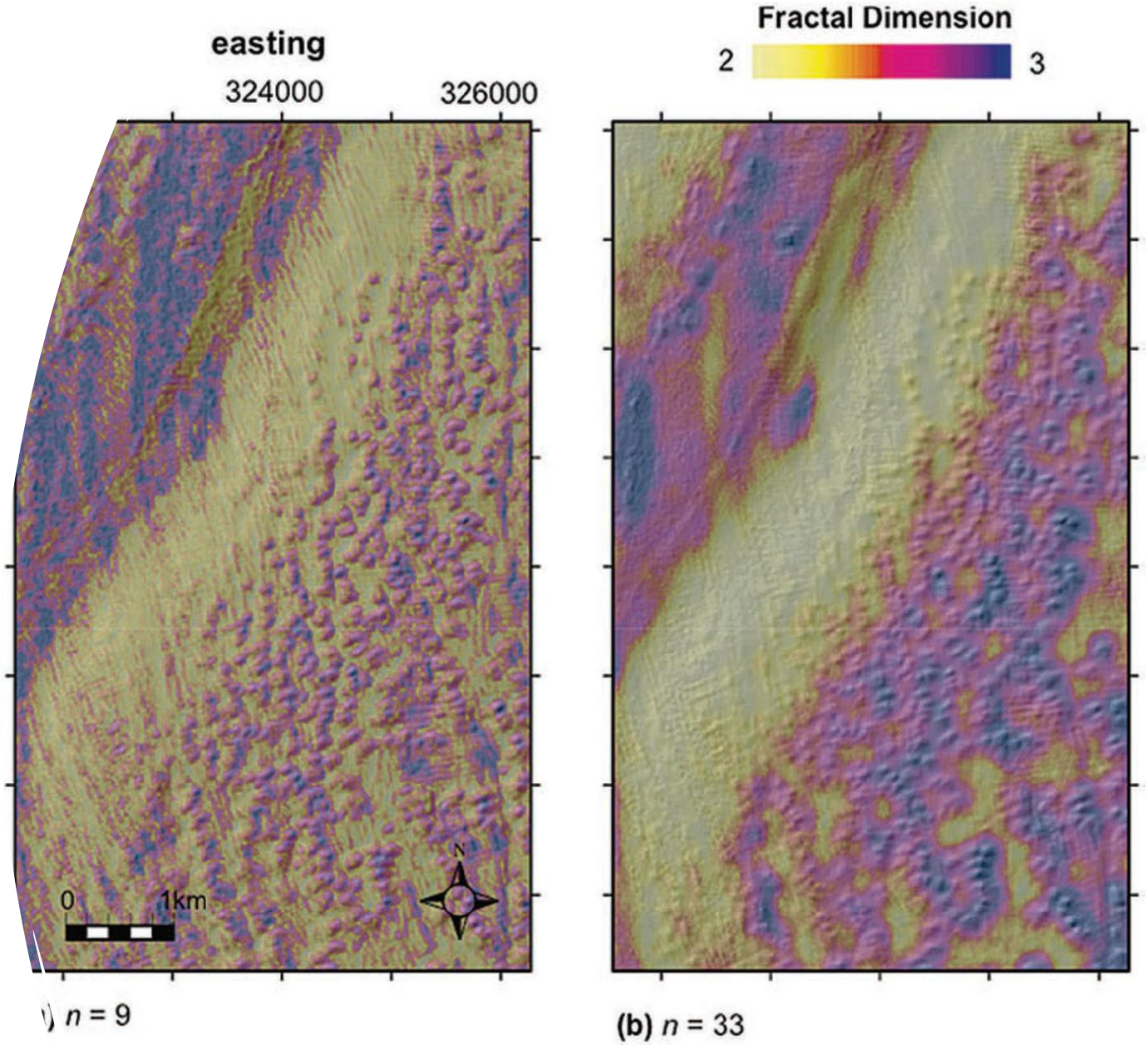
- Elevation: that's easy
- Slope: rise/run
- Aspect: What direction does the slope face?



Wilson et al (2007) Multiscale Terrain Analysis of Multibeam Bathymetry Data for Habitat Mapping on the Continental Slope

Roughness

- Roughness Indices
- Fractal Dimension



Fractal dimension calculated at (a) 9 × 9 analysis window (b) 33 ×33 analysis window.

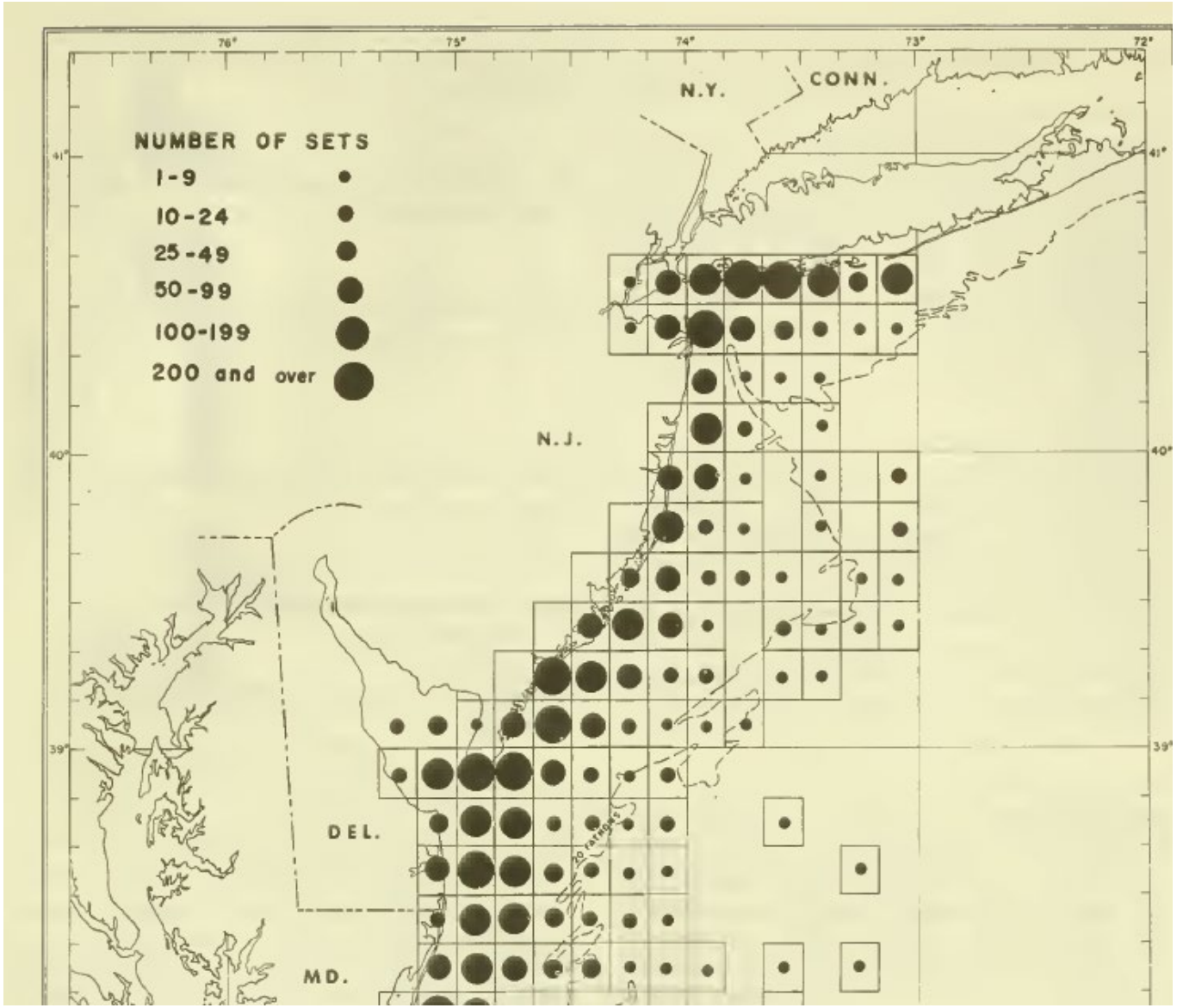
Wilson et al (2007)

Suitability analyses require DATA

DATA	SOURCE
Land cover and land use maps	Satellite image classification, aerial photo interpretation
Climate	Weather stations, interpolation
Census map	Mailed surveys
Topographic maps	Surveys, RADAR
LULC Change maps	Digitized paper maps, aerial photo interpretation
Occurrence/abundance	Surveys, collars

Scanning & interpreting historic aerial photos

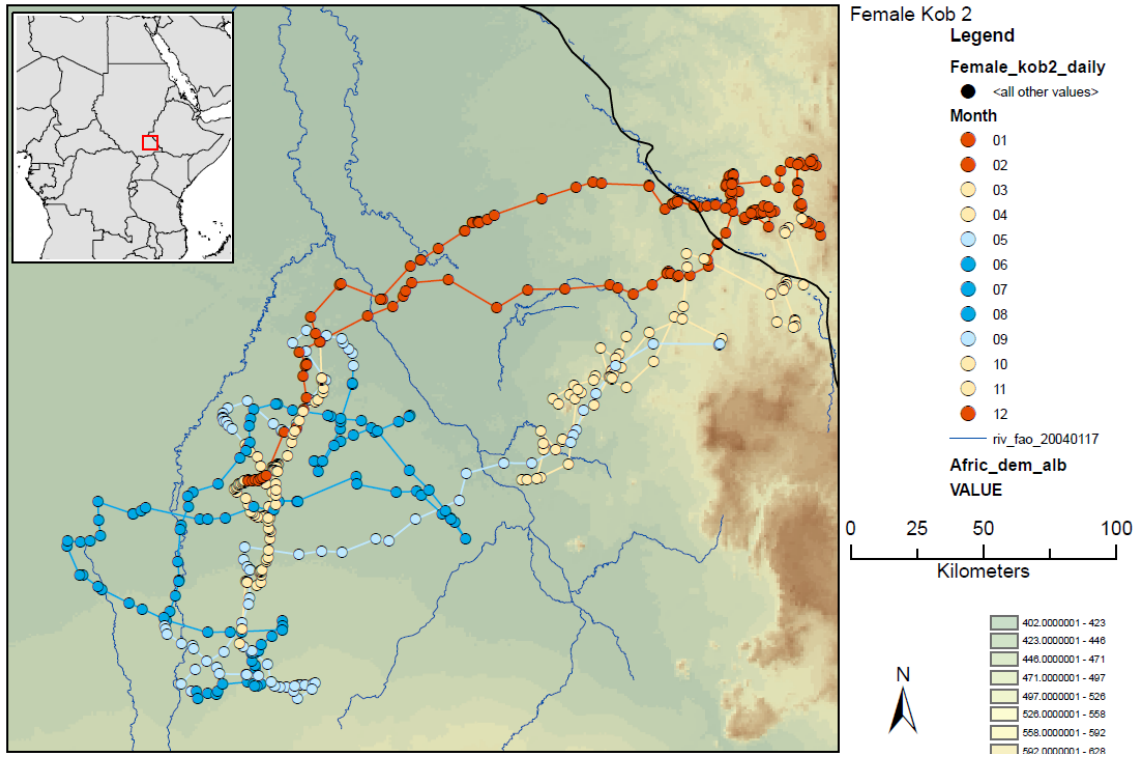
Digitizing historic maps



Suitability analyses require DATA

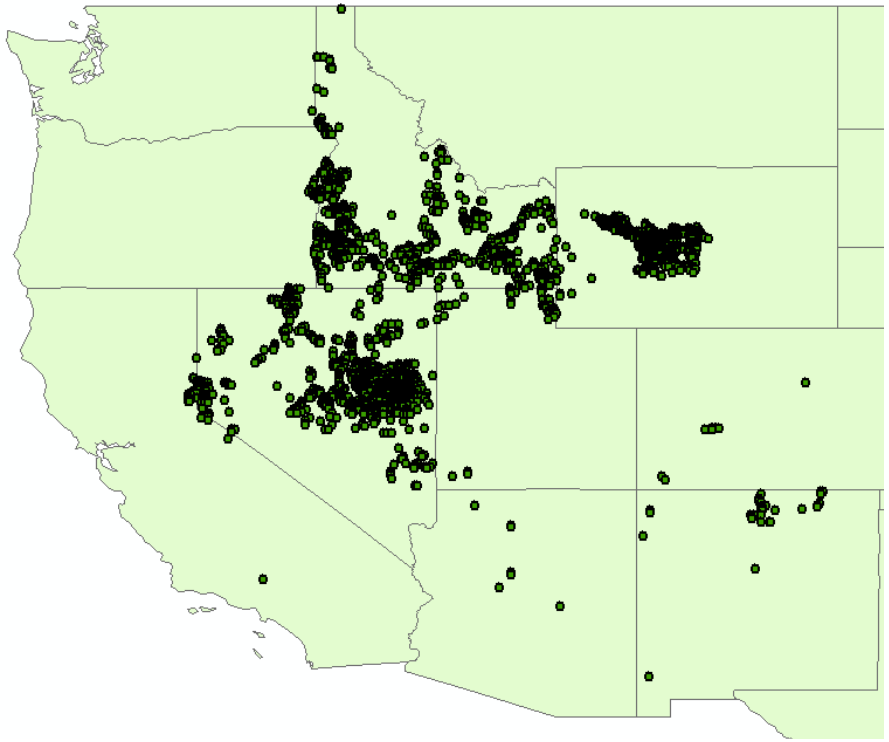
DATA	SOURCE
Land cover and land use maps	Satellite image classification, aerial photo interpretation
Climate	Weather stations, interpolation
Census map	Mailed surveys
Topographic maps	Surveys, RADAR
LULC Change maps	Digitized paper maps, aerial photo interpretation
Occurrence/abundance	Surveys, collars

GPS collar data – migration routes



Herbarium records

Spatial display of XY Coordinates

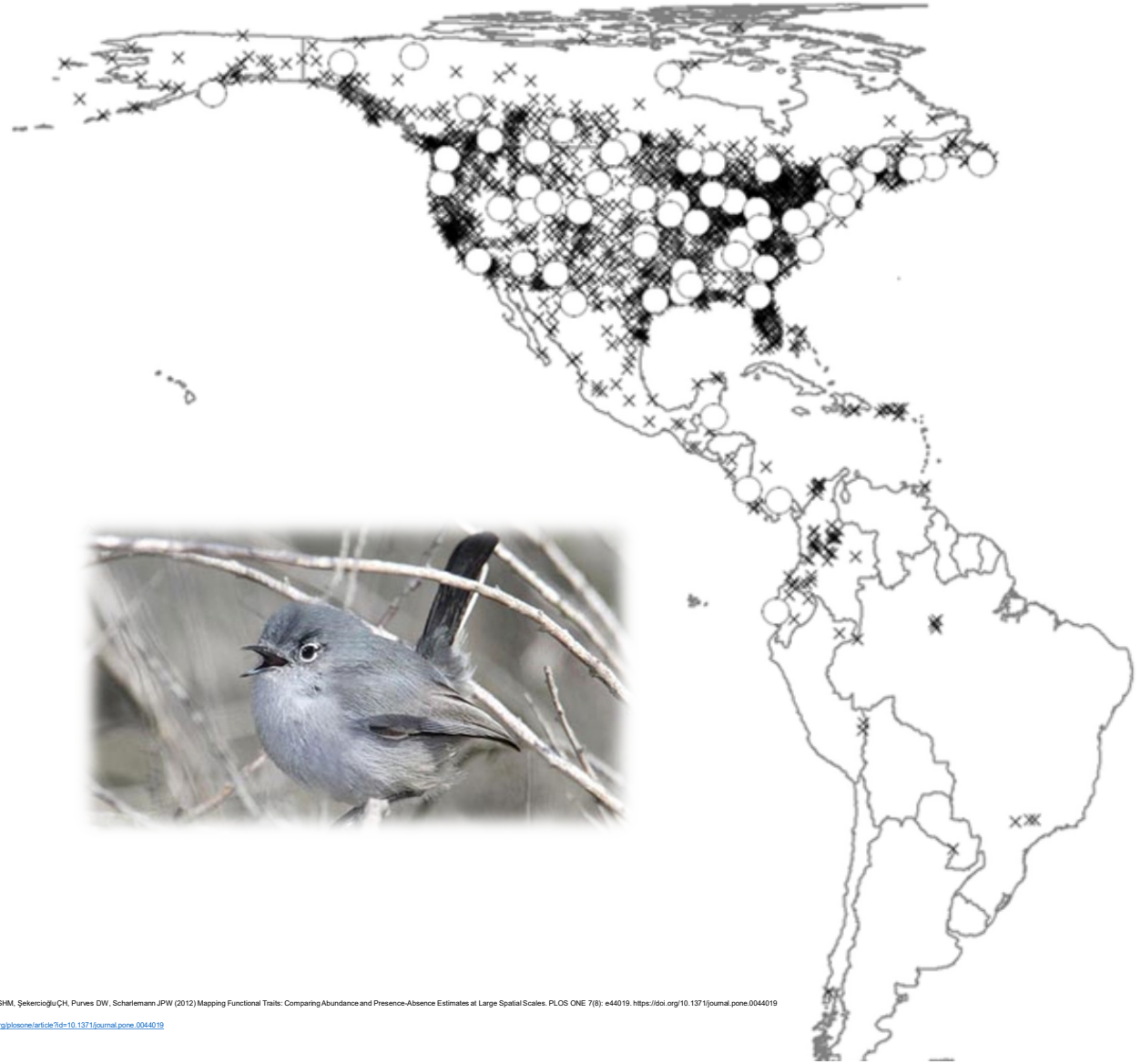


	A	B	C
1	ID	Latitude	Longitude
2	ciar_gbif	48.7502	-122.875
3	ciar_gbif	48.5835	-123.042
4	ciar_gbif	48.5835	-123
5	ciar_gbif	48.5835	-122.958
6	ciar_gbif	48.5419	-122.958
7	ciar_gbif	48.5419	-122.833
8	ciar_gbif	48.4585	-122.958
9	ciar_gbif	48.8335	-121.917
10	ciar_gbif	48.7918	-116.458
11	ciar_gbif	48.7918	-105.542
12	ciar_gbif	48.7918	-103.5
13	ciar_gbif	48.7918	-102.5
14	ciar_gbif	48.7918	-100.833
15	ciar_gbif	48.7502	-98.4583
16	ciar_gbif	48.7502	-97.5416
17	ciar_gbif	48.7085	-116.292
18	ciar_gbif	48.7085	-113
19	ciar_gbif	48.7085	-104.5

Presence, Absence, and Abundance

Beware sampling bias: Why are there so many rare plants near major universities?

Absence vs. pseudoabsence



Newbold T, Butchart SHM, Şekercioğlu ÇH, Purves DW, Scharlemann JPW (2012) Mapping Functional Traits: Comparing Abundance and Presence-Absence Estimates at Large Spatial Scales. PLOS ONE 7(8): e44019. https://doi.org/10.1371/journal.pone.0044019

https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0044019

Data Take-Homes

Data Quality and Metadata

- Look for the metadata in pre-existing datasets.
- Create useful metadata for any data you collect.
- Descriptive folder and file names are a type of metadata.
- Anticipate the types of errors that your data might contain.
- Do a reality check: Does your polygon of Rhode Island show up off the west coast of Africa?

Beware Human Error

Do categories make sense in the context of your project?

Always collect data with your goals in mind.

Take home messages:

Data come from *many* different sources

- Can’t find the data you need? Create your own!

All data are subject to error