

Appendix B

Detailed Study Methods

Chapter 2.0 in the guide provides a brief summary of the overall methods used to develop the information used as the basis for the selection of habitat blocks as conservation targets. This appendix provides greater detail on our methods for certain analyses including predicted habitat mapping for rare species and the rules used to define ‘contiguous’ vernal pool habitat blocks to identify the 20 largest remaining blocks in each Region.

Predicted Habitat Mapping

A key element of our analysis was developing predicted habitat mapping for the core species in each Vernal Pool Region (described in detail below). We then layered this mapping using ArcGIS software (ESRI 2020) to develop a ‘hotspot’ map showing the variable concentration of predicted habitat for the core species across the mapped vernal pool habitat in the Region. We considered the hotspot mapping when selecting target habitat blocks, favoring areas with a concentration of predicted habitat for multiple species.

We developed our predicted habitat mapping using a technique developed by Holland and Hollander (2007) and refined by Vollmar et al. (2016). These papers provide detailed descriptions of this technique. The methods described below are adapted from Vollmar et al. (2016), tailored to the specific techniques we used for this guide. We also assessed an alternative method using maximum entropy (MaxEnt) (Phillips et al. 2006; Phillips and Dudik 2008). After comparison of the outputs from the two techniques, we opted to use the Vollmar et al. (2016) technique based on our determination that the predicted habitat mapping was more predictive and targeted. The MaxEnt technique is briefly described below along with our methods and results for comparing the accuracy of the two techniques.

Background

Species distribution is a function of at least four classes of factors: 1) abiotic conditions, 2) biotic factors including competition, 3) regions that are accessible to dispersal, and 4) the ability of the species to adapt to new conditions (Soberon and Peterson 2005). The regional and local distribution of a species is determined by the dynamic interaction of these factors in space and time. Determining the specific factors that control the distribution of a particular species can be extremely challenging. As an example from this study, midvalley fairy shrimp (*Branchinecta mesovallensis*) can readily occupy a range of pool sizes and depths and can tolerate a range of pool temperature conditions. Yet, it has a very limited and sporadic distribution within two localized areas in the Great Valley. This may be due to competition with the more widespread vernal pool fairy shrimp (*Branchinecta lynchi*), preference for vernal pool habitats near the valley basin due to some climate or soil conditions, or some other unknown factor.

Predictive (or predicted) habitat modeling is a rapidly growing discipline that is being used to develop predictive maps of the distributions of species and habitats, often focusing on rare or sensitive resources of conservation interest. It has been marked by a steady increase in analytic capacity due to improvements in technology. Franklin (1995) cites the origin of predictive modeling with the refining of ecological

niche theory and gradient analysis. The ability to analyze digital geographic information and apply statistical methods has given ecologists the capacity to use the information about a species’ current distribution in a model that can predict other potential locations (Fertig and Reiners 2002). This process is driven by the ability to project environmental variables across a landscape and make predictions and calculations based on a range of parameters given by the habitat requirements of the study species.

Most predictive modeling methods use either the mechanistic or correlative approach (Fertig and Reiners 2002; Soberon and Peterson 2005). The mechanistic approach measures the species’ response to physical parameters and then projects those parameters across the study area to produce the predictive map. The correlative approach uses known occurrence points to determine the physical and biological parameters that the species currently lives within and then projects these parameters across the study area to produce the predictive map (Soberon and Peterson 2005).

Our method is correlative and based on the assumption that there is a meaningful relationship between the known points of occurrence and the environmental parameters with which they coincide (Soberon and Peterson 2005). Our study builds on Holland and Hollander’s (2007) original study which was also correlative and was inspired by earlier correlative studies by Raxworthy et al. (2003) and Nix (1986). We also found guidance from Skov and Borchsenius (1997) as well as several papers included in the proceedings of a symposium on predicting species occurrences (Scott et al. 2002).

All predictive maps require an underlying habitat model of some type. When research-grade presence and absence data from a well-designed study are available, a correlative model can be created based on a statistical analysis of the habitat correlation of the presences and absences in the study. Techniques such as MaxEnt can even create habitat models based on presence-only data. However, presence-only data collected in an ad-hoc manner (such as California Natural Diversity Database (CNDDB) occurrence records) is not the same as data collected in a random sampling design. Sampling bias in these datasets (e.g., over-sampling of mitigation properties and public lands, under-sampling of private and hard-to-access lands) creates significant auto-correlation in the measured environmental variables used to create the habitat model. This in turn distorts the habitat model, and reduces the accuracy of the predictive map. Auto-correlation created by over-sampling in particular can have the effect of training the model to over-value those areas which are well-studied, and under-value areas which are under-studied. This completely obviates the purpose of the predictive map, which is to assign value to areas which have not been directly sampled. The Holland and Hollander method is a non-statistical method, which is not vulnerable to this form of sampling bias (2007). It is unable to predict occurrences in locations which have unique habitat conditions under which the species has never before been documented. In exchange for this lack of flexibility, though, it is completely unaffected by over-sampling of any particular habitat. This method provides a valuable tool when only ad-hoc occurrence data is available, and time-sensitive conservation issues need to be addressed.

Using the correlative method as well as a non-randomized/presence-only occurrence dataset introduced limitations to our study analyses. Each occurrence data point includes both environmental and biological factors that cannot be separated. Therefore, it was generally difficult to ascribe distribution patterns to specific habitat factors. Also, the occurrence data were not collected using a randomized methodology, are not uniformly distributed, and do not include data points for species absence. Statistical analysis of the data is therefore likely to be misleading.

The predictive habitat mapping produced by our study is intended to provide general guidance on the potential occurrence of the analyzed species within the study areas. Actual occurrence in a specific location or pool can depend on a variety of factors such as local soil inclusions below the minimum mapping unit, the apparently random manner in which species tend to occupy individual pools in a complex, undetected past disturbances, etc. Also, the predicted mapping relies on occurrence records and habitat data layers that may include some errors. As such, actual occurrence or absence of the analyzed species in a particular location would need to be verified by field sampling.

Analytic Methods Used in this Study

For this study, we analyzed available, vetted rare species occurrence records within our study areas using ArcGIS software to predict habitat distributions for our selected core species. For each species, we developed the predicted habitat mapping for the entire Great Valley region, using all available vetted occurrence records. We then clipped this mapping to an individual Vernal Pool Region when doing the regional analysis and hotspot mapping for the core species identified in the Region. The extent of predicted habitat for each species was based on the range of the occurrence locations within a cumulative set of habitat parameters related to soils, topography, and climate (detailed below). The output of predicted habitat for each species was limited to the subset of areas that were spatially within the range of all the combined habitat parameters. We did not perform predictive habitat mapping for the two rare amphibians (California tiger salamander [*Ambystoma californiense*] and western spadefoot [*Spea hammondi*]). The distribution of these species is heavily influenced by the presence of individual suitable breeding ponds which are not indicated in the available vernal pool habitat mapping. Using the other habitat parameters without including presence of suitable breeding ponds led to inaccurate predicted habitat mapping for these species.

Compilation and Vetting of Occurrence Records

We compiled species occurrence records from two sources: CDFW (2021) and limited additional occurrence records obtained from Vollmar Natural Lands Consulting and Helm Biological Consulting. The consultant occurrence records had precise location data, creating a more robust dataset for analysis. All of the consultant data consisted of either point data or vernal pool polygon data collected using professional GPS units (with submeter precision) or points marked on high-resolution aerial photograph maps clearly showing vernal pool signatures and the specific occupied pools. We only used oc-

currence records located within our study area (Great Valley) using the boundary shown on figures within the guide.

The records contained in the CNDDDB are submitted mostly on a voluntary basis by individuals who have conducted field surveys either in a professional or amateur capacity. These records range from 'non-specific' records with a one-mile precision, to 'specific' point data with an 80-meter precision, to 'specific' polygon data encompassing a suitable habitat area within which the species occurs in one or more locations. All of the available CNDDDB records for the rare species within our study area were initially obtained for review and potential inclusion in the final dataset.

Data from both sources (CNDDDB and consultants) were evaluated by applying the process and exclusion criteria described below to ensure that all occurrence records used met basic levels of accuracy and consistency. All occurrence records were merged into a master dataset in ArcGIS software by employing the steps described below. All CNDDDB and consultant records consisting of a single point were directly entered into the master dataset. All CNDDDB and consultant records consisting of a single polygon were exported to a centroid point file and then entered into the master dataset. Some CNDDDB records consist of multiple mapped point occurrences or polygons combined into a single 'record'. All such records were 'exploded' (separated) into solitary ('single-part') polygons or points. These single-part point records were entered into the master dataset. The single-part polygon records were exported as a geodatabase layer, then exported to a centroid point file. This point file was then entered into the master dataset. All relevant attribute fields were maintained during the merge including fields describing attributes relevant to the four exclusion criteria listed below.

The final master dataset of individual point records was overlaid onto the most recent available ArcGIS aerial background imagery (ESRI 2020) to review site conditions associated with each occurrence. Individual occurrences were excluded from the analysis subset if they had any of the following characteristics:

- Not located in the Great Valley study area;
- Classified as a 'created' or 'constructed' pool, including stock ponds;
- Occurrence description was so general in nature that it could not be reasonably verified to be in the vicinity of its assigned location; or
- Assigned occurrence location was mapped on a soil series or phase or in a topographic position (such as a steep slope or creekbed) that was not suitable for vernal pools and thus likely erroneous.

All CNDDDB 'specific' point occurrences with a specific-area, 80-meter, or 0.10-mile precision were automatically included in the dataset for analysis (as long as they were not otherwise excluded by other characteristics listed above). The remaining, less accurate CNDDDB occurrences (again, as long as they were not otherwise by other characteristics listed above) were included for all parameters except elevation, slope, and soil since these parameters can vary substantially on a local scale. This approach was important because it allowed many non-precise occurrences to still be included as part of the anal-

ysis for the other parameters and also to be depicted on maps as part of the species' distribution. In nearly all cases, these non-precise occurrences were in the vicinity of soils and slopes represented by other, more precise occurrences so the loss of data with regards to these parameters was minimal. In the case of duplicate data when consultant data had also been reported to CNDDDB and therefore occurred twice in the dataset, the CNDDDB record was excluded and the consultant data were included in the data set since they were assumed to be the most accurate.

The data used in our study were collected over many years by numerous individuals using a wide array of survey and occurrence representation methods, and, therefore are not generally suitable for rigorous statistical analysis to determine the relative quality of predicted habitat for a given species.

Selection of Habitat Parameters

We used nine individual habitat parameters for our predicted habitat analysis: soil, elevation, slope, mean annual precipitation, mean minimum January temperature, mean maximum July temperature, mean July precipitation, mean relative summer humidity, and low slope continuity. We used all eight of the habitat parameters used by Holland and Hollander (2007) (the first eight listed above) for all analyzed species. We also added a new parameter developed for the Vollmar et al. (2016) study termed 'low slope continuity' (see description below) for midvalley fairy shrimp and vernal pool tadpole shrimp (*Lepidurus packardii*) only. We considered using additional parameters including mapped vernal pool habitat density, cover, and/or diversity but could not justify any differentiation of predicted habitat due to the potential bias inherent in the occurrence data and the coarseness of the vernal pool habitat mapping data.

Numerous studies have shown a correlation between the distribution of vernal pools as well as vernal pool endemic species with specific geologic surfaces and their associated soils (Holland and Dains 1990; Dittes and Guardino 2002; Helm and Vollmar 2002; Vollmar et al. 2016). This association was a basic premise in both Holland and Hollander's (2007) study and the Vollmar et al. (2016) study. There are three different levels of available soil mapping for the Great Valley: the State Soil Geographic (STATSGO) database, Soil Survey Geographic (SSURGO) database soil series, and SSURGO soil phases. The STATSGO data (USDA 1994) are the most generalized, mapping common landscape units often comprised of a set of related soil series. This dataset was produced on a statewide basis. Examples of STATSGO map units include 'Redding-Pentz-Corning' and 'Fresno-Dinuba-Lewis'. Holland and Hollander (2007) used the STATSGO data since they were doing statewide analysis and the more precise SSURGO data were not yet available on a statewide basis. Vollmar et al. (2016) found that the STATSGO data often greatly exaggerated the extent of predicted substrate when applied on a regional scale. The SSURGO data (USDA 2012) were developed by digitizing county soil surveys and are very detailed (particularly in the Great Valley and other agricultural regions), down to the soil phase. Examples of SSURGO map units (soil phases) are 'Redding gravelly loam, 0-8% slope' and 'Lewis loam, moderately saline-alkali, 0-1% slopes'. Vollmar et al. (2016) found

that using SSURGO soil phase data sometimes resulted in under-mapping of predictive substrate, especially when there are a limited number of species occurrences. For example, while a species may have the capacity to occur on the full range of Redding soil phases, limited surveys may have only documented the species on a subset of the different soil phases associated with the series. We were concerned that limiting the predictive substrate to only those specific soil phases on which the species has been documented might arbitrarily exclude other related soil types which were also likely suitable for the species. Also, slope is a component of soil map units. Since we used slope as a separate habitat parameter, it was not necessary to distinguish slope within the soil data layer. In consideration of these factors, we agglomerated individual SSURGO soil phases into common soil series such as 'Redding' or 'Lewis'. We chose this approach to err on the side of inclusion versus exclusion of actual suitable habitat using the SSURGO data. This approach provided a broader, more inclusive predicted substrate without going as broad as the STATSGO data.

We used 30-meter Digital Elevation Model (DEM) grids (USGS 1997) to analyze elevation and slope. We investigated using 10-meter DEM grids but found them to be too fine-grained which introduced erroneous slope ranges (excessively steep). Since the slope of an individual occurrence is calculated as the average of a 30-meter grid scale, there is a high potential that a small number of occurrences have erroneously high calculated slopes where there is a nearby bluff, creek bank, or other steep feature not reflective of the overall landscape-level slope trend or the local site where the occupied vernal pool occurs. This potential exaggeration of slope could be addressed by eliminating slope 'outliers', though we did not do this for this parameter since we found it had only a very minor effect on the extent of predicted habitat when considered on a study area-wide scale.

'Low slope continuity' is a derived parameter that we developed for our study to represent an area that is contiguous with the basin of the Great Valley up to a selected slope maximum. Many expert vernal pool shrimp biologists (Vollmar pers. obs.; Witham pers. obs.; Schweitzer pers. obs.; Helm pers. comm.; Helm and Vollmar 2002) have noted the apparent association of midvalley fairy shrimp and vernal pool tadpole shrimp with lower slope areas that are contiguous with the Great Valley basin. This association may be related to elevated ambient humidity in the Great Valley basin (though the available coarse climate data do not show this), geologic history, or other unknown factors. We developed this parameter to address this apparent association. Low slope continuity was an important added parameter that excluded significant areas where midvalley fairy shrimp and vernal pool tadpole shrimp are not known or expected to occur but otherwise would have been included as predicted habitat using the soil, slope, and elevation parameters. This parameter excluded terraces on high terrain around the edge of the Great Valley that are separated or disjunct from the Great Valley basin by moderate to steep slopes. This parameter includes areas that are directly contiguous with the basin of the Great Valley up to a selected slope maximum. Areas that are not directly connected through a continuous slope below the maximum threshold were excluded. For example, while there are extensive, contiguous low gradient terraces

in the northeast of Merced County, these areas are separated from the Great Valley basin by steeper terrain and were thus excluded. The upper end of the slope range was set at the slope that included 95 percent of occurrences extending up from the valley basin. For example, we calculated that 95 percent of the vernal pool tadpole shrimp occurrences used in our analysis occurred on slopes between 0.0 and 4.2 percent. Thus, the low slope continuity parameter for vernal pool tadpole shrimp included all areas that were directly contiguous with the valley basin via slopes that were at or below a 4.2 percent slope.

For climate parameters, we used the same dataset used by Holland and Hollander (2007) since no higher accuracy climate data were available. This dataset was obtained from state-wide 1.25 arc-minute Parameter-elevation Regressions on Independent Slopes Model (PRISM) climatic data, re-sampled to 100-meter grids.

Analysis of Predicted Habitat

We calculated the extent of predicted habitat for each analyzed rare species throughout the Great Valley using ArcGIS software by first calculating the habitat range of each parameter associated with the specific available and vetted occurrence locations and then extrapolating to include all areas in the Great Valley within the range of that parameter. The individual habitat parameter range layers were then combined to produce the predicted habitat for each species throughout the Great Valley. The final output of predicted habitat was limited to the subset of areas that were spatially within the range of all combined habitat parameters.

Calculation of Predicted Habitat

The acreage of extant predicted habitat for each analyzed rare species was determined by overlaying the current vernal pool habitat mapping (Witham 2021) over the predicted habitat layer and calculating the acreage of predicted habitat within the vernal pool habitat polygons. Thus, the extant predicted species habitat was a subset of the larger mapped extant vernal pool habitat polygons. This approach assumes that the extant predicted species habitat is limited to areas where vernal pools still exist since all of the analyzed species are essentially vernal pool endemics.

MaxEnt Technique

As mentioned above, we evaluated an alternate method for developing rare species predicted habitat mapping. The current standard algorithms for species distribution modeling are based on maximum entropy (MaxEnt) (Phillips et al. 2006). MaxEnt models are widely used because of their ability to use presence-only data and multivariate habitat data to predict species occurrence. These methods work by creating a set of “pseudo absence” or “background” points, and comparing the habitat metrics at these points to the habitat metrics at documented occurrence points for the species. MaxEnt is used to generate response curves of the species to each of the variables, and then output a map of predicted likelihood of species occurrence based on the environmental variables.

We conducted MaxEnt analyses on all of the rare vernal pool species addressed in our guide (except California tiger

salamander and western spadefoot for the reasons discussed earlier). We used a similar, but much larger, set of predictor variables in the MaxEnt analyses than were used in the revised Holland and Hollander technique (Vollmar et al. 2016). The MaxEnt analyses used as inputs: the 19 BioClim variables (Fick and Hijmans 2017), the diurnal temperature range in each month (derived from Fick and Hijmans 2017), total precipitation in each month (ibid), the percent of annual precipitation in each month (ibid), average incident solar radiation in each month (ibid), average, average minimum, and average maximum temperature in each month (ibid), average vapor pressure in each month (ibid), average wind speed in each month (ibid), geologic formation (see **Table 3.4, Chapter 3**), elevation (USGS 1997), percent slope (ibid), terrain roughness (ibid), soil series and texture (USDA 2012), Vernal Pool Region (Keeler-Wolf et al. 1998), vernal pool types present in the Region (ibid), and vernal pool cover, density, size diversity, disturbance, and large pool presence (Witham et al. 2014; Witham 2021). The MaxEnt algorithm includes a regularization multiplier which penalizes models with many predictor variables, so most variables were not actually used in the final MaxEnt model created for each species.

MaxEnt can select the “background” values that are used as pseudo-absences either at random, or from selected locations where the species is believed to be absent. The list of these selected background locations is referred to as a “bias” file. We created bias files for each species by using locations which were known to have been sampled for similar species. For example, the bias file for each specific vernal pool branchiopod species was based on all known occurrences of all vernal pool branchiopods (CDFW 2021). Therefore, this represents locations which are known to have been sampled for vernal pool branchiopods (and thus, where a specific species was not found). Similar bias files were created for groups of plants.

MaxEnt is sensitive to several input parameters, and the quality of its output can usually be improved by tuning these parameters (Searcy and Shaffer 2016). We experimented with a wide range of parameters to attempt to find the highest predictive value for each species. For each input species, we ran models with regularization multipliers between 0.5 and 10. We tested models using differing bias files, and no bias files. We tested models using only selected subsets of habitat variables. We tested models constructed using both permutation importance and percent contribution for each variable.

We selected the models with the best omission lines and areas under the curve for each species, representing the highest predictive value. Model parameters for each species varied, but typical selected models were based on a 20% omission random test with 20 bootstrapped replicates, and a regularization multiplier between 3 and 5. We also tested different threshold rules to determine the cut-off for what was and was not modeled as habitat for each species. We settled on using the 10% omission threshold to create models which focus on the most important habitat for each species.

After final selection of MaxEnt models, we compared the MaxEnt mapping to the Vollmar et al. 2016 mapping. We conducted two types of comparison: visual evaluation of maps by species experts (John Vollmar and Brent Helm), and quantitative comparisons of area predicted vs. prediction rate against

test data. We compared only three species (vernal pool tadpole shrimp, succulent owl's-clover [*Castilleja campestris* var. *succulenta*], and San Joaquin Valley Orcutt grass [*Orcuttia inaequalis*]) due to time constraints, the large number of species being analyzed, and limited availability of species experts to consult on all species. Species experts generally preferred the Vollmar et al. method-based mapping to the MaxEnt-based mapping based on their understanding of the species' actual habitat needs and distribution. In the quantitative analysis, the Vollmar et al. analysis had a higher prediction rate of test samples for two out of three species. The Vollmar et al. analysis had a lower predicted area (i.e., was more specific/less general) for all three species. Based on these results, we opted to use the Vollmar et al. method to develop the predicted habitat mapping for the overall block analysis.

Rules for Defining 'Contiguous' Vernal Pool Habitat Blocks

For the purposes of this guide, 'vernal pool habitat' is defined as the complex of pool basins and swales along with the surrounding upland habitat that form local pool watersheds, and together supports the overall ecology of vernal pools (Witham et al. 2014). We used the most current available vernal pool habitat mapping available for the Great Valley (Witham et al. 2014; Witham 2021) as the basis for the habitat analyses that we conducted for this study.

The available mapping delineates existing vernal pool habitat within individual polygons. In some cases, these polygons may be transected by roads, railroad tracks, or canals, have some areas with low-density (rural residential) housing development, or have other developments. Adjacent polygons that are near each other may be separated by development, farmlands, or natural habitat. To identify and map 'contiguous' vernal pool habitat blocks, we developed a set of rules based on the habitat and land use setting within and between individual mapped polygons. We then applied these rules using ArcGIS software to determine if adjacent polygons were considered part of the same block or constituted different blocks for the purposes of our analysis.

Adjacent polygons were merged to form a single habitat block if they were 400 meters or less apart and the intervening area fell into any of the following categories:

- Grassland
- Disked grassland/hay crop (not irrigated)
- Woodland or forest
- Natural streams
- Ponds and lakes
- Natural barren soil or rock outcrops

- Utility line corridors
- Natural wetlands
- Artificial/enhanced wetlands (e.g., canal leakage wetlands)
- Naturalistic open space with or without public access
- Dirt roads (any width)
- Paved roads up to 4 lanes max/40m wide max (without a median)
- Railroads up to 2 tracks max/20m wide max
- Paved hiking/biking trails up to 20m wide max
- Canals (of any size)
- Airport runways up to 40m wide max

Any combination of paved roads (up to 4 lanes without a median), railroads (up to 2 tracks/20m wide), trails (up to 20m wide max), canals (any size), or runways (up to 40m wide max) up to a combined 80m wide

Additionally, a connecting corridor at least 400 meters wide of the types listed above must be present for adjacent polygons to be merged into a habitat block. For example, this prevented two adjacent polygons which were diagonally opposite ('kitty-corner') each other to be merged into a habitat block, where species would have to travel to the corner of one polygon to reach the other polygon of habitat. It was determined that this 'bottle-necking' effect did not represent habitat continuity. Adjacent polygons were not merged to form a habitat block if they were more than 400 meters apart or the intervening area fell into any of the following categories:

- Any area of earth moving/grading more than 40m wide (meant to capture in-progress roads and developments)
- Any amount of urban/industrial/parking lot
- Any amount of residential/rural residential neighborhood/subdivision (less than 40-acre lot size)
- Any amount of obviously irrigated pasture (green in summer aerals)
- Any amount of row crop
- Any amount of permanent crop (orchard/vineyard)
- Any amount of obvious fallow cropland
- Any amount of non-naturalistic park/lawn/ball field
- Any amount of high-density enclosed livestock rearing (horse paddocks, feedlots, turkey ranches)
- Any amount of mine/quarry/dredge piles

Once all individual extant habitat polygons were merged into habitat blocks, the acreage of each resulting block was calculated. The top 20 largest blocks were then identified and assessed to determine the target blocks for conservation.