Environmental factors affecting ecological niche of Coccidioides species and spatial dynamics of valley fever in the western United States

Weaver, Elizabeth A.,[a] Kolivras, Korine N,.[b] Thomas, R. Quinn,[b] Thomas, Valarie A.[b] Abbas, Kaja M.[c]

a. U.S. Military Academy, 745 Brewerton Road, West Point, NY 10996

b. Virginia Polytechnic Institute and State University, 115 Major Williams Hall, Blacksburg, VA 24061

c. London School of Hygiene & Tropical Medicine, 119 LSHTM Keppel Street, London, WC1E 7HT, United Kingdom

Abstract

Coccidioidomycosis is an understudied infectious disease acquired by inhaling fungal spores of *Coccidioides* species. While historically connected to the southwestern United States, the endemic region for this disease is not well defined. This study's objective was to estimate the impact of climate, soil, elevation and land cover on the *Coccidioides* species' ecological niche. This research used maximum entropy ecological niche modeling based on disease case data from 2015 to 2016. Results found mean temperature of the driest quarter, and barren, shrub, and cultivated land covers influential in characterizing the niche. In addition to hotspots in central California and Arizona, the Columbia Plateau ecoregion of Washington and Oregon showed more favorable conditions for fungus presence than surrounding areas. The identification of influential spatial drivers will assist in future modeling efforts, and the potential distribution map generated may aid public health officials in watching for potential hotspots, assessing vulnerability, and refining endemicity.

Key Words: Coccidioides; Valley Fever; Maxent; Niche Modeling

1. Introduction

Coccidioidomycosis, commonly referred to as valley fever, is a fungal disease that affects humans and mammals in the western United States. The fungal species that cause this disease, *Coccidioides immitis* and *Coccidioides posadasii*, live and grow in the soil, but environmental conditions and human disturbances can cause the fungal spores to become airborne. These spores will either return to the soil or become inhaled. If inhaled, the fungus can become parasitic and cause symptoms such as fever, cough, chest discomfort, and fatigue; in a small percentage of the population, endospores disseminate throughout the body and cause more serious illness (Saubolle, McKellar, & Sussland, 2007). Valley fever is not contagious; inhalation of spores is the only method of infection, with a few rare exceptions (Pappagianis, 1988). The public health burden associated with this disease has been increasing as cases have been steadily rising over the past 20 years with approximately 14,000 cases reported in the U.S. in 2017 (Centers for Disease Control and Prevention, 2018). Approximately 200 people a year die from this disease (Huang, Bristow, Shafir, & Sorvillo, 2012).

The environmental factors influencing the geographic range and distribution of the *Coccidioides* spp. is not well understood because it is very difficult to find and isolate the fungus from the soil (Barker, Tabor, Shubitz, Perrill, & Orbach, 2012; Greene, Koenig, Fisher, & Taylor, 2000). Until recently, the established and suspected endemic region for this disease in the U.S. was thought to be confined to states in the southwest (Pappagianis, 1988). In the early 2000s, the discovery of multiple *Coccidioides* spp. growth sites in Washington state, well outside what had previously been considered the suspected endemic region, put into question the actual and potential geographic range and distribution of the pathogen (Litvintseva et al., 2015; Marsden-Haug et al., 2013). It is uncertain whether the spores found growing in Washington had been present for an extended period, or whether they had been recently introduced (Litvintseva et al., 2015), but researchers believe that a new niche for *Coccidioides* spp. has established or is establishing in eastern Washington (Marsden-Haug et al., 2013). This expansion of the suspected endemic area leads to questions regarding what controls the distribution of this disease (Benedict, Thomspon, Deresinski, & Chiller, 2015). Species distributions are often limited by climate conditions and physical environment features (Raghavan et al., 2016). Based on previous research and our understanding of the *Coccidioides* species' lifecycle, we hypothesize that climate, soil, elevation, and land cover influence this spatial distribution.

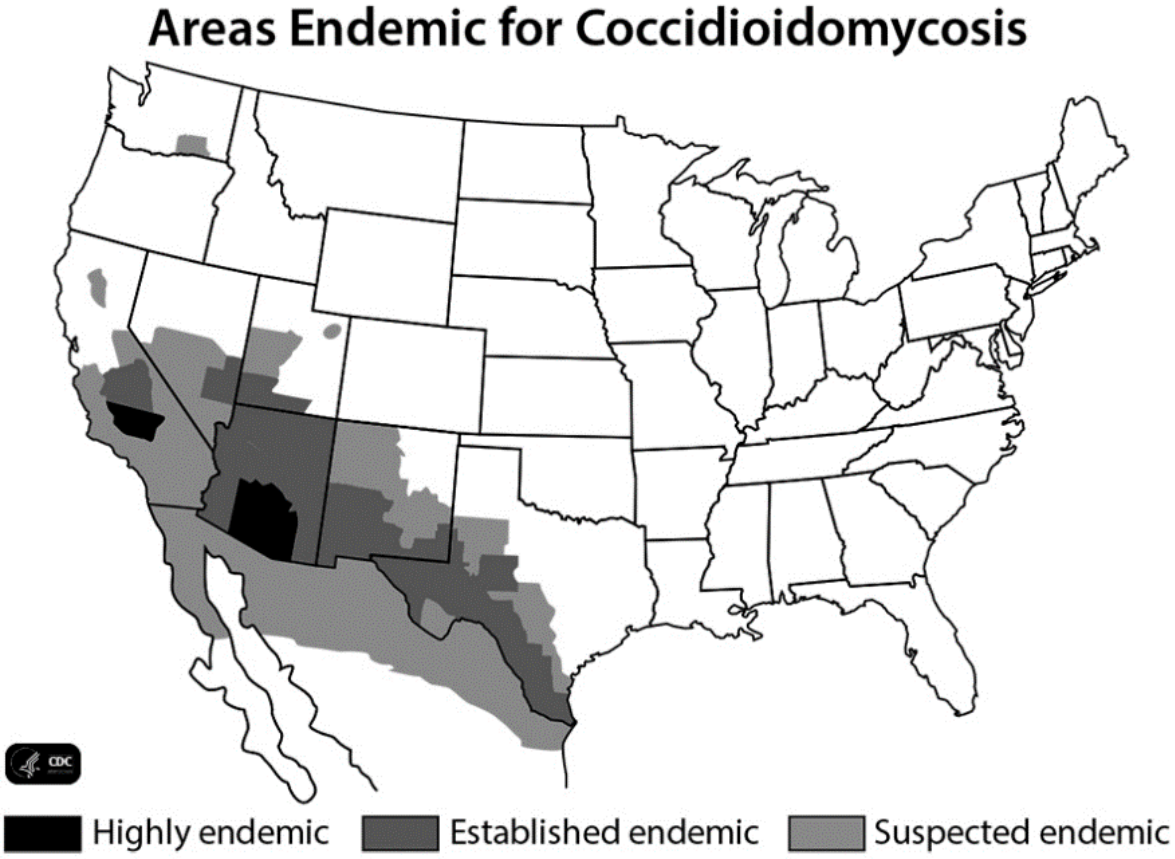

Figure 1. CDC map showing approximate areas where *Coccidioides* spp. are known or suspected to live in the U.S. and Mexico. This map is based on studies performed in the late 1940s and 1950s and also on locations of more recent outbreaks and cases (source: https://www.cdc.gov/fungal/diseases/coccidioidomycosis/causes.html).

A relatively unexplored approach for assessing the ecological and geographic distribution of *Coccidioides* spp. is ecological niche modeling (ENM). Ecological niche models use a mathematical representation of known species distribution points, as represented by environmental variables, to estimate the probability of occurrence at any site (Peterson, 2006). Ecological niche modeling of *Coccidioides* spp. has been limited, presumably, due to the relatively small number of positive soil samples. Ideally, the spatial distribution of the disease could be assessed at fine-scales based on positive soil isolations, but the organism is difficult to detect and has only been recovered from natural settings in a limited number of studies in small geographic regions (i.e. Barker et al., 2012; Elconin, Egeberg, & Egeberg, 1964; Greene et al., 2000; Swatek, Omieczynski, & Plunkett, 1967). Baptista-Rosas et al. (2007) developed an ecological niche model for *Coccidioides* spp. using a Genetic Algorithm for Rule Set Production (GARP) approach based on reports of 18 point-sites of known positive isolations dating from 1960-2002 to generate a predictive model that identified hotpots in Mexico, California, Arizona, and Texas. They concluded that the most probable fundamental ecological niche is the arid North American deserts, providing a methodological basis for further characterization of realized niches.

Building off this previous research, the present study used a maximum entropy (MaxEnt) ecological niche modeling approach to estimate the environmental impact of climate, soil, elevation, and land cover affecting the ecological niche of *Coccidioides* spp. and the spatial dynamics of valley fever in the United States. We addressed the limited number of soil samples that have tested positive for *Coccidioides* spp. by using recent human case data (2015-2016) and expanded the region of analysis to include seven states in the western U.S. A statistical explanation of MaxEnt can be found in Elith et al. (2011), but in summary, the algorithm works with presence-only data by comparing the environmental data found at species presence locations to data from across the entire study area. It estimates the ratio of f_1/f where f_1 is the probability distribution describing the environmental characteristics at species occurrence sites and f is the probability distribution describing characteristics of the whole environment, including presence and absence sites (Guillera-Arriota, Lahoz-Monfort, & Elith, 2014). To estimate the probability distribution for species occurrence sites, MaxEnt applies the maximum entropy principle, seeking a distribution as close as possible to a uniform distribution, while deviating only as minimally as possible in order to explain the observations (Guillera-Arriota et al., 2014; Jaynes, 1957). While there are multiple options available for niche modeling, the MaxEnt algorithm was chosen because of its ability to handle presence-only data, explore complex/interacting relationships, and generate predictions that compare favorably with other models (Elith et al., 2006); as far as we are aware, this is the first time MaxEnt has been used to assess the ecological niche of *Coccidioides* spp.

2. Materials and Methods

2.1 Study Area

The study area for this research includes established endemic and suspected endemic states for valley fever to include California, Nevada, Arizona, Utah, New Mexico, and Washington along with the neighboring state of Oregon. Oregon is not considered suspected endemic by the CDC, but it is included as it lies geographically in between two states with known growth sites; therefore, this research assumes that the pathogen is potentially present in Oregon given the pathogen's airborne nature and the state's location between two endemic or suspected endemic states. Texas is the only endemic state that does not require valley fever to be reported and therefore was excluded from this study.

2.2 Presence Data

Annual valley fever case totals from 2015 to 2016 were obtained for all states in the study area from the respective state health departments. Valley fever reporting was mandatory for all states in the study area during this time frame. Data was available at the county-scale for California, Arizona, New Mexico, Oregon, and Washington and at the health-district scale for Nevada and Utah. Yearly disease incidence rates per 100,000 population were calculated using linearly interpolated annual population estimates based on the 2010 census count and the 5-year American Community Survey 2015 population estimate (U.S. Census Bureau, 2010, 2015). Annual incidence rates were averaged over the study period to obtain average annual incidence per year for each county (see Figure 2).

County/health-district valley fever case reports were used in place of known *Coccidioides* spp. presence points. The use of this proxy was necessary for two reasons: 1) currently there is no consolidated, georeferenced database of the positive soil samples that have been retrieved and 2) required valley fever reporting spans the endemic region while the limited soil samples do not. While case data are not a direct measure of pathogen presence in the soil of a county, we felt it was reasonable to assume that most cases reported were from those who were exposed to the disease in the same county where they live and receive medical care (Benedict et al., 2018), but we acknowledge that this contributes to model uncertainty. Washington is the only state that records whether cases were believed to be locally acquired or travel related; because we are interested using case data as a proxy for fungal presence, only the locally acquired cases were included.

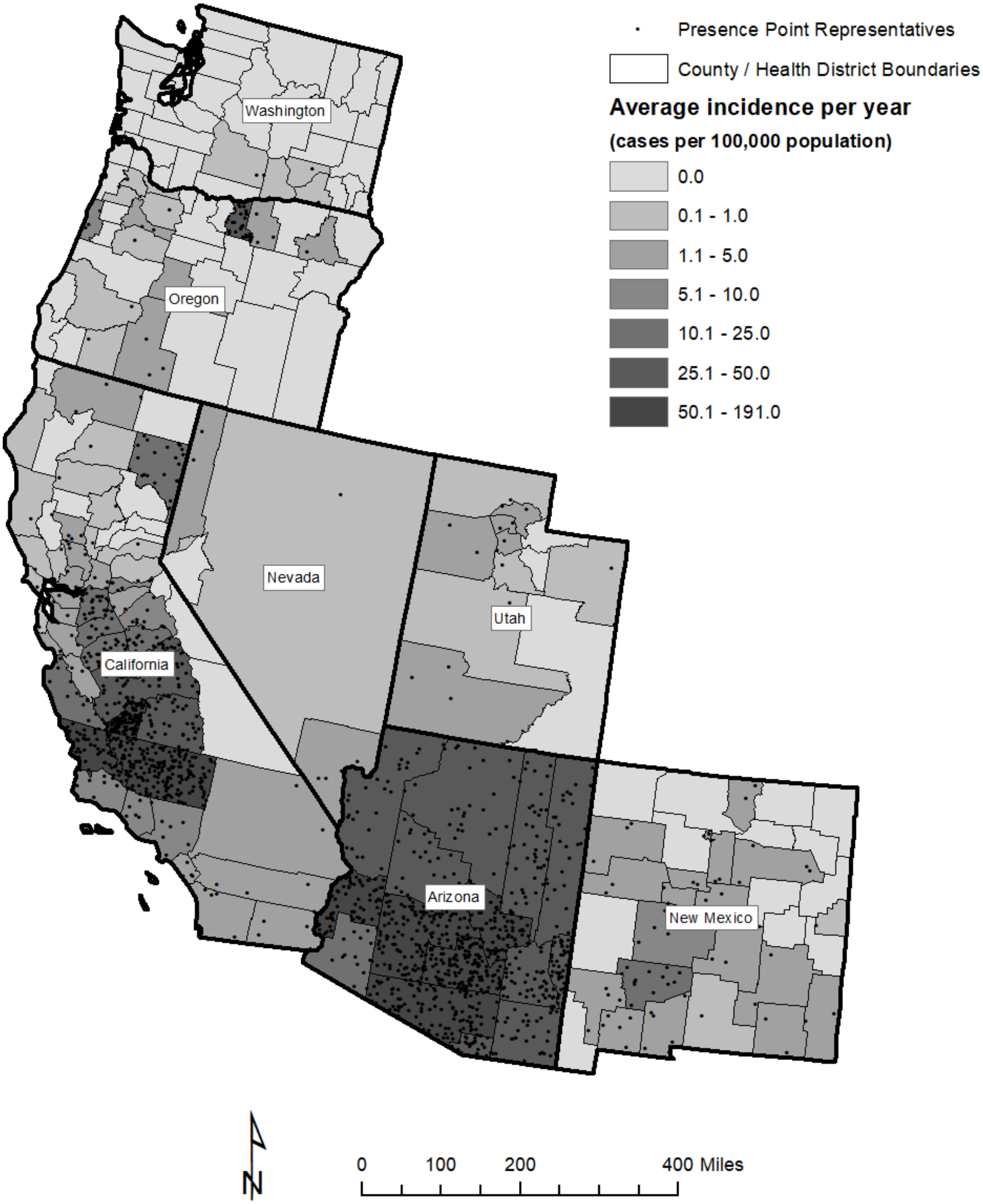
Presence Point Representatives
County / Health District Boundaries
Average incidence per year
(cases per 100,000 population)
0.0
0.1 - 1.0
1.1 - 5.0
5.1 - 10.0
10.1 - 25.0
25.1 - 50.0
50.1 - 191.0
Washington
Oregon
California
Nevada
Utah
Arizona
New Mexico
N
0
100
200
400 Miles

Figure 2. Average valley fever incidence based on data from 2015-2016, with one set of randomly-generated representative presence points.

2.3 Environmental Data

The environmental data used in the niche model represents ecological dimensions that are hypothesized to be relevant to the distribution of *Coccidioides* spp. (see Table 1). The climate data include 19 bioclimatic variables and downward solar radiation from the WorldClim Version 2 dataset at 30 arc-second resolution (Fick & Hijmans, 2017). Soil data are from the Regridded Harmonized World Soil Database v1.2 and include 0.05-degree resolution rasters for pH, percent clay, sand, silt, and topsoil carbon content (Wieder, Boehnert, Bonan, & Langseth, 2014). While the coarse resolution of this dataset is not ideal, this data source has the relevant attributes in a gridded format that does not require significant preprocessing, allowing for easy replication if more precise occurrence data become available. Elevation data, including elevation, slope, aspect, and Compound Topographic Index (also referred to as the Wetness Index) are from the USGS HYDRO1k data set derived from the USGS' 30 arc-second digital elevation model (U.S. Geological Survey, 2015). Land cover data are from the 2011 National Land Cover Database at a spatial resolution of 30-meters (Homer et al., 2015). The land cover categories were aggregated into seven broad habitat classes: water/wetlands, developed, barren, forest, shrubland, herbaceous, and cultivated. County/health district land cover percentages for each class were calculated in ArcMap using zonal statistics and rasters with these values were then created. A raster of percent total land cover change from 2001 to 2011 for each county/health district was also created. While the original land cover data was available at a much higher resolution than the other environmental variables, it was converted to continuous data, aggregated at the county/health-district scale to match the health data, so that the land cover types could be directly compared with the other continuous variables. All environmental data were projected to USA Contiguous Albers Equal Area Conic (NAD 1983) and resampled using nearest-neighbor to a 30 arc-second resolution.

TABLE 1. Environmental Data

Variable	Resolution	Units
Climatic Variables[1]		
Annual Mean Temperature (BIO1)*	30 arc-second	°C*100
Mean Diurnal Range (BIO2)	30 arc-second	°C*100
Isothermality (BIO3)*	30 arc-second	NA
Temperature Seasonality (BIO4)	30 arc-second	°C
Max Temp of Warmest Month (BIO5)	30 arc-second	°C*100
Min Temp of Coldest Month (BIO6)	30 arc-second	°C*100
Temp Annual Range (BIO7)*	30 arc-second	°C*100
Mean Temp of Wettest Quarter (BIO8)*	30 arc-second	°C*100
Mean Temp of Driest Quarter (BIO9)*	30 arc-second	°C*100
Mean Temp of Warmest Quarter (BIO10)	30 arc-second	°C*100
Mean Temp of Coldest Quarter (BIO11)	30 arc-second	°C*100
Annual Precipitation (BIO12)	30 arc-second	Mm
Precip of Wettest Month (BIO13)	30 arc-second	Mm
Precip of Driest Month (BIO14)*	30 arc-second	Mm
Precip Seasonality (BIO15)*	30 arc-second	Mm
Precip of Wettest Quarter (BIO16)*	30 arc-second	Mm
Precip of Driest Quarter (BIO17)	30 arc-second	Mm
Precip of Warmest Quarter (BIO18)*	30 arc-second	Mm
Precip of Coldest Quarter (BIO19)	30 arc-second	Mm
Downward Radiation	30 arc-second	$kJ\ m^{-2}\ day^{-1}$
Soil Variables[2]		
Topsoil clay fraction*	0.05 degree	percent weight
Topsoil silt fraction*	0.05 degree	percent weight
Topsoil sand fraction*	0.05 degree	percent weight
Topsoil carbon content*	0.05 degree	kg C m-2
Topsoil pH (in H_2O)*	0.05 degree	-log(H+)
Elevation Variables[3]		
DEM (elevation)*	1-kilometer	Meters
Slope*	1-kilometer	Degree
Aspect*	1-kilometer	degree (0-360°)
Compound Topographic Index*	1-kilometer	NA
Land Cover Variables[4]		
Proportion Shrub/Scrub*	30-meter	percent area

Proportion Barren*	30-meter	percent area
Proportion Herbaceous*	30-meter	percent area
Proportion Forest*	30-meter	percent area
Proportion Water & Wetland*	30-meter	percent area
Proportion Developed*	30-meter	percent area
Proportion Cultivated*	30-meter	percent area
Proportion Changed from 2001 – 2011*	30-meter	percent area

[1] http://worldclim.org/version2

[2] https://daac.ornl.gov/SOILS/

[3] https://lta.cr.usgs.gov/HYDRO1K

[4] https://www.mrlc.gov/nlcd11data.php

[*] Included in final model

2.4 Data Processing

Previous ecological niche model studies have handled county-level disease presence data by either assigning the geographic coordinates of county centroids or population density centers to each occurrence (Peterson, Pereira, & Neves, 2004; Zeimes et al., 2015), or by plotting random points within each county polygon to represent each occurrence (Nakazawa et al., 2007, 2010; Peterson, Lash, Carroll, & Johnson, 2006). This study used the latter method as it is better suited to represent the variability found in the large counties where the disease is endemic. Specifically, we used 25 random points generated for each case per 100,000 population to develop 25 sets of covariates, to be used in 25 replicate niche models. This case per 100,000 value was based on the previously calculated average incidence per year, rounded up to the nearest whole number to capture all counties that had cases during the study period. This approach represents counties with high incidence more precisely and counties with low incidence with more spatial variation in representative points. The random occurrence points were generated in R using the 'spsample' command from the SP package and verified in ArcGIS version 10.5 (Bivand, Pebesma, & Gomez-Rubio, 2013; Environmental Systems Research Institute, 2017; Pebesma & Bivand, 2005; RStudio Team, 2016). The MaxEnt default of 10,000 background points were randomly selected from the study area. Predictor variables considered for this analysis included the 37 environmental raster layers listed in Table 1. High correlations among predictors can lead to misleading results in the MaxEnt variable contribution reports (Merow, Smith, & Silander, 2013; Phillips, Anderson, & Schapire, 2006), therefore variables were assessed for collinearity and reduced based on a Pearson correlation coefficient $r > 0.80$ ($\alpha = 0.05$). Of the highly correlated variables (all climatic), those retained were selected based on findings from past research or ease in model interpretation. This resulted in the inclusion of 26 predictor variables (denoted by * in Table 1).

2.5 Ecological Niche Modeling

The MaxEnt modeling in this study was performed in R using the DISMO package (Hijmans, Phillips, Leathwick, & Elith, 2017). The model was run through 25 iterations, once for each of the 25 sets of occurrence points; the background points used to represent the overall study area were the same for each model iteration. The MaxEnt default features were limited to linear, quadratic, and product to remove highly nonlinear variable response curves and improve our ability to interpret the species' response to the predictor of interest (Merow et al., 2013). We also increased the regularization coefficients by 10% to force the algorithm to focus on the most important features and reduce overfitting by relaxing the empirical constraints on the model (Merow et al., 2013). Each model was evaluated using a 10-fold cross validation (0.01 convergence limit and 1000 maximum iterations). Overall model quality was assessed through analysis of the mean area under the receiver operating characteristic curve (or AUC). AUC is a measure of how well the model separates presence and background locations; a value of 0.5 indicates that the model performs no better than a random model and 1 indicates perfect accuracy. AUC is commonly used in ecological niche model comparisons (Elith et al., 2006; Frans et al., 2018; Padalia, Srivastava, & Kushwaha, 2014; Phillips et al., 2006; Warren & Seifert, 2011), but has been critiqued with concerns that it lacks an indication of model fit and has biased values with larger background extents generally having higher AUC values (Jiménez-Valverde, Acevedo, Barbosa, Lobo, & Real, 2013; Lobo, Jiménez-Valverde, & Real, 2008). Predictor variables were assessed by averaging the 25 reports generated by MaxEnt including 1) variable percent contributions and permutation importance, 2) jackknife tests of model gains for three scenarios (without variable, with only one variable, and with all variables), and 3) variable response curves. Finally, the raw output data from the model predictions was combined by averaging each pixel to produce an estimate of potential distribution. The raw output represents the probability, given the species is present, that is it found at the respective site; it should be interpreted as a relative suitability, not an occurrence probability (Guillera-Arriota et al., 2014; Phillips et al., 2006).

3. Results

3.1 Variable Assessment

Three MaxEnt reports were used to estimate the effect of each variable on the spatial distribution of *Coccidioides* spp. The results reported here focus on the top five variables in each assessment. The first assessment report showed which variables contributed the most to the model. This report had two categories: variable contribution to the final model and permutation importance. Permutation importance indicates variables that contributed the most regardless of the order in which they added to the model. Variables common to both categories included downward radiation, mean temperature of driest quarter, and proportion shrub land cover (see Table 2). The second variable assessment report included results of jackknife tests on the variables. When the model was run with only one selected variable at a time, climate variables were the most influential, meaning these variables contain the most information by themselves (see Table 3). The model was also run with all variables except the selected variable. Model performance decreases if the selected variable contains information not found in other variables. Model performance decreased for proportion barren, downward radiation, proportion shrub, and proportion cultivated (see Table 4). The final variable assessment report consisted of variable response curves that show how the predicted probability of presence changes as the environmental variable changes. Two sets of response curves were assessed- one with other variables held constant, and one using only the selected variable (not shown). All precipitation variables except seasonality had generally negative relationships, and all temperature variables had positive relationships, with the exception of temperature range which had no relationship with probability of presence. Land cover relationships showed a negative curve between probability of presence and proportion barren, and a positive curve for proportion cultivated and proportion shrub.

Table 2. Top variables contributing to model development.

Rank	Variable Contribution	%	Permutation Importance	%
1	Mean Temperature of Driest Quarter	22.8	Downward Radiation	41.3
2	Annual Mean Temp	18.5	Mean Temperature of Driest Quarter	19.2
3	Downward Radiation	14.7	Proportion Cultivated	14.4
4	Precipitation Seasonality	9.5	Proportion Barren	12.7

5	Proportion Shrub	7.8	Proportion Shrub	4.3

Table 3. Top variables from model jackknife tests using only the selected variable.

Rank	Training	Gain	Testing	Gain	AUC	Value
1	Annual Mean Temp	0.27	Annual Mean Temp	0.31	Annual Mean Temp	0.73
2	Mean Temperature of Driest Quarter	0.27	Precipitation of Driest Month	0.31	Precipitation of the Driest Month	0.72
3	Precipitation of Driest Month	0.22	Mean Temperature of Driest Quarter	0.30	Mean Temperature of Driest Quarter	0.71
4	Down. Radiation	0.18	Down. Radiation	0.21	Isothermality	0.69
5	Precipitation Seasonality	0.13	Precipitation Seasonality	0.15	Proportion Developed	0.68

Table 4. Top variables from model jackknife tests without the selected variable.

Rank	Training	Gain Lost	Testing	Gain Lost
1	Down. Radiation	0.05	Proportion Barren	0.06
2	Proportion Barren	0.04	Down. Radiation	0.04
3	Proportion Shrub	0.04	Proportion Shrub	0.04
4	Mean Temperature of Driest Quarter	0.02	Proportion Cultivated	0.03
5	Proportion Cultivated	0.02	Herbaceous	0.02

3.2 Potential Distribution

The raw output from the MaxEnt prediction is equivalent to the relative occurrence rate, it shows relative habitat suitability for each pixel. The map produced from averaging the raw output captures the areas around the counties with the highest incidence rates, particularly in south central Arizona and central California, as having the highest relative occurrence rates (see Figure 3). Southcentral Washington and parts of northern Oregon stand out from the rest of the Pacific Northwest as having slightly higher rates than surrounding areas. Low rates of relative occurrence can be seen in the Mojave Desert of southeastern California, the northern coastlines, and throughout most of Nevada, Utah, Oregon, and western Washington. While the output from the prediction is continuous, some county borders are visible due to the summation and influence of land cover variables at the county/health district spatial scale.

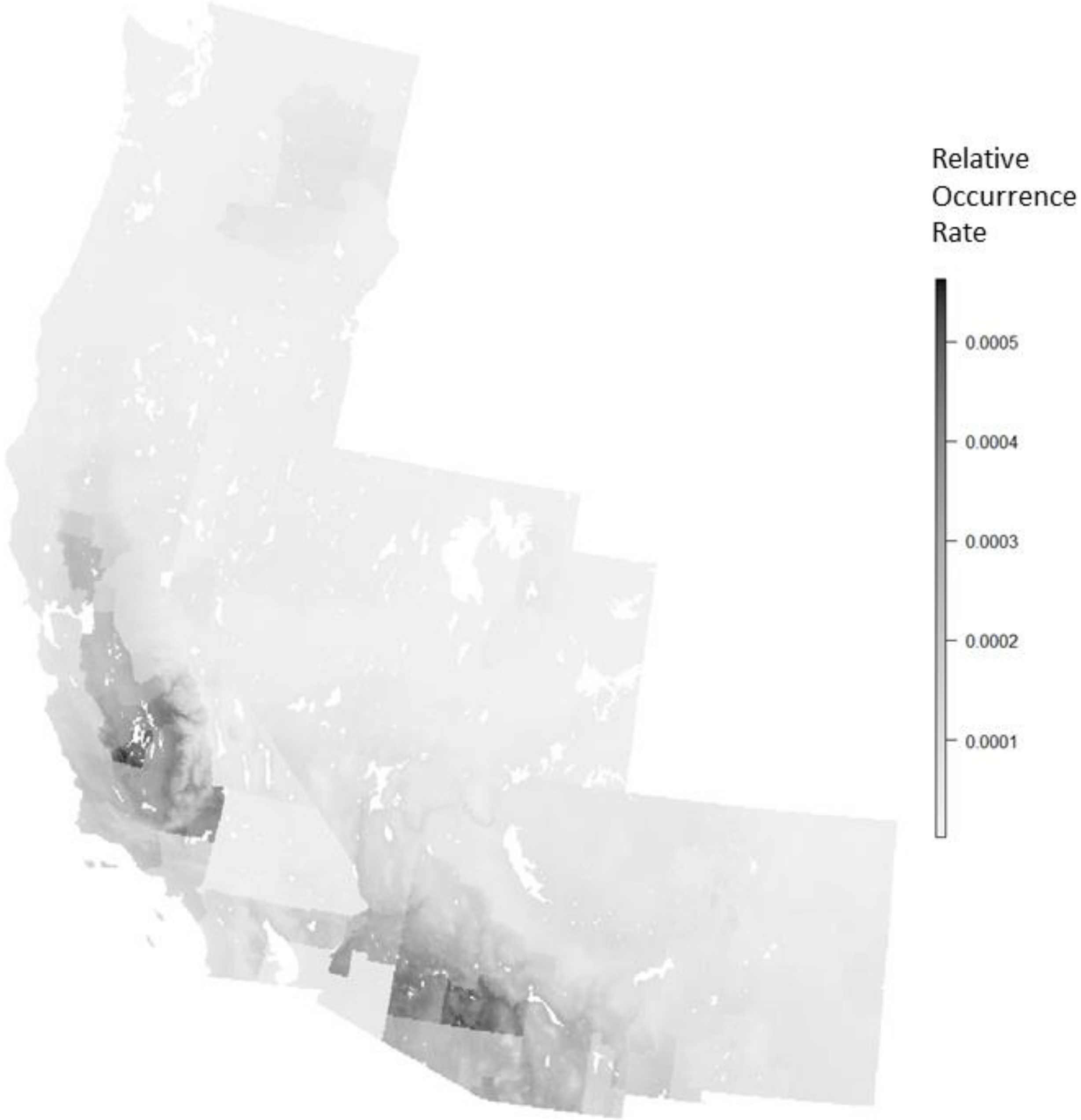

Figure 3. Map of potential distribution based on average raw output from model predictions. Values indicate the probability, given the species is present, that is it found at that location; the individual pixel values sum to unity over the entire landscape.

3.3 Model Evaluation

Model evaluation results show that the average testing AUC value of 0.819 was only slightly below the average training value of 0.821; these are considerably higher than the null model of 0.5. These results indicate that the environmental variables used in this study were able to partially explain the spatial distribution of valley fever. Additionally, there was consistency between the 25 sets of models as indicated by low standard deviations values for both testing and training AUCs.

4. Discussion

This research aimed to improve our understanding of factors affecting *Coccidioides* species' ecological niche and the spatial distribution of valley fever through ecological niche modeling. This modeling framework allowed us to conduct an informed assessment of environmental factors influencing *Coccidioides* spp. occurrence and to produce a potential distribution map based on environmental inputs. Significant findings include the identification of specific land cover types and climatic variables, including downward radiation, mean temperature of driest quarter, and proportion shrub land cover as influential factors. A significant finding from the potential distribution map is that southeastern Washington and northcentral Oregon, generally aligning with the Columbia Plateau ecoregion, have higher rates of relative occurrence than surrounding areas indicating environmental conditions more suitable for *Coccidioides* spp. This is the first time Oregon and Washington have been included in such an assessment. Based on the results of this research, we can characterize the probable realized niche for *Coccidioides* spp., and therefore the probable U.S. endemic region for valley fever, as areas in the western U.S. that are semi-arid with a hot-dry season supporting shrub vegetation and/or cultivated land cover.

The variable assessment results showed that of the four broad environmental categories hypothesized to affect *Coccidioides* species' ecological niche, climate and land cover had the greatest effects, while soil and elevation variables were less influential. Specifically, temperature-related variables accounted for over 50% of the variable contribution to model development and permutation importance. Researchers have hypothesized that *Coccidioides* spp. grow best in areas where a hot and dry season sterilizes the top layer of soil, making it inhospitable to many microorganisms (Egeberg & Ely, 1956; Egeberg, 1962; Maddy, 1965). It is believed *Coccidioides* spp. survive hot and dry seasons by moving deeper into the soil, then return to the relatively competitor-free surface when rains return (Sorensen, 1967). The presence of the mean temperature of the driest quarter, precipitation of the driest month, and precipitation seasonality variables as top contributors in many of the assessments, with a positive response curve for temperature and a negative response curve for precipitation, support this hypothesis. This also aligns with a "grow and blow" hypothesis that theorizes alternating cool/wet and hot/dry seasons support pathogen growth and disturbance/dispersal, respectively (Comrie & Glueck, 2007). Solar radiation was also a very prominent variable in many of the assessments. Researchers have observed that when *Coccidioides* spp. is in the form of an arthroconidia, spherule, or endospore, it has a biological defense, an ability to deposit melanin within its cell walls, that protects it from extreme temperatures and UV radiation (Nosanchuk, Yu, Hung, Casadevall, & Cole, 2007; Taborda, da Silva, Nonsanchuk, & Travassos, 2008). This likely gives the fungus an additional survival advantage that other competitors may not have during the hot and dry season; competitors die off without protection from radiation and then *Coccidioides* spp. are able to grow in a relatively competitor-free environment.

Land cover variables were also prominent in some of the variable assessment reports; they accounted for approximately 7% of model development and 30% permutation importance. In a comprehensive review of attributes from nine sites that have tested positive for *Coccidioides* spp. presence in the soil, Fisher et al. (2007) found that there was no definitive vegetation types or densities common among sites. But landcover classes, though many are defined by vegetation, represent broader ecologies with interactions among vegetation, soil, climate, and human activity, all of which may affect *Coccidioides* species' lifecycle. The variable response curves showed positive relationships between probability of presence and proportions shrub and cultivated land cover. The relationship with cultivated land cover aligns with recent research (Colson et al., 2017; Gorris, Cat, Zender, Treseder, & Randerson, 2017), but is at odds with past findings that concluded *Coccidioides* spp. does not grow well in cultivated soils possibly due to microbial competitors or fungicides (Maddy, 1958; Pappagianis, 1988; Swatek, 1970). It might be that fallow agricultural fields in these regions are supporting pathogen growth. Counties with a significant proportion of barren land cover, meaning little to no green vegetation, had reduced probability of presence. It is possible that this land cover may not contain the types of nutrients that *Coccidioides* spp. need for survival and growth and/or that this land cover is associated with climates that are too hot and dry for the fungus. Of note, the greatest model gains were not from individual variables, but the result of interactions, mainly between various climatic variables and between climate and land cover types; such interactions warrant further study. Variables not significant to model development or not found to contain significant information by themselves include most of the soil and elevation variables; the resolution of the soil variables may have been too coarse, or they may just not be as influential on the distribution of *Coccidioides* spp. as climate and land cover at the scale of this analysis.

The potential distribution map generated from the MaxEnt model provides an indication of *Coccidioides* species' realized niche and allows generation of new hypothesis regarding other factors that might influence the disease distribution. Although evaluation metrics show the model was able to satisfactorily differentiate between presence and background locations, this map should not be interpreted as the definitive range of *Coccidioides* spp., but rather as a guide for further evaluations and field studies. A visual assessment of the map shows that it captured the most endemic regions of central California and southcentral Arizona as having the highest relative rates of occurrence. Based on the patterns shown, we can characterize the probable realized niche for *Coccidioides* spp. as North American shrublands, including those that have been altered for cultivation, with semi-arid climates that include a very hot, dry season. Overall, this characterization is fairly similar to the findings from 50 years ago made based on observations (Maddy & Coccozza, 1964), but provides more specific information on probable land cover associations. It should be noted that because land cover percentages were calculated at the county/health district scale, results in the probable distribution map highlight counties with similar land cover proportions rather than the actual geographic locations of associated land covers.

Of interest to this study were the rates of relative occurrence in Oregon and Washington, which have received little attention in valley fever studies. The region in Oregon and Washington that shows higher relative occurrence rates generally aligns with the Columbia Plateau ecoregion, characterized by a semi-arid climate that supports native shrub-steppe and other drought-tolerant plant communities with over half of the native shrub-steppe currently converted to agriculture (https://waconnected.org/columbia-plateau-ecoregion/). Not surprisingly, these characteristics are similar to what we find in the highly endemic areas of central California and southcentral Arizona, making this region a great candidate for increased valley fever surveillance and awareness campaigns. The probable distribution map can also be used to make informed hypotheses regarding other factors that may be influential that were not included in this study. For example, based on patterns observed, future studies might consider including variables such as predominant winds and proximity to stream networks.

The findings of this study are subject to several limitations. In using case data, we assume that reported cases are geographically connected to pathogen presence, but it is likely that some of the cases are travel-related and not acquired in the county in which they were reported. As such distinctions are not maintained in most states, this adds uncertainty to our model. Counties represented in the model with presence points that may have had only travel related cases might show areas within the county as having erroneously high relative habitat suitability. Conversely, there may be counties that have suitable habitat not indicated on the map if exposure occurred in that county, but the report was made elsewhere. For context, a recent study by Benedict et al. (2018), reported on enhanced surveillance of coccidioidomycosis in which they conducted in-depth interviews with patients; 64 patients were from Nevada, New Mexico, or Utah, and 26 of them (37.5%) reported traveling to known endemic areas in the 4 months before symptom onset. Based on this study, the implications are possible overestimations of potential suitable areas, and some inaccuracies in variable importance rankings, but by using presence points based on incidence rates, we believe the model captured the most endemic areas with the highest incidence rates very well with variable importance results highly dependent on these locations.

Another limitation to this study is the scale of analysis. Ideally, research would be conducted at the scale at which the organism interacts with limiting environmental resources (Cushman & Huettmann, 2010). Because we are working with case data collected at the county/health district level, we are restricted to working only at courser resolutions, though this study attempted to address this scale issue by using random points within the counties to represent pathogen presence at a finer scale. The implications of using a coarser scale are that precise features of distribution can be lost and these resolutions also tend to overestimate potential suitable areas when compared to predictions at finer-scales (Wiens, Stralberg, Jongsomjit, Howell, & Snyder, 2009). Additionally, there is a slight discrepancy between temporal scale of disease data and the climatic data used in this research; the climate variables were averaged from 1970 to 2000 and the disease data were averaged from 2015-2016. While climate is typically averaged over 30-year periods and averages tend to change slowly, the magnitudes of climate-disease connections found in this study may not reflect current associations and adaptations, but general (positive or negative) relationships are not expected to be affected.

It was our intent to be cautious and conservative in interpreting the results of this research, as a significant assumption is being made in using valley fever case data to assess *Coccidioides* species' niche. Additionally, with ecological niche modeling, as with all modeling, results are dependent on assumptions made in model selection and parametrization; different inputs and assumptions will produce different results. Our aim was to draw conclusions regarding the spatial and ecological distribution of valley fever and the identification of factors influencing the distribution that are generalizable, and not the result of the modeling process. Results could be refined in the future if there are more precise locations of exposure recorded, widespread skin testing, or improved environmental detection of the fungus as technologies continue to advance (Benedict et al., 2015).

5. Conclusion

This study confirmed the importance of environmental drivers such as temperature and land cover on the spatial distribution of this disease. The MaxEnt algorithm used the provided environmental variables to capture the spatial patterns observed in valley fever case reporting, and it identified the understudied Columbia Plateau region as a possible habitat for *Coccidioides* spp. This information can be applied by stakeholders ranging from other coccidioidomycosis researchers to public health officials in assessing vulnerability, refining endemicity, and in watching for potential hotspots. Furthermore, the uncovered relationships between the spatial distribution of *Coccidioides* spp. and the environmental variables used in this study will be informative to the development of predictive models in assessing how disease distribution may change with varying climate, land cover, and population.

References

Barker, B. M., Tabor, J. A., Shubitz, L. F., Perrill, R., & Orbach, M. J. (2012). Detection and phylogenetic analysis of Coccidioides posadasii in Arizona soil samples. *Fungal Ecology*, *5*, 163–176. https://doi.org/10.1016/j.funeco.2011.07.010

Benedict, K., Ireland, M., Weinberg, M. P., Gruninger, R. J., Weigand, J., Chen, L., … Jackson, B. R. (2018). Enhanced Surveillance for Coccidioidomycosis, 14 US States, 2016. *Emerging Infectious Diseases*, *24*(8).

Benedict, K., Thomspon, G. R. I., Deresinski, S., & Chiller, T. (2015). Mycotic Infections Acquired outside Areas of Known. *Emerging Infectious Diseases*, *21*(11), 1935–1941. https://doi.org/10.3201/eid2111.141950

Bivand, R. S., Pebesma, E. J., & Gomez-Rubio, V. (2013). *Applied spatial data analysis with R, Second edition*. (Springer, Ed.). New York. Retrieved from http://www.asdar-book.org/

Centers for Disease Control and Prevention. (2018). Valley Fever (Coccidioidomycosis) Statistics. Retrieved January 4, 2018, from https://www.cdc.gov/fungal/diseases/coccidioidomycosis/statistics.html

Colson, A. J., Vredenburgh, L., Guevara, R. E., Rangel, N. P., Kloock, C. T., & Lauer, A. (2017). Large-Scale Land Development, Fugitive Dust, and Increased Coccidioidomycosis Incidence in the Antelope Valley of California, 1999-2014. *Mycopathologia*, *182*, 1–20. https://doi.org/10.1007/s11046-016-0105-5

Comrie, A. C., & Glueck, M. F. (2007). Assessment of Climate-Coccidioidomycosis Model: Model Sensitivity for Assessing Climatologic Effects on the Risk of Acquiring Coccidioidomycosis. *Annals of the New York Academy of Sciences*, *1111*, 83–95. https://doi.org/10.1196/annals.1406.024

Cushman, S. A., & Huettmann, F. (2010). Spatial complexity, informatics, and wildlife conservation. In *Spatial Complexity, Informatics, and Wildlife Conservation* (pp. 1–458). https://doi.org/10.1007/978-4-431-87771-4

Egeberg, R. O. (1962). Factors Influencing the Distribution of Coccidioides Immitis in Soil. In *Recent Progress in Microbiology, 6th Meeting* (pp. 652–655).

Egeberg, R. O., & Ely, A. F. (1956). Coccidioides Immitis in the SOil of hte Southern San Joaquin Valley. *The American Journal of the Medical Sciences*, *231*(2), 151–154.

Elconin, A. F., Egeberg, R. O., & Egeberg, M. C. (1964). Significance of soil salinity on the ecology of Coccidioides immitis. *Journal of Bacteriology*, *87*(3), 500–503.

Elith, J., Graham, C. H., Anderson, R. P., Dudík, M., Guisan, A., Hijmans, R. J., … Zimmermann, N. E. (2006). Novel Methods Improve Prediction of Species' Distributions from Occurrence Data. *Ecography*, *29*(2), 129–151.

Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., & Yates, C. J. (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, *17*(1), 43–57. https://doi.org/10.1111/j.1472-4642.2010.00725.x

Environmental Systems Research Institute. (2017). ArcMap Release 10.5. Redlands, CA.

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12), 4302–4315. https://doi.org/10.1002/joc.5086

Fisher, F. S., Bultman, M. W., Johnson, S. M., Pappagianis, D., & Zaborsky, E. (2007). Coccidioides niches and habitat parameters in the southwestern United States: A matter of scale. *Annals of the New York Academy of Sciences*, *1111*, 47–72. https://doi.org/10.1196/annals.1406.031

Frans, V. F., Augé, A. A., Edelhoff, H., Erasmi, S., Balkenhol, N., & Engler, J. O. (2018). Quantifying apart what belongs together: A multi-state species distribution modelling framework for species using distinct habitats. *Methods in Ecology and Evolution*, *9*(1), 98–108. https://doi.org/10.1111/2041-210X.12847

Gorris, M. E., Cat, L. A., Zender, C. S., Treseder, K. K., & Randerson, J. T. (2017). Coccidioidomycosis dynamics in relation to climate in the southwestern United States. *GeoHealth*, 1–19. https://doi.org/10.1002/2017GH000095

Greene, D., Koenig, G., Fisher, M. C., & Taylor, J. W. (2000). Soil isolation and molecular identification of Coccidioides immitis. *Mycologia*, *92*(3), 406–410.

Guillera-Arriota, G., Lahoz-Monfort, J. J., & Elith, J. (2014). Maxent is not a presence – absence method : a comment on Thibaud et al . *Methods in Ecology and Evolution*, *5*, 1192–1197. https://doi.org/10.1111/2041-210X.12252

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

Homer, C. G., Dewitz, J. A., Yang, L., Jin, S., Danielson, P., Xian, G., … Megown, K. (2015). Completion of the 2011 National Land Cover Database for the conterminous United States-Representing a decade of land cover change information. *Photogrammetric Engineering and Remote Sensing*, *81*(5), 345–354.

Huang, J. Y., Bristow, B., Shafir, S., & Sorvillo, F. (2012). Coccidioidomycosis-associated Deaths, United States, 1990–2008. *Emerging Infectious Diseases*, *18*(11), 1723–1728. https://doi.org/10.3201/eid1811.120752

Jaynes, E. T. (1957). Information Theory and Statistical Mechanics. *Physical Review*, *108*, 171–190.

Jiménez-Valverde, A., Acevedo, P., Barbosa, A. M., Lobo, J. M., & Real, R. (2013). Discrimination capacity in species distribution models depends on the representativeness of the environmental domain. *Global Ecology and Biogeography*, *22*(4), 508–516. https://doi.org/10.1111/geb.12007

Litvintseva, A. P., Marsden-Haug, N., Hurst, S., Hill, H., Gade, L., Driebe, E. M., … Chiller, T. (2015). Valley Fever: Finding New Places for an Old Disease: Coccidioides immitis Found in Washington State Soil Associated With Recent Human Infection. *Clinical Infectious Diseases*, *60*(1), e1-3. https://doi.org/10.1093/cid/ciu681

Lobo, J. M., Jiménez-Valverde, A., & Real, R. (2008). AUC: A misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, *17*(2), 145–151. https://doi.org/10.1111/j.1466-8238.2007.00358.x

Maddy, K. T. (1958). The geographic distribution of Coccidioides Immitis and possible ecologic implications. *Arizona Medicine: Journal of Arizona State Medical Association*, *15*(3), 178–

188. https://doi.org/10.1017/S000748530002229X

Maddy, K. T. (1965). Observations on Coccidioides Immitis Found Growing Naturally in Soil. *Arizona Medicine: Journal of Arizona State Medical Association*, *22*, 281–288. https://doi.org/10.1017/S000748530002229X

Maddy, K. T., & Coccozza, J. (1964). The Probable Geographic Distribution of Coccidioides Immitis in Mexico. *Boletin de La Oficina Sanitaria Panamericana. Pan American Sanitary Bureau*, *57*, 44–54. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14175564

Marsden-Haug, N., Goldoft, M., Ralston, C., Limaye, A. P., Chua, J., Hill, H., … Chiller, T. (2013). Coccidioidomycosis acquired in Washington State. *Clinical Infectious Diseases*, *56*(6), 847–850. https://doi.org/10.1093/cid/cis1028

Merow, C., Smith, M. J., & Silander, J. A. (2013). A practical guide to MaxEnt for modeling species' distributions: What it does, and why inputs and settings matter. *Ecography*, *36*(10), 1058–1069. https://doi.org/10.1111/j.1600-0587.2013.07872.x

Nakazawa, Y., Williams, R. A. J., Peterson, A. T., Mead, P. S., Kugeler, K. J., & Petersen, J. M. (2010). Ecological niche modeling of Francisella tularensis subspecies and clades in the United States. *American Journal of Tropical Medicine and Hygiene*, *82*(5), 912–918. https://doi.org/10.4269/ajtmh.2010.09-0354

Nakazawa, Y., Williams, R., Peterson, A. T., Mead, P., Staples, E., & Gage, K. L. (2007). Climate Change Effects on Plague and Tularemia in the United States. *Vector-Borne and Zoonotic Diseases*, *7*(4), 529–540.

Nosanchuk, J., Yu, J., Hung, C., Casadevall, A., & Cole, G. (2007). Coccidioides posadasii produces melanin in vitro and during infection. *Fungal Genet Biol*, *44*(6), 517–20.

Padalia, H., Srivastava, V., & Kushwaha, S. P. S. (2014). Modeling potential invasion range of alien invasive species, Hyptis suaveolens (L.) Poit. in India: Comparison of MaxEnt and GARP. *Ecological Informatics*, *22*, 36–43. https://doi.org/10.1016/j.ecoinf.2014.04.002

Pappagianis, D. (1988). Epidemiology of Coccidioidomycosis. *Current Topics in Medical Mycology*, *2*, 199–238. https://doi.org/10.1017/S000748530002229X

Pebesma, E. J., & Bivand, R. S. (2005). Classes and methods for spatial data in R. *R News*, *5*(2).

Peterson, a. (2006). Ecologic Niche Modeling and Spatial Patterns of Disease Transmission. *Emerging Infectious Diseases*, *12*(12), 1822–1826. https://doi.org/10.3201/eid1212.060373

Peterson, A. T., Lash, R. R., Carroll, D. S., & Johnson, K. M. (2006). Geographic potential for outbreaks of Marburg hemorrhagic fever. *American Journal of Tropical Medicine and Hygiene*, *75*(1), 9–15. https://doi.org/10.4269/ajtmh.2006.75.1.0750009

Peterson, T. A., Pereira, R., & Neves, V. (2004). Using epidemiological survey data to infer geographic distributions of leishmaniasis vector species. *Revista Da Sociedade Brasileira de Medicina Tropical*, *37*(1), 10–14. https://doi.org/10.1590/s0037-86822004000100003

Phillips, S. B., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, *190*, 231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

Raghavan, R. K., Goodin, D. G., Hanzlicek, G. A., Zolnerowich, G., Dryden, M. W., Anderson, G. A., & Ganta, R. R. (2016). Maximum Entropy-Based Ecological Niche Model and Bio-Climatic Determinants of Lone Star Tick (*Amblyomma americanum*) Niche. *Vector-Borne and Zoonotic Diseases*, *16*(3), 205–211. https://doi.org/10.1089/vbz.2015.1837

RStudio Team. (2016). RStudio: Integrated Development for R. Version 1.1.423. Boston, MA: RStudio, Inc. Retrieved from http://www.rstudio.com/

Saubolle, M. A., McKellar, P. P., & Sussland, D. (2007). Epidemiologic, clinical, and diagnostic aspects of coccidioidomycosis. *Journal of Clinical Microbiology*, *45*(1), 26–30. https://doi.org/10.1128/JCM.02230-06

Sorensen, R. H. (1967). Survival Characteristics of Diphasic Coccidioides Immitis Exposed to the Rigors of a Simulated Natural Environment. In *Proc. 2nd Coccidioidomycosis Symp.* (pp. 313–317). Tucson, AZ: Univ. of Arizona Press.

Swatek, F. E. (1970). Ecology of Coccidioides immitis. *Mycopathologia et Mycologia Applicata*, *40*(1–2), 3–12.

Swatek, F. E., Omieczynski, D. T., & Plunkett, O. A. (1967). Coccidioides immitis in California.

In *Papers from the 2nd Symposium on Coccidioidomycosis.* (pp. 255–265). Tucson, AZ: University of Arizona Press.

Taborda, C., da Silva, M., Nonsanchuk, J., & Travassos, L. (2008). Melanin as a virulence factor of Paracoccidioides brasiliensis and other dimorphic pathogenic fungi: a minireview. *Mycopathologia*, *165*(4–5), 331.

U.S. Census Bureau. (2010). Profile of General Population Characteristics: 2010, 2010 Census Summary File 1 Data. Retrieved June 5, 2017, from https://factfinder.census.gov/

U.S. Census Bureau. (2015). ACS Demographic and Housing Estimates, 2011-2015 American Community Survey 5-Year Estimates. Retrieved June 5, 2017, from https://factfinder.census.gov/

U.S. Geological Survey. (2015). HYDRO1K Dataset. Retrieved September 10, 2018, from https://lta.cr.usgs.gov/HYDRO1K

Warren, D. L., & Seifert, S. N. (2011). Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. *Ecological Applications*, *21*(2), 335–342.

Wieder, W. R., Boehnert, J., Bonan, G. B., & Langseth, M. (2014). Regridded Harmonized World Soil Database v1.2. ORNL DAAC, Oak Ridge, Tennessee, USA. Retrieved from https://doi.org/10.3334/ORNLDAAC/1247

Wiens, J. A., Stralberg, D., Jongsomjit, D., Howell, C. A., & Snyder, M. A. (2009). Niches, models, and climate change: Assessing the assumptions and uncertainties. *Proceedings of the National Academy of Sciences*, *106*(Supplement_2), 19729–19736. https://doi.org/10.1073/pnas.0901639106

Zeimes, C. B., Quoilin, S., Henttonen, H., Lyytikäinen, O., Vapalahti, O., Reynes, J.-M., … Vanwambeke, S. O. (2015). Landscape and Regional Environmental Analysis of the Spatial Distribution of Hantavirus Human Cases in Europe. *Frontiers in Public Health*, *3*(March). https://doi.org/10.3389/fpubh.2015.00054

Acknowledgments:

We thank the Arizona, California, Nevada, New Mexico, Oregon, Utah, and Washington Departments of Health for providing us with valley fever case data from their respective state health agencies. Valley fever data may be obtained from the affiliated state health agencies. KMA is supported by NIH R01GM109718.

Declaration of Interest: The authors declare no competing financial interests.

Appendix A – List of Tables and Figures

FIGURE 1. CDC map showing approximate areas where Coccidioides spp. are known or suspected to live in the U.S. and Mexico. This map is based on studies performed in the late 1940s and 1950s and also on locations of more recent outbreaks and cases.

FIGURE 2. Average valley fever incidence based on data from 2015-2016, with one set of randomly generated representative presence points.

FIGURE 3. Map of potential distribution based on average raw output from model predictions. Values indicate the probability, given the species is present, that is it found at that location; the individual pixel values sum to unity over the entire landscape.

TABLE 1. Environmental Data.

TABLE 2. Top variables contributing to model development.

TABLE 3. Top variables from model jackknife tests using only the selected variable.

TABLE 4. Top variables from model jackknife tests without the selected variable.