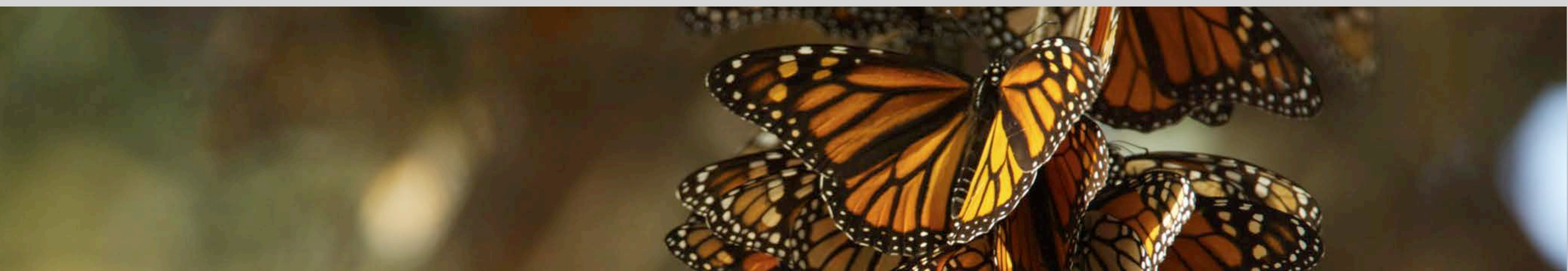
# Western milkweed and monarch breeding habitat suitability modeling review

Joe Engler, Liz Cruz, & Madeline Steele,

February 18th, 2016, 8:00 am Pacific

U.S. FISH & WILDLIFE SERVICE

THE XERCES SOCIETY FOR INVERTEBRATE CONSERVATION

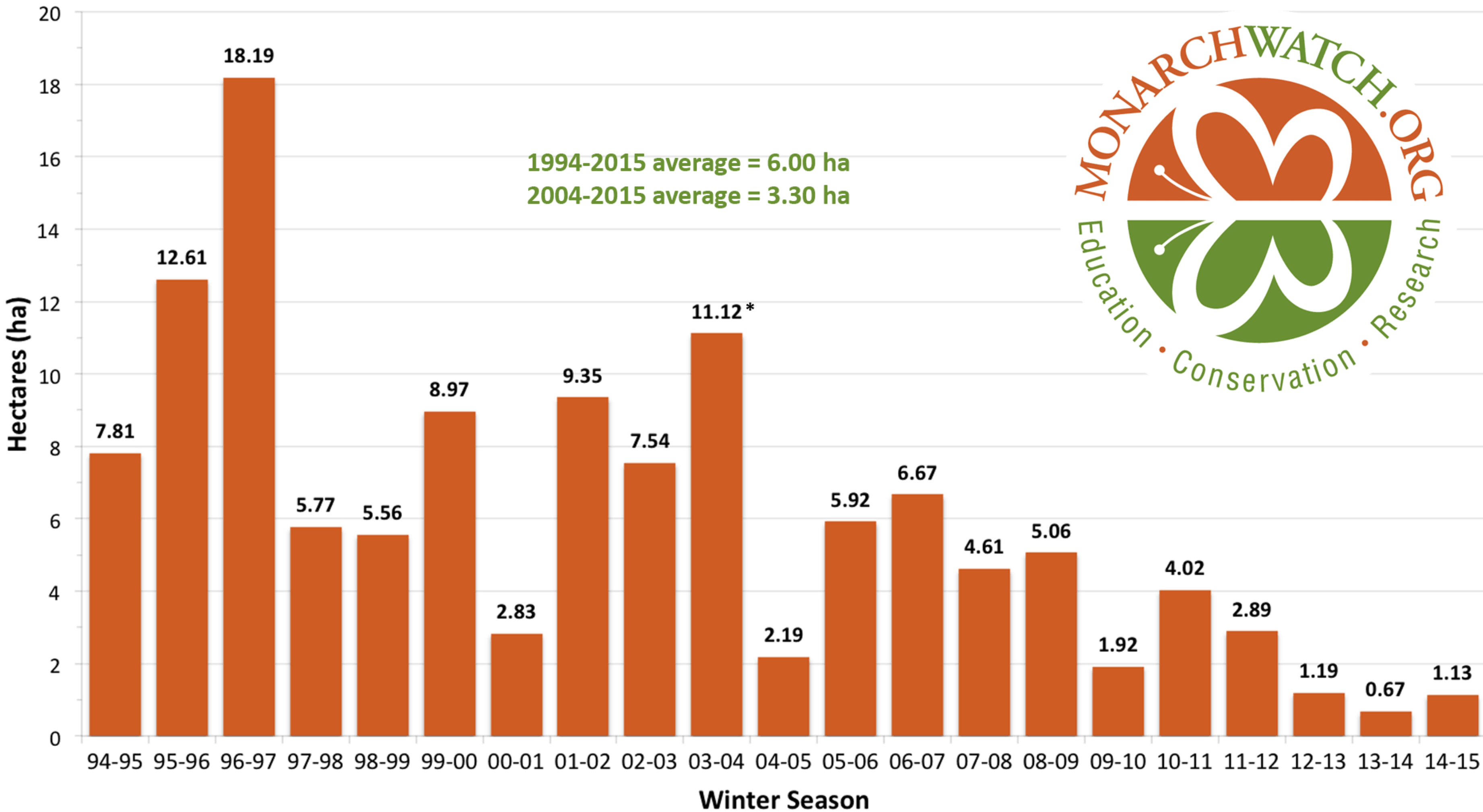


# Meeting outline

- Motivation for project
- Milkweed and monarch database development
- Modeling methods
  - MaxEnt
  - Mitigating for sampling bias
  - Environmental variable selection
  - Defining the background
  - Model tuning
- Results review
- Discussion

Photos by Elissa Buttermore, USFWS 2015

# Total Area Occupied by Monarch Colonies at Overwintering Sites in Mexico



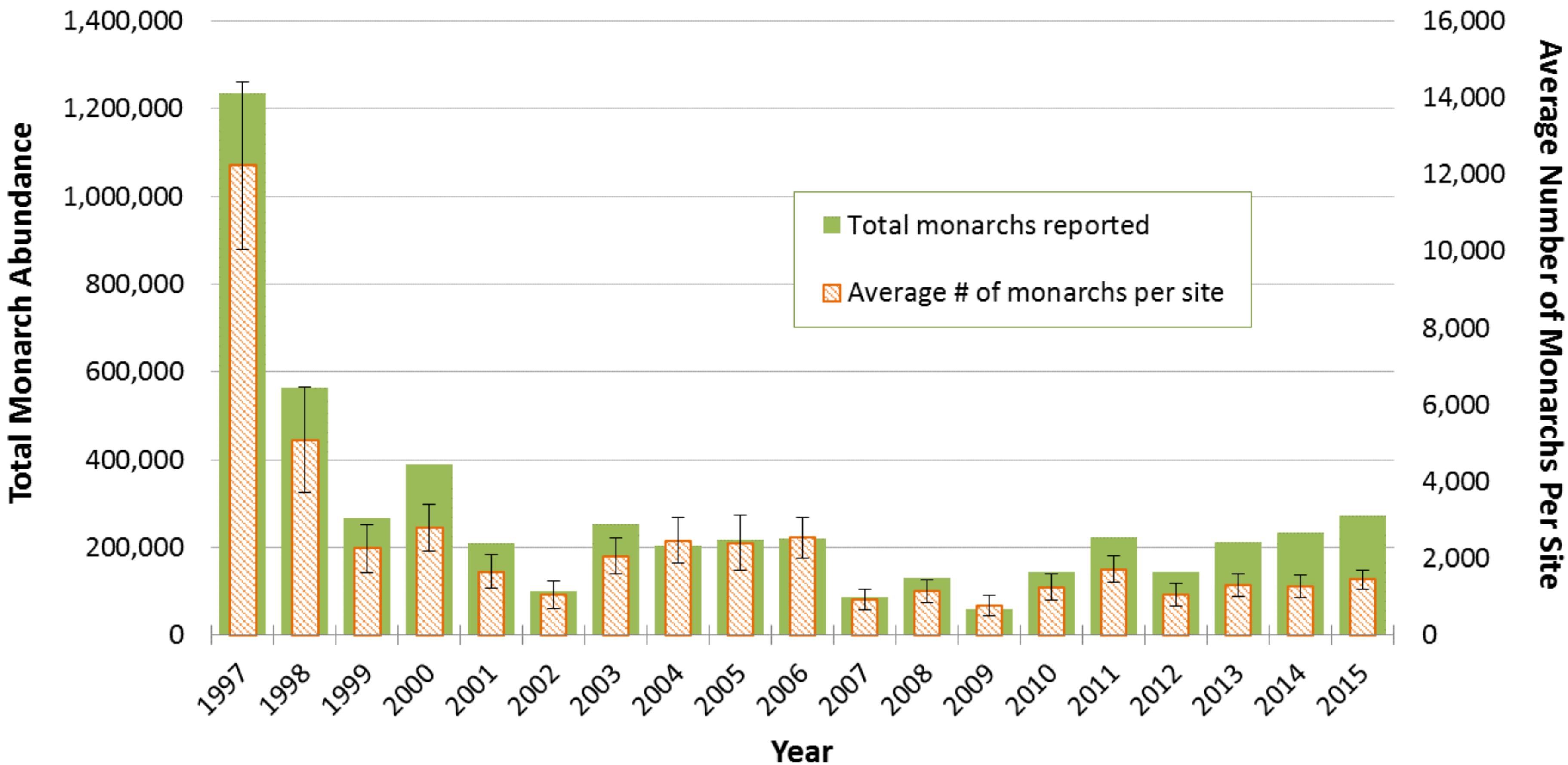

Data for 1994-2003 collected by personnel of the Monarch Butterfly Biosphere Reserve (MBBR) of the National Commission of Natural Protected Areas (CONANP) in Mexico. Data for 2003-2014 collected by World Wildlife Fund Mexico in coordination with the Directorate of the MBBR.

* Represents colony sizes measured in November of 2003 before the colonies consolidated. Measures obtained in January 2004 indicated the population was much smaller, possibly 8-9 hectares. CT

# Western Monarch Thanksgiving Count

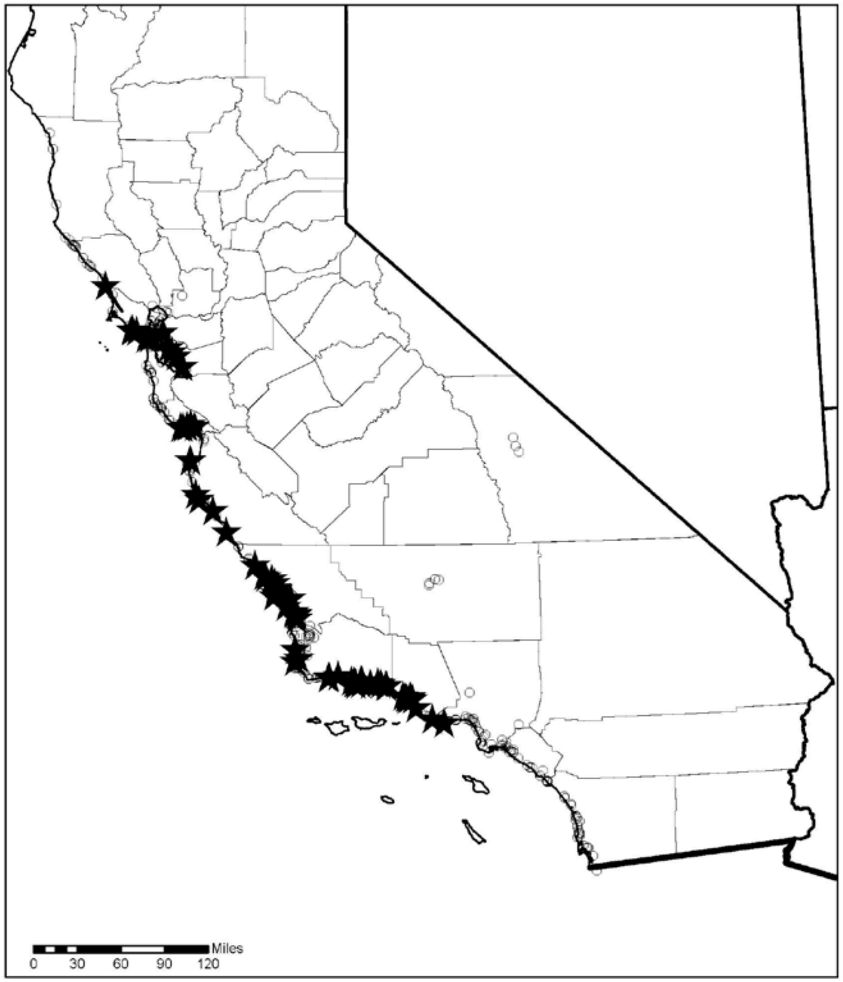
Total and Average Abundance Estimates w/ Standard Error of the Means
at 76-187 Overwintering Sites from 1997-2015
(Monroe *et al.* 2016)





# Western Monarch Overwintering Habitat

In the fall and winter, monarchs are found clustering in hundreds of coastal California groves



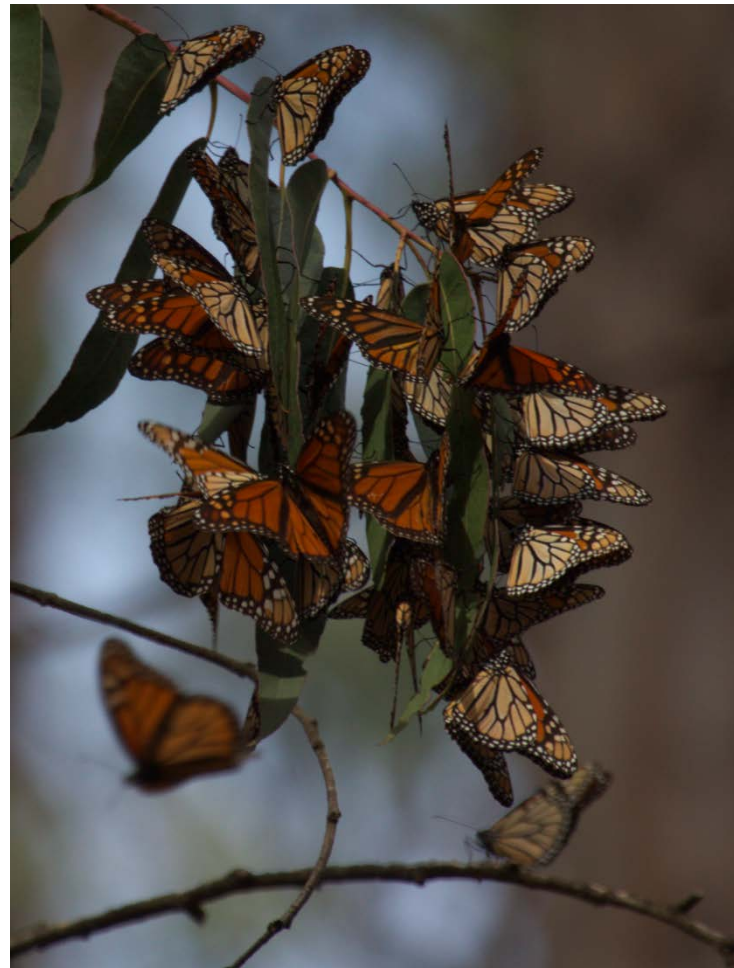
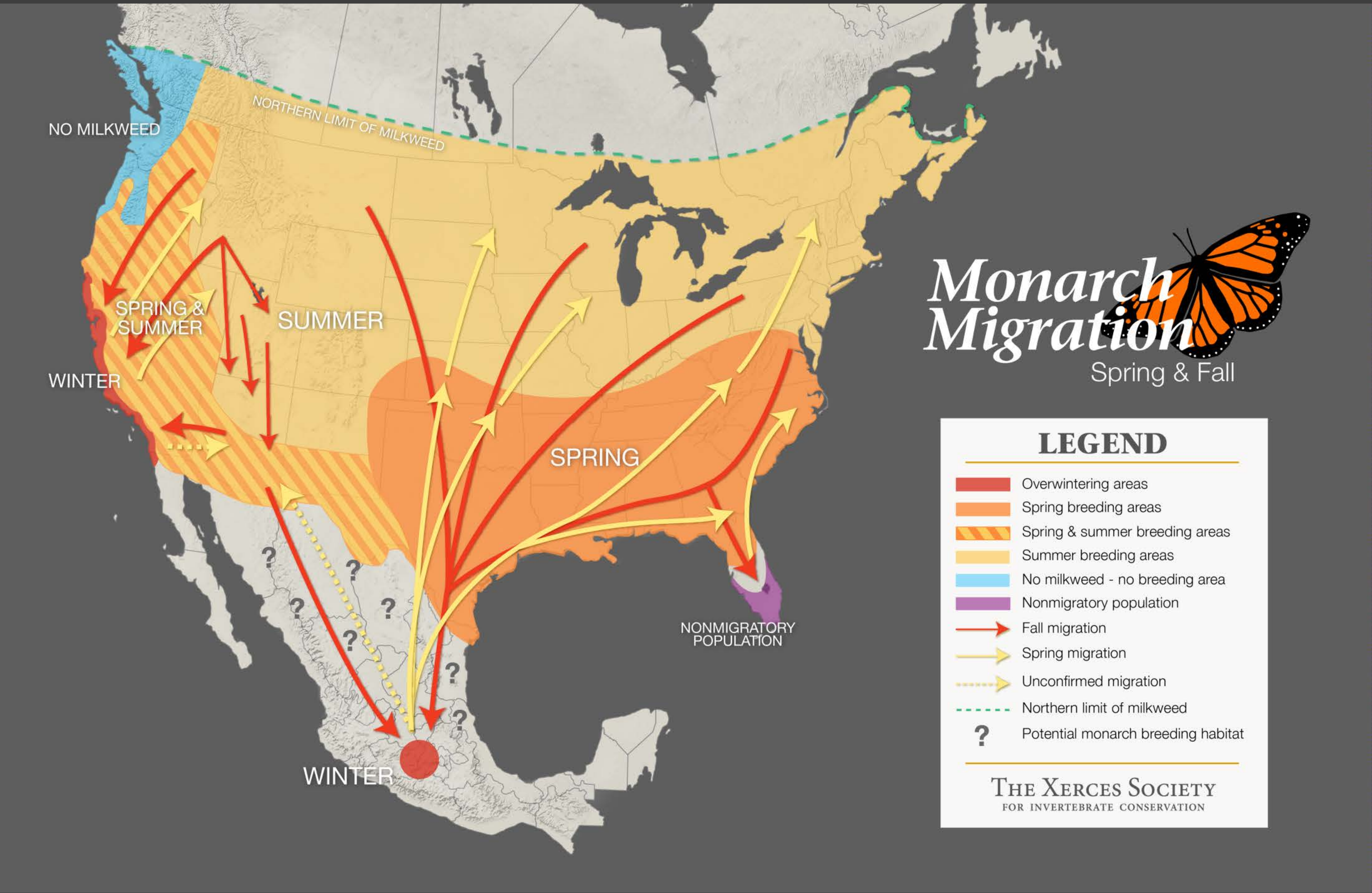



Photo: The Xerces Society/Carly Voight


Monarch Migration
Spring & Fall
NO MILKWEED
NORTHERN LIMIT OF MILKWEED
SPRING & SUMMER
SUMMER
WINTER
SPRING
NONMIGRATORY POPULATION
WINTER
LEGEND
Overwintering areas
Spring breeding areas
Spring & summer breeding areas
Summer breeding areas
No milkweed - no breeding area
Nonmigratory population
Fall migration
Spring migration
Unconfirmed migration
Northern limit of milkweed
? Potential monarch breeding habitat
The Xerces Society
FOR INVERTEBRATE CONSERVATION

# Dingle et al. 2005

Distribution of western monarchs throughout the year

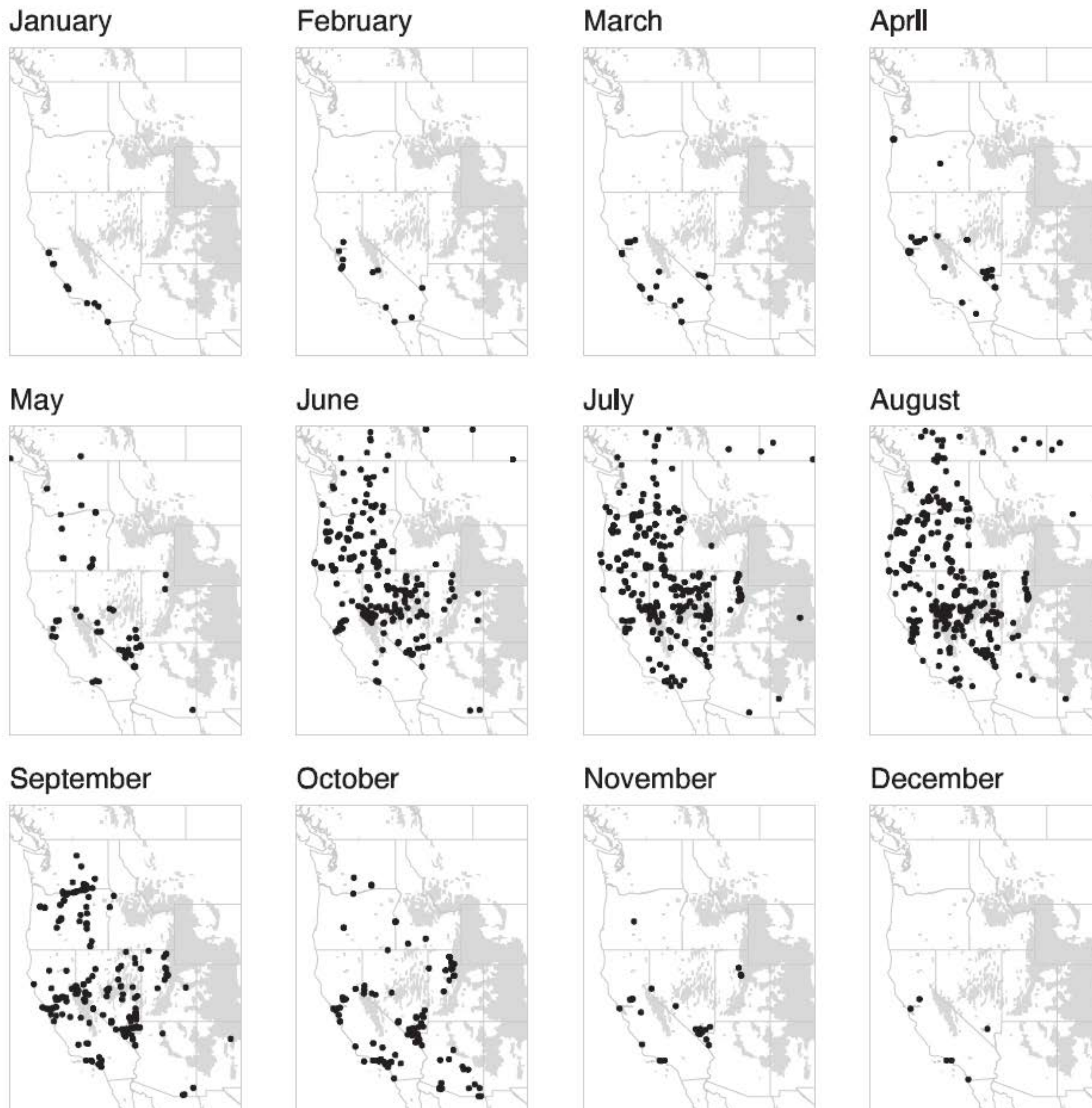
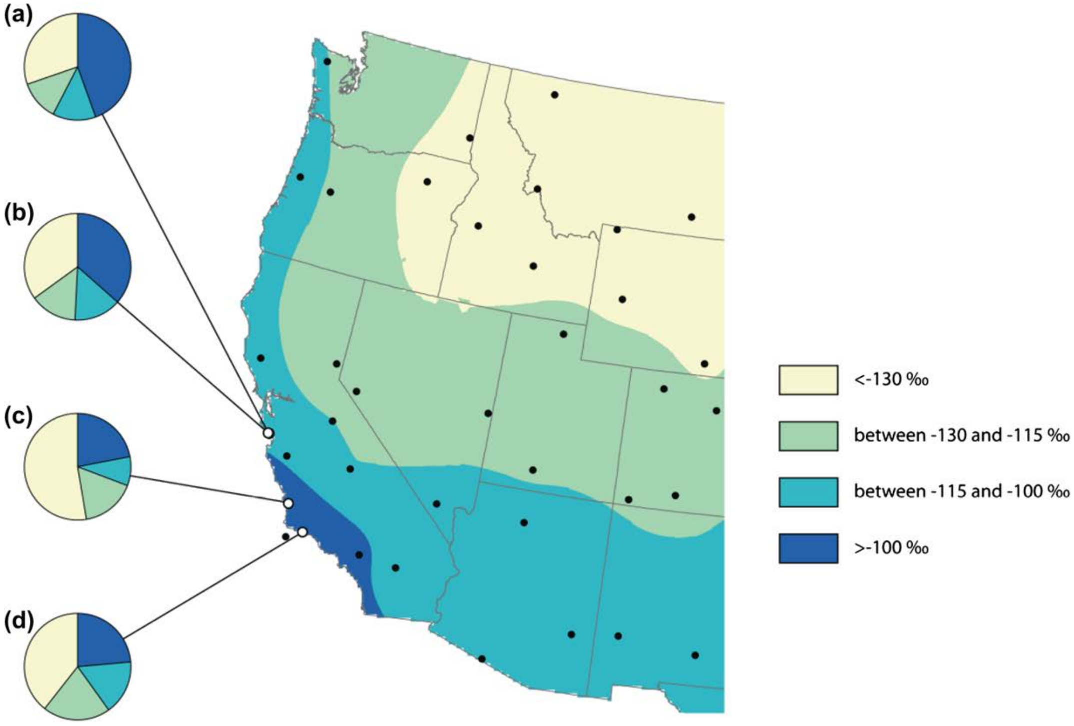



**Figure 1.** Collection records for monarch butterflies in western North America. Grey shaded areas are above 2000 m elevation.

# Monarch Natal Habitat

We understand general regional patterns in monarch natal habitat suitability, but higher resolution maps would be very useful




Yang et al. 2015

Stephens and Frey 2010

# The need for a habitat suitability model

- Identification of priority monarch areas needed for efficient use of limited funds
- We know that geographic features, climate variability, and milkweed availability influence monarch habitat distribution
- Species distribution models can quantify these relationships and help us prioritize landscapes

# Project objectives

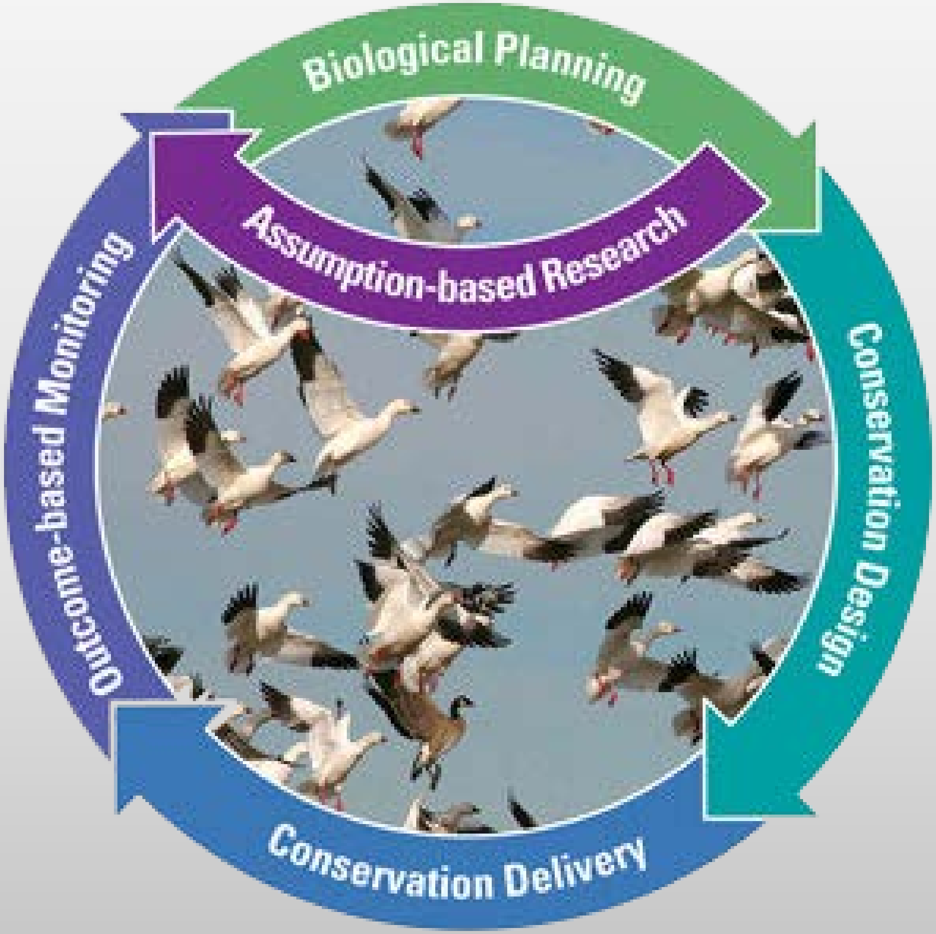
- Consolidate existing western milkweed and monarch records
- Collect new data for milkweed and monarchs (2015-2020)
- Compile geospatial data layers that may influence milkweed & monarch distributions

# Objectives, continued

- Produce coarse-scale models that predict important regions for key milkweed species and for monarchs themselves.

- Use models to help prioritize landscapes and to guide surveys that can refine next year's models

- **An iterative process**



# Primary Objectives

Build a database to house the large, and growing, collection of milkweed and monarch breeding observation records held by The Xerces Society.

Collect and incorporate USFWS milkweed field surveys and western milkweed and monarch breeding records from a variety of other sources.

Convert data to spatial format to be used in the western milkweed habitat suitability model.

# How did we get the data?

- Request for data sent to researchers, herbariums, and other (FWS) regional biologists
- Obtained through agreements with Xerces/FWS
- Returned FWS field survey forms and other field surveys
- Downloads from online consortiums and herbariums
- ArcGIS online web mapping application
- Social Media

# Main Providers:

- The Xerces Society
- US Fish & Wildlife
- SEINet
- Bureau of Land Managment
- US Forest Service
- National Park Service
- Consortium of Pacific Northwest Herbaria
- Consortium of California Herberia
- iNaturalist
- National Phenology Network

- BISON (USGS)
- OSU Herbarium (Oregon Plant Atlas)
- Journey North
- Hugh Dingle (UC Davis)
- Monarch Larva Monitoring Project (MLMP)
- Flickr
- GBIF

# Original Source Data:

65 Universities

13 Government agencies

10 Botanical Gardens

7 Consortiums

6 Conservation Organizations

6 Museums

5 Research Organizations

6 Other

# Major Challenges:

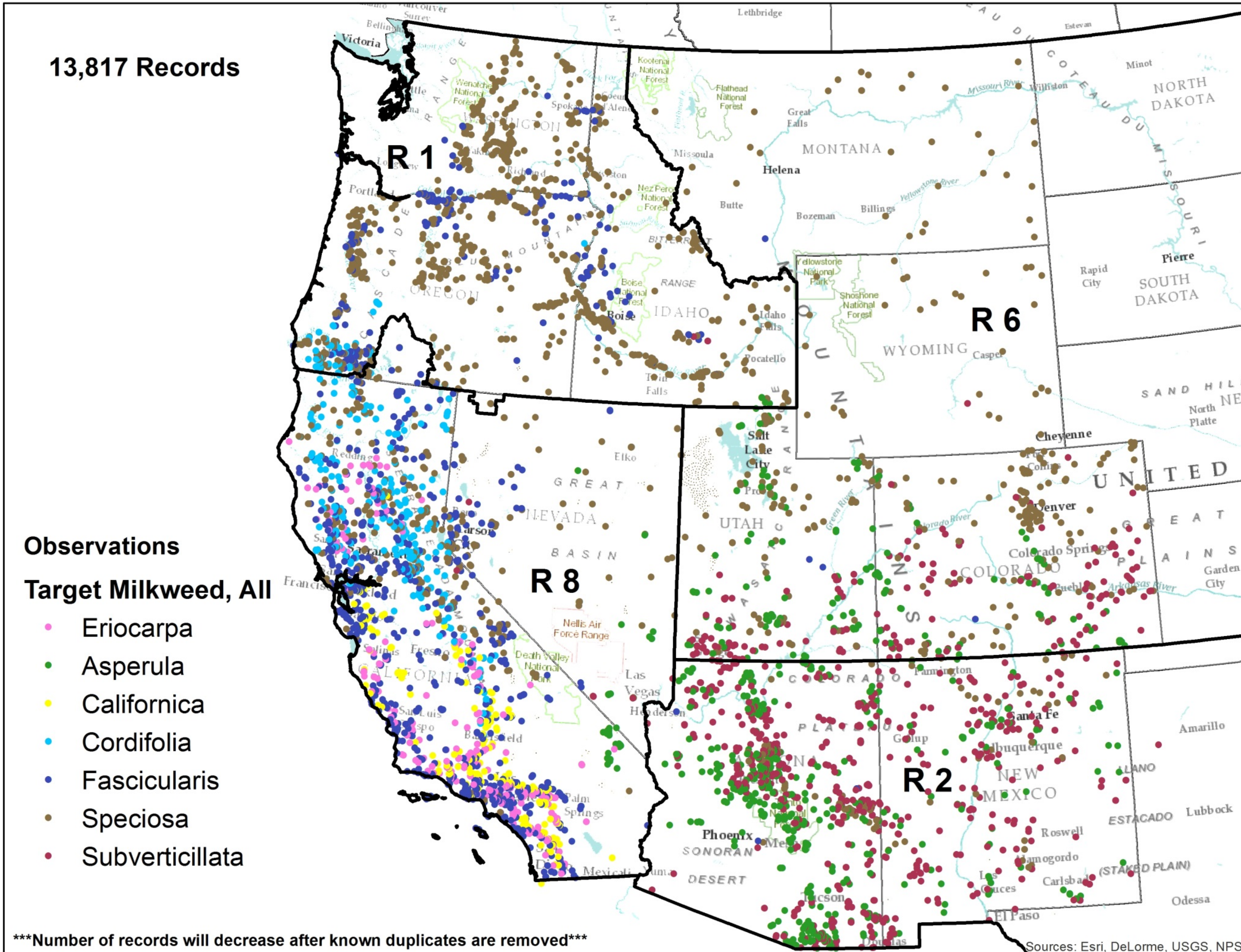
- 92 input files in dozens of unique formats
- Cross-walking to new database structure
- Interpreting data (no explanation of fields)
- Inconsistent data values
- Poor or unknown location accuracy
- Datasets lacking information we sought
- Duplicates (mostly overlaps between data sources)

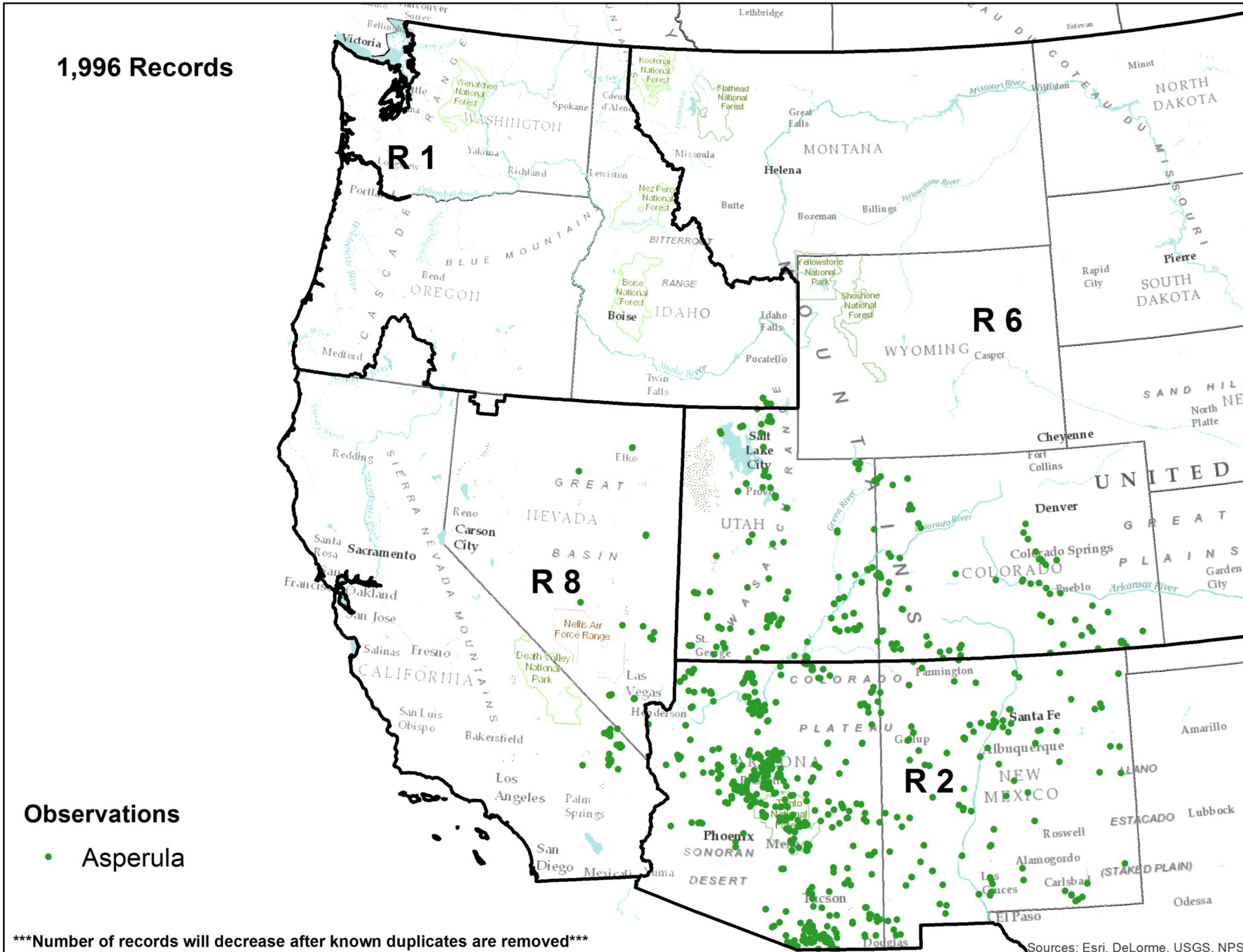
# The Numbers

Began with a single spreadsheet of ~8,000 records shared by Xerces and ended with roughly…

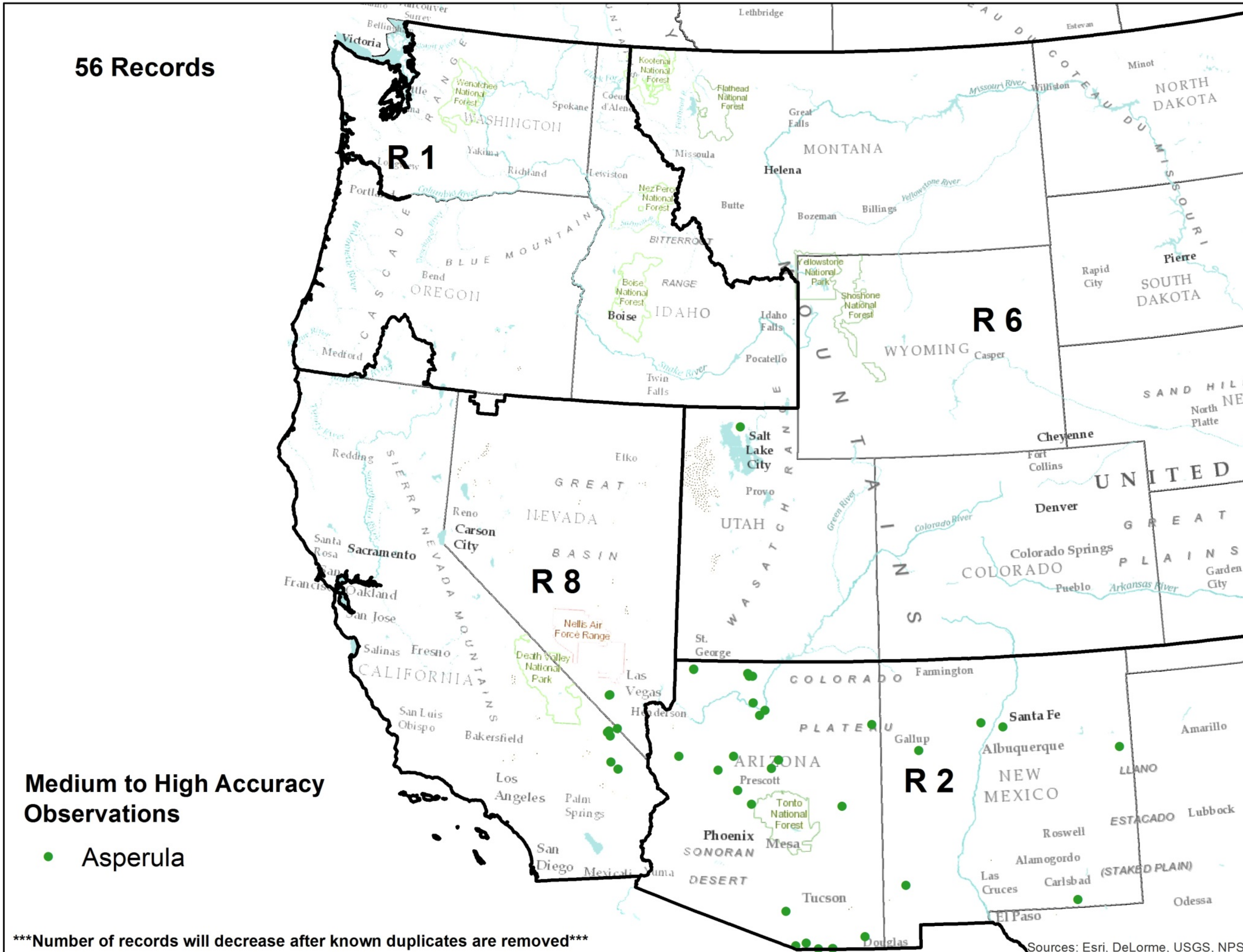
- 28,000 total records (still many known duplicates to contend with)
- 60 data providers
- 90 input files
- 61% of records are for our target milkweed species and 26% of those records have a high enough accuracy level to be used in model
- 19% of records for monarch, 24% of those have good accuracy
- Only 12% of the total “good” monarch records indicate breeding activity

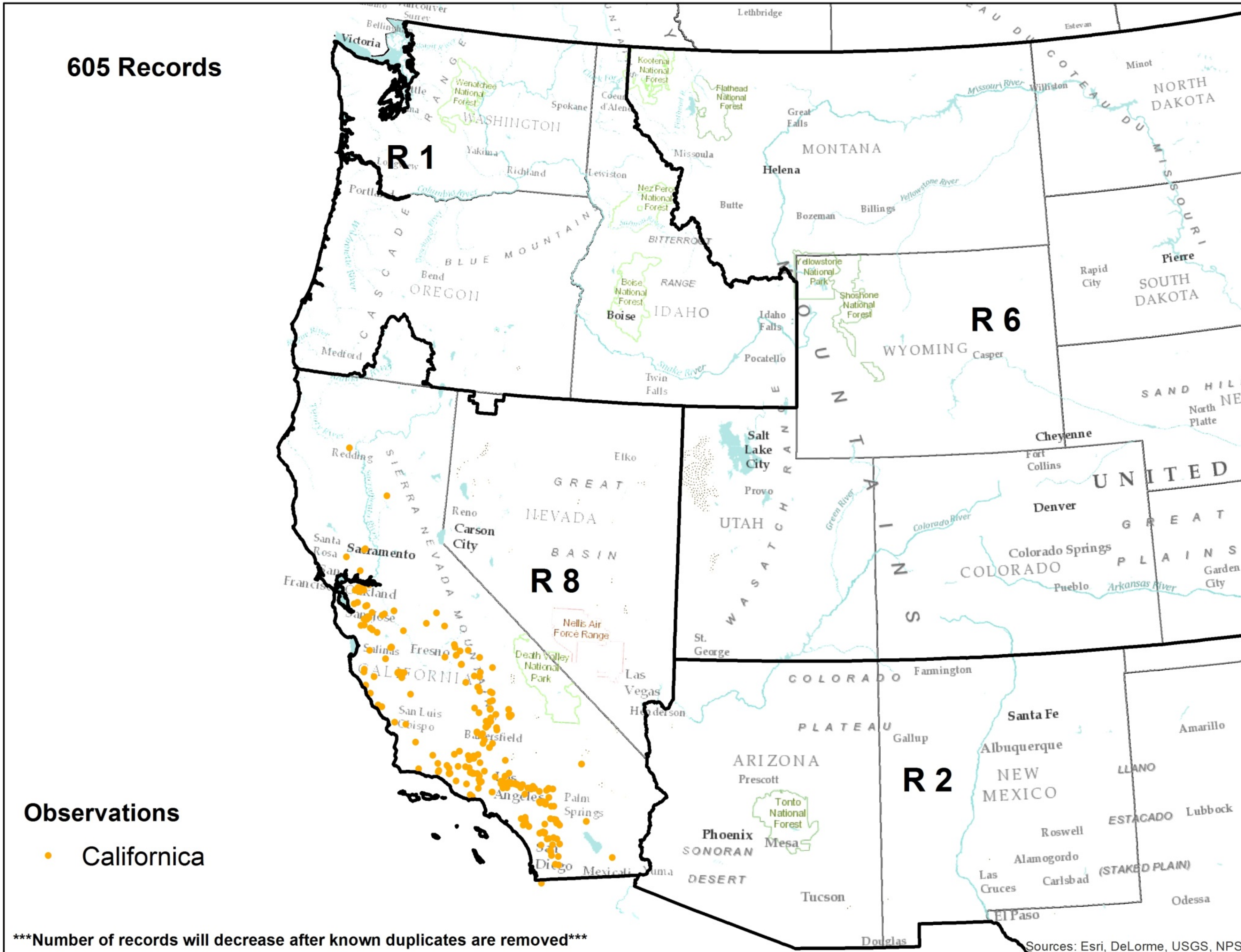
| | | | |
|---|---|---|---|
| ProviderName | SourceFile | LocAccuracyCode | WindSpeed |
| RecordID | RecordDate_vbtm | LocationRemarks | Precipitation |
| ProviderContact | RecordDate | LocReview | CloudCover_pct |
| ProviderNotes | RecordMonth | CoordinateUncertainty_m | Phenology |
| DatasetDate | RecordDay | GeoreferencedBy | PlantStructure |
| AddedBy | RecordYear | Habitat | PatchSize_m2 |
| AddedDate | RecordedBy | IsUrban_YN | AvgPlantHeight_in |
| CompilationNotes | RecordedBy_Lastname | VegRemarks | PlantCount |
| DataRemarks | RecordedBy_Firstname | Breeding_YN | StemCount |
| Share_YN | RecordedBy_Contact | Lifestage | MaturePlantCount |
| ConsortiumDatabase | Recorder_Affiliation | Count | ImmaturePlantCount |
| ConsortiumUniqueID | Locality | AdultCount | FloweringPlantCount |
| OriginalSource | NearestCity | LarvaeCount | PlantCount_ImmaturePods |
| OriginalSourceID | State | PupaeCount | PlantCount_MaturePods |
| CatalogNum | County | EggCount | PodCollectionCount |
| CollectionNum | LandMgr | ObservationTime | EstimationMethod |
| BasisofRecord | LandMgrType | MonarchPopRemarks | PlantPopRemarks |
| SurveyType | Latitude | BehaviorNotes | OtherRemarks |
| Genus | Longitude | NectarFamilyUsed | Literature_YN |
| Species | VerbatimElevation | NectarSpeciesUsed | |
| ScientificName | DataProjection | NectarFamilyObserved | |
| VernacularName | Datum | NectarSpeciesObserved | |
| | GeoreferenceSource | Temperature_F | |

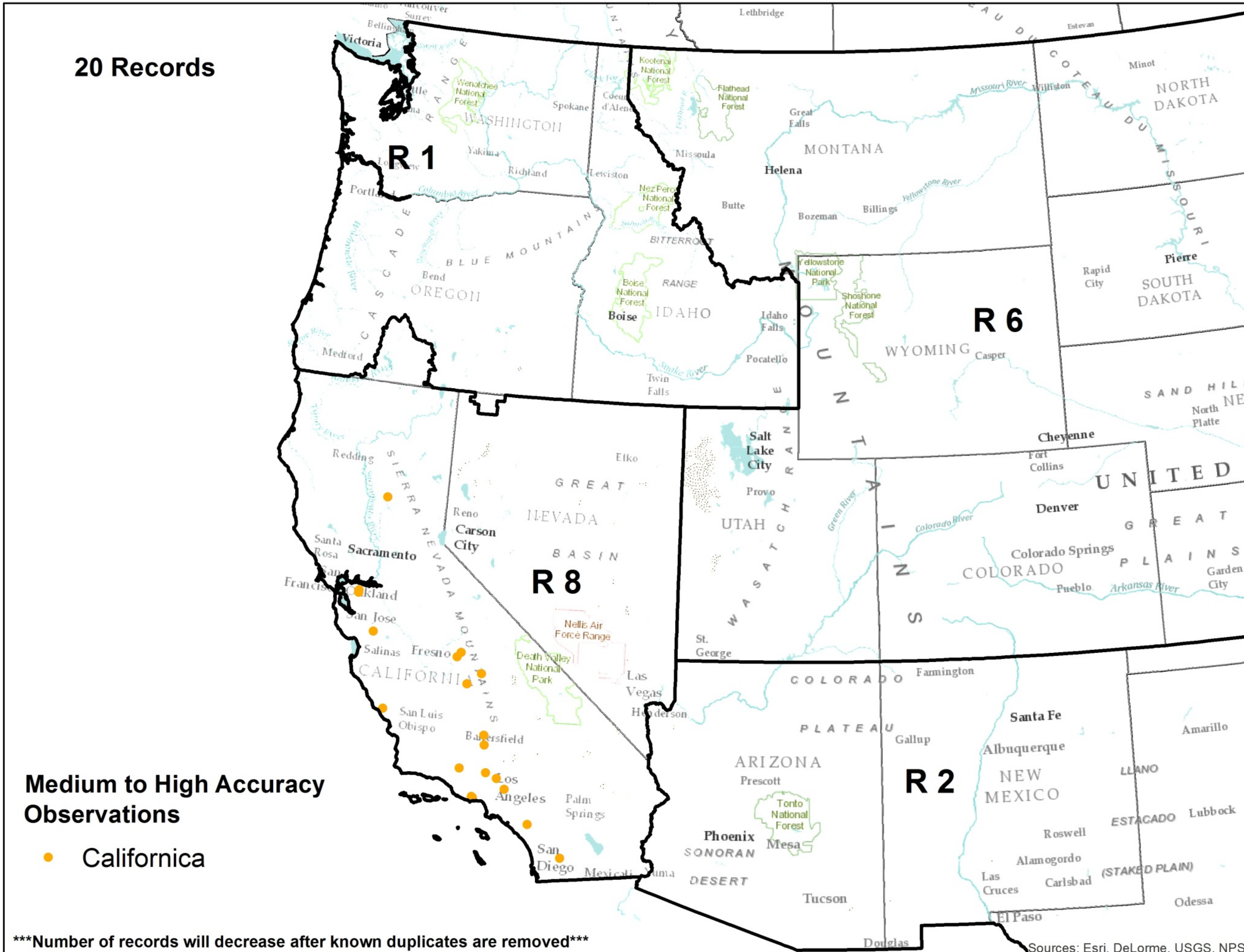

13,817 Records
Observations
Target Milkweed, All
Eriocarpa
Asperula
Californica
Cordifolia
Fascicularis
Speciosa
Subverticillata
R 1
R 6
R 8
R 2
WASHINGTON
OREGON
IDAHO
MONTANA
WYOMING
NORTH DAKOTA
SOUTH DAKOTA
NEVADA
UTAH
COLORADO
CALIFORNIA
ARIZONA
NEW MEXICO
***Number of records will decrease after known duplicates are removed***
Sources: Esri, DeLorme, USGS, NPS

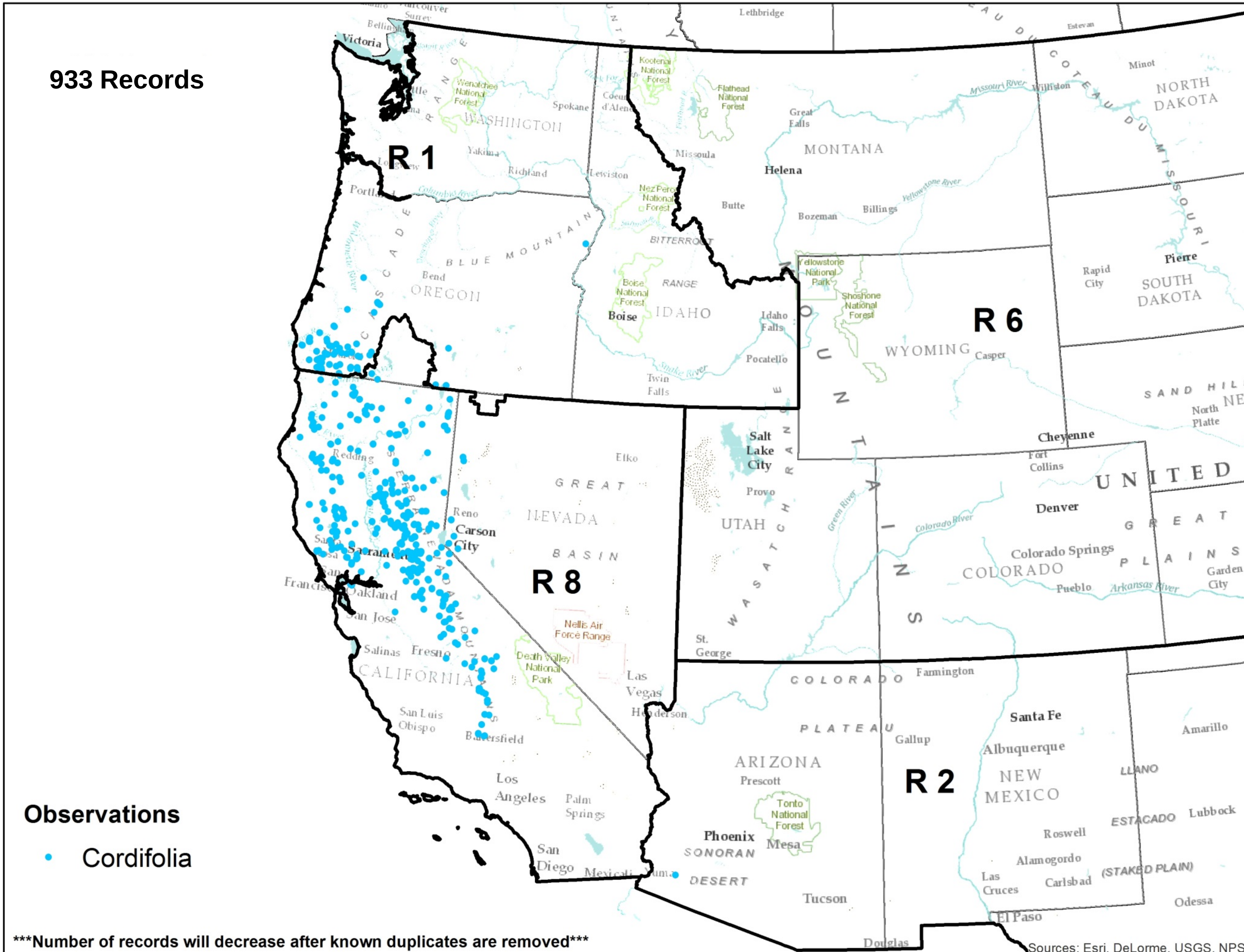


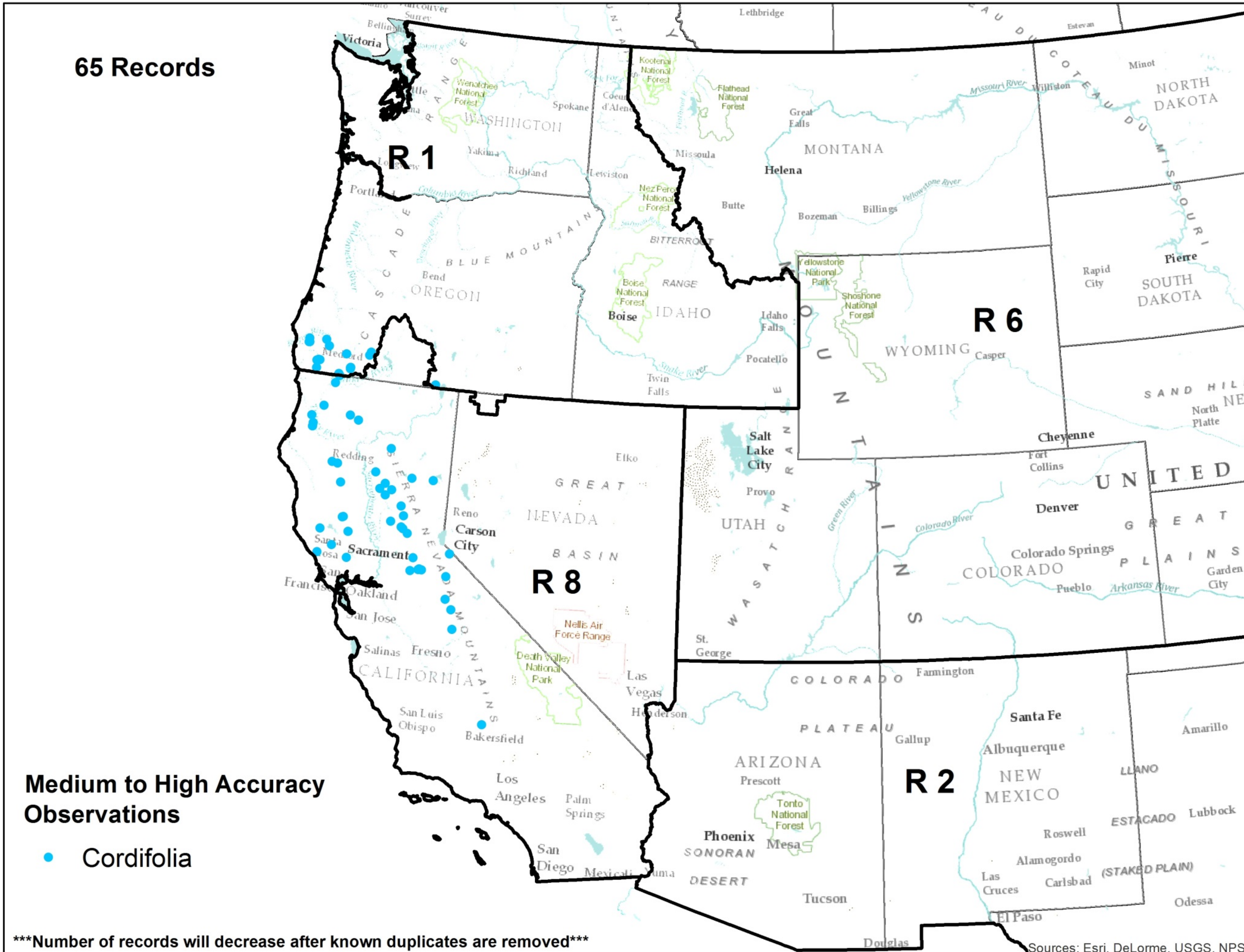
1,996 Records
R 1
R 6
R 8
R 2
Observations
Asperula
***Number of records will decrease after known duplicates are removed***
Sources: Esri, DeLorme, USGS, NPS
WASHINGTON
OREGON
IDAHO
MONTANA
WYOMING
NORTH DAKOTA
SOUTH DAKOTA
NEVADA
UTAH
COLORADO
CALIFORNIA
ARIZONA
NEW MEXICO
GREAT BASIN
COLORADO PLATEAU
SONORAN DESERT
GREAT PLAINS
SIERRA NEVADA MOUNTAINS
BLUE MOUNTAINS
CASCADE RANGE
BITTERROOT RANGE
ROCKY MOUNTAINS
WASATCH RANGE
SAND HILLS
LLANO ESTACADO (STAKED PLAIN)
UNITED

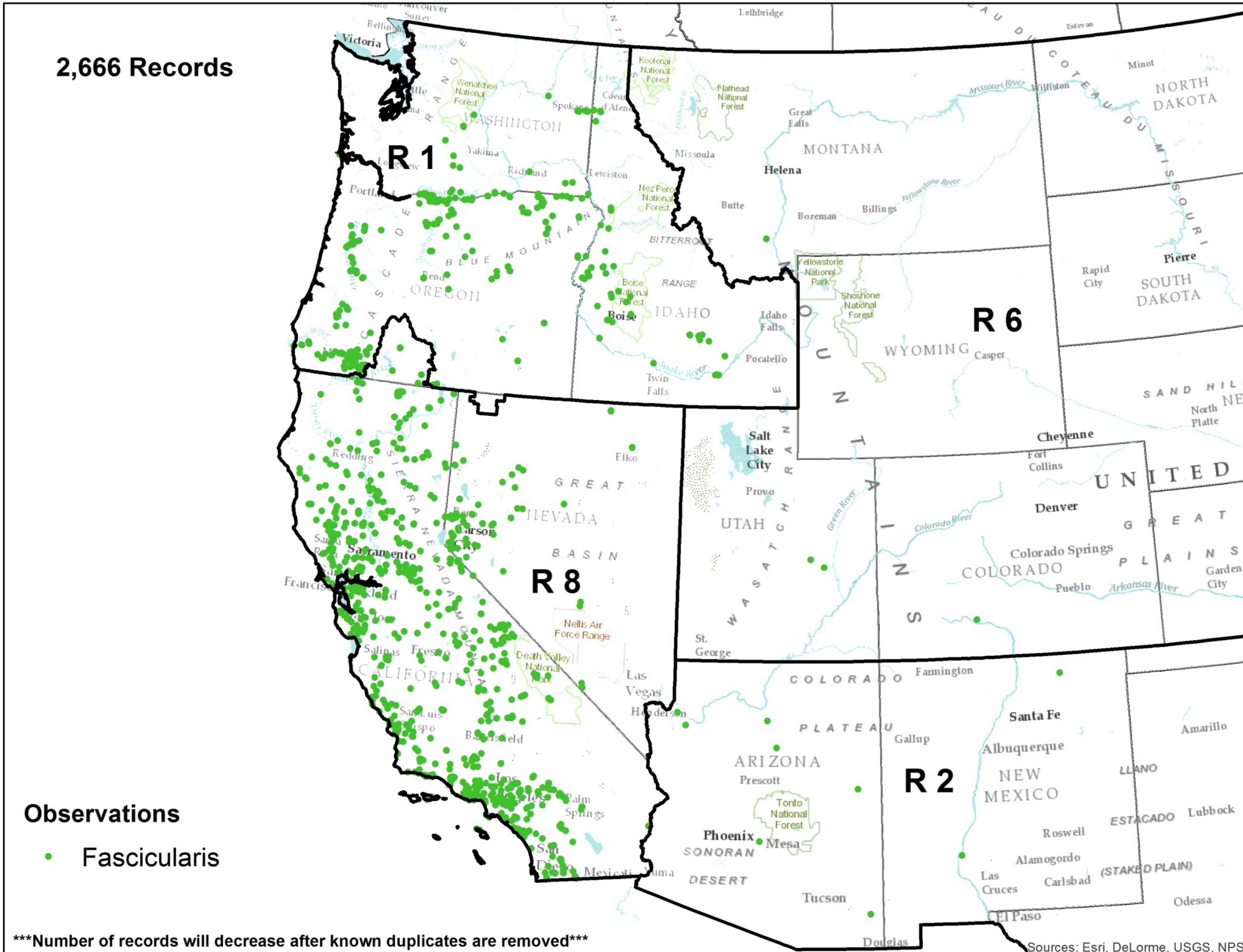

56 Records
R 1
R 6
R 8
R 2
Medium to High Accuracy Observations
Asperula
***Number of records will decrease after known duplicates are removed***
Sources: Esri, DeLorme, USGS, NPS
WASHINGTON
OREGON
IDAHO
MONTANA
WYOMING
NORTH DAKOTA
SOUTH DAKOTA
NEVADA
UTAH
COLORADO
CALIFORNIA
ARIZONA
NEW MEXICO
UNITED
Victoria
Lethbridge
Spokane
Yakima
Richland
Portland
Bend
Medford
Lewiston
Missoula
Helena
Great Falls
Butte
Bozeman
Billings
Williston
Minot
Pierre
Rapid City
Boise
Idaho Falls
Pocatello
Twin Falls
Casper
Cheyenne
Fort Collins
Denver
Colorado Springs
Pueblo
Garden City
North Platte
Salt Lake City
Provo
St. George
Elko
Reno
Carson City
Las Vegas
Henderson
Redding
Santa Rosa
Sacramento
San Francisco
Oakland
San Jose
Salinas
Fresno
San Luis Obispo
Bakersfield
Los Angeles
Palm Springs
San Diego
Farmington
Gallup
Santa Fe
Albuquerque
Prescott
Phoenix
Mesa
Tucson
Douglas
Roswell
Alamogordo
Las Cruces
Carlsbad
El Paso
Amarillo
Lubbock
Odessa
Kootenai National Forest
Flathead National Forest
Wenatchee National Forest
Nez Perce National Forest
Boise National Forest
Yellowstone National Park
Shoshone National Forest
Nellis Air Force Range
Death Valley National Park
Tonto National Forest
BLUE MOUNTAINS
CASCADE RANGE
BITTERROOT RANGE
ROCKY MOUNTAINS
WASATCH RANGE
GREAT BASIN
SIERRA NEVADA MOUNTAINS
COLORADO PLATEAU
SONORAN DESERT
GREAT PLAINS
LLANO ESTACADO
(STAKED PLAIN)
SAND HILLS
COTEAU DU MISSOURI
Missouri River
Yellowstone River
Columbia River
Snake River
Colorado River
Green River
Arkansas River

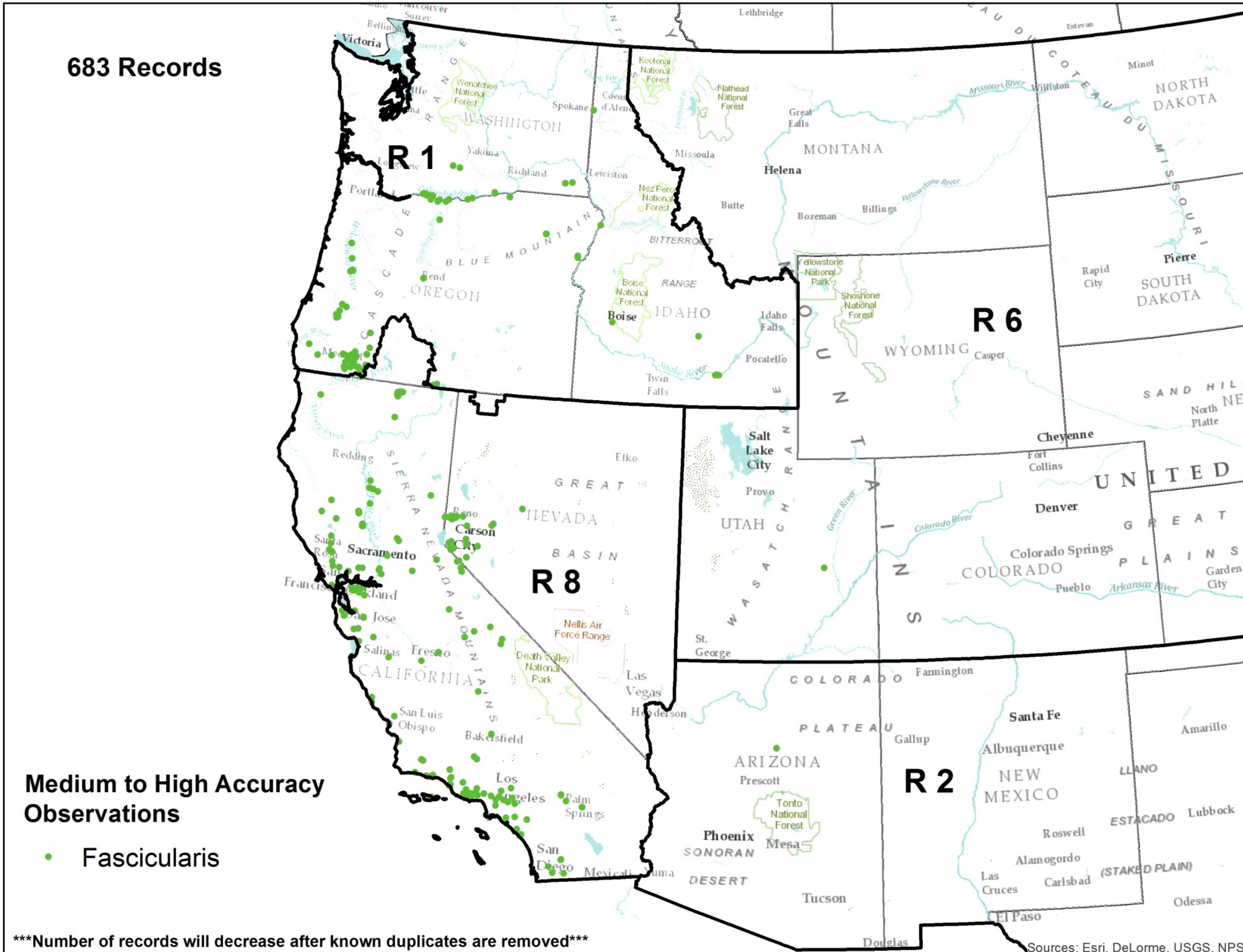

605 Records
R 1
R 6
R 8
R 2
WASHINGTON
OREGON
IDAHO
MONTANA
WYOMING
NORTH DAKOTA
SOUTH DAKOTA
NEVADA
UTAH
COLORADO
ARIZONA
NEW MEXICO
CALIFORNIA
Observations
Californica
***Number of records will decrease after known duplicates are removed***
Sources: Esri, DeLorme, USGS, NPS

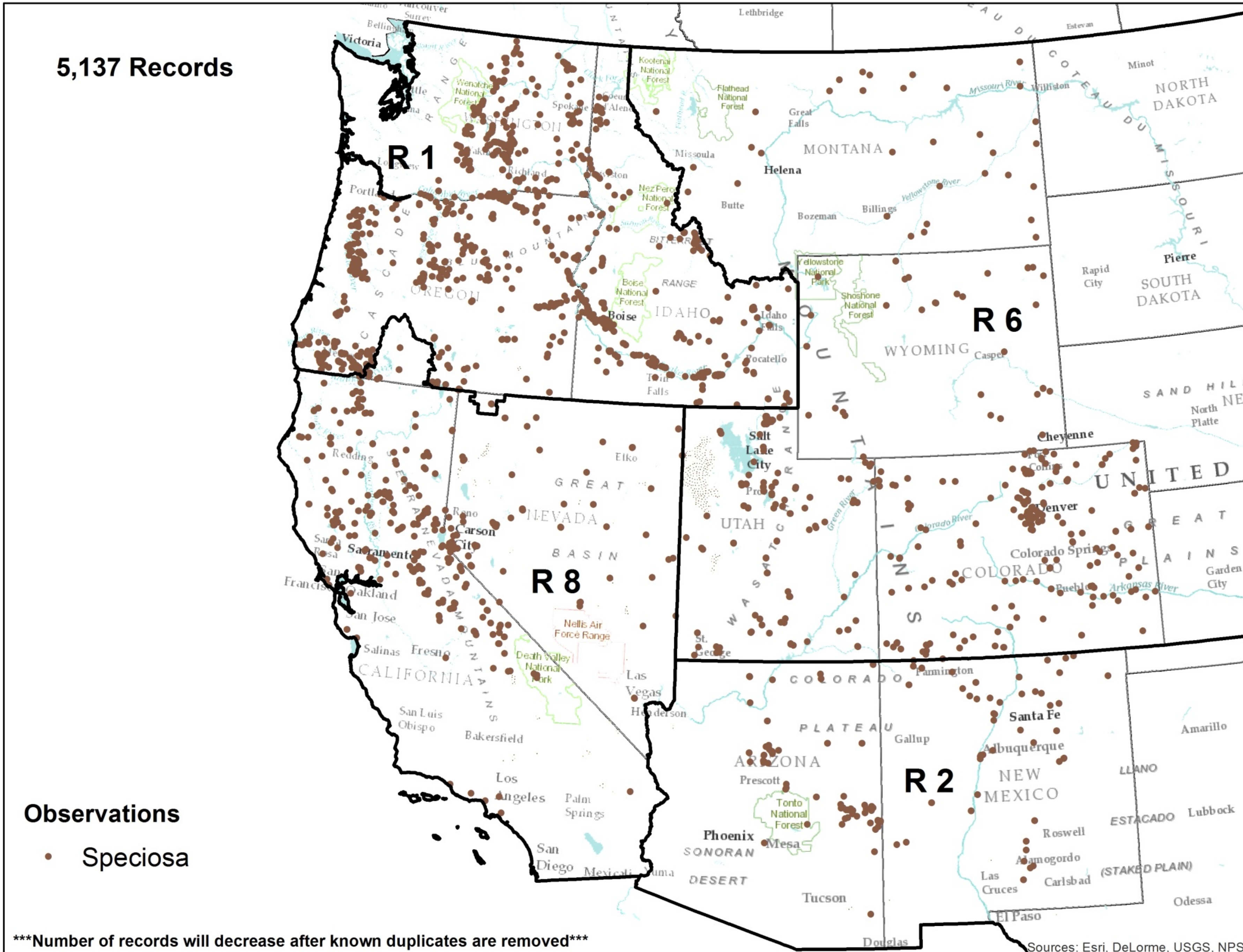

20 Records
R 1
R 6
R 8
R 2
WASHINGTON
OREGON
IDAHO
MONTANA
WYOMING
NORTH DAKOTA
SOUTH DAKOTA
NEVADA
UTAH
COLORADO
CALIFORNIA
ARIZONA
NEW MEXICO
Medium to High Accuracy Observations
Californica
***Number of records will decrease after known duplicates are removed***
Sources: Esri, DeLorme, USGS, NPS


933 Records
R 1
R 6
R 8
R 2
WASHINGTON
OREGON
IDAHO
MONTANA
WYOMING
NORTH DAKOTA
SOUTH DAKOTA
NEVADA
UTAH
COLORADO
CALIFORNIA
ARIZONA
NEW MEXICO
Observations
Cordifolia
***Number of records will decrease after known duplicates are removed***
Sources: Esri, DeLorme, USGS, NPS

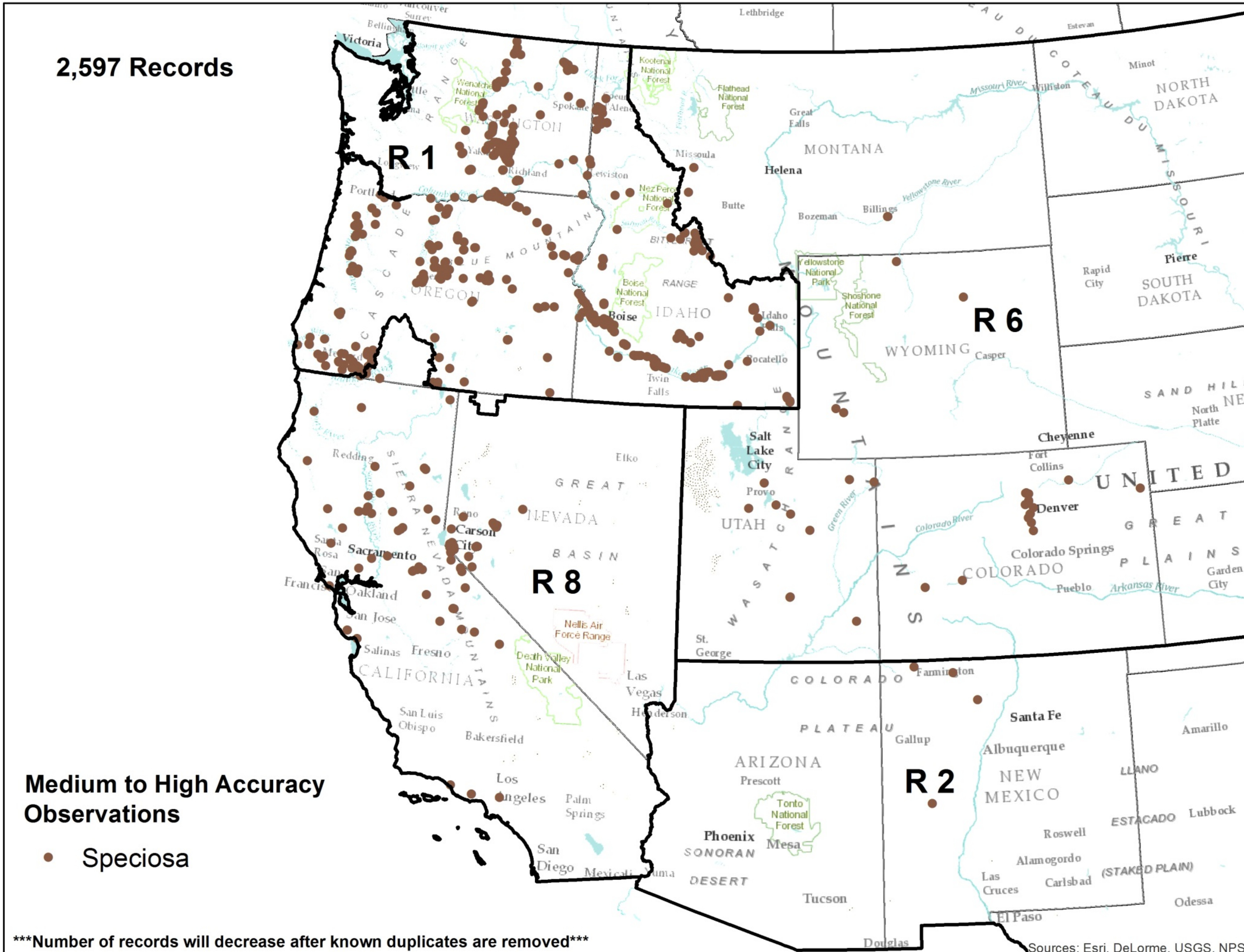

65 Records
R 1
R 6
R 8
R 2
WASHINGTON
OREGON
IDAHO
MONTANA
WYOMING
NEVADA
UTAH
COLORADO
CALIFORNIA
ARIZONA
NEW MEXICO
NORTH DAKOTA
SOUTH DAKOTA
Medium to High Accuracy Observations
Cordifolia
***Number of records will decrease after known duplicates are removed***
Sources: Esri, DeLorme, USGS, NPS

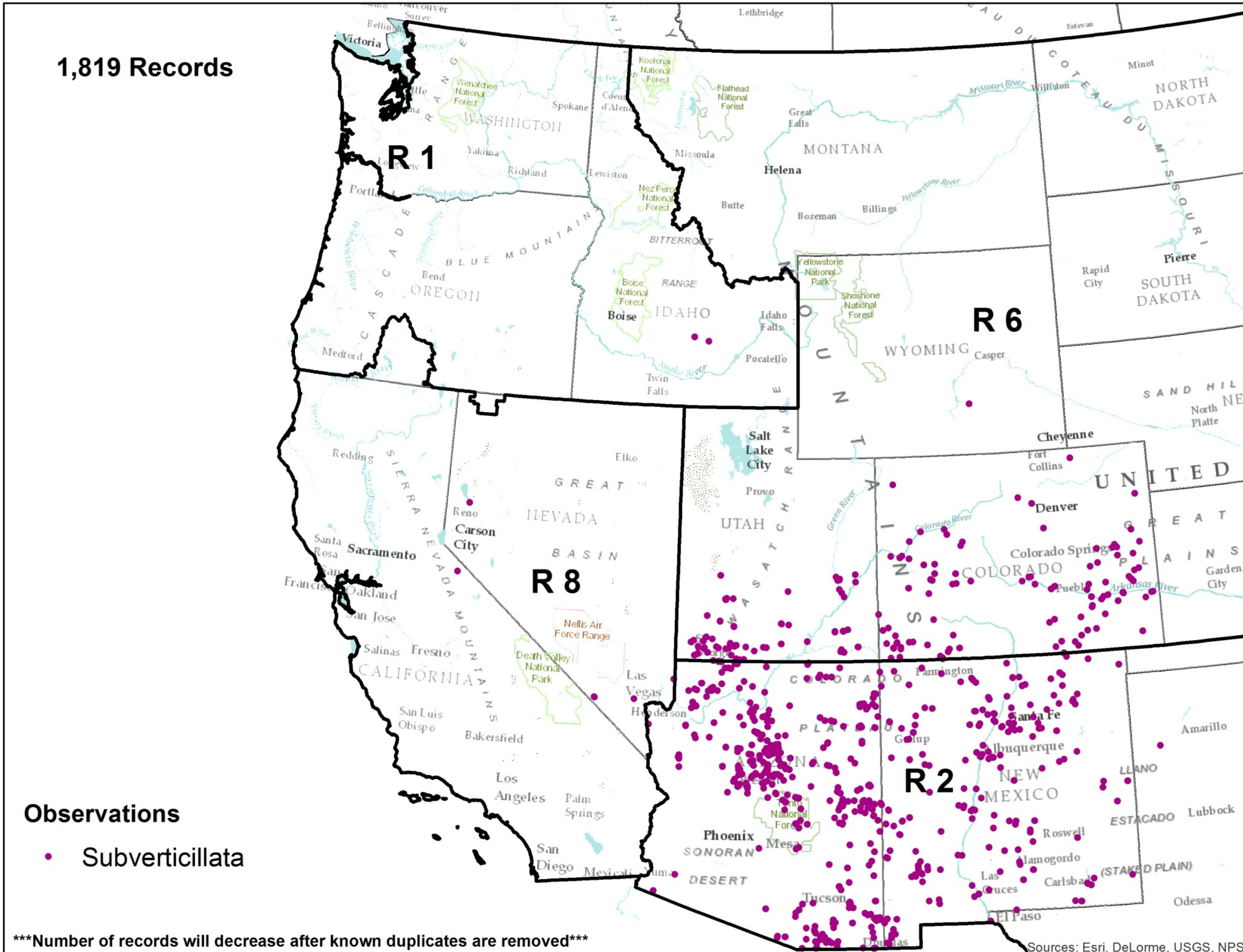

2,666 Records
R 1
R 6
R 8
R 2
Observations
Fascicularis
***Number of records will decrease after known duplicates are removed***
Sources: Esri, DeLorme, USGS, NPS
WASHINGTON
OREGON
IDAHO
MONTANA
WYOMING
NORTH DAKOTA
SOUTH DAKOTA
NEVADA
UTAH
COLORADO
CALIFORNIA
ARIZONA
NEW MEXICO
UNITED
Victoria
Lethbridge
Spokane
Yakima
Richland
Lewiston
Missoula
Great Falls
Helena
Butte
Bozeman
Billings
Williston
Minot
Portland
Bend
Boise
Idaho Falls
Pocatello
Twin Falls
Rapid City
Pierre
Casper
Cheyenne
North Platte
Fort Collins
Denver
Colorado Springs
Pueblo
Garden City
Salt Lake City
Provo
Elko
Redding
Sacramento
Carson City
Salinas
Fresno
Bakersfield
Las Vegas
Henderson
St. George
Farmington
Santa Fe
Gallup
Albuquerque
Amarillo
Prescott
Phoenix
Mesa
Tucson
Douglas
Roswell
Alamogordo
Las Cruces
Carlsbad
Lubbock
Odessa
El Paso
Palm Springs
Wenatchee National Forest
Kootenai National Forest
Flathead National Forest
Nez Perce National Forest
Boise National Forest
Yellowstone National Park
Shoshone National Forest
Tonto National Forest
Death Valley National Park
Nellis Air Force Range
Missouri River
Yellowstone River
Snake River
Green River
Colorado River
Arkansas River
BLUE MOUNTAINS
BITTERROOT RANGE
GREAT NEVADA BASIN
COLORADO PLATEAU
SONORAN DESERT
GREAT PLAINS
LLANO ESTACADO
(STAKED PLAIN)
SAND HILLS
WASATCH RANGE
SIERRA NEVADA

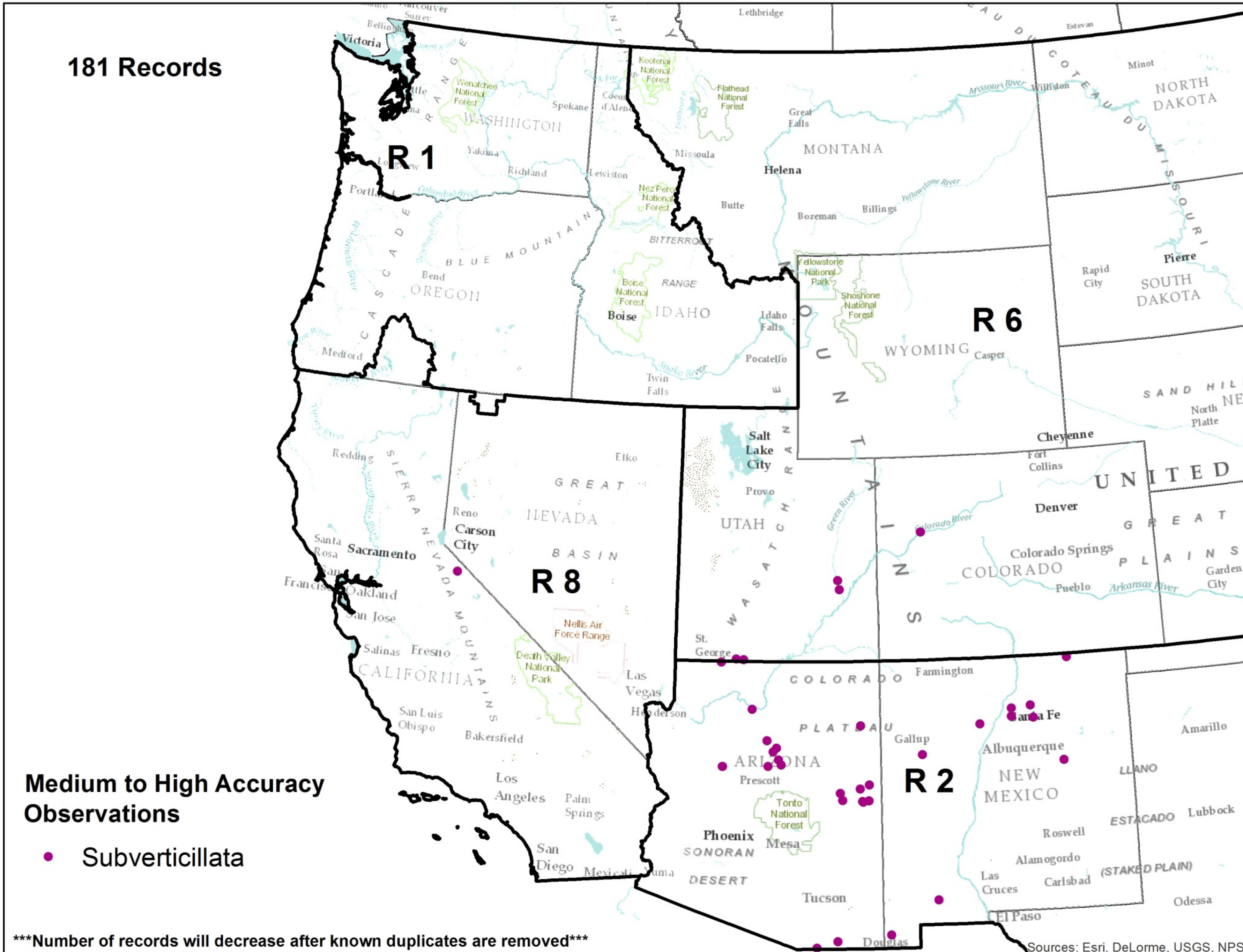

683 Records
R 1
R 6
R 8
R 2
WASHINGTON
OREGON
CALIFORNIA
NEVADA
IDAHO
MONTANA
WYOMING
UTAH
COLORADO
ARIZONA
NEW MEXICO
NORTH DAKOTA
SOUTH DAKOTA
Medium to High Accuracy Observations
Fascicularis
***Number of records will decrease after known duplicates are removed***
Sources: Esri, DeLorme, USGS, NPS

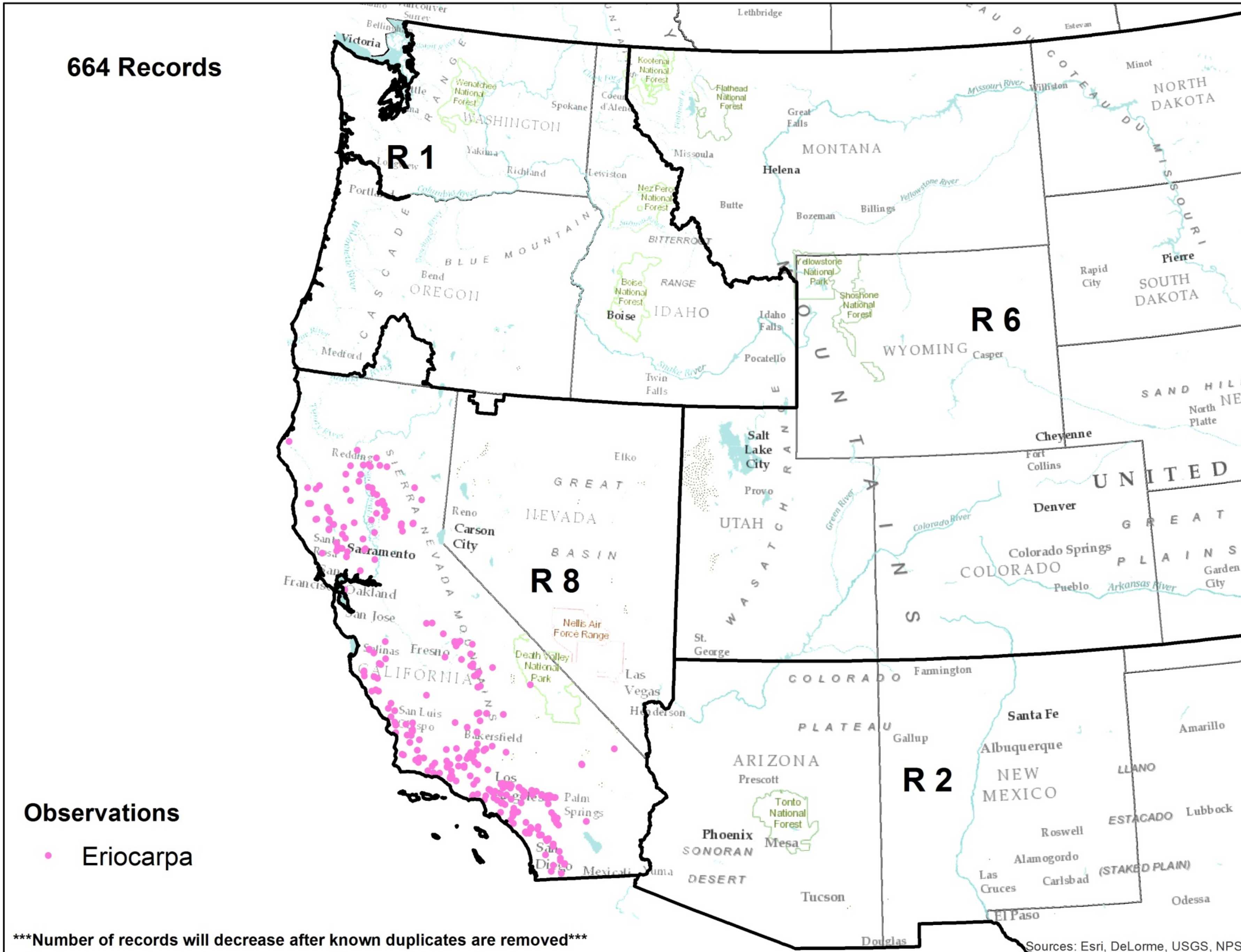


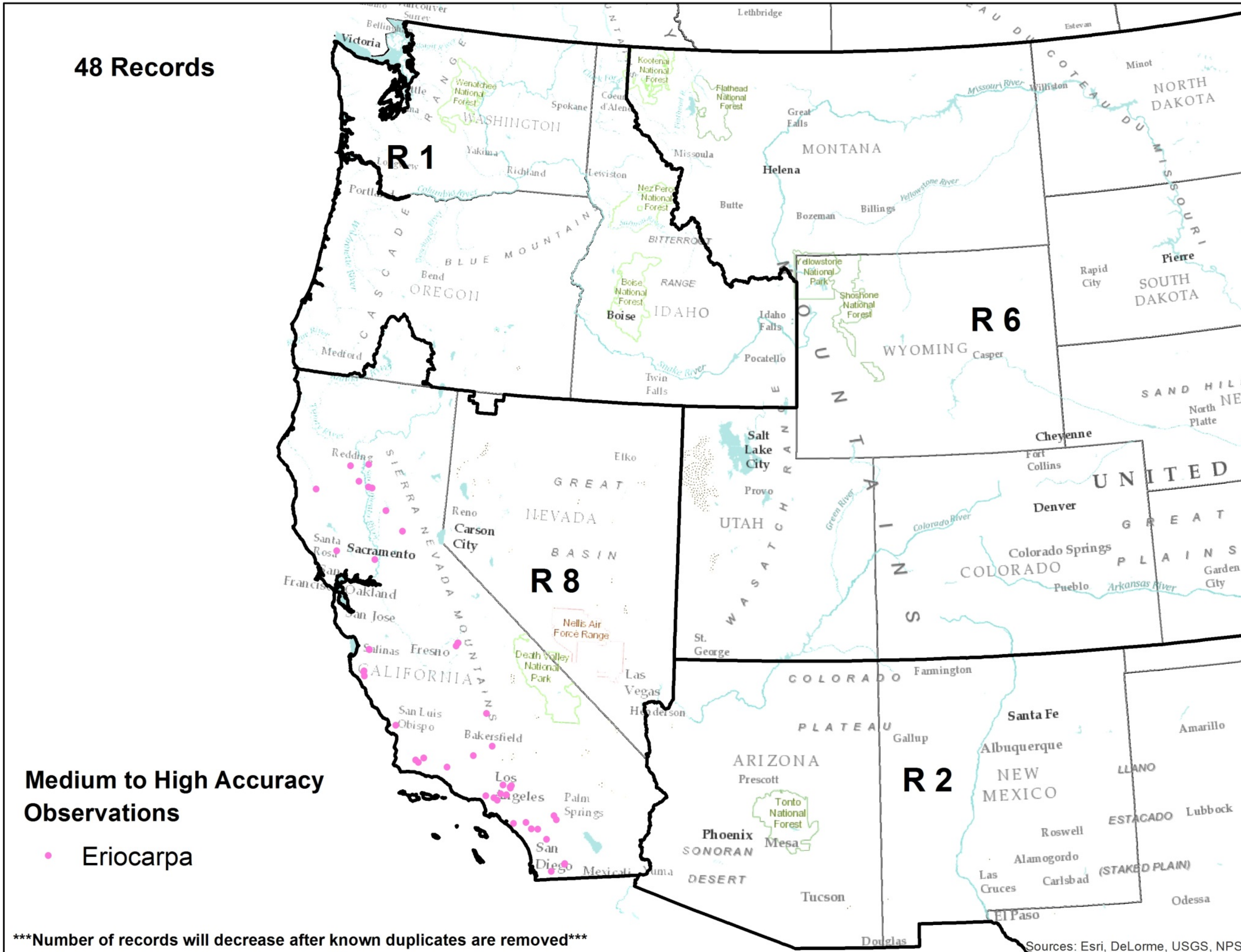
5,137 Records
R 1
R 2
R 6
R 8
WASHINGTON
OREGON
CALIFORNIA
NEVADA
IDAHO
MONTANA
WYOMING
UTAH
COLORADO
ARIZONA
NEW MEXICO
NORTH DAKOTA
SOUTH DAKOTA
GREAT BASIN
Observations
Speciosa
***Number of records will decrease after known duplicates are removed***
Sources: Esri, DeLorme, USGS, NPS

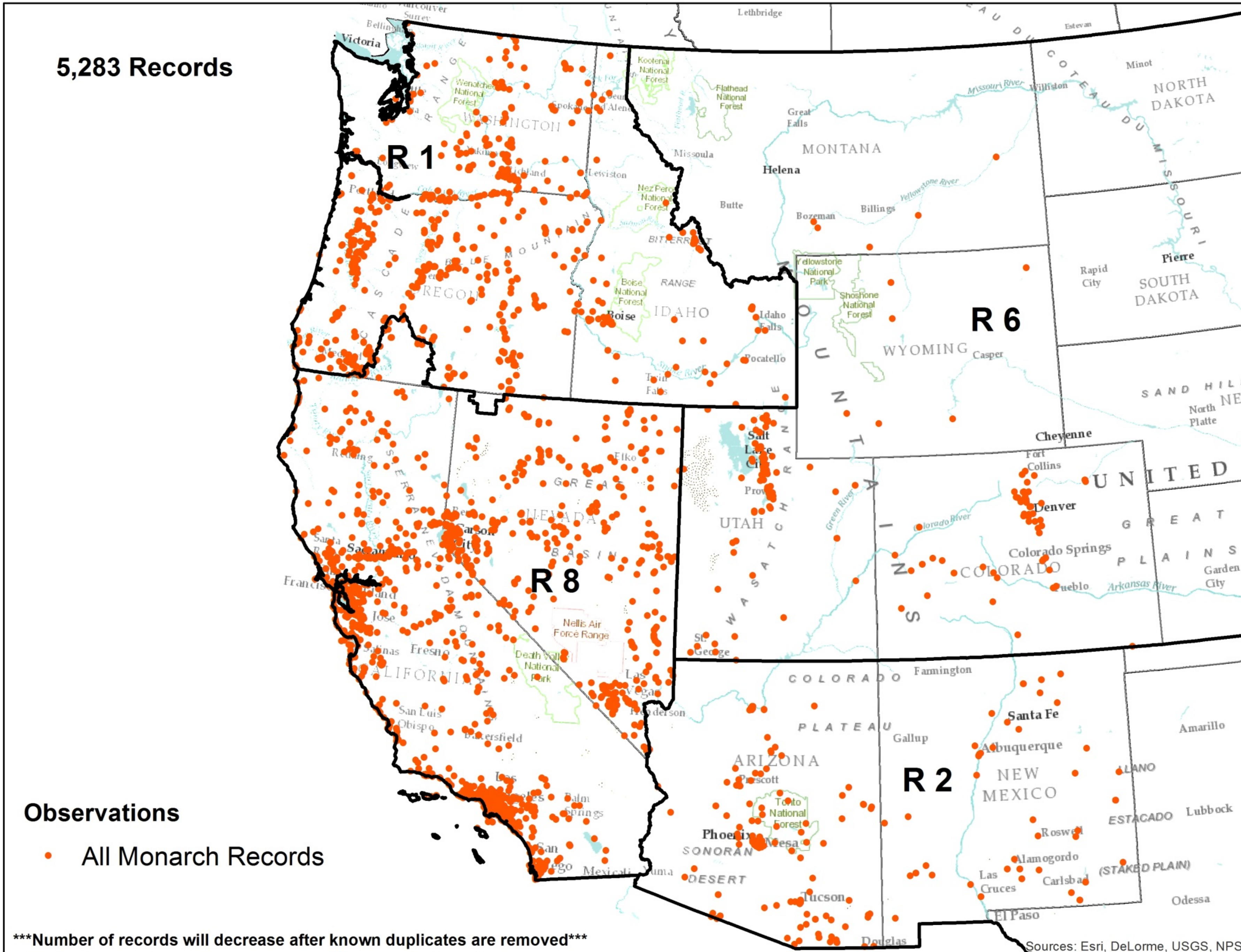

2,597 Records
R 1
R 6
R 8
R 2
WASHINGTON
OREGON
IDAHO
MONTANA
WYOMING
NORTH DAKOTA
SOUTH DAKOTA
NEVADA
UTAH
COLORADO
CALIFORNIA
ARIZONA
NEW MEXICO
UNITED
Victoria
Seattle
Spokane
Coeur d'Alene
Yakima
Richland
Lewiston
Portland
Missoula
Helena
Great Falls
Butte
Bozeman
Billings
Williston
Minot
Lethbridge
Estevan
Pierre
Rapid City
Boise
Idaho Falls
Pocatello
Twin Falls
Casper
Cheyenne
Fort Collins
Denver
Colorado Springs
Pueblo
Garden City
North Platte
Salt Lake City
Provo
St. George
Elko
Reno
Carson City
Redding
Sacramento
Santa Rosa
San Francisco
Oakland
San Jose
Salinas
Fresno
San Luis Obispo
Bakersfield
Los Angeles
Palm Springs
San Diego
Mexicali
Yuma
Las Vegas
Henderson
Farmington
Gallup
Santa Fe
Albuquerque
Amarillo
Lubbock
Roswell
Alamogordo
Las Cruces
Carlsbad
El Paso
Odessa
Prescott
Phoenix
Mesa
Tucson
Douglas
Kootenai National Forest
Flathead National Forest
Wenatchee National Forest
Nez Perce National Forest
Boise National Forest
Yellowstone National Park
Shoshone National Forest
Nellis Air Force Range
Death Valley National Park
Tonto National Forest
CASCADE RANGE
BLUE MOUNTAINS
BITTERROOT
RANGE
ROCKY MOUNTAINS
WASATCH RANGE
SIERRA NEVADA MOUNTAINS
GREAT BASIN
GREAT PLAINS
COTEAU DU MISSOURI
SAND HILLS
COLORADO PLATEAU
SONORAN DESERT
LLANO ESTACADO
(STAKED PLAIN)
Missouri River
Yellowstone River
Columbia River
Green River
Colorado River
Arkansas River
Medium to High Accuracy Observations
Speciosa
***Number of records will decrease after known duplicates are removed***
Sources: Esri, DeLorme, USGS, NPS

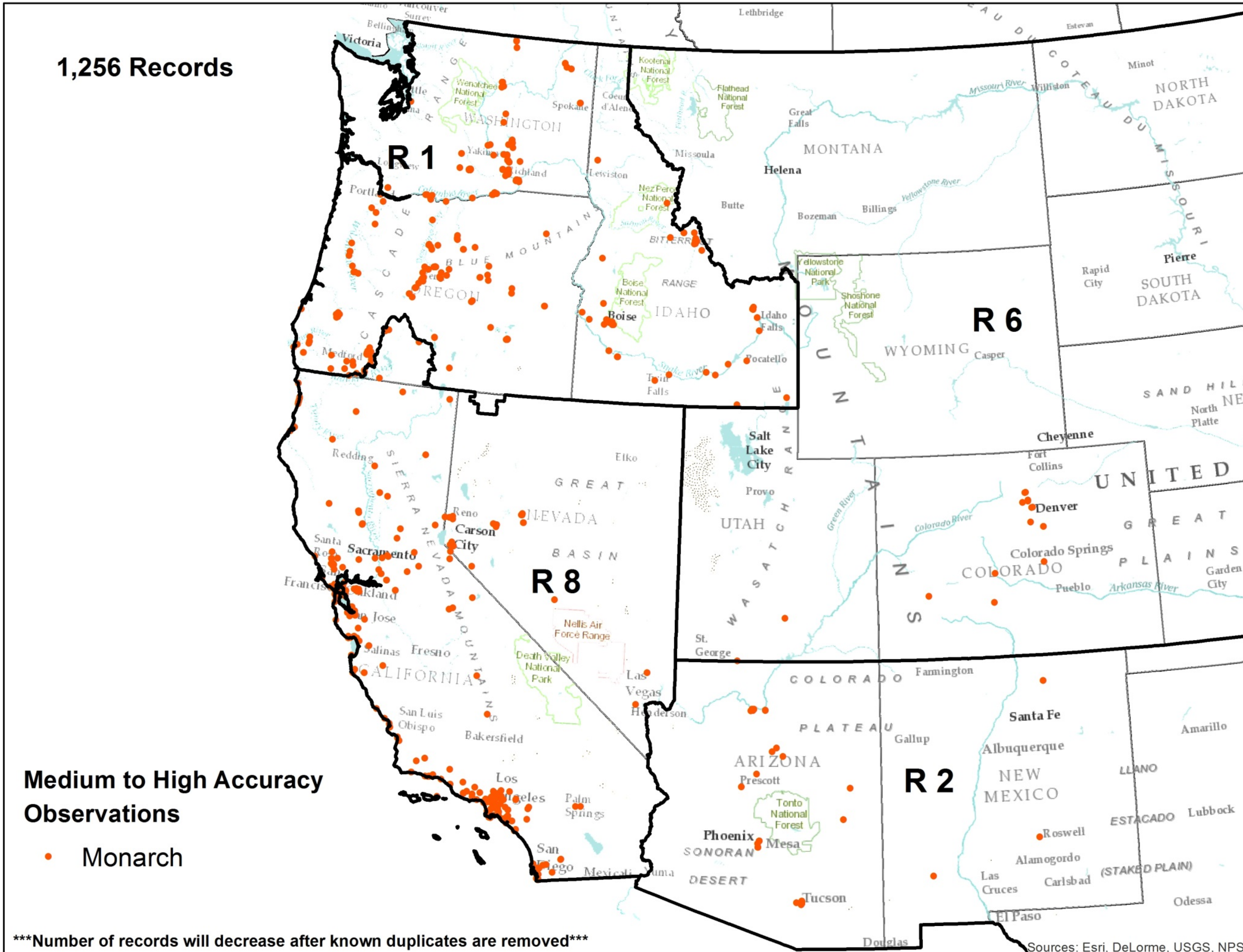

1,819 Records
R 1
R 6
R 8
R 2
WASHINGTON
OREGON
IDAHO
MONTANA
WYOMING
NORTH DAKOTA
SOUTH DAKOTA
NEVADA
UTAH
COLORADO
CALIFORNIA
ARIZONA
NEW MEXICO
Observations
Subverticillata
***Number of records will decrease after known duplicates are removed***
Sources: Esri, DeLorme, USGS, NPS

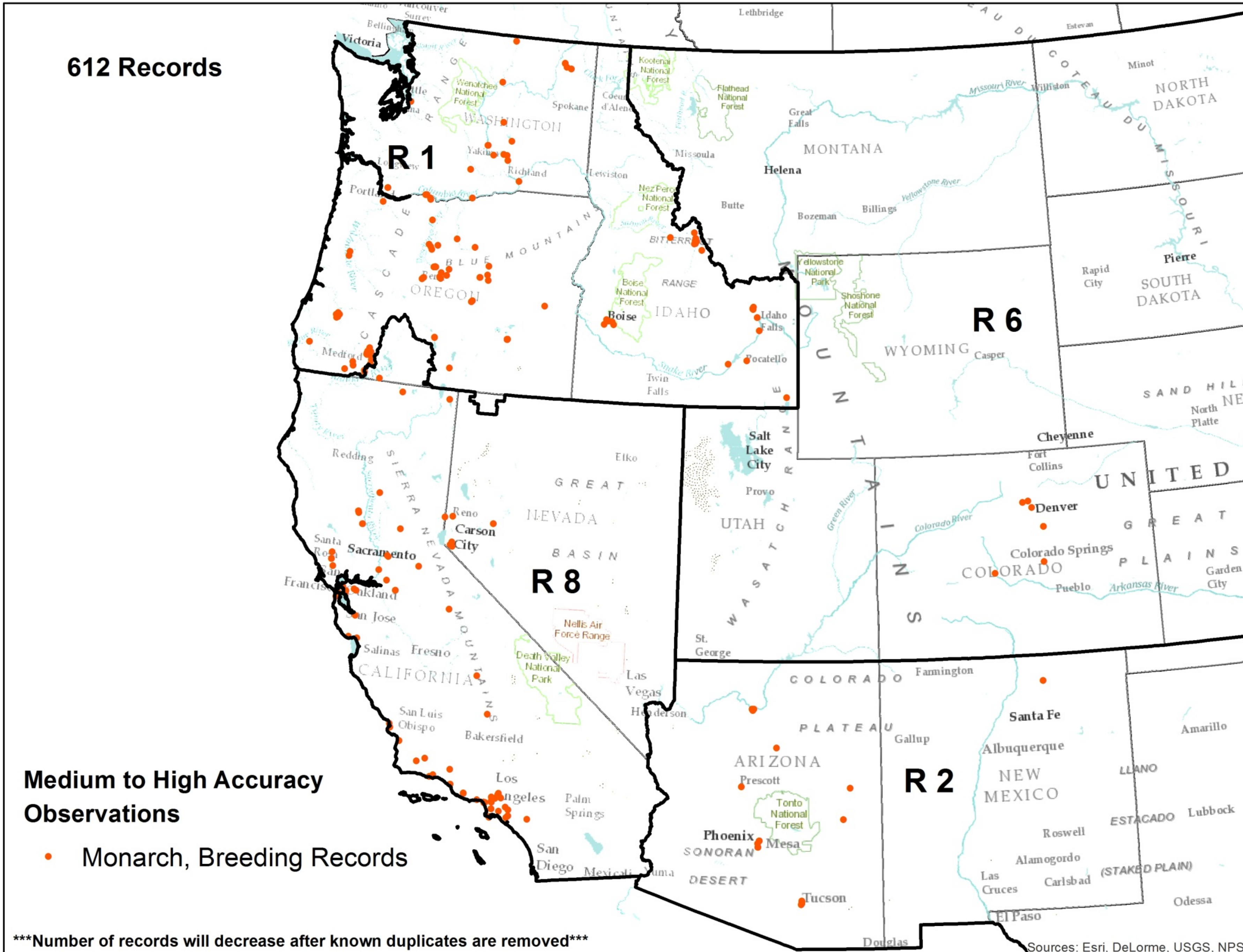

181 Records
R 1
R 6
R 8
R 2
Medium to High Accuracy Observations
Subverticillata
***Number of records will decrease after known duplicates are removed***
Sources: Esri, DeLorme, USGS, NPS


664 Records
R 1
R 2
R 6
R 8
WASHINGTON
OREGON
IDAHO
MONTANA
WYOMING
NEVADA
UTAH
COLORADO
CALIFORNIA
ARIZONA
NEW MEXICO
NORTH DAKOTA
SOUTH DAKOTA
Victoria
Spokane
Portland
Bend
Medford
Boise
Idaho Falls
Pocatello
Twin Falls
Missoula
Helena
Great Falls
Butte
Bozeman
Billings
Williston
Minot
Pierre
Rapid City
Casper
Cheyenne
Fort Collins
Denver
Colorado Springs
Pueblo
Garden City
North Platte
Salt Lake City
Provo
St. George
Elko
Reno
Carson City
Las Vegas
Henderson
Redding
Sacramento
Oakland
San Jose
Fresno
Salinas
San Luis Obispo
Bakersfield
Palm Springs
Farmington
Gallup
Santa Fe
Albuquerque
Amarillo
Lubbock
Roswell
Alamogordo
Las Cruces
Carlsbad
Odessa
El Paso
Prescott
Phoenix
Mesa
Tucson
Douglas
Lethbridge
Lewiston
Richland
Yakima
Kootenai National Forest
Flathead National Forest
Wenatchee National Forest
Nez Perce National Forest
Boise National Forest
Yellowstone National Park
Shoshone National Forest
Death Valley National Park
Nellis Air Force Range
Tonto National Forest
BLUE MOUNTAIN
BITTERROOT RANGE
GREAT BASIN
SIERRA NEVADA MOUNTAINS
ROCKY MOUNTAINS
WASATCH RANGE
COLORADO PLATEAU
SONORAN DESERT
GREAT PLAINS
LLANO ESTACADO
(STAKED PLAIN)
SAND HILLS
UNITED
Missouri River
Yellowstone River
Colorado River
Arkansas River
Green River
Snake River
Columbia River
Observations
Eriocarpa
***Number of records will decrease after known duplicates are removed***
Sources: Esri, DeLorme, USGS, NPS


48 Records
Medium to High Accuracy Observations
Eriocarpa
***Number of records will decrease after known duplicates are removed***
Sources: Esri, DeLorme, USGS, NPS
R 1
R 6
R 8
R 2
WASHINGTON
OREGON
IDAHO
MONTANA
WYOMING
NORTH DAKOTA
SOUTH DAKOTA
NEVADA
UTAH
COLORADO
CALIFORNIA
ARIZONA
NEW MEXICO
UNITED
Victoria
Spokane
Yakima
Richland
Portland
Bend
Medford
Lewiston
Missoula
Helena
Butte
Bozeman
Billings
Great Falls
Williston
Minot
Pierre
Rapid City
Boise
Idaho Falls
Pocatello
Twin Falls
Casper
Cheyenne
Fort Collins
Denver
Colorado Springs
Pueblo
Garden City
North Platte
Salt Lake City
Provo
St. George
Elko
Reno
Carson City
Las Vegas
Henderson
Redding
Santa Rosa
Sacramento
Oakland
San Jose
Salinas
Fresno
San Luis Obispo
Bakersfield
Los Angeles
Palm Springs
San Diego
Farmington
Gallup
Santa Fe
Albuquerque
Amarillo
Lubbock
Roswell
Alamogordo
Las Cruces
Carlsbad
Odessa
El Paso
Prescott
Phoenix
Mesa
Tucson
Douglas
Lethbridge
Estevan
Kootenai National Forest
Flathead National Forest
Wenatchee National Forest
Nez Perce National Forest
Boise National Forest
Yellowstone National Park
Shoshone National Forest
Nellis Air Force Range
Death Valley National Park
Tonto National Forest
CASCADE RANGE
BLUE MOUNTAINS
BITTERROOT RANGE
ROCKY MOUNTAINS
WASATCH RANGE
GREAT BASIN
SIERRA NEVADA MOUNTAINS
COLORADO PLATEAU
SONORAN DESERT
GREAT PLAINS
SAND HILLS
LLANO ESTACADO
(STAKED PLAIN)
COTEAU DU MISSOURI
Missouri River
Yellowstone River
Columbia River
Snake River
Colorado River
Green River
Arkansas River


5,283 Records
R 1
R 6
R 8
R 2
WASHINGTON
OREGON
IDAHO
MONTANA
WYOMING
NORTH DAKOTA
SOUTH DAKOTA
NEVADA
UTAH
COLORADO
CALIFORNIA
ARIZONA
NEW MEXICO
UNITED
GREAT PLAINS
GREAT BASIN
COLORADO PLATEAU
SONORAN DESERT
LLANO ESTACADO
(STAKED PLAIN)
Victoria
Lethbridge
Estevan
Minot
Williston
Great Falls
Missoula
Helena
Butte
Bozeman
Billings
Lewiston
Boise
Idaho Falls
Pocatello
Twin Falls
Casper
Rapid City
Pierre
Cheyenne
North Platte
Fort Collins
Denver
Colorado Springs
Pueblo
Garden City
St George
Farmington
Santa Fe
Albuquerque
Gallup
Amarillo
Lubbock
Roswell
Alamogordo
Las Cruces
Carlsbad
El Paso
Odessa
Prescott
Phoenix
Mesa
Tucson
Douglas
Fresno
Salinas
San Luis Obispo
Bakersfield
Palm Springs
Henderson
Kootenai National Forest
Flathead National Forest
Wenatchee National Forest
Nez Perce National Forest
Boise National Forest
Yellowstone National Park
Shoshone National Forest
Tonto National Forest
Death Valley National Park
Nellis Air Force Range
Missouri River
Yellowstone River
Colorado River
Green River
Arkansas River
Snake River
Observations
All Monarch Records
***Number of records will decrease after known duplicates are removed***
Sources: Esri, DeLorme, USGS, NPS


1,256 Records
R 1
R 6
R 8
R 2
WASHINGTON
OREGON
IDAHO
MONTANA
WYOMING
NORTH DAKOTA
SOUTH DAKOTA
NEVADA
UTAH
COLORADO
CALIFORNIA
ARIZONA
NEW MEXICO
Medium to High Accuracy Observations
Monarch
***Number of records will decrease after known duplicates are removed***
Sources: Esri, DeLorme, USGS, NPS

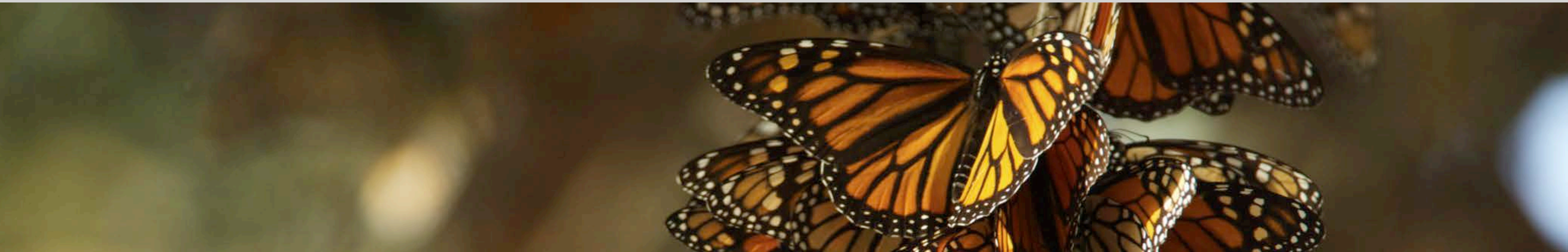

612 Records
R 1
R 6
R 8
R 2
WASHINGTON
OREGON
IDAHO
MONTANA
WYOMING
NORTH DAKOTA
SOUTH DAKOTA
NEVADA
UTAH
COLORADO
CALIFORNIA
ARIZONA
NEW MEXICO
Medium to High Accuracy Observations
Monarch, Breeding Records
***Number of records will decrease after known duplicates are removed***
Sources: Esri, DeLorme, USGS, NPS

# What’s next?

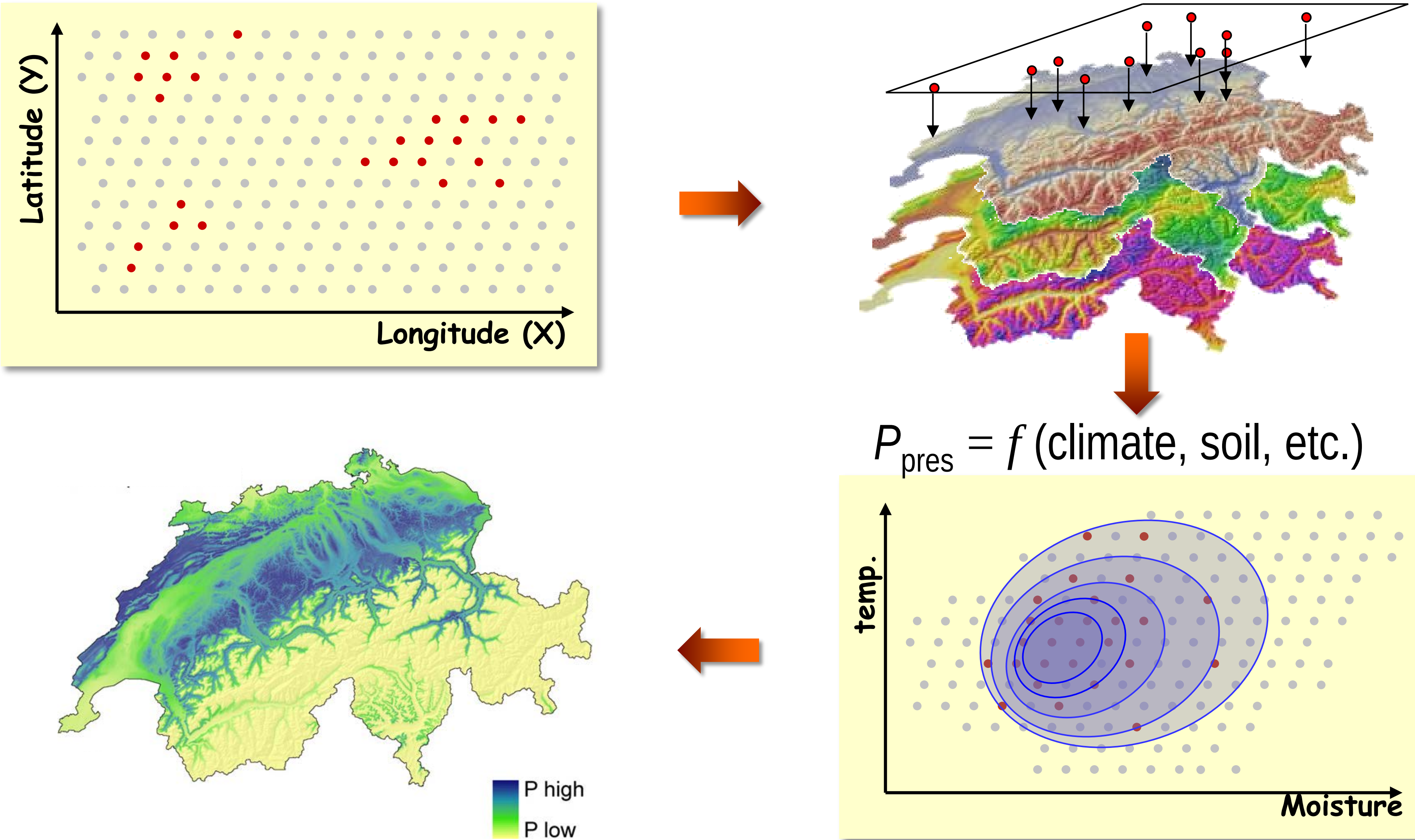
- Continue to standardize data
- Remove duplicates
- Break up into relational tables (normalize data)
- Create useful queries for end-users
- Fill the data gaps
- Compile notes and document the process
- **Increase number of georeferenced records**
- Share!

# Modeling methods

- MaxEnt, the modeling tool
- Defining the analysis area for each species
- Mitigating for sampling bias
- Environmental covariate selection
- Model calibration and selection



# Species Distribution Modeling Basics




Slide credit: Niklaus Zimmermann and Thomas Edwards: IALE 2015 Species Distribution Modeling using R course

# MaxEnt, the modeling tool

"The idea of Maxent is to estimate a target probability distribution by finding the probability distribution of maximum entropy (i.e., that is most spread out, or closest to uniform), subject to a set of constraints that represent our incomplete information about the target distribution. The information available about the target distribution [is]… a set of real-valued variables… and the constraints are that the expected value of each feature should match its empirical average."

- Philips et al. 2006

# Why MaxEnt

- Presence only data
- Explore complex relationships with environment (with user control of allowed complexity)
- Considered good for small data sets
- Performs well relative to other machine learning  models (e.g., Random Forests) when they are run with presence/pseudo-absence data (Elith et al. 2006)
- Fairly well documented and studied

# Defining the geographic background for each species

- Known to have strong effect on the model outputs
- Using a calibrated buffer approach described in VanDerWal et al. 2009

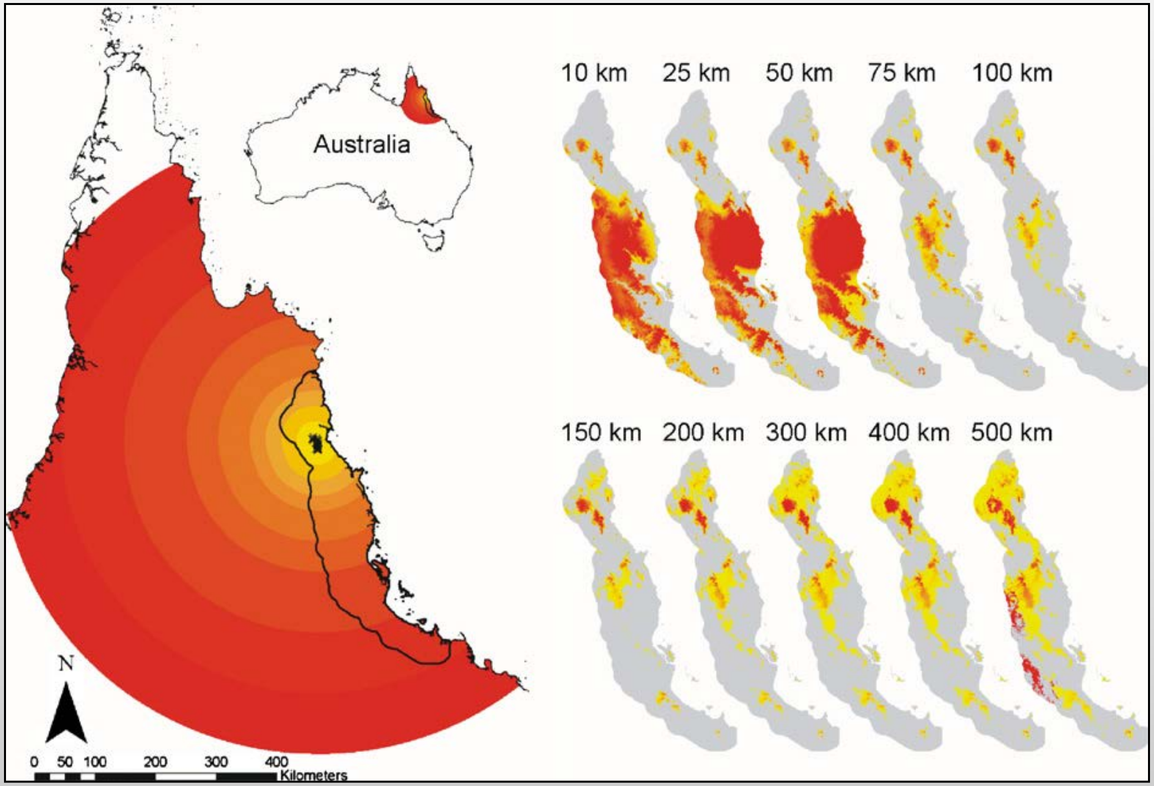



Fig. 1, VanderWal et al. 2009, Ecological Modelling

# Across 12 species, results improve; then plateaued

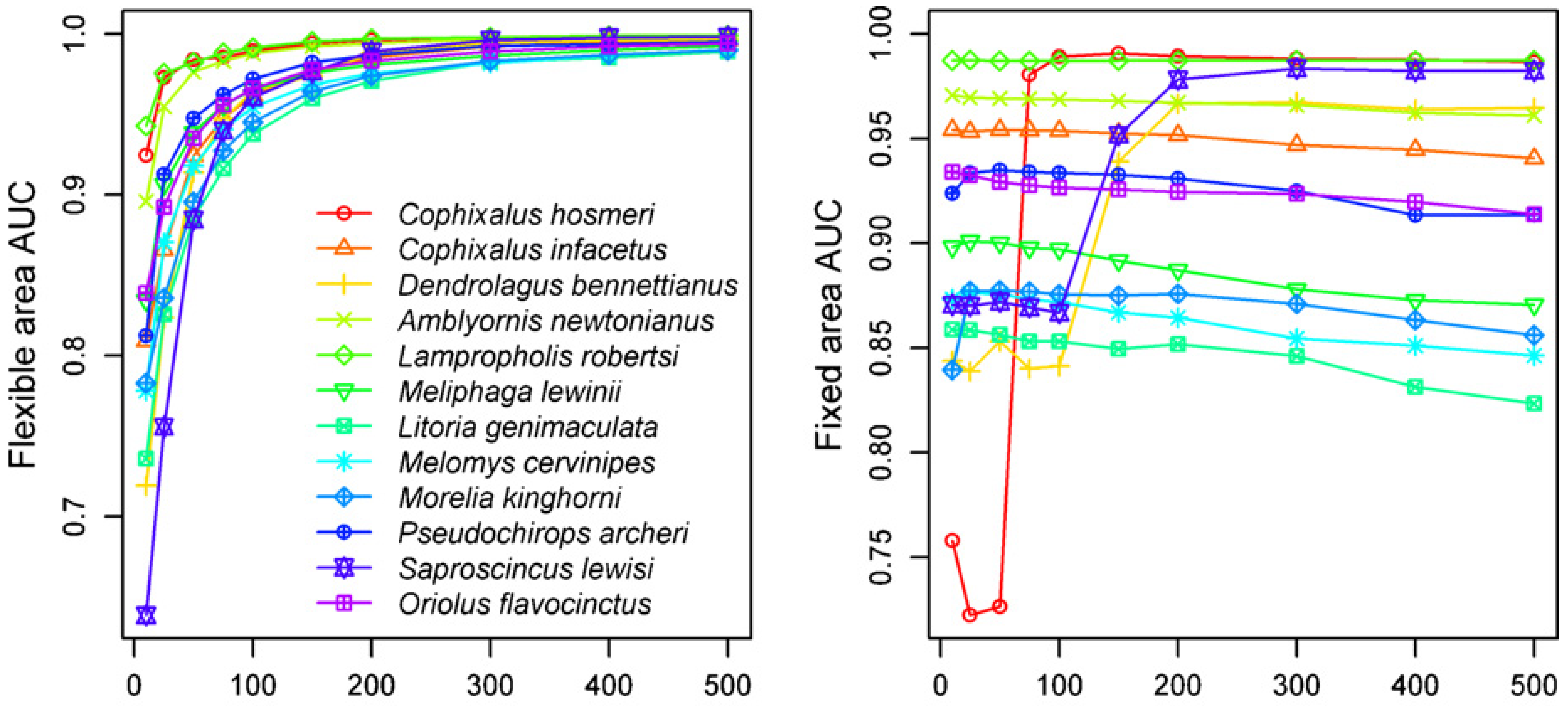



Fig. 2, VanderWal et al. 2009, Ecological Modelling

# Similar findings in Iturbide et al. 2015

And they released an R package that makes this buffer calibration much easier!

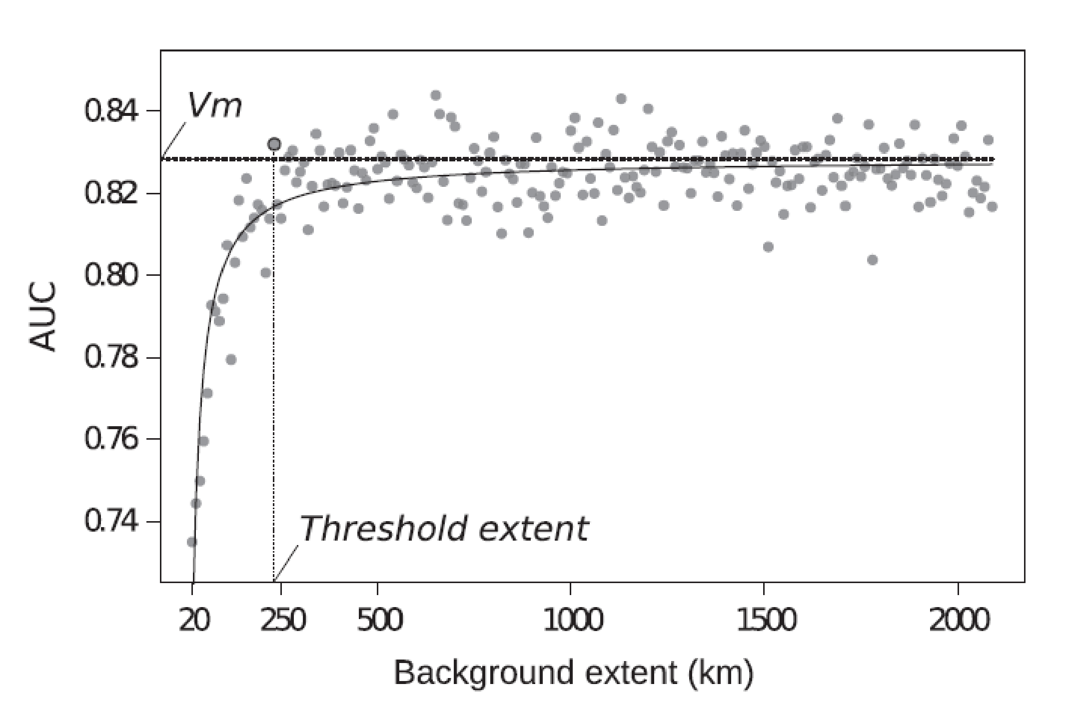
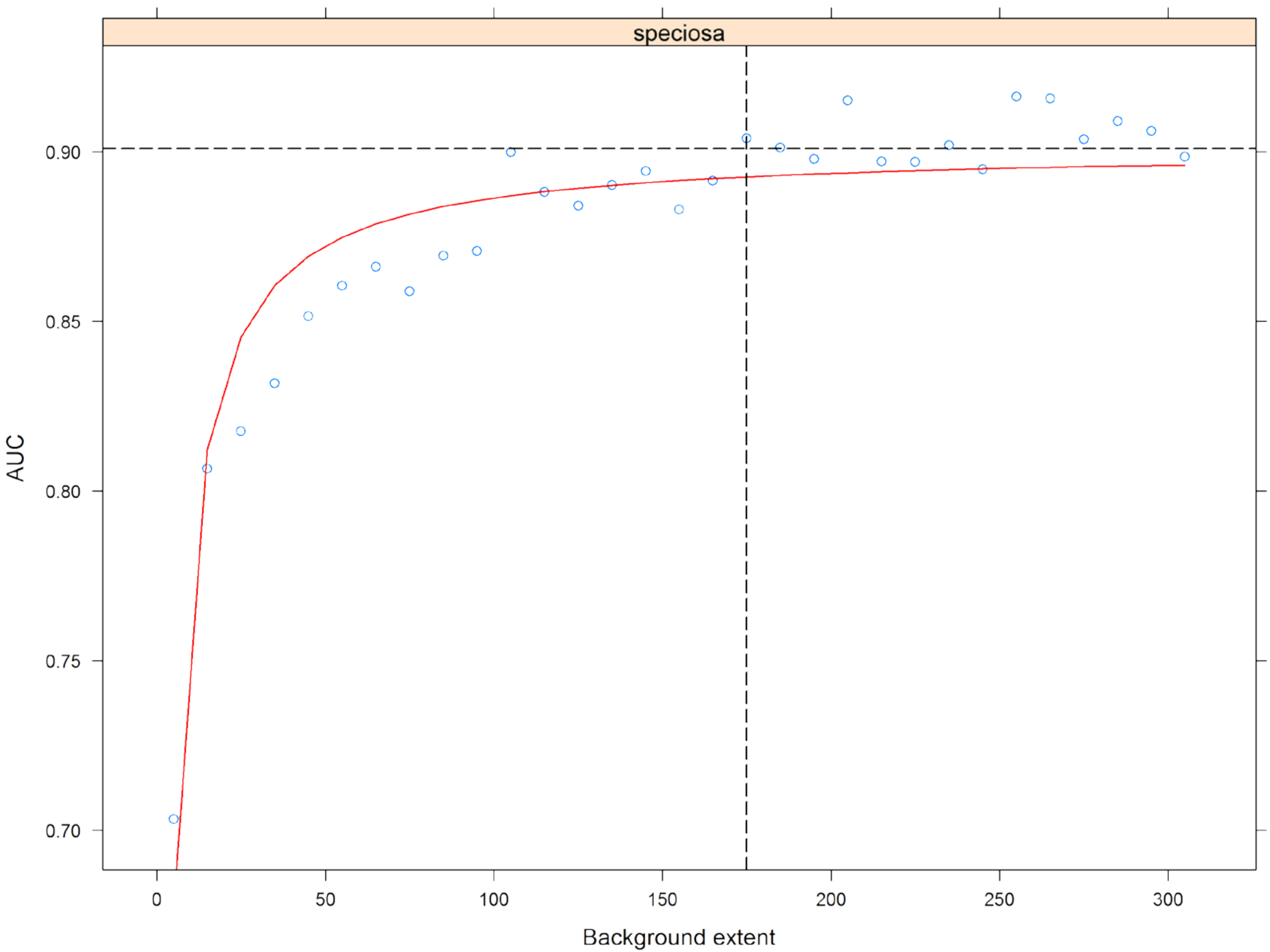
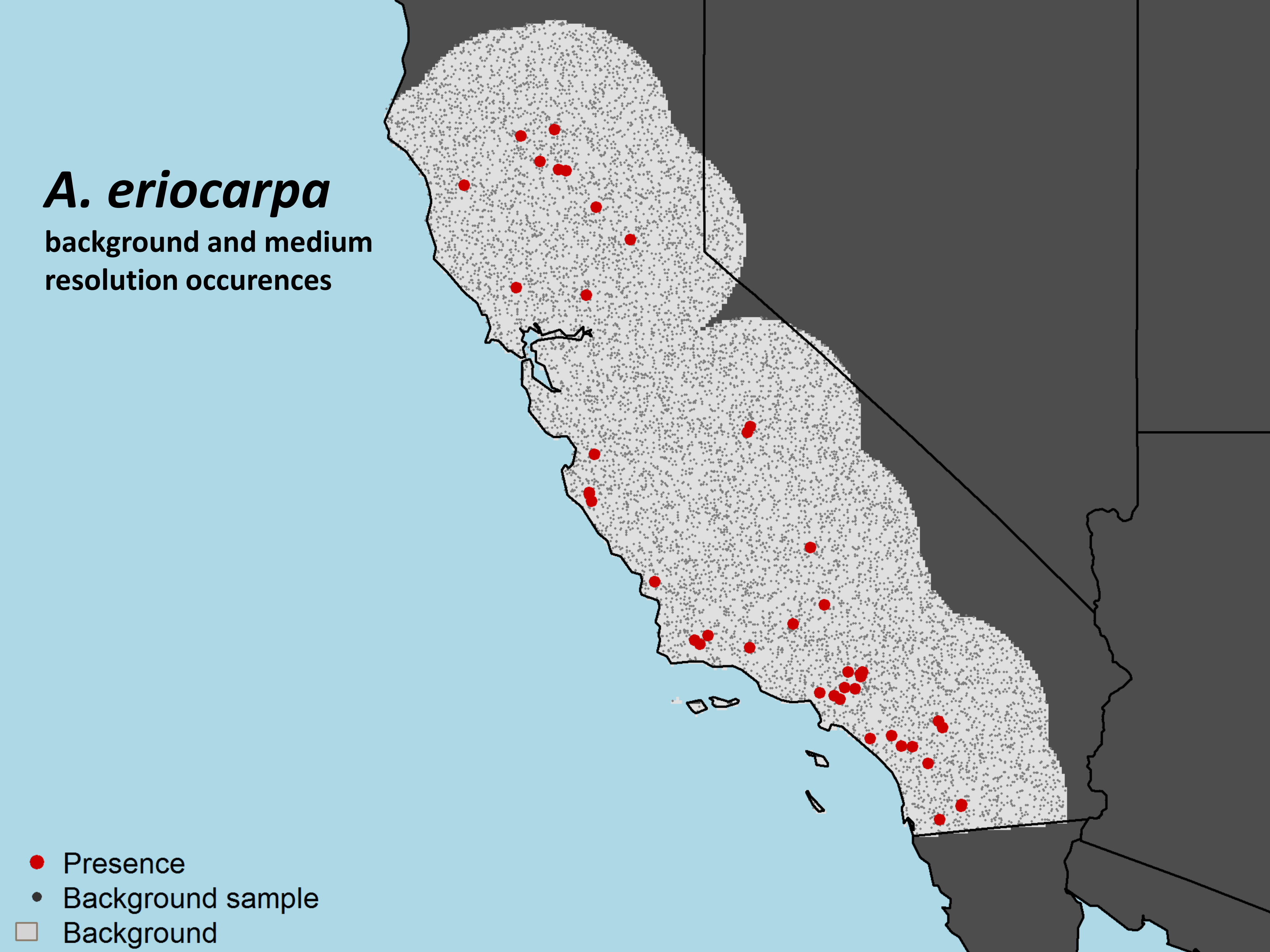



Fig. 3, Iturbide et al. 2015, Ecological Modelling

# Example background extent selection with our data




A. eriocarpa
background and medium resolution occurences
Presence
Background sample
Background

# Correction for sampling bias

- Serious problem in many data sets, including ours
- Fourcade et al. did a particularly thorough exploration:
  - Developed 36 artificially biased datasets (3 species X 4 types of biases X 3 intensity levels)
  - For each, tried 5 different bias correction methods

# Mitigating for sampling bias: results from Fourcade et al. 2014

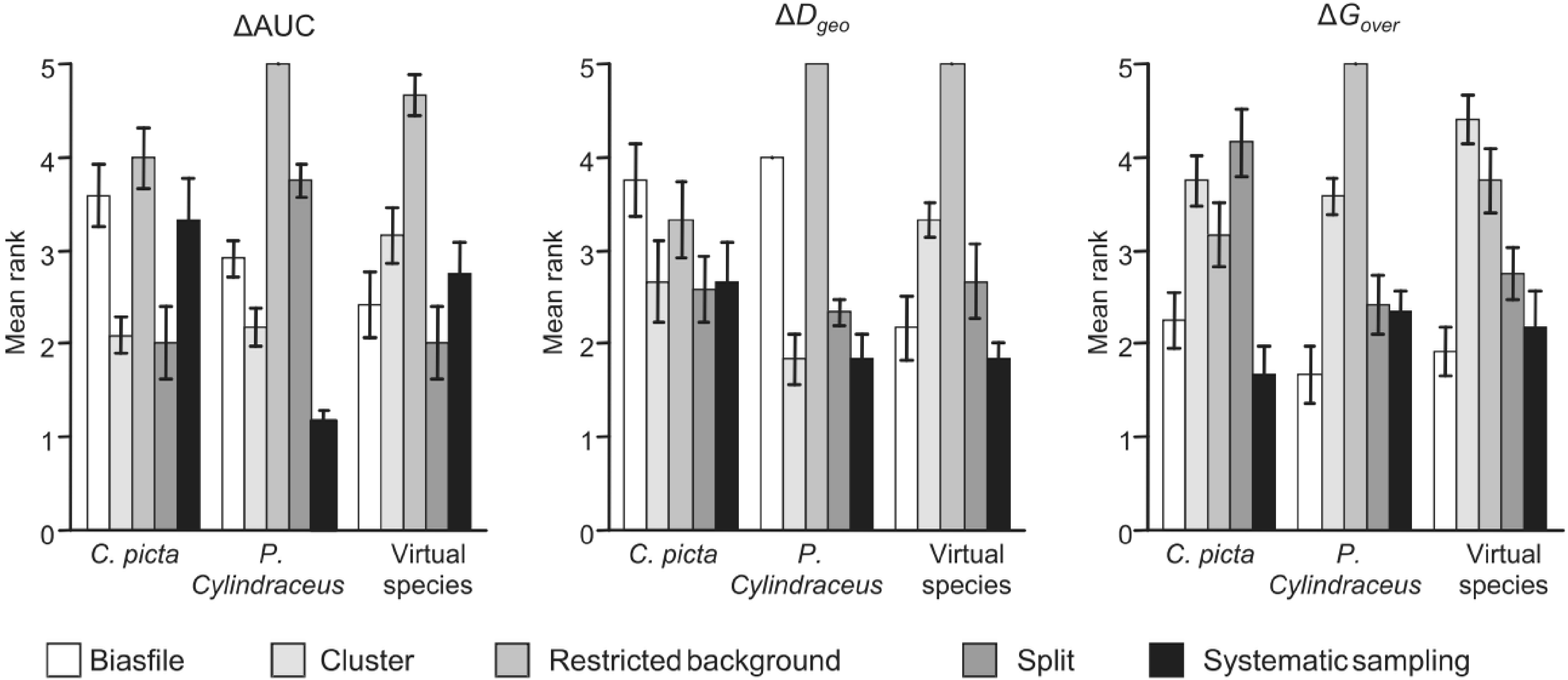



**Figure 5. Rank of each method to correct sampling bias.** Mean ranks ± standard-error for the performance of each method to correct sampling bias for each species (*Chrysemys picta*: left, *Plethodon cylindraceus*: center, *virtual species*: right), following 3 measures of correction performance: ΔAUC (left), $\Delta D_{geo}$ (centre), and $\Delta G_{over}$ (right). For each type of bias and bias intensity, the method which results in the most efficient correction is set to 1 whereas the least powerful method is set to 5. The plotted values are the mean rank across the 4 types of bias and 3 intensities.
doi:10.1371/journal.pone.0097122.g005

# Environmental Covariates

- Based on expert input and data availability
- Study domain is conterminous US, west of continental divide
- 25 total in current candidate set
  - (More on selection of best subset for each species later)
- Experimented with 3 resolutions: 90m, 270m, and 900m pixels
- 270 seemed most reasonable given resolution of inputs and accuracy of points

# Covariates: PRISM Climate Variables

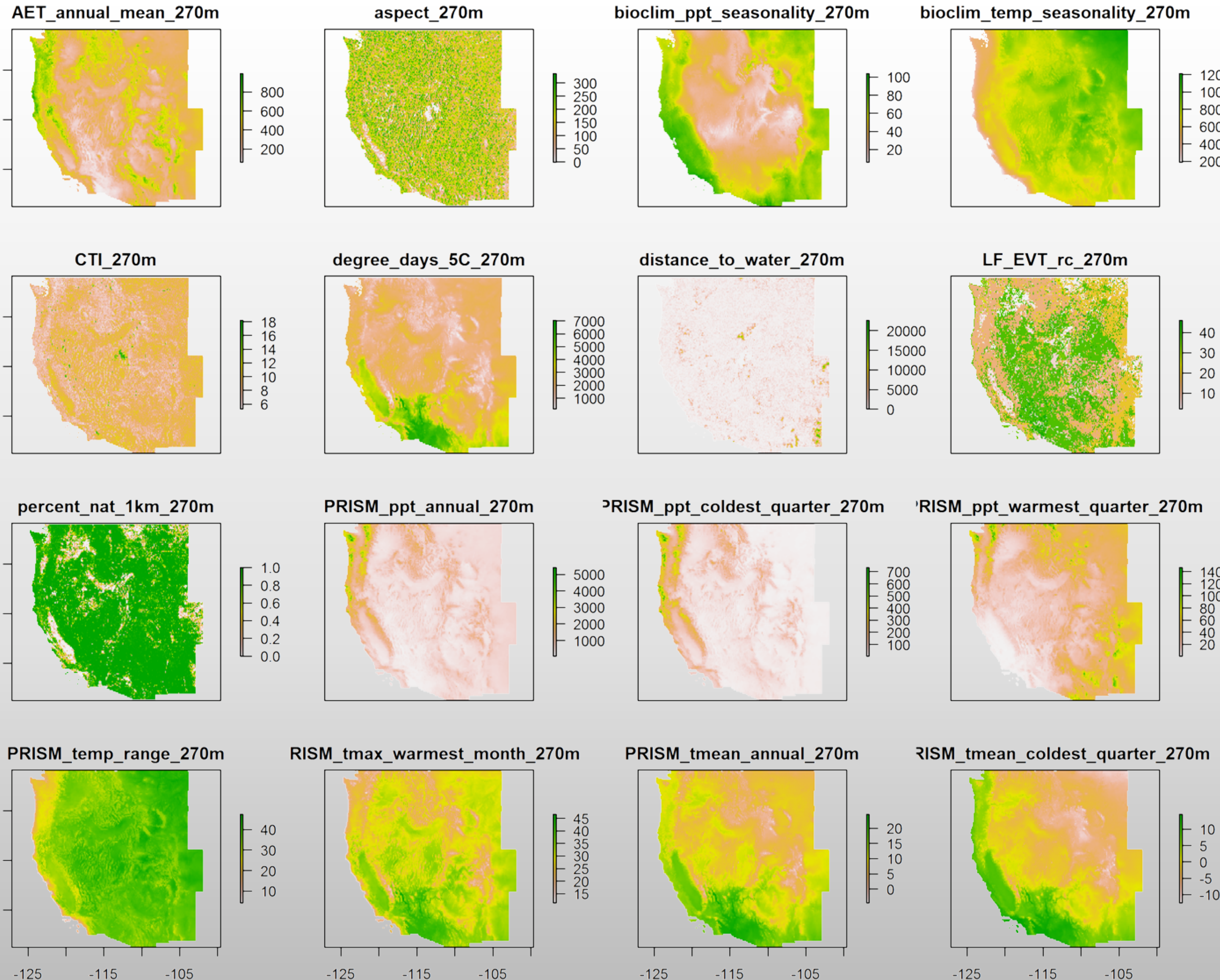
| | Variable name | Source | Original Resolution | Time period |
|---|---|---|---|---|
| 1 | Mean Annual temp. | PRISM | ~900m pixels | 30 year avg: 1981-2010 |
| 2 | Mean Annual precip. | PRISM | ~900m pixels | 30 year avg: 1981-2010 |
| 3 | Mean precip. of the Warmest Quarter | PRISM | ~900m pixels | 30 year avg: 1981-2010 |
| 4 | Mean precip. of the Coldest Quarter | PRISM | ~900m pixels | 30 year avg: 1981-2010 |
| 5 | Mean temp. of the Warmest Quarter | PRISM | ~900m pixels | 30 year avg: 1981-2010 |
| 6 | Mean temp. of the coldest quarter | PRISM | ~900m pixels | 30 year avg: 1981-2010 |
| 7 | Max temp. of the warmest month | PRISM | ~900m pixels | 30 year avg: 1981-2010 |
| 8 | Minimum temp. of the coldest month | PRISM | ~900m pixels | 30 year avg: 1981-2010 |
| 9 | temp. annual range | PRISM | ~900m pixels | 30 year avg: 1981-2010 |

# Covariates: Land cover, soil, and other

| | Variable name | Source | Original Resolution | Time period |
|---|---|---|---|---|
| 10 | Topsoil (0-30cm) bulk density | SSURGO patched with STATSGO | varies | n/a |
| 11 | Topsoil (0-30cm) clay fraction | SSURGO patched with STATSGO | varies | n/a |
| 12 | Topsoil (0-30cm) sand fraction | SSURGO patched with STATSGO | varies | n/a |
| 13 | Topsoil (0-30cm) silt fraction | SSURGO patched with STATSGO | varies | n/a |
| 14 | Topsoil (0-30cm) pH | SSURGO patched with STATSGO | varies | n/a |
| 15 | Slope | Derived from 30m DEM (3DEP) | 30m pixels | n/a |
| 16 | Aspect | Derived from 30m DEM (3DEP) | 30m pixels | n/a |
| 17 | Compound Topographic Index (CTI) | Derived from 90m DEM (SRTM) | 90m pixels | n/a |
| 18 | Distance to water | Derived from (NHD) | Calculated at 90m resolution | n/a |
| 19 | Reclassified LANDFIRE Existing Vegetation Type | LANDFIRE | 30m pixels | 2011 |
| 20 | Percent natural within 1 km | LANDFIRE | 30m pixels | 2011 |

# Covariates: Other Climate Variables

| | Variable name | Source | Original Resolution | Time period |
|---|---|---|---|---|
| 21 | Mean Actual Evapotranspiration (AET) | AdaptWest | ~900m pixels | 30 year avg: 1981-2010 |
| 22 | Mean Climatic Water Deficit, 1980-2009 | AdaptWest | ~900m pixels | 30 year avg: 1981-2010 |
| 23 | Mean precip. Seasonality (Coefficient of Variation) | BIOCLIM | ~900m pixels | 50 year avg: 1950-2000 |
| 24 | Mean temp. Seasonality (standard deviation *100) | BIOCLIM | ~900m pixels | 50 year avg: 1950-2000 |
| 25 | Annual Mean Degree Days | ClimateWNA | ~900m pixels | 30 year avg: 1961-1990 |


AET_annual_mean_270m
aspect_270m
bioclim_ppt_seasonality_270m
bioclim_temp_seasonality_270m
CTI_270m
degree_days_5C_270m
distance_to_water_270m
LF_EVT_rc_270m
percent_nat_1km_270m
PRISM_ppt_annual_270m
PRISM_ppt_coldest_quarter_270m
PRISM_ppt_warmest_quarter_270m
PRISM_temp_range_270m
RISM_tmax_warmest_month_270m
PRISM_tmean_annual_270m
RISM_tmean_coldest_quarter_270m
800 600 400 200
300 250 200 150 100 50 0
100 80 60 40 20
1200 1000 8000 6000 4000 2000
18 16 14 12 10 8 6
7000 6000 5000 4000 3000 2000 1000
20000 15000 10000 5000 0
40 30 20 10
1.0 0.8 0.6 0.4 0.2 0.0
5000 4000 3000 2000 1000
700 600 500 400 300 200 100
140 120 100 80 60 40 20
40 30 20 10
45 40 35 30 25 20 15
20 15 10 5 0
10 5 0 -5 -10
45 40 35
-125 -115 -105

# Selection of uncorrelated subsets of covariates for each species

- Multicollinearity can lead to unpredictable results and make models difficult to interpret
- Did initial runs with all variables to get "permutation importance" scores for each
- Reran with an uncorrelated subset
  - From each correlated group of variables, selected the one with highest permutation importance

# Example variable selection: *A. eriocarpa*

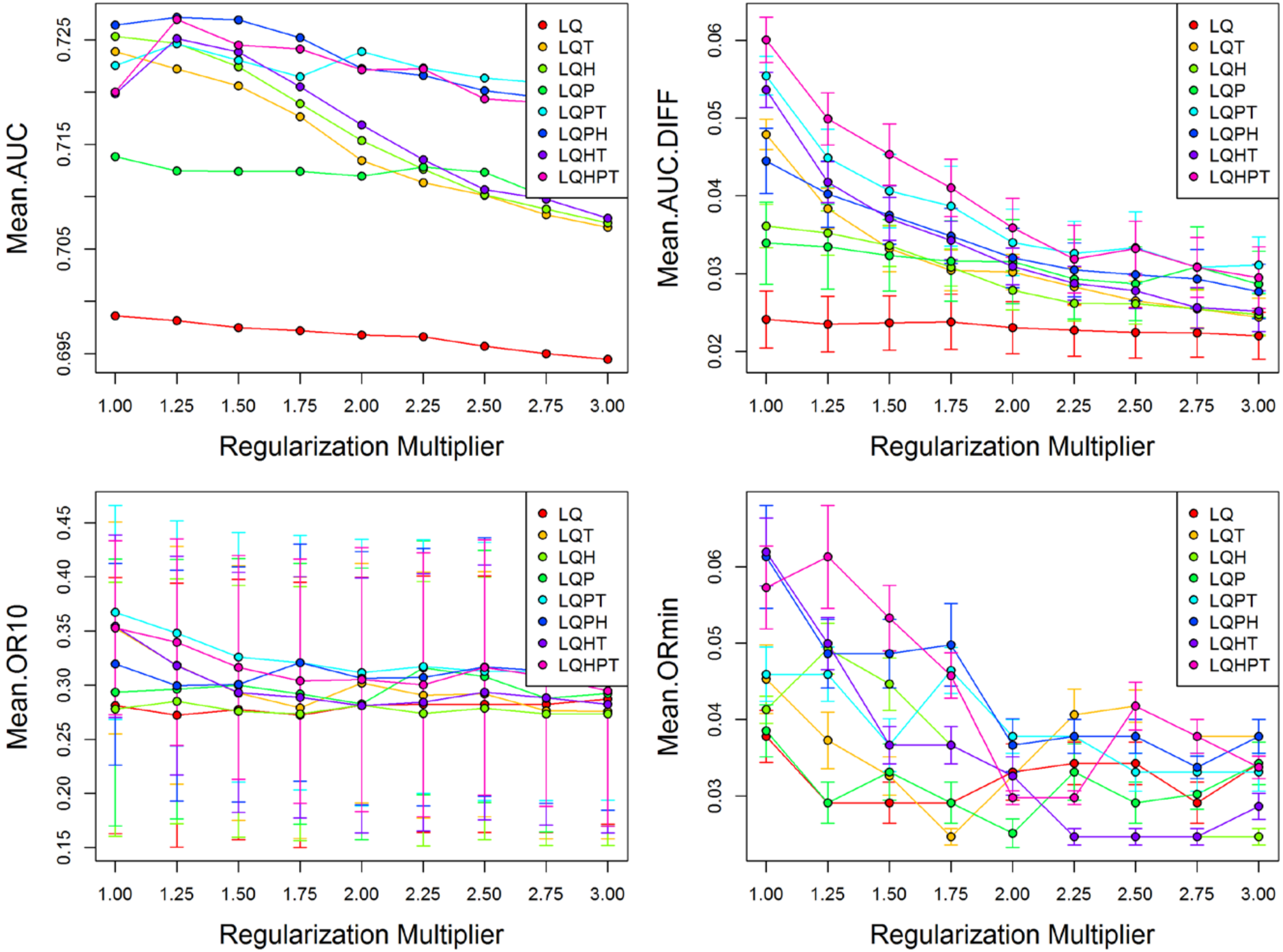
- **water deficit annual: 43.89%**
  - *degree days 5C: 2.63%*
  - *tmean annual: 2.16%*
  - *tmax warmest month: 0%*
  - *tmean warmest quarter: 0.11%*
  - *AET annual mean: 0.97%*
- **ppt warmest quarter: 12.74%**
- **ppt annual: 11.28%**
  - *tmean coldest quarter: 0%*
  - *Topsoil pH: 0%*
- **temp range: 5.22%**
  - *temp seasonality: 4.81%*
- **LANDFIRE reclass: 5.07%**
- **ppt seasonality: 3.64%**
  - *tmin coldest month: 0.72%*
- **aspect: 2.20%**
- **slope: 1.62%**
- **distance to water: 1.44%**
- **topsoil clay fraction: 1.42%**

- topsoil silt fraction: 0.09%
- combined topographic index: 0%
- ppt coldest quarter: 0%
- topsoil bulk density: 0%
- topsoil sand fraction: 0%

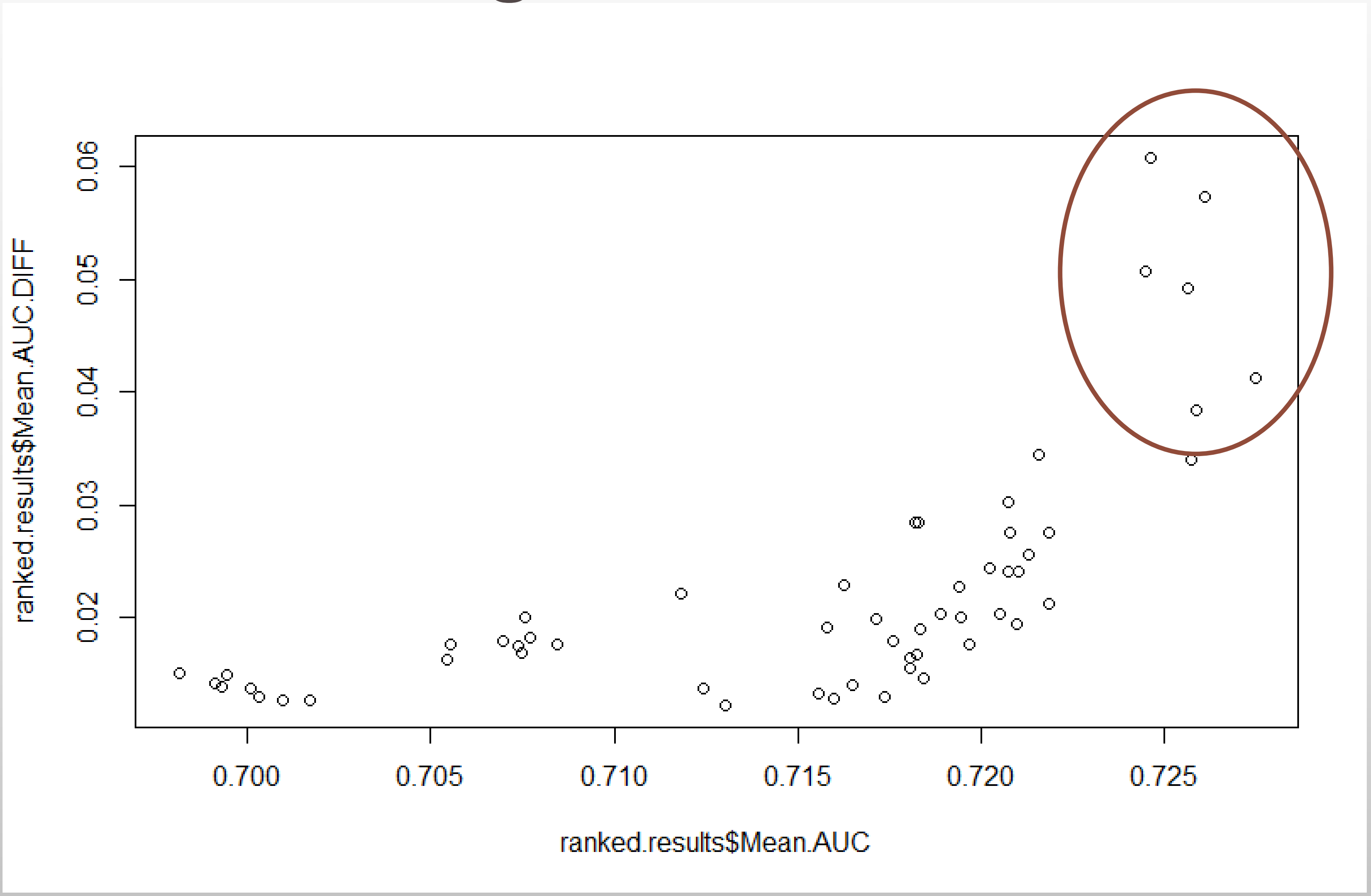
# Calibrating MaxEnt

- Two important parameters
  - Regularization Multiplier (RM)
    - Tightness of fit
  - Allowed "Feature Classes" (FCs)
    - Types of mathematical relationships explored by MaxEnt
      - Include Linear, Quadratic, Hinge, Product, and Threshold
  - Numerous recent emphasize the importance of tuning/calibrating these parameters
  - ENMeval (Muscarella et al. 2014) now makes it relatively easy to do so.
    - For each run, tried 8 FC combos X 9 RMs = 72 models

# Example ENMeval calibration outputs for monarch breeding habitat suitability



# AUC vs. AUC DIFF: Finding overfit models



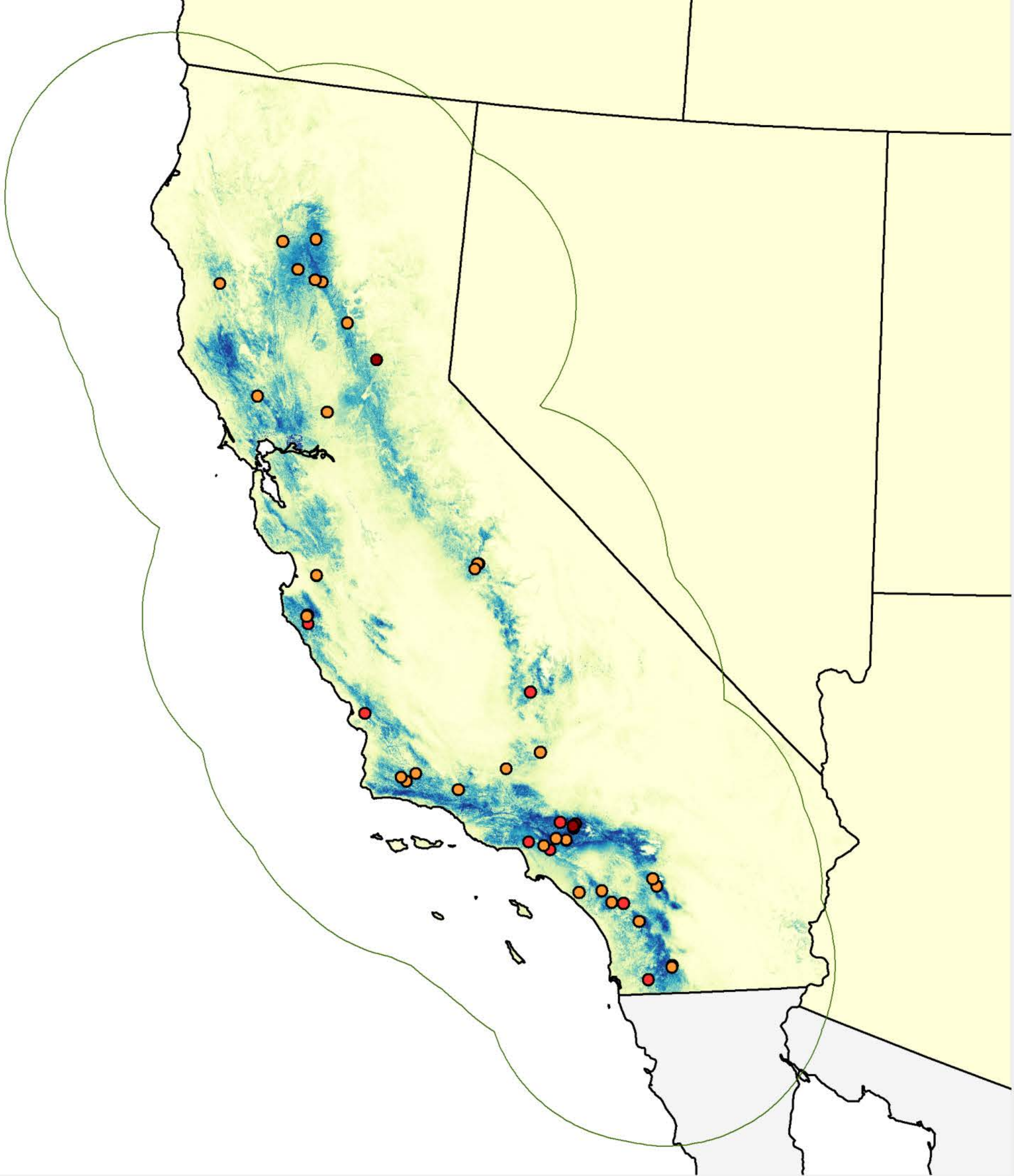
# Model selection

- No accepted standard metric.
- AUC is often used on its own, but does not adequately penalize overfitting
- We are selecting models by
  - First removing anything below a 0.7 AUC
  - Removing models that are in the bottom quadrant for any of the three overfitting metrics
  - Sorting those that remain by AUC, and inspecting the top several versions (usually very similar)

# *A. eriocarpa*

AUC = 0.71 AUC.diff=0.08, RM=1.5

water deficit annual: 43.0%
ppt annual: 26.8%
temp range: 7.3%
ppt seasonality: 6.0%
ppt warmest quarter: 4.2%
slope: 3.9%
LANDFIRE reclass: 3.1%
aspect: 2.5%
Topsoil Clay Fraction: 2.4%
distance to water: 0.7%



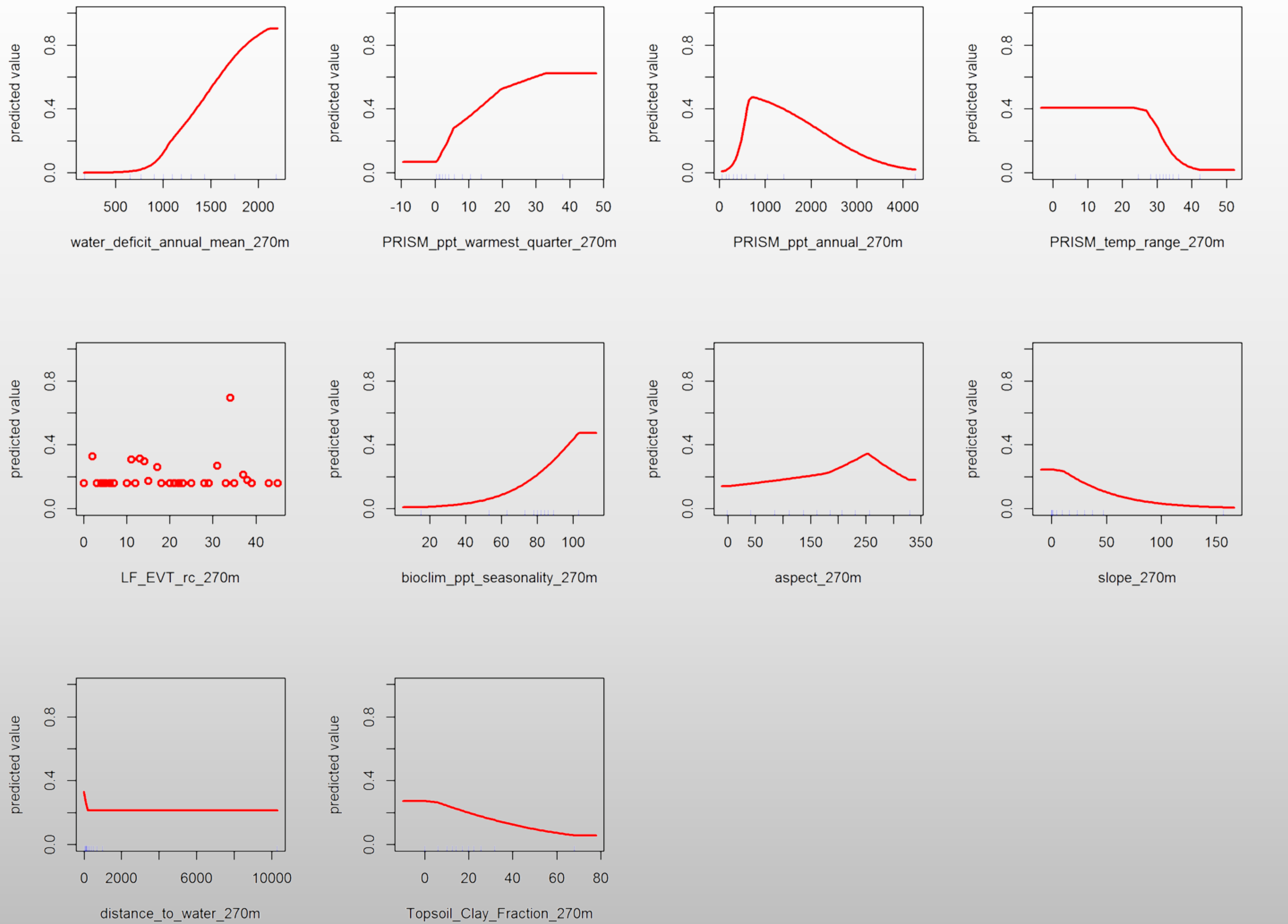
**Points by accuracy class**

- High
- Medium-High
- Medium
- Modeling extent

**Relative Habitat Suitability**

High : 0.98

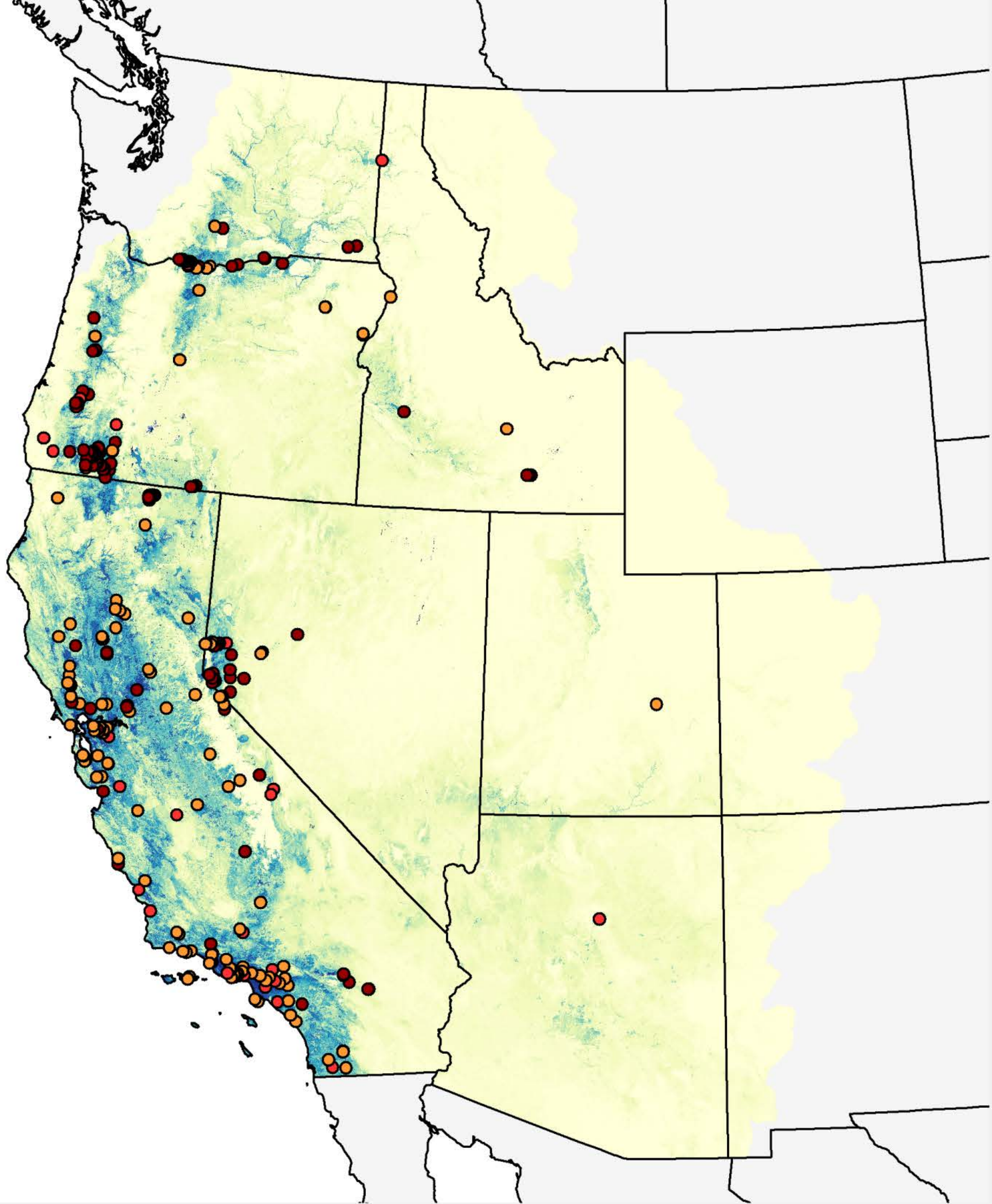
Low : 0


predicted value
0.0
0.4
0.8
500
1000
1500
2000
water_deficit_annual_mean_270m
-10
0
10
20
30
40
50
PRISM_ppt_warmest_quarter_270m
0
1000
2000
3000
4000
PRISM_ppt_annual_270m
0
10
20
30
40
50
PRISM_temp_range_270m
0
10
20
30
40
LF_EVT_rc_270m
20
40
60
80
100
bioclim_ppt_seasonality_270m
0
50
150
250
350
aspect_270m
0
50
100
150
slope_270m
0
2000
6000
10000
distance_to_water_270m
0
20
40
60
80
Topsoil_Clay_Fraction_270m

# *A. fascicularis*

AUC = 0.73 AUC.diff=0.02, RM=2.25

ppt coldest quarter: 23.7%
tmean coldest quarter: 21.5%
ppt warmest quarter: 17.8%
LANDFIRE reclass: 9.7%
Topsoil Silt Fraction: 7.8%
ppt seasonality: 6.1%
Topsoil Bulk Density: 5.5%
slope: 5.0%
distance to water: 2.8%



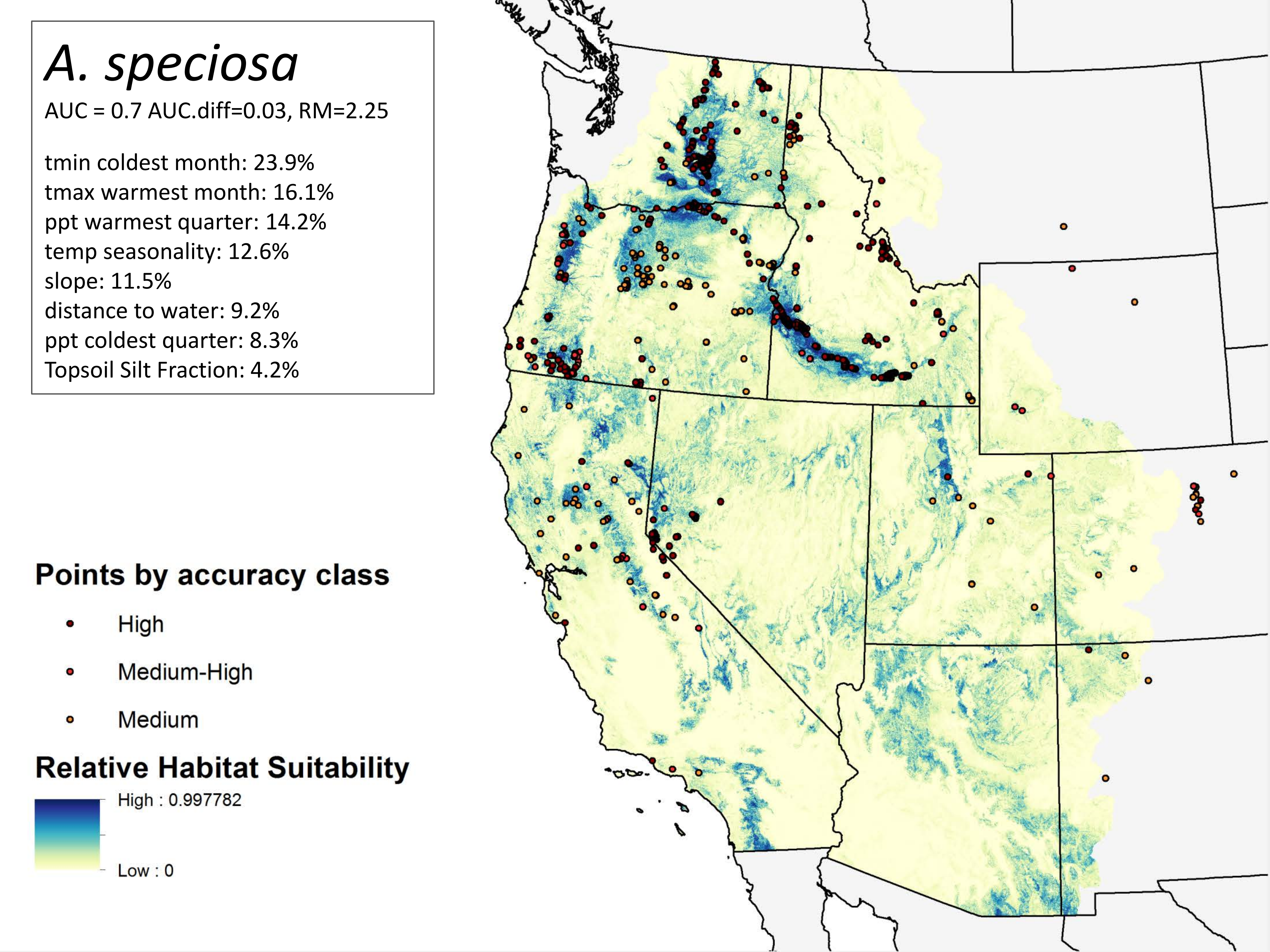
**Points by accuracy class**

- High
- Medium-High
- Medium

**Relative Habitat Suitability**

High : 0.999567

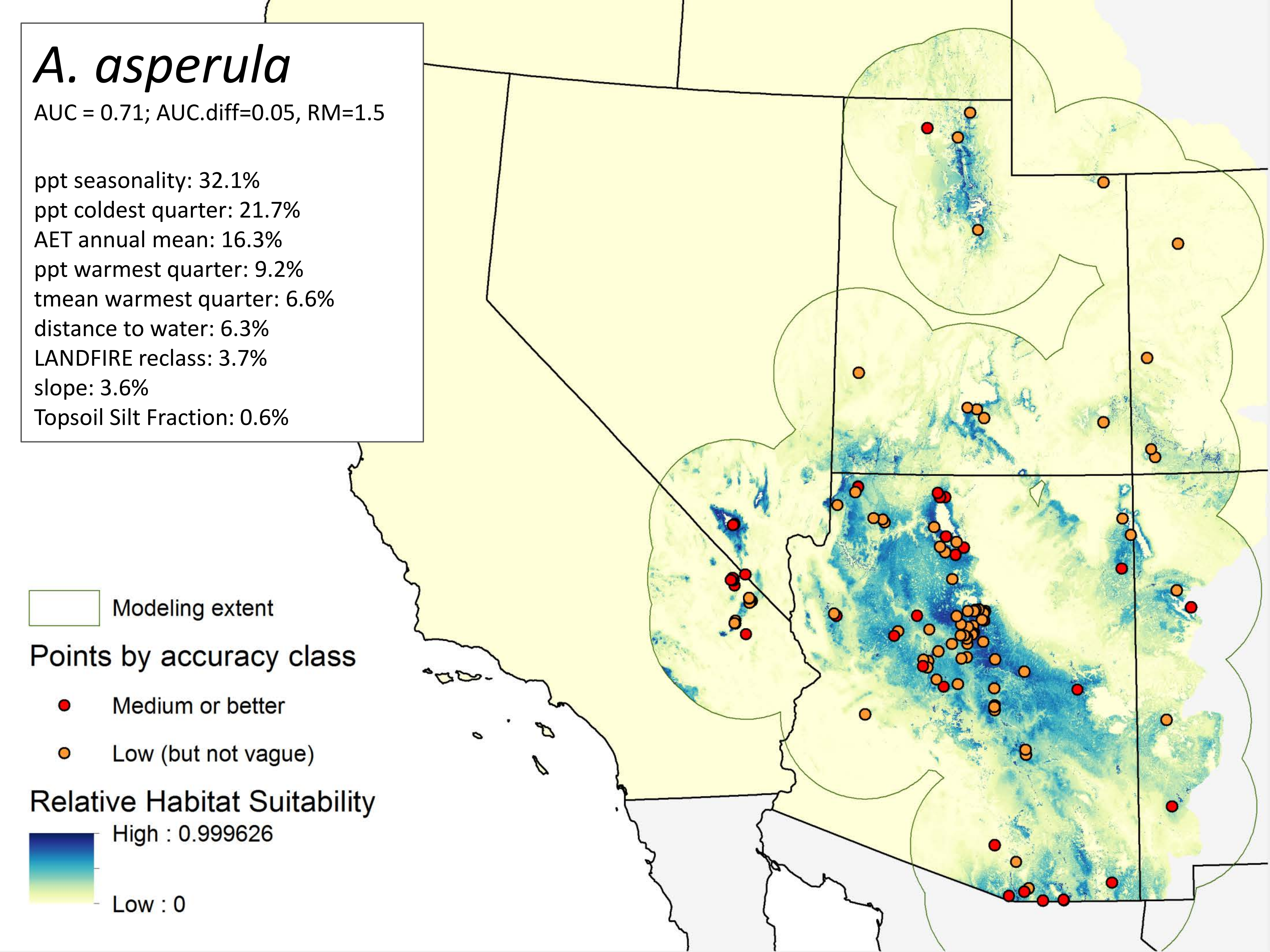
Low : 0



A. speciosa
AUC = 0.7 AUC.diff=0.03, RM=2.25
tmin coldest month: 23.9%
tmax warmest month: 16.1%
ppt warmest quarter: 14.2%
temp seasonality: 12.6%
slope: 11.5%
distance to water: 9.2%
ppt coldest quarter: 8.3%
Topsoil Silt Fraction: 4.2%
Points by accuracy class
High
Medium-High
Medium
Relative Habitat Suitability
High : 0.997782
Low : 0

# Effect of spatial inaccuracy in data on species distribution models

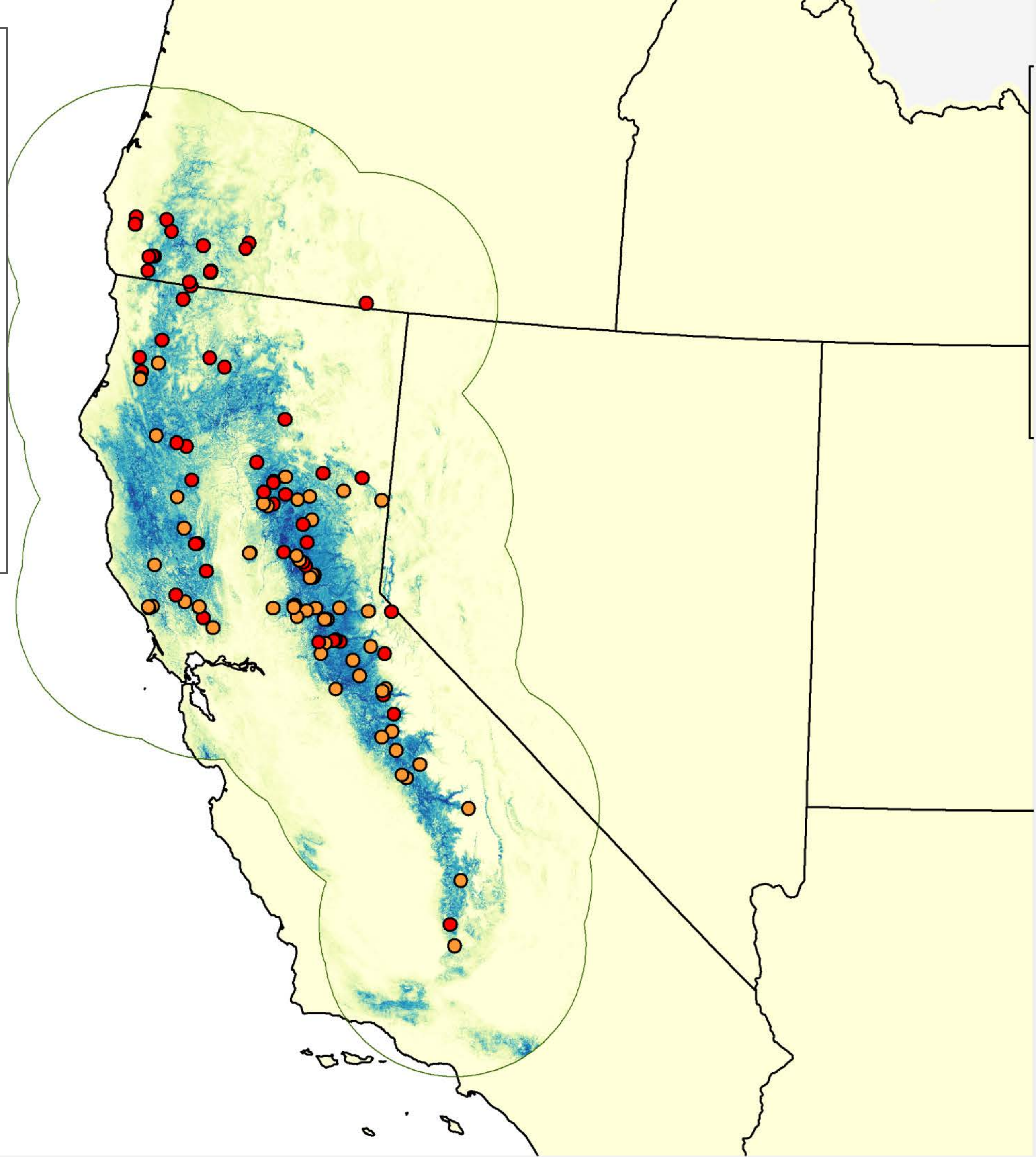
- Graham et al 2008
  - Moved each coordinate with a random number drawn from the normal distribution with a mean of zero and a standard deviation of 5 km
  - Found MaxEnt was largely unaffected by these errors
  - But, many of the errors in the milkweed and monarch database could be far greater that 5km.


A. asperula
AUC = 0.71; AUC.diff=0.05, RM=1.5
ppt seasonality: 32.1%
ppt coldest quarter: 21.7%
AET annual mean: 16.3%
ppt warmest quarter: 9.2%
tmean warmest quarter: 6.6%
distance to water: 6.3%
LANDFIRE reclass: 3.7%
slope: 3.6%
Topsoil Silt Fraction: 0.6%
Modeling extent
Points by accuracy class
Medium or better
Low (but not vague)
Relative Habitat Suitability
High : 0.999626
Low : 0

# *A. cordifolia*

AUC = 0.72; AUC.diff=0.03, RM=1.75

ppt annual: 28.2%
tmin coldest month: 15.0%
temp range: 12.9%
water deficit annual : 11.9%
ppt warmest quarter: 8.6%
slope: 6.1%
Topsoil Silt Fraction: 6.1%
LANDFIRE reclass: 5.1%
distance to water: 4.5%
aspect: 1.5%



Modeling extent

Points by accuracy class

- Medium or better
- Low (but not vague)

Relative Habitat Suitability

High : 0.975831

Low : 0

# Monarch breeding

AUC = 0.7; AUC.diff=0.06, RM=1.25

fascicularis model: 41.6%
speciosa model: 22.7%
tmean warmest quarter: 9.7%
AET annual mean: 5.5%
ppt warmest quarter: 4.1%
eriocarpa model: 3.4%
ppt seasonality: 2.9%
asperula model: 2.8%
slope: 2.5%
Topsoil Sand Fraction: 1.5%
cordifolia model: 1.3%
Topsoil Clay Fraction: 1.1%
distance to water: 0.9%

Modeling extent

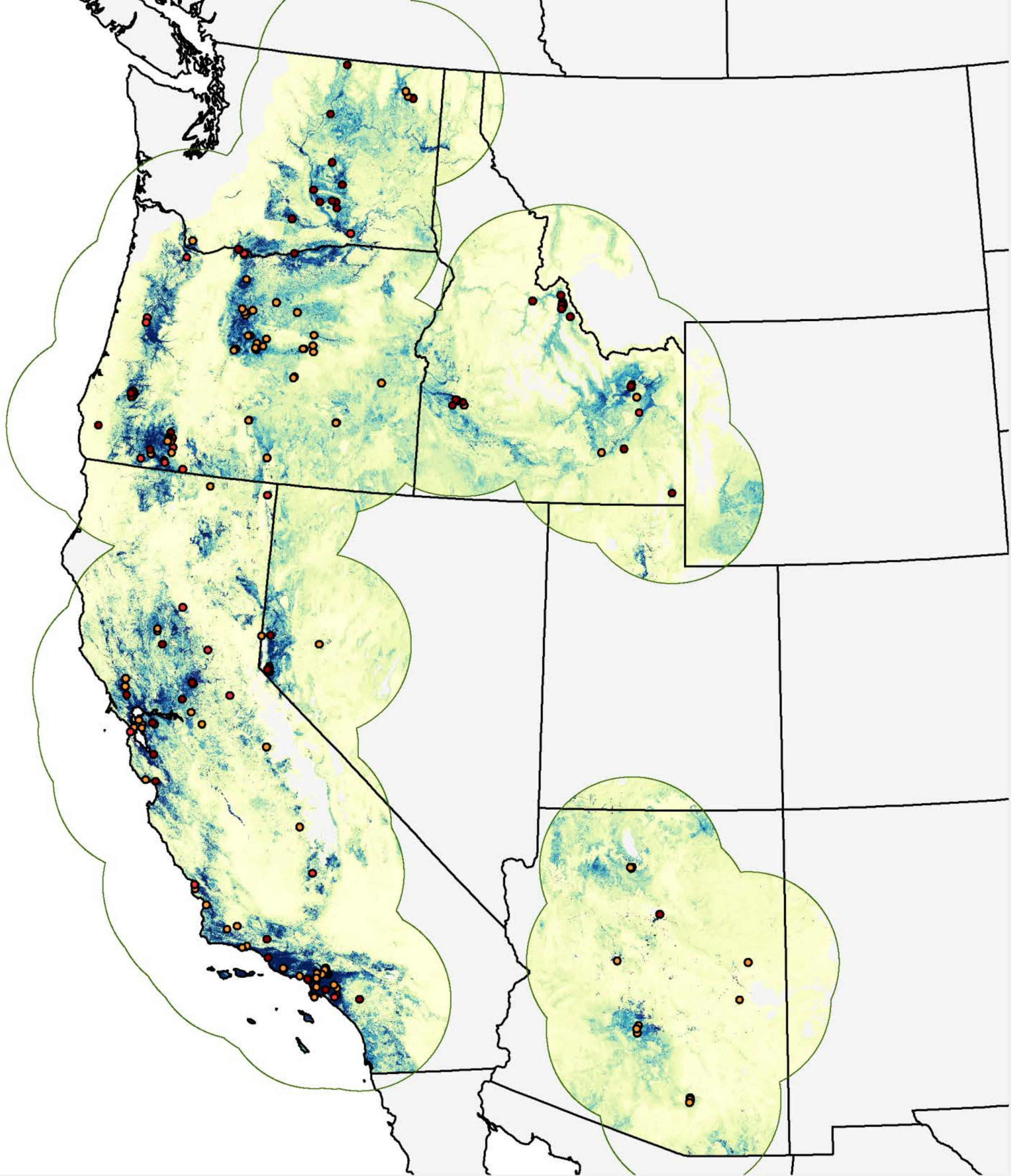
**Points by accuracy class**

- High
- Medium-High
- Medium

**Relative Habitat Suitability**

High : 1

Low : 0



# Potential improvements

- Near term
  - Can run with different variable subsets
- Long term
  - **More species data -> better models**
  - Improve variables
    - Rivers instead of all minor waterways
    - Recent climate variables instead of long-term avgs.
  - Ensemble of model types
  - Model regions separately
  - Finer models of key regions
  - Seasonal suitability models
  - Climate change modeling

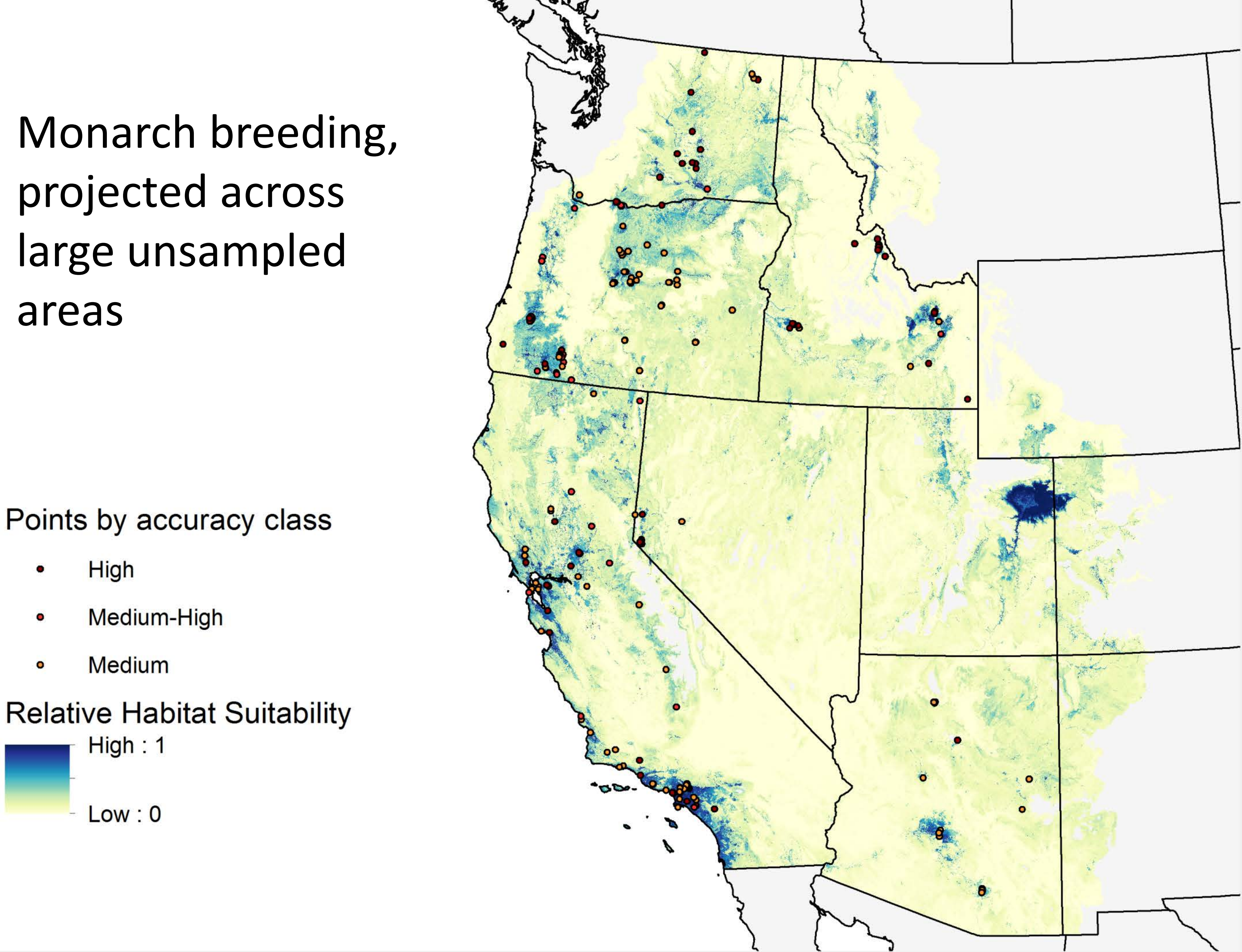
# Potential uses

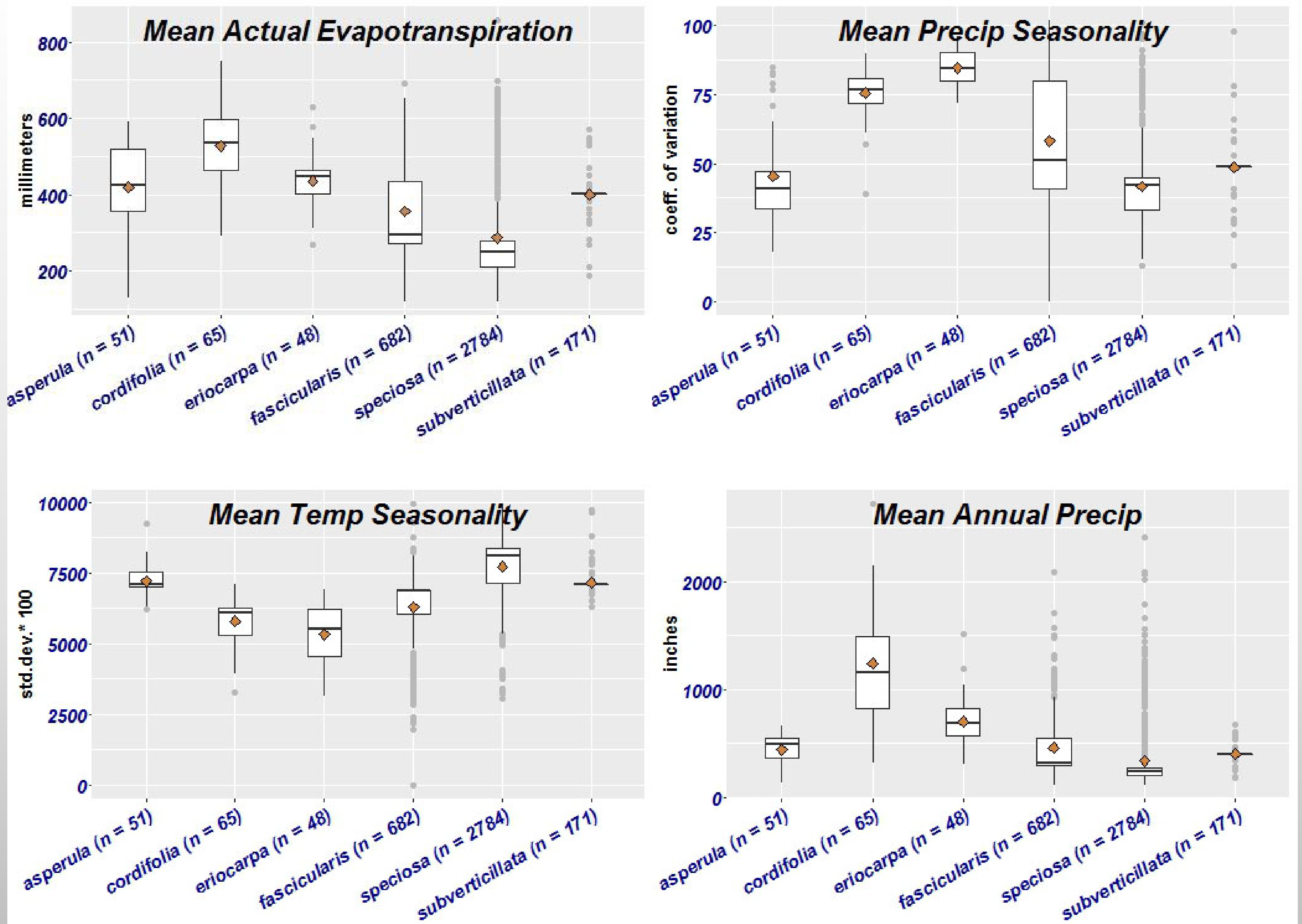
- Guide surveys
  - For example, speciosa model could be mapped with roads and public lands in NV to help inform sampling locations this summer
- Help point to regions of higher potential value for restoration
  - This sort of application probably only appropriate in CA and PNW with these models
- Need to be validated, and new runs


Thank you!

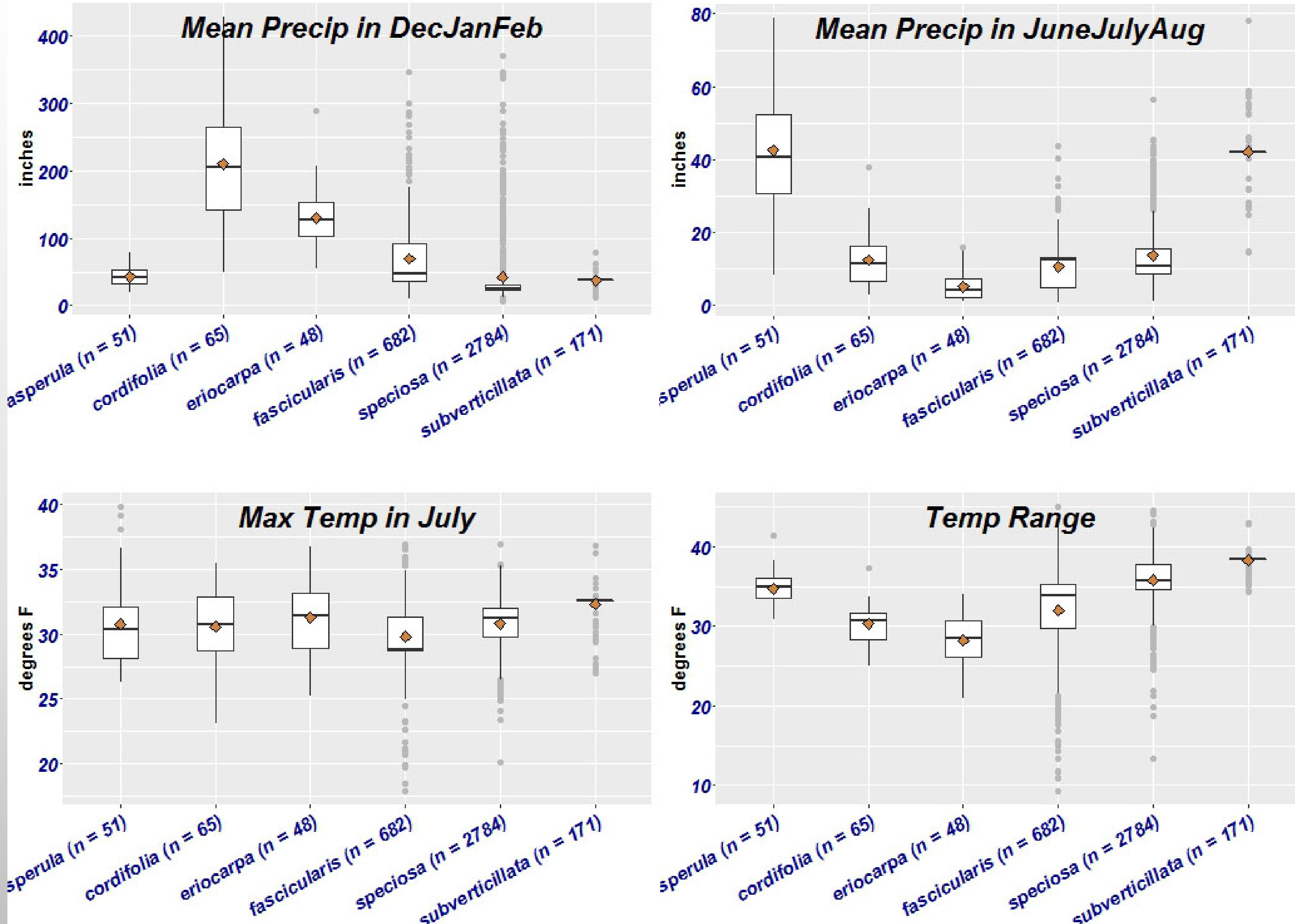
# Appendix Slides

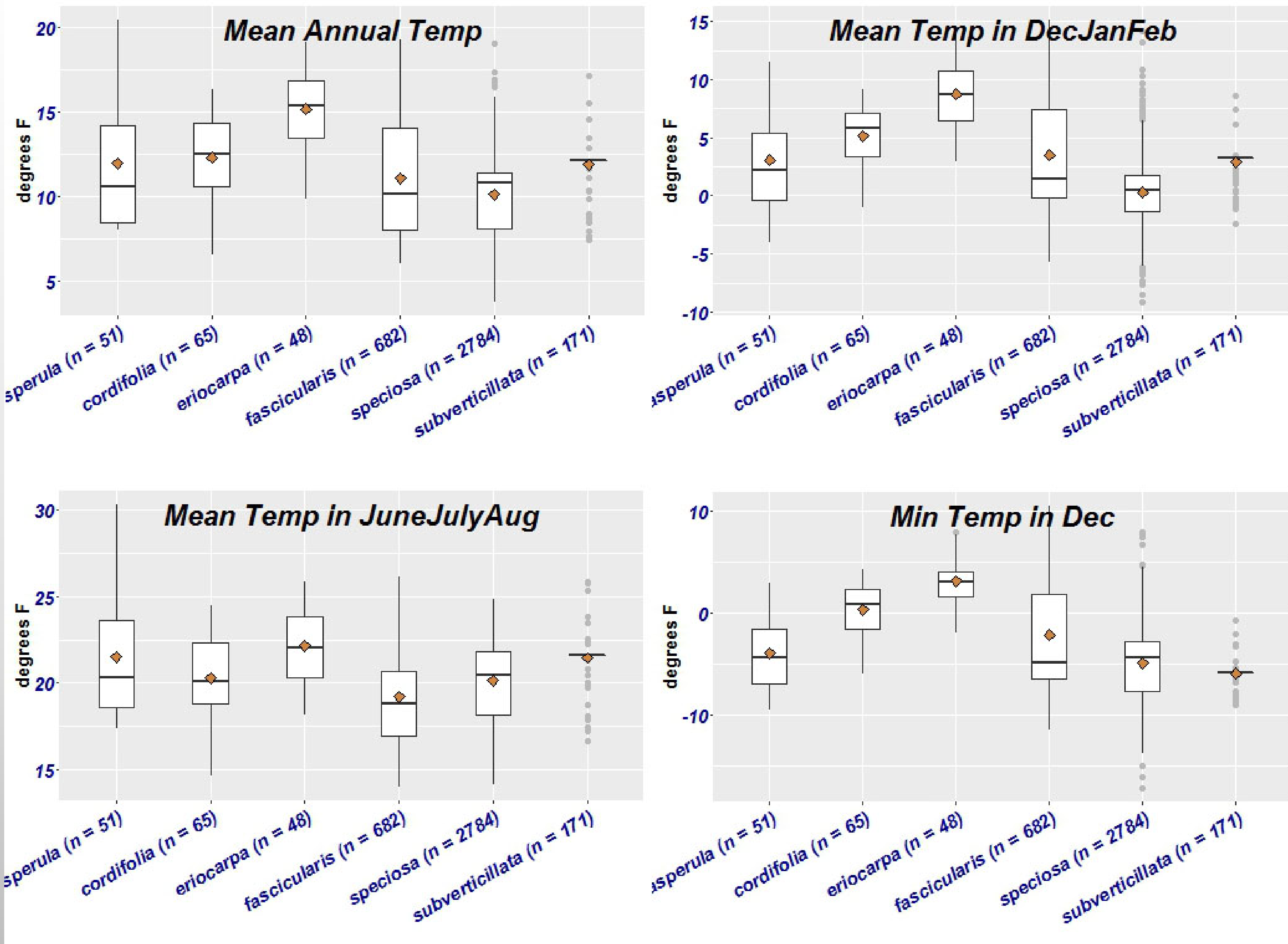
Monarch breeding, projected across large unsampled areas



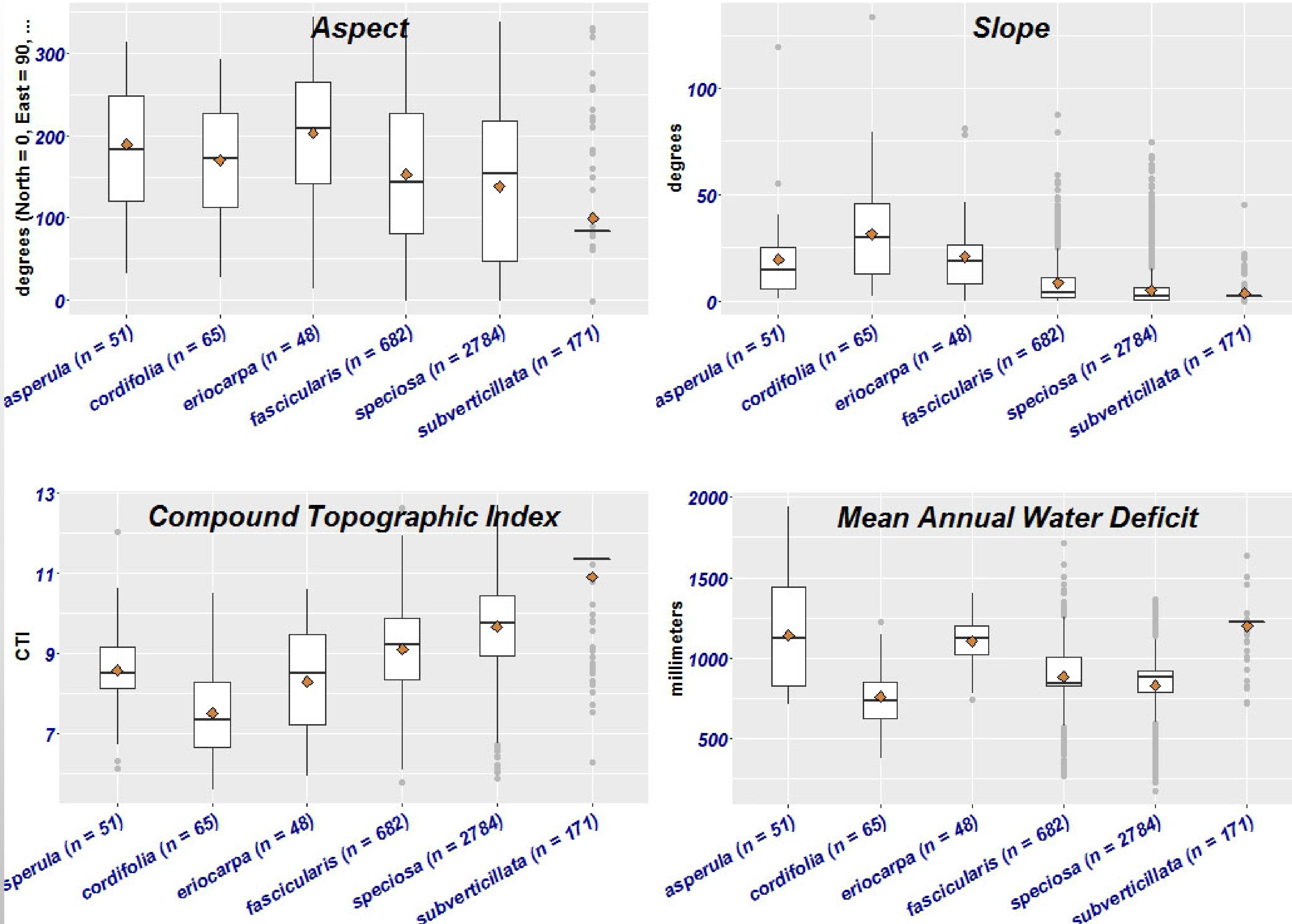


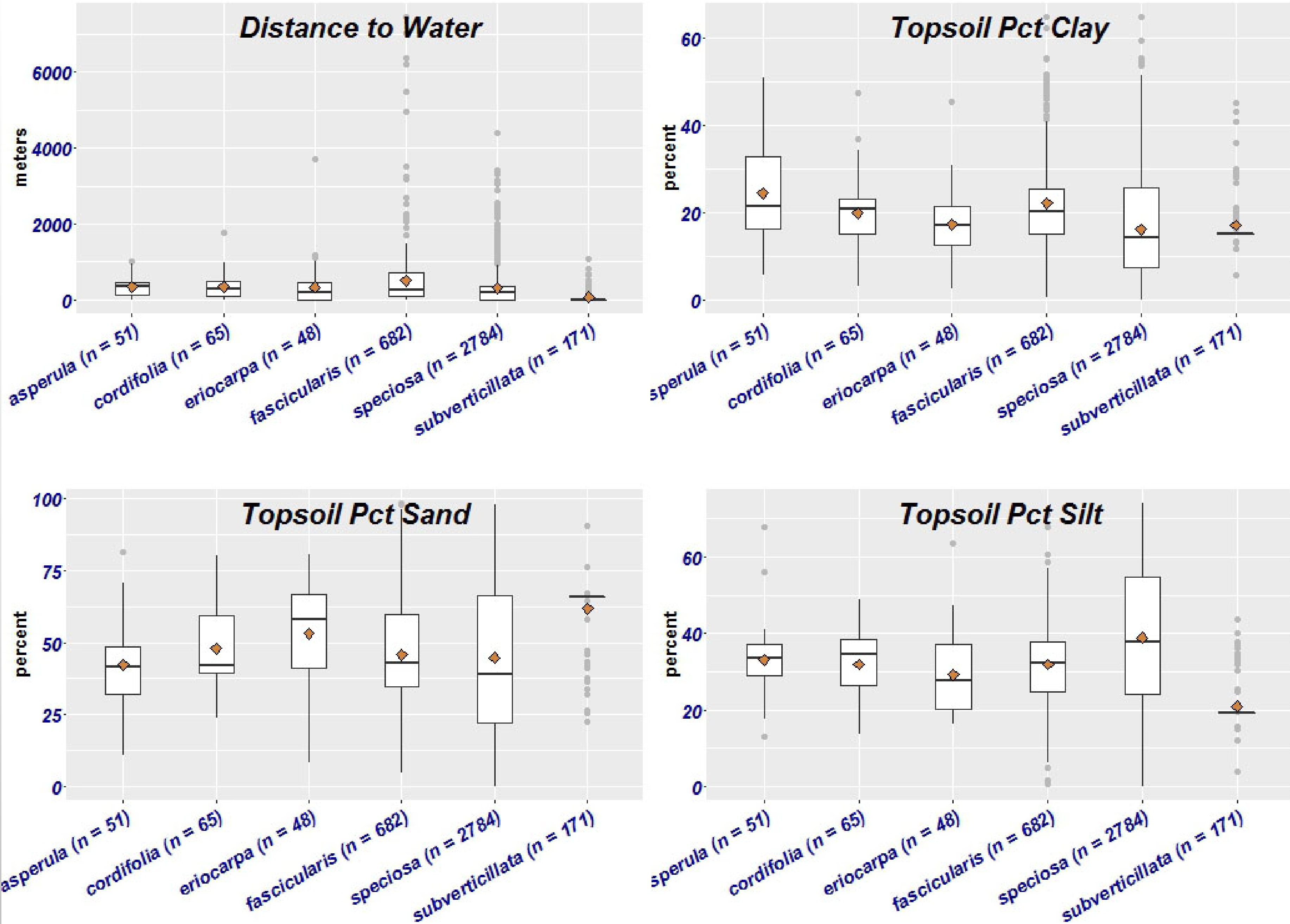
Mean Actual Evapotranspiration
millimeters
800
600
400
200
asperula (n = 51)
cordifolia (n = 65)
eriocarpa (n = 48)
fascicularis (n = 682)
speciosa (n = 2784)
subverticillata (n = 171)
Mean Precip Seasonality
coeff. of variation
100
75
50
25
0
asperula (n = 51)
cordifolia (n = 65)
eriocarpa (n = 48)
fascicularis (n = 682)
speciosa (n = 2784)
subverticillata (n = 171)
Mean Temp Seasonality
std.dev.* 100
10000
7500
5000
2500
0
asperula (n = 51)
cordifolia (n = 65)
eriocarpa (n = 48)
fascicularis (n = 682)
speciosa (n = 2784)
subverticillata (n = 171)
Mean Annual Precip
inches
2000
1000
0
asperula (n = 51)
cordifolia (n = 65)
eriocarpa (n = 48)
fascicularis (n = 682)
speciosa (n = 2784)
subverticillata (n = 171)


Mean Precip in DecJanFeb
inches
0
100
200
300
400
asperula (n = 51)
cordifolia (n = 65)
eriocarpa (n = 48)
fascicularis (n = 682)
speciosa (n = 2784)
subverticillata (n = 171)
Mean Precip in JuneJulyAug
inches
0
20
40
60
80
Max Temp in July
degrees F
20
25
30
35
40
Temp Range
degrees F
10
20
30
40

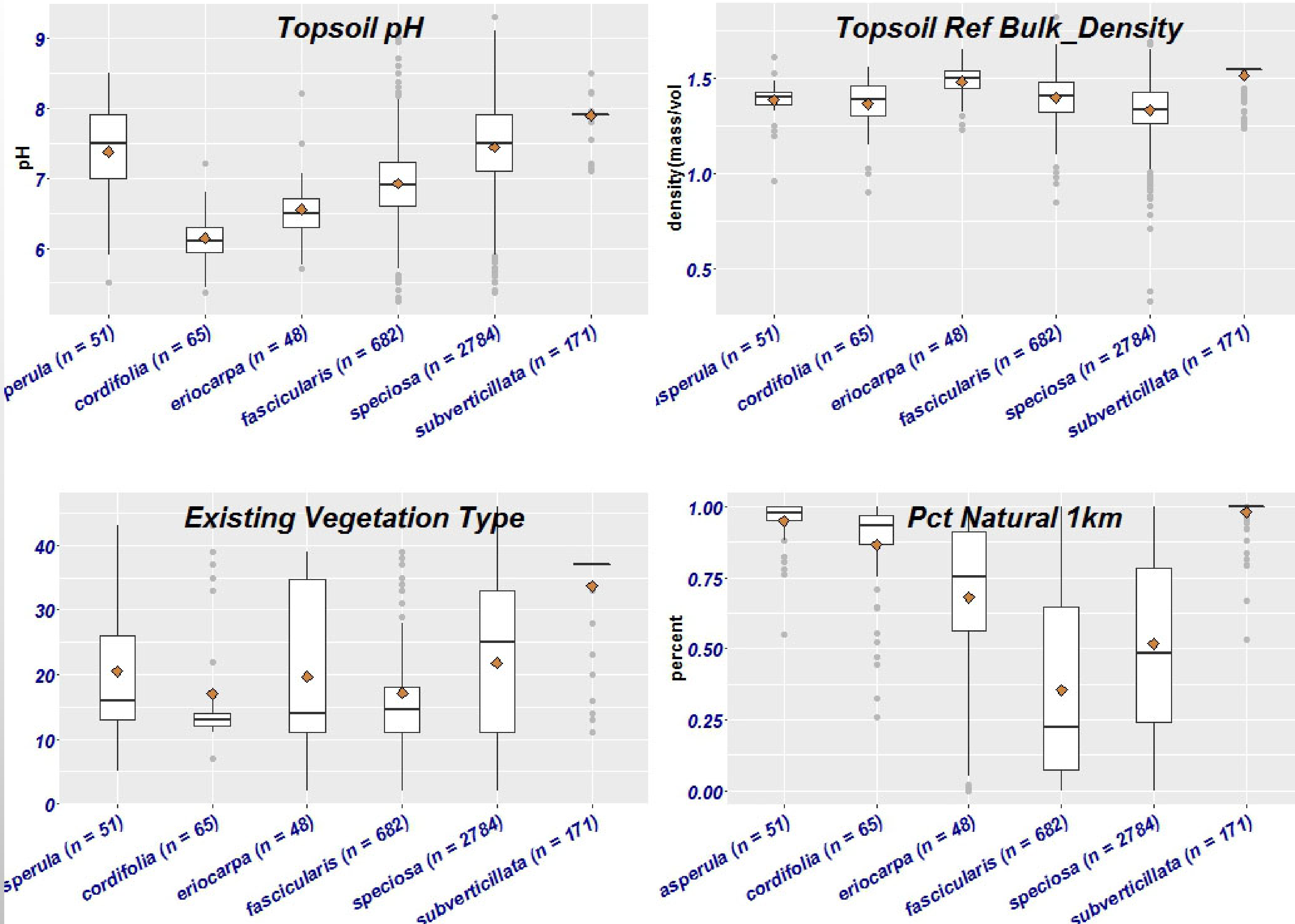


Mean Annual Temp
Mean Temp in DecJanFeb
Mean Temp in JuneJulyAug
Min Temp in Dec
degrees F
asperula (n = 51)
cordifolia (n = 65)
eriocarpa (n = 48)
fascicularis (n = 682)
speciosa (n = 2784)
subverticillata (n = 171)


Aspect
degrees (North = 0, East = 90, ...
0
100
200
300
asperula (n = 51)
cordifolia (n = 65)
eriocarpa (n = 48)
fascicularis (n = 682)
speciosa (n = 2784)
subverticillata (n = 171)
Slope
degrees
0
50
100
asperula (n = 51)
cordifolia (n = 65)
eriocarpa (n = 48)
fascicularis (n = 682)
speciosa (n = 2784)
subverticillata (n = 171)
Compound Topographic Index
CTI
7
9
11
13
sperula (n = 51)
cordifolia (n = 65)
eriocarpa (n = 48)
fascicularis (n = 682)
speciosa (n = 2784)
subverticillata (n = 171)
Mean Annual Water Deficit
millimeters
500
1000
1500
2000
asperula (n = 51)
cordifolia (n = 65)
eriocarpa (n = 48)
fascicularis (n = 682)
speciosa (n = 2784)
subverticillata (n = 171)



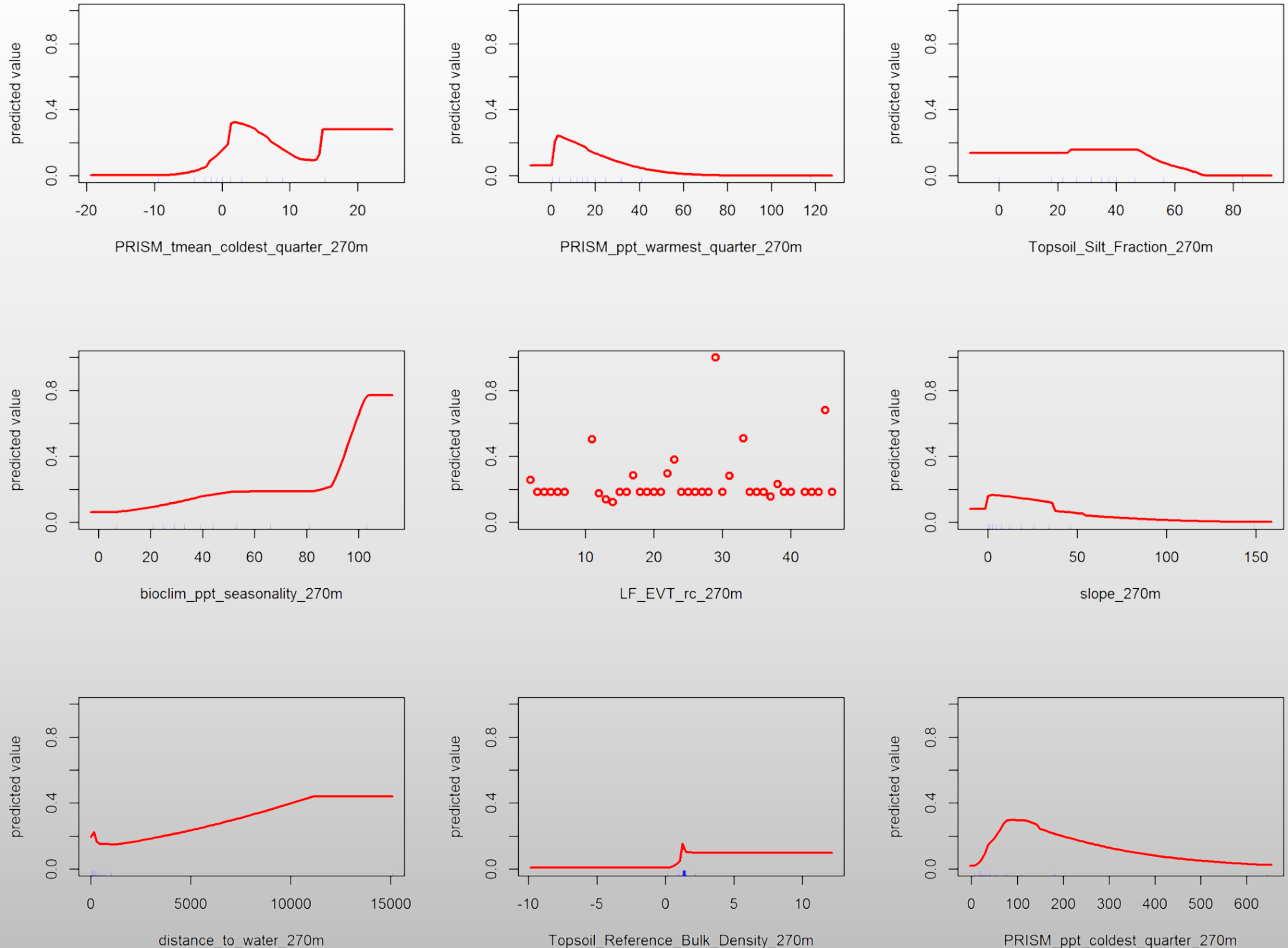
Distance to Water
meters
6000
4000
2000
0
Topsoil Pct Clay
percent
60
40
20
0
Topsoil Pct Sand
percent
100
75
50
25
0
Topsoil Pct Silt
percent
60
40
20
0
asperula (n = 51)
cordifolia (n = 65)
eriocarpa (n = 48)
fascicularis (n = 682)
speciosa (n = 2784)
subverticillata (n = 171)


Topsoil pH
pH
9
8
7
6
perula (n = 51)
cordifolia (n = 65)
eriocarpa (n = 48)
fascicularis (n = 682)
speciosa (n = 2784)
subverticillata (n = 171)
Topsoil Ref Bulk_Density
density(mass/vol
1.5
1.0
0.5
asperula (n = 51)
cordifolia (n = 65)
eriocarpa (n = 48)
fascicularis (n = 682)
speciosa (n = 2784)
subverticillata (n = 171)
Existing Vegetation Type
40
30
20
10
0
sperula (n = 51)
cordifolia (n = 65)
eriocarpa (n = 48)
fascicularis (n = 682)
speciosa (n = 2784)
subverticillata (n = 171)
Pct Natural 1km
percent
1.00
0.75
0.50
0.25
0.00
asperula (n = 51)
cordifolia (n = 65)
eriocarpa (n = 48)
fascicularis (n = 682)
speciosa (n = 2784)
subverticillata (n = 171)

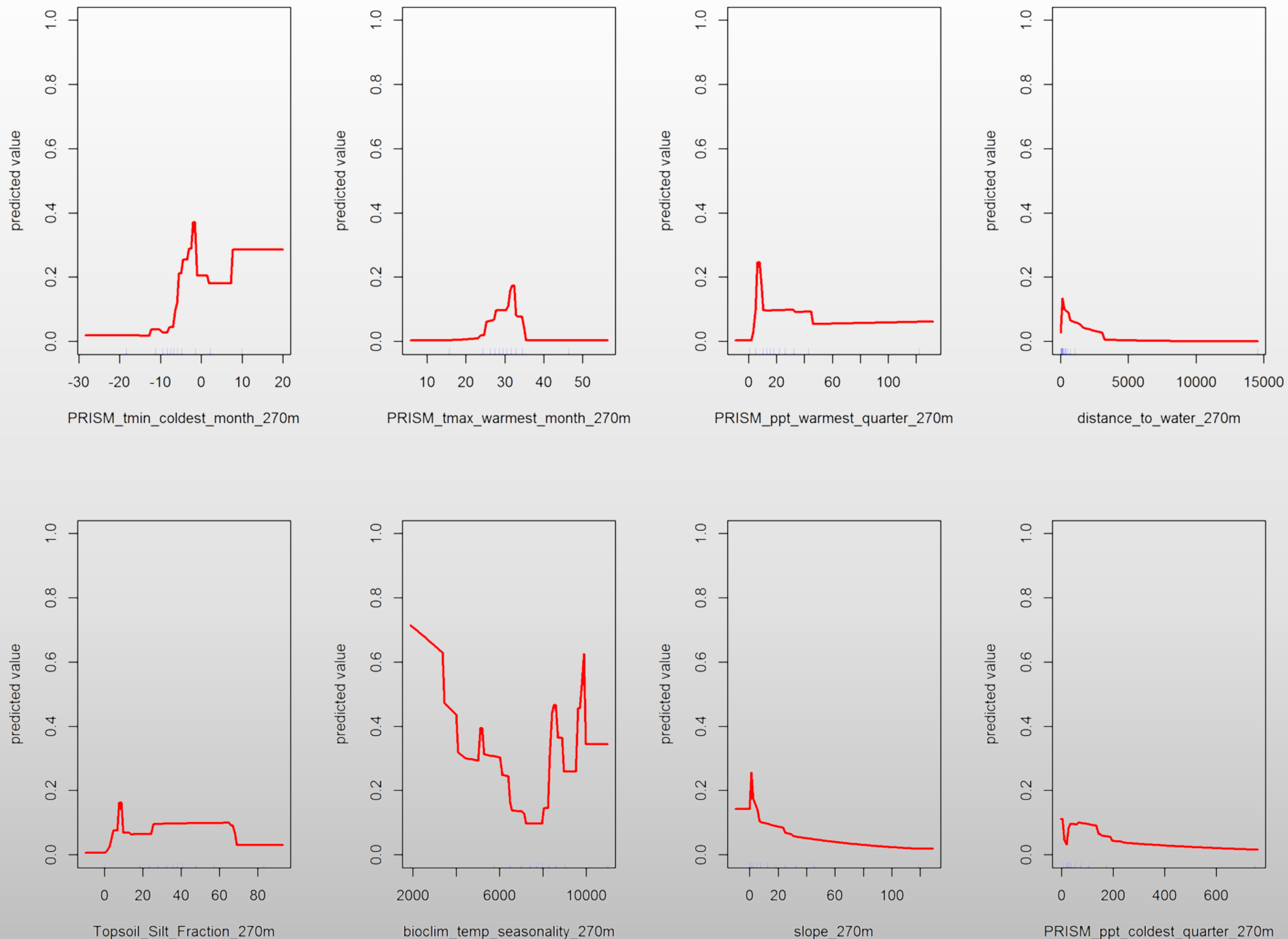
# Model evaluation metrics

| Metric | Description |
|---|---|
| AUC Test | Measures goodness of fit. Probability that a randomly-drawn presence has a *higher score than a randomly drawn background location*. Tends to be lower for generalist species. |
| AUC Diff | “The difference between the AUC value based on training localities (i.e. AUCTRAIN) and AUCTEST (AUCTRAIN  AUCTEST). If AUCTRAIN < AUCTEST, the returned value is zero. Value of AUCDIFF is expected to be positively associated with the degree of model  overfitting” (Muscarella et a. 2014) |
| ORMTP (‘Minimum Training Presence’ omission rate) | “indicates the proportion of test localities with suitability values (MAXENT relative occurrence rates) lower than that associated with the lowest-ranking training locality. Omission rates greater than the expectation of zero typically indicate model overfitting. |
| OR10 (10% training omission rate) | indicates the proportion of test localities with suitability values (MAXENT relative occurrence rates) lower than that excluding the 10% of training localities with the lowest predicted suitability. Omission rates greater than the expectation of 10% typically indicate model overfitting |

# *A. fascicularis*

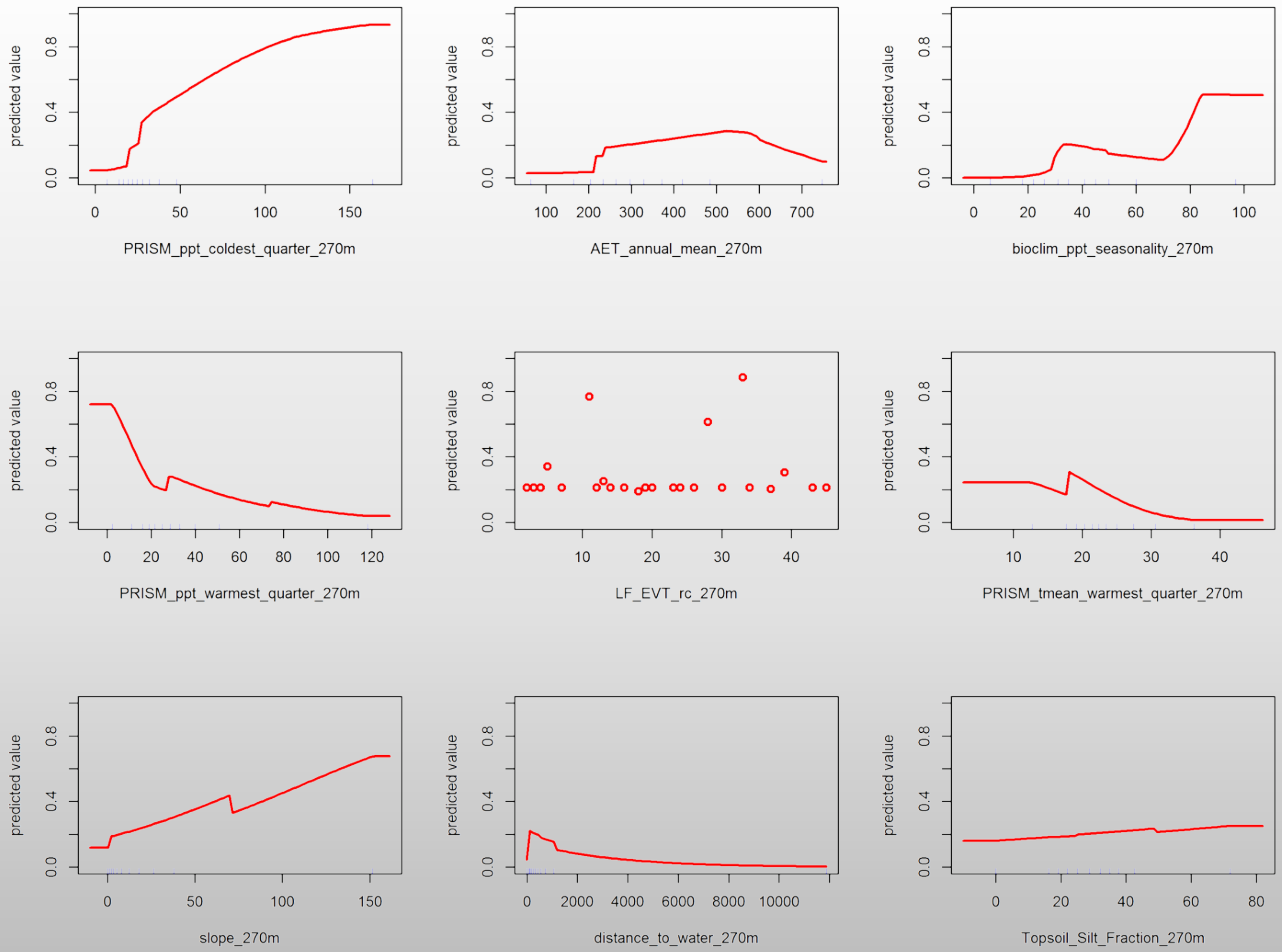


# *A. speciosa*



predicted value
0.0
0.2
0.4
0.6
0.8
1.0
-30
-20
-10
0
10
20
PRISM_tmin_coldest_month_270m
10
20
30
40
50
PRISM_tmax_warmest_month_270m
0
20
60
100
PRISM_ppt_warmest_quarter_270m
0
5000
10000
15000
distance_to_water_270m
0
20
40
60
80
Topsoil_Silt_Fraction_270m
2000
6000
10000
bioclim_temp_seasonality_270m
0
20
60
100
slope_270m
0
200
400
600
PRISM_ppt_coldest_quarter_270m

## *A. asperula*



predicted value
0.0
0.4
0.8
0
50
100
150
PRISM_ppt_coldest_quarter_270m
100
200
300
400
500
600
700
AET_annual_mean_270m
0
20
40
60
80
100
bioclim_ppt_seasonality_270m
0
20
40
60
80
100
120
PRISM_ppt_warmest_quarter_270m
10
20
30
40
LF_EVT_rc_270m
10
20
30
40
PRISM_tmean_warmest_quarter_270m
0
50
100
150
slope_270m
0
2000
4000
6000
8000
10000
distance_to_water_270m
0
20
40
60
80
Topsoil_Silt_Fraction_270m

# *A. cordifolia*