Technical Report of Data Sources and Analyses for Identifying Locations for Seed Increase Gardens in the Los Angeles Basin

Team Members:
Hannah Crispi
Jaclyn Ha
Emma Lauterbach
Maddie McKee
Amir Patel
Michelle Pham
Keely Watland

Advisor: Andrew Kleinhesselink

University of California, Los Angeles

Overview

The goal of the Seed LA practicum was to identify areas that were both ecologically suitable and technically feasible to grow native plants within the urban Los Angeles Basin. As the attached executive summary "Determining suitable locations for urban seed banks in Los Angeles" describes, our geographic analyses combine three components:

1) We used species distribution models (SMDs) to map the ecological suitability for native plants across the LA Basin. These take into account current and future climate and soil conditions.

2) We used the National Land Cover Database (NLCD) to pinpoint areas within Los Angeles that would be suitable for planting, i.e. those areas that are not currently covered in buildings, pavement or intensive development.

3) We used land-ownership data to find areas that are publicly owned or otherwise maintained for public services.

By intersecting the results from each of these analyses, we produce a detailed map of the LA Basin that shows areas that are ecologically suitable, unbuilt and publicly owned. This data product is a powerful tool for our client Seed LA to use to plan and develop urban seed gardens.

In this document, we describe the datasets and analytical tools used for our analysis in more detail so that our results are reproducible.

I. Species Distribution Modeling

A key aspect of this project is identifying areas that can act as suitable habitat for the twelve study species. In order to uncover these areas, we used species distribution modeling tools in the R statistical program (R Core Team, 2013). By matching species occurrences to the environmental factors at those locations, we can project other areas with similar conditions the species would be predicted to thrive in.

Study Area

Our study area focuses on the Los Angeles Basin which consists of parts of the San Gabriel Valley, the San Fernando Valley, the Santa Monica Mountain Range, and most of the Los Angeles River watershed from Pasadena to Long Beach (Li et al, 2019). This region is located within the California Floristic Province (CFP) which supports a wide range of ecosystems and endemic plant species (California Floristic Province, 2020). Also known as a biodiversity hotspot, the CFP is uniquely characterized by its Mediterranean climate, diversity, and intersection with the human factor (United States Forest Service, 2018). Although a small fraction of the CFP, the Los Angeles Basin is populated by nearly 10 million people, a factor that has heightened the competition for land between people and plants (Census Bureau QuickFacts, 2020).

Occurrence Data

We acquired geo-referenced occurrence data in the form of .csv files for the twelve target species from the CalFlora database (CalFlora - Search for Plants, 2020). In R, we completed the required preparatory steps to run the data in the model. These steps included: filtering the occurrence data for only California, setting the projection, and cleaning the data of erroneous points. This process can be found at the beginning of "fit_SDM.Rmd."

Species	Common Name
Acmispon glaber	Deerweed
Artemisia californica	California Sagebrush
Diplacus longiflorus	Southern Bush Monkeyflower
Eriogonum fasciculatum var. foliolosum	Leafy California Buckwheat
Frangula californica	Coffee Berry
Hesperoyucca whipplei	Chaparral Yucca
Heteromeles arbutifolia	Toyon
Mimulus cardinalis	Scarlet Monkeyflower
Ribes aureum var. gracillimum	Golden Currant
Salvia mellifera	Black Sage
Stipa lepida	Foothill Needlegrass
Stipa pulchra	Purple Needlegrass

Table 1: Target Species

Current Climate Data

We utilized WorldClim's bioclim rasters for 30 year average climate data at a resolution of 30 arc-seconds (WorldClim, 2020). This was obtained directly through R using the "getdata" function from the raster package (Hijmans, 2020). Of the 19 bioclimatic variables, we chose 9 (table 2) that seem to minimize correlation and make the model fit best for our study area (Williams et. al, 2009).

Future Climate Data

To predict how species suitability would change in the future due to climate change, we downloaded spatial climate projections for five general circulation models (GCM) included in the Climate Model Intercomparison Project (CMIP5). The GCM's we chose were the CCSM4, CNRM-CM5, HadGem2-ES, IPSL-CM5A-LR, and MIROC-ESM. Projections from these GCMs were chosen because they represent a variety of projections for precipitation and temperature change in Southern California and have been used in other SDM modeling (Riordan et. al, 2018). Although there is a CMIP6 in the works, fine-scale data is not yet available. As a result, we chose to use the 30-second data from CMIP5 to match the current distribution model resolution. The models chosen were for 50 years in the future (2070) and the representative concentration pathway 8.5, which represents a "business-as-usual" projection continuing with high emissions. This pathway is considered to be a worst case scenario for climate change, so our SDM projections based on this scenario should be taken to represent the most extreme change in local conditions. We downloaded bioclim rasters for each of the five GCMs from the worldclim.org website. Each bioclimatic variable comes as an individual raster layer, so in order to mimic the format of the current climate data, the "Make_CA_future_bioclim.Rmd" stacks the bioclim variables together and crops the resulting raster stack to California's extent.

Bioclimatic Variable	Description
Bio1	Annual Mean Temperature
Bio2	Mean Diurnal Range
Bio3	Isothermality
Bio4	Temperature Seasonality
Bio7	Temperature annual range
Bio12	Annual precipitation
Bio13	Precipitation of wettest month
Bio14	Precipitation of driest month
Bio15	Precipitation seasonality

Table 2: Bioclimatic variable key

Soil Data

In addition to the climate variables, we utilized soil data sourced from the USDA Web Soil Survey (Web Soil Survey - Home, 2020). We downloaded STATSGO data for the entire state of California. We then calculated average soil sand, clay and available water capacity in the upper 1 m of the soil for each soil survey map unit ("MUKEY"). These average soil properties were then converted to rasters with the same spatial extent, projection and grid as the climate variables described above. We subsequently used these average soil properties along with climate in the distribution models. See the "rasterize_soil.R and state_soil.Rmd for the full process.

Preparing the Occurrence Data

Unlike other models, the Maxent modeling system does not require pseudo-absences. Because the CalFlora data used only included occurrences, areas of the species' presences and absences had to be distinguished. In order to make this distinction, we first extracted the coordinates for each species' occurrence. We then generated "background" data that randomly sampled 2000 grid cells in the 'CA_bioclim' raster consisting of the occurrence points and current climate data. By giving cells a value of 0 when no presence was detected and a value of 1 with species presence, a presence absence data frame was built that kept cells with no species presence as "background" points. Climate data for both the present and background points were extracted and plotted on a map of California. After this, we split the occurrences into separate training and testing datasets using a 75% training to 25% testing split. We fit the model to the training data and used the testing data to evaluate model performance after the model had been fitted. With all the preparation complete, we were able to run the Maxent predict function for all 12 species using the training data points for both presence and absence. (Species distribution modeling - R Spatial, 2020)

Fitting the Maxent Model

For each of the twelve target species, we ran SDMs using the statistical modeling software Maxent in the statistical program **R** (Hijmans et al., 2017). Maxent was chosen as it performed the best when compared to traditionally used models and had flexibility in which variables were incorporated (Williams et al., 2009). After fitting the models to the training data sets we projected suitability for the held out testing dataset. The training data served as the source for the model-generated predictive cell values and the testing data evaluated the model performance (Williams et al., 2009). We ran the models for each species under both current and future climate conditions. We evaluated model performance by examining Area Under the Curve (AUC) scores for the projections. An AUC score of 1 signifies a perfect ability to predict the occurrences in the testing dataset, whereas an AUC score of 0.5 would be expected if the model were no better than randomly predicting occurrences (William et al., 2009). We used the "evaluate" function in the dismo package to calculate and plot the AUC scores (Hijmans et al., 2017).

Modeling Future Climate Suitability

To model future climate suitability, we generated projections from the fitted maxent models for each of the five GCM outputs described above (see future climate data section). This resulted in five habitat suitability projections per species for Los Angeles County (all used the same soil data). We then averaged the suitability scores from each model projection for each raster cell in Los Angeles County. This ensemble average takes into account variation among GCM's in their projections for the future climate of Los Angeles.

II. Geographic Analysis

The goal of this geographic analysis was to determine the location of vacant, public land within Los Angeles County that would be suitable for planting native seed gardens. We refined the focus of our analysis to the Los Angeles Basin using a boundary previously described by Li et al. (2019). Since Los Angeles County has numerous ecosystems and geographies over a large area, this helped narrow our project's scope to specific urban areas that would benefit from these gardens and were most feasible for obtaining permissions for planting.

Land Ownership

We included datasets in our analysis that encompassed public land ownership in order to locate potential parcels for the twelve target species. Our primary data sources for land ownership included the Los Angeles County GIS Portal, the Los Angeles GeoHub, the California Department of Transportation, and the Los Angeles City Controller website. We then gathered data on areas of interest for Seed LA including public parks, open space, major freeways, major railroads, Los Angeles Neighborhood Councils, and Los Angeles Department of Water and Power owned parcels (Los Angeles County GIS Data Portal, 2020; LA GeoHub, 2020; Caltrans GIS Data, 2020; Los Angeles City Controller Ron Galperin, 2020).

We exported the necessary datasets into an .mxd file on ArcGIS and clipped them to the Los Angeles Basin shapefile from Li et al. (2019). Since railways were located on top of land parcels, we represented those parcels in our geographic analysis. Freeways did not overlap with land parcels, so a 250-foot buffer was created around each major freeway to represent the average dimensions of a freeway, its median, and non-developed green space on either side of the thoroughfare. Once all of this data was displayed, we overlaid the parcels with boundaries delineating the Los Angeles Neighborhood Councils in order to represent local governance.

Land Use

After accumulating data on our areas of interest, we filtered out parcels with unsuitable land cover using data from the National Land Cover Database (Multi-Resolution Land Characteristics Consortium, 2016). We filtered the land cover to five classifications that encompassed non-developed to low intensity developed land (table 4).

Classification	Description
Cultivated Crops	Areas that are actively tilled or used to produce annual crops
Grassland	Areas dominated by gramaroid or herbaceous vegetation at levels above 80% of total vegetation
Shrubland	Areas dominated by shrubs less than 5 meters tall at levels above 20% of total vegetation
Developed, Open Space	Areas with constructed materials but mostly vegetation, with impervious surfaces accounting for <20% of total cover
Low Intensity Development	Areas with both constructed materials and vegetations, with impervious surfaces accounting for 20-49% of total cover

Table 3. National Land Cover Database Classifications

Suitability Mapping

Finally, we overlaid the filtered parcels with the SDMs. Each raster was clipped to the LA Basin, resampled to a smaller pixel size to better fit the shape of small parcels, and reclassified to demonstrate areas above and below average suitability. Average suitability was calculated using zonal statistics. Our final suite of maps included:

- Map of areas of interest in the Los Angeles Basin
- Map of land cover in the Los Angeles Basin
- Map for each species demonstrating current suitability projection for 2020 and current observation data
- Map for each species demonstrating projected suitability for 2070

III. Results

Species Distribution Modeling

The table below lists the Area Under the Curve (AUC) scores produced from the Maxent modeling for each of the target species. To reiterate from the “Fitting the Maxent Model” section, the closer an AUC score is to 1, the better the model is performing.

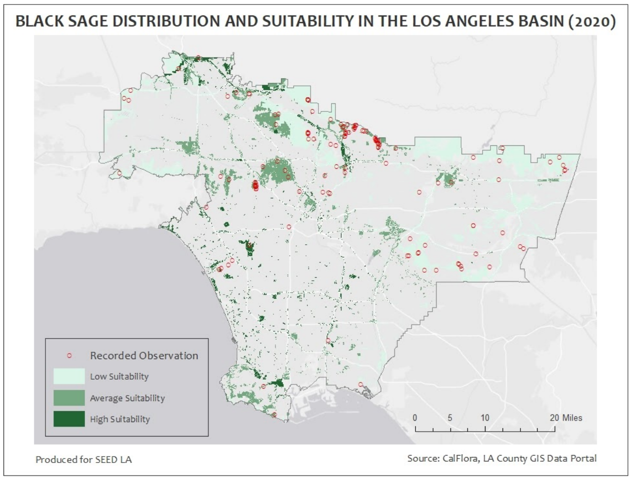
Species (Common Name)	Out of Sample AUC Score
Deerweed	0.905
California Sagebrush	0.949
Southern Bush Monkeyflower	0.952
Leafy California Buckwheat	0.955
Coffee Berry	0.914
Chaparral Yucca	0.918
Toyon	0.929
Scarlet Monkeyflower	0.863
Golden Currant	0.980
Black Sage	0.953
Foothill Needlegrass	0.949
Purple Needlegrass	0.945

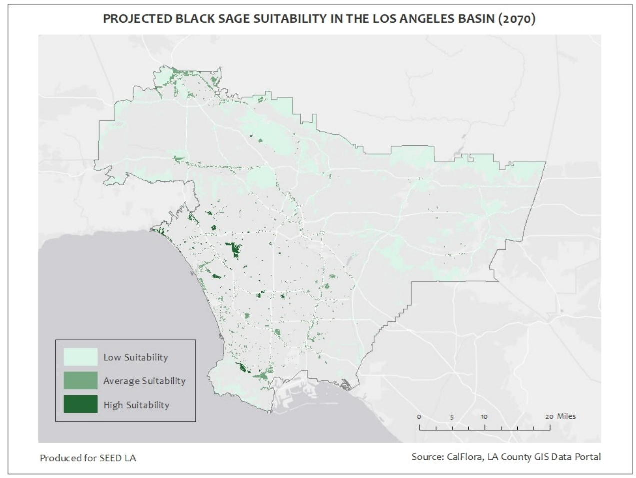
Table 4. Species Distribution Model performance.

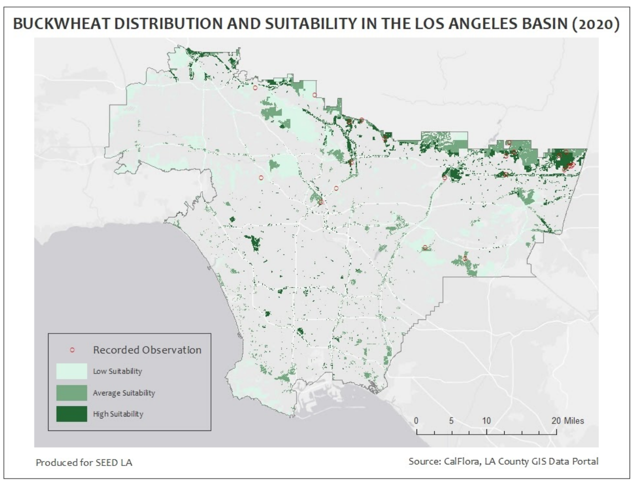
Geographic Analysis

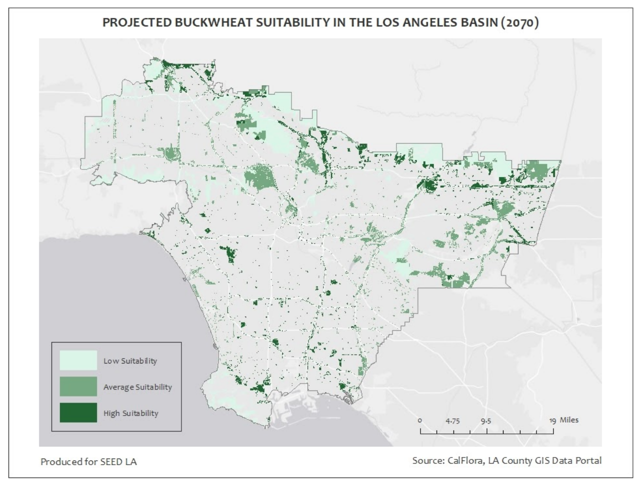
On the following pages are examples of the final suitability maps for each of the target species. The maps on the left illustrate the current suitability projections for the year 2020 as well as the occurrence points sourced from CalFlora (CalFlora - Search for Plants, 2020) and on the right are future suitability projections for the year 2070 under the rcp 8.5, “business as usual” emissions scenario (WorldClim, 2020) . The suitability was given scores ranging from 0 (least suitable) to 1 (most suitable) which are indicated by the various shades of green. The scale for both years is kept the same to highlight the change in suitability over time. The darker green indicates areas of the highest suitability in which to grow the species while the lighter, mint color indicates areas of low suitability. For the maps of 2020, the red circles indicate species occurrences.

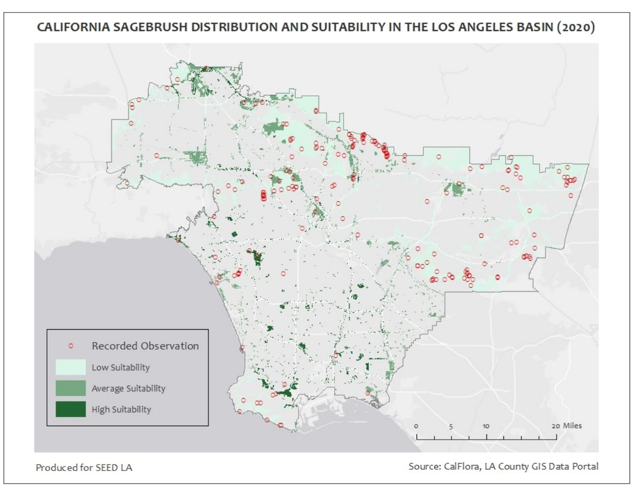
Fig. 1: Final Suitability Maps (2020 and 2070)

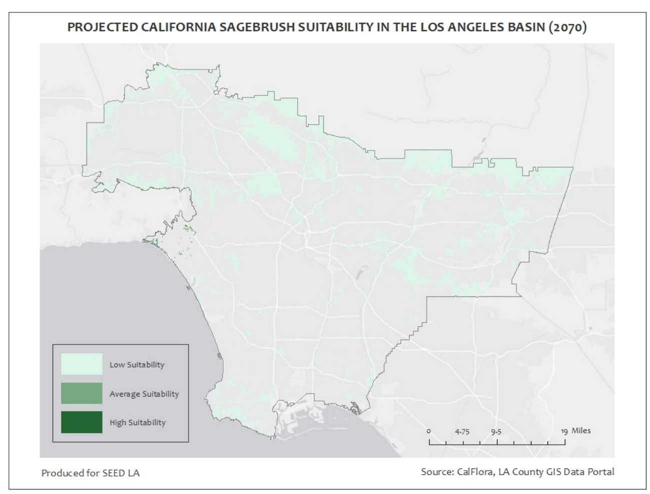


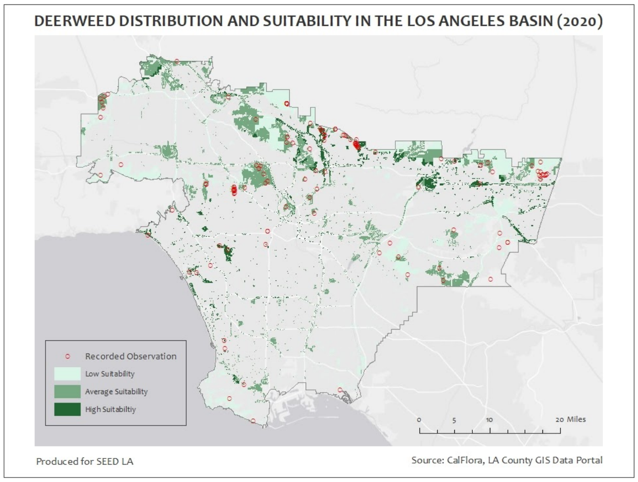


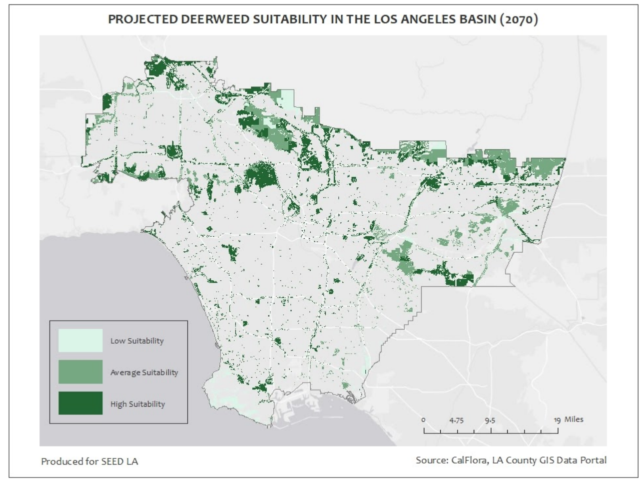


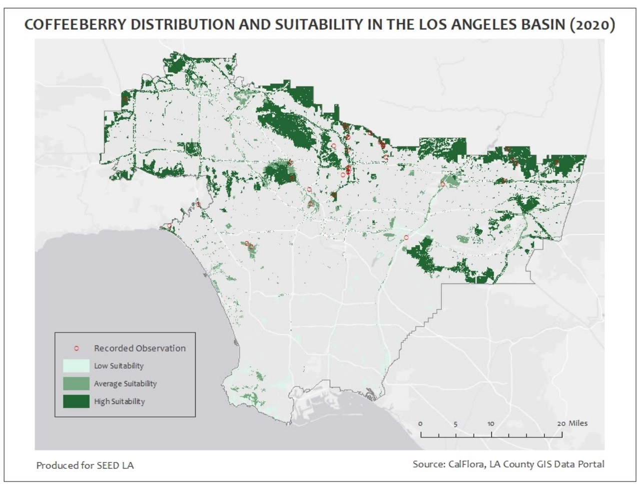


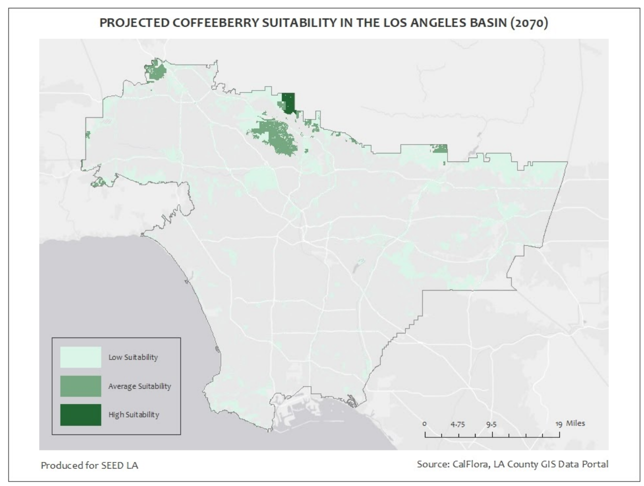


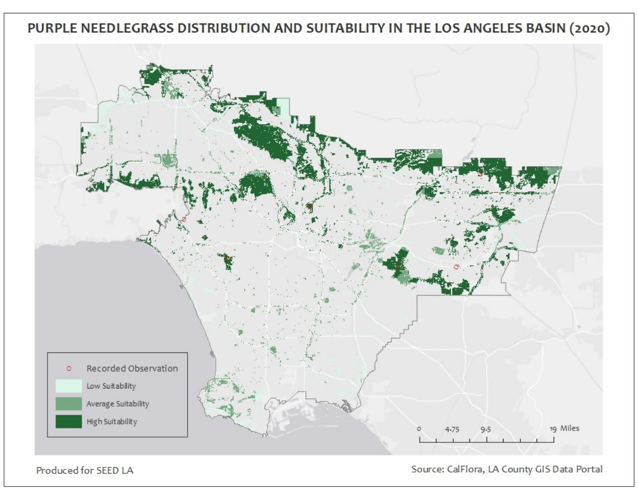


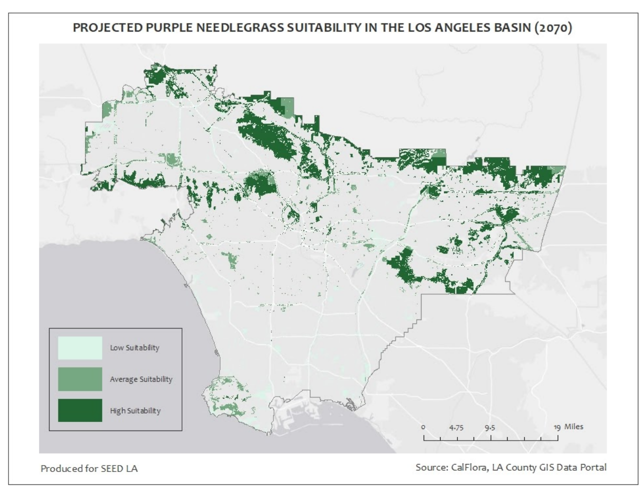

DEERWEED DISTRIBUTION AND SUITABILITY IN THE LOS ANGELES BASIN (2020)
Recorded Observation
Low Suitability
Average Suitability
High Suitability
0 5 10 20 Miles
Produced for SEED LA
Source: CalFlora, LA County GIS Data Portal


PROJECTED DEERWEED SUITABILITY IN THE LOS ANGELES BASIN (2070)
Low Suitability
Average Suitability
High Suitability
0 4.75 9.5 19 Miles
Produced for SEED LA
Source: CalFlora, LA County GIS Data Portal

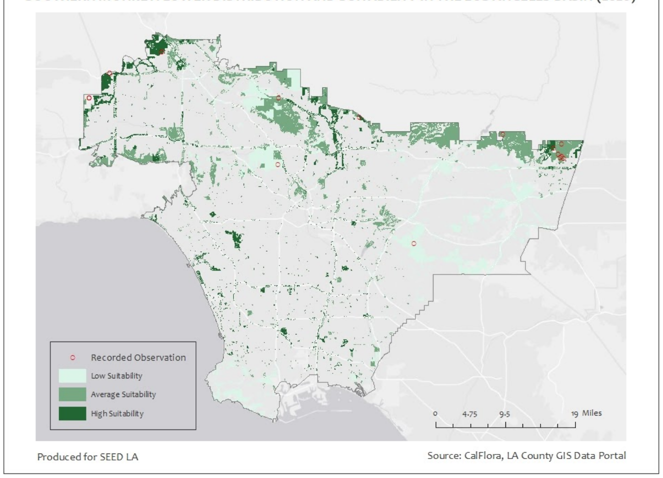

COFFEEBERRY DISTRIBUTION AND SUITABILITY IN THE LOS ANGELES BASIN (2020)
Recorded Observation
Low Suitability
Average Suitability
High Suitability
0 5 10 20 Miles
Produced for SEED LA
Source: CalFlora, LA County GIS Data Portal


PROJECTED COFFEEBERRY SUITABILITY IN THE LOS ANGELES BASIN (2070)
Low Suitability
Average Suitability
High Suitability
0 4.75 9.5 19 Miles
Produced for SEED LA
Source: CalFlora, LA County GIS Data Portal

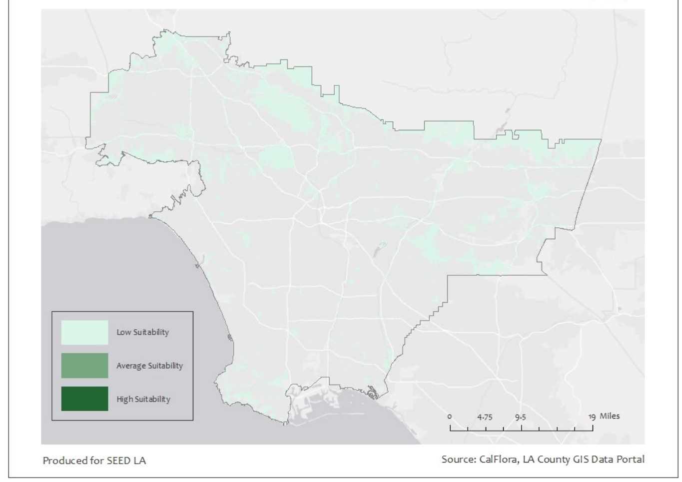

PURPLE NEEDLEGRASS DISTRIBUTION AND SUITABILITY IN THE LOS ANGELES BASIN (2020)
Recorded Observation
Low Suitability
Average Suitability
High Suitability
0 4.75 9.5 19 Miles
Produced for SEED LA
Source: CalFlora, LA County GIS Data Portal


PROJECTED PURPLE NEEDLEGRASS SUITABILITY IN THE LOS ANGELES BASIN (2070)
Low Suitability
Average Suitability
High Suitability
0 4.75 9.5 19 Miles
Produced for SEED LA
Source: CalFlora, LA County GIS Data Portal

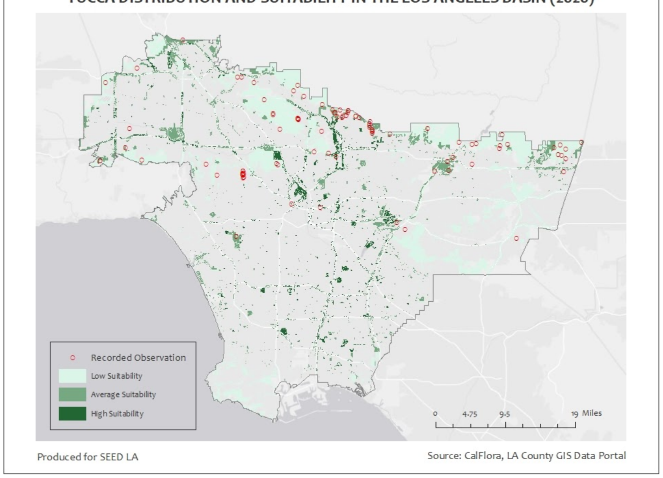
SOUTHERN MONKEYFLOWER DISTRIBUTION AND SUITABILITY IN THE LOS ANGELES BASIN (2020)



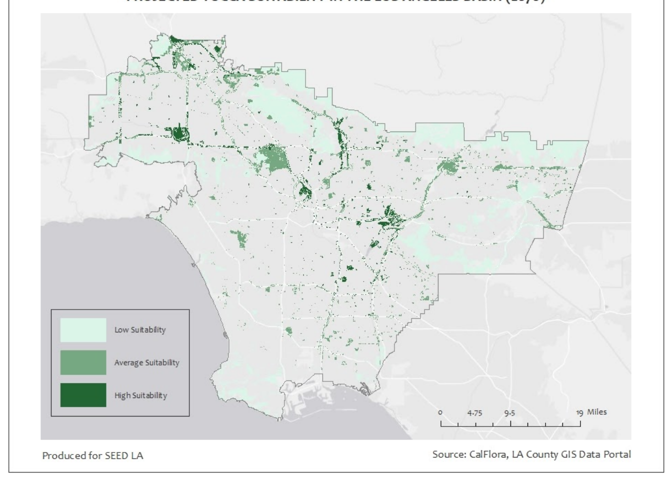
PROJECTED SOUTHERN MONKEYFLOWER SUITABILITY IN THE LOS ANGELES BASIN (2070)



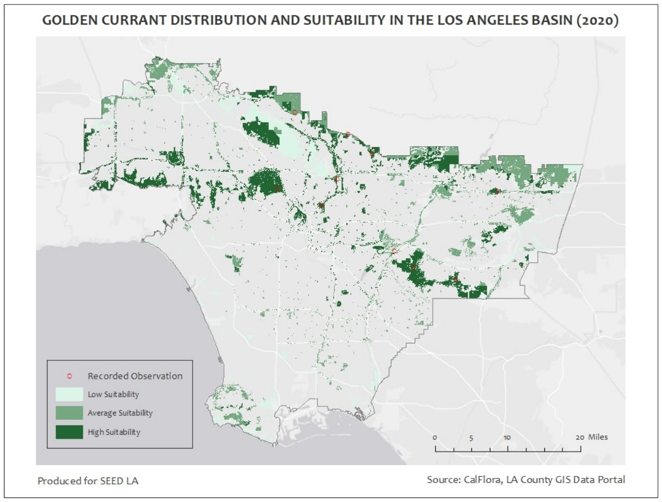
YUCCA DISTRIBUTION AND SUITABILITY IN THE LOS ANGELES BASIN (2020)



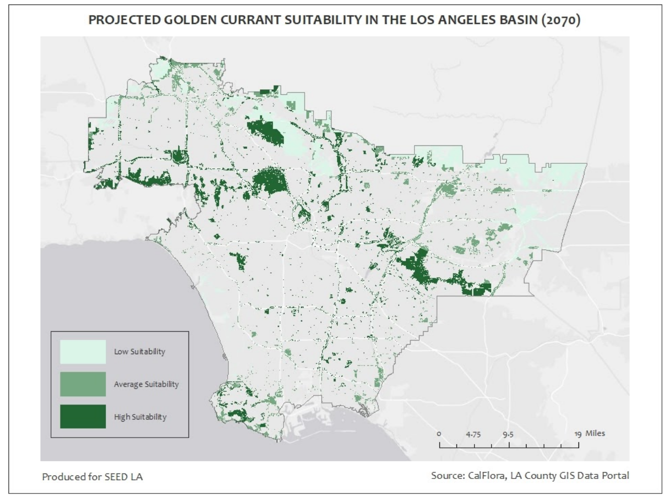
PROJECTED YUCCA SUITABILITY IN THE LOS ANGELES BASIN (2070)

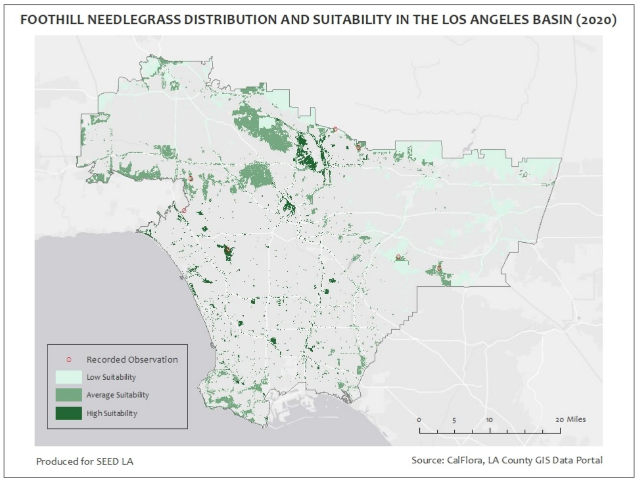


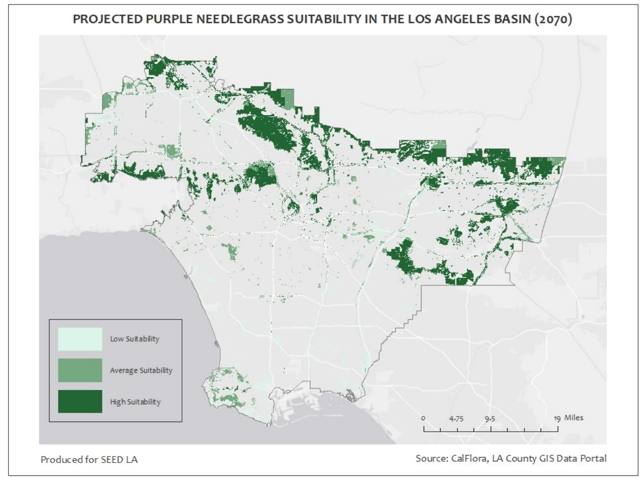
GOLDEN CURRANT DISTRIBUTION AND SUITABILITY IN THE LOS ANGELES BASIN (2020)

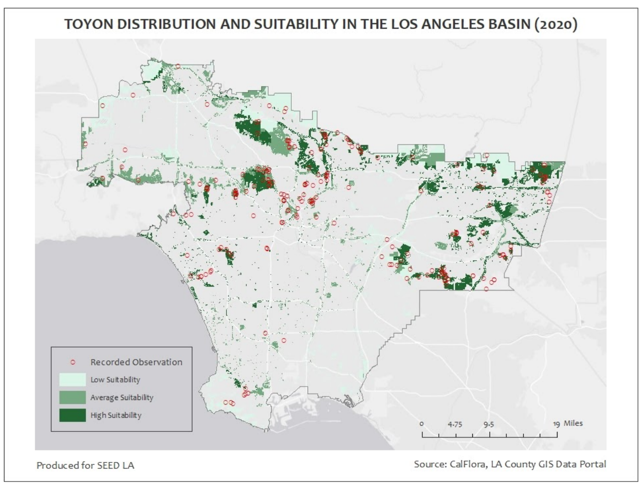


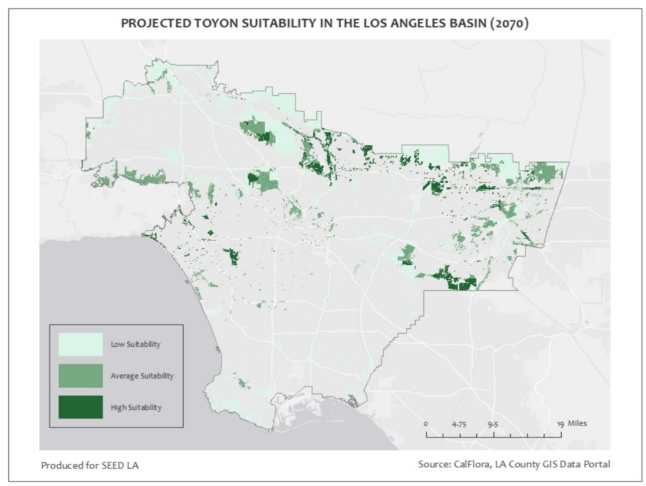
PROJECTED GOLDEN CURRANT SUITABILITY IN THE LOS ANGELES BASIN (2070)

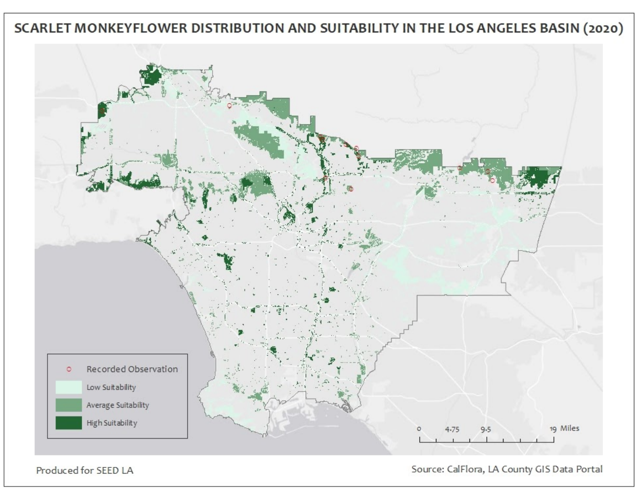


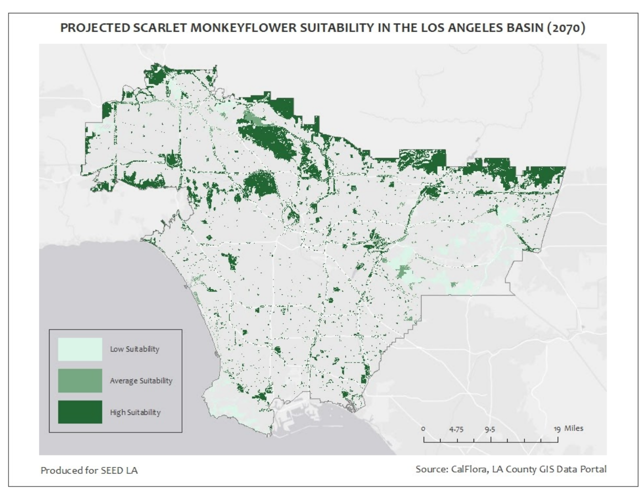

FOOTHILL NEEDLEGRASS DISTRIBUTION AND SUITABILITY IN THE LOS ANGELES BASIN (2020)
Recorded Observation
Low Suitability
Average Suitability
High Suitability
0 5 10 20 Miles
Produced for SEED LA
Source: CalFlora, LA County GIS Data Portal


PROJECTED PURPLE NEEDLEGRASS SUITABILITY IN THE LOS ANGELES BASIN (2070)
Low Suitability
Average Suitability
High Suitability
0 4.75 9.5 19 Miles
Produced for SEED LA
Source: CalFlora, LA County GIS Data Portal


TOYON DISTRIBUTION AND SUITABILITY IN THE LOS ANGELES BASIN (2020)
Recorded Observation
Low Suitability
Average Suitability
High Suitability
0 4.75 9.5 19 Miles
Produced for SEED LA
Source: CalFlora, LA County GIS Data Portal


PROJECTED TOYON SUITABILITY IN THE LOS ANGELES BASIN (2070)
Low Suitability
Average Suitability
High Suitability
0 4.75 9.5 19 Miles
Produced for SEED LA
Source: CalFlora, LA County GIS Data Portal


SCARLET MONKEYFLOWER DISTRIBUTION AND SUITABILITY IN THE LOS ANGELES BASIN (2020)
Recorded Observation
Low Suitability
Average Suitability
High Suitability
0 4.75 9.5 19 Miles
Produced for SEED LA
Source: CalFlora, LA County GIS Data Portal


PROJECTED SCARLET MONKEYFLOWER SUITABILITY IN THE LOS ANGELES BASIN (2070)
Low Suitability
Average Suitability
High Suitability
0 4.75 9.5 19 Miles
Produced for SEED LA
Source: CalFlora, LA County GIS Data Portal

References

Calflora.org. 2020. *Calflora - Search For Plants*. [online] Available at: <https://www.calflora.org/> [Accessed 10 June 2020].

Caltrans. 2020. Caltrans GIS Data. [online] Available at: <https://gisdata-caltrans.opendata.arcgis.com> [Accessed 20 May 2020].

Census Bureau QuickFacts. 2020. *U.S. Census Bureau Quickfacts: Los Angeles City, California; Los Angeles County, California; California*. [online] Available at: <https://www.census.gov/quickfacts/fact/table/losangelescitycalifornia,losangelescountycalifornia,CA/PST045219> [Accessed 21 May 2020].

Cepf.net. 2020. *California Floristic Province*. [online] Available at: <https://www.cepf.net/our-work/biodiversity-hotspots/california-floristic-province/species> [Accessed 21 May 2020].

Egis-lacounty.hub.arcgis.com. 2020. *County of Los Angeles Enterprise Geographic Information Systems*. [online] Available at: <https://lacounty.maps.arcgis.com/home/webmap/viewer.html?useExisting=1&layers=e808f076a0ef4e96bd9f139f10b2b2c> [Accessed 30 April 2020].

Geohub.lacity.org. 2020. *Los Angeles GeoHub.* [online] Available at: <http://geohub.lacity.org> [Accessed 21 May 2020].

Hijmans, R. (2020). raster: Geographic Data Analysis and Modeling. R package version 3.0-12. https://CRAN.R-project.org/package=raster

Hijmans, R., Phillips, S., Leathwick, J., Elith, J. (2017). dismo: Species Distribution Modeling. R package version 1.1-4. https://CRAN.R-project.org/package=dismo

Lacontroller.org. *Property Panel*. 2020. [online] Available at: <https://lacontroller.org/data-stories-and-maps/propertypanel/> [Accessed 27 May 2020].

Li, E., Parker, S., Pauly, G., Randall, J., Brown, B., Cohen, B. 2019. An Urban Biodiversity Assessment Framework That Combines an Urban Habitat Classification Scheme and Citizen Science Data. Frontiers in Ecology, Evol. 7:277 [Accessed 28 April 2020].

Loarie, S., Carter, B., Hayhoe, K., McMahon, S., Moe, R., Knight, C. and Ackerly, D., 2008. Climate Change and the Future of California's Endemic Flora. *PLoS ONE*, 3(6), p.e2502.

Mrld.gov. 2016. *National Land Cover Database*. [online] Available at: <https://www.mrlc.gov/data?f%5B0%5D=category%3Aland%20cover> [Accessed 30 April 2020].

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Riordan, E., Montalvo, A., Beyers, J., 2018. *Using Species Distribution Models With Climate Change Scenarios To Aid Ecological Restoration Decisionmaking For Southern California Shrublands.* United State Department of Agriculture.

Rspatial.org. 2020. *Species Distribution Modeling — R Spatial*. [online] Available at: <https://rspatial.org/raster/sdm/index.html> [Accessed 10 June 2020].

Websoilsurvey.sc.egov.usda.gov. 2020. *Web Soil Survey - Home*. [online] Available at: <https://websoilsurvey.sc.egov.usda.gov/App/HomePage.htm> [Accessed 10 June 2020].

Williams, J., Seo, C., Thorne, J., Nelson, J., Erwin, S., O'Brien, J. and Schwartz, M., 2009. Using species distribution models to predict new occurrences for rare plants. *Diversity and Distributions*, 15(4), pp.565-576.

Worldclim.org. 2020. *Worldclim*. [online] Available at: <https://worldclim.org/> [Accessed 10 June 2020].