Investigating the Valley Fever – Environment Relationship in the Western U.S.

Elizabeth Ann Weaver

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State
University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Geospatial and Environmental Analysis

Korine N. Kolivras, Chair
Kaja M. Abbas
Valerie A. Thomas
R. Quinn Thomas

March 26, 2019
Blacksburg, Virginia

Keywords: Valley fever, coccidioides, modeling, disease distribution, climate and health

Investigating the Valley Fever – Environment Relationship in the Western U.S.

Elizabeth Ann Weaver

ABSTRACT

Valley fever, or coccidioidomycosis, is a disease caused by the *Coccidioides* immitis and *Coccidioides* posadasii fungal species that dwell in the soil but can become airborne and infect a human or mammalian host through their respiratory tract. Disease rates in the western U.S. have significantly increased over the past two decades, creating an emerging public health burden. Studies have been conducted that attempt to elucidate the association between environmental conditions and the growth and dispersal of the pathogen, yet the specific ecology of and environmental precursors to the disease remain uncertain.

This research project investigates the relationship between environmental variables and valley fever by modeling the spatial and temporal dynamics of the disease using varying techniques. Chapter 1 discusses relevant literature before discussing the challenges associated with studying valley fever. Chapter 2 analyzes the temporal relationships between valley fever and climatic variables, focusing on Kern County, California, an understudied region in the U.S. where valley fever is highly endemic. Chapter 3 focuses on a regional spatial analysis using ecological niche modeling to better understand the environmental factors that influence the overall spatial distribution of valley fever in the U.S. Finally, combining both spatial and temporal components, Chapter 4 uses a hierarchical Bayesian spatio-temporal model to investigate the patterns and drivers of this disease, focusing on state of California, which saw an approximate 200% increase in cases from 2014 to 2018.

Cumulatively, this work offers new insights on relationships between climate, landcover, and valley fever disease risk. Significant findings include climate variables explaining up to 76% of valley fever variability in Kern County, California, the significance of both climatic and landcover variables in characterizing the geographic distribution of the disease, and identification of patterns increasing risk in geographic regions of California not currently considered highly endemic. These findings advance scholarly understandings of valley fever's environmental disease drivers. The results of this research can be applied by public health officials in the allocation of surveillance and public education resources, focusing upon regions that are most likely to encounter the illness.

Investigating the Valley Fever – Environment Relationship in the Western U.S.

Elizabeth Ann Weaver

GENERAL AUDIENCE ABSTRACT

Valley fever is a fungal disease that causes illness in over ten thousand people in the western U.S. every year. Disease rates have been increasing for the past two decades for unknown reasons, although previous research suggests that climatic variations are likely contributing factors. This research evaluated environmental factors with hypothesized relationships to valley fever disease rates. First, this dissertation explored time-series relationships between climatic factors and valley fever incidence in an understudied county in California. Research findings identified that climatic factors including precipitation from previous seasons and temperature were significantly associated with valley fever incidence in this county. Second, this dissertation assessed where valley fever is found in the western U.S. The likely spatial distribution for the disease was mapped and environmental variables influential to this distribution were identified; they included both climate and landcover variables. Finally, a model was developed to analyze patterns of disease risk in California that considered both space and time, and environmental risk factors potentially contributing to the observed patterns were assessed. Counties with increased risk were identified and significant environmental relationships with valley fever risk were confirmed. The results of this research can be applied by public health officials in allocating surveillance and public education resources, focusing upon regions that are most likely to encounter the illness.

Acknowledgements

I would like to express my sincerest gratitude to all the people who helped me at various stages of my studies. First, I want to thank my advisor and mentor, Dr. Korine Kolivras, whose guidance, encouragement, and support were instrumental to my progress. I am also grateful to my committee members, Dr. R. Quinn Thomas, Dr. Kaja Abbas, and Dr. Valerie Thomas, whose insightful comments and suggestions motivated me to think more deeply about my research. I am especially grateful to my husband and children who never complained about the hours I spent at my desk and were a constant source of encouragement and support; my friend Kathryn who regularly called to check in, listen, and reassure; my friends Misty, Katie, and Stef who gave me perspective; and the crew from Williamsburg who encouraged me to take breaks and have fun.

Attribution

Several committee members aided in the research presented as part of this dissertation. A brief description of their contributions is included here.

Chapter 2: Investigating the climate – valley fever relationship.
Dr. Korine Kolivras served as co-author on this paper and helped with research design and editing.

Chapter 3: Environmental factors affecting the ecological niche of Coccidioides species.
Dr. Valerie Thomas, Dr. Kaja Abbas, and Dr. Kolivras assisted with research design and editing. Dr. Quinn Thomas provided coding assistance.

Chapter 4: Spatiotemporal modeling of relative risk of valley fever.
Dr. Quinn Thomas assisted with research design and model development.

Table of Contents

Chapter 1. Introduction

1.1 Background

This research investigates environmental drivers influencing the spatial and temporal distribution of a fungal disease called coccidioidomycosis, commonly referred to as valley fever, affecting over ten thousand people in the U.S. each year. Per Centers for Disease Control and Prevention statistics, in 2017 there were over 14,000 reported cases of valley fever in the U.S. with a steady rise in cases observed over the past two decades (CDC, 2018). Valley fever symptoms typically include fever, cough, fatigue, and other flu-like symptoms (Jin, 2013). Between 1% and 5% of the infected population will experience a more severe form of the disease where the spherules and endospores disseminate through the lymphatics and bloodstream to the skin, bones, joints, or other parts of the body (Saubolle, McKellar, & Sussland, 2007). In these rare cases, valley fever can be fatal (Ampel, 2007). In addition to the health burden, there is also a significant economic burden associated with this disease. An enhanced surveillance study conducted in Arizona for 2007 - 2008 estimated that valley fever health hospital costs for that year totaled over $85 million (Tsang et al., 2010). Overall, it is believed that valley fever is underdiagnosed and underreported, which is likely due to a combination of lack of recognition among health care providers, the mildness of most cases, and the requirement of a laboratory-confirmed diagnosis (Ampel, 2010; Tsang et al., 2013).

Valley fever is caused by the *Coccidioides* fungal species. A basic understanding of *Coccidioides'* species lifecycle is helpful in understanding its relationship with the environment and how the fungus causes illness. The *Coccidioides'* species lifecycle has two different phases. During the first phase, the pathogen gets nutrients from inorganic or decaying organic matter in the soil. After periods of heavy precipitation, a hyphal form (strand) grows in the soil (Thompson III, 2011). The hyphae then break apart into microscopic spores called arthroconidia, which can easily become airborne (Thompson III, 2011). The arthroconidia range in size from 5 to 30 μm in length and 1.5 to 4.5 μm in width (Kolivras, Johnson, Comrie, & Yool, 2001). Airborne arthroconidia can be inhaled by a host or can return to the soil and grow into more hyphae. The fungus transitions to the parasitic phase when it enters a human or animal host through inhalation (Cole, Hurtgen, & Hung, 2012). This is the primary method of infection; valley fever is not spread from person to person. The transition to the second phase occurs inside the host; the arthroconidia transform into a spherule that separates into many small endospores, with each endospore potentially growing into a new spherule and causing illness (Cole et al., 2012). The illness then manifests as primary coccidioidomycosis, chronic pulmonary coccidioidomycosis, or acute/chronic coccidioidomycosis (Laniado-Laborin, 2007).

Valley fever is endemic in the semi-arid regions of the southwestern U.S., parts of Mexico, and in Central and South America, and has recently been found in south-central Washington state (Marsden-Haug et al., 2013). The two different *Coccidioides* species that cause valley fever have different geographical distributions, though there is some overlap. *C. immitis* is predominantly in central California while *C. posadasii* has a wider distribution that includes Arizona and other regions in the southwestern U.S., Mexico, and South America (Fisher, Koenig, White, & Taylor, 2002; Thompson III, 2011). Pathogenically, there is little difference between these two species; their infections lead to similar outcomes (Engelthaler et al., 2016). Suitable environmental conditions, such as temperature, moisture, and soil conditions, are necessary for the growth of *Coccidioides* spp., though the specifics of these conditions are not well understood in the natural environment (Fisher et al., 2007). Skin testing that occurred over 50 years ago forms the basis of our understanding of the disease's endemic range (Edwards & Palmer, 1957). The most infective areas in the U.S. were found to be Kern County in California; Pima, Pinal, and Maricopa Counties in Arizona; and a few counties in southwestern Texas (Maddy, 1958).

This research is situated in medical geography, a sub-discipline of geography that began in the 1950s with a focus on the mapping of infectious diseases (May, 1950), and broadly focuses on the spatial patterns and processes associated with health and disease. Specifically, this research is influenced by the landscape epidemiology framework within medical geography. In the 1960s, E. Pavlovsky, a Russian parasitologist, defined the nidus or foci of infectious disease by analyzing association between vegetation, fauna, soils, climate, and other environmental factors (Pavlovsky, 1966). It was found that landscape patterns could be used to assess the presence of disease agents. The ability to assess the natural landscape for conditions permissible for disease transmission allowed for preventative measures to be taken based on geographic analysis. Building upon this initial framework, the significance of human contributions to this permissible environment were recognized (Meade, 1977). Today, the landscape epidemiology framework considers the temporal and spatial interactions between the host, pathogen, vector, and environment in the enabling of disease transmission (Reisen, 2010). This research uses this framework in developing studies to better understand the specific conditions, interactions, and environments that combine to produce a pathway for transmission of valley fever. Overall, this research contributes to medical geography by improving our understanding of the spatial patterns and processes associated with an emerging infectious disease.

1.2 Explanation of the Problem

1.2.1 Environmental Risk Factors

For over 50 years, *Coccidioides'* spp. relationship with environmental factors has been questioned and examined, yet we are still uncertain of what environmental risk variables are associated with this disease. *Coccidioides* spp. research in the U.S. began in early 1900s (Hirschmann, 2007), but many insights into the disease and its distribution were not discovered until the early 1940s when military members began training for WWII in the San Joaquin Valley of California and elsewhere around the southwestern U.S. (Smith, Beard, Rosenberger, & Whiting, 1946). One of the first environmental connections made was that between disease incidence and dust exposure. Smith et al. (1946) found that new trainees became infected at rates of 8-25% per year, with infection rates reduced after dust control measures had been put in place. In the 1950s, Maddy (1957) published research on the geographical distribution of *C. immitis*, connecting the disease to the Lower Sonoran Life Zone's environmental and climatic characteristics (Maddy, 1957). In addition to the observational studies occurring in the 1950s, Friedman et al. (1956) performed laboratory testing to determine the temperature and humidity limits of *C. immitis,* concluding that it was well-suited to arid habitats. Research on valley fever and the environment picked up again in the early 2000s, the same time that a valley fever outbreak at Dinosaur National Monument in Utah drew interest (Petersen et al., 2004). Research during the 2000s included studies that sought to quantify the relationship between climate and valley fever incidence as well as descriptive studies that explored environmental factors where positive *Coccidioides* samples had been found.

Much of what we know about *Coccidioides*' spp. ecological niche comes from limited field observations and laboratory experiments. Regarding climate and soil conditions, *Coccidioides* spp. are typically found in drier soils and is not found in saturated soils (Fisher et al., 2007); rainfall greater than 20 inches per year decreases the prevalence of the fungus in the soil (Reed, 1960). Conversely, *Coccidioides* spp. seem to do well in areas where there is a hot and dry period that sterilizes the soil surface and removes competitors (Maddy, 1957). In a laboratory setting, researchers found that *C. immitis'* growth was more abundant in soil that had been sterilized (Maddy, 1965). Laboratory testing has shown that *Coccidioides* spp. can survive temperatures ranging from 15°C to 37°C (Friedman et al., 1956), but the optimal range is suspected to be between 20-40°C (Fisher et al., 2007). In taking repeated soil samples from multiple depths at a known growth site over the course of a few years, Maddy (1965) found that during the hottest months, the fungus had moved deeper into the soil.

Since the identification of valley fever in the early 1900s, there have been multiple events and occurrences that have connected this disease with dust inhalation. For example, a dust storm in central California in 1977 resulted in a valley fever outbreak that infected hundreds as far north as San Francisco (Pappagianis & Einstein, 1978). A 1994 earthquake in Northridge, California also created a cloud of dust that settled in Ventura County along the California coast, where valley fever is not considered highly endemic, and infected over 200 people with the disease (Schneider et al., 1997). There was also a valley fever outbreak in 2001 at Dinosaur National Monument, Utah among a group of workers digging at an archeological site (Petersen et al., 2004). Finally, between 2011 and 2014, there was a valley fever outbreak among solar farm workers in California with research concluding that solar farm workers who only infrequently wore respirators had significantly increased odds for coccidioidomycosis (Cooksey et al., 2017).

Consolidating information learned from nine locations with positive soil isolations, Fisher et al. (2007) wrote a review of the environmental conditions in which *Coccidioides* spp. has been found. He summarized the key factors influencing the growth of the *Coccidioides* spp. as: "amount and timing of rainfall and available moisture, soil humidity, soil temperature, soil texture, alkalinity, salinity, organic content of soils, degree of exposure to sunlight and ultraviolet light, and competition with other microorganisms or plant species or both" (p. 50). His review found that temperature, precipitation, and soil textures were indicative variables for presence of *Coccidioides* spp., with the most common shared feature among sites being very fine sand and silt. While much has been learned in the last decade, there is still a lot of uncertainty surrounding valley fever's ecological niche, mainly due to research challenges, including difficulty isolating of the fungus from the soil, and data limitations.

The significant limitations related to direct analysis of *Coccidioides* spp. have led researchers to use valley fever incidence data to make inferences about the relationship between environmental conditions and risk of infection from *Coccidioides* spp. Research over the past 15 years has found correlations between valley fever incidence and environmental factors such as precipitation, wind, temperature, soil moisture, and dust, but the findings have varied in time and location, along with more nuanced differences in model lag times and degrees of explanation. Kolivras and Comrie (2003) created a statistical model for Pima County, AZ that found antecedent temperature and precipitation to be important predictors of incidence. Park et al. (2005) conducted a state-wide study in AZ finding the following climate variables to be statistically significant: drought indices, wind, mean temperature, dust, and rainfall. They also noted that hot, dry conditions had the strongest association. Zender et al. (2006) found that in Kern County, CA, climate factors explained very little of monthly incidence variability. Tamerius and Comrie (2011) found precipitation to be associated with rates of exposure for Maricopa and Pima Counties in Arizona with overall models explaining between 54 and 69% of the valley fever variance. Stacy, Comrie, and Yool (2012) used a vegetation index as a proxy for soil moisture and found moist soils in spring correlated with increases in incidence up to one year later. Using soil moisture sensors, Coopersmith et al. (2017) found that in California, summer and fall incidence was related to soil moisture from previous the previous winter/spring and in Arizona, winter/spring incidence was related to the previous summer's soil moisture. Finally, Gorris et al. (2017) conducted a regional analysis that found annual associations between valley fever incidence and temperature, precipitation, soil moisture, surface dust concentration, and cropland area, but seasonal and monthly associations varied by region.

Looking across previous research efforts and disease occurrences, a couple of gaps are identified. First, there is a relative lack of research in California, the state with the second highest incidence rates in the U.S. The results from the few studies that have focused on California have been inconsistent regarding the significance of associations between environmental factors and valley fever incidence. Second, the unexpected 2010 discovery of a *Coccidioides* spp. growth site in Washington State, an area far to the north of the commonly accepted endemic region for valley fever, along with concurrent human cases there, highlights that environmental conditions that control the spatial distribution of *Coccidioides* remain unclear (Marsden-Haug et al., 2013). Finally, past studies have mainly approached valley fever research from either a purely spatial or purely temporal perspective. Research is needed that assesses potential environmental risk factors that may interact in both space and time.

1.2.2. Research Challenges

Coccidioides spp. are difficult to analyze directly because it is very difficult and costly to isolate the fungus from the soil or detect it in air samples (Baptista-Rosas, Hinojosa, & Riquelme, 2007; Barker, Tabor, Shubitz, Perrill, & Orbach, 2012; Chow, Griffin, Barker, Loparev, & Litvintseva, 2016; Greene, Koenig, Fisher, & Taylor, 2000). A study done in the 1960s took over 700 soil samples, from 13 different suspected sites, with only 23 testing positive for *Coccidioides* spp. (Swatek, Omieczynski, & Plunkett, 1967). In a more recent study, Greene et al. (2000) found only 4 genetically distinct isolates from a sample size of 720 taken from the San Joaquin Valley in California. The ability to detect *Coccidioides* spp. in the soil improved in the early 21st century with polymerase chain reaction (PCR) techniques that test the soil directly and allow microbial detection and isolation of DNA, but positive samples have still proven to be patchy and hard to find (Fisher et al., 2007).

There are quite a few limitations to using incidence data to make inferences about the environmental niche and phenology of *Coccidioides* spp. One of the limitations to using incidence data is how far removed it is from the environmental variables acting on the pathogen. Environmental factors influence the growth and distribution of the fungus, but then individuals have to inhale the spores, become symptomatic, visit a doctor, get tested for valley fever, and have the results reported. Temporal analyses using incidence data would ideally be based on exposure date, but there are time gaps between date of exposure, onset of symptoms, and valley fever diagnosis with physicians not always recording or being able to determine exposure dates. Comrie (2005) thoroughly analyzed over 3,000 records and found that they were inconsistent in recording estimated symptom onsets dates and that no simple adjustments could be made for all records to improve accuracy in estimating exposure dates; thus, estimating exposure dates required more complicated data manipulation.

Other concerns with the use of incidence data include changes in reporting criteria, improved disease awareness, changes in testing methods, changes in the susceptible population, and inconsistent data availability across political boundaries (Comrie, 2005; Kolivras & Comrie, 2003). Some increases in incidence may be attributed to increased awareness of the disease, however, in a 2006 study, only a small proportion of patients (2-13%) in a valley fever endemic area with symptoms similar to those of valley fever were found to be tested for the disease, which indicates the disease is greatly underreported (Tsang et al., 2013). Changes in definitions, reporting requirements, and testing procedures can also affect incidence levels. For example, California began using a laboratory-based reporting system in 2010, which may explain some of the increase in cases seen in 2011 (Tsang et al., 2013). Arizona might have had artificially high numbers around the same period due to a commercial laboratory that changed their definition to only require positive enzyme immunoassay (EIA) results (Tsang et al., 2013). Valley fever is a reportable disease in many, but not all, of the states where it is endemic. The differences in reporting requirements, and in how long each state has been collecting data, makes the use of incidence data challenging when conducting studies that cross political boundaries. All these factors contribute to uneven, heterogeneous data that must be taken into consideration when conducting valley fever studies (Comrie, 2005). While the use of incidence data is not ideal, it is the only time series data currently available to compare to temporal environmental conditions.

1.3 Statement of Purpose, Significance, and Dissertation Outline

1.3.1. Statement of Purpose and Significance of Research

This research aims to contribute to the body of literature seeking to better understand and quantify how environmental factors are related to valley fever incidence. This work is organized around three different research questions, each leading to three different objectives of the dissertation project. Through these, the associations between environmental variables and valley fever incidence is explored and models are developed to assess and quantify relationships.

- **Research Question 1:** *Are the statistically significant relationships between valley fever incidence rates and climatic factors in Pima County, Arizona also significant in Kern County, CA?*

- **Research Question 2:** *Which environmental variables are the most influential in explaining the probable distribution of valley fever in the western U.S.?*
- **Research Question 3:** *What are the spatio-temporal trends of valley fever risk in California? Do variations in climate and land cover explain these trends?*

The objective of the first research question is to examine and quantify climatic relationships in an understudied valley fever endemic area. The second question aims to improve our ecological understanding of the *Coccidioides* habitat and characterize the current disease distribution area across the western U.S. The third question identifies areas of current and future concern and assesses drivers of spatio-temporal patterns in California. Overall, this research improves our understanding of underlying ecological relationships that connect humans, the valley fever pathogens, and the environment, and it brings us a step closer to forecasting disease distributions and emergence events based on environmental conditions. The results of this research can be applied by public health officials in the development of an early warning system and in the allocation of scarce surveillance and public education resources.

1.3.2 Structure of the Dissertation

The dissertation is organized in five chapters: Chapter 1 provides a review of the literature regarding the valley fever and its environmental connections along with identified gaps in past analyses. Chapter 2 assesses the temporal relationship between valley fever and climate in an understudied region. Chapter 3 focuses on a regional spatial assessment of *Coccidioides'* ecological niche, expanding the region of study beyond previous analyses. Chapter 4 details a new method used in the study of this disease, a hierarchical Bayesian spatio-temporal model developed to analyze observed patterns and assess environmental drivers of valley fever in California. Chapter 5 provides the overall conclusions reached as a result of the previously described research. It also identifies research obstacles and limitations of this work and includes future avenues for research.

1.4 References

Ampel, N. M. (2007). The complex immunology of human coccidioidomycosis. Annals of the New York Academy of Sciences, 1111, 245–258. https://doi.org/10.1196/annals.1406.032

Ampel, N. M. (2010). What 's Behind the Increasing Rates of Coccidioidomycosis in Arizona and California? Current Infectious Disease Reports, 12, 211–216. https://doi.org/10.1007/s11908-010-0094-3

Baptista-Rosas, R. C., Hinojosa, A., & Riquelme, M. (2007). Ecological Niche Modeling of Coccidioides spp. in Western North American Deserts. Annals of the New York Academy of Sciences, 1111, 35–46. https://doi.org/10.1196/annals.1406.003

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

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

Chow, N. A., Griffin, D., Barker, B., Loparev, V. N., & Litvintseva, A. P. (2016). Molecular detection of airborne coccidioides in tucson, arizona. Medical Mycology, 54(6), 584–592. https://doi.org/10.1093/mmy/myw022

Comrie, A. C. (2005). Climate Factors Influencing Coccidioidomycosis Seasonality and Outbreaks. Environmental Health Perspectives, 113(6), 688–692. https://doi.org/10.1289/ehp.7786

Cooksey, G. L. S., Wilken, J. A., Mcnary, J., Gilliss, D., Shusterman, D., Materna, B. L.,

& Vugia, D. J. (2017). Dust Exposure and Coccidioidomycosis Prevention Among Solar Power Farm Construction Workers in California, 107(8), 1296–1304. https://doi.org/10.2105/AJPH.2017.303820

Coopersmith, E. J., Bell, J. E., Benedict, K., Shriber, J., McCotter, O., & Cosh, M. H. (2017). Relating coccidioidomycosis (valley fever) incidence to soil moisture conditions. GeoHealth, 51–63. https://doi.org/10.1002/2016GH000033

Edwards, P. Q., & Palmer, C. E. (1957). Prevalence of sensitivity to coccidioidin, with special reference to specific and nonspecific reactions to coccidioidin and to histoplamin. Dis. Chest, 31, 355–360.

Engelthaler, D. M., Chandler, R., Litvintseva, A., Keim, P., Driebe, E. M., Schupp, J. M., … Barker, B. (2016). Local Population Structure and Patterns of Western Hemisphere Dispersal for Coccidioides spp., the Fungal Cause of Valley Fever. MBIO, 7(2), 1–37. https://doi.org/10.1101/024778

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

Fisher, M. C., Koenig, G. L., White, T. J., & Taylor, J. W. (2002). Molecular and phenotypic description of Coccidioides posadasii sp. nov., previously recognized as the non-California population of Coccidioides immitis. Mycologia, 94(1), 73–84.

Friedman, L., Smith, C. E., Pappagianis, D., & Berman, R. J. (1956). Survival of Coccidioides immitis under controlled conditions of temperature and humidity. American Journal of Public Health, 46, 1317–1324.

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.

Hirschmann, J. V. (2007). The Early History of Coccidioidomycosis: 1892-1945. Clinical Infectious Diseases, 44(9), 1202–1207. https://doi.org/10.1086/513202

Jin, J. (2013). Valley Fever (Coccidioidomycosis). Journal of the American Medical Association, 310(22), 2470.

Kolivras, K. N., & Comrie, A. C. (2003). Modeling valley fever (coccidioidomycosis) incidence on the basis of climate conditions. International Journal of Biometeorology, 47, 87–101. https://doi.org/10.1007/s00484-002-0155-x

Kolivras, K. N., Johnson, P. S., Comrie, A. C., & Yool, S. R. (2001). Environmental variability and coccidioidomycosis (valley fever). Aerobiologia, 17, 31–42.

Laniado-Laborin, R. (2007). Expanding understanding of epidemiology of coccidioidomycosis in the western hemisphere. Annals of the New York Academy of Sciences, 1111, 19–34. https://doi.org/10.1196/annals.1406.004

Maddy, K. T. (1957). Ecological Factors of the Geographic Distribution of Coccidioides Immitis. Journal of the American Medical Association, 130(11), 475–476. https://doi.org/10.1017/S000748530002229X

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., & 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

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

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

May, J. M. (1950). Medical Geography: Its Methods and Objectives. Geographical Review, 40(1), 9–41. https://doi.org/http://dx.doi.org/10.1016/0037-7856(77)90157-3

Meade, M. S. (1977). Medical Geography as Human Ecology: The Dimension of Population Movement. The Geographical Review, 67(4), 379–393.

National Oceanic and Atmospheric Administration (NOAA). (2018). Climate of Texas. Retrieved Febrary 25, 2019 from https://www.ncdc.noaa.gov/climatenormals/clim60/states/Clim_TX_01.pdf

Pappagianis, D., & Einstein, H. E. (1978). Tempest from Tehachapi takes toll or Coccidioides conveyed aloft and afar. The Western Journal of Medicine, 129, 527–530.

Pavlovsky, E. (1966). Natural Nidality of Transmissible Diseases with Special Reference to the Landscape Ecology of Zooanthroponses (University). Urbana, IL.

Petersen, L. R., Marshall, S. L., Barton-Dickson, C., Hajjeh, R. A., Lindsley, M. D., Warnock, D. W., … Morgan, J. (2004). Coccidioidomycosis among Workers at an Archeological Site, Northeastern Utah. Emerging Infectious Diseases, 10(4), 637–642. https://doi.org/10.3201/eid1004.030446

Reed, R. E. (1960). Ecology and Epizoology of Coccidioidomycosis. In Reports of papers delivered at the 1959-1960 Conventions of the Intermountain Veterinary Medical Association.

Reisen, W. K. (2010). Landscape epidemiology of vector-borne diseases. Annual Review of Entomology, 55, 461–483. https://doi.org/10.1146/annurev-ento-112408-085419

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

Schneider, E., Hajjeh, R. A., Spiegel, R. A., Jibson, R. W., Harp, E. L., Marshall, G. A., … Werner, S. B. (1997). A Coccidioidomycosis Outbreak Following the Northridge, Calif, Earthquake. The Journal of the American Medical Association, 277(11), 904–908. https://doi.org/10.1001/jama.277.11.904

Smith, C. E., Beard, R. R., Rosenberger, H. G. M., & Whiting, E. G. M. (1946). Effect of Season and Dust Control on Coccidioidomycosis. The Journal of the American Medical Association, 132(14), 833–838.

Stacy, P. K. R., Comrie, A. C., & Yool, S. R. (2012). Modeling Valley Fever Incidence in Arizona Using a Satellite-Derived Soil Moisture Proxy. GIScience & Remote Sensing, 49(2), 299–316. https://doi.org/10.2747/1548-1603.49.2.299

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.

Talamantes, J., Behseta, S., & Zender, C. S. (2007). Statistical modeling of valley fever data in Kern County, California. International Journal of Biometeorology, 51(4), 307–313. https://doi.org/10.1007/s00484-006-0065-4

Tamerius, J. D., & Comrie, A. C. (2011). Coccidioidomycosis incidence in Arizona predicted by seasonal precipitation. PLoS ONE, 6(6). https://doi.org/10.1371/journal.pone.0021009

Thompson III, G. R. (2011). Strategies to combat coccidioidomycosis: Are we making any progress? Current Fungal Infection Reports, 5(4), 215–223. https://doi.org/10.1007/s12281-011-0066-6

Tsang, C. A., Tabnak, F., Vugia, D. J., Benedict, K., Chiller, T., & Park, B. J. (2013). Increase in Reported Coccidiodomycosis - United States, 1998-2011. Morbidity and Mortality Weekly Report, 62(12), 217–221. https://doi.org/http://dx.doi.org/10.1108/17506200710779521

Zender, C. S., & Talamantes, J. (2006). Climate controls on valley fever incidence in Kern County, California. International Journal of Biometeorology, 50(3), 174–182. https://doi.org/10.1007/s00484-005-0007-6

Chapter 2. Investigating the Relationship between Climate and Valley Fever (Coccidioidomycosis)

2.1 Introduction

This chapter assesses and quantifies relationships between valley fever and temperature, precipitation, wind, and dust in a California county. It uses a time-series analysis, regressing the climatic variables with valley fever incidence in Kern County, CA at the seasonal temporal scale using data from 2001 to 2015. The objective is to identify whether variables found to be significantly related to valley fever in highly endemic Arizona counties are also significant in the most endemic California county.

2.2 Publication

The manuscript related to this chapter was published in EcoHealth journal and can be found in Appendix A.

Weaver, E. A., and Kolivras, K. N. (2018). Investigating the Relationship Between Climate and Valley Fever (Coccidioidomycosis). *EcoHealth*, *15*(4), 840-852.

Chapter 3. Environmental factors affecting ecological niche of *Coccidioides* (spp.) and spatial dynamics of valley fever in the United States

3.1 Introduction

This chapter analyzes the spatial distribution of valley fever in the western United States. Maximum Entropy ecological niche modeling is used to identify the probable spatial distribution of valley fever based on climate, elevation, soil, and land cover variables using disease case data from 2001 to 2016. The aim is to characterize the disease's spatial distribution and identify environmental factors influential in the model's development.

3.2 Publication

Appendix B contains the draft of the original manuscript submitted to an academic journal for review on 18 December 2018.

Weaver, E.A., Kolivras, K., Thomas, V., Thomas, R.Q., and Abbas, K. Under Review. Environmental factors affecting ecological niche of *Coccidioides* (spp.) and spatial dynamics of valley fever in the United States . *Spatial and Spatio-temporal Epidemiology*.

Chapter 4. Spatiotemporal modeling of relative risk of coccidioidomycosis in California

4.1 Introduction

This chapter details the development of a hierarchical Bayesian spatio-temporal model used to analyze patterns of disease risk in California using valley fever case data from 2001-2018. The objective of this study is to identify patterns in disease risk that may not be apparent in purely spatial or purely temporal analyses and to evaluate hypothesized environmental risk factors in accounting for the patterns observed. The covariates assessed include temperature, precipitation, cultivated land cover, and temperature of the driest quarter.

4.2 Publication

The draft manuscript for this analysis is found in Appendix C. It is currently being revised for submission to a peer-reviewed journal.

Weaver, Elizabeth A and Thomas, R.Q. In preparation. Spatiotemporal modeling of relative risk of coccidioidomycosis in California

Chapter 5. Conclusions

Valley fever rates are increasing in the United States, creating a public health burden in affected areas. This research contributes to our knowledge regarding how valley fever incidence rates are related to varying environmental conditions. Focusing on the western U.S., this dissertation broadly aimed to: (1) assess the temporal relationship between climate and valley fever in an understudied area; (2) identify factors influencing the regional spatial distribution of the disease; and (3) analyze the spatio-temporal distribution of valley fever risk along with environmental factors hypothesized to influence that distribution. The dissertation research resulted in several findings with public health applications.

First, this dissertation found that some climatic variables have significant relationships with valley fever in the most endemic counties in Arizona also had significant relationships in Kern County, California when assessed at the seasonal scale. Specifically, winter precipitation was identified as a significant factor in explaining valley fever variability in multiple models. This is an important insight that can be used in the future development of early warning systems. With previous research focused mainly in Arizona, the significant relationships identified in this California study created a basis for further regional analysis regarding valley fever-environment associations.

Second, this dissertation contributed to a better understanding of the U.S. endemic region for this disease. Using a new ecological niche modeling technique and expanding the study's range to take into consideration the discovery of *Coccidioides* spp. growth sites in the state of Washington, the probable distribution of valley fever was identified. The distribution based on environmental variables was characterized and the Columbia

Plateau region of Washington and Oregon was identified as having more favorable conditions for fungus presence than surrounding areas. This identification of potential disease habitats is important for public health officials watching for potential outbreaks and looking to allocate educational and awareness resources. This research also identified climatic and land cover variables important in explaining the spatial distribution of the disease, which can be used in future predictive models to assess how environmental changes will affect disease distribution.

Finally, this research developed a Bayesian spatio-temporal model that identified counties at increased risk in California by assessing geographic patterns in temporal trends. Counties outside of the known highly endemic areas were identified for elevated risk, which may be useful in prioritization of public health resources. This methodological framework can be replicated in other states to assess disease risk, and it provides a foundation for future forecasting of valley fever risk based on environmental conditions.

Overall, this dissertation used various modeling techniques to advance our knowledge regarding the relationship between valley fever and environmental risk factors. This research expanded on past studies by assessing relationships in new regions, assessing new environmental variables, and assessing variable interactions in space and time. The major limitations to this research involved the use of case data as a substitute for pathogen presence and abundance. In using case data, assumptions were made that disease report locations were geographically connected to exposure locations and that report dates were associated with exposure dates. To minimize the impact of any error in these assumptions, relatively course spatial and temporal scales were used in capturing trends and identifying significant associations. Overall, these findings support past research regarding the significance of alternating cool/wet and hot/dry seasons. Added to this body of research are the significance of land cover variables, specifically barren and cultivated land covers, and the identification areas of concern based on spatio-temporal patterns. Future research can use these quantified associations to forecast where and when disease emergence may occur based on environmental conditions, allowing for more focused disease prevention, awareness, education, and monitoring efforts.

Appendix – Publications related to this research

Appendix A. Weaver and Kolivras (2018) Climate and valley fever paper

Investigating the Climate - Valley Fever Relationship in Kern County, California, U.S.A.

Abstract

Valley fever (coccidioidomycosis) is a disease caused by inhalation of spores from the soil-dwelling *Coccidioides* fungal species. The disease is endemic to semi-arid areas in the western United States and parts of Central and South America. The region of interest for this study, Kern County, California, accounts for approximately 14% of the reported valley fever cases in the U.S. each year. It is hypothesized that the weather conditions that foster the growth and dispersal of the fungus influence the number of cases in the endemic area. This study uses regression-based analysis to model and assess the seasonal relationships between valley fever incidence and climatic variables including concurrent and lagged precipitation, temperature, Palmer Drought Severity Index, wind speed, and PM_{10} using data from 2000 – 2015. We find statistically significant links between disease incidence and climate conditions in Kern County, California. The best performing seasonal model explains up to 76% of the variability in fall valley fever incidence based on concurrent and antecedent climate conditions. Findings are consistent with previous studies suggesting that antecedent precipitation is an important predictor of disease. The significant relationships found support the "grow and blow" hypothesis for climate-related coccidioidomycosis incidence risk that was originally developed for Arizona.

Introduction

Valley fever, or coccidioidomycosis, is a fungal disease that can infect a human or other mammalian host through their respiratory tract. The fungi responsible, *Coccidioides immitis and Coccidioides posadasii*, normally dwell in the soil but can become airborne, and thus inhaled, if the soil is disturbed. Approximately 40 percent of those who inhale the fungal spores (arthroconidia) will become ill (Thompson III, 2011). Common symptoms associated with valley fever include fever, cough, chest discomfort, and fatigue; the typical time between exposure and onset of symptoms is 7 to 21 days (Saubolle, McKellar, & Sussland, 2007). In a small portion of the population, fungal endospores disseminate from the lungs to other parts of the body, including the skin, lymph nodes, and skeletal system (Saubolle et al., 2007). These endospores can also spread to the brain causing meningeal disease, though this is less common. Valley fever is not spread from person to person; inhalation of fungal spores is the only method of infection other than rare cases in which it has been acquired through an open cut or wound (Eckmann, Schaefer, & Huppert, 1964). In rare cases, the disease can be fatal, especially if not treated (Hector et al., 2011). Valley fever rates in the U.S. have been generally increasing over the past two and a half decades (Centers for Disease Control and Prevention, 2018), creating an emerging public health burden. In California alone, there were over 1,000 deaths attributed to the disease between 2000 and 2013, equating to 2.2 deaths per 1 million population (Sondermeyer, Lee, Gilliss, & Vugia, 2016).

Valley fever is endemic in the western U.S., northern and central Mexico, and in parts of Central and South America. These regions are characterized as arid and semi-arid with hot summers and few winter freezes (Maddy & Coccozza, 1964). It is hypothesized that the weather conditions that foster the growth and dispersal of the fungus influence the number of cases in the endemic area (i.e. Comrie, 2005; Coopersmith et al., 2017; Kolivras, Johnson, Comrie, & Yool, 2001; Park et al., 2005). This hypothesis is based on our understanding of the *Coccidioides* spp. dimorphic lifecycle. During the saprobic phase of the pathogen's lifecycle, the fungus lives in the top layer of the soil and gets nutrients from inorganic or decaying organic matter. To grow, *Coccidioides* spp. require suitable temperature, moisture, and soil conditions that allow it to develop into a hyphal form (strand) (Fisher, Bultman, Johnson, Pappagianis, & Zaborsky, 2007). During the hot/dry season, the hyphal strands can break apart into arthroconidia, or microscopic spores that range in size from 5 to 30μm in length and 1.5 to 4.5μm in width (Pappagianis, 1988). *Coccidioides* spp. can spend their whole life cycle in the soil, but if the soil is disturbed by wind gusts or another disturbance mechanism, the arthroconidia can easily become airborne. These spores either return to the soil where they can grow into more hyphae, or they can be inhaled by a host. In the host, the arthroconidia begin the parasitic phase of the lifecycle as they transform into a spherule that can separate into many small endospores and potentially initiate another generation of spherules within the host (Cole, Hurtgen, & Hung, 2012). Thus, the "grow and blow" hypothesis suggests that soil moisture facilitates the growth of the fungus, subsequent dry conditions result in spore fragmentation, and a disturbance event enables the spores to become airborne, ultimately resulting in human exposures (Comrie & Glueck, 2007; Tamerius & Comrie, 2011).

Alternating wet and hot/dry seasons are thought to be critical for the growth and dispersal of the fungal spores. In direct observations, it was found that during the hottest months of the year, *Coccidioides* spp. survive by moving deeper into the soil (Maddy, 1965). High summer temperatures likely remove competitors in the top layer of soil, effectively sterilizing the soil, allowing for more abundant *Coccidioides* spp. growth when moisture returns to optimal levels (Maddy, 1965). Areas with approximately 130 to 500 mm of rainfall seem to have the best conditions for fungal growth (Maddy, 1958). *Coccidioides* spp. are not found in soils that are frequently or continually saturated (Fisher et al., 2007); if the soil is too moist, it is speculated that competitors of the *Coccidioides* spp. thrive and decrease its prevalence (Reed, 1960). *Coccidioides* spp. can survive a wide range of temperatures in a laboratory setting (Friedman, Smith, Pappagianis, & Berman, 1956), but it is believed that the optimal temperature for growth is between 20-40 degrees Celsius (Fisher et al., 2007). The counties in California and Arizona with the highest valley fever rates have temperature and precipitation ranges that exhibit these climate conditions along with an alternating wet and hot/dry seasonal behavior, though California has a single wet season in the winter and a very dry summer, while Arizona has a bimodal pattern with precipitation in both the summer and winter (Gorris, Cat, Zender, Treseder, & Randerson, 2017).

While observed relationships between climate conditions and valley fever incidence have been documented in previous studies (i.e. Flynn et al., 1979; Pappagianis & Einstein, 1978; Smith, Beard, Rosenberger, & Whiting, 1946), only a few studies have attempted quantify this relationship. Multiple studies in Arizona have found significant relationships between climatic conditions and the timing of reported cases (Comrie, 2005; Kolivras & Comrie, 2003; Komatsu et al., 2003; Park et al., 2005; Pianalto, 2013; Stacy, Comrie, & Yool, 2012; Tamerius & Comrie, 2011), but these same relationships were much weaker, or not found at all, in two Kern County, California studies (Talamantes, Behseta, & Zender, 2007b; Zender & Talamantes, 2006). The earlier of these two Kern County studies investigated the relationship between incidence fluctuations and temperature, precipitation, surface pressure, and wind anomalies with data from 1980 to 2002 (Zender & Talamantes, 2006). These researchers found that, after correcting for annual cycles and autocorrelations, 8-month lagged precipitation was the only highly statistically significant ($p<0.01$) variable, explaining only 4% of monthly incidence variability. The second study tested three different generalized auto regressive moving average (GARMA) models with various inputs including temperature, precipitation, wind, and prior incidence (Talamantes et al., 2007b). They found that dependence of incidence rate fluctuations on weather parameters was weak. Overall, these researchers suspected that human activities or biological processes may play a more influential role than climate in explaining anomalies and that weather fluctuations in Kern County might be too small to explain fluctuations in incidence.

The seasonal behavior of valley fever incidence in Kern County (see FIGURE 1) and its precise relationship with climatic variables remains uncertain. This study investigates the incidence-climate relationship through regression techniques using multiple climatic variables, including several that have not previously been assessed in Kern County. Clarifying and quantifying the relationships between seasonal climatic changes and disease dynamics will enhance our understanding of what drives valley fever and allow for future forecasting of valley fever health risks.

FIGURE 1. Study Area, Kern County, California, U.S.A.

Climate Variables

Climate variables used in this study include precipitation, drought, temperature, wind, wind gust, and air quality measurements (see TABLE 1). Cumulative precipitation and the Palmer Drought Severity Index (PDSI) are used as measures of soil moisture. The role of moisture is hypothesized to be two-fold: a wet period is required for fungal growth, and a dry period is required for autolyzation into spores and spore dispersal (Pappagianis, 1988). Because precipitation influences spore growth in months and seasons prior to exposure, precipitation data is assessed concurrently with incidence, but also with lags up to one-year prior to account for this growth. The PDSI uses temperature and precipitation data to determine the accumulated water excess or deficit, indicating the severity of a wet or dry spell (Palmer, 1965). The index generally ranges from -6 to +6, with negative values denoting dry spells and positive values indicating wet spells (National Centers for Environmental Information, 2014). The PDSI is intrinsically autocorrelated, with the index changing slowly and smoothly over time, and does not compare well between locations, therefore it is only used in the initial univariate analysis (Alley, 1984). Air temperature data are used as proxies for soil temperature. Soil temperatures generally reflect ambient air temperature, with a lag that increases with soil depth (Fisher et al., 2007). High summer temperatures can sterilize the soil and create an environment, relatively free of competitors, that supports *Coccidioides* species' growth. Wind and wind gust data are used as indicators of potential disturbance that lead to increased spore concentration in the air. PM_{10} is a measure of airborne dust and is used as a more direct proxy for the windblown spore concentration in the air. PM_{10} is a mixture of various substances in the form of solid particles or liquid drops measuring 10 micrometers or less in the atmosphere. This range includes the size of a typical *Coccidioides* spp. spore. A variety of emission sources and meteorological conditions contribute to ambient PM_{10} (California Air Resources Board, 2017).

TABLE 1. Climatic Variables.

Climatic Variable	Source
Average temperature^1	Western Regional Climate Center
Average temperature of previous 3 months^1	Western Regional Climate Center
Average wind speed^1	National Centers for Environmental Information
Average 2-minute wind gust^1	National Centers for Environmental Information
Average 5-second wind gust^1	National Centers for Environmental Information
Average PM_{10}^2	California Air Resources Board
Average PDSI^3	National Centers for Environmental Information
Average Cumulative precipitation (concurrent)^1	Western Regional Climate Center
Cumulative precipitation from previous 12 months^1	Western Regional Climate Center
Cumulative precipitation lagged seasonally up to 12 months^1	Western Regional Climate Center

Spatial scale: ^{1}**Bakersfield Airport Weather Station;** ^{2}**San Joaquin Valley air basin;** ^{3}**San Joaquin Valley drainage basin**

Disease Data

Monthly valley fever incidence data (case counts) for Kern County were obtained from the California Department of Public Health from January 2000 to December 2015. The monthly case incidence rate per 100,000 of the general population was determined using linearly interpolated annual population estimates based on the 2000 and 2010 census counts and the 5-year American Community Survey 2015 population estimate (U.S. Census Bureau, 2000, 2010, 2015). A one-month offset in incidence data was used to allow for incubation time and diagnosis/reporting lags (Comrie & Glueck, 2007). Kern County accounted for over 40% of all cases in California during the study period with a total of 20,366 cases over the 16-year period. There is a slight upward trend to the data from 2000 to 2015, with incidence peaking in 2011 at 2,565 cases after a low of 599 cases in 2009 (see FIGURE 2).

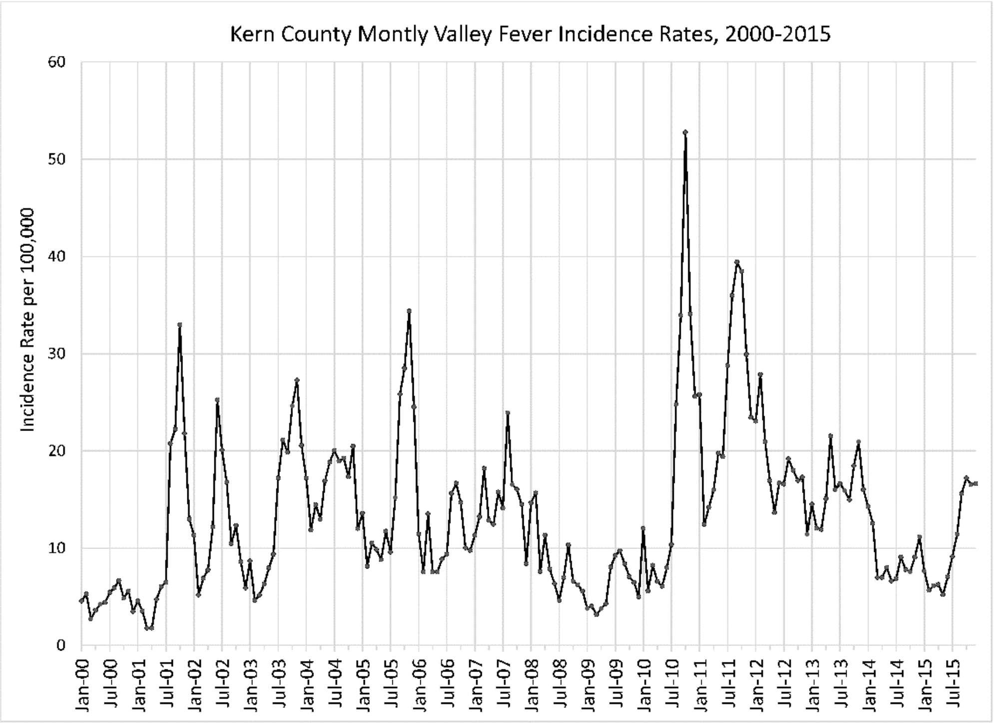


FIGURE 2. Monthly Valley Fever Incidence Rates for Kern County, CA from 2000-2015.

Monthly climatic data were obtained from the Western Regional Climate Center, the National Center for Environmental Information, and the California Air Resources Board. Precipitation, wind, and temperature data were obtained for the Bakersfield Airport weather observation station, which has continuous data back to 1938. The Bakersfield Airport station is located centrally within the county, approximately three miles north of the Bakersfield city center (see FIGURE 1). Since Bakersfield is the largest city in Kern County, with approximately 40% of the population within the city boundaries, the Bakersfield Airport station data reasonably represents climate conditions in the county and is applied to countywide valley fever data. Daily PM_{10} readings from four different San Joaquin Valley (SJV) Air Basin sites near Bakersfield were averaged to obtain an average monthly PM_{10} concentration value.

Methods

To better understand the temporal relationships between valley fever and climate in Kern County, univariate and multivariable linear regression was performed with incidence as the response variable and the climate data as the predictor variables (Hidalgo & Goodman, 2013). Multivariable regression models for each season were constructed using a forward stepwise selection approach with minimum AICc as the selection criteria. The use of AICc is a standard way of evaluating a model that takes into account fit and complexity with a bias correction term for small sample sizes (Johnson & Omland, 2004) and has been used in previous studies on valley fever (Pianalto, 2013; Talamantes et al., 2007b). Tests for normality, constant variance, collinearity, and autocorrelation were performed on each model. A leave-one-out cross-validation method was used to assess the robustness of correlations identified in the multivariable linear regression analyses.

The monthly variables were grouped into seasons prior to conducting the regression analysis. The benefit of a seasonal analysis, summarized by Comrie (2005), includes: (1) a useful way of dividing the year into wet and dry periods, (2) a simplified analysis that avoids monthly variability, (3) a reduction in error due to inconsistencies in reported onset dates, (4) an easy way to identify recurring times of the year that are important, and (5) an analytically and conceptually simpler way to compute and understand seasonal lag relationships. The monthly valley fever and climatic averages presented in FIGURE 3 enabled the definition of seasonal groupings based on periods of maximums and minimums. Data were grouped into four seasons, with winter including data from the months Dec-Feb, spring including Mar-May, summer including Jun-Aug, and fall including Sep-Nov. These time blocks were chosen because they captured the highest two and lowest two months of both average precipitation and valley fever in a single season. Each season was modeled separately as climate conditions can vary significantly between the different valley fever seasons (Comrie, 2005; Kolivras & Comrie, 2003; Stacy et al., 2012), with lower rates of valley fever from Feb-May and higher rates from Aug-Nov.

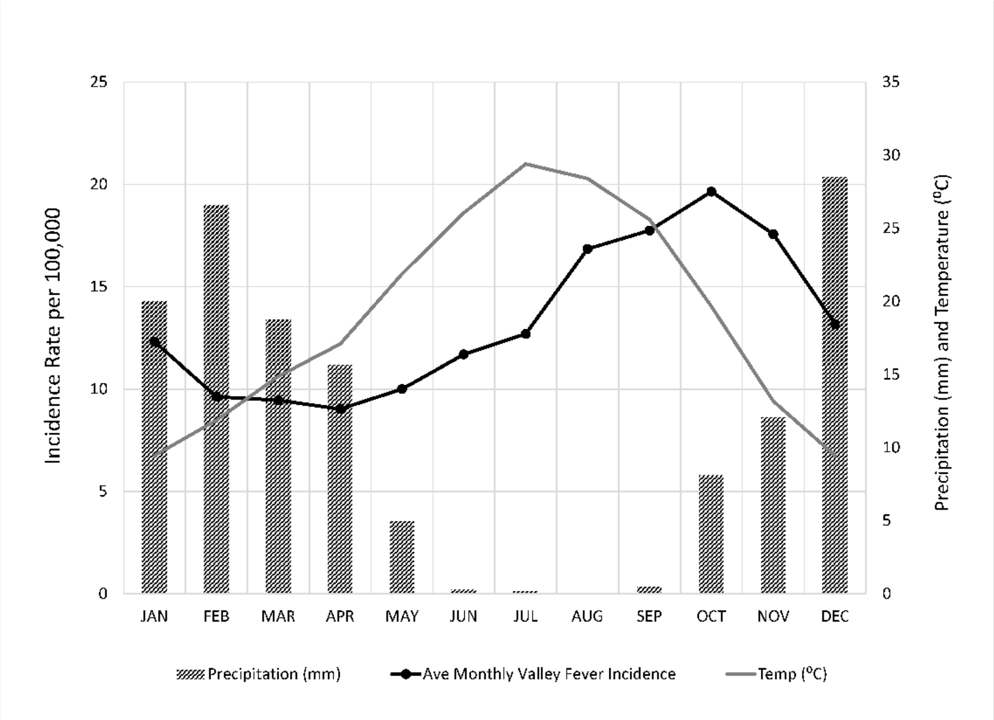


FIGURE 3. This plot shows disease and climate seasonality based on 2000-2015 monthly averages with the incidence having a single peak in the month of October, precipitation having a single peak in the month of February, and temperature having a single peak in July.

Results

Univariate Analysis

In the univariate analysis, there were no consistent significant variables (p-value < 0.05) across all four seasonal models (see TABLE 2). The winter model had no significant variables and the spring model only had one significant variable, precipitation lagged four seasons, i.e. precipitation from the previous spring. This variable had a direct, positive relationship with incidence. In the summer model, of the six variables with statistically significant relationships with valley fever, precipitation from the previous winter had the strongest relationship with an R-squared value of 0.43 ($p = 0.006$). The average 2-minute wind gust speed and PM_{10} had negative relationships with the incidence rate. For the fall model, when incidence is generally highest, the most significant variable was precipitation the previous winter with a positive relationship and an R-squared of 0.60 ($p < 0.001$), the highest value of any variable in the univariate analysis. PM_{10} also had a negative relationship with incidence rate in this season as well.

TABLE 2. Univariate linear regression significant correlations ($p < 0.05$) for all seasonal models.

Seasonal Model	Significant Variables	Coefficient	t-Ratio	P-Value	R-Squared
Winter	N/A	N/A	N/A	N/A	N/A
Spring	Average Precipitation Previous Spring	0.30	2.26	0.04	0.26
Summer	Average Temp. Previous 3 Months	-1.69	-2.15	0.05	0.25
	Average 2-Minute Wind Gust	-10.72	-2.66	0.02	0.34
	PDSI	1.04	2.40	0.03	0.29
	PM_{10}	-0.50	-2.34	0.03	0.28
	Total Precipitation Previous 12 Months	0.08	3.19	0.006	0.42
	Average Precipitation Previous Winter	0.32	3.25	0.006	0.43
Fall	PM_{10}	-0.43	-2.18	0.04	0.35
	Total Precipitation Previous 12 Months	0.12	3.83	0.002	0.51
	Average Precipitation Previous Winter	0.52	4.58	<0.001	0.60

Multivariable Analysis

In the multivariable analysis, the winter model explains 58% of the variation in incidence ($p = 0.007$) and involves four significant variables, including positive relationships with average temperature over the past 3 months, average 5-sec wind gust, and precipitation from the previous winter and spring (see TABLE 3). The spring model explains 43% of the variation in incidence ($p = .01$) and has two significant variables: precipitation from the previous fall and previous spring, both with positive relationships. The summer model explains 59% of the variability in incidence ($p = .001$). This model has two significant variables including an inverse relationship with average 2-minute wind speed, and a positive relationship with precipitation from the previous winter. The fall model is the most explanatory, with an adjusted R-squared of 0.76 ($p = 0.001$). This model has three significant variables including an inverse relationship with precipitation from the previous summer, and positive relationships with average temperature and precipitation from the previous winter. The time series of reported and modelled valley fever incidence for each season is shown in FIGURES 4a-4d.

TABLE 3. Multivariable linear regression results.

Model	Model Adj. R-Squared	AICc	RMSE	P-Value
Winter	0.58	101.37	3.56	0.007
Spring	0.43	96.00	3.75	0.01
Summer	0.59	102.04	4.53	0.001
Fall	0.76	106.60	4.74	0.001

Model	Variables	Standardized Coefficient	t-Ratio	P-Value
Winter	Average Temp Previous 3 Months	3.03	2.96	0.01
	Average 5-Second Wind Gust	2.14	2.24	0.05
	Precipitation Previous Spring	2.19	2.35	0.04
	Precipitation Previous Winter	3.26	3.24	0.008
Spring	Precipitation Previous Fall	2.72	2.52	0.03
	Precipitation Previous Spring	3.77	3.50	0.004
Summer	Average 2-Minute Wind Gust	-3.30	-2.78	0.02
	Precipitation Previous Winter	3.96	3.33	0.005
Fall	Average Temperature	3.47	2.72	0.02
	Precipitation Previous Summer	-3.64	-2.63	0.02
	Precipitation Previous Winter	10.04	7.00	<0.001

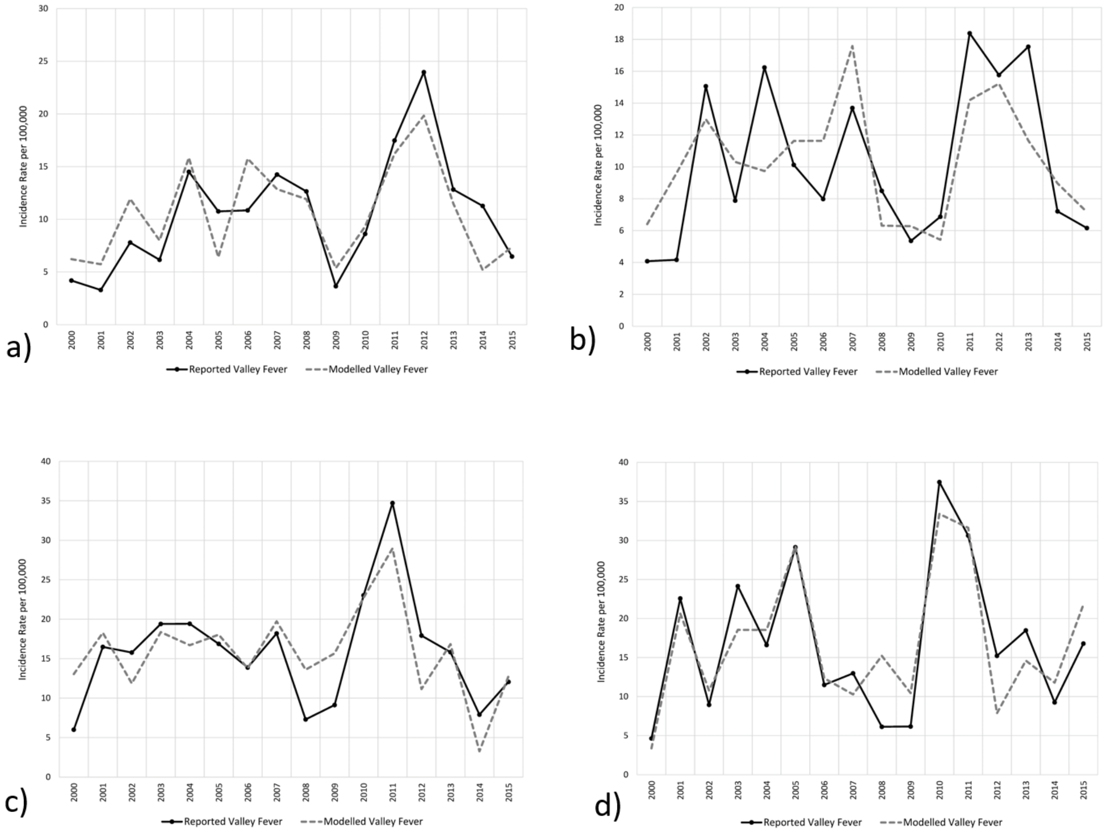


FIGURE 4. Reported valley fever incidence in Kern County, CA, and modelled valley fever incidence based on seasonal multivariable linear regression models: a) winter, b) spring, c) summer, d) fall.

Model Diagnostics

All models exhibited a near normal distribution of residuals, considering the low sample size, and had low autocorrelation based on Durbin-Watson statistics. In the multivariable models, assessment of variance inflation factors found low correlation among variables (VIF < 2.0). Constant variance was visually assessed using scatterplots of residuals vs. predicted values. A few of the univariate regression models showed homoscedasticity violations; PM_{10} in the summer and fall models, and PDSI in the fall model had fan shaped distributions with residuals growing larger as a function of the predicted value. While the estimates will still be unbiased, the presence of heteroscedasticity can cause the estimates for the variance of the slope to be inaccurate. Multivariable models were assessed using a leave-one-out cross validation method. This validation was done in JMP by selecting the *k*-fold cross-validation option and inputting the total number of observations (16) as the *k* value for each model (JMP, 2017). The output is a R-squared *k*-fold statistic, which is the average of the R-squared validation values for the *k* folds (see TABLE 4). The R-squared *k*-fold statistic for the spring model was significantly lower than the direct fit R-squared for the spring model, indicating that this model may be overfit, or may be tailored to specific data points and not generalizable outside the sample. The fall model had an R-squared *k*-fold statistic closest to the fitted R-squared values, 0.69, indicating predictions in closer agreement with observations.

TABLE 4. Model Evaluation Statistics

Season	**Model R-Squared**	**Model Adj. R-Squared**	**k-Fold R-Squared**
Winter	0.69	0.58	0.38
Spring	0.51	0.43	0.19
Summer	0.64	0.59	0.39
Fall	0.81	0.76	0.69

Discussion

This study found significant relationships between valley fever incidence rates and climatic factors in Kern County, CA. Specifically, we found highly significant correlations ($p < 0.01$) between valley fever incidence and precipitation from the previous winter in five of the eight models. This relationship implies that winter rains provide the moisture needed in the soil for the fungus to grow. In winter, the semi-permanent high-pressure area in the north Pacific Ocean moves southward and permits storm centers to move across California, bringing widespread, moderate precipitation. Occasionally, heavy rains and flooding occur when changes in the typical circulation pattern permit storm centers to approach from a southwesterly direction (Western Regional Climate Center, 2017). The only multivariable seasonal model that did not include winter precipitation was the spring model, in which precipitation from the previous spring (not winter) was significant, implying that the timing of the precipitation may be important, with a lag of at least two seasons needed to allow time for fungal growth, fragmentation, and disturbance.

Lagged precipitation variables stood out in this analysis as the most commonly significant climatic factor related to valley fever incidence among the different seasonal models. The significant, positive relationship between lagged precipitation and valley fever is consistent with earlier research by Zender and Talamantes (2006) but explains a higher proportion of variance which may be due to significant differences in methodologies and in the time-scales assessed. This finding supports recent research by Gorris et al. (2017) who assessed monthly, seasonal, and annual relationships between climate and valley fever in the San Joaquin Valley of California. Their research similarly found positive correlations between valley fever and precipitation, soil moisture, and NDVI, with a lag of 7 to 9 months. In the fall multivariable model, precipitation from the previous summer had a significant, negative relationship with incidence. This finding is consistent with the Arizona studies of Kolivras and Comrie (2003), Comrie (2005), and Tamerius and Comrie (2011) with the explanation that rain in the typically dry season likely inhibits dust production, lowering chances for inhalation of *Coccidioides'* spores. Overall, these findings support the hypothesized double-role of precipitation, with a wet period required for fungal growth, and a dry period required for spore formation and dispersal. It is likely that winter rains provide moisture that facilitate *Coccidioides* spp. growth, then subsequent dry conditions result in fragmentation of the hyphae after which the arthroconidia can easily become airborne, ultimately resulting in human exposures that occur in the seasons following.

Significant relationships between valley fever incidence and temperature were found in only a few of the models. Average temperature was significant in one model, the fall multivariable model, in which it was positively correlated with valley fever. Average temperature from the previous three months was negatively correlated with valley fever in the univariate summer model, but positively correlated with valley fever in the multivariable winter model. Gorris et al. (2017) calculated monthly climate and valley fever anomalies for the endemic region and found that in the San Joaquin Valley, temperature was negatively correlated with incidence six months later. The negative correlation finding contrasts with Park et al. (2005) who found a positive correlation between incidence and the average temperature from the preceding three months for Arizona. Kolivras and Comrie (2003) also only found positive correlations in Pima County, Arizona, with summer temperature positively correlated with the subsequent season's incidence rate rather than the preceding season. It is possible that these discrepancies may be due to regional differences as these latter two studies took place in Arizona where there is a different climate pattern. It is also possible that phenotypic differences between the two *Coccidioides* species may play a role as they have different geographical distributions (Fisher, Koenig, White, & Taylor, 2002). *C. immitis* is mainly found in the Central Valley region of California while *C. posadasii* is mainly found in Arizona and other parts of the world (Fisher et al., 2002; Thompson III, 2011). A field study that examined a known *Coccidioides* spp. soil site on the border between Inyo and Kern Counties in California found that during August, October, and November (September was not included), the fungus was not present in the top layer of soil (Plunkett & Swatek, 1957). It is thought that the *Coccidioides* spp. survive in the hotter, drier months by moving deeper into the soil (Maddy, 1965). Though the months don't align perfectly with the field study, the movement of the *Coccidioides* spp. into deeper soil in the hotter months and back to the surface in cooler months may partially explain the negative relationship between temperature and valley fever in the summer and the positive relationship in the winter. It is also possible that some of the correlations found are due to chance as the significance level for this study was set to five percent.

Unexpected findings from this study include the inverse relationship found between PM_{10} and incidence in the summer and fall univariate models and the inverse relationship found between wind and incidence in the univariate and multivariable summer models. These findings were unexpected as we hypothesized that increased disturbances would relate directly to increased spore exposure, and wind and PM_{10} are proxies for disturbance and spore concentration in the air. The negative correlation between dust and incidence is contradictory to the Arizona studies by Park et al. (2005) and Comrie (2005), which found positive relationships between PM_{10} and valley fever incidence. Conversely, this finding is supports Tong et al. (2017) who found that PM_{10} was negatively correlated with valley fever in the two Arizona counties with the largest number of cases. The inverse relationship in the current study may be due to high PM_{10} contributions from anthropogenic emissions, road dust, or farming operations in Kern County, all of which do not contribute to fungal spore distribution (Talamantes, Behseta, & Zender, 2007a; Tong et al., 2017). The relationship between PM_{10} and valley fever in California should be investigated further, as should other proxies for spore concentration in the atmosphere, such as the number and frequency of dust storms (Tong et al., 2017). Other disturbance mechanisms, such as anthropogenic soil disturbance from construction, may play a more significant role than climatic factors in causing spores to become airborne.

Model performance varied between the seasonal models. Figures 4a through 4d show that the models all predicted a decrease in cases around 2009 prior to a significant rise around 2011, but there were smaller variations that were missed, such as a spike in incidence in spring 2004. These incorrect predictions are likely a result of the complicated ecology surrounding this disease and indicate that there are other factors influencing incidence that were not considered in these models. The fall models in the univariate and multivariable analyses were the best performing models possibly because fall has the highest incidence rates. The methods used in the present study are similar to those used by Comrie (2005) who also found that his best performing seasonal model, the foresummer season that includes the months of May through July, was the season with the highest valley fever rates in Pima County, Arizona. Kolivras and Comrie (2003) likewise found that their best monthly models, in terms of explained variance, were those months that had a high percent of the total annual incidence. This finding is valuable given that future models using this information may be best able to predict valley fever in the season with the highest incidence rates.

There are a few limitations and assumptions for this research associated with the use valley fever incidence data. First, changes in provider awareness, diagnosis, and/or reporting have presumably improved over time as Kern County's public health department has an active valley fever campaign. While this may account for some of the long term increasing trend, we do not believe this significantly affects the seasonal variations observed and assessed in this study (Ampel, 2010). Second, there are likely inconsistencies in lag times between disease exposure dates and the date of report to the local health department; this was addressed by incorporating a one-month lag in the incidence data and by conducting this study at sufficiently broad time-scale (i.e, seasonally) in attempt to reduce such error. Next, this research assumes that a majority of the cases reported are from those who acquired the disease in the same county where they live and receive medical care; we believe it is unlikely that variations in incidence are heavily affected by those traveling in and out of the county due to the consistent and high volume of cases in the county. Finally, in 2010 California transitioned to a laboratory-based reporting system which may have affected (increased) the number of reported cases. Kern County had been using a lab-based reporting system prior to 2010, so it is unlikely that administrative changes are predominantly responsible for the spike seen in 2011, but is still possible that some of the increase seen was artifactual (Tsang et al., 2013). This research did not consider changes in population characteristics, exposures, or comorbidities.

Conclusion

This study builds on existing literature to better understand the impact of climate and seasonality on the disease ecology of the *Coccidioides* spp. The main objective of this research was to examine and quantify relationships between valley fever and climatic factors in Kern County, California. We confirm findings of previous studies that found higher precipitation amounts to be related to higher incidence rates in subsequent seasons, and we provide an indication of seasonal timing between precipitation and incidence. While other factors such as anthropogenic soil disturbance require further research, we conclude that seasonal cycles of valley fever can be partially explained by climatic variables. These findings clarify some of the uncertainty regarding relationships between climate and valley fever in Kern County, California.

An improved understanding of the climatic mechanisms that may contribute to the growth and dispersal of the *Coccidioides* spp. has practical health applications. Specifically, the ability to define winter precipitation as a critical event related to fall incidence in Kern County may provide a useful public health tool for disease prevention. It also opens possibilities for the future development of an advanced warning system for disease outbreaks. By quantifying the climatic variables correlated to valley fever rates in Kern County, California, this research has created a basis for further research and predictive analysis as valley fever rates continue to trend upward.

References

Alley, W. M. (1984). The Palmer Drought Severity Index: Limitations and Assumptions. Journal of Climate and Applied Meteorology. https://doi.org/10.1175/1520-0450(1984)023<1100:TPDSIL>2.0.CO;2

Ampel, N. M. (2010). What 's Behind the Increasing Rates of Coccidioidomycosis in Arizona and California? Current Infectious Disease Reports, 12, 211–216. https://doi.org/10.1007/s11908-010-0094-3

California Air Resources Board. (2017). Inhalable Particulate Matter and Health (PM2.5 and PM10). Retrieved January 4, 2018, from https://www.arb.ca.gov/research/aaqs/common-pollutants/pm/pm.htm

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

Cole, G. T., Hurtgen, B. J., & Hung, C. Y. (2012). Progress toward a human vaccine against coccidioidomycosis. Current Fungal Infection Reports, 6, 235–244. https://doi.org/10.1007/s12281-012-0105-y

Comrie, A. C. (2005). Climate Factors Influencing Coccidioidomycosis Seasonality and Outbreaks. Environmental Health Perspectives, 113(6), 688–692. https://doi.org/10.1289/ehp.7786

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

Coopersmith, E. J., Bell, J. E., Benedict, K., Shriber, J., McCotter, O., & Cosh, M. H. (2017). Relating coccidioidomycosis (valley fever) incidence to soil moisture conditions. GeoHealth, 51–63. https://doi.org/10.1002/2016GH000033

Eckmann, B., Schaefer, G. L., & Huppert, M. (1964). Bedside Interhuman Transmission of Coccidioidomycosis Via Growth on Fomites. An Epidemic Involving Six Persons. American Review of Respiratory Disease, 89, 175–185.

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

Fisher, M. C., Koenig, G. L., White, T. J., & Taylor, J. W. (2002). Molecular and phenotypic description of Coccidioides posadasii sp. nov., previously recognized as the non-California population of Coccidioides immitis. Mycologia, 94(1), 73–84.

Flynn, N. M., Hoeprich, P. D., Kawachi, M. M., Lee, K. K., Lawrence, R. M., Goldstein, E., … Wong, G. A. (1979). An Unusual Outbreak of Windborne Coccidioidomycosis. The New England Journal of Medicine, 301(7), 358–361.

Friedman, L., Smith, C. E., Pappagianis, D., & Berman, R. J. (1956). Survival of Coccidioides immitis under controlled conditions of temperature and humidity. American Journal of Public Health, 46, 1317–1324.

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

Hector, R. F., Rutherford, G. W., Tsang, C. A., Erhart, L. M., McCotter, O., Anderson, S. M., … Galgiani, J. N. (2011). The Public Health Impact of Coccidioidomycosis in Arizona and California. International Journal of Environmental Research and Public Health, 8, 1150–1173. https://doi.org/10.3390/ijerph8041150

Hidalgo, B., & Goodman, M. (2013). Multivariate or multivariable regression? American Journal of Public Health, 103(1), 39–40. https://doi.org/10.2105/AJPH.2012.300897

JMP. (2017). K-Fold Cross Validation. Retrieved January 5, 2018, from https://www.jmp.com/support/help/13-2/K-Fold_Cross_Validation.shtml

Johnson, J. B., & Omland, K. S. (2004). Model selection in ecology and evolution. Trends in Ecology and Evolution, 19(2), 101–108. https://doi.org/10.1016/j.tree.2003.10.013

Kolivras, K. N., & Comrie, A. C. (2003). Modeling valley fever (coccidioidomycosis) incidence on the basis of climate conditions. International Journal of Biometeorology, 47, 87–101. https://doi.org/10.1007/s00484-002-0155-x

Kolivras, K. N., Johnson, P. S., Comrie, A. C., & Yool, S. R. (2001). Environmental variability and coccidioidomycosis (valley fever). Aerobiologia, 17, 31–42.

Komatsu, K., Vaz, V., McRill, C., Colman, T., Comrie, A., Sigel, K., … Park, B. (2003). Increase in Coccidioidomycosis — Arizona, 1998-2001. Journal of the American Medical Association, 289(12), 1500–1501.

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

National Centers for Environmental Information. (2014). Divisional Temperature-Precipitation-Drought. Retrieved January 4, 2018, from ftp://ftp.ncdc.noaa.gov/pub/data/cirs/climdiv/divisional-readme.txt

Palmer, W. C. (1965). Meteorological Drought (Vol. 30). Washington DC: Department of Commerce, Weather Bureau.

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

Pappagianis, D., & Einstein, H. E. (1978). Tempest from Tehachapi takes toll or Coccidioides conveyed aloft and afar. The Western Journal of Medicine, 129, 527–530.

Park, B. J., Sigel, K., Vaz, V., Komatsu, K., Mcrill, C., Phelan, M., … Hajjeh, R. A. (2005). An epidemic of coccidioidomycosis in Arizona associated with climatic changes, 1998-2001. The Journal of Infectious Diseases, 191(11), 1981–1987. https://doi.org/10.1086/430092

Pianalto, F. S. (2013). Estimating sources of valley fever pathogen propagation in southern Arizona: a remote sensing approach. The University of Arizona. Retrieved from http://arizona.openrepository.com/arizona/handle/10150/311322

Plunkett, O. A., & Swatek, F. E. (1957). Ecological Studies of Coccidioides immitis. In Proceedings of the Symposium on Coccidioidomycosis (pp. 158–160). Phoenix, AZ: US Public Health Services.

Reed, R. E. (1960). Ecology and Epizoology of Coccidioidomycosis. In Reports of papers delivered at the 1959-1960 Conventions of the Intermountain Veterinary Medical Association.

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

Smith, C. E., Beard, R. R., Rosenberger, H. G. M., & Whiting, E. G. M. (1946). Effect of Season and Dust Control on Coccidioidomycosis. The Journal of the American Medical Association, 132(14), 833–838.

Sondermeyer, G. L., Lee, L. A., Gilliss, D., & Vugia, D. J. (2016). Coccidioidomycosis-Associated Deaths in California, 2000 – 2013. Public Health Reports, 131(4), 531–535.

Stacy, P. K. R., Comrie, A. C., & Yool, S. R. (2012). Modeling Valley Fever Incidence in Arizona Using a Satellite-Derived Soil Moisture Proxy. GIScience & Remote Sensing, 49(2), 299–316. https://doi.org/10.2747/1548-1603.49.2.299

Talamantes, J., Behseta, S., & Zender, C. S. (2007a). Fluctuations in climate and incidence of coccidioidomycosis in Kern County, California. Annals of the New York Academy of Sciences, 1111, 73–82. https://doi.org/10.1196/annals.1406.028

Talamantes, J., Behseta, S., & Zender, C. S. (2007b). Statistical modeling of valley fever data in Kern County, California. International Journal of Biometeorology, 51(4), 307–313. https://doi.org/10.1007/s00484-006-0065-4

Tamerius, J. D., & Comrie, A. C. (2011). Coccidioidomycosis incidence in Arizona predicted by seasonal precipitation. PLoS ONE, 6(6). https://doi.org/10.1371/journal.pone.0021009

Thompson III, G. R. (2011). Strategies to combat coccidioidomycosis: Are we making any progress? Current Fungal Infection Reports, 5(4), 215–223. https://doi.org/10.1007/s12281-011-0066-6

Tong, D. Q., Wang, J. X. L., Gill, T. E., Lei, H., & Wang, B. (2017). Intensified dust storm activity and Valley fever infection in the southwestern United States. Geophysical Research Letters, 44, 4304–4312. https://doi.org/10.1002/2017GL073524

Tsang, C. A., Tabnak, F., Vugia, D. J., Benedict, K., Chiller, T., & Park, B. J. (2013). Increase in Reported Coccidiodomycosis - United States, 1998-2011. Morbidity and Mortality Weekly Report, 62(12), 217–221. https://doi.org/http://dx.doi.org/10.1108/17506200710779521

U.S. Census Bureau. (2000). Profile of General Demographic Characteristics: 2000 Census, 2000 Summary File 1 (SF 1) 100% Data. Retrieved June 5, 2017, from https://factfinder.census.gov/

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/

Western Regional Climate Center. (2017). Climate of California. Retrieved January 5, 2018, from https://wrcc.dri.edu/narratives/CALIFORNIA.htm

Zender, C. S., & Talamantes, J. (2006). Climate controls on valley fever incidence in Kern County, California. International Journal of Biometeorology, 50(3), 174–182. https://doi.org/10.1007/s00484-005-0007-6

Appendix B. Coccidioides ecological niche manuscript submitted to SSTE

Environmental factors affecting ecological niche of Coccidioides (spp.) and spatial dynamics of valley fever in the United States

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 2001 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. 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 et al., 2007). Valley fever cases have been steadily rising over the past 20 years with approximately 12,000 cases reported in the U.S. in 2016 (Centers for Disease Control and Prevention, 2018). Approximately 200 people a year die from this disease (Huang, Bristow, Shafir, & Sorvillo, 2012).

Much of what we know about the range of the U.S. endemic area for coccidioidomycosis (Figure 1) is based on coccidioidin skin testing that occurred over 50 years ago (Edwards & Palmer, 1957). These skin tests along with valley fever case data are used to define the endemic range of this disease because it is very difficult to isolate *Coccidioides* species (spp.) from the soil (Barker et al., 2012; Greene et al., 2000). Based on the skin tests, Kern County in California and Pima, Pinal, and Maricopa Counties in Arizona were found to be the most infective areas in the U.S. along with counties in Texas from the southeast border of New Mexico to Laredo, TX (Maddy & Coccozza, 1964). The known and suspected endemic areas remained focused on the southwestern U.S. until the early 2000s when the discovery of multiple *Coccidioides* spp. growth sites in Washington state 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 have been present for an extended period, or whether they have 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, Goodin, Hanzlicek, et al., 2016). Based on previous research and our understanding of the *Coccidioides'* spp. lifecycle, we hypothesize that climate, soil, elevation, and land cover influence this spatial distribution.

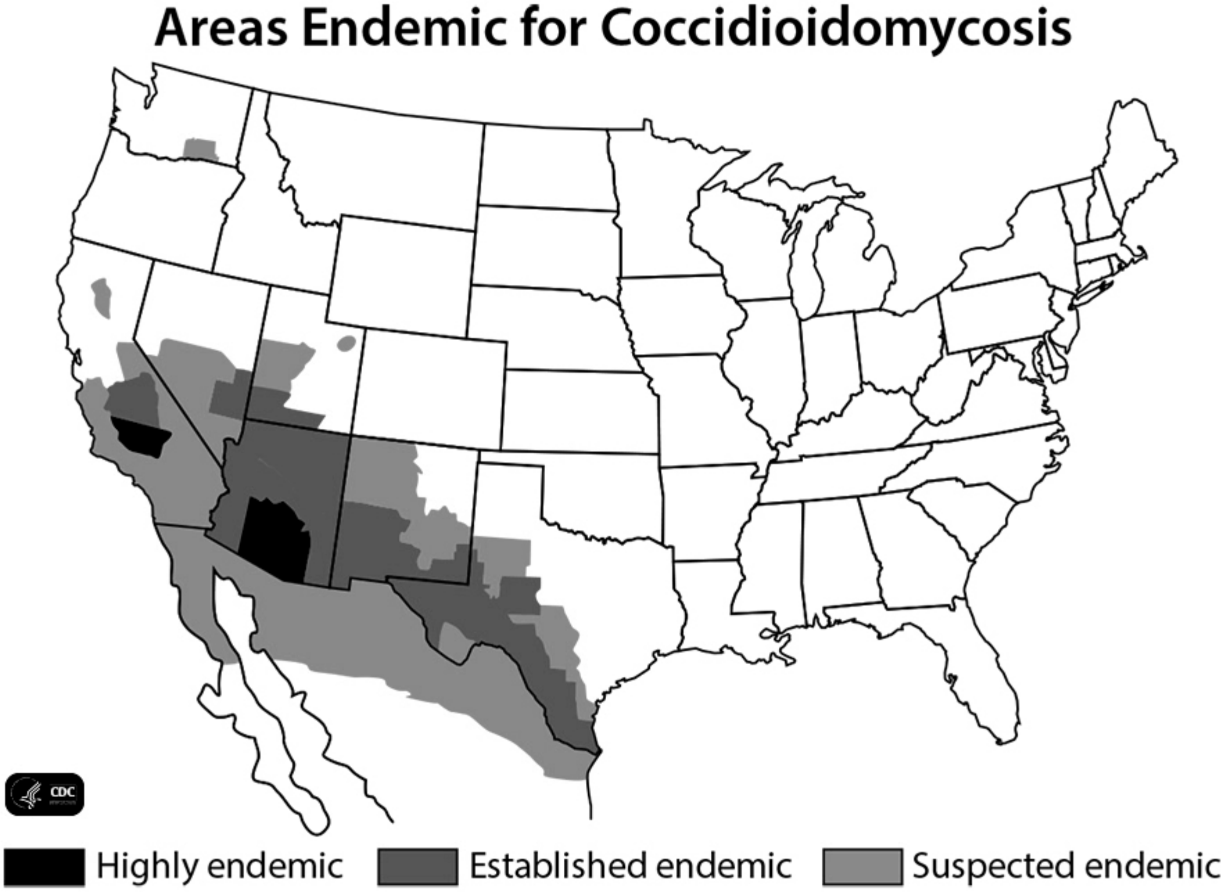


Figure 1. CDC map showing areas where *Coccidioides* spp. are known or suspected to live in the U.S. and Mexico. (source: https://www.cdc.gov/fungal/diseases/coccidioidomycosis/maps.html)

Ecological niche modeling is a relatively unexplored approach that can be used to assess the ecological and geographic distribution of *Coccidioides* species. 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. Once links between geographic occurrences and environmental characteristics are established and characterized, the model can produce a map of the potential distribution of a species (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 (2001-2016) and expanded the region of analysis to include seven states in the western U.S. - California, Nevada, Arizona, Utah, New Mexico, and Washington along with the neighboring state of Oregon. 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 background locations consisting of a random sample of points from across the entire study area. 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 Presence Data

Annual valley fever case totals from 2001 to 2016 were obtained for all states in the study area from the respective state health departments with a few exceptions; Washington data included locally acquired cases from 2014 to 2016 and Oregon data included cases from 2015 to 2016 as valley fever reporting was not mandatory in those states prior to those years. Data was available at the county-scale for California, Arizona, New Mexico, 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 2000 and 2010 census counts and the 5-year American Community Survey 2015 population estimate (U.S. Census Bureau, 2000, 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 for most cases, the county where the report was made was the same county where the disease was acquired, but we acknowledge that this contributes to model uncertainty (Benedict et al., 2018).

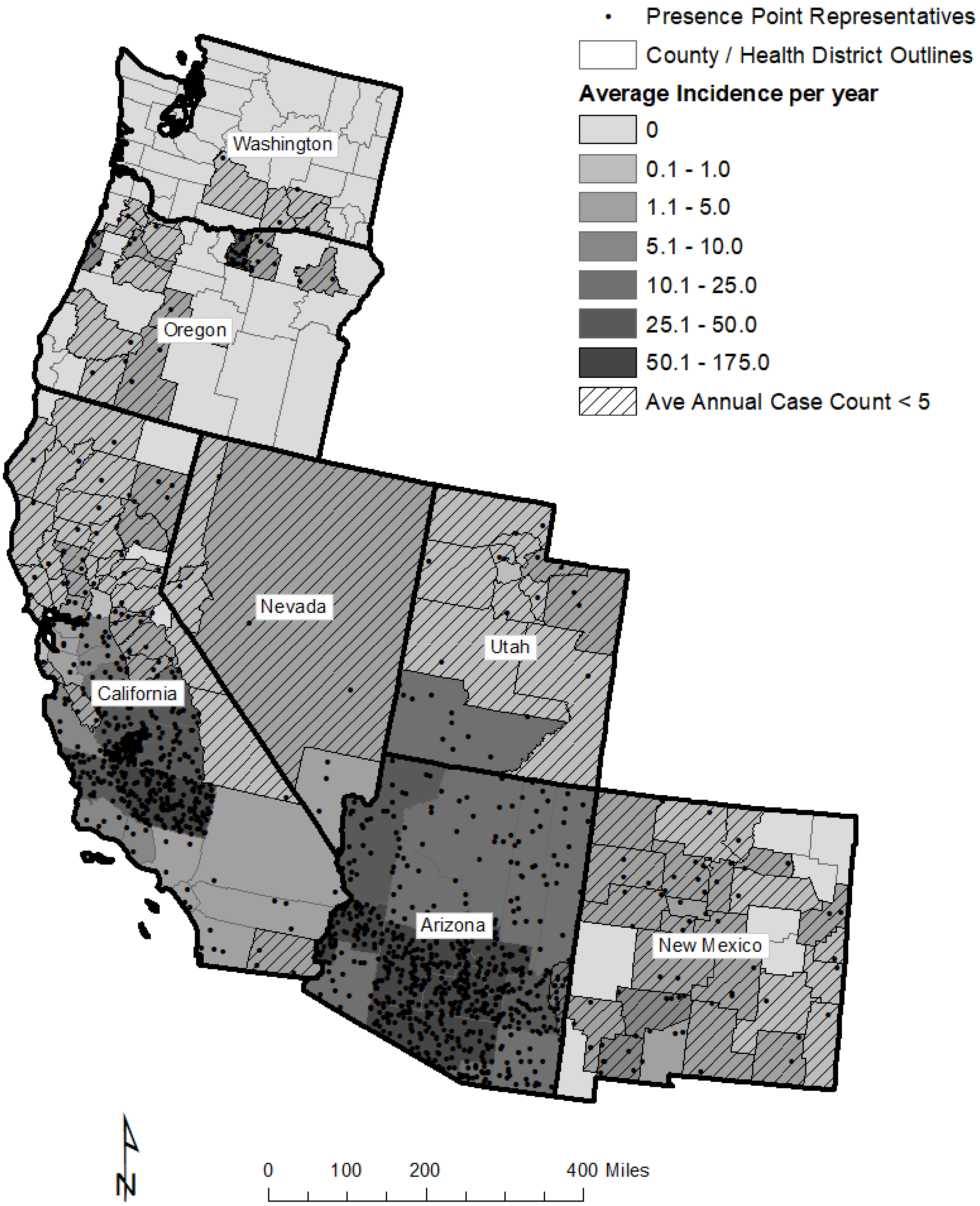


Figure 2. Average valley fever incidence per 100,000 population per year, based on data from 2001-2016, with one set of randomly-generated representative presence points.

2.2 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	$\mathrm{kJ\ m^{-2}\ day^{-1}}$
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

2.3 Data Processing

Previous ENM studies have handled county-level 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).

2.4 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 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) with over 1200 training samples and 135 testing samples. 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). All the models in this study use the same background, therefore AUC should be suitable for comparisons. Predictor variables were assessed by averaging 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.

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; complete results can be found in Appendix A. The first assessment report showed what variables contributed most to the model; these include downward radiation, mean temperature of driest quarter, proportion barren land cover, 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, see Appendix A). 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. The only soil and elevation variables that contribute significant information were silt and elevation, both having a negative relationship with probability of presence when assessed individually.

Table 2. Top variables contributing to model development.

Rank	Variable Contribution	%	Permutation Importance	%
1	Annual Mean Temp	28.2	Downward Radiation	38.5
2	Mean Temperature of Driest Quarter	19.0	Mean Temperature of Driest Quarter	18.8
3	Downward Radiation	11.3	Proportion Barren	15.5
4	Proportion Shrub	7.4	Proportion Cultivated	15.5
5	Proportion Barren	7.0	Proportion Shrub	4.6

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

Rank	Training	Gain	Testing	Gain	AUC	Value
1	Annual Mean Temp	0.33	Annual Mean Temp	0.37	Annual Mean Temp	0.76
2	Mean Temperature of Driest Quarter	0.32	Mean Temperature of Driest Quarter	0.36	Mean Temperature of Driest Quarter	0.75
3	Precipitation of Driest Month	0.25	Precipitation of Driest Month	0.33	Mean Temperature of Driest Quarter	0.74
4	Down. Radiation	0.19	Down. Radiation	0.23	Developed	0.74
5	Silt Fraction	0.14	Elevation	0.16	Down. Radiation	0.74

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

Rank	Training	Gain Lost	Testing	Gain Lost
1	Barren	0.06	Barren	0.06
2	Down. Radiation	0.05	Down. Radiation	0.04
3	Shrub	0.05	Shrub	0.03
4	Cultivated	0.03	Cultivated	0.03
5	Mean Temperature of Driest Quarter	0.02	Herbaceous	0.01

3.2 Potential Distribution

The raw output from the MaxEnt prediction is equivalent to the relative occurrence rate; it shows the probability that a cell is contained in a collection of presence samples (Merow et al., 2013). The map produced from averaging the raw output for the simplified model 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 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.

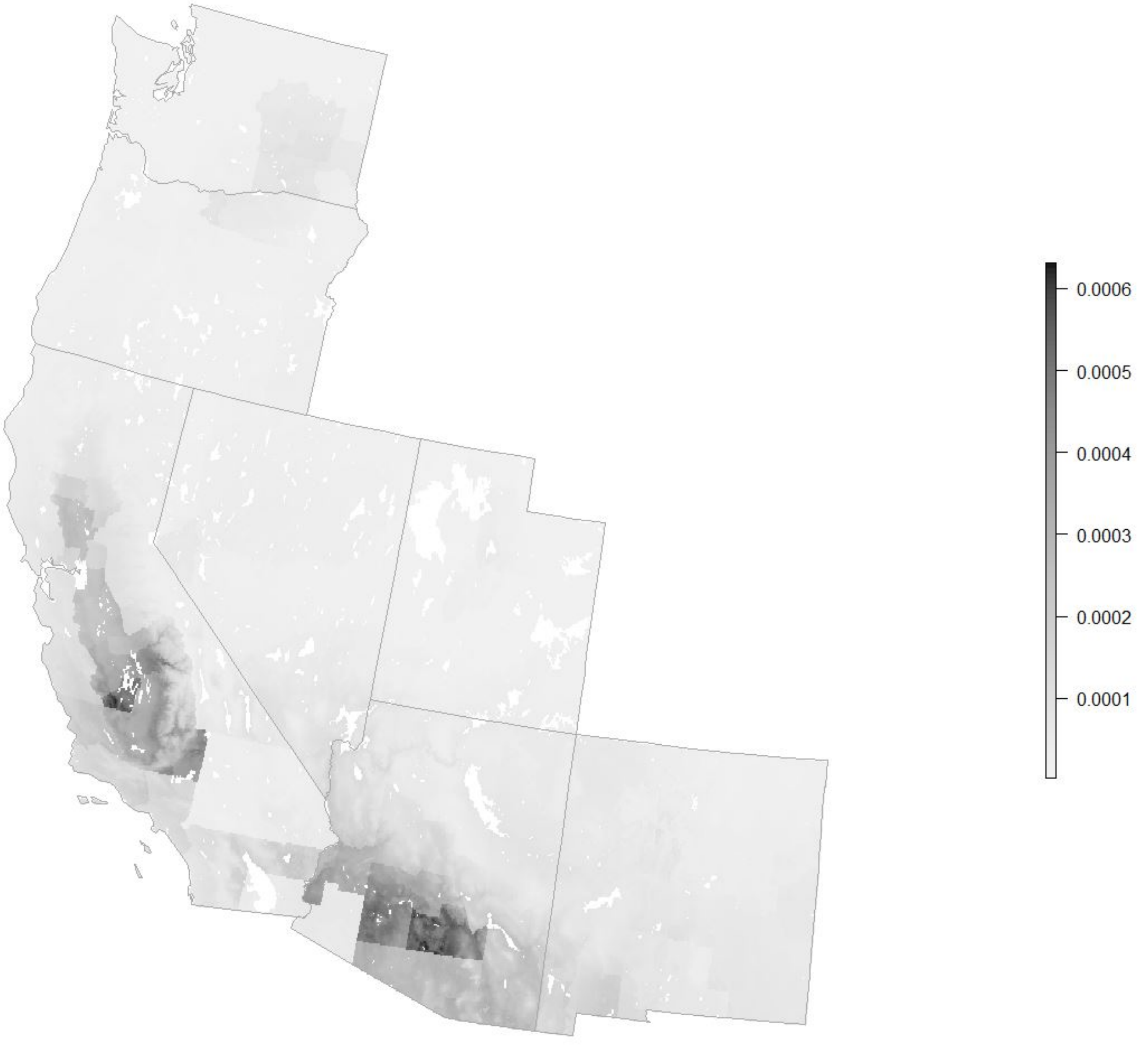


Figure 3. Map of potential distribution based on average raw output from model predictions. Higher values indicate a higher relative occurrence rate.

3.3 Model Evaluation

Model evaluation results show that model testing performance was only slightly below training performance (Tables 5). Testing AUC values for the model ranged from 0.823 to 0.829, which is considerably higher than the null model of 0.5 indicating model performance no better than random. 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 the low standard deviations in both testing and training AUC values.

Table 5. Model Evaluation Results

	Model Average	**Standard Deviation**
Training AUC	0.828	0.002
Testing AUC	0.826	0.002

4. Discussion

This research aimed to improve our understanding of factors affecting *Coccidioides'* spp. 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 proportion barren and mean temperature of the driest quarter, 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'* spp. 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 and precipitation of the driest month 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.

Land cover variables were also prominent in the development of the model, accounting for approximately 20% of model development and 35% 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'* spp. 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'* spp. 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. 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. The implications of this are possible overestimations of potential suitable areas, and inaccuracies in variable importance rankings. Another complication in using case data is that administrative changes can affect reporting requirements. In 2008 the Council of State and Territorial Epidemiologists changed the definition of coccidioidomycosis to allow a single detection of coccidioidal immunoglobulin G antibody to serve as laboratory confirmation of infection, and in 2010, California transitioned to a laboratory-based reporting system; these administrative changes may have increased or decreased the number of reported cases in those states compared to prior years (Tsang et al., 2013), but it is unlikely that these changes significantly impact this study with the broad spatial and temporal scales used.

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 disease occurrence 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 2001-2016. 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'* spp. 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

Valley fever has been considered endemic to the southwestern U.S. for over 50 years, but over the past decade, *Coccidioides* spp. growth sites have been discovered in an unexpected region, south central Washington State (Marsden-Haug et al., 2013). These discoveries put into question what we thought we knew about the endemic area of valley fever and the drivers influencing the niche of *Coccidioides* spp. (Baptista-Rosas et al., 2007; Benedict et al., 2015). This study used ecological niche modeling to further our understanding of *Coccidioides'* spp. realized niche and to define its probable distribution. The model 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

Baptista-Rosas, R. C., Hinojosa, A., & Riquelme, M. (2007). Ecological Niche Modeling of Coccidioides spp. in Western North American Deserts. Annals of the New York Academy of Sciences, 1111, 35–46. https://doi.org/10.1196/annals.1406.003

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

Edwards, P. Q., & Palmer, C. E. (1957). Prevalence of sensitivity to coccidioidin, with special reference to specific and nonspecific reactions to coccidioidin and to histoplamin. Dis. Chest, 31, 355–360.

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.

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

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.

Tsang, C. A., Tabnak, F., Vugia, D. J., Benedict, K., Chiller, T., & Park, B. J. (2013). Increase in Reported Coccidiodomycosis - United States, 1998-2011. Morbidity and Mortality Weekly Report, 62(12), 217–221. https://doi.org/http://dx.doi.org/10.1108/17506200710779521

U.S. Census Bureau. (2000). Profile of General Demographic Characteristics: 2000 Census, 2000 Summary File 1 (SF 1) 100% Data. Retrieved June 5, 2017, from https://factfinder.census.gov/

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.

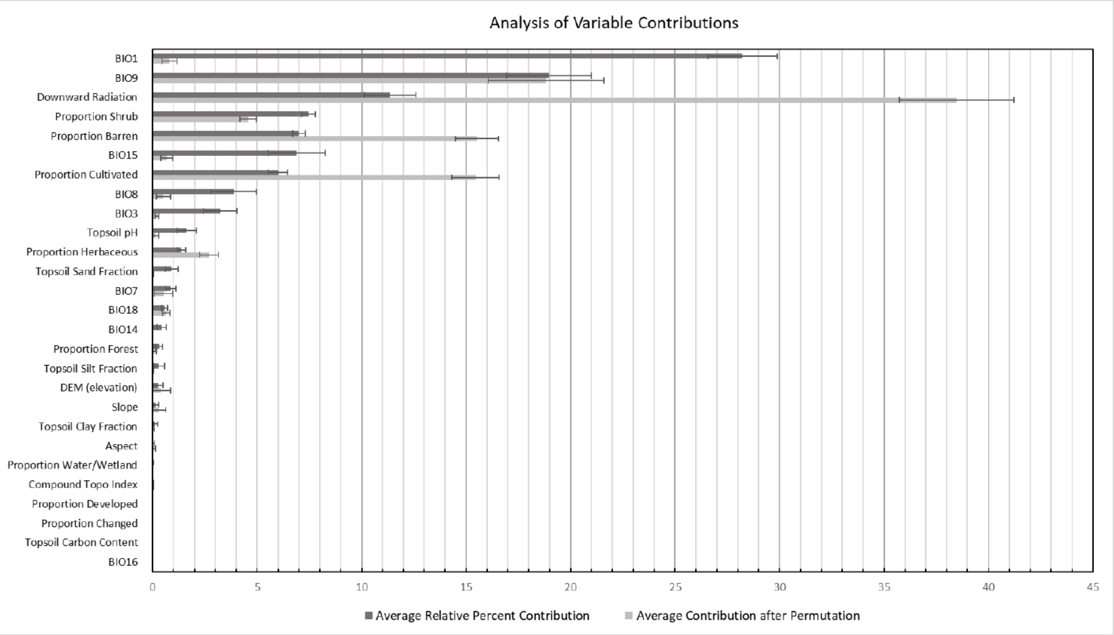
Wieder, W. R., Boehnert, J., Bonan, G. B., & Langseth, M. (2014). Regridded Harmonized World Soil Database v1.2. ORNL DAAC, Oak Ridge, Tennessee, USA. Retrieved from https://doi.org/10.3334/ORNLDAAC/1247

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

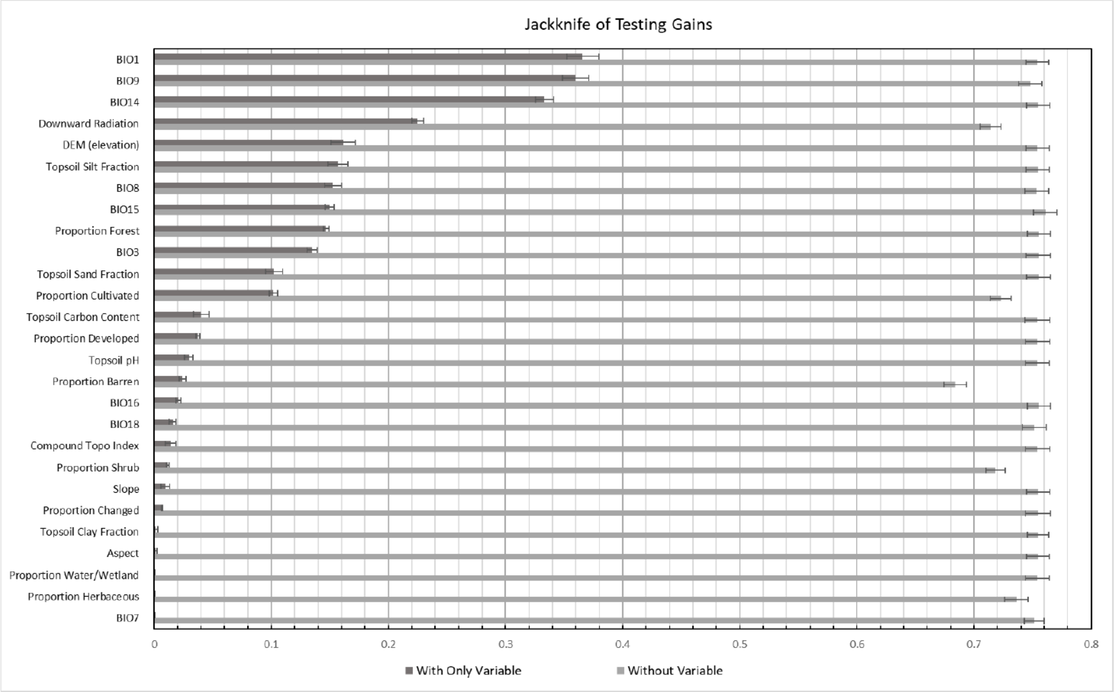
Supplementary Material - Appendix A

The following chart gives estimates of relative contributions of the environmental variables to the MaxEnt model. The bars represent the average of the 25 sets while the error bars show the standard deviation.

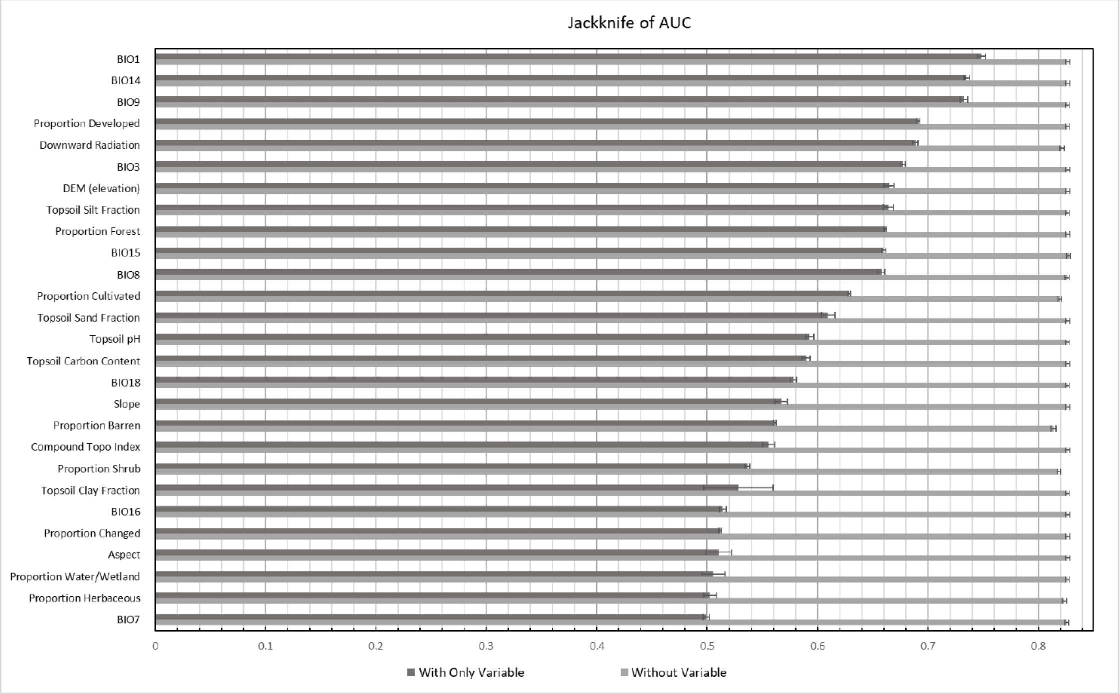


The following chart shows the results of the jackknife test of variable importance based on the training dataset. The bars represent the average of the 25 sets while the error bars show the standard deviation.

This chart shows the same jackknife test, using test gain instead of training gain. The bars represent the average of the 25 sets while the error bars show the standard deviation.



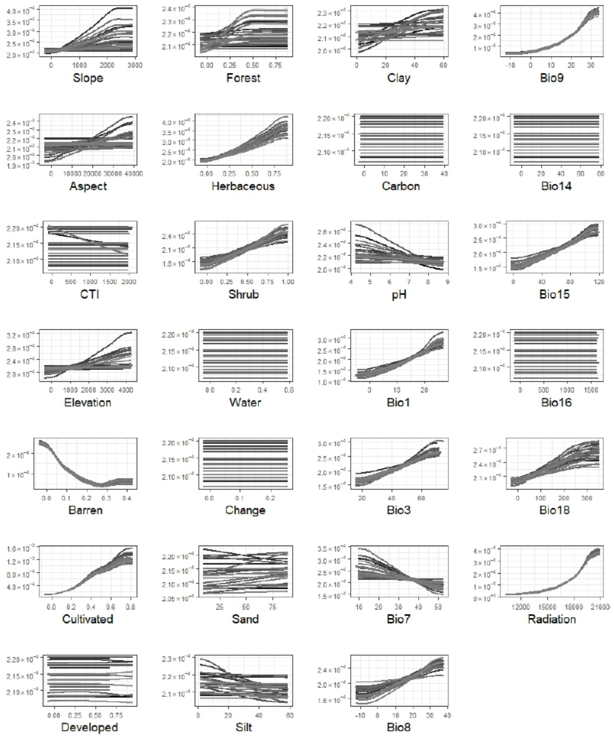
This chart shows the same jackknife test, using AUC on test data. The bars represent the average of the 25 sets while the error bars show the standard deviation.



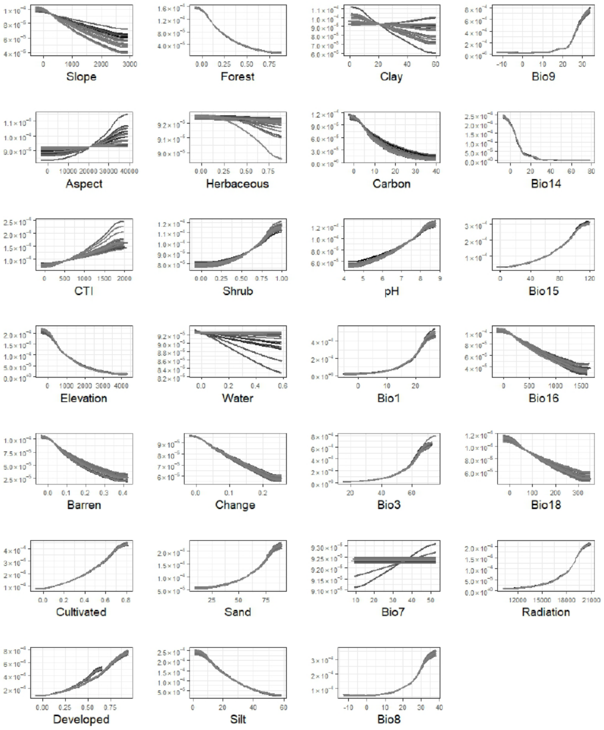
Model performance metrics for the 25 sets of models.

Model Run	**Regularized training gain**	**Unregularized training gain**	**Training AUC**	**Test gain**	**Test AUC**	**AUC Standard Deviation**
Set 1	0.601	0.749	0.825	0.740	0.824	0.016
Set 2	0.623	0.773	0.829	0.764	0.828	0.016
Set 3	0.627	0.777	0.831	0.766	0.829	0.016
Set 4	0.619	0.769	0.829	0.762	0.828	0.016
Set 5	0.630	0.776	0.830	0.768	0.829	0.016
Set 6	0.605	0.757	0.827	0.748	0.825	0.016
Set 7	0.614	0.765	0.829	0.753	0.826	0.016
Set 8	0.612	0.760	0.827	0.751	0.826	0.016
Set 9	0.623	0.773	0.830	0.763	0.828	0.016
Set 10	0.634	0.788	0.833	0.779	0.831	0.016
Set 11	0.620	0.764	0.828	0.756	0.826	0.016
Set 12	0.613	0.758	0.827	0.748	0.825	0.016
Set 13	0.612	0.762	0.828	0.752	0.826	0.016
Set 14	0.621	0.764	0.828	0.755	0.827	0.016
Set 15	0.607	0.752	0.825	0.742	0.824	0.016
Set 16	0.618	0.764	0.828	0.754	0.827	0.016
Set 17	0.607	0.754	0.826	0.745	0.824	0.016
Set 18	0.618	0.767	0.829	0.759	0.827	0.016
Set 19	0.626	0.775	0.831	0.768	0.829	0.016
Set 20	0.602	0.747	0.825	0.735	0.823	0.016
Set 21	0.611	0.760	0.829	0.753	0.827	0.016
Set 22	0.606	0.752	0.826	0.740	0.824	0.016
Set 23	0.621	0.767	0.829	0.759	0.827	0.016
Set 24	0.626	0.771	0.829	0.763	0.828	0.016
Set 25	0.615	0.759	0.828	0.751	0.826	0.016
Average	**0.616**	**0.764**	**0.828**	**0.755**	**0.826**	**0.016**

These curves show how each environmental variable affects the MaxEnt prediction. The curves show how the predicted probability of presence changes as each environmental variable is varied, keeping all other environmental variables at their average sample value. The curves show the marginal effect of changing exactly one variable, whereas the model may take advantage of sets of variables changing together. The individual lines depict the result of each of the 25 sets of replicate models.



In contrast to the previous marginal response curves, each of the following curves represents a different model, namely, a MaxEnt model created using only the corresponding variable. These plots reflect the dependence of predicted suitability both on the selected variable and on dependencies induced by correlations between the selected variable and other variables. The individual lines depict the result of each of the 25 sets of replicate models.



Appendix C. Spatio-temporal relative risk modeling manuscript

Spatiotemporal modeling of relative risk of coccidioidomycosis in California

Manuscript for 'Applied Geography'

Abstract

Valley fever, or coccidioidomycosis, cases are rapidly rising in California and the causes are unclear. Spatio-temporal disease mapping can be used to better understand where valley fever risk is increasing and to identify environmental risk factors associated with the disease. This study applies a hierarchical Bayesian spatio-temporal model to analyze county-scale valley fever relative risk in California from 2001 to 2018. Additionally, we evaluated the contribution of annual temperature, annual precipitation, precipitation from the previous year, mean temperature of the driest quarter, and percent cultivated land cover in explaining the patterns observed in relative risk. Results show a steady increase in valley fever risk over the study period and increasing relative risk trends along the central coast and into the northern Central Valley of California. Eight counties were classified as high risk and four counties were classified as medium risk. Statistically significant covariates include both precipitation variables and percent cultivated land cover. The findings from this study can be applied to focusing resources on counties classified as high or medium risk with increasing temporal trends. The application of a Bayesian spatio-temporal model and identification of significant environmental covariates may additionally be used in future studies that forecast valley fever based on environmental conditions.

1. Introduction

Valley fever or coccidioidomycosis is an understudied fungal disease affecting populations in the western U.S., as well as in parts of Central and South America. Over the last two decades case counts in the U.S. have been increasing, but over the last four years there has been a dramatic spike in the number of cases in California with an increase from approximately 2,300 in 2014 to a record high of 7,446 in 2017 (California Department of Public Health, 2018). Common symptoms of valley fever include flu-like symptoms such as fever and cough, but the disease can also disseminate to other parts of the body and cause serious, lifelong illnesses such as chronic meningitis (Rosenstein et al., 2001).

The reason for the sudden increase in valley fever cases in California is uncertain but it could in part be due to changes in environmental conditions. Researchers have hypothesized that there is a relationship between valley fever and environmental conditions based on our knowledge of the fungus' lifecycle. Valley fever is caused by inhalation of fungal spores from *Coccidioides immitis* or *Coccidioides posadasii*. *Coccidioides* species typically live and get their nutrients from the soil where they take a hyphal form (Barker, 2017). As part of their lifecycle, they also fragment into microscopic spores known as arthroconidia that can easily become airborne. When these airborne arthroconidia are inhaled by a human or other mammalian host they change into a spherule form and become pathogenic (Barker, 2017). The spherules break apart inside the host into tiny endospores, which then grow into spherules, allowing the cycle to repeat and causing illness. Researchers have hypothesized that wet environmental conditions foster growth of the fungus in the soil, then subsequent dry periods promote fragmentation and allow spores to become airborne resulting in valley fever illnesses (Comrie & Glueck, 2007; Kolivras & Comrie, 2004; Tamerius & Comrie, 2011). Research over the past 15 years has found and quantified statistically significant valley fever – environment relationships in regional and single-county studies focused mainly in Arizona and California (Comrie, 2005; Coopersmith et al., 2017; Gorris, Cat, Zender, Treseder, & Randerson, 2017; Park et al., 2005; Stacy, Comrie, & Yool, 2012; Tamerius & Comrie, 2011; Weaver & Kolivras, 2018). Results from these studies have not been consistent; they vary with the spatial and temporal scales used and locations assessed. There are also a few studies that found only weak correlations or a lack of statistically significant relationships between climate and valley fever (Talamantes, Behseta, & Zender, 2007; Zender & Talamantes, 2006).

Past valley fever studies include descriptive analyses (Maddy, 1965; Maddy & Coccozza, 1964; Petersen et al., 2004), time-series analyses (Comrie, 2005; Stacy et al., 2012; Talamantes, Behseta, & Zender, 2007), and ecological or spatial analyses (Baptista-Rosas et al., 2007; Gorris et al., 2017; Grayzel, Martínez-López, & Sykes, 2017), that are often aggregated over time or area and are focused on disease incidence rather than risk. Epidemiological disease mapping techniques that assess both the spatial and temporal patterns of disease risk are useful in developing risk maps for surveillance and forecasting. Such techniques have not yet been applied to valley fever; therefore, this research investigates county-level, spatio-temporal valley fever relative risk trends in California and evaluates potential drivers of those trends using data from 2001-2018. A Bayesian spatio-temporal framework is well-suited for this type of research because it allows risk to be divided into spatial, temporal, and spatial-temporal interaction effects (Kim & Kim, 2018). Bayesian spatio-temporal models can account for observation and process uncertainty, and spatial and temporal autocorrelation in the data (Dietze, 2017; Khana, Rossen, Hedegaard, & Warner, 2018). Bayesian spatio-temporal modeling has been used to assess disease risk and covariate relationships for diseases such as dengue (Lowe et al., 2014; Martınez-Bello, Lopez-Quilez, & Prieto, 2018), scrub typhus (Kim & Kim, 2018), Rocky Mountain spotted fever (Raghavan, Goodin, Neises, Anderson, & Ganta, 2016) and hand-foot-mouth disease (Tian et al., 2018) but this research is the first to apply these methods to valley fever.

This study has three aims to address gaps in the literature: 1) to develop a model to investigate spatio-temporal trends of valley fever in California, 2) to identify high risk areas and temporal changes in those areas, and 3) to evaluate environmental risk factors hypothesized to be associated with valley fever. The understanding of relative risk patterns is important for targeting awareness and educational campaigns and promoting diagnostic testing in areas where a disease is emerging. Identification of environmental risk factors could help establish the utility of environmental monitoring as an aide for valley fever surveillance. Furthermore, this study demonstrates how Bayesian spatio-temporal disease mapping can be applied by those studying disease distributions to obtain valuable information regarding both where disease risk is greatest and how that risk has evolved through space and time.

2. Materials and Methods

2.1 Disease data and environmental covariates

The study area for this research includes all counties in California (see FIGURE 1). California has the second highest case counts and incidence rates in the U.S. and accounted for approximately half of the 2017 cases. Valley fever reporting was required statewide during the study period, 2001-2018. The annual county-scale data used in this study was acquired from the California Department of Public Health (CDPH, 2019); the 2018 data is provisional at the time of this writing. County population data was from the U.S. Census Bureau and was linearly interpolated based on 2001, 2010, 2015 census data (U.S. Census Bureau, 2000, 2010, 2015).

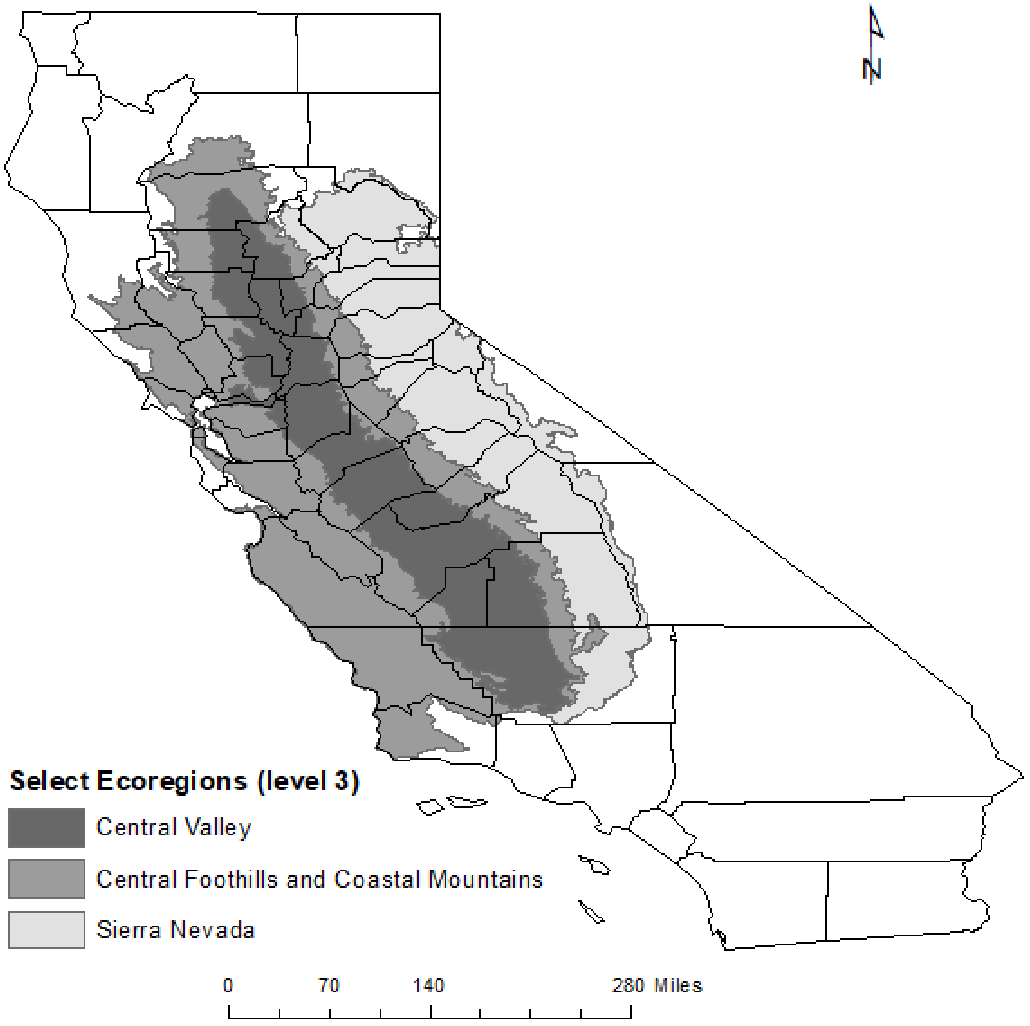


FIGURE 1. Study area showing select U.S. Environmental Protection Agency Level III Ecoregions in California (EPA, 2018).

The environmental covariates included in the model were selected based on associations found in previous research and have known or hypothesized ecological relationships with valley fever (Baptista-Rosas et al., 2007; Comrie, 2005; Gorris et al., 2017; Weaver & Kolivras, 2018). Most of the previous studies used monthly or seasonal temporal scales in their analyses; this research evaluates whether these relationships hold when aggregated to the annual scale. The covariates assessed that vary over time included annual precipitation (concurrent), annual precipitation (lagged one year), mean annual temperature, and mean temperature of the driest quarter. The data for these covariates was from the Multivariate Adaptive Constructed Analogs (MACA) dataset (https://climate.northwestknowledge.net/MACA/). The MACA dataset covers the study area, was available for the entire study period, and has climate projections needed for disease forecasting, a potential future use of this model. The only temporally static covariate assessed was percentage of landcover that is cultivated. The landcover data is from the 2011 National Land Cover Database (https://www.mrlc.gov/data). This set was chosen because it is closest to the mid-point of the temporal range of the study period. R was used to transform the covariate data to the same spatial resolution as the disease data and to calculate zonal statistics (R Core Team, 2003). All covariates were standardized to have a mean of zero and a standard deviation of one for comparability.

2.2 Bayesian Modeling

Using a Bayesian framework, Poisson regression, commonly used for count data, was used to quantify spatio-temporal variation in valley fever and associations with selected environmental factors. The general equation is given in formula (1) below. It can be thought of as having a component representing the background disease risk (e_{it}) and a component representing the excess risk in a specific area-time unit (θ_{it}). The excess (relative) risk is the focus of the modeling for this study (Lawson, 2018).

(1) $y_{it} \sim Poisson(\lambda_{it})$; $\lambda_{it} = e_{it}\theta_{it}$

- y_{it} is the observed number cases
- λ_{it} is the modeled average number of cases
- e_{it} is the expected number of cases
- θ_{it} is the relative risk
- *i* is the *i*th area (*i* = 1, …58)
- *t* is the *t*th time unit (*t* = 1, …16)

The expected number of cases is computed using equation (2). Local expected cases are calculated by multiplying the local population by standardized incidence rate. The standardized incidence rate is the overall disease count divided by total population (of the study area).

(2) $e_{it} = n_{it}\frac{\sum_i \sum_t y_{it}}{\sum_i \sum_t n_{it}}$

- n_{it} is the population in the *i* – *t*th area-time unit.

This research will estimate the relative risk using two models. The first model is termed the null model and does not contain any covariates. The second model is termed the full model and contains the selected covariates. The null model uses random effects to represent unobserved risk factors that vary over space or time or both space and time. It is a log link function which contains spatial, temporal, and spatio-temporal random effects terms to represent the disease occurrence patterns. The null model used is given by equation (3).

(3) $\log(\theta_{it}) = \alpha + u_i + v_i + \varphi_t + \delta_{it}$

- α is the intercept term or the mean log risk for the overall study area and study period.
- u_i is a structured area random effect to account for spatial clustering or spatial dependency. It uses an intrinsic conditional autoregressive (iCAR) prior. The conditional expectation of u_i is equal to the mean of the random effects in neighboring areas; the conditional variance is inversely proportional to the number of neighbors (Besag, York, & Mollie, 1991; Lawson, 2018). Prior: $u_i|u_{-i}, W, \tau_u \sim N(\frac{\sum_{h\sim i} u_h}{n_i}, \frac{1}{\tau_u n_i})$
    - u_{-i} is the vector of spatial effects omitting county i
    - W is a county spatial adjacency matrix. Weights are assigned to each county according to adjacency where neighboring counties (counties that share a border) receive a weight of one and non-neighboring counties a weight of zero (Source: US Census).
    - τ_u represents the precision (reciprocal of variance) that controls the amount of variation between the random effects. $\tau_u \sim$ Gamma(1,0.00005).
    - $h \sim i$ represents neighboring counties
    - n_i is the number of neighbors of county i
- v_i is an unstructured area random effect to account for unknown spatial factors. It is given a normal prior: $v_i \sim N(0,\tau_v)$. Precision $\tau_v \sim$ Gamma(1,0.00005).
- φ_t is an unstructured temporal random effect to account for unknown factors or temporal noise. It is given a normal prior: $\varphi_t \sim N(0,\tau_\varphi)$. $\tau_\varphi \sim$ Gamma(1,0.00005).

- δ_{it} is an interaction random effect between area and time. This research explores interactions between unstructured area and time effects (a Type I interaction). These can be thought of as unobserved covariates for each area-time unit that do not have any structure in space*time (Knorr-Held, 2000). It is given a normal prior: $\delta_{it} \sim N(0, \tau_\delta)$. Precision τ_δ = Gamma(1,0.00005).

The full model expands the null model to incorporate environmental covariates to assist in explaining the spatial and temporal patterns. The full model is given by equation (4):

(4) $\log(\theta_{it}) = \alpha + u_i + v_i + \varphi_t + \delta_{it} + \sum_{\boldsymbol{k=1}} \boldsymbol{\beta_k X_{itk}}$

- k is the kth covariate (k = 1,…5)
- X are the covariates
- β is the corresponding vector of regression coefficients. They are given normal priors: $\beta \sim N(0,100000)$.

The models were fit using Integrated Nested Laplace Approximation (INLA), rather than Markov Chain Monte Carlo (MCMC) methods, built and fitted in R (Bivand, Gomez-Rubio, & Rue, 2015). R-INLA uses numerical integration for fixed effects and Laplace integral approximation for the random effects; it does not require sampler convergence (Rue, Martino, & Chopin, 2009). A fundamental difference between these model fitting techniques is that MCMC proves exact inference while INLA provides an approximation to the relevant posterior distribution (Blangiardo, Cameletti, Baio, & Rue, 2013). The benefits of using R-INLA are that it is computationally more efficient than MCMC and provides estimations of regression parameters comparable to other Bayesian software and/or methods (Martino & Rue, 2010; Rue et al., 2009; Schrödle & Held, 2011). R-INLA default values were used for priors; these priors are vague, intended to have little influence on posterior distributions allowing results to be primarily derived from the data (Rue & Martino, 2009). The models were assessed using cross validation and posterior predictive indices, specifically the probability integral transform and posterior predictive distribution (Blangiardo & Cameletti, 2015).

County-level risk classifications were determined using information from the whole posterior distribution rather than just the posterior mean; cutoffs were based on a simulation study from Richardson et al. (2004) that assessed sensitivity and specificity of posterior relative risk estimates in disease mapping. A county was classified as high-risk for valley fever if there was at least an 80% probability of the county-specific relative risk being greater than two, indicating twice the average risk. A moderate risk county was classified as having at least an 80 percent probability of relative risk being greater than one, or above the average risk, and all others were classified as low risk. Temporal trends were graphed for all counties classified as high and medium risk.

3. Results

3.1 Relative Risk Trends.

From 2001 to 2018 a total of 63,612 valley fever cases were reported in CA with 2001 having the lowest count and 2017 having the highest count. FIGURE 2 shows the clearly increasing temporal trend of valley fever risk. Specifically, in 2018 the relative risk in California was approximately 5 times higher than that in 2001, with a sharp rise evident from 2014 to 2017. Valley fever risk had peaks in 2006, 2011, and 2017 while 2008 and 2014 were below average.

FIGURE 2. Overall time trend of the relative risk of valley fever in California with 95% credible interval.

When viewed spatially, relative risk trends show regional distinctions. FIGURE 3 presents the overall relative risk during the study period as estimated from the posterior means of the null model. The counties in California's southern Central Valley show the highest relative risk (see FIGURE 1 for regions). Kern and King Counties (two darkest shades) stand out from the rest of the state with posterior means indicating 44 and 23 times higher risk than average, respectively. This map also indicates that there are counties along the central coast and in the center section of the Central Valley that also have elevated risk. Lower relative risk areas include many of the counties in northern California, along with the counties on the eastern side of the Sierra Nevada mountain range.

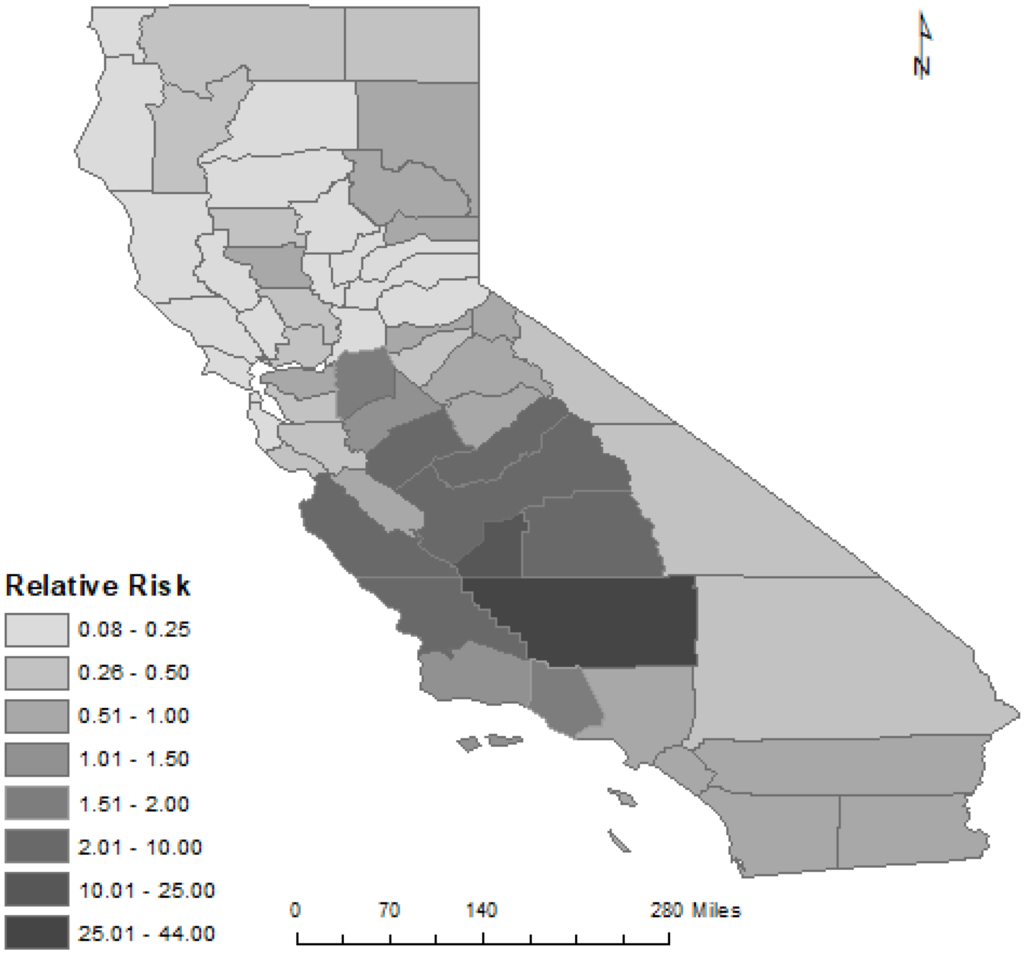


FIGURE 3. The relative risk of valley fever in California (2001-2018). Individuals in the darkest two counties have over ten times higher chance of being infected by valley fever than average.

A key output of this research is the spatio-temporal trend for valley fever. FIGURE 4 shows the spatio-temporal trends from the last five years which captures the sharp increase in valley fever risk seen in the purely temporal trend. Maps for all years can be found in the supplemental material. These maps let us explore how the relative risk of valley fever is changing in both time and space, which is critical understanding disease evolution and trends, but is not often modeled. These maps show whether temporal risk trends are increasing, decreasing, or average for each county. Counties along the central coast of California have had increasing or above average risk for the past three years. There is also a pattern of above average risk in the multiple counties in the center of the state. The two counties in the southern Central Valley that showed the highest relative risk in the spatial assessment have had decreasing temporal trends for the past five years.

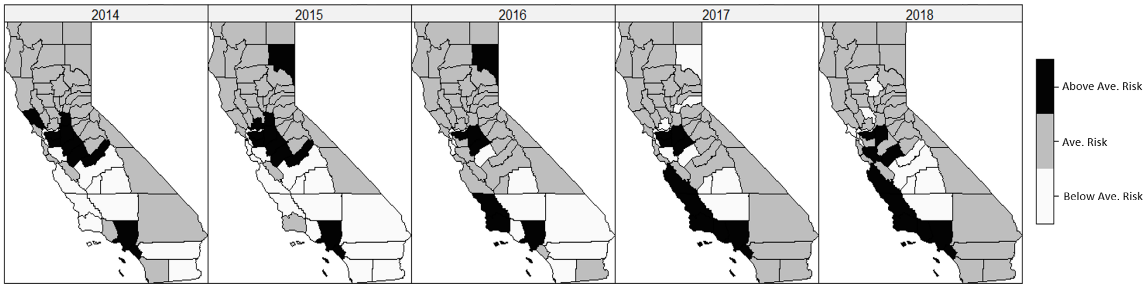


FIGURE 4. Spatio-temporal trends of relative risk of valley fever in California (2014-2018). The darkest shade indicates higher than the average relative risk and the lightest shade indicates lower than the average risk based on an 80% credible interval.

3.2 Risk Classification and County Temporal Trends

All counties in California were classified into risk categories based off the classification standards outlined in Section 2.2. Of the 58 counties in California, 8 (14%) were recognized as high risk, 4 (7%) as medium risk, and 46 (79%) as low risk. The risk classification map is shown in FIGURE 5. The high risk counties include all counties in the southern Central Valley along with the neighboring coastal counties of San Luis Obispo and Monterey. The medium risk counties are all adjacent to the high risk counties, either in the Central Valley or along the coast. Temporal relative risk trends of the high and medium risk counties are shown in FIGURES 6 and 7 respectively. Monterey County, along the coast, and Merced County, in the Central Valley, are both classified as high risk and show increasing temporal trends. San Joaquin County, classified as medium risk, also shows increasing temporal trends. Kern and Tulare Counties show decreasing temporal trends.

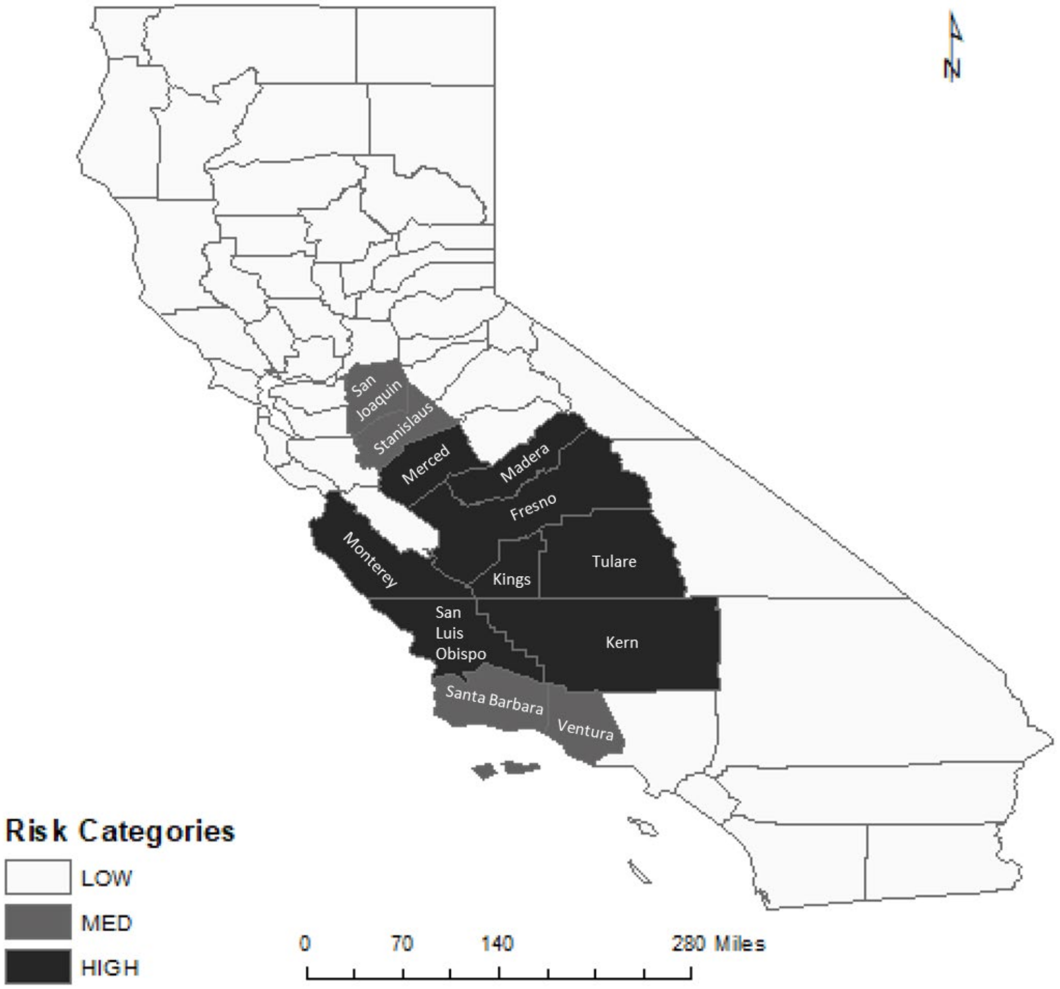


FIGURE 5. Risk Classification. High risk is based on 80% of the posterior distribution indicating at least double the risk of being infected. Medium risk is based on 80% of the posterior distribution indicating above average risk for being infected.

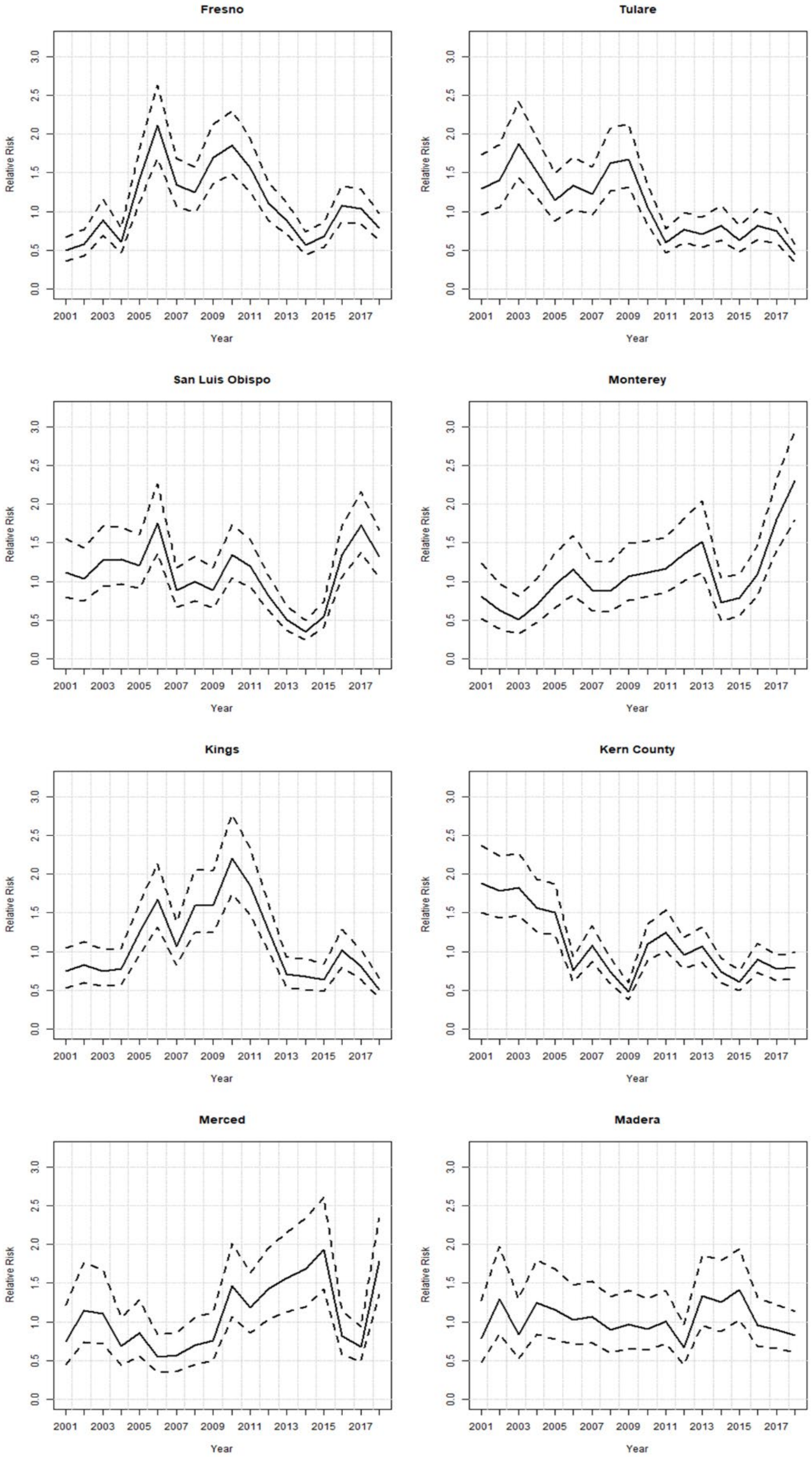


FIGURE 6. Temporal relative risk trends for high risk counties with dashed lines indicating 95% credible interval.

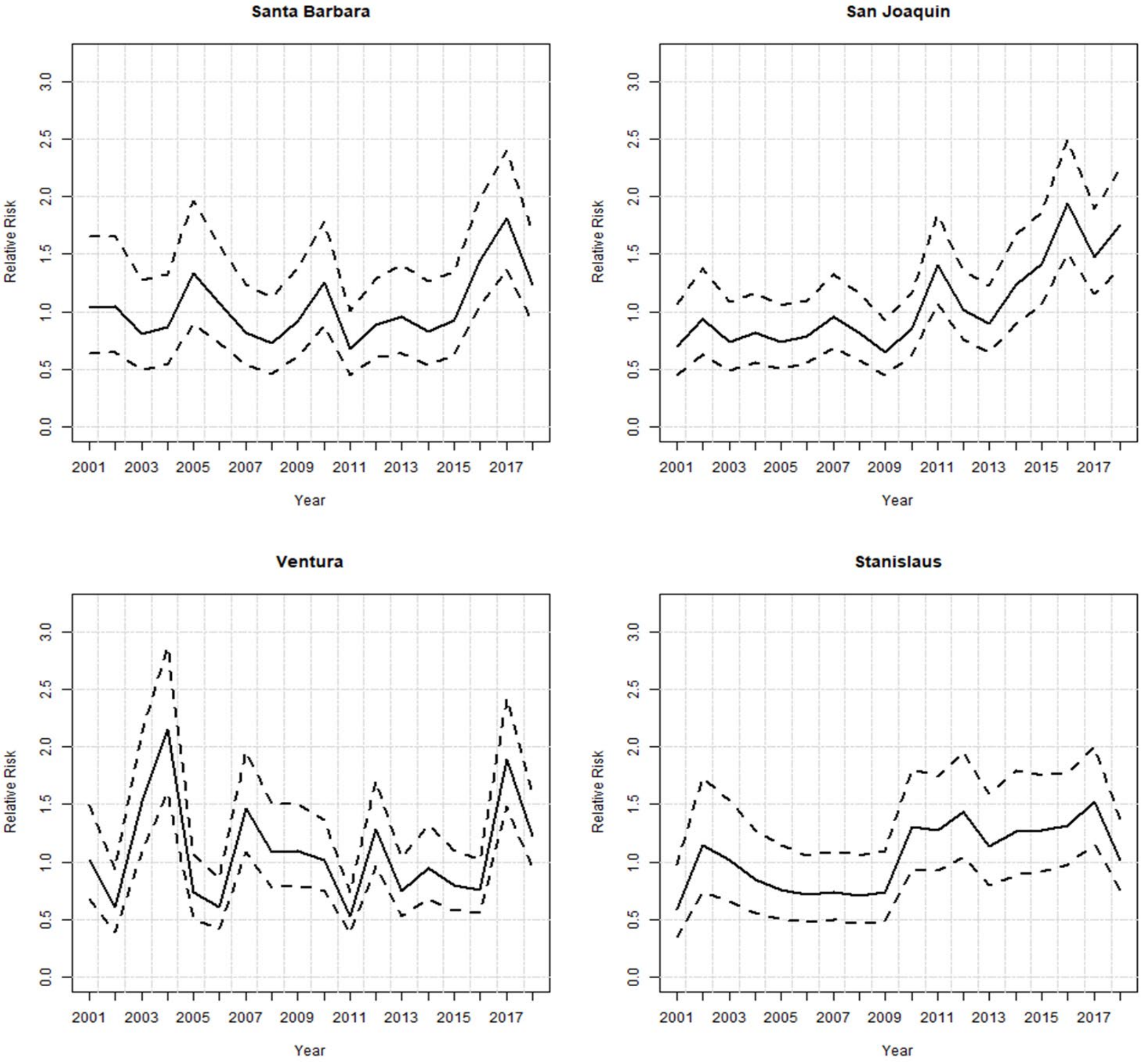


FIGURE 7. Temporal relative risk trends for medium risk counties with dashed lines indicating 95% credible interval.

3.3. Evaluation of Environmental Risk Factors.

Five covariates were added to the null model to assist in explaining the geographic and temporal variations in valley fever risk. TABLE 1 shows the regression coefficients and their exponentially transformed values. The results indicate that annual precipitation, precipitation from the previous year, and percentage of cultivated landcover display significant correlations with valley fever risk. Annual and lagged precipitation were negatively correlated with valley fever risk and percentage of cultivated landcover was positively associated. Landcover had the strongest significant association while lagged precipitation was comparatively weak.

TABLE 1. The estimated beta coefficient from the regression model and exponential transformations for relative risk in California. Statistically significant variables are in bold.

Covariate	Posterior estimate	
	Beta (95% CI)	Relative Risk (95% CI)
Annual Precipitation	**-0.22 (-0.34, -0.11)**	**0.80 (0.71, 0.89)**
Annual Mean Temperature	-0.05 (-0.27, 0.17)	0.94 (0.76, 1.18)
Lagged Precipitation (1-year lag)	**-0.11 (-0.22, -0.003)**	**0.89 (0.79, 0.99)**
Percent Land Cover Cultivated	**0.38 (0.19, 0.47)**	**1.45 (1.21, 1.60)**
Mean Temperature of Driest Quarter	-0.007 (-0.08, 0.06)	0.99 (0.92, 1.06)

4. Discussion

This study applied a hierarchical Bayesian spatio-temporal approach to assess county-level valley fever risk in California based on annual data from 2001-2018. We mapped the relative risk, classified counties based on model results, and presented temporal trends for the counties at elevated risk. The relative risks presented in this study indicate that counties in central California, including central-coastal counties, are at increased risk for valley fever. While Kern County has long been recognized as a valley fever hot spot, these results show that valley fever risks are present and increasing in counties to the north and west of Kern County. Spatio-temporal results show that disease risk has been increasing faster than average in counties in the Central Valley centered around San Joaquin County for the past nine years. Over the past two years, coastal counties from Monterey in the north to Orange County in the south have seen increased risk. Some of these areas are classified as medium or low risk in this study, but with increasing trends, they deserve additional attention.

The overall modeled temporal risk trend generally aligns with case count observations. Two depressions apparent in the trend line coincide with drought periods in California – 2007-2009 and 2012-2016 (California Water Science Center, 2019). Both of these low points were followed by significant increases in valley fever risk, seen in 2011 and 2017. In the early 1990s, a similar pattern occurred in Kern County, where a five-year drought was followed by above average seasonal rainfall that coincided with spike in valley fever from 1991 to 1995 (Jinadu, 1995). Lauer et al. (2014) hypothesized that droughts may decimate the less heat-tolerant competitors of *Coccidioides* spp., and then an above average wet season could help *Coccidioides*' spores bloom in a less competitive environment. This echoes the "grow and blow" hypothesis proposed by Comrie and Glueck (2007) studying valley fever in Arizona. The patterns observed here support this hypothesis, warranting further study.

The counties classified as high risk are centered in the southern Central Valley of California and along the central coast. The Central Valley is characterized by average annual precipitation between 5-15 inches per year and the central coast with up to 25 inches per year (Western Regional Climate Center, 2019). Average temperatures in the coastal counties are slightly cooler than in the Central Valley and the seasonal climatic pattern consists of a dry, hot summer and a wet winter/early spring in both areas (WRCC, 2019). Climates of both regions align with a review done by Fisher et al. (2007) that assessed habitat attributes suitable for *Coccidioides* spp. growth based on locations where the soil had tested positive for the pathogen. The temporal risk trends in these counties indicate that valley fever may be establishing itself along the coastline, which the CDC considers to be suspected endemic rather than established endemic, and further north in the Central Valley than is currently thought to be established endemic (Centers for Disease Control and Prevention, 2019). Additional surveillance and soil sampling efforts should be focused on these counties to determine if the established endemic range of this disease is expanding.

This research also assessed five environmental covariates, finding three to be statistically significant: two precipitation variables and one land cover variable. Both annual precipitation and precipitation from the previous year had inverse relationships with relative risk. This relationship was expected for annual (current year) precipitation as *Coccidioides* spp. are not found in saturated soils (F. S. Fisher et al., 2007), and precipitation likely prevents the spores from become airborne; therefore, risk is expected to be decreased in years with heavy or consistent precipitation. The negative relationship between relative risk and lagged precipitation runs counter to previous research from Arizona and California (Comrie, 2005; Coopersmith et al., 2017; Weaver & Kolivras, 2018), though the only other study to assess this variable at an annual scale also found a positive relationship between 1-year lagged precipitation and valley fever incidence specifically in the southern Central Valley (Gorris et al., 2017). Because rainfall in California is received seasonally, the ecological relationship between precipitation and valley fever risk may be obscured at annual scales or it may be that heavy precipitation hampers *Coccidioides* spp. growth and dispersal for periods longer than a year. Although these variables were found to be significant, additional factors are needed to better account for the spatio-temporal patterns observed.

The Bayesian spatio-temporal approach used this research allowed us to observe geographic patterns in the temporal trends that might not otherwise be noticed while avoiding the common problems associated with classical risk estimators such as standardized morbidity ratios. Crude standardized morbidity ratios are imprecise for rare diseases and/or in areas with small populations and can have extreme values with low precision (Thompson, Carozza, & Zhu, 2007). Bayesian smoothing assumes that the relative risks come from a common distribution and borrows from estimates in other areas of the map that are likely to be similar. The resulting estimates of ‘true’ relative risk are a combination of the average of the observed risk and parameters reflecting the state-wide distribution, weighted on the population at risk (Lawson, 2018). Bayesian spatio-temporal models are also commonly used for forecasting (Dietze, 2017). Future research can improve upon this model to forecast how valley fever patterns will change as climate and other environmental variables change.

There are a few limitations to this study that should be addressed. First is a possible discrepancy between exposure location and where the disease was reported. Because valley fever is contracted through inhalation of spores, the exposure site can occur in counties other than where the reports are made, typically the home address of the case. Ideally, the study would use cases where exposure sites were known or where soil testing has had positive results, which is not possible with currently available data. Such analyses may be possible in the future as technology to detect *Coccidioides* spp. in the soil improves. Next, the land use variable, percent cultivated land cover, did not have a temporal component, impacting the ability to model disease variability due to changes in landcover. Next, this research focused on assessing environmental risk factors, but the inclusion of additional variables, such as those that capture anthropogenic soil disturbance, may help explain more of the spatial and temporal patterns observed. Finally, the model might be improved in future studies with consideration of more explicit prior distributions, finer temporal scales to capture seasonality, and a weighted spatial adjacency matrix that takes into account factors such as length of county borders.

5. Conclusion

This study makes broad contributions to the discipline of geography through spatio-temporal disease mapping, and more specifically to valley fever research through the application of spatio-temporal modeling of relative risk to valley fever.

Through the application of Bayesian spatio-temporal modeling, we have a better understanding of counties at risk for valley fever risk in California. The resulting risk maps and temporal graphs produced can be used by state and county public health officials in decision making and prioritization for public health awareness and educational campaigns. This research also investigated potential risk factors present at an annual scale, finding three significant environmental relationships. Quantifying the relationships between environmental variables and the distribution of valley fever is needed to not only better understand ecology of the disease, but to discern how future environmental changes may impact disease distribution.

References

Baptista-Rosas, R. C., Hinojosa, A., & Riquelme, M. (2007). Ecological Niche Modeling of Coccidioides spp. in Western North American Deserts. Annals of the New York Academy of Sciences, 1111, 35–46. https://doi.org/10.1196/annals.1406.003

Barker, B. M. (2017). The Changing Epidemiology and Diagnosis of Valley Fever. Clinical Microbiology Newsletter, 39(20), 159–164. https://doi.org/10.1016/j.clinmicnews.2017.09.007

Besag, J., York, J., & Mollie, A. (1991). Bayesian Image Restoration, with Two Applications in Spatial Statistics. Ann Inst Statist Math, 43(1), 1–59.

Bivand, R. S., Gomez-Rubio, V., & Rue, H. (2015). Spatial Data Analysis with R-INLA with Some Extensions. Journal of Statistical Software, 63.

Blangiardo, M., & Cameletti, M. (2015). Spatial and Spatio-temporal Bayesian Models with R-INLA. (Wiley, Ed.). West Sussex, UK.

Blangiardo, M., Cameletti, M., Baio, G., & Rue, H. (2013). Spatial and Spatio-temporal Epidemiology Spatial and spatio-temporal models with R-INLA. Spatial and Spatio-Temporal Epidemiology, 7, 39–55. https://doi.org/10.1016/j.sste.2013.07.003

California Department of Public Health. (2018). Epidemiologic Summary of Coccidioidomycosis in California, 2017. Retrieved from https://www.cdph.ca.gov/Programs/CID/DCDC/CDPH Document Library/CocciEpiSummary2017.pdf

California Department of Public Health. (2019). Coccidioidomycosis (Valley Fever). Retrieved March 3, 2019, from https://www.cdph.ca.gov/Programs/CID/DCDC/Pages/Coccidioidomycosis.aspx

California Water Science Center. (2019). 2012-2016 California Drought: Historical Perspective. Retrieved March 19, 1BC, from https://ca.water.usgs.gov/california-drought/california-drought-comparisons.html

Centers for Disease Control and Prevention. (2019). Valley Fever Maps. Retrieved March 19, 1BC, from https://www.cdc.gov/fungal/diseases/coccidioidomycosis/maps.html

Comrie, A. C. (2005). Climate Factors Influencing Coccidioidomycosis Seasonality and Outbreaks. Environmental Health Perspectives, 113(6), 688–692. https://doi.org/10.1289/ehp.7786

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

Coopersmith, E. J., Bell, J. E., Benedict, K., Shriber, J., McCotter, O., & Cosh, M. H. (2017). Relating coccidioidomycosis (valley fever) incidence to soil moisture conditions. GeoHealth, 1(1), 51–63. https://doi.org/10.1002/2016GH000033

Czado, C., Gneiting, T., & Held, L. (2009). Predictive Model Assessment for Count Data. Biometrics, 65(4), 1254–1261. https://doi.org/10.1111/j.l541-0420.2009.01191.x

Dietze, M. (2017). Ecological Forecasting. Princeton, NJ: Princeton University Press.

EPA. (2018). Level III and IV Ecoregions of the Continental United States. Retrieved March 3, 2019, from https://www.epa.gov/eco-research/level-iii-and-iv-ecoregions-continental-united-states

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

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

Grayzel, S. E., Martínez-López, B., & Sykes, J. E. (2017). Risk Factors and Spatial Distribution of Canine Coccidioidomycosis in California, 2005–2013. Transboundary and Emerging Diseases, 64(4), 1110–1119. https://doi.org/10.1111/tbed.12475

Jinadu, B. A. (1995). Valley Fever Task Force Report on the Control of Coccidioides immitis. Bakersfield, CA.

Khana, D., Rossen, L. M., Hedegaard, H., & Warner, M. (2018). A BAYESIAN SPATIAL AND TEMPORAL MODELING APPROACH TO MAPPING GEOGRAPHIC VARIATION IN MORTALITY RATES FOR SUBNATIONAL AREAS WITH R-INLA. Journal of Data Science, 18, 147–182.

Kim, S., & Kim, Y. (2018). Hierarchical Bayesian modeling of spatio-temporal patterns of scrub typhus incidence for 2009 – 2013 in South Korea. Applied Geography, 100(August 2016), 1–11. https://doi.org/10.1016/j.apgeog.2018.08.008

Knorr-Held, L. (2000). Bayesian modelling of inseparable space-time variation in disease risk. Statistics in Medicine, 19, 2555–2567.

Kolivras, K. N., & Comrie, A. C. (2004). Climate and infectious disease in the southwestern United States. Progress in Physical Geography, 3, 387–398. Retrieved from http://ppg.sagepub.com/cgi/content/abstract/28/3/387%5Cnpapers3://publication/uuid/B3F1F77C-A82F-4E39-8162-CD5B5FCA5E60

Lawson, A. B. (2018). Bayesian Disease Mapping - Hierarchical Modeling in Spatial Epidemiology (Third Edit). Boca Raton, FL: Taylor & Francis.

Lowe, R., Barcellos, C., Coelho, C. A. S., Bailey, T. C., Coelho, G. E., Graham, R., … Ramalho, W. M. (2014). Dengue outlook for the World Cup in Brazil: an early warning model framework driven by real-time seasonal climate forecasts. Lancet, 14, 619–626. https://doi.org/10.1016/S1473-3099(14)70781-9

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

Martino, S., & Rue, H. (2010). Case Studies in Bayesian Computation using INLA. In P. Mantovan & P. Secchi (Eds.), Complex Data Modeling and Computationally Intensive Statistical Methods. Milano: Springer.

Martınez-Bello, D., Lopez-Quilez, A., & Prieto, A. T. (2018). Spatiotemporal modeling of relative risk of dengue disease in Colombia. Stoc, 32, 1587–1601. https://doi.org/10.1007/s00477-017-1461-5

Park, B. J., Sigel, K., Vaz, V., Komatsu, K., McRill, C., Phelan, M., … Hajjeh, R. a. (2005). An epidemic of coccidioidomycosis in Arizona associated with climatic changes, 1998-2001. The Journal of Infectious Diseases, 191(11), 1981–7. https://doi.org/10.1086/430092

Petersen, L. R., Marshall, S. L., Barton-Dickson, C., Hajjeh, R. A., Lindsley, M. D., Warnock, D. W., … Morgan, J. (2004). Coccidioidomycosis among Workers at an Archeological Site, Northeastern Utah. Emerging Infectious Diseases, 10(4), 637–642. https://doi.org/10.3201/eid1004.030446

R Core Team (2003). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org/

Raghavan, R. K., Goodin, D. G., Neises, D., Anderson, G. A., & Ganta, R. R. (2016). Hierarchical Bayesian Spatio-Temporal Analysis of Climatic and Socio-Economic Determinants of Rocky Mountain Spotted Fever. PLoS ONE, 11(3). https://doi.org/10.1371/journal.pone.0150180

Richardson, S., Thomson, A., Best, N., & Elliott, P. (2004). Interpreting Posterior Relative Risk Estimates in Disease-Mapping Studies. Environmental Health Perspectives, 112(9), 1016–1025. https://doi.org/10.1289/ehp.6740

Rosenstein, N. E., Emery, K. W., Werner, S. Ben, Kao, A., Johnson, R., Rogers, D., … Hajjeh, R. A. (2001). Risk Factors for Severe Pulmonary and Disseminated

Coccidioidomycosis : Kern County, California, 1995 – 1996. Clinical Infectious Diseases, 32(5), 708–714.

Rue, H., & Martino, S. (2009). INLA: Functions which allow to perform a full Bayesian analysis of structured additive models using Integrated Nested Laplace Approximation. R Package.

Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. Journal of the Royal Statistical Society. Series B: Statistical Methodology, 71(2), 319–392. https://doi.org/10.1111/j.1467-9868.2008.00700.x

Schrödle, B., & Held, L. (2011). A primer on disease mapping and ecological regression using INLA, 241–258. https://doi.org/10.1007/s00180-010-0208-2

Stacy, P. K. R., Comrie, A. C., & Yool, S. R. (2012). Modeling Valley Fever Incidence in Arizona Using a Satellite-Derived Soil Moisture Proxy. GIScience & Remote Sensing, 49(2), 299–316. https://doi.org/10.2747/1548-1603.49.2.299

Talamantes, J., Behseta, S., & Zender, C. S. (2007). Statistical modeling of valley fever data in Kern County, California. International Journal of Biometeorology, 51(4), 307–313. https://doi.org/10.1007/s00484-006-0065-4

Tamerius, J. D., & Comrie, A. C. (2011). Coccidioidomycosis incidence in Arizona predicted by seasonal precipitation. PLoS ONE, 6(6). https://doi.org/10.1371/journal.pone.0021009

Thompson, J. A., Carozza, S. E., & Zhu, L. (2007). An evaluation of spatial and multivariate covariance among childhood cancer histotypes in Texas (United

States). Cancer Causes Control, 18, 105–113. https://doi.org/10.1007/s10552-006-0085-8

Tian, L., Liang, F., Xu, M., Jia, L., Pan, X., & Clements, A. C. A. (2018). Spatio-temporal analysis of the relationship between meteorological factors and hand-foot-mouth disease in Beijing, China. BMC Infectious Diseases, 18(158).

U.S. Census Bureau. (2000). Profile of General Demographic Characteristics: 2000 Census, 2000 Summary File 1 (SF 1) 100% Data. Retrieved June 5, 2017, from https://factfinder.census.gov/

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/

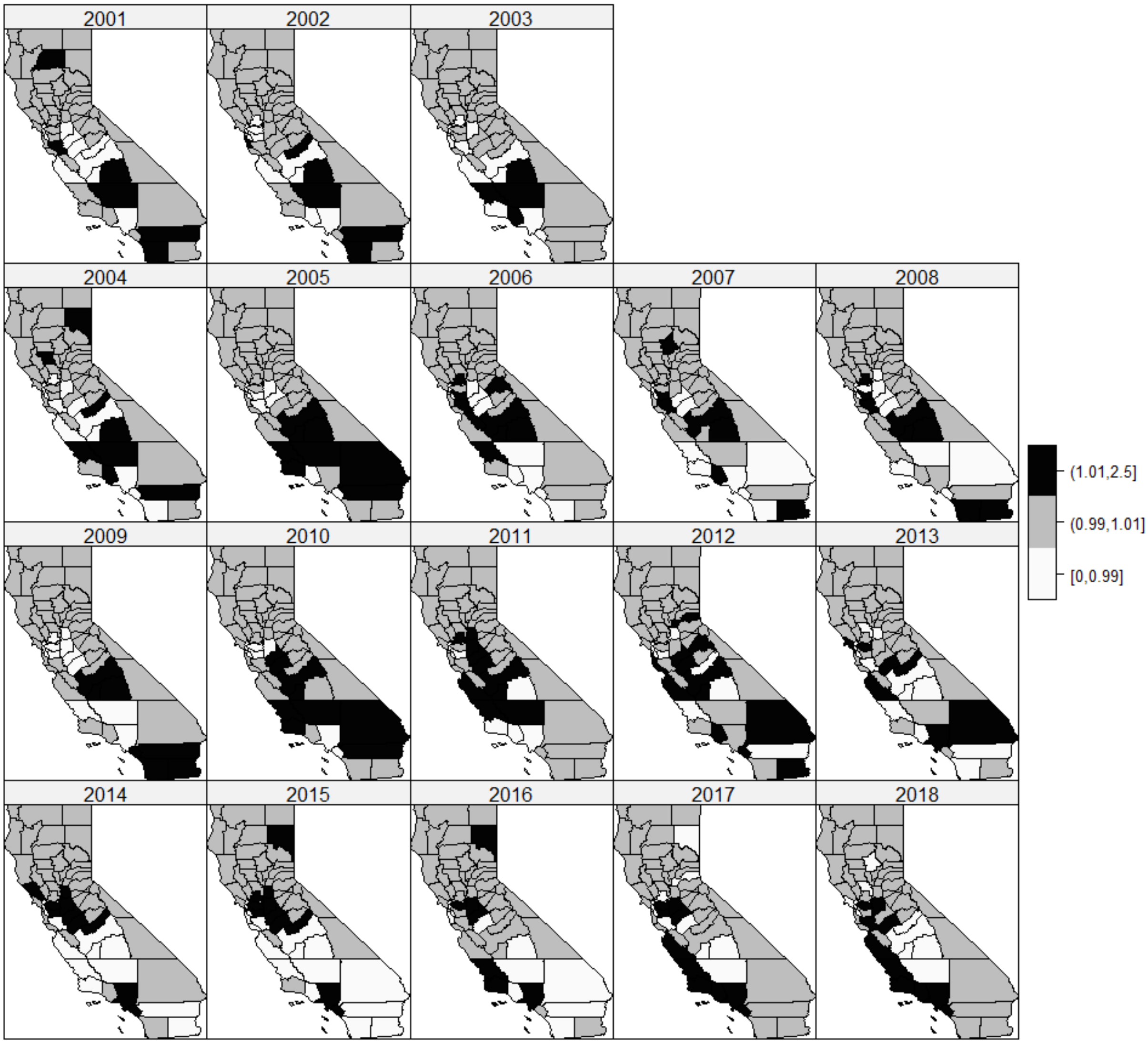
Watanabe, S. (2010). Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory. Journal of Machine Learning Research, 11, 3571–3594.

Weaver, E. A., & Kolivras, K. N. (2018). Investigating the Relationship Between Climate and Valley Fever (Coccidioidomycosis). EcoHealth, 15(4), 840–852. https://doi.org/10.1007/s10393-018-1375-9

Zender, C. S., & Talamantes, J. (2006). Climate controls on valley fever incidence in Kern County, California. International Journal of Biometeorology, 50(3), 174–182. https://doi.org/10.1007/s00484-005-0007-6

SUPPLEMENTAL MATERIAL

Spatiotemporal Maps

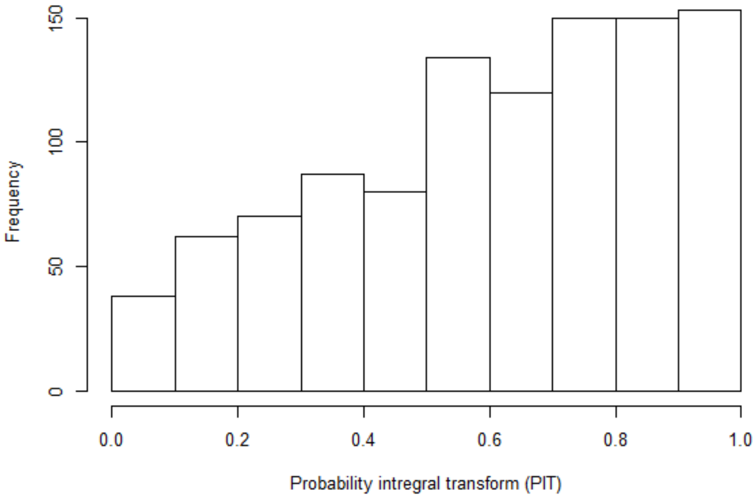


Hyperparameters (null model)

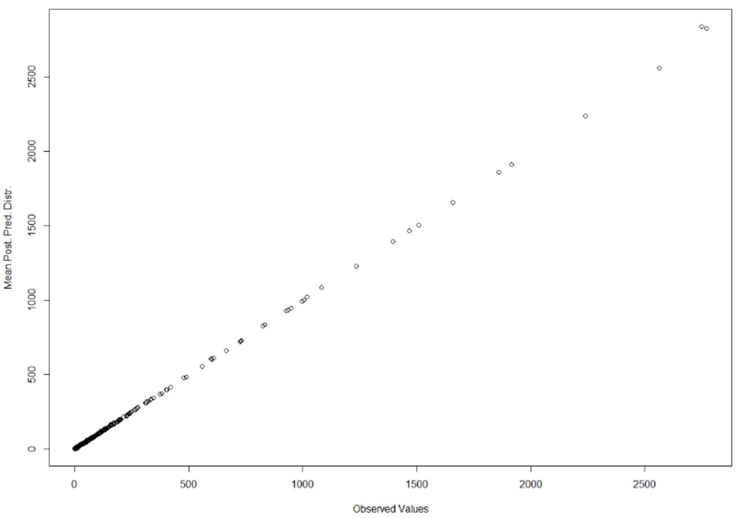
	Mean	**SD**	**0.025**	**0.5**	**0.975**	**Mode**
τ_v	0.61	0.079	0.433	0.621	0.725	0.685
τ_u	15.75	2.046	11.138	16.041	18.741	17.707
τ_φ	3.79	1.342	1.728	3.606	6.943	3.246
τ_δ	7.91	0.638	6.703	7.902	9.202	7.897

The random effect variance (note: precision shown here) quantifies the variability across groups, such as year or county.

Model Checking



$PIT = p(y_i^* \leq y_i | y_f)$ * represents a future occurrence; y_f is the group of samples used to fit the model. Interpretation - the PIT histogram is adequate considering plot is not expected to be perfectly uniform for count data. See Czado, Gneiting, & Held (2009).



The posterior predictive distribution represents the likelihood of a replicate observation y_i^* having observed the data y: $p(y_i^*|y) = \int p(y_i^*|\theta_i)p(\theta_i|y)\mathrm{d}\theta_i$
Interpretation – the prediction is very close to the observed values as is desired.

Model Code

```
# Null Model
CA.formula.mod.1<- y ~ 1 +
  f(CA.county.ID, model="bym", graph=CA.adj) +
  f(CA.year.ID,model="iid") +
  f(CA.county.year.ID,model="iid")

CA.mod.1 <- inla(CA.formula.mod.1,family="poisson",data=CA.data,offset=log(E),
              control.predictor=list(compute=TRUE),
              control.inla=list(strategy="laplace",npoints=30),
              control.compute=list(dic=TRUE,cpo=TRUE,waic=TRUE))

# Full Model
CA.formula.mod.2<- y ~ 1 +
  f(CA.county.ID, model="bym", graph=CA.adj) +
  f(CA.year.ID, model="iid") +
  f(CA.county.year.ID,model="iid") +
  x1 + x2 + x3 + x4 + x5

CA.mod.2 <- inla(CA.formula.mod.2,family="poisson",data=CA.data,offset=log(E),
              control.fixed=list(mean=0,prec=0.00001,
                             mean.intercept=0,
                             prec.intercept=0.00001),
              control.predictor=list(compute=TRUE),
              control.inla=list(strategy="laplace",npoints=30),
              control.compute=list(dic=TRUE,cpo=TRUE,waic=TRUE))
```