Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Ian David Buller Date

On estimating the spatial distribution of *Yersinia pestis* in the United States using a wide-ranging sentinel species and spatial statistics with sampling considerations

By

Ian Buller
Doctor of Philosophy

Environmental Health Sciences

Lance Waller, Ph.D.
Advisor

Yang Liu, Ph.D.
Committee Member

Stefanie Sarnat, Sc.D.
Committee Member

Howard Chang, Ph.D.
Committee Member

A. Townsend Peterson, Ph.D.
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of James T. Laney School of Graduate Studies

Date

On estimating the spatial distribution of *Yersinia pestis* in the United States using a wide-ranging sentinel species and spatial statistics with sampling considerations

By

Ian David Buller
M.A., University of Colorado, 2013

Advisor: Lance Waller, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Environmental Health Sciences
2019

Abstract

On estimating the spatial distribution of *Yersinia pestis* in the United States
using a wide-ranging sentinel species and spatial statistics with sampling considerations
By Ian David Buller

Plague is a highly consequential disease caused by the bacterium *Yersinia pestis* that infects multiple mammal host species, including humans providing a concern for public health and conservation. However, the precise locations where the disease is transmitted within the United States remain unknown and uncertain. While human cases are rare in the United States, plague is maintained within small mammal populations, namely wild rodents, and their flea vectors; therefore, plague surveillance systems typically focus on monitoring animal species, but are often faced with administrative, logistical, and biological challenges. In collaboration with plague surveillance agencies, I aim to: 1) identify sources of data uncertainties and bias in plague surveillance systems, 2) predict the spatial distribution of enzootic plague in the western United States, and 3) evaluate the association between enzootic plague and historical human plague cases in the United States. I collate an extensive set of human, animal, and environmental information and discuss the quality of these data. I develop an ecological niche modeling method that uses climatological variables and coyote (*Canis latrans*) specimens tested for exposure to *Y. pestis* to predict the spatial distribution of enzootic plague in California and the western United States, even in areas that have not historically been monitored for plague activity. I identify areas of the United States that are sensitive to some types of data uncertainty and bias, including positional uncertainty in the sampling location of coyote specimens and sampling effort bias of agencies that monitor plague activity. Finally, I use a spatial statistical framework, integrated nested Laplace approximation, to estimate historical human plague risk across the western United States and demonstrate that enzootic plague is positively associated with human plague risk at the county level (relative risk: 1.20; 95% credibility interval: 1.16 – 1.24). I work closely with plague surveillance agencies, so my results will have immediate impact on plague surveillance such as prioritizing future laboratory testing. I close by proposing future directions, including tangible advancements to the developed method to test biogeographical hypotheses and applications to other disease systems and fields of science.

On estimating the spatial distribution of *Yersinia pestis* in the United States using a wide-ranging sentinel species and spatial statistics with sampling considerations

By

Ian David Buller
M.A., University of Colorado, 2013

Advisor: Lance Waller, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Environmental Health Sciences
2019

Acknowledgements

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1444932 and a National Science Foundation Graduate Research Internship (NSF 18-069; formerly NSF 16-015) under the same grant number. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. The Environmental Health Sciences (EHS) program at Emory University, in particular Dr. Jeremy Sarnat, Dr. Gary Miller, and Dr. Paige Tolbert, has provided exceptional safeguards to complete my academic training. Thank you to my friends and fellow doctoral students within the EHS program for the emotional and social support throughout my time at Emory University. I am indebted to the constant motivation from the Rollins School of Public Health Dean's Office including Adrienne Schwartz, Dean Surbey, and Claudia Paez-Ellett. Each member of my Dissertation Committee has greatly broadened my scientific perspective of combining ecology, statistics, and public health. I would like to especially thank my Doctoral Advisor and fellow green chile connoisseur, Dr. Lance Waller, for imparting his optimistic vision and philosophy of science. His unwavering trust enabled me to assemble a remarkable crew of experts to navigate my own dissertation route.

This work would not have been possible without my scientific collaborators and colleagues. At the California Department of Public Health (CDPH), I would like to thank Dr. Mark Novak, Jim Tucker, Dr. Greg Hacker, and Dr. Vicki Kramer. At the U.S. Department of Agriculture (USDA), I would like to thank Dr. Sarah Bevins and Dr. Jeff Chandler. At the U.S. Centers for Disease Control and Prevention (CDC), I would like to thank Dr. Ken Gage, CAPT Rusty Enscore, Dr. Becky Eisen. Thank you to LCDR Danielle Buttke (National Park Service) and CDR Jeffrey McCollum (Indian Health Service) who elucidated the nuances of the plague surveillance system in the United States. Special thanks to CDR

Joanna Gaines (CDC) and Tom Chapel (CDC) for their guidance through the small but labyrinthine network of public health professionals in order to make initial contact with collaborators. The initial conception of this work sprung up from conversations with Dr. Katie Richgels (U.S. Geological Survey) shortly after our time working with Dr. Pieter Johnson (University of Colorado). Support also came from academic colleagues who were subject matter experts for plague ecology (Dr. Dan Salkeld, Colorado State University, and Dr. Sean Maher, Missouri State University) or ecological niche modeling (the Biodiversity and Macroecology Laboratory at the University of Kansas Biodiversity Institute). In addition, I am grateful to the unsung USDA and state health department field biologists for collecting samples, USDA/CDPH/CDC laboratory specialists for testing samples, citizen scientists and independent investigators who recorded coyote observations in museum repositories, as well as data managers who digitized and provided the secondary data for this work.

Finally, I would like to dedicate this work to my parents, Dr. David and Mary Buller, for inspiring a curious love of science and instilling the moral compass to discover where to apply it and to my sister, Katherine Buller, for setting the pace by embodying the family motto *aquila non captat muscas*.

Contents

1 Introduction

"Surgeon General W. X. Van Reypen of the navy is an authority on the bubonic plague. In discussing the probability of its reaching San Francisco he says: 'The climatic conditions of the United States preclude the possibility of the plague ever getting within this country...'"

- *San Francisco Call* on March 7, 1900 (*1*)

"Bubonic plague has appeared at widely separated spots throughout the world, and, while it has not attained the proportions of an epidemic anywhere except certain places in the Orient, it is entirely possible for it to become as endemic as the influenza did during the past two years."

- *Sacramento Union* on June 20, 1920 (*2*)

Emerging and re-emerging infectious diseases (EIDs) increasingly burden global public health, threaten species conservation, and negatively impact local and global economies (*3*, *4*). The majority of EIDs originate from animals (*5*) and these diseases are called zoonoses. Zoonoses are of particular concern because humans are often accidental hosts that are naive to infection and experience extreme pathology. Notable EID outbreaks were observed in humans including the 1993 Sin Nombre virus (Hantavirus) outbreak (*6*) and the 2003 severe acute respiratory syndrome coronavirus outbreak (*7*), but also in animal species such as, for example, bird die-offs due to West Nile virus (*8*) and loss of entire prairie dog (*Cynomys* spp.) colonies due to bubonic plague (*9*). Zoonoses are maintained in non-human reservoir species, which presents a challenge for control and eradication programs, especially if a zoonosis has multiple reservoir species or is transmitted by an insect vector (*10*). Public health and conservation efforts focus on disease burden reduction or local eradication based on an understanding of the ecological conditions favorable for transmission.

Zoonoses are originally infectious diseases of animals, including wildlife, and are generally environmentally mediated. Climate and landscape combine to create a heterogeneous scattering of ecosystems within which hosts, vectors, and their pathogens can thrive. Optimal environmental combinations promote survivorship and ecological fitness of a species (a.k.a., a "fundamental abiotic ecological niche") whereby a species can survive without immigration of individuals from another population (*11*). Geographic areas characterized by an optimal combination of environmental conditions for a pathogen define an ecological

niche where a pathogen can maintain itself inside its hosts (*12*). Enzootic areas occur if transmission is ongoing in regions of the ecological niche (*13*). An epizootic (i.e., disease outbreak within an animal population) event can arise in enzootic areas when host demographic changes amplify pathogen transmission (*14*, *15*), a pathogen experiences changes (e.g., increased transmissibility or virulence) due to a genetic change (*16*), or a pathogen spreads to a highly susceptible host species (*17*). Much scientific attention has focused on the conditions leading up to an epizootic event, but understanding the locations of enzootic transmission is a primary step in forecasting epizootic events. Statistical tools commonly used in ecology are available to study the ecological niche of diseases (*12*, *13*, *18–20*) and they rely on rich disease surveillance data.

However, surveillance for zoonoses often is challenging, especially in large-scale administrative settings where such systems are necessary for setting and monitoring public health policies but constrained by annual operating budgets. Zoonoses are particularly challenging to monitor because of the limited opportunity to observe the disease before transmission to a human host occurs (*21*). For example, animal hosts can be cryptic, highly susceptible to infection, or difficult to collect. Large-scale administrative surveillance systems provide an opportunity to explore the development of targeted statistical analysis and data science tools in order to take full advantage of their surveillance data, and meet mandated responsibilities, budgetary constraints, and regulations governing data sharing and reporting (*22–26*). A promising combination for enhanced disease surveillance involves traditional disease surveillance systems (i.e., clinical and laboratory data) and big data resources (i.e., digital data products) (*27*). In the Big Data Era, mapping diseases has becoming an increasingly powerful and available tool for infectious disease surveillance (*28*). Maps of infectious diseases can inform disease control programs by identifying areas for additional surveillance, anticipating disease control resource needs, and strategically deploying prevention in at-risk locations. For example, control efforts for diseases transmitted human to human by mosquitoes, such as malaria, have had great success monitoring areas with disease (*29*, *30*) because an established sophisticated surveillance network provides accurate identification of areas of high human incidence and prevalence. Maps can also predict high-risk areas for spillover events of zoonoses where environmental conditions can cause an epizootic event in

an enzootic area.

In this dissertation, I take a particular look at *Yersinia pestis*, the causative agent of the zoonotic disease plague. By applying an ecological niche modeling framework, I make informed predictions of the spatial distribution of enzootic plague in the western United States (Arizona, California, Colorado, Idaho, Kansas, Montana, Nebraska, Nevada, New Mexico, North Dakota, Oklahoma, Oregon, Texas, Utah, South Dakota, Washington, and Wyoming). I specifically collaborate with a state health department (California Department of Public Health) and a national agency (U.S. Department of Agriculture) for their expertise and source of plague surveillance data. I implement spatial statistical methods to improve plague surveillance by making use of multiple types of data at multiple spatial scales. I begin this chapter with a brief background on plague and then introduce the specific research aims of the dissertation.

1.1 *Yersinia pestis*: Plague

Plague is caused by the gram-negative bacterium *Yersinia pestis*. *Y. pestis* is a generalist pathogen, found in over 200 species of mammals. Infection has been reported in rodents, lagomorphs, ungulates, carnivores, insectivores, hydraxes, primates, and marsupials with varying levels of susceptibility (*31*). Humans are highly susceptible to infection when exposed to plague bacteria. Plague manifests in humans in three forms: 1) Bubonic plague (about 85% of cases) where patients develop a painful swelling of lymph nodes; 2) Septicemic plague (about 10% of cases) where infection spreads through the bloodstream to other organs; and 3) Pneumonic plague (about 5% of cases) where infection starts in the lungs from aerosol exposure to infected droplets or where untreated bubonic or septicemic forms spread to the lungs (*32*). Mortality is high (40%-70%) for bubonic plague and usually fatal for pneumonic and septicemic cases without prompt antibiotic treatment, primarily streptomycin (16% mortality with treatment) (*33*).

Transmission of plague is complex, involving various exposure pathways and routes. The primary transmission route for plague is believed to be arthropod vector-borne, through fleas (*31*). Fleas live on the outside of mammals and birds (ectoparasite) and survive by feeding on the blood of their hosts (hematophagous). Contemporary human cases are

likely contracted via a bite from an infected flea that has previously fed on a rodent (*34*–*36*). Humans can encounter infected fleas from interactions with rodents and domesticated animals that have had contact with rodents and their fleas. Rabbits, cats, and dogs can harbor infected rodent fleas and can bring these vectors into close contact with humans (*37*, *38*). If a flea bloodmeal from an infectious host contains plague bacilli, the bacilli multiply and form a synthesized biofilm that can block the proventriculus of the flea (*39*). The proventriculus is the first organ after the esophagus where food is ground into fine particles and acts as a valve regulating the flow of food into the midgut where digestion and absorption occur. Blocking the proventriculus starves the flea host, which seeks more bloodmeals during which plague bacilli can be regurgitated into the animal host transmitting plague (*39*). Plague can be transmitted by unblocked flea species (*40*), but blockage can increase a flea species' transmissibility of plague (*41*).

In North America, *Oropsylla montana* and *Xenopsylla cheopis* are primary flea vectors of plague from rodents to humans (a.k.a., "bridging vector"). Both species can become infectious immediately after feeding on an infected host, experience lower mortality from infection than other species, and remain infectious for many weeks (*40*, *41*). *Oropsylla montana* is almost exclusively found on ground squirrel species in the genus *Spermophilus* while *X. cheopis* is found on rat species of the genus *Rattus* and predominantly found in sub-tropical/tropical climates. These and other flea species are responsible for transmission of *Y. pestis* within animal populations (*42*). Over 200 flea species are known to carry *Y. pestis*, including *Xenopsylla brasiliensis*, *X. astia*, *Nosopsyllus fasciatus*, and poorly understood "wild rodent fleas" (*43*–*45*). Over 85% of these fleas specifically feed on rodents and about 23% are found in North America (*46*).

In addition to vector-borne transmission, human infection can occur directly from another infected host through inhalation, ingestion, or soft-tissue absorption of *Y. pestis* (*31*, *47*, *48*). Infectious respiratory droplets can be passed between humans during the pneumonic form of plague (*43*); however, this mode of transmission is rare in the United States with the most recent confirmed case of pneumonic human-to-human plague occurring in 1924 (one potential case was observed in Colorado in 2014 (*48*)). Domesticated animals, in particular *Felis catus*, succumb to plague infection, express respiratory symptoms, and can

transmit plague pneumonically to humans (*37*, *48*, *49*). Consuming infected animal meat from guinea pigs and camels has also lead to human cases worldwide (*50–54*). Finally, direct contact with animals, including dead carcasses handled during hunting and butchering, can transmit plague to humans (*55*, *56*).

Human plague cases have often been linked to a nearby recent or on-going epizootic plague event (*57*). Climate and landscape define the environmental conditions suitable for both epizootic and enzootic *Y. pestis* transmission but linking these as factors contributing to human plague risk is complex (*58–74*). In general, human cases have likely occurred in a major enzootic area (i.e., enzootic foci) of plague when environmental conditions were favorable for an epizootic event; however, the full extent of enzootic foci of plague is not fully known across the entire western United States (see Map 3.1 in *75*). In addition, the mechanism by which *Y. pestis* persists in enzootic foci is not completely known. The bacterium is likely maintained in a sylvatic cycle of rodent-to-rodent transmission via fleabites between rodent populations (see reviews by *76*, *77*). Soils contaminated with *Y. pestis* may be another mechanism by which plague can persist in the environment between epizootic periods. Eisen and colleagues (*78*) found *Y. pestis* can survive 24 days in soil, and, although rare (1 out of 103 individuals), a susceptible animal host can become infected from exposure to contaminated soil (*79*). An amoeba (*Dictyostelium discoideum*) has demonstrated ability to be an environmental reservoir for *Y. pestis* (*80*) and may play a role in the persistence of plague in underground burrowed habitats of communal rodent species and their fleas.

Human plague cases in the United States since plague became enzootic (∼1950s) were likely from contact with an infected mammal or its fleas. Since 1950, human plague cases have primarily occurred in the southwestern United States (i.e., Arizona, Colorado, New Mexico, and Utah). An effort led by the U.S. Centers for Disease Control and Prevention (CDC) estimated areas of high human plague risk in the southwestern United States (*71–73*). Between 1957 and 2004, over 92% of human cases with a discernible site of exposure and route of transmission (n = 165 of 180 cases from a total of 346 cases) were from direct exposure to a wild animal or its fleas (*72*). Data used in Eisen and colleagues' (*72*) report account for over 66% of human plague cases in the western United States used in this dissertation (1950–2017; Unpublished data courtesy of Ken Gage at the CDC) and

similar characteristics were found in a review of by Kugeler and colleagues (*33*). While risk of infection is small, plague is a high-consequence disease, for which risk mapping can be beneficial to determine areas of enhanced surveillance and targeted prevention. Human cases of plague have occurred in 13 western United States, but case numbers are small, which presents data limitations that challenge the prediction of high-risk areas.

Non-human plague information is also limited which challenges estimates of where *Y. pestis* is enzootic in the United States. Temperature can affect transmission efficiency and survivorship of fleas (*63*, *81*, *82*) and seasonal climate explains flea population patterns (*83*, *84*). Therefore, the spatial distribution and ecology of plague vector species are likely important determinants of plague (*85*), but data are spatially limited and spatially biased by sampling effort (*86*). Because *Y. pestis* is a generalist pathogen, identifying a single mammal species to monitor for human plague transmission has proved challenging, unlike monitoring deer mice for Hantavirus or crow die-off events for West Nile Virus (*8*, *47*, *87*, *88*). Holt and colleagues (*89*) restricted their prediction using rodent plague data to California. Maher and colleagues (*85*) used multiple animal species across the United States, but their predictions were limited by the unavailability of absence information (i.e., where plague was tested but not observed). Walsh and Haseeb (*90*) used a small sample ($n = 66$) of plague-positive deer mice (*Peromyscus maniculatus*), which are non-essential hosts for the plague lifecycle and likely only indicate epizootic conditions (*91–94*). Domesticated pets and coyotes (*Canis latrans*) are used to monitor epizootic plague activity in New Mexico and Arizona (*57*, *95*), respectively, but not enzootic plague. Here, I use an extensive set of location and disease status information of coyotes that act as a wide-ranging sentinel species for sylvatic plague across the western United States.

1.2 Dissertation aims

In this dissertation, I aim to make extended use of available surveillance data to understand two things. First, where is plague transmission occurring and being maintained within a sylvatic cycle in the United States? What information do we need and how do we sample these data to predict and understand the spatial distribution of enzootic plague? Second, what is the connection between enzootic plague and human plague risk? Are humans at

increased risk of *Y. pestis* infection in enzootic plague areas compared to non-enzootic plague areas?

I designed the subsequent chapters to achieve my research aims. In Chapter 2, I collate the various sources of data used in the dissertation and discuss potential data uncertainties and biases. I account for these sources of sampling uncertainties and biases in subsequent analyses to produce robust predictions for the spatial distribution of enzootic plague and human plague risk. In Chapter 3, I propose a method that adapts a spatial statistic to estimate associations between *Y. pestis* occurrence and environmental variables. The proposed method accounts for sampling effort bias and can predict enzootic locations of plague even into areas with no historical sampling. In Chapter 4, I examine the impact of location uncertainty on my proposed coyote-based plague surveillance tools of coyote-based plague surveillance. I identify areas of my predicted spatial distribution of enzootic plague that are sensitive to not knowing the true location where coyotes were collected. Chapter 5 examines the variation in sampling effort from multiple sources of sampling data of administrative plague surveillance systems. When I adjust my predictions by sampling effort, I create a conservative estimate of the spatial distribution of enzootic plague in the western United States. In Chapter 6, I investigate the association between human plague cases in the United States and my predictions from Chapter 5. I find the spatial distribution of enzootic plague can explain a portion of why human cases have occurred in the western United States, but epizootic conditions likely better explain local human risk (*57*, *59*, *95*).

1.3 References

1. California Digital Newspaper Collection, Center for Bibliographic Studies and Research, University of California, Riverside, *San Francisco Call* **87**, (2019; http://cdnc.ucr.edu) (Mar. 7, 1900).
2. California Digital Newspaper Collection, Center for Bibliographic Studies and Research, University of California, Riverside, *Sacramento Union* **214**, (2019; http://cdnc.ucr.edu) (June 27, 1920).
3. P. Daszak *et al.*, *Science* **287**, 443–449, ISSN: 0036-8075, DOI 10.1126/science.287.5452.443 (2000).
4. P. N. Fonkwo, *European Molecular Biology Organization Reports* **9**, S13–S17, ISSN: 1469-221X, 1469-3178, DOI 10.1038/embor.2008.110 (2008).
5. K. E. Jones *et al.*, *Nature* **451**, 990–993, ISSN: 0028-0836, DOI 10.1038/nature06536 (2008).
6. L. Simonsen *et al.*, *The Journal of Infectious Diseases* **172**, 729–733, ISSN: 0022-1899, DOI 10.1093/infdis/172.3.729 (1995).
7. R. M. Anderson *et al.*, *Philosophical Transactions of the Royal Society B: Biological Sciences* **359**, 1091, ISSN: 0962-8452, 1471-2954, DOI 10.1098/rstb.2004.1490 (2004).
8. M. Eidson *et al.*, *Emerging Infectious Diseases* **7**, 615, ISSN: 1080-6059, 1080-6040, DOI 10.3201/eid0704.017402 (2001).
9. J. F. Cully Jr, in *Proceedings of the Symposium on the management of prairie dog complexes for the reintroduction of the black-footed ferret*, ed. by J. L. Oldemyer *et al.* (U.S. Department of the Interior Fish and Wildlife Service, Washington, D.C., 1993), vol. Biological Report, pp. 38–48.
10. M. E. Hugh-Jones *et al.*, *Zoonoses: recognition, control, and prevention* (John Wiley & Sons, Ltd., Hoboken, New Jersey, 2008), ISBN: 0813825427.
11. J. Grinnell, *The Auk* **34**, 427–433, ISSN: 938-4254, DOI 10.2307/4072271 (1917).
12. A. T. Peterson *et al.*, *Ecological niches and geographic distributions (MPB-49)* (Princeton University Press, Princeton, New Jersey, 2011), vol. 56, ISBN: 0691136882.
13. A. T. Peterson, *Naturwissenschaften* **95**, 483–491, ISSN: 0028-1042, 1432-1904, DOI 10.1007/s00114-008-0352-5 (2008).
14. S. S. Morse, *Emerging Infectious Diseases* **1**, 7–15, ISSN: 1080-6059, 1080-6040, DOI 10.3201/eid0101.950102 (1995).
15. J. G. Morris Jr, M. Potter, *Emerging Infectious Diseases* **3**, 435, ISSN: 1080-6059, 1080-6040, DOI 10.3201/eid0304.970404 (1997).
16. S. Cleaveland *et al.*, *Philosophical Transactions of the Royal Society B: Biological Sciences* **356**, 991–999, ISSN: 0962-8452, 1471-2954, DOI 10.1098/rstb.2001.0889 (2001).
17. R. J. Eisen, K. L. Gage, *Veterinary Research* **40**, 1, ISSN: 1993-5412, DOI 10.1051/vetres:2008039 (2009).
18. A. T. Peterson, *Emerging Infectious Diseases* **12**, 1822, ISSN: 1080-6059, 1080-6040, DOI 10.3201/eid1212.060373 (2006).
19. A. T. Peterson, *Mapping disease transmission risk: enriching models using biogeography and ecology* (Johns Hopkins University Press, Baltimore, Maryland, 2014), ISBN: 1421414737.
20. L. E. Escobar, M. E. Craft, *Frontiers in Microbiology* **7**, 1174, ISSN: 1664-302X, DOI 10.3389/fmicb.2016.01174 (2016).

21. J. E. Childs, E. R. Gordon, *Mount Sinai Journal of Medicine* **76**, 421–428, ISSN: 1931-7581, DOI 10.1002/msj.20133 (2009).

22. W. G. van Panhuis *et al.*, *BMC Public Health* **14**, 1144, ISSN: 1471-2458, 1471-2458, DOI 10.1186/1471-2458-14-1144 (2014).

23. P. Hitchcock *et al.*, *Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science* **5**, 206–227, ISSN: 1538-7135, 1557-850X, 1538-7135, DOI 10.1089/bsp.2007.0041 (2007).

24. B. M. Althouse *et al.*, *EPJ Data Science* **4**, 17, ISSN: 2193-1127, DOI 10.1140/epjds/s13688-015-0054-0 (2015).

25. S. B. Thacker *et al.*, *Morbidity and Mortality Weekly Report supplements* **61**, 3–9, ISSN: 2380-8950, 2380-8942 (2012).

26. M. Ryser-Degiorgis, *BMC Veterinary Research* **9**, 223, ISSN: 1746-6148, DOI 10.1186/1746-6148-9-223 (2013).

27. L. Simonsen *et al.*, *The Journal of Infectious Diseases* **214**, S380–S385, ISSN: 0022-1899, DOI 10.1093/infdis/jiw376 (2016).

28. S. I. Hay *et al.*, *PLoS Medicine* **10**, e1001413, ISSN: 1549-1277, DOI 10.1371/journal.pmed.1001413 (2013).

29. P. W. Gething *et al.*, *Malaria Journal* **10**, 378, ISSN: 1475-2875, DOI 10.1186/1475-2875-10-378 (2011).

30. P. W. Gething *et al.*, *PLoS Neglected Tropical Diseases* **6**, e1814, ISSN: 1935-2735, 1935-2727, DOI 10.1371/journal.pntd.0001814 (2012).

31. K. L. Gage, M. Y. Kosoy, *Annual Review of Entomology* **50**, 505–528, ISSN: 0066-4170, DOI 10.1146/annurev.ento.50.071803.130337 (2005).

32. N. Kwit *et al.*, *Morbidity and Mortality Weekly Report* **64**, 918–919, ISSN: 0149-2195, 1545-861X (2015).

33. K. J. Kugeler *et al.*, *Emerging Infectious Diseases* **21**, 16, ISSN: 1080-6059, 1080-6040, DOI 10.3201/eid2101.140564 (2015).

34. R. Pollitzer, *Bulletin of the World Health Organization* **7**, 231, ISSN: 0042-9686 (1952).

35. K. L. Gage *et al.*, *Journal of Mammalogy* **76**, 695–715, ISSN: 0022-2372, DOI 10.2307/1382741 (1995).

36. C. E. Levy, K. L. Gage, *Infections in Medicine* **16**, 54–64, ISSN: 0749-6524 (1999).

37. J. M. Doll *et al.*, *American Journal of Tropical Medicine and Hygiene* **51**, 109–114, ISSN: 0002-9637, 1476-1645, DOI 10.4269/ajtmh.1994.51.109 (1994).

38. H. L. Gould *et al.*, *Zoonoses and Public Health* **55**, 448–454, ISSN: 1863-1959, DOI 10.1111/j.1863-2378.2008.01132.x (2008).

39. C. Darby, *Trends in Microbiology* **16**, 158–164, ISSN: 0966842X, DOI 10.1016/j.tim.2008.01.005 (2008).

40. R. J. Eisen *et al.*, *Journal of Medical Entomology* **44**, 678–682, ISSN: 0022-2585, DOI 10.1603/0022-2585(2007)44[678:ETOYPB]2.0.CO;2 (2007).

41. R. J. Eisen *et al.*, *Proceedings of the National Academy of Sciences* **103**, 15380–15385, ISSN: 0027-8424, DOI 10.1073/pnas.0606831103 (2006).

42. R. J. Eisen *et al.*, *Journal of Medical Entomology* **46**, 737–744, ISSN: 0022-2585, DOI 10.1603/033.046.0403 (2009).

43. R. Pollitzer, *Bulletin of the World Health Organization* **23**, 313, ISSN: 0042-9686 (1960).

44. J. D. Poland, A. M. Barnes, in *CRC handbook series in zoonoses: Section A, Bacterial, rickettsial and mycotic diseases*, ed. by J. H. Steele *et al.* (CRC Press, Boca Raton, Florida, 1979), vol. 1, pp. 515–597, ISBN: 0849329067.

45. O. S. Serzhan, V. S. Ageyev, *Karantinye iZoonoznye Infektsii v Kazakhstane* **2**, 183–192 (2000).

46. R. J. Eisen *et al.*, *Journal of Medical Entomology* **44**, 672–677, ISSN: 0022-2585, DOI 10.1093/jmedent/44.4.672 (2007).

47. J. L. Lowell *et al.*, *Journal of Vector Ecology* **34**, 22–31, ISSN: 1081-1710, DOI 10.1111/j.1948-7134.2009.00004.x (2009).

48. J. K. Runfola *et al.*, *Morbidity and Mortality Weekly Report* **64**, 429–434, ISSN: 0149-2195, 1545-861X (2015).

49. K. L. Gage *et al.*, *Clinical Infectious Diseases* **30**, 893–900, ISSN: 1058-4838, DOI 10.1097/00006454-200011000-00032 (2000).

50. J. Gabastou *et al.*, *Transactions of the Royal Society of Tropical Medicine and Hygiene* **94**, 387–391, ISSN: 0035-9203, DOI 10.1016/S0035-9203(00)90114-7 (2000).

51. V. N Fedorov, *Bulletin of the World Health Organization* **23**, 275–281, ISSN: 0042-9686 (1960).

52. A. B. Christie *et al.*, *The Journal of Infectious Diseases* **141**, 724–726, ISSN: 0022-1899, DOI 10.1093/infdis/141.6.724 (1980).

53. A. Arbaji *et al.*, *Annals of Tropical Medicine and Parasitology* **99**, 789–793, ISSN: 0003-4983, DOI 10.1179/136485905X65161 (2005).

54. A. A. Bin Saeed *et al.*, *Emerging Infectious Diseases* **11**, 1456–1457, ISSN: 1080-6059, 1080-6040, DOI 10.3201/eid1209.050081 (2005).

55. C. F. von Reyn *et al.*, *American Journal of Tropical Medicine and Hygiene* **25**, 626–629, ISSN: 0002-9637, 1476-1645, DOI 10.4269/ajtmh.1976.25.626 (1976).

56. C. F. von Reyn *et al.*, *American Journal of Epidemiology* **104**, 81–87, ISSN: 0002-9262, DOI 10.1093/oxfordjournals.aje.a112276 (1976).

57. H. E. Brown *et al.*, *American Journal of Tropical Medicine and Hygiene* **82**, 95–102, ISSN: 0002-9637, 1476-1645, DOI 10.4269/ajtmh.2010.09-0247 (2010).

58. T. Ben-Ari *et al.*, *PLoS Pathogens* **7**, e1002160, ISSN: 1553-7374, 1553-7366, DOI 10.1371/journal.ppat.1002160 (2011).

59. T. Ben-Ari *et al.*, *American Journal of Tropical Medicine and Hygiene* **83**, 624–632, ISSN: 0002-9637, 1476-1645, DOI 10.4269/ajtmh.2010.09-0775 (2010).

60. T. Ben-Ari *et al.*, *Biology Letters* **4**, 737–740, ISSN: 1744-9561, 1744-957X, DOI 10.1098/rsbl.2008.0363 (2008).

61. T. Ben-Ari *et al.*, *Proceedings of the National Academy of Sciences* **109**, 8196–8201, ISSN: 1091-6490, DOI 10.1073/pnas.1110585109 (2012).

62. R. E. Enscore *et al.*, *American Journal of Tropical Medicine and Hygiene* **66**, 186–196, ISSN: 0002-9637, 1476-1645, DOI 10.4269/ajtmh.2002.66.186 (2002).

63. D. C. Cavanaugh, J. D. Marshall, *Journal of Wildlife Diseases* **8**, 85–94, ISSN: 0090-3558, DOI 10.7589/0090-3558-8.1.85 (1972).

64. L. Xu *et al.*, *Proceedings of the National Academy of Sciences of the United States of America* **108**, 10214–10219, ISSN: 1091-6490, 0027-8424, DOI 10.1073/pnas.1019486108 (2011).

65. K. MacMillan *et al.*, *American Journal of Tropical Medicine and Hygiene* **86**, 514–523, ISSN: 0002-9637, 1476-1645, DOI 10.4269/ajtmh.2012.11-0569 (2012).

66. K. S. Kreppel *et al.*, *PLoS Neglected Tropical Diseases* **8**, ISSN: 1935-2735, 1935-2727, DOI 10.1371/journal.pntd.0003155 (2014).

67. K. MacMillan *et al.*, *American Journal of Tropical Medicine and Hygiene* **84**, 435–442, ISSN: 0002-9637, 1476-1645, DOI 10.4269/ajtmh.2011.10-0571 (2011).

68. S. B. Neerinckx *et al.*, *International Journal of Health Geographics* **7**, 54, ISSN: 1476-072X, DOI 10.1186/1476-072X-7-54 (2008).

69. S. B. Neerinckx *et al.*, *American Journal of Tropical Medicine and Hygiene* **82**, 492–500, ISSN: 0002-9637, 1476-1645, DOI 10.4269/ajtmh.2010.09-0426 (2010).

70. Y. Nakazawa *et al.*, *Vector-Borne and Zoonotic Diseases* **7**, 529–540, ISSN: 1557-7759, 1530-3667, DOI 10.1089/vbz.2007.0125 (2007).

71. R. J. Eisen *et al.*, *American Journal of Tropical Medicine and Hygiene* **77**, 999–1004, ISSN: 0002-9637, 1476-1645, DOI 10.4269/ajtmh.2007.77.999 (2007).

72. R. J. Eisen *et al.*, *Journal of Medical Entomology* **44**, 530–537, ISSN: 0022-2585, DOI 10.1603/0022-2585(2007)44 (2007).

73. R. J. Eisen *et al.*, *American Journal of Tropical Medicine and Hygiene* **77**, 121–125, ISSN: 0002-9637, 1476-1645, DOI 10.4269/ajtmh.2007.77.121 (2007).

74. A. M. Winters *et al.*, *American Journal of Tropical Medicine and Hygiene* **80**, 1014–1022, ISSN: 0002-9637, 1476-1645, DOI 10.4269/ajtmh.2009.80.1014 (2009).

75. M. Anker, D. Schaaf, in *WHO report on global surveillance of epidemic-prone infectious diseases* (World Health Organization, Geneva, Switzerland, 2000), chap. 3, pp. 25–37.

76. R. J. Eisen, K. L. Gage, in *Yersinia: systems biology and control*, ed. by E Carniel, B. J. Hinnesbusch (Caister Academic Press, Norfolk, United Kingdom, 2012), pp. 169–182, ISBN: 1908230058.

77. D. J. Salkeld *et al.*, *BioScience* **66**, 118–129, ISSN: 2579-308X, DOI 10.1093/biosci/biv179 (2016).

78. R. J. Eisen *et al.*, *Emerging Infectious Disease* **14**, 941, ISSN: 1080-6059, 1080-6040, DOI 10.3201/eid1406.080029 (2008).

79. K. A. Boegler *et al.*, *Vector-Borne and Zoonotic Diseases* **12**, 948–952, ISSN: 1557-7759, 1530-3667, DOI 10.1089/vbz.2012.1031 (2012).

80. D. W. Markman *et al.*, *Emerging Infectious Diseases* **24**, 294–302, ISSN: 1080-6059, 1080-6040, DOI 10.3201/eid2402.171065 (2018).

81. A. M. Schotthoefer *et al.*, *Journal of Medical Entomology* **48**, 411–417, ISSN: 0022-2585, DOI 10.1603/ME10155 (2011).

82. D. C. Cavanaugh, *American Journal of Tropical Medicine and Hygiene* **20**, 264–273, ISSN: 0002-9637, 1476-1645, DOI 10.4269/ajtmh.1971.20.264 (1971).

83. J. A. Hubbart *et al.*, *Journal of Vector Ecology* **36**, 117–123, ISSN: 1081-1710, DOI 10.1111/j.1948-7134.2011.00148.x (2011).

84. M. E. Metzger, M. K. Rust, in *Proceedings of the 3rd International Conference on Urban Pests*, ed. by W. H. Robinson, G. W. Rambo, pp. 235–240.

85. S. P. Maher *et al.*, *American Journal of Tropical Medicine and Hygiene* **83**, 736–742, ISSN: 0002-9637, 1476-1645, DOI 10.4269/ajtmh.2010.10-0042 (2010).

86. J. C. Z. Adjemian *et al.*, *Journal of Medical Entomology* **43**, 93–103, ISSN: 0022-2585, DOI 10.1093/jmedent/43.1.93 (2006).

87. J. E. Childs *et al.*, *The Journal of Infectious Diseases* **169**, ISSN: 0022-1899, DOI 10.1093/infdis/169.6.1271 (1994).

88. K. G. Julian *et al.*, *Vector-Borne and Zoonotic Diseases* **2**, 145–155, ISSN: 1557-7759, 1530-3667, DOI 10.1089/15303660260613710 (2002).

89. A. C. Holt *et al.*, *International Journal of Health Geographics* **8**, 38, ISSN: 1476-072X, DOI 10.1186/1476-072X-8-38 (2009).

90. M. Walsh, M. A. Haseeb, *PeerJ* **3**, e1493, ISSN: 2167-8359, DOI 10.7717/peerj.1493 (2015).

91. D. J. Salkeld, P. Stapp, *Vector-Borne and Zoonotic Diseases* **6**, 231–239, ISSN: 1557-7759, 1530-3667, DOI 10.1089/vbz.2006.6.231 (2006).
92. R. J. Eisen *et al.*, *Journal of Medical Entomology* **45**, 1160–1164, ISSN: 0022-2585, DOI 10.1093/jmedent/45.6.1160 (2008).
93. J. D. Lang, *Journal of Vector Ecology* **29**, 236–247, ISSN: 1081-1710 (2004).
94. K. L. Gage, in *Advances in Yersinia research* (Springer, New York, New York, 2012), pp. 79–94, ISBN: 1461435600.
95. H. E. Brown *et al.*, *Vector-Borne and Zoonotic Diseases* **11**, 1439–1446, ISSN: 1557-7759, 1530-3667, DOI 10.1089/vbz.2010.0196 (2011).

2 Surveillance data for *Yersinia pestis* in the United States

"Just as you write in part in order to figure out what you trying to say you are in the business of revealing your thoughts to yourself, even persuading yourself, on the way to persuade others, so you do statistics not just to learn from data, but also to learn what you can learn from data, and to decide how to gather future data to help resolve key uncertainties."

- "Ethics in statistical practice and communication: Five recommendations" in *Significance* by Andrew Gelman, Ph.D. (*96*)

2.1 Introduction

Effective zoonotic disease surveillance relies on rich data from observational studies and laboratory analyses, building on dedicated efforts by data collectors, including public health scientists, field biologists, laboratory technicians, and medical professionals. Animal reservoirs must be sampled, processed, and tested and human cases must be admitted, diagnosed, and treated, before a public health scientist can detect patterns in the disease. However, surveillance can be imperfect and faces challenges for zoonoses caused by pathogens transmitted directly or indirectly from animal to human hosts, especially if infected animal hosts are difficult to observe or if there is limited opportunity to observe the disease until after transmission to human hosts occurs (*21*). Linking various data sources may help overcome imperfect data collection by informing reasons for missingness, accounting for confounding factors, or identifying potential biases in the data collection.

This dissertation proposes analytic methods to predict the risk of a rare, deadly zoonotic infectious disease in humans using a combination of environmental variables and the location and disease exposure status of a wildlife host species. Here, I collate various sources of secondary data to explore the spatial distribution of enzootic *Yersinia pestis*, the pathogen causing plague, in the United States. Data used in the dissertation do not involve primary data collection (i.e., collected by investigator) because of the availability of secondary data (i.e., collected by various public health, biological, and climatological agencies). While pooling secondary data is an advantage for zoonotic disease surveillance studies, one must

take into consideration specific quality issues as outlined in this chapter. Sources of potential uncertainty and bias are different in each dataset, and I carefully catalog the sources, features, and motivations for the various data sources supporting the methods and analyses in subsequent chapters. The chapter begins with a listing of multiple sources of data used in my analysis in the dissertation (Figure 1), which includes human case data, animal case data, and environmental data. The dissertation primarily focuses on coyote (*Canis latrans* Say, 1823) data from ongoing plague surveillance systems for the aim of predicting where *Y. pestis* is being maintained in the western United States by cycling between animal hosts. The second part of the chapter discusses the quality, uncertainty, and potential biases of the data.

2.2 Cataloging data sources

From a One Health perspective (*97–100*), the dynamics of a pathogen occur at the intersection of humans, animals, and the environment. Understanding the dynamics of plague requires analyses that are data-intensive for every actor at this intersection. It is important to catalog data sources because data from multiple sources are often collected for differing purposes. To focus the discussion, I carefully describe and review my sources for human data, animal data, and environmental data.

2.2.1 Human data

Plague is a Category A infectious disease caused by the bacterium *Y. pestis* and human cases are reportable to the World Health Organization (WHO), U.S. Centers for Disease Control and Prevention (CDC), and state health authorities. Plague is potentially deadly and can be used as a biological weapon (*101*). The WHO case definition of plague is found in Panel 1. Diagnostic tests for *Y. pestis* are found in Panel 2. Human plague surveillance is conducted by local, state, and federal health agencies and human plague cases are rare in the United States (*33*). Epidemiological investigations seek to identify the source of infection and exposure locations in order to minimize risk of additional cases. I use human-plague case data and human population data in Chapter 6 to predict the relative risk of plague in humans across counties of the western United States (1950–2017).

State health departments The CDC receives aggregated human plague case data from state health departments on an annual basis. Data are reported monthly at the county level to protect the privacy of cases. Since its introduction in 1900, *Y. pestis* has infected 1,045 people in the United States (1900–2017; Table 1 and Figure 2a). Plague arrived in California in 1900 and quickly jumped from commensal rats into native rodent populations; it spread eastward and stopped approximately at the 100th Meridian, which passes through the middle of Kansas (*102*). Human cases before the 1930s typically concentrated in port cities (e.g., San Francisco, CA and Los Angeles, CA), but by 1950, plague became locally enzootic in the western United States and human cases occurred primarily in the country's interior. Since 1950, human plague cases have occurred in 13 western states (521 total cases), but over 80% of these cases have occurred in the Four Corners states (Arizona, Colorado, New Mexico, and Utah; Figure 2b; (*32*, *33*, *103*)). Since 2000, there have been an average of 6 cases per year (less than 1 death per year on average). The majority of human plague cases occur April through September. Cases reported between the October and February "Off-Season" have been linked to rabbit-hunting season (*104*). Since 1983, 26 human plague cases have been reported in California, primarily in mountainous and foothill areas of the Sierra Nevada, Cascades, and Transverse mountain ranges (unpublished data, California Department of Public Health).

United States Census Bureau Every 10 years, the U.S. Census Bureau conducts a count to determine the number of people living in the United States. Socio-demographic data associated with each household are recorded. Additionally, since 2005, the U.S. Census Bureau conducts the American Community Survey (ACS) which collects additional information about 3.5 million households in the United States each year. In 2010, the ACS replaced the (decennial) long form of the United States census for high-resolution geographic information about the United States population. The human population density of the western United States in 2010 appears in Figure 3a and the percent change in human population of the western United States (1950–2010) appears in Figure 3b.

2.2.2 Animal data

Surveillance systems of high consequence pathogens are important because human treatment often relies on early case detection, but when human infections are rare it can be challenging to predict and monitor high-risk areas. Plague is a potentially deadly zoonotic pathogen, an infectious disease of animal origin where *Y. pestis* is maintained by rodent communities and their fleas. While rare in humans, plague infections are usually linked to epizootics, or outbreaks of plague in rodents (*57*), which supports the use of animal case data to inform human plague dynamics.

Animal-based surveillance for plague focuses on rodent species because plague is primarily a disease of rodents. Epizootics are more readily detected in rodent species that are highly susceptible to plague. For example, prairie dogs (*Cynomys* spp.) suffer dramatic die-off events caused by plague marked by up to 98% mortality of an affected population (*105*, *106*). Active surveillance of rodent populations and their fleas is conducted in the United States (*107*), typically in plague enzootic regions and focusing on areas of high public health concern such as recreational areas with historical plague activity (*108*). Enhanced environmental surveillance generally occurs in response to a reported human plague case. However, human disease is rare (i.e., typically low risk), so active rodent plague surveillance is intermittent. Active rodent plague surveillance is costly and frequently yields few positive detections (*107*, *109–111*), so it is not implemented systematically across or within states.

Instead, animal-based plague surveillance systems monitor animal species that can act as a sentinel of plague in rodents. Coyotes are often used as a sentinel species of plague in the United States (*95*, *112–116*) (but see *91*, *117*, and Panel 3). Plague-positive coyote locations are associated ecologically with plague-positive rodent species locations in California (*89*) and coyote plague cases are temporally associated with human plague cases between 1974 and 1998 in Arizona (*95*). Coyotes are carnivores that scavenge carcasses and prey on potentially infectious rodents, and typically survive plague infection (*109*). In response to exposure to *Y. pestis*, surviving coyotes develop long-lasting antibodies (*91*, *112*, *118*); via this immunological evidence, coyotes effectively act as wide-ranging indicators of plague activity. For coyote specimens used in this dissertation, whole blood is collected in the

field on strips of filter paper, called "Nobuto strips," allowed to dry, and then transported to a laboratory where the dried strips are eluted and tested for plague antibodies (*119*). Diagnostic tests for *Y. pestis* appear in Panel 2. A positive test result indicates that the coyote was exposed to and developed an antibody response to *Y. pestis*. The antibody titer is also recorded where high antibody titer (1:256 or greater) is assumed to indicate recent plague exposure (*95*).

In this dissertation, I use coyote-plague case data (Chapters 3-5) and historical coyote observation records (Chapter 5) to estimate the ecological niche of plague in coyotes and predict its enzootic spatial distribution in California and the western United States. A "niche" is the range of biotic and abiotic factors and interactions that permit a population of a species to persist in a given habitat (e.g., collection of resources) without immigration from another population, and outside of these habitats with permissive factors a species faces extinction (*111*). A range of agencies conducts animal-based plague surveillance using multiple animal species. A review by Bevins and colleagues (*107*) found that multiple state health departments, research universities, the CDC, and the U.S. Department of Agriculture (USDA) monitor plague. This dissertation collaborated with three of these agencies. An overview of each agency and their data sources are detailed below.

United States Department of Agriculture In 2005, the USDA Animal and Plant Health Inspection Service (APHIS) National Wildlife Disease Program (NWDP) began conducting continuous animal-based plague surveillance across the United States, including Alaska. Collected blood samples were sent to the CDC for screening using passive hemagglutination (Panel 2) until 2016, at which point the USDA developed a novel screening test for *Y. pestis* exposure (*120*). Between 2005 and 2010, the USDA tested over 25,000 samples for plague from more than 70 species primarily from taxonomic groups with previously documented *Y. pestis* exposure (*107*). Of these, coyotes had the highest average seroprevalence (*107*), consequently I focus on the coyote data.

In this dissertation, I use data from a 13-year (2005–2017) plague-surveillance system conducted by the USDA. Over 28,000 coyotes were screened for *Y. pestis* exposure. Plague antibodies were detected in 12.7% of coyotes in 15 plague enzootic states ($n = 3,665$) across

all 13 years of surveillance. In order to protect landowner privacy, county-level USDA coyote sampling appears in Figure 4 and counties with at least one coyote observed with plague antibodies appear in Figure 5. I do not present crude rates of seropositive coyotes across the western United States in order to avoid problematic interpretation. The reported rate of positive tests does not necessarily reflect the true prevalence of plague due to the nature of wildlife sampling for plague as defined below.

California Department of Public Health Since 1900, the California Department of Public Health (CDPH) has recorded the geographic location of animals screened for plague exposure by CDPH and local California health department partners. A current, digitized database includes data beginning in 1983. The CDPH conducts active surveillance of plague in rodent populations across California, including repeat sampling in plague enzootic areas (Figure 6) with priority given to higher risk locations such as heavily used campgrounds and other outdoor recreational areas (*108*). The CDPH typically conducts enhanced rodent and flea surveillance in response to and in the vicinity of any observed human plague cases to ascertain the potential source of human plague exposure. Passive surveillance occurs when the public or recreational area staff reports animals (primarily rodent carcasses) that are suspect for plague infection. The CDPH will test these specimens, and their fleas, for *Y. pestis* (*108*). Additionally, in conjunction with livestock and wildlife damage management (i.e., depredation) activities conducted by USDA APHIS Wildlife Services (WS), coyotes are sampled and tested for plague by the CDPH.

The CDPH uses coyote observations as surrogates for rodent surveillance in regions of California with lower resources or access. Coyote data are also retrospectively analyzed to corroborate regional increases in plague activity. For example, coyote-plague surveillance is either used to confirm what has been seen in rodent-plague surveillance or as an indicator for plague activity in areas not captured by rodent-plague surveillance. Between 1983 and 2015, the CDPH tested 8,119 coyotes for *Y. pestis* antibodies via the CDPH Vector-Borne Disease Laboratories (Table 2). Confirmation of *Y. pestis* exposure is made by direct florescent antibody (FA) test (Panel 2). In order to protect landowner privacy, county-level CDPH coyote sampling appears in Figure 7a and counties with at least one coyote observed

with plague antibodies appear in Figure 7b. Similar to the USDA plague data, I do not present crude rates of coyotes that tested positive for plague antibodies across California.

Other state health departments State health departments partner with the CDC to monitor plague in wildlife. Animals are sampled using active and passive surveillance in response to a suspicious animal or nearby human plague case. State health departments send blood samples to the CDC in Fort Collins, Colorado for laboratory testing using either passive hemagglutination assay, F-1 inhibition test, FA test, or bacterial culture (Panel 2). Test results are recorded in a massive, paper-based archive of animal-based plague surveillance in North America maintained by the CDC National Center for Emerging and Zoonotic Infectious Disease Division of Vector-Borne Diseases Bacterial Diseases Branch. A digitization effort by Maher and colleagues (*85*) geocoded 3,777 occurrence points between 1998 and 2008 that included generic or specific reference to host species for 75 mammal taxa, primarily *C. latrans* ($n = 2,516$; 67% of samples). County-level state health department coyote sampling appear in Figure 8. I use these data in Chapter 5 to externally validate the predicted spatial distribution of enzootic plague in the western United States.

Museum collections I collated locations of coyotes observed in the United States to determine the background coyote population. These data are an approximation of coyote habitat preference. When compared to coyote-based plague surveillance data, these data help account for sampling effort of coyote-based plague surveillance systems in the analysis in Chapter 6. Coyotes are observed by individuals including university principal investigators, biologists, citizen scientists, and conservation agencies that record observations in databases maintained by various museums and government agencies. Here, I assemble coyote observations across the United States from three main databases: 1) United States Geological Survey–Biodiversity Information Serving Our Nation (*121*), 2) Global Biodiversity Information Facility (*122*), and 3) VertNet (*123*). Across the United States, 13,972 unique coyote occurrences were reported (Table 3 and Figure 9). These coyotes are not necessarily tested for plague antibodies but are included in my analysis because they have accompanying location information.

2.2.3 Environmental data

I use environmental data in Chapter 3, Chapter 4, and Chapter 5 to estimate the ecological niche of plague in coyotes and predict its enzootic spatial distribution in California and the western United States. An ecological niche manifests within a "predictor space" (or "environmental space" or "state space"), which includes climatological variables such as temperature and precipitation (*124*).

PRISM Climate Group Oregon State University has created a climatological data product using a network of ground measurements. The measurements are used in a Parameter elevation Regression on Independent Slopes Model (PRISM) statistical mapping system that follows a weighted regression framework relying on digital elevation models (DEM; *125*). I use 30-year average normals (1981–2010) at a 2.5 arcminute (∼4 kilometer by 4 kilometer; ∼16 square kilometers) resolution for the western United States. A list of variables I use in the analysis appears in Table 4. I chose the PRISM data because the temporal range overlapped with CDPH plague surveillance (1983–2015) and PRISM is the official climatological data of the USDA. Data are available from the PRISM portal (http://prism.oregonstate.edu) or via the **prism** package (*126*) in the statistical software R (*127*).

National Aeronautics and Space Administration The National Aeronautics and Space Administration Shuttle Radar Topology Mission (NASA-SRTM) provides Digital Terrain Elevation Data for the entire world (*128*). The data are freely available through the U.S. Geological Survey via the National Map Seamless Data Distribution System (https://www.usgs.gov/core-science-systems/ngp/tnm-delivery) or via the **raster** package (*129*) in the statistical software R (*127*). In order to examine sampling uncertainty of CDPH coyote-based plague surveillance I aggregate the elevation of California from a resolution of 90 m to a 4 km resolution and record the variability in elevation within each 16 km^2 grid cell. I use this layer to categorize coyotes in Chapter 4.

2.3 Cataloging data quality

Data I use in this dissertation are from various sources and are collected for differing purposes using differing methods. Drawing the best inference requires an understanding of the strengths and weaknesses of each dataset for the purposes of my dissertation.

2.3.1 Quality of human data

Assessing the uncertainties and biases of human data is challenging due to the small numbers of human cases each year. I focus on the more numerous animal cases to provide insight on local human risk in this dissertation, but it is important to note data quality issues of human data. Human plague case data in this dissertation are generally considered exhaustive, but the spatial resolution of available data leads to important analytical considerations. The vast majority of human plague cases has been captured by United States human disease surveillance systems because plague symptoms are severe. There is evidence of asymptomatic human plague cases outside the United States (*130*, *131*), which suggests the possibility, albeit small, of unreported human cases in the United States. Additionally, plague is a rare disease and cases are likely identifiable to the probable site of exposure at the sub-county level. Therefore, I sought finer spatial resolution data from the CDC and state health department collaborators, but they are ultimately unavailable due to data privacy reasons. Counties, especially in the western United States, can represent large areas (e.g., 610 counties in United States are larger than the state of Rhode Island by area; *132*). This presents a challenge for information aggregation.

State health departments Uncertainty arises about the location, source, and transmission pathway of exposure during epidemiological investigations. Some human cases may not be from residents of counties but are imported (see *133*) and information about the source or transmission pathway of human cases is not available for all cases. Nearly 17% (38 of 224) of plague cases in New Mexico (1960–2003) were not exposed near their homes (*73*) and the site of exposure was not determined for about 9% (30 of 356 cases) of human plague cases in the southwestern United States (1957-2004; *72*).

In Chapter 6, I assess how well my prediction of where *Y. pestis* is located in the en-

vironment on an annual average explains where human cases were located in the western United States (1950–2017). I propose analytic adjustments to the predicted relative risk of human plague for these uncertainties in various ways. I expand the inclusion criteria for plague cases to the county-of-exposure instead of near a person's home, which allows for a larger sample size and captures a greater activity space of where a human may have been exposed to *Y. pestis*. Also, I assume all cases were exposed from a vector (flea or mammal; wildlife or domesticated animal) in an area suitable for plague transmission. There has only been one documented human-to-human transmission event in the United States since 1924 (*48*).

United States Census Bureau There are known uncertainties associated with the United States census data, as recently reviewed by Spielman and colleagues (*134*). Both the decennial long form and American Community Survey are sample surveys and have accompanying sampling error. Three main sources of sampling error are 1) sample size, 2) heterogeneity of the population being measured, and 3) response rate (*135*). Additionally, uncertainty from the geographic resolution of census estimates increases as resolution increases, especially in sparsely populated areas. Sampled areas are pooled across multiple years (typically five years) to counteract this source of uncertainty. The ACS has larger margins of error than the 2000 long-form estimate (on average 75 times larger; *136*), where about 25% of this increase is due to the decrease in sample size from the long-form census to ACS (*137*). This uncertainty is heterogeneous across the country and is the subject of ongoing research (*134*). The U.S. Census Bureau has made great efforts to reduce bias in both forms of the census, so the estimates are, on average, reasonable approximations of the United States population. However, the estimates are based on the sampling design and implementation and I do not consider this source of uncertainty in the dissertation because available disease mapping techniques do not account for this margin of error.

2.3.2 Quality of animal data

Wildlife disease surveillance often consists of multiple sources of uncertainty and potential sampling biases because large-scale administrative wildlife disease surveillance systems aim

for three broad, but non-mutually exclusive goals: 1) to fulfill administrative mandates, 2) to remain within budgetary constraints, and 3) to contribute to wider scientific knowledge. For coyote-based plague surveillance, these goals converge because agencies aim to optimize the detection of plague cases (i.e., number of cases) by adaptively deciding where surveillance operations (i.e., which coyotes to sample and test from depredation efforts) should occur while staying within budget. Budget limitations influence the sampling design employed for surveillance by first meeting administrative goals (e.g., protect public health by focusing on high risk areas) in the most scientifically rigorous manner possible (e.g., balanced random sampling across entire study region), then, where feasible, by meeting scientific goals (e.g., gain insight into comprehensive plague dynamics in wildlife throughout a study region) that may be outside of the initial administrative purview. Because I link multiple coyote-based plague surveillance datasets from different agencies, it is important to catalog and consider the differing administrative motivations and constraints. For example, coyotes sampled by the USDA are drawn from coyotes managed on agricultural land for the purpose of livestock harassment abatement or other reasons. Therefore, sampling is not designed specifically for plague detection. In contrast, coyote sampling by state health departments and testing by the CDC (not the coyotes submitted by the USDA 2005–2016) mainly occurs in human populated areas for the purpose of protecting human health. Therefore, applying these data to the examination of broader questions must be accompanied by an understanding of the underlying motivation of data collection in order to account or adjust for potential statistical biases and uncertainties. As outlined in the Chapter 3, accurate analysis will require statistical modeling of both the observation process and the underlying population of the hosts and the pathogen. These priorities drive the methodology development in this dissertation.

This dissertation aims to extract improved value of information from coyote-based plague surveillance in the United States. While data collection methods of large-scale administrative surveillance systems of plague are not statistically perfect, it is important to recognize that the data are not absent of scientific information. The complex and heterogeneous data provide opportunities to identify the potential sources of statistical biases and uncertainties that may affect this and other zoonotic disease surveillance systems. Identifying these

biases and uncertainties allows statistical modeling of the observational process, which is of interest to disease surveillance agencies in order to inform current and future operational decisions. Additionally, "bias" is not necessarily used as a negative statement but rather as a term that reflects the error from the focused nature of data collection to meet administrative mandates. This investigation of plague in the United States requires analytic methods that describe and understand *both* the underlying ecology of plague as well as the realities of a multi-source coyote-based data collection process. The elements of this framework are not novel, but new analytic methods acknowledge the importance of accounting for the observation process of data collection when examining the underlying disease process. A simple example is the expansion of inverse probability weighting (*138*) in observational health studies (*139*). Methodological analogs can be found in ecology and environmental health, too (*140–143*). Subsequent chapters draw from these varied examples in order to model the observational process of coyote-based plague and assess the effect of some, but not all, potential sources of biases and uncertainties on the prediction of the underlying wildlife plague system.

Drawing from the statistical literature, phrases such as "imperfect sampling" are used to reflect variations from simple random sampling, a data collection method that would be infeasible to conduct at a large spatial scale (i.e., statewide). Also, an example of an imperfect sampling method is "preferential sampling," which is used here to reflect data collection driven by multiple priorities described above (Figure 10). Using preferentially sampled coyote-based plague surveillance by the USDA as an example, higher frequency of human-coyote or livestock-coyote interactions increase the chance the USDA will be tasked with managing those individuals. Therefore, the coyotes screened for plague are not evenly spatially distributed in a study region nor are they uniformly sampled from the population of coyotes at risk for plague. Additionally, not all coyotes are tested for plague exposure. Plague is a relatively rare disease of high consequence and surveillance is costly, monetarily and politically. While detecting fewer cases is generally considered ideal (i.e., less disease), surveillance systems with low detection may generate skepticism (i.e., potentially missing the disease when truly present) or fiscal concerns. Therefore, agencies may have criteria to prioritize or inform which coyotes to test to optimize detection.

Coyote natural history, ecology, and immunological response to *Y. pestis* add uncertainty around the location of disease exposure. The exact location and time of plague exposure by an individual coyote is not discernible because of the potentially large home range size of individual coyotes (*144*, *145*) and laboratory testing of plague relies on antibodies that may be lifelong and not from a recent plague exposure. Indeed, coyotes can maintain low levels of plague antibodies for long periods (*118*, *146*). The sampling location of a coyote is often not precisely identifiable in the field especially for historical observations before the Global Positioning System became readily accessible by public health agencies. The relationship between the sampling location and the exposure location of a coyote observation is neither known nor explored in this dissertation.

Potential reinfection adds uncertainty to the location and time of exposure of a coyote. Epizootic events (i.e., disease outbreak in animal populations) are infrequent with inter-epizootic periods ranging from a little over 12 months (*147*) to 50 years (*148*). In the United States, black-tailed prairie dog (*Cynomys ludivicianus*) die-offs caused by plague appear to follow the 3-7 year El Niño Southern Oscillation (*149*). The average lifespan of a coyote in the wild is 6–8 years (13.5 max; *150*) so an adult coyote will likely survive beyond an inter-epizootic period and has the potential to be reinfected with plague or be exposed during enzootic periods. Investigating this source of observation uncertainty is not addressed in this dissertation, but methods developed in the dissertation could feasibly be extended to this question in future work.

Knowledge of the background population of the wildlife species helps to more accurately determine spatial distribution of a disease in wildlife, but determining the size of a wildlife population has its own challenges. Preferential sampling of coyotes across the United States can lead to spatial statistical bias where areas of higher human population density (*151*) or park visitations (*152*) may lead to more human-coyote interactions and more recorded coyote observations. Human development influences coyote population ranges, activity, and density (*153–156*), but the direction of these relationships is not clear and may be confounded by factors such as the study region or health status of individual coyotes (*157*).

The identification and correction for sources of statistical biases and uncertainties in wildlife disease surveillance is an active area of study (*158*, *159*). However, plague surveil-

lance involving coyotes has been neglected by some public health agencies because of its statistical biases and uncertainties in addition to the high cost of surveillance for a low priority infectious disease. This dissertation uses four sources of coyote data and each has various nuances that require careful consideration before linking data sources for my analyses.

United States Department of Agriculture Not all coyotes managed by USDA APHIS WS are tested for *Y. pestis* antibodies. USDA plague testing primarily occurs in regions with previously documented plague activity. The vast majority (99.6%) of tested coyotes (2005–2017) was observed in states with previously documented plague activity and plague was not detected outside of known plague enzootic states ($n = 136$ tested coyotes outside western states). In 2016 alone, the USDA managed 70,719 total coyotes of which 2,449 were sampled and ultimately 1,130 were screened for *Y. pestis* antibodies (*160*). However, there is little risk of sampling an individual multiple times during the analysis because the vast majority of samples is taken after the USDA has euthanized a coyote (*160*).

I conducted a geocoding effort for specimens without latitude and longitude coordinates using location descriptions, county data, and base maps. There were 1,480 coyotes with missing or erroneous coordinates (5.1% of coyotes). I geocoded 915, but excluded 565 from my analysis due to a lack of location information (Table 5). Over 95% of excluded coyotes were located in one county (Yakima County, Washington) and I exclude only two plague-positive coyotes from my analysis because they had an indeterminate location (Table 5).

California Department of Public Health The majority of coyotes provided by USDA APHIS WS to CDPH are tested by the CDPH Vector-Borne Disease Laboratories. The CDPH prioritizes testing for specimens in areas with historical plague observations or in regions of California with lower resources or access in order to optimize the detection of *Y. pestis* with limited laboratory resources. However, various efforts to test coyotes in historically untested areas of California have been conducted when resources were available. The vast majority of coyotes did not have precise accompanying sampling location information. I geocoded specimens without latitude and longitude coordinates recorded in the field using location descriptions, county data, and base maps. I was unable to locate 2%

of tested coyotes ($n = 164$) below the county-level and I ignored these specimens in my analysis (Table 5). The effect of uncertainty in recorded location of coyote observations on the prediction of the spatial distribution of enzootic *Y. pestis* is the focus of Chapter 4.

Due to ecological, immunological, and observational factors, the CDPH considers coyote-plague surveillance as a regional (and trailing) indicator of rodent-plague. The large home range size, long-term immunity, frequent opportunities for repeat exposures to plague create uncertainty in the location of plague. Determining the true location and time of plague exposure based on ecological, immunological, and observational factors is outside the scope of this dissertation but is an active area of investigation by the USDA.

Other state health departments State health departments partner with the CDC to monitor plague in wildlife, including coyotes. Plague-positive coyote location data are provided by Maher and colleagues (*85*) containing 93.8% of coyote locations from the original analysis ($n = 2,360$). All specimens in this dataset are seropositive since seronegative specimens were not digitized. Additionally, data for coyotes tested before 1998 are maintained by the CDC in Fort Collins, Colorado, but were not digitized. Specimens tested since 2008 are digitized and maintained by the CDC in Fort Collins, Colorado, but were unobtainable at the time of analysis. While Maher and colleagues (*85*) used a statistical method that can draw inference from only seropositive cases, adjustments for sampling effort are limited without digitized seronegative coyotes. Methods using both positive and negative data are generally preferred to positive-only analyses when data are available (*161–164*).

A consideration when using state health department coyote data for external validation when predicting the spatial distribution of enzootic plague in the western United States is the mismatch in the main objective of the coyote-plague surveillance between the USDA (i.e., opportunistic plague analysis of coyotes handled/euthanized for non-disease related reasons) and state health departments (i.e., public health). Coyotes used to train the prediction models later in the dissertation (USDA data) were not sampled in the exact same manner as the coyotes used to test the prediction (CDC data), which may lead to poorer validation statistics if the climate is dissimilar between rural and urban areas of the western United States. However, at the spatial scale of my analysis climate is similar between rural

and urban areas of the western United States and evaluation data must be independent of the calibration data, including being collected by independent means and protocols (*12*). Thus using CDC coyote data for external validation is a thorough and rigorous evaluation of the prediction using USDA coyote data. There are duplicates between the state health department data (1998–2008) and the USDA data (2005–2017) because the CDC tested coyotes for the USDA until 2012 ($n = 552$ duplicates within $n = 2,360$ total seropositive coyotes tested by the CDC). Complete metadata are not available from the CDC, which makes it challenging to identify duplicates. I consider duplicate coyote records between the CDC and the USDA as coyotes with exact spatial coordinates using the **sp** package (*165*, *166*) in the statistical software R (*127*). This is not a perfect duplicate detection method because there is a risk of excluding coyotes that may not truly be duplicates (false-positive error). However, the method is conservative because I maximize the chance that the state health department data and the USDA data are independent.

Museum collections Biodiversity databases improve data access and availability for investigations of species distributions; however, these data are often spatially biased. Sampling effort, spatial scale, type of data collected, and data storage protocols distort large-scale biodiversity patterns (*167–169*). Not all organisms are observed, of those that are observed not all are recorded, and of those recorded not all have accompanying geolocation information. Not all organisms are systematically observed *a priori* and sampling effort is unbalanced, both spatially and by species, for many reasons including the difficulty of observing a cryptic species (*19*, *170*).

This dissertation focuses on one species common throughout the United States and with an expansive species home range. Not all coyote records in the databases have coordinates and I did not conduct a geocoding effort for coyotes with missing coordinates. However, coyote records with coordinates pass quality control measures conducted by the repository to ensure accurate geolocation information. Additionally, observations are not restricted to the present study period (1983–2017) and include historical coyote occurrences (early 1900s) in order to capture as many records as possible to balance animal-based plague surveillance sampling effort. Coyotes are found across the majority of North America (*171*) and are not

homogeneously distributed within their species home range. Coyote habitat preferences or individual coyote home range information are better indicators of where coyotes reside than species home range, but are unavailable for the spatial extent of my analysis.

2.3.3 Quality of environmental data

The environmental data sources I use in this dissertation are data products with their own uncertainties and biases. Because the scientific community frequently use these open sources, great efforts, both internally and externally via peer-review, have been made to validate these data products as well as to assess their statistical biases and measure their uncertainties. A non-exhaustive review of the known biases and uncertainties of these data sources is included below.

PRISM Climate Group The Oregon State PRISM data products have been extensively peer-reviewed (*125*, *172–175*). PRISM develops local climate-elevation regression functions for each DEM grid cell, while weighting stations on their similarity to the target grid cell (*172*, *176*). The regressions can account for 1) spatially varying elevation relationships, 2) effectiveness of terrain as barriers, 3) terrain-induced climate transitions, cold air drainage, and inversions, and 4) coastal effects (*177*).

However, PRISM does have a few uncertainties and limitations. The DEMs rely on the accuracy of the underlying elevation data taken from the Defense Mapping Agency at one-degree resolution. PRISM has an accuracy of 130 m circular error with 90% probability. Additionally, long-term datasets such as the ones used in the analysis use all weather stations in the network to calculate the data surface regardless of known systematic sampling bias such as time of observation. Daily readings may occur in the afternoons (higher temperatures) or the mornings (lower temperatures) leading to observation bias. PRISM data products at smaller temporal resolution adjust for this bias, but the annual average normals I use in my analysis do not.

National Aeronautics and Space Administration NASA-SRTM has collected data from 80% of all land on the Earth. Here, I use these data for the state of California,

which was collected by NASA-SRTM at least once if not multiple times over the course of the mission. The NASA-SRTM targeted a vertical accuracy of 16 m absolute error at 90% confidence (*178*) and a Linear Error 90 equivalent to the root mean squared error of 9.73 meters. This target was successfully achieved (*179*), although there is some evidence more recent elevation data are more accurate than NASA-SRTM in extremely mountainous regions (e.g., the Himalayas; *180*). This level of uncertainty is not a large concern because the data are aggregated from a 90 m to a 4 km resolution.

2.4 Summary

Zoonotic disease surveillance is challenging and requires careful curation of multiple, heterogeneous data sources of various information and varying quality. This effort involves close collaborations with multiple U.S. government agencies at both the state and federal level. Beyond data sharing relationships, these collaborations are essential to understand the intricacies and context of real-world zoonotic disease surveillance systems. By collaborating with federal and state agencies that monitor plague activity I enhance the interpretation of my results by re-evaluating the analytical constraints, important considerations about data privacy, and policy implications. Throughout this dissertation, a balance is struck between data openness and privacy. Efforts were made for data and method transparency while being considerate of how each collaborating agency is mandated, especially if the aims appear tangential from the original scope of data collection. Unless otherwise specified, the data I use in this dissertation are available through data usage agreements with each collaborating agency.

2.5 References

12. A. T. Peterson *et al.*, *Ecological niches and geographic distributions (MPB-49)* (Princeton University Press, Princeton, New Jersey, 2011), vol. 56, ISBN: 0691136882.

19. A. T. Peterson, *Mapping disease transmission risk: enriching models using biogeography and ecology* (Johns Hopkins University Press, Baltimore, Maryland, 2014), ISBN: 1421414737.

21. J. E. Childs, E. R. Gordon, *Mount Sinai Journal of Medicine* **76**, 421–428, ISSN: 1931-7581, DOI 10.1002/msj.20133 (2009).

32. N. Kwit *et al.*, *Morbidity and Mortality Weekly Report* **64**, 918–919, ISSN: 0149-2195, 1545-861X (2015).

33. K. J. Kugeler *et al.*, *Emerging Infectious Diseases* **21**, 16, ISSN: 1080-6059, 1080-6040, DOI 10.3201/eid2101.140564 (2015).

48. J. K. Runfola *et al.*, *Morbidity and Mortality Weekly Report* **64**, 429–434, ISSN: 0149-2195, 1545-861X (2015).

57. H. E. Brown *et al.*, *American Journal of Tropical Medicine and Hygiene* **82**, 95–102, ISSN: 0002-9637, 1476-1645, DOI 10.4269/ajtmh.2010.09-0247 (2010).

72. R. J. Eisen *et al.*, *Journal of Medical Entomology* **44**, 530–537, ISSN: 0022-2585, DOI 10.1603/0022-2585(2007)44 (2007).

73. R. J. Eisen *et al.*, *American Journal of Tropical Medicine and Hygiene* **77**, 121–125, ISSN: 0002-9637, 1476-1645, DOI 10.4269/ajtmh.2007.77.121 (2007).

85. S. P. Maher *et al.*, *American Journal of Tropical Medicine and Hygiene* **83**, 736–742, ISSN: 0002-9637, 1476-1645, DOI 10.4269/ajtmh.2010.10-0042 (2010).

86. J. C. Z. Adjemian *et al.*, *Journal of Medical Entomology* **43**, 93–103, ISSN: 0022-2585, DOI 10.1093/jmedent/43.1.93 (2006).

89. A. C. Holt *et al.*, *International Journal of Health Geographics* **8**, 38, ISSN: 1476-072X, DOI 10.1186/1476-072X-8-38 (2009).

91. D. J. Salkeld, P. Stapp, *Vector-Borne and Zoonotic Diseases* **6**, 231–239, ISSN: 1557-7759, 1530-3667, DOI 10.1089/vbz.2006.6.231 (2006).

95. H. E. Brown *et al.*, *Vector-Borne and Zoonotic Diseases* **11**, 1439–1446, ISSN: 1557-7759, 1530-3667, DOI 10.1089/vbz.2010.0196 (2011).

96. A. Gelman, *Significance* **15**, 40–43, ISSN: 1740-9705, DOI 10.1111/j.1740-9713.2018.01193.x (2018).

97. L. J. King *et al.*, *Journal of the American Veterinary Medical Association* **233**, 259–261, ISSN: 1943-569X, DOI 10.2460/javma.233.2.259 (2008).

98. J. Zinsstag *et al.*, *Preventive Veterinary Medicine* **101**, 148–156, ISSN: 0167-5877, DOI 10.1016/j.prevetmed.2010.07.003 (2011).

99. G. McNew, in *Plant Pathology: An Advanced Treatise*, ed. by J. G. Horsfall, A. E. Dimond (Academic Press, New York, New York, 1960), vol. 2, pp. 19–69.

100. K. G. Scholthof, *Nature Reviews Microbiology* **5**, 152, ISSN: 1740-1526, DOI 10.1038/nrmicro1596 (2007).

101. T. V. Inglesby *et al.*, *Journal of the American Medical Association* **283**, 2281–2290, ISSN: 0098-7484, DOI 10.1001/jama.283.17.2281 (2000).

102. J. Z. Adjemian *et al.*, *American Journal of Tropical Medicine and Hygiene* **76**, 365–375, ISSN: 0002-9637, 1476-1645, DOI 10.4269/ajtmh.2007.76.365 (2007).

103. R. B. Craven *et al.*, *Journal of Medical Entomology* **30**, 758–761, ISSN: 0022-2585, DOI 10.1093/jmedent/30.4.758 (1993).

104. U.S. Centers for Disease Control and Prevention, *Morbidity and Mortality Weekly Report* **33**, 145, ISSN: 0149-2195, 1545-861X (1984).
105. D. E. Biggins, M. Y. Kosoy, *Journal of Mammalogy* **82**, 906–916, ISSN: 0022-2372, DOI 10.1644/1545-1542(2001)082<0906:IOIPON>2.0.CO;2 (2001).
106. D. E. Biggins, M. Y. Kosoy, *Journal of the Idaho Academy of Science and Engineering* **37**, 62–66 (2001).
107. S. N. Bevins *et al.*, *Integrative Zoology* **7**, 99–109, ISSN: 1749-4877, DOI 10.1111/j.1749-4877.2011.00277.x (2012).
108. J. R. Tucker *et al.*, *California compendium of plague control*, 2015, (2018; https://www.cdph.ca.gov/Programs/CID/DCDC/CDPH%20Document%20Library/CAPlagueCompendium.pdf).
109. K. L. Gage *et al.*, in *Proceedings of the Sixteenth Vertebrate Pest Conference*, pp. 200–206.
110. B. Thiagarajan *et al.*, *Journal of Wildlife Diseases* **44**, 731–736, ISSN: 0090-3558, DOI 10.7589/0090-3558-44.3.731 (2008).
111. R. D. Holt, M. Barfield, in *Spatial ecology*, ed. by S. Cantrell *et al.* (CRC Press, Boca Raton, Florida, 2010), pp. 189–211, ISBN: 1420059858.
112. P. W. Willeberg *et al.*, *American Journal of Epidemiology* **110**, 328–334, ISSN: 0002-9262, DOI 10.1093/oxfordjournals.aje.a112818 (1979).
113. C. U. Thomas, P. E. Hughes, *Journal of Wildlife Diseases* **28**, 610–613, ISSN: 0090-3558, DOI 10.7589/0090-3558-28.4.610 (1992).
114. N. W. Dyer, L. E. Huffman, *Journal of Wildlife Diseases* **35**, 600–602, ISSN: 0090-3558, DOI 10.7589/0090-3558-35.3.600 (1999).
115. A. Malmlov *et al.*, *Journal of Wildlife Diseases* **50**, 946–950, ISSN: 1943-3700, DOI 10.7589/2014-03-065 (2014).
116. B. R. Hoar *et al.*, *Preventive Veterinary Medicine* **56**, 299–311, ISSN: 0167-5877, DOI 10.1016/S0167-5877(02)00194-0 (2003).
117. R. J. Brinkerhoff *et al.*, *Vector-Borne and Zoonotic Diseases* **9**, 491–497, ISSN: 1557-7759, 1530-3667, DOI 10.1089/vbz.2008.0075 (2009).
118. L. A. Baeten *et al.*, *Journal of Wildlife Diseases* **49**, 932–939, ISSN: 0090-3558, DOI 10.7589/2013-02-040 (2013).
119. K. L. Wolff, B. W. Hudson, *Applied and Environmental Microbiology* **28**, 323–325, ISSN: 0003-6919 (1974).
120. J. C. Chandler *et al.*, *Journal of Clinical Microbiology* **56**, e00273–18, ISSN: 0095-1137, 1098-660X, DOI 10.1128/JCM.00273-18 (2018).
121. Core Science Analytics, Synthesis, & Libraries Program of the U.S. Geological Survey (USGS), *Biodiversity Information Serving Our Nation (BISON)*, Washington, D.C., 2012.
122. GBIF.org, *GBIF occurrence download*, Accessed: 1 February 2018, DOI 10.15468/dl.7dnihd.
123. VertNet, *Query: specificepithet:latrans genus:Canis mappable:1*, Accessed: 30 January 2018, (http://vertnet.org).
124. H. R. Pulliam, *Ecology Letters* **3**, 349–361, ISSN: 1461-0248, DOI 10.1046/j.1461-0248.2000.00143.x (2000).
125. C. Daly *et al.*, *International Journal of Climatology* **28**, 2031–2064, ISSN: 1097-0088, DOI 10.1002/joc.1688 (2008).
126. E. M. Hart, K. Bell, *prism: download data from the Oregon PRISM project*, R package version 0.0.6, DOI 10.5281/zenodo.33663, (http://github.com/ropensci/prism).

127. R Core Team, *R: a language and environment for statistical computing*, R Foundation for Statistical Computing (Vienna, Austria, 2018), (https://www.R-project.org/).
128. J. J. van Zyl, *Acta Astronautica* **48**, 559–565, ISSN: 0094-5765, DOI 10.1016/S0094-5765(01)00020-0 (2001).
129. R. J. Hijmans, *raster: geographic data analysis and modeling*, R package version 2.6-7, (https://CRAN.R-project.org/package=raster).
130. R. Pollitzer, World Health Organization, *Plague. WHO Monograph Series 22* (Geneva, Switzerland, 1954), ISBN: 9241400226.
131. M. Li *et al.*, *Emerging Infectious Diseases* **11**, 1494, ISSN: 1080-6059, 1080-6040, DOI 10.3201/eid1109.041147 (2005).
132. U.S. Census Bureau, *2010 census*, U.S. Department of Commerce, 2011.
133. C. Myron, *Elmore County child recovering after plague infection*, Central District Health Department (2018; http://www.cdhd.idaho.gov/pdfs/News/2018/06-12-18-human-plague-case.pdf).
134. S. E. Spielman *et al.*, *Applied Geography* **46**, 147–157, ISSN: 0143-6228, DOI 10.1016/j.apgeog.2013.11.002 (2014).
135. J. N. K. Rao, I. Molina, *Small area estimation* (John Wiley & Sons, Ltd., Hoboken, New Jersey, 2015), ISBN: 0471413747.
136. C. H. Alexander, *Survey Methodology* **28**, 35–42, ISSN: 0033-5207 (2002).
137. A. Navarro, presented at the Measuring People in Place Conference.
138. D. G. Horvitz, D. J. Thompson, *Journal of the American Statistical Association* **47**, 663–685, ISSN: 0162-1459, 1537-274X, DOI 10.1080/01621459.1952.10483446 (1952).
139. L. Li *et al.*, *Statistical Methods in Medical Research* **22**, 14–30, ISSN: 0962-2802, DOI 10.1177/0962280211403597 (2013).
140. S. Narduzzi *et al.*, *Epidemiologia e Prevenzione* **38**, 335–341, ISSN: 1120-9763, 1120-9763 (2014).
141. N. T. Hobbs, M. B. Hooten, *Bayesian models: a statistical primer for ecologists* (Princeton University Press, Princeton, New Jersey, 2015), ISBN: 1400866557, 1400874422.
142. C. K. Wikle, *International Statistical Review* **71**, 181–199, ISSN: 1751-5823, DOI 10.1111/j.1751-5823.2003.tb00192.x (2003).
143. J. A. Royle, R. M. Dorazio, *Hierarchical modeling and inference in ecology: the analysis of data from populations, metapopulations and communities* (Elsevier, Amsterdam, Netherlands, 2008), ISBN: 0123740977, DOI 10.1016/B978-0-12-374097-7.50001-5.
144. R. M. Nowak, *Walker's mammals of the world* (Johns Hopkins University Press, Baltimore, Maryland, 1999), vol. 1, ISBN: 0801857899.
145. V. Howard Jr, G. G. DelFrate, in *Great Plains Wildlife Damage Control Workshop Proceedings* (University of Nebraska-Lincoln, Lincoln, Nebraska, 1991), pp. 39–49.
146. A. M. Barnes, *Symposia of the Zoological Society of London* **50**, 237–270 (1982).
147. C. R. Smith *et al.*, *Journal of Vector Ecology* **35**, 1–12, ISSN: 1081-1710, DOI 10.1111/j.1948-7134.2010.00051.x (2010).
148. E. Bertherat *et al.*, *Emerging Infectious Diseases* **13**, 1459–1462, ISSN: 1080-6059, 1080-6040, DOI 10.3201/eid1310.070284 (2007).
149. P. Stapp *et al.*, *Frontiers in Ecology and the Environment* **2**, 235–240, ISSN: 1540-9295, DOI 10.1890/1540-9295(2004)002[0235:POEIPD]2.0.CO;2 (2004).
150. M. Bekoff, *Mammalian Species*, 1–9, ISSN: 0076-3519, DOI 10.2307/3503817 (1977).
151. S. A. Poessel *et al.*, *The Journal of Wildlife Management* **77**, 297–305, ISSN: 1937-2817, DOI 10.1002/jwmg.454 (2013).

152. D. L. Bounds, W. W. Shaw, *Natural Areas Journal* **14**, 280–284, ISSN: 0885-8608 (1994).
153. N. McClennon *et al.*, *American Midland Naturalist* **146**, 27–36, ISSN: 0003-0031, 1938-4238, DOI 10.1674/0003-0031(2001)146[0027:TEOSAA]2.0.CO;2 (2001).
154. S. D. Gehrt *et al.*, *Journal of Mammalogy* **90**, 1045–1057, ISSN: 0022-2372, DOI 10.1644/08-MAMM-A-277.1 (2009).
155. V. M. Lukasik, S. M. Alexander, *Cities and the Environment (CATE)* **4**, 8, ISSN: 1932-7048, DOI 10.15365/cate.4182011 (2011).
156. J. M. Fedriani *et al.*, *Ecography* **24**, 325–331, ISSN: 0906-7590, DOI 10.1034/j.1600-0587.2001.240310.x (2001).
157. M. Murray *et al.*, *Proceedings of the Royal Society B: Biological Sciences* **282**, 20150009, ISSN: 0962-8452, DOI 10.1098/rspb.2015.0009 (2015).
158. L. Gerardo-Giorda *et al.*, *Journal of the Royal Society, Interface* **10**, 20130418, ISSN: 1742-5662, 1742-5662, DOI 10.1098/rsif.2013.0418 (2013).
159. S. M. Nusser *et al.*, *Journal of Wildlife Management* **72**, 52–60, ISSN: 0022-541X, DOI 10.2193/2007-317 (2008).
160. U.S. Department of Agriculture, *2016 program data reports*, 2017, (2018; https://www.aphis.usda.gov/aphis/ourfocus/wildlifedamage/sa_reports/sa_pdrs/ct_pdr_home_2016).
161. L. Brotons *et al.*, *Ecography* **27**, 437–448, ISSN: 0906-7590, DOI 10.1111/j.0906-7590.2004.03764.x (2004).
162. T. Hastie, W. Fithian, *Ecography* **36**, 864–867, ISSN: 0906-7590, DOI 10.1111/j.1600-0587.2013.00321.x (2013).
163. A. M. Gormley *et al.*, *Journal of Applied Ecology* **48**, 25–34, ISSN: 1365-2664, DOI 10.1111/j.1365-2664.2010.01911.x (2011).
164. V. H. F. Gomes *et al.*, *Scientific Reports* **8**, 1003, ISSN: 2045-2322, DOI 10.1038/s41598-017-18927-1 (2018).
165. E. J. Pebesma, R. S. Bivand, *R News* **5**, 9–13, ISSN: 1609-3631, (https://CRAN.R-project.org/doc/Rnews/) (2005).
166. R. S. Bivand *et al.*, *Applied spatial data analysis with R* (Springer, New York, New York, ed. 2, 2013), ISBN: 1461476177, 1461476184, DOI 10.1007/978-1-4614-7618-4, (http://www.asdar-book.org/).
167. E. H. Boakes *et al.*, *PLoS Biology* **8**, e1000385, ISSN: 1545-7885, DOI 10.1371/journal.pbio.1000385 (2010).
168. W. Yang *et al.*, *Journal of Biogeography* **40**, 1415–1426, ISSN: 0305-0270, DOI 10.1111/jbi.12108 (2013).
169. J. Beck *et al.*, *Ecological Informatics* **19**, 10–15, ISSN: 1574-9541, DOI 10.1016/j.ecoinf.2013.11.002 (2014).
170. J. Troudet *et al.*, *Scientific Reports* **7**, 9132, ISSN: 2045-2322, DOI 10.1038/s41598-017-09084-6 (2017).
171. R. Kays, *Canis latrans*, 2018, (10.2305/IUCN.UK.2018-2.RLTS.T3745A103893556.en).
172. C. Daly *et al.*, *Climate Research* **22**, 99–113, ISSN: 1616-1572, 0936-577X, DOI 10.3354/cr022099 (2002).
173. C. Daly, presented at the Proceedings of the 13th American Meteorological Society Conference on Applied Climatology, pp. 177–180.
174. C. Daly *et al.*, presented at the Proceedings of the 14th American Meteorological Society Conference on Applied Climatology.

175. C. Daly *et al.*, *Journal of Applied Meteorology* **33**, 140–158, ISSN: 0021-8952, DOI 10.1175/1520-0450(1994)033<0140:ASTMFM>2.0.CO;2 (1994).
176. C. Daly *et al.*, *International Journal of Climatology* **23**, 1359–1381, ISSN: 1097-0088, DOI 10.1002/joc.937 (2003).
177. C. Daly, *International Journal of Climatology* **26**, 707–721, ISSN: 1097-0088, DOI 10.1002/joc.1322 (2006).
178. T. G. Farr *et al.*, *Reviews of Geophysics* **45**, ISSN: 1944-9208, DOI 10.1029/2005RG000183 (2007).
179. E. Rodriquez *et al.*, *Photogrammetric Engineering and Remote Sensing* **72**, 249–260, ISSN: 2374-8079, DOI 10.14358/PERS.72.3.249 (2006).
180. M. Mukul *et al.*, *Scientific Reports* **7**, 41672, ISSN: 2045-2322, DOI 10.1038/srep41672 (2017).
181. M. C. Chu, *Laboratory manual of plague diagnostic tests* (World Health Organization, Geneva, Switzerland, 2000).
182. T. J. Quan *et al.*, in *Diagnostic procedures for bacterial, mycotic, and parasitic infections*, ed. by A. Barlows, W. J. Hausler (American Public Health Association, Washington, D.C., ed. 6, 1981), pp. 723–745, ISBN: 0875530869.
183. B. W. Hudson *et al.*, *Epidemiology & Infection* **60**, 443–450, ISSN: 0950-2688, DOI 10.1017/S002217240002057X (1962).
184. J. C. C. Neale, B. N. Sacks, *Oikos* **94**, 236–249, ISSN: 0030-1299, DOI 10.1034/j.1600-0706.2001.940204.x (2001).
185. G. A. Feldhamer *et al.*, Eds., *Wild mammals of North America: biology, management, and conservation* (Johns Hopkins University Press, Baltimore, Maryland, ed. 2, 2003), ISBN: 0801874165.
186. S. Manson *et al.*, *IPUMS National Historical Geographic Information System: Version 13.0 nhgis0001*, Minneapolis, Minnesota, 2018, DOI 10.18128/D050.V13.0.
187. U.S. Census Bureau, *1960 census*, U.S. Department of Commerce, 1961.
188. U.S. Census Bureau, *1990 census*, U.S. Department of Commerce, 1991.

2.6 Appendices

2.6.1 Appendix A: Panels

Panel 1: World Health Organization case definition for plague surveillance

Disease characterized by rapid onset of fever, chills, headache, severe malaise, prostration, with

- Bubonic form: extreme painful swelling of lymph nodes (buboes)
- Pneumonic form: cough with blood-stained sputum, chest pain, difficult breathing

Note: Both forms can progress to a septicemic form with toxemia. Sepsis without evident buboes rarely occurs.

Laboratory criteria for diagnosis

- Isolation of *Yersinia pestis* in cultures from buboes, blood, cerebrospinal fluid or sputum or
- Passive hemagglutination (PHA) test, demonstrating an at least fourfold change in antibody titer, specific for F1 antigen of *Y. pestis*, as determined by the hemagglutination inhibition test (HI) in paired sera.

Case classification

- Suspected: A case compatible with the clinical description. May or may not be supported by laboratory finding of Gram stain negative bipolar coccobacilli in clinical material (bubo aspirate, sputum, tissue, blood).
- Probable: A suspected case with Positive direct fluorescent antibody (FA) test for *Y. pestis* in clinical specimen; or passive hemagglutination test, with antibody titer of at least 1:10, specific for the F1 antigen of *Y. pestis* as determined by the hemagglutination inhibition test (HI); or epidemiological link with a confirmed case.
- Confirmed: A suspected or probable case that is laboratory-confirmed.

Panel 2: Diagnostic tests for *Yersinia pestis* in wildlife

The following are four of the recognized diagnostic tests for *Yersinia pestis* infection

Passive hemagglutination Assay (PHA) & F1-inhibitation (PHI) test

- Antigen-antibody reactivity is visible qualitatively using F1 antigen coated on glutaraldehyde-fixed sheep red blood cells as a sensitizing antigen. If serum reacts positively, a PHI test is run to verify the specificity of the PHA test that calculates the titer.
- Positive Test: At least fourfold change in antibody titer, specific for F1 antigen of *Y. pestis*.
- Strength: Sensitive and generally reproducible.
- Limitations: Relies on unstable reagent, interpretation is subjective, and prone to nonspecific reactivity of natural antibodies.
- More detailed procedures found in (*181*).

F1 Luminex Plague Assay (F1-LPA)

- Semi-automated bead-based flow cytometric assay, specific for the F1 antigen of *Y. pestis*.
- Positive Test: With a baseline background noise of 250 mean fluorescent intensity (MFI), a Signal to Noise Ratio (S/N) $\geq$ 10 (or $\geq$ 2,500 MFI) is considered a positive test.
- Strength: More sensitive (x64) than PHA-PHI, fewer false negative results.
- Limitation: Has not been assessed for human tissue.
- More detailed procedures found in (*120*).

Fluorescent Antibody (FA)

- Smears of suspected tissue are prepared with plague antiserum and examined via fluorescent microscopy.
- Positive Test: Smear brightly fluoresce as "apple green-colored hollow rods" (Or less brightly if conducting a fluorescence inhibition test).
- Strength: Quick assay and requires a small amount of test material.
- Limitations: Sensitive with fresh tissue samples, not optimized for field collection.
- More detailed procedures found in (*182*, *183*).

Bacterial Culture

- Suspected tissue samples are suspended in blood agar and examined for growth daily for at least 7-10 days.
- Positive Test: Bacterial growth appears, typically a "stalactite"-type pattern or colonies about 4-7 mm in diameter.
- Strength: The gold-standard assay.
- Limitation: Time intensive. *Y. pestis* is slow growing.
- More detailed procedures found in (*182*).

Panel 3: Considerations for coyotes as a sentinel species for plague activity

Coyotes (*Canis latrans*) have been used as a sentinel species of plague (*Yersinia pestis*) activity in rodent populations of North America (*95*, *112–116*), but there are limitations for focusing on coyote-based plague surveillance data. I outline some notable limitations below:

- Coyotes are a wide-ranging species (*144*, *145*, *171*) and effectively sample a broad variety of habitats in which plague occurs and thus sample a considerable extent of the ecological niche of plague (*85*). However, coyotes are not observed homogeneously across North America (Chapter 5) and other species may be better indicators of plague activity in some habitats (e.g., rodent species for alpine regions of California). Therefore, without rodent data, I am estimating a the fundamental ecological niche of plague in coyotes, which I am considering as an approximation for the entire fundamental ecological niche of plague.
- The location where an individual coyote was truly exposed to *Y. pestis* is unknown and challenging to discern (*91*, *117*). Coyotes are mobile, encountering many rodents across their expansive individual home range (*144*, *145*, *184*) and whose plague antibodies can last for many months (*91*, *112*, *118*). Reinfection of coyotes is also probable and so a coyote can only indicate its most recent plague exposure. Rodent species have smaller individual home range sizes (*185*) and are the gold standard indicator of current or recent plague activity (*108*).
- Coyote specimens are not collected evenly across the United States and instead are collected opportunistically (*107*). Coyotes tested for plague exposure are primarily collected in conjunction with ongoing livestock/wildlife damage management operations conducted by the U.S. Department of Agriculture Animal and Plant Health Inspection Service Wildlife Services. Agricultural and urban areas may be more sampled for plague than other land-use types and their respective habitats. Including other species monitored for plague activity may help overcome this potential spatial sampling effort bias.

2.6.2 Appendix B: Tables

Table 1: Human plague cases in United States by state of exposure (1900–2017). (Unpublished data from Ken Gage, Ph.D. at the U.S. Centers for Disease Control and Prevention, Fort Collins, Colorado).

State	Number of Human Cases			
	Total	Pre-1950	Post-1949	Post-1983
Arizona	68	0	68	27
California	492	441	48	26
Colorado	67	0	67	50
Florida	10	10	0	0
Idaho	5	1	4	3
Illinois*	1	0	1	1
Louisiana	51	51	0	0
Maryland*	1	0	1	0
Michigan	1	1	0	0
Montana	2	0	2	2
Nevada	6	0	6	3
New Mexico	283	3	280	140
Oklahoma	2	0	2	2
Oregon	19	1	18	9
Texas	35	31	4	3
Utah	16	1	15	10
Washington	9	8	1	1
Wyoming	5	0	5	3
Unknown	2	1	1	1
Total	1,045	522	523	281

*Laboratory Acquired

Table 2: Summary of *Canis latrans* testing for antibodies against *Yersinia pestis* by California Department of Public Health (CPDH; 1983-2015), U.S. Department of Agriculture (USDA; 2005-2017) and U.S. Centers for Disease Control and Prevention (CDC; 1998-2008).

Agency	Plague Result	Count Tested	Count With Location	Prevalence
CDPH	Positive	705	704	8.8%
	Negative	7,414	7,251	
USDA	Positive	3,665	3,648	13.0%
	Negative	25,082	24,381	
CDC	Positive	2,516	2,360	
	Negative	Not digitized	Not digitized	

Table 3: Sources of independent coyote observations in the United States

Database	Coyote Observations
Angelo State Natural History Collections	100
California Academy of Sciences	59
California State University, Long Beach	2
Chicago Academy of Sciences	11
Cornell University Lab of Ornithology	42
Cornell University Museum of Vertebrates	36
Denver Museum of Nature & Science	83
Fort Hays Sternberg Museum of Natural History	70
Florida Museum of Natural History	120
Humboldt State University	1
iNaturalist.org	4,485
Illinois State University	3
James R. Slater Museum of Natural History	78
Louisiana State University Museum of Natural Science	43
Michigan State University Museum	91
Museum of Cultural and Natural History–Central Michigan University	4
Museum of Comparative Zoology, Harvard University	4
Museum of Southwestern Biology	458
Museum of Texas Tech University	250
Museum of Vertebrate Zoology	987
Natural History Museum of Geneva	3
Natural History Museum of Los Angeles County	166
naturgucker.de	11
New Mexico Museum of Natural History and Science	1
National Museum of Natural History, Smithsonian Institution	1
National Parks Service	207
Natural History Museum of Utah	186
New York State Museum	528
North Carolina Museum of Natural Sciences	86
Northern Michigan University	5
The Ohio State University Borror Lab of Bioacoustics	139
The Ohio State University Museum of Biological Diversity	363
Royal Ontario Museum	46
Sam Noble Oklahoma Museum of Natural History	556
Santa Barbara Museum of Natural History	37
Tall Timbers Research Station and Land Conservancy	1
Texas A&M University Biodiversity Research and Teaching Collections	44
United States Geological Survey (USGS)	30
USGS Western Ecological Research Center San Diego Field Station	2,899
University of Alaska Museum of the North	93
University of Arizona Museum of Natural History	31
University of California–Los Angeles Dickey Collection	48
University of California–Santa Barbara Marine Science Institute	62
University of Colorado Museum of Natural History	1
University of Kansas Biodiversity Institute	676
University of Michigan Museum of Zoology	257
University of Texas at El Paso Biodiversity Collections	409
University of Washington Burke Museum	90
University of Wyoming Museum of Vertebrates	9
Yale University Peabody Museum	59

Table 4: Oregon State University Parameter elevation Regression on Independent Slopes Model (PRISM) 30-Year Average Annual Normals (1981–2010). Modeled using a combination of a digital elevation model (DEM) and climatologically-aided interpolation (CAI)*. Variables are modeled at 30 arcseconds (∼800 meters) resolution and aggregated to 2.5 arcminutes (∼4 kilometers). See (*125*) for more details.

Variable	Units	Derivation
Precipitation	millimeters (mm)	Modeled; Summing monthly averages (rain + melted snow)
Maximum Temperature	°Celsius (°C)	Modeled; Averaging over all months using a DEM as the predictor grid
Mean Temperature	°Celsius (°C)	Derived; Average of Maximum Temperature and Minimum Temperature
Minimum Temperature	°Celsius (°C)	Modeled; Averaging over all months using a DEM as the predictor grid
Mean Dewpoint Temperature	°Celsius (°C)	Modeled; CAI used minimum temperature as the predictor grid
Maximum Vapor Pressure Deficit	hectopascal (hPA)	Modeled; CAI used mean dewpoint temperature and maximum temperature as the predictor grids
Minimum Vapor Pressure Deficit	hectopascal (hPA)	Modeled; CAI used mean dewpoint temperature and minimum temperature as the predictor grids

*Accuracy of these data is based on the original specification of the Defense Mapping Agency (DMA) one-degree digital elevation models (DEM). The stated accuracy of the original DEMs is 130-meter circular error with 90% probability. Datasets use all weather stations, regardless of time of observation.

Table 5: Missingness of coyote specimens by U.S. Department of Agriculture (USDA; 2005–2017) and California Department of Public Health (CDPH; 1983-2015). Specimens are considered missing if textual geographic location information is unavailable or indiscernible to geocode. Fewer than 2.5% of specimens are missing and ignored in the analysis.

State	Missing	Total
Arizona	33	3,097
California*	164	8,199
Colorado	3	2,161
Idaho	1	576
Kansas	2	240
Montana	43	6,272
Nebraska	5	586
New Mexico	42	5,998
Nevada	9	3,269
Oklahoma	12	1,177
Oregon	3	543
South Dakota	5	521
Texas	16	1,084
Utah	3	487
Washington†	540	590
Wyoming	0	1,224

*Missing specimens are located in 28 counties (∼48% of counties in California). Over 30% of missing specimens are located in Santa Clara County, California and San Luis Obispo County, California.
†All missing specimens are located in Yakima County, Washington.

2.6.3 Appendix C: Figures

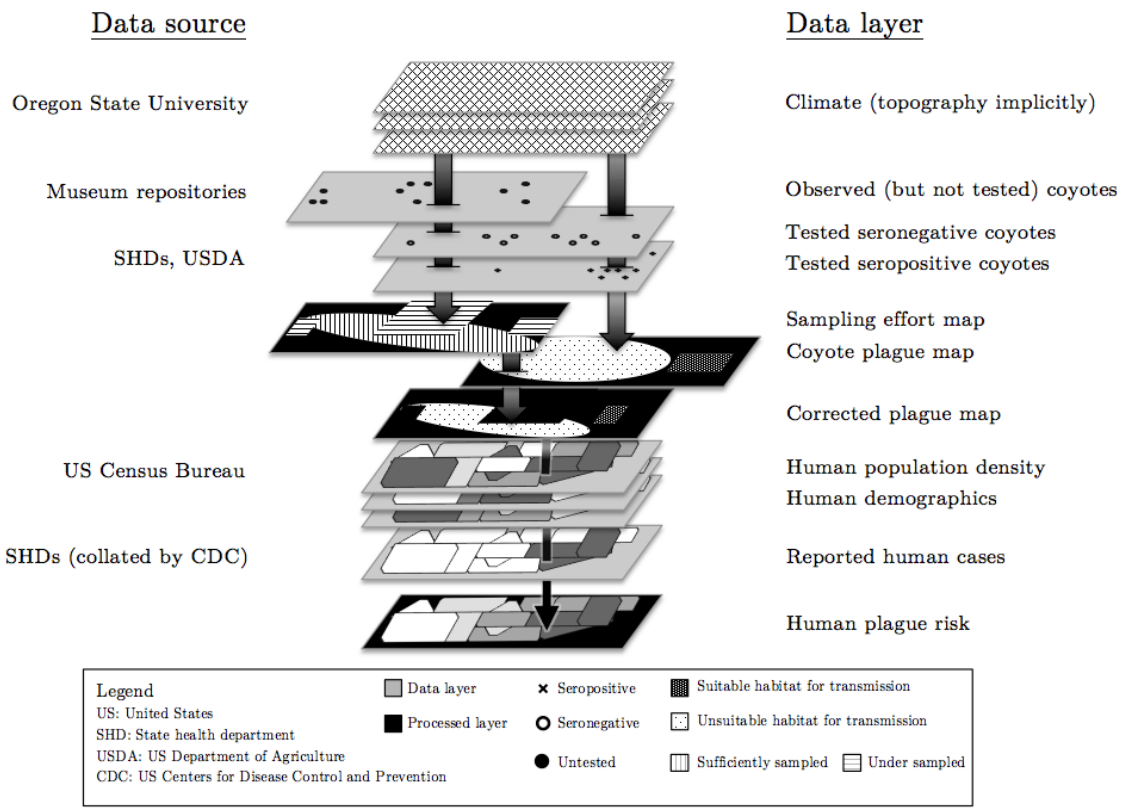
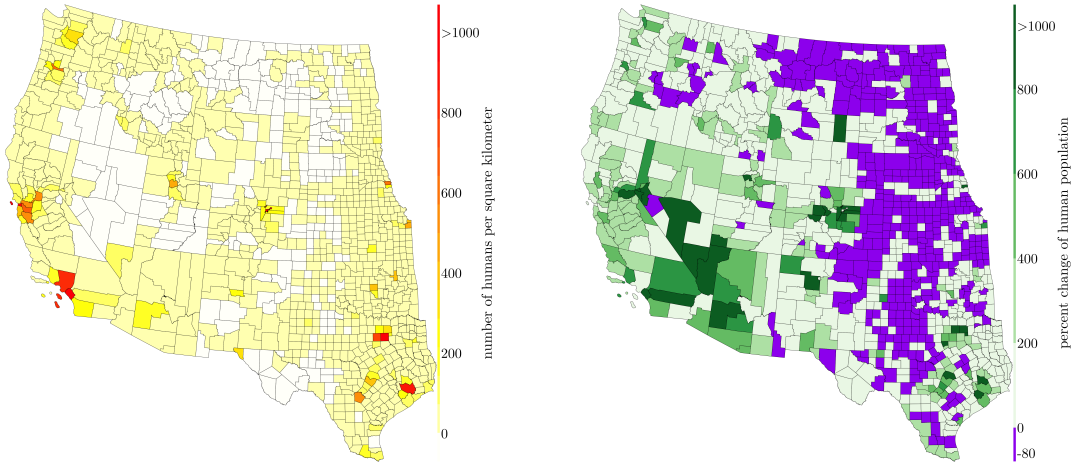


Figure 1: Data layers of dissertation and their sources. Data are structured from point locations to polygons and raster grids. Four layers are predicted (black color) using data layers (grey color). Arrows represent the layers involved for each predicted layer.

(a) **(b)**

Figure 2: **(a)** Human plague cases in contiguous United States since introduction (1900-2017). The early period (1900–1930) contained outbreaks in port cities (e.g., San Francisco, CA and Los Angeles, CA), but by 1950, plague became locally enzootic in the western United States and human cases occurred in the country interior only. **(b)** Human plague cases in western United States (1950-2017) by county of exposure. Plague spread from California to Kansas by the 1940s (*86*). Since 1950, human plague cases have occurred in 13 western states (521 total cases) where over 80% of these cases have occurred in the Four Corner States (Arizona, Colorado, New Mexico, and Utah;(*32*, *33*, *103*)). Data are unpublished and courtesy of Kenneth Gage, Ph.D. at the U.S. Centers for Disease Control and Prevention, Fort Collins, Colorado.



(a) **(b)**

Figure 3: **(a)** Human population density of the western United States (2010) per square kilometer by county (*132*). **(b)** Percent change of human population in the western United States 1950 to 2010 by county (*132*, *186*). Counties without 1950 census data ($n = 3$) are given earliest population estimate. Broomfield County, Colorado was created in 2001 from Boulder County, Colorado but existed as a town since 1961 ($n = 4,535$ people; *187*). Cibola County, New Mexico was created in 1981 from Valencia County, New Mexico and its first census was in 1990 ($n = 23,794$ people; *188*). La Paz County, Arizona was created in 1983 from Yuma County, Arizona and its first census was in 1990 ($n = 13,844$ people; *188*).

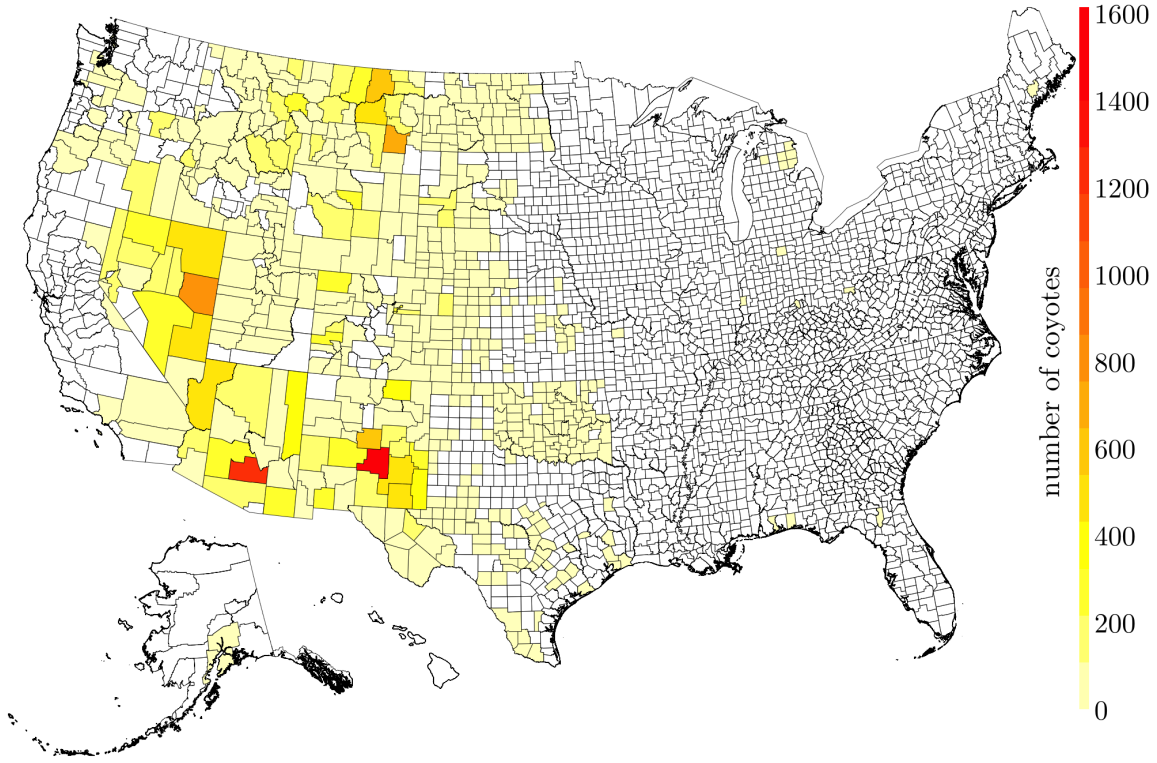


Figure 4: Number of coyotes tested by U.S. Department of Agriculture Animal and Plant Health Inspection Service National Wildlife Disease Program for antibodies against *Yersinia pestis* (2005–2017). Sampling is heterogeneous across the United States, primarily in historic plague enzootic areas. Data for coyotes from California are managed by the California Department of Public Health and are not included with these data.

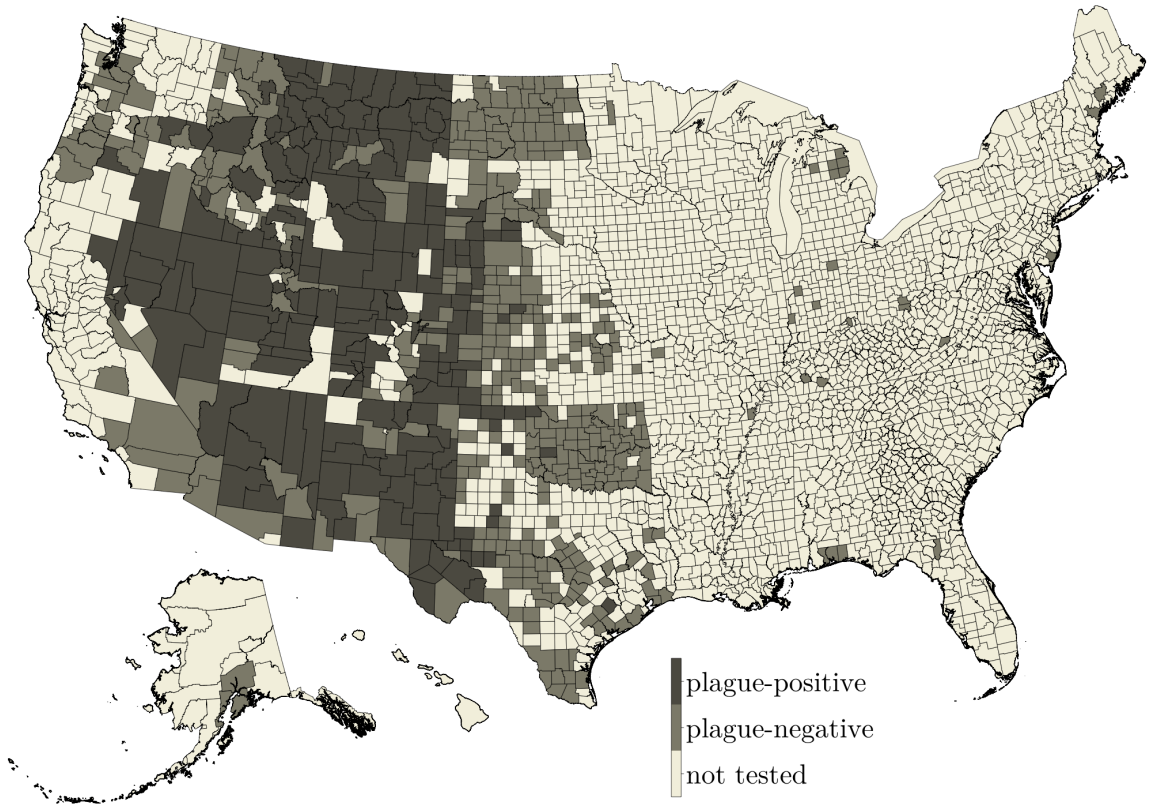


Figure 5: Results of coyotes tested by U.S. Department of Agriculture Animal and Plant Health Inspection Service National Wildlife Disease Program for antibodies against *Yersinia pestis* (2005–2017). A coyote with plague antibodies is not observed in every county sampled. Counties with a coyote that tested plague-positive are found throughout the western region of the United States. A figure of crude rates is not included.

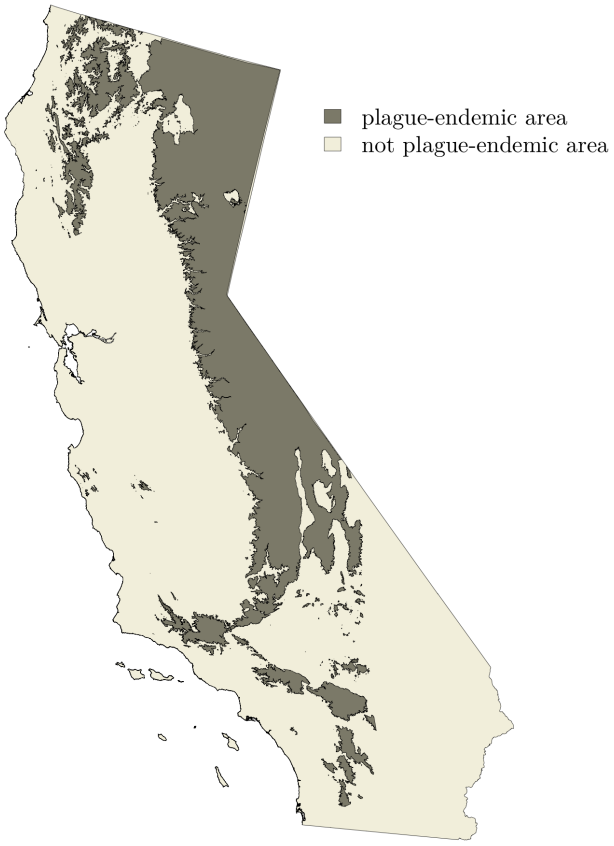
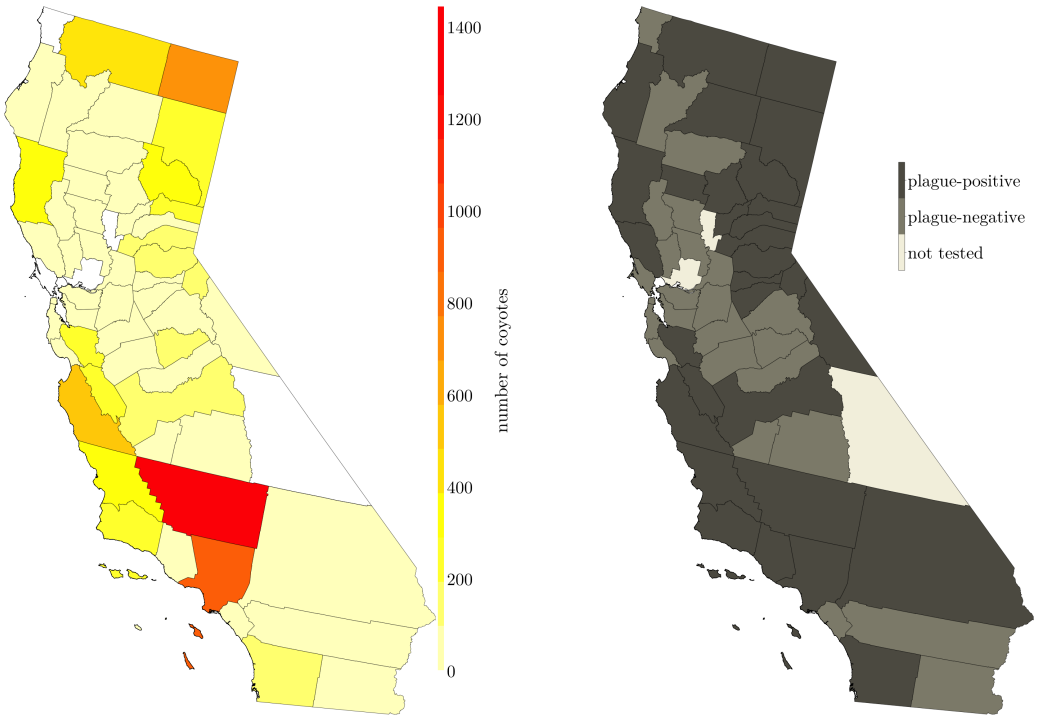


Figure 6: Plague-enzootic regions of California. Area primarily based on rodent plague surveillance and historical knowledge by California Department of Public Health. Area bounded by a minimum elevation of 4,000 feet. See (*108*) for more details.



(a) **(b)**

Figure 7: **(a)** Number of coyotes tested by California Department of Public Health (CDPH) for antibodies against *Yersinia pestis* (1983–2015). Sampling is heterogeneous across the state of California, primarily in historic plague enzootic areas and areas of high concern. Limited sampling in the Mojave and Sonoran Deserts. **(b)** Results of coyotes tested by CDPH for antibodies against *Y. pestis* (1983–2015). A coyote with plague antibodies is not observed in every county sampled. Only 2% of observations are unable to be geolocated and are ignored in the analysis.

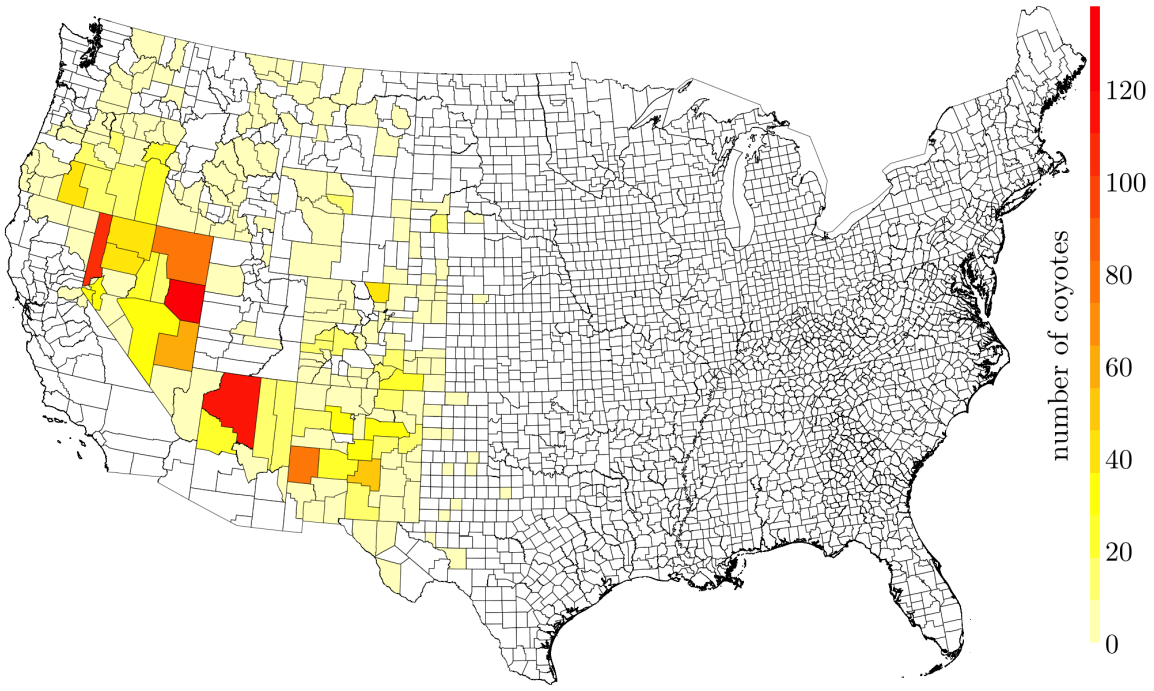


Figure 8: Number of coyotes sampled by state health departments and tested by the U.S. Centers for Disease Control and Prevention for antibodies against *Yersinia pestis* (1998–2008). All coyotes in this data set tested positive for plague antibodies because coyotes that tested negative were not electronically available. See (*85*) for more details. Possible duplicate coyote specimens between the USDA and the CDC were removed from the CDC data and are not presented here.

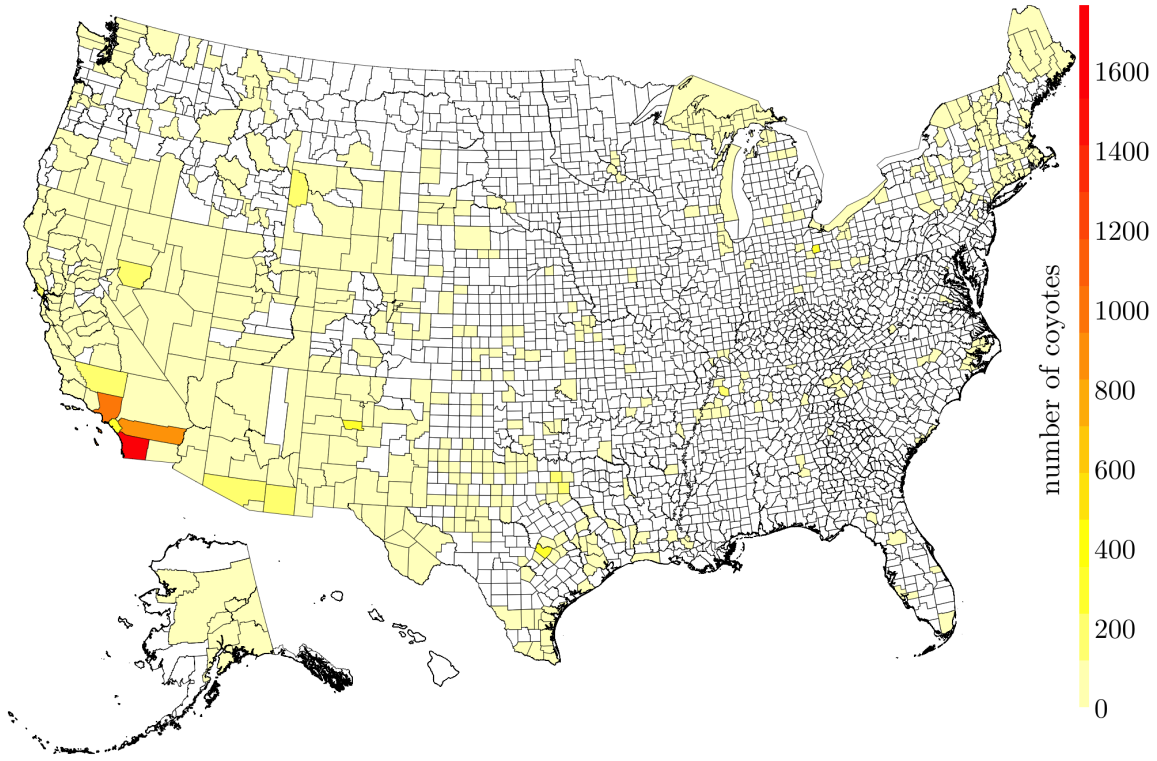


Figure 9: Number of coyote observations with geolocation information collated by various biodiversity repositories (1900–2017). These coyotes were not tested for antibodies against *Yersinia pestis*. Coyotes are found throughout the United States, including Alaska, with higher observations recorded in Southern California and the western United States, in general. See (*121–123*) for more details.

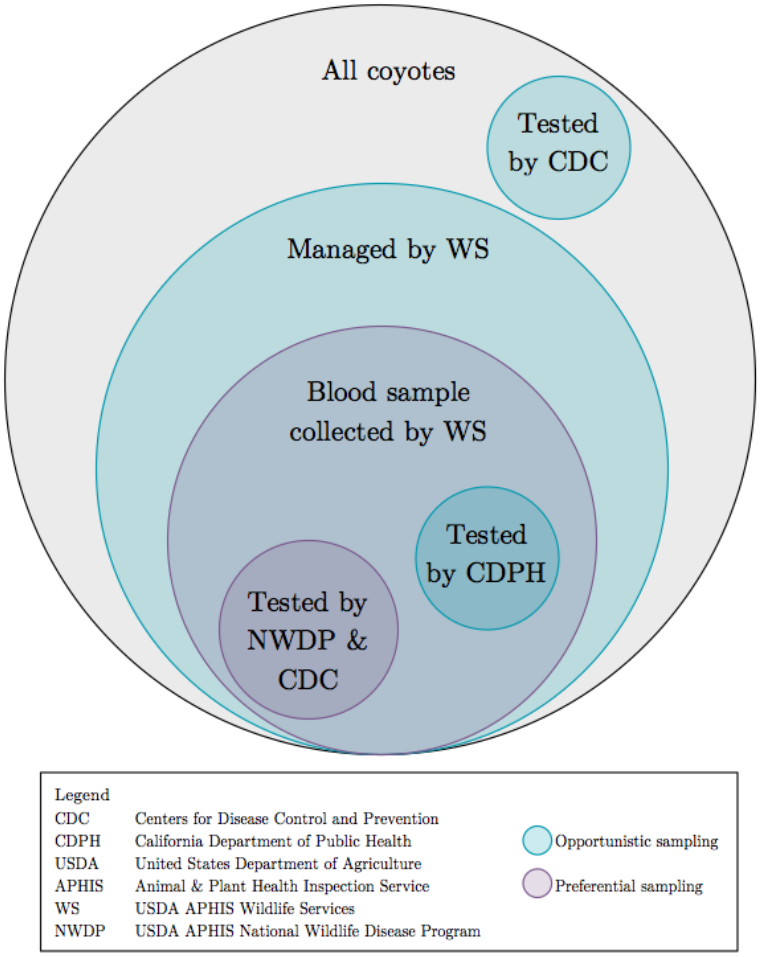


Figure 10: The sources of coyote specimens used in the dissertation as well as their sampling strategy. Coyotes tested for the presence of *Yersinia pestis* antibodies are not collected from the field randomly. Instead, coyotes are primarily collected from ongoing livestock or wildlife damage management activities (i.e., opportunistic sampling) by the U.S. Department of Agriculture Animal and Plant Health Inspection Service Wildlife Services (WS). In addition, only a subset of these collected specimens are tested for plague, namely from plague enzootic areas (i.e., preferential sampling) of the western United States. The coyotes tested by the California Department of Public Health are collected by WS in California. The coyotes tested by the U.S. Centers for Disease Control and Prevention are submitted by state health departments.

3 Examining spatial patterns in principal component space to identify suitable habitat for enzootic plague transmission in California, U.S.A.

"Nature's dice are always loaded; that in her heaps and rubbish are concealed sure and useful results."

- *Nature* by Ralph Waldo Emerson (*189*)

3.1 Introduction

Emerging infectious diseases (EIDs) are a significant threat to public health (*190*). About 60% of EIDs originate in animals (a.k.a., zoonotic diseases) (*5*, *191*), primarily from wildlife (*5*, *192*). Early detection of spillover events into human populations is essential for timely response (*193*), yet these events are rare (*194*) and can be difficult to monitor (*159*), creating challenges for zoonotic disease surveillance systems. The presence of known animal reservoir hosts that can act as potential sources of infection that increases the chance of observing a zoonotic disease outbreak in human populations (*195*). Knowledge of the geographic distribution of reservoir hosts can enhance zoonotic disease surveillance via identifying areas where humans interact with these hosts. This chapter outlines a novel method to use the geographic distribution of infection within an animal host to define areas of higher human risk (see Chapter 6).

Plague is a zoonotic infectious disease caused by the gram-negative bacterium *Yersinia pestis*. Humans are highly susceptible to infection and can experience severe symptoms. Mortality is high (40%–90%) for plague without prompt antibiotic treatment (16% mortality with treatment; *33*). The most common route of transmission to humans in the United States is via the bite of an infected rodent flea (*33*). Plague arrived in California in 1900 and quickly established in native rodent populations, and then spread eastward to approximately the 100th Meridian by 1960 (*102*). Human plague risk has been estimated for select regions of the United States using human case location information, frequency of infections, and environmental factors (*59*, *71–73*). However, the vast majority of human cases since 1960

was acquired via fleabite or direct contact with an infectious animal, with only one case via human-to-human transmission documented in the United States since 1924 (*48*). Human plague infections frequently are linked to recent or ongoing epizootic events (i.e., disease outbreak in animal populations; *57*). Therefore, geographic surveillance of zoonotic disease activity provides a strategy for identifying areas that present a high risk for human exposure.

Since human disease data are sparse, wildlife disease data have been used to predict the potential spatial distribution of plague in parts of the United States (*70*, *85*, *89*, *90*). Previous investigations employed spatial statistical methods that used the location of wildlife plague cases and environmental factors to predict the potential spatial distribution of plague. These *ecological niche models* (ENMs) increasingly have been used to identify the environmental conditions associated with the occurrence of a pathogen and then predict the spatial distribution of pathogens and their hosts (*20*), especially arthropod vector host species. ENMs are appealing for estimating disease risk as a way to circumvent data limitations of surveillance systems where information about cases (positives, presences) is available, but controls (negatives, absences) are lacking (*12*). When data are available, methods using both case and control data are mostly preferred to case-only analyses (*162–164*) because sampling effort can be accounted for in the predictions (*161*).

Here, this chapter develops and evaluates an analytic strategy using a set of both case and control data of coyotes (*Canis latrans* Say, 1823) screened for exposure to plague, coupled with long-term, landscape-level climatological variables (Figure 11). This approach uses dimension reduction (via principal component analysis) of the climate variables to identify signature patterns associated with high enzootic plague risk. More specifically, my proposed method adapts a spatial cluster algorithm originally designed for spatial cancer epidemiology to identify spatial clusters of climate signatures *within the principal component space*. My proposed method offers a mechanism of environmental interpolation (*19*) to predict areas of disease occurrence that match statistically the environmental conditions where a disease was observed, even in areas not historically sampled or from which data were unavailable. Statistical inference can be drawn from the resulting prediction regarding the spatial distribution of sylvatic plague (i.e., cycling in rodent populations) for California via a species (coyotes) which is not a reservoir but functions as a wide-ranging sentinel for

plague activity. Regions presenting conditions conducive to the enzootic maintenance (i.e., endemicy) of plague are likely the location of human and animal exposure to *Y. pestis*-carrying hosts and vectors (*94*, *196*). I compare the results of my proposed method to previous predictions of plague risk in California (*89*). Definitions for key terms in this chapter appear in Panel 4.

3.2 Data and methods

3.2.1 Wildlife disease data

Beginning in 1983 the California Department of Public Health (CDPH) Vector-Borne Disease Section digitized information for multiple animal species during animal-based plague surveillance across California. In addition to active plague surveillance of rodents and epidemiological investigations in response to human plague cases, the CDPH conducts passive surveillance of plague in non-rodent species, primarily coyotes (*108*). The CDPH uses coyote observations as sentinels for rodent-based surveillance in regions of California with lower access or resources as well as to corroborate regional increases in plague activity indicated by rodent plague surveillance data (but see Panel 5). Between 1983 and 2015, the CDPH screened 8,119 coyote blood samples for *Y. pestis* antibodies. An individual is considered a case if it tests positive for plague exposure (i.e., seropositive) and is considered a control if it tests negative for plague exposure (i.e., seronegative) via direct florescent antibody (FA) tests (*182*, *183*).

The CDPH tests coyote specimens for plague in conjunction with the U.S. Department of Agriculture (USDA) Animal and Plant Health Inspection Service (APHIS) Wildlife Services (WS). Therefore, specimens are collected for purposes other than disease surveillance (e.g., livestock and wildlife damage management) in predominantly agricultural/rural areas of California. Additionally, blood samples are not collected from all coyotes taken during livestock/wildlife damage management operations. Instead, WS specialists are requested to collect coyote blood samples from areas of suspected plague risk (e.g., historical case locations; *107*).

The vast majority of coyotes do not have precise accompanying sampling location in-

formation. I geocode specimens without geographic coordinates recorded in the field using location descriptions provided by the collector, county data, and base maps. Not all tested specimens have an identifiable location below the county level. I am unable to identify the location of 164 coyotes (2% of tested coyotes, $n = 1$ seropositive) and I ignore these specimens in subsequent analysis. Of all 7,955 coyotes with location information, about 8.8% ($n = 704$) are seropositive for plague exposure in California. Figure 12a presents county-level CDPH coyote sampling and the counties with at least one coyote observed with plague antibodies appears in Figure 12b. I present data at the county level to protect landowner privacy.

3.2.2 Environmental data processing

Extrinsic, abiotic factors influence environmentally-mediated infectious diseases, such as vector-borne diseases. These factors can directly affect the rate of transmission between hosts or indirectly affect either pathogen-host interactions, the population dynamics (e.g., survival) of hosts, or both (*20*). Abiotic factors include temperature and humidity at long-term temporal scales and at various spatial scales. Although the relationship between plague and climate is complex (*58*), past studies have established links between *Y. pestis* occurrence and climate factors (*70*, *85*, *89*, *90*).

To summarize climatic variation across the western United States, I employ the Oregon State University PRISM (Parameter elevation Regression on Independent Slopes Model) statistical mapping system. PRISM follows a weighted regression framework that relies on digital elevation models built from a network of ground measurements (*125*). I select the 30-year average normals (1981–2010) at a 2.5 arcminute (∼4 kilometer by ∼4 kilometer; ∼16 square kilometers) resolution. The variables used in my analysis appear in Table 6. The temporal range of the data mostly overlap with CDPH plague surveillance (1983–2015). I obtain PRISM data via the **prism** package (*126*) in the statistical software R (*127*).

Environmental data sets often contain many different measurements, some highly correlated (e.g., elevation and precipitation), so I avoid collinearity by reducing the number of variables in my proposed method by conducting a principal component analysis (PCA). A PCA transforms multiple predictors into a set of linearly uncorrelated variables called prin-

cipal components. I range standardize PRISM variables using the Gower metric (*197*) as the inputs in the PCA analysis because PRISM variables are measured on different scales. The first two principal components (PC1 and PC2) of the PRISM variables account for over 96% of the total variance across the United States (Table 7) and I use PC1 and PC2 as the environmental covariates for my analysis (Figure 13). I use the **RStoolbox** package (*198*) in the statistical software R (*127*) to perform a PCA.

3.2.3 Spatial analysis in the principal component space

ENMs provide key information in the study of relationships between species and factors that support or limit the existence or fitness (i.e., survivorship, fecundity) of a species. These factors are then located in a geographical area of interest to assess where a species may occur. *Habitat* is difficult to define, but here I consider the combination of environmental characteristics that describe a particular location (*199*). Certain species are better adapted to and have higher fitness in certain habitats (*200*). The response of a species to the combination of abiotic conditions (e.g., climate) is called a *fundamental ecological niche* (*201*, *202*). Biotic influences on a species fitness (e.g., competition) limit the fundamental ecological niche of a species into a smaller *realized ecological niche* (*201*, *203*). The concept of the ecological niche is an intensely debated but foundational principle in ecology (*11*, *201*, *204*). Here, I use the *Grinnellian* definition of an ecological niche (*11*) represented by an n-dimensional hypervolume (*201*, *205*) of only abiotic variables.

Once a species' ecological niche is determined based on ecological data (e.g., environmental measurements related to case-control locations), a set of other areas where the species niche are located is often predicted. This predicted spatial distribution can be interpreted as locations where a species may already exist but were not previously observed or could potentially exist if introduced to these areas (i.e., invasion or emergence). However, this method of prediction can be controversial because there could be non-ecological factors limiting the spatial distribution of a species (e.g., a geographic boundary; *12*). This limitation is not concerning for subsequent analysis because plague has been endemic in California since the 1930s (*102*).

A common consideration across both epidemiology and ecology studies that aim to pre-

dict the spatial distribution of a disease (species) is to account for the underlying population dispersion by comparing the seropositive case (presence) locations to the at-risk seronegative control (absence) locations. Logistic regression or other forms of generalized liner models (*206*) are standard statistical approaches for these studies. However, a nonparametric approach is often more appropriate to capture the nonlinear relationships between disease (species) occurrence and environmental variables (*207*). I propose a flexible, robust, nonparametric approach for environmental interpolation that uses both case and control location information in the form of a nonparametric multiplicative regression (NPMR; *208*, *209*).

Here, I propose a geographical epidemiology model as an ENM in order to predict the risk of a pathogen across an area of interest. Detecting disease clusters is a foundational exercise in spatial epidemiology and the development of cluster detection methods remains an active area of biostatistical and epidemiological research (*210*). A tool for describing spatial variations in disease risk is using a relative risk surface (*211–216*) (i.e., the ratio of estimated spatial density of cases to that of controls) to interpolate (predict) the risk of disease across an area of interest. These approaches typically treat the observed point locations of cases and controls as realizations of spatial Poisson processes. The kernel-based estimation of relative risk by Kelsall & Diggle (*211*, *212*) is expanded to include predictors because predicting risk solely on spatial interpolation is prone to sampling bias effects and may predict low risk for areas with no data (*20*). Environmental interpolation is an alternative approach to predict disease risk (*19*) because it incorporates the background heterogeneity of environmental variables into the prediction. The spatial distribution of a species is an environmental interpolation using its ecological niche.

Spatial interpolation, the relative risk surface Consider two sets of n bivariate mutually independent observations $X_1, X_2, ..., X_n$ and $Y_1, Y_2, ..., Y_n$ drawn from unknown spatial densities f and g, respectively, each arising as a partial realization of a Poisson process at a location i across the study area. These observations are made in geographical space and represent coordinates of disease-positive case (presence) and disease-negative

control (absence) locations, respectively.

Using kernel smoothing, the unknown density surfaces f and g can be estimated by $\hat{f}$ and $\hat{g}$, respectively, given by:

$$\hat{f}(z) = \frac{1}{n}\sum_{i=1}^{n} h_i^{-2} K\left(\frac{z - X_i}{h_i}\right)$$

$$\hat{g}(z) = \frac{1}{n}\sum_{i=1}^{n} h_i^{-2} K\left(\frac{z - Y_i}{h_i}\right),$$

where K is a kernel function and h_i is the smoothing parameter (bandwidth) for the ith observation. It is generally accepted that the choice of kernel function is not as critical (*217*) as the value of the smoothing parameter so the kernel is typically a radially symmetric probability density function (*212*). The smoothing parameter must be chosen more carefully because its value can have a large effect on the resulting estimate and is discussed in further detail below. Of note, the same smoothing parameter is used for both density surfaces of cases and controls in order to provide a fair comparison for the log relative risk surface (*218*). The relative risk surface (*215*, *216*) is calculated by:

$$\rho(z) = \frac{f(z)}{g(z)},$$

but estimated nonparametrically by their respective kernel estimates. In practice, the estimate is often stabilized by symmetrizing the two densities (*211*, *212*) and I define the log relative risk surface as:

$$\hat{\rho}(z) = ln\left(\frac{\hat{f}(z)}{\hat{g}(z)}\right).$$

The Grinnellian definition of an ecological niche (*11*, *201*, *205*) can be interpreted as areas in "predictor space" (a.k.a., "environmental space" of n-dimension environmental predictors) that are more optimal for a species fitness than other areas, or relative clustering. Of interest is the determination of where disease-positive case (presence) locations and disease-negative control (absence) locations appear to be more clustered in the log relative risk surface than expected under spatial constant risk. When identifying significantly differ-

ent fluctuations in the log relative risk surface it is often desirable to discern whether or not a given peak (or valley) in the estimated relative risk surface represents significantly heightened (lowered) risk or arises by random variation. Originally proposed by Kelsall & Diggle (*212*) and expanded in Waller (*218*), this can be approached as a two-tailed hypothesis test because areas with significantly high and low log relative risk are of interest.

$$H_0 : \rho(z) = 0$$

$$H_A : \rho(z) \neq 0$$

Pointwise significance levels can be calculated for the log relative risk surface ρ using Monte Carlo (MC) permutations for fixed bandwidth risk surfaces comparing the observed log relative risk surface to one corresponding to a null hypothesis of uniform risk (*212*). However, this approach is computationally expensive and prone to overestimating risk hotspots in areas with little to no data (*219*). An alternative framework uses asymptotic normality of a kernel density estimate (*220*) and asymptotic approximations for the variances of the risk surfaces, which are then interpretable with respect to a standard normal distribution. This framework is incorporated for risk surfaces (fixed: *219*; adaptive: *221*) and is both computationally less expensive and more stable in areas with sparse data than the MC method. Only local test statistics are calculated, because a good global test statistic with an asymptotic normality framework is currently unavailable (see *219*, *221*, for more details). Here, two-tailed asymptotic tolerances are calculated at two significance levels ($\alpha = 0.05$ & $\alpha = 0.01$).

My proposed approach involves three steps: 1) data rearrangement and dimension reduction, 2) spatial cluster detection of the ecological niche *within* the principal component predictor space, and 3) back-transformation and prediction of the detected ecological niche in geographic space. First, instead of calculating relative clustering of cases and controls by their geospatial positions, I calculate the relative clustering of cases and controls by their position in a combination of predetermined principal components (i.e., instead of longitude and latitude, arrange by PC1 and PC2). After rearranging case locations and control locations by their predictor values, I estimate the relative risk of cases and controls using the

Kelsall & Diggle (*211*, *212*) method to determine areas of "predictor space" that are more likely to observe a case versus more likely to observe a control. The relative risk clustering of "cases" is the ecological niche of the disease in predictor space. All pixels in the geographic study region correspond to locations in predictor space. Once the ecological niche of the disease is identified by the clustering of high relative risk within predictor space, I identify all pixels falling within the significantly elevated relative risk surface in predictor space and back transform to their respective location in geographic space. The spatial clustering in predictor space identifies all other areas where this ecological niche may be located in geographic space (but may not have been previously monitored). This rearrangement is possible because spatially-defined predictor data are available for areas for which prediction is desired. After arranging the prediction data set by the same principal components used to identify the ecological niche of a disease, the locations of the prediction data set that fall within the relative clustering of cases are more likely to have the ecological conditions for the disease than the locations of the prediction data set that fall within the relative clustering of controls. My proposed approach is summarized in Figure 14.

I conduct model performance and statistical inference on the predicted spatial distribution of plague in California. Predictive model performance is evaluated by n-fold cross-validation. Controls are randomly undersampled to balance plague prevalence (0.5) of each iteration. I calculate the average Area Under the Receiver Operating Characteristic Curve (AUC) and a 95% confidence interval across iterations (*222*) as well as a precision-recall curve (*223*) to evaluate the prediction. I draw statistical inference from the log relative risk and significance level surfaces when predicted into geographic space by determining associations between predicted values and the PRISM 30-year climate annual average normals across the study region. After values are resampled at each spatial pixel (i.e., grid cell), I estimate smooth univariate trends using generalized additive models (*224*) for both the predicted log relative risk and significance level values. Additionally, I compare the difference in mean PRISM values between suitable and unsuitable habitat using two-tailed Student's t-tests.

Important considerations There are two primary considerations for my proposed method. First, choosing the smoothing parameter h_i is a central issue for kernel density estimation because it has a large influence on the estimation (*225*). The smoothing parameter (bandwidth) controls how much averaging is occurring in predictor space. If the bandwidth is small, the density estimator function will be under smoothed with high variability, but if the bandwidth is large, the density estimator function will be over smoothed likely not representative of the true pattern in the data. Finding an optimal bandwidth size is desirable for kernel density estimation-based models. In predictor space, an ecological niche appears as a convex shape (*12*). A smaller smoothing parameter is more likely able to detect small changes in the shape (i.e., edges) of an ecological niche but creates a statistically uneven surface that may lead to the main body of the ecological niche to be inaccurately detected. On the contrary, a large smoothing parameter will detect the main body of the ecological niche but will not be as precise around the edges.

My proposed method approaches this crucial issue in multiple ways. Cases and controls may have differing sample sizes, especially for rare diseases where disease-negative controls are observed in higher frequency than cases. However, the same smoothing parameter is used for the density surfaces of cases and controls in order to avoid variations in the spatial flexibility of the relative risk surface due solely to different levels of smoothness induced by different bandwidths (*211*, *212*, *218*). The smoothing parameter can be fixed (*211*, *212*) or adaptive (*226*) and can be chosen by expert judgment (statistical and ecological) or computationally using various methods (*211*, *212*, *227–229*). Here, I use a fixed-bandwidth kernel density estimator for the log relative risk surface with a smoothing parameter chosen with the maximal smoothing principle (*229*). The utility of an adaptive-bandwidth kernel density estimator for ecological niche modeling represents a promising area for future study. Additionally, I conduct a sensitivity analysis comparing the smoothing parameter chosen via the maximal smoothing principle ($h_i = 0.052$ coefficient units) to a value half (0.026 coefficient units) and a value double (0.104 coefficient units) the chosen value.

Second, my proposed method adjusts for edge effects within the principal component predictor space. First, I use an adjusted density estimator (*212*, *230*) to correct for edge effects. Second, in predictor space, the spatial density surfaces for cases ($f(z)$) and controls

($g(z)$) extend beyond areas of predictor space where observations are located, area A. Here, ratio of $f(z)$ and $g(z)$ is of interest only in the area A because the log relative risk surface outside of area A tends to be distorted by edge effects. I use a conservative edge correction that limits my analysis within the area A. More specifically, I calculate a concave hull around the observed data in predictor space (area A) with a very small buffer ($\sim$1/10 bandwidth) to include all observed data within the area A. Any relative risk estimates outside of area A are labeled as "sparse data" and are translated to locations of the study area presenting environmental conditions where no observation data are recorded. This conservative method may exclude habitats around the edges of area A that may be in the ecological niche of a disease.

3.3 Results

Statistically significant relative clustering of seronegative coyotes (controls; $n = 7251$) and seropositive coyotes (cases; $n = 704$) by the CDPH (1983–2015) was detected using the kernel-density estimation based approach in predictor space (bandwidth = 0.052 coefficient units). The estimated intensity surfaces of cases and controls in predictor space appear in Figure 15 and calculated asymptotic tolerances in Figure 16 reveal evidence of statistically significant relative clustering of cases and controls.

Spatial interpolation of the estimated log relative risk (Figure 17) and significance level surfaces (Figure 18) revealed spatial patterns of plague occurrence in California. Positive log relative risk (cases) was predicted in the Sierra Nevada, Transverse Ranges, and Peninsular Ranges as well as locations within the Klamath Mountains, Cascade Range, and Modoc Plateau. Negative log relative risk (controls) was predicted in the Coastal Ranges, Central Valley, and desert regions, including the Mojave, Sonoran, and Colorado Deserts. Sparse data were available for a large portion of eastern California, especially the Basin and Range regions, and coastal areas of the Channel Islands and Humboldt County as well as high alpine zones of the Sierra Nevada. Areas my proposed method deemed too similar to distinguish between plague-suitable habitat and plague-unsuitable habitat were predicted in mountainous regions as well as the central coastline, the western Mojave Desert, and regions of the Klamath Mountains(Figure 18). The prediction of log relative risk across

California had an average Area Under the Receiver Operating Characteristic Curve (AUC) of 0.846 (95% CI: 0.825–0.867) and an acceptable precision-recall curve after a 25-fold cross-validation (Figure 19). Areas of California with sparse observation data were predicted when my proposed method was expanded beyond the area of predictor space with observations (Figure 20). After back-transformation, additional geographic areas of California with log relative risk values (Figure 21) and significance level values (Figure 22) were predicted, namely the eastern Sierra Nevada, Mojave Desert, and coastal regions. The model could not predict into the Sonoran Desert and Death Valley.

The ecological niche in predictor space was generally the same across the three smoothing parameters with a greater smoothing parameter value associated with a smoother surface and a smaller smoothing parameter value with a more heterogeneous surface (Figure 23). The average AUC of the log relative risk in California did not vary significantly across the three smoothing parameters where all 95% confidence intervals contained the average scores (Figure 25). However, an increase in the smoothing parameter corresponded to a reduction in the areas predicted as insignificantly different between plague-suitable and plague-unsuitable habitat as the smoothing parameter increased in value (Figure 24). The Sierra Valley flipped from plague-unsuitable to plague-suitable habitat as the smoothing parameter value increased.

Univariate relationships between the predictions and PRISM variables showed a strong relationship with temperature variables. Positive log relative risk (cases) was associated with lower temperatures (maximum, minimum, average, and dew point) while negative log relative risk (controls) was associated with higher temperatures and extremely low temperatures (Figure 26). Precipitation had a cyclical relationship with the predicted log relative risk. Lower temperatures were associated with statistically low significance level values (p-value < 0.025) of the log relative risk surface (Figure 27). On average, the areas of California with a statistically low significance level value (p-value < 0.025) were about eight degrees Celsius cooler, had about 125 millimeters more precipitation, and about seven hectopascal fewer maximum vapor pressure deficit than areas of California with a statistically high significance level value (p-value > 0.975; Figure 28).

3.4 Discussion

My results illustrate the value of the spatial nonparametric ENM for predicting the spatial distribution of a zoonotic pathogen (*Y. pestis*) in an sentinel species across an endemic region. My proposed method accounted for spatial variations in sampling effort by the CDPH by moderating the spatial patterns of case (presence) locations by the spatial patterns of control (absence) locations within a predictor space of principal components of climate variables. After a thorough search of the relevant literature, this represents the first analysis of the ecological niche of plague in coyotes using both case and control data. Presence-only (case-only) ENMs have been employed to predict the spatial distribution of *Y. pestis* in California by using plague-positive animals (including coyotes) (*85*, *89*), human plague cases (*70*), and flea vectors without known infection status (*86*). Hoar and colleagues (*116*) conducted a spatial cluster analysis of plague-positive coyotes irrespective of climatic variables. None of these studies accounted for sampling effort, which was accounted for in the present analysis by using control locations.

My results are in line with Holt and colleagues (*89*) who used plague-positive rodent locations collected by the CDPH with a presence-only ENM where plague-suitable habitat was predicted in the Sierra Nevada, Transverse Ranges, and Peninsular Ranges. However, unlike Holt and colleagues (*89*), my results showed no statistically significant plague-suitable habitat in the Southern Coast Ranges suggesting a potential limitation of using only coyotes to predict plague-suitable habitat. Sampling design affects wildlife disease surveillance (*159*) and differences between these two analyses may be an artifact of the different sampling design of rodents and coyotes – rodents are actively monitored in high consequence or historically observed areas whereas coyotes are passively monitored in conjunction with livestock/wildlife damage management operations (*108*). My analysis also identified areas of California with sparse observation data (e.g., Mojave Desert) and predicted plague-suitable habitat in these areas. Holt and colleagues (*89*) neither identified these areas nor discussed this data limitation in their analysis. Although these data-sparse locations were not frequently predicted as plague-suitable habitat, the areas that were (e.g., Mono County) and that have high human plague exposure risk (e.g., campsites) can be targets for additional

plague surveillance.

The relationship between climatic variables and the occurrence of *Y. pestis* predicted by my proposed method in California is consistent with previous studies. The best candidate models in the Holt and colleagues (*89*) analysis contained two temperature variables (maximum temperature of warmest month and temperature annual range). Plague outbreaks in animals (*57*, *149*, *231*) and humans (*60*, *62*) have been linked to milder temperatures associated with lower winter rodent mortality (*232*), higher flea vector survivorship (*63*), and higher flea transmission rate (*81*, *82*). Positive log relative risk values predicted across California with my proposed method were associated with milder temperatures. Additionally, low local significance level values (i.e., plague-suitable habitat) were associated with milder temperatures variable values (about 8°Celsius lower on average across the state) and high local significance level values (i.e., plague-unsuitable habitat) were associated with high temperatures. Precipitation has a more complex and temporally varying relationship with plague (*58–60*, *64*) likely by influencing the reproductive rate of rodent hosts (*232–235*). Results revealed a sinusoidal pattern between the predicted log relative risk values and precipitation across California where many precipitation values were associated with a positive log relative risk value, although no precipitation value was associated with low significance level values (plague-suitable habitat). In addition, high local p-values (i.e., plague-unsuitable habitat) were associated with extreme values of precipitation, especially in exceedingly dry climates (i.e., deserts). However, the average precipitation variable and the cross-sectional design was not designed to detect the annual or inter-annual precipitation variations associated with plague occurrence.

A nonparametric modeling approach is attractive in order to capture the nonlinear, complex relationships between disease occurrence and environmental variables (*207*). This is the first adaptation of a log relative risk function within the predictor space of a NPMR. NPMR has been used as an ENM (*208*, *236*), but this is the first application of the Kelsall & Diggle (*211*, *212*) method as an ENM. One main advance of this method is that the output is more readily interpretable than presence-only species distribution models. For example, the "logistic" output in MaxEnt (*237*), a common presence-only species distribution model, requires major assumptions to predict the probability of presence in a study region

(*238*, *239*) whereas the relative risk output from my proposed method does not require the same questionable assumptions. Another advance is the identification of statistically significant areas of predictor space (i.e., suitable and unsuitable habitat), which is selected via statistical inference rather than by an ad hoc selection of commission error cutoff. My proposed method is flexible, but conservative for areas of sparse data. Additionally, the predictions are robust as evident by cross-validation and a sensitivity analysis of the choice of smoothing parameter value. The cross-validation statistics are also within range of the training-test split cross-validation employed by Holt and colleagues (*89*). My proposed method requires few computational resources, which when paired with a large sample size (using both case and control data) allowed for k-fold cross-validation even for the relatively large environmental data set of 16 km^2 pixels across the state of California. Methods using both case and control data can account for sampling effort in the predictions (*161*), which addresses a major criticism of presence-only ENMs (*240–242*).

There are notable limitations to my proposed method, and it remains under development and evaluation. My proposed method is flexible for and sensitive to the choice of environmental variables used in my analysis. While this is not unique to the method (*243*), it is important to choose biologically relevant variables that are standardized. Multicollinearity of climatic variables is a common concern for ENMs as well as for my proposed method. Because my proposed method is designed for a two-dimensional predictor space, a dimension reduction technique (i.e., PCA) is employed instead of choosing two PRISM variables. My proposed method is limited to two variables (dimensions), but nonparametric multiplicative regression methods can be easily extended to many dimensions (*208*) and future studies can extend this method to more than two dimensions for applications in which more than two variables are necessary. Lastly, the log relative risk function is a pairwise comparison of case and control data; therefore, my proposed method cannot compare more than two groups. For analyses that assess ecological niche overlap between more than two groups (e.g., three or more species), further extension is required.

There were notable limitations for the plague prediction results. The analyses are a cross-sectional study design to predict the spatial distribution of endemic plague in California. These predictions used 30-year climate annual averages to account for climate

variability across recent time (1981–2010). Here, only abiotic climate variables were used for the analyses. Disease transmission and ecological niches are likely limited by both biotic and abiotic factors (*20*). Past studies of plague in California have primarily focused on climatic variables (*85*, *89*), but future studies could incorporate biotic variables into the principal component analysis. Coyote ecology and immunological response to *Y. pestis* add uncertainty around the location of disease exposure. The exact location and time of plague exposure by an individual coyote is not discernible because of the large home range size of coyotes (*144*, *145*) and laboratory testing of plague relies on antibodies that may be lifelong and not from a recent plague exposure. Indeed, evidence suggests that coyotes can maintain low levels of plague antibodies for long periods (*146*). Future studies can assess the effect of this type of uncertainty on the ecological niche of plague in coyotes.

Predicting the spatial distribution of plague solely relying on coyote observations is not recommended because of the problem of scale (*244*) in ecological analyses (Panel 5). The scale of the pathogen within a host, scale of the flea vector, scale of the transmission between hosts, scale of the coyote host behavior, scale of the habitat, scale of surveillance, and scale of the environmental variables in my analysis are not commonly aligned. Further, ENMs translate from geographic space to predictor space and back to geographic space, so the scale of the ecological niche and the spatial distribution of a disease are likely dissimilar. Plague transmission occurs in close proximity and likely happens within a 2.5 arcminute pixel of the principal components, but because a coyote can range larger than the size of a pixel, a coyote can be exposed to *Y. pestis* in one pixel but be observed in another. This is problematic in areas where neighboring pixel values (and habitat) are dissimilar because a plague occurrence would be observed in the incorrect grid cell for the ENM analysis.

One area of California can demonstrate the problem of scale for coyote plague surveillance: the Sierra Valley in Plumas and Sierra counties. The Sierra Valley is surrounded by mountains ranging up to 1 kilometer higher than the valley floor and thus have different climates. Coyote plague surveillance was conducted frequently in the agricultural areas of the valley floor where livestock damage management was conducted by WS. Seropositive coyotes were likely exposed to *Y. pestis* in the mountainous regions but were observed by CDPH surveillance on the valley floor. The mountainous regions have a climate that

matches other areas of the state with observed seropositive coyotes. The habitat of the valley floor is similar enough to the habitat of the surrounding mountainous regions that my proposed approach with a large smoothing parameter predicted the valley floor is plague-suitable habitat, but as the edges of the ecological niche were more refined with a smaller smoothing parameter a model predicted the valley floor was not plague-suitable habitat. Additional rodent plague surveillance is warranted in this area to confirm plague absence in the Sierra Valley because rodents are more localized within a pixel of my proposed analysis and can provide a more precise measurement of plague occurrence. Similarly, areas to target rodent plague surveillance are zones my proposed method could not distinguish between plague-suitable habitat or plague-unsuitable habitat. Further investigation of the problem of scale is favorable to more precisely identify high risk areas of *Y. pestis* exposure.

There are further enhancements for my proposed method that are of future interest. First, while my analysis is limited to a fixed-bandwidth kernel density estimator for the log relative risk surface, the utility of adaptive-bandwidth kernel density estimators is of interest as well. A sensitivity analysis of the smoothing parameter revealed the prediction using a fixed bandwidth was robust. A fixed bandwidth can perform poorly with highly spatially heterogeneous observations which an adaptive bandwidth can overcome (*226*), but this is yet to be tested in predictor space on estimating an ecological niche. Second, sample size considerations have not been examined here and future sensitivity analyses could assess the robustness of the predictions based on sample size. Third, investigations on how and where an ecological niche may change or shift through time, especially in the face of global climate change, is a foundation of biogeography (*245*, *246*). Spatio-temporal log relative risk estimation is available in geographic space and its application in predictor space is an attractive avenue of investigation.

Effective zoonotic disease surveillance relies on rich data from observational studies and laboratory analyses. This chapter provides an additional ecological niche modeling approach for analysis in circumstances where the researcher has both case (presence) and control (absence) data available. The present study was conducted in collaboration with the CDPH and the predicted spatial distribution of endemic plague in California will have immediate impact on state-wide plague surveillance. The predictions can inform ongoing rodent

surveillance locations, better define future coyote-based plague surveillance decisions, and identify a seropositive coyote that may have ranged into plague-unsuitable habitat areas. Extensions to my proposed method to allow for the different sampling intensities associated with other species (i.e., rodents) remains of interest. My proposed method also has potential to fill gaps in spatially heterogeneous sampling of other diseases collected by large-scale administrative surveillance systems.

My proposed method is extended and applied to two large data settings: the state of California (Chapter 4) and the western United States (Chapter 5). The basic structure defined and assessed generally above is used to customize analyses for both of these chapters. In Chapter 4, I use my proposed method to assess how the predicted ecological niche of *Y. pestis* in coyotes is affected by a type of data uncertainty. In Chapter 5, I use my proposed method to predict where the ecological niche of *Y. pestis* in coyotes is located in the western United States while accounting for sampling effort by coyote-based plague surveillance agencies. While not directly used in Chapter 6, I examine if the predicted spatial distribution of enzootic plague from Chapter 5 explains where human plague cases have occurred in the United States (1950–2017) in Chapter 6.

3.5 References

5. K. E. Jones *et al.*, *Nature* **451**, 990–993, ISSN: 0028-0836, DOI 10.1038/nature06536 (2008).

11. J. Grinnell, *The Auk* **34**, 427–433, ISSN: 938-4254, DOI 10.2307/4072271 (1917).

12. A. T. Peterson *et al.*, *Ecological niches and geographic distributions (MPB-49)* (Princeton University Press, Princeton, New Jersey, 2011), vol. 56, ISBN: 0691136882.

19. A. T. Peterson, *Mapping disease transmission risk: enriching models using biogeography and ecology* (Johns Hopkins University Press, Baltimore, Maryland, 2014), ISBN: 1421414737.

20. L. E. Escobar, M. E. Craft, *Frontiers in Microbiology* **7**, 1174, ISSN: 1664-302X, DOI 10.3389/fmicb.2016.01174 (2016).

33. K. J. Kugeler *et al.*, *Emerging Infectious Diseases* **21**, 16, ISSN: 1080-6059, 1080-6040, DOI 10.3201/eid2101.140564 (2015).

48. J. K. Runfola *et al.*, *Morbidity and Mortality Weekly Report* **64**, 429–434, ISSN: 0149-2195, 1545-861X (2015).

57. H. E. Brown *et al.*, *American Journal of Tropical Medicine and Hygiene* **82**, 95–102, ISSN: 0002-9637, 1476-1645, DOI 10.4269/ajtmh.2010.09-0247 (2010).

58. T. Ben-Ari *et al.*, *PLoS Pathogens* **7**, e1002160, ISSN: 1553-7374, 1553-7366, DOI 10.1371/journal.ppat.1002160 (2011).

59. T. Ben-Ari *et al.*, *American Journal of Tropical Medicine and Hygiene* **83**, 624–632, ISSN: 0002-9637, 1476-1645, DOI 10.4269/ajtmh.2010.09-0775 (2010).

60. T. Ben-Ari *et al.*, *Biology Letters* **4**, 737–740, ISSN: 1744-9561, 1744-957X, DOI 10.1098/rsbl.2008.0363 (2008).

62. R. E. Enscore *et al.*, *American Journal of Tropical Medicine and Hygiene* **66**, 186–196, ISSN: 0002-9637, 1476-1645, DOI 10.4269/ajtmh.2002.66.186 (2002).

63. D. C. Cavanaugh, J. D. Marshall, *Journal of Wildlife Diseases* **8**, 85–94, ISSN: 0090-3558, DOI 10.7589/0090-3558-8.1.85 (1972).

64. L. Xu *et al.*, *Proceedings of the National Academy of Sciences of the United States of America* **108**, 10214–10219, ISSN: 1091-6490, 0027-8424, DOI 10.1073/pnas.1019486108 (2011).

70. Y. Nakazawa *et al.*, *Vector-Borne and Zoonotic Diseases* **7**, 529–540, ISSN: 1557-7759, 1530-3667, DOI 10.1089/vbz.2007.0125 (2007).

71. R. J. Eisen *et al.*, *American Journal of Tropical Medicine and Hygiene* **77**, 999–1004, ISSN: 0002-9637, 1476-1645, DOI 10.4269/ajtmh.2007.77.999 (2007).

72. R. J. Eisen *et al.*, *Journal of Medical Entomology* **44**, 530–537, ISSN: 0022-2585, DOI 10.1603/0022-2585(2007)44 (2007).

73. R. J. Eisen *et al.*, *American Journal of Tropical Medicine and Hygiene* **77**, 121–125, ISSN: 0002-9637, 1476-1645, DOI 10.4269/ajtmh.2007.77.121 (2007).

81. A. M. Schotthoefer *et al.*, *Journal of Medical Entomology* **48**, 411–417, ISSN: 0022-2585, DOI 10.1603/ME10155 (2011).

82. D. C. Cavanaugh, *American Journal of Tropical Medicine and Hygiene* **20**, 264–273, ISSN: 0002-9637, 1476-1645, DOI 10.4269/ajtmh.1971.20.264 (1971).

85. S. P. Maher *et al.*, *American Journal of Tropical Medicine and Hygiene* **83**, 736–742, ISSN: 0002-9637, 1476-1645, DOI 10.4269/ajtmh.2010.10-0042 (2010).

86. J. C. Z. Adjemian *et al.*, *Journal of Medical Entomology* **43**, 93–103, ISSN: 0022-2585, DOI 10.1093/jmedent/43.1.93 (2006).

89. A. C. Holt *et al.*, *International Journal of Health Geographics* **8**, 38, ISSN: 1476-072X, DOI 10.1186/1476-072X-8-38 (2009).
90. M. Walsh, M. A. Haseeb, *PeerJ* **3**, e1493, ISSN: 2167-8359, DOI 10.7717/peerj.1493 (2015).
91. D. J. Salkeld, P. Stapp, *Vector-Borne and Zoonotic Diseases* **6**, 231–239, ISSN: 1557-7759, 1530-3667, DOI 10.1089/vbz.2006.6.231 (2006).
94. K. L. Gage, in *Advances in Yersinia research* (Springer, New York, New York, 2012), pp. 79–94, ISBN: 1461435600.
95. H. E. Brown *et al.*, *Vector-Borne and Zoonotic Diseases* **11**, 1439–1446, ISSN: 1557-7759, 1530-3667, DOI 10.1089/vbz.2010.0196 (2011).
102. J. Z. Adjemian *et al.*, *American Journal of Tropical Medicine and Hygiene* **76**, 365–375, ISSN: 0002-9637, 1476-1645, DOI 10.4269/ajtmh.2007.76.365 (2007).
107. S. N. Bevins *et al.*, *Integrative Zoology* **7**, 99–109, ISSN: 1749-4877, DOI 10.1111/j.1749-4877.2011.00277.x (2012).
108. J. R. Tucker *et al.*, *California compendium of plague control*, 2015, (2018; https://www.cdph.ca.gov/Programs/CID/DCDC/CDPH%20Document%20Library/CAPlagueCompendium.pdf).
112. P. W. Willeberg *et al.*, *American Journal of Epidemiology* **110**, 328–334, ISSN: 0002-9262, DOI 10.1093/oxfordjournals.aje.a112818 (1979).
113. C. U. Thomas, P. E. Hughes, *Journal of Wildlife Diseases* **28**, 610–613, ISSN: 0090-3558, DOI 10.7589/0090-3558-28.4.610 (1992).
114. N. W. Dyer, L. E. Huffman, *Journal of Wildlife Diseases* **35**, 600–602, ISSN: 0090-3558, DOI 10.7589/0090-3558-35.3.600 (1999).
115. A. Malmlov *et al.*, *Journal of Wildlife Diseases* **50**, 946–950, ISSN: 1943-3700, DOI 10.7589/2014-03-065 (2014).
116. B. R. Hoar *et al.*, *Preventive Veterinary Medicine* **56**, 299–311, ISSN: 0167-5877, DOI 10.1016/S0167-5877(02)00194-0 (2003).
117. R. J. Brinkerhoff *et al.*, *Vector-Borne and Zoonotic Diseases* **9**, 491–497, ISSN: 1557-7759, 1530-3667, DOI 10.1089/vbz.2008.0075 (2009).
118. L. A. Baeten *et al.*, *Journal of Wildlife Diseases* **49**, 932–939, ISSN: 0090-3558, DOI 10.7589/2013-02-040 (2013).
125. C. Daly *et al.*, *International Journal of Climatology* **28**, 2031–2064, ISSN: 1097-0088, DOI 10.1002/joc.1688 (2008).
126. E. M. Hart, K. Bell, *prism: download data from the Oregon PRISM project*, R package version 0.0.6, DOI 10.5281/zenodo.33663, (http://github.com/ropensci/prism).
127. R Core Team, *R: a language and environment for statistical computing*, R Foundation for Statistical Computing (Vienna, Austria, 2018), (https://www.R-project.org/).
144. R. M. Nowak, *Walker's mammals of the world* (Johns Hopkins University Press, Baltimore, Maryland, 1999), vol. 1, ISBN: 0801857899.
145. V. Howard Jr, G. G. DelFrate, in *Great Plains Wildlife Damage Control Workshop Proceedings* (University of Nebraska-Lincoln, Lincoln, Nebraska, 1991), pp. 39–49.
146. A. M. Barnes, *Symposia of the Zoological Society of London* **50**, 237–270 (1982).
149. P. Stapp *et al.*, *Frontiers in Ecology and the Environment* **2**, 235–240, ISSN: 1540-9295, DOI 10.1890/1540-9295(2004)002[0235:POEIPD]2.0.CO;2 (2004).
159. S. M. Nusser *et al.*, *Journal of Wildlife Management* **72**, 52–60, ISSN: 0022-541X, DOI 10.2193/2007-317 (2008).
161. L. Brotons *et al.*, *Ecography* **27**, 437–448, ISSN: 0906-7590, DOI 10.1111/j.0906-7590.2004.03764.x (2004).

162. T. Hastie, W. Fithian, *Ecography* **36**, 864–867, ISSN: 0906-7590, DOI 10.1111/j.1600-0587.2013.00321.x (2013).
163. A. M. Gormley *et al.*, *Journal of Applied Ecology* **48**, 25–34, ISSN: 1365-2664, DOI 10.1111/j.1365-2664.2010.01911.x (2011).
164. V. H. F. Gomes *et al.*, *Scientific Reports* **8**, 1003, ISSN: 2045-2322, DOI 10.1038/s41598-017-18927-1 (2018).
171. R. Kays, *Canis latrans*, 2018, (10.2305/IUCN.UK.2018-2.RLTS.T3745A103893556.en).
182. T. J. Quan *et al.*, in *Diagnostic procedures for bacterial, mycotic, and parasitic infections*, ed. by A. Barlows, W. J. Hausler (American Public Health Association, Washington, D.C., ed. 6, 1981), pp. 723–745, ISBN: 0875530869.
183. B. W. Hudson *et al.*, *Epidemiology & Infection* **60**, 443–450, ISSN: 0950-2688, DOI 10.1017/S002217240002057X (1962).
184. J. C. C. Neale, B. N. Sacks, *Oikos* **94**, 236–249, ISSN: 0030-1299, DOI 10.1034/j.1600-0706.2001.940204.x (2001).
185. G. A. Feldhamer *et al.*, Eds., *Wild mammals of North America: biology, management, and conservation* (Johns Hopkins University Press, Baltimore, Maryland, ed. 2, 2003), ISBN: 0801874165.
189. R. W. Emerson, *Nature* (Gutenberg Project, 2009), (2019; http://www.gutenberg.org/ebooks/29433).
190. D. M. Morens, A. S. Fauci, *mBio* **3**, e00494–12, ISSN: 2150-7511, DOI 10.1128/mBio.00494-12 (2012).
191. L. H. Taylor *et al.*, *Philosophical Transactions of the Royal Society B: Biological Sciences* **356**, 983–989, ISSN: 0962-8452, 1471-2954, DOI 10.1098/rstb.2001.0888 (2001).
192. T. Allen *et al.*, *Nature Communications* **8**, 1124, ISSN: 2041-1723, DOI 10.1038/s41467-017-00923-8 (2017).
193. In, *Sustaining Global Surveillance and Response to Emerging Zoonotic Diseases*, ed. by G. T. Keusch *et al.* (National Academies Press, Washington, D.C., 2009).
194. R. K. Plowright *et al.*, *Nature Reviews Microbiology* **15**, 502, ISSN: 1740-1526, DOI 10.1038/nrmicro.2017.45 (2017).
195. M. Voinson *et al.*, *Journal of Theoretical Biology* **457**, 37–50, ISSN: 0022-5193, DOI 10.1016/j.jtbi.2018.08.017 (2018).
196. J. L. Lowell *et al.*, *Journal of Clinical Microbiology* **43**, 650–656, DOI 10.1128/JCM.43.2.650-656.2005 (2005).
197. J. C. Gower, *Biometrics*, 857–871, ISSN: 1541-0420, 0006-341X, DOI 10.2307/2528823 (1971).
198. B. Leutner *et al.*, *RStoolbox: tools for remote sensing data analysis*, R package version 0.2.3, (https://CRAN.R-project.org/package=RStoolbox).
199. A. Guisan, N. E. Zimmermann, *Ecological Modelling* **135**, 147–186, ISSN: 0304-3800, DOI 10.1016/S0304-3800(00)00354-9 (2000).
200. A. H. Hirzel, G. Le Lay, *Journal of Applied Ecology* **45**, 1372–1381, ISSN: 1365-2664, DOI 10.1111/j.1365-2664.2008.01524.x (2008).
201. G. E. Hutchinson, *Cold Spring Harbor Symposia on Quantitative Biology* **22**, 415–427, ISSN: 0091-7451, 1943-4456, DOI 10.1101/SQB.1957.022.01.039 (1957).
202. M. P. Austin, *Oikos*, 170–178, ISSN: 0030-1299, DOI 10.2307/3546582 (1999).
203. J. Soberón, B. Arroyo-Peña, *PLoS One* **12**, e0175138, ISSN: 1932-6203, DOI 10.1371/journal.pone.0175138 (2017).

204. M. P. Austin *et al.*, *Journal of Vegetation Science* **5**, 215–228, ISSN: 1654-1103, DOI 10.2307/3236167 (1994).
205. G. E. Hutchinson, in *The ecological theatre and the evolutionary play* (Yale University Press, New Haven, Connecticut, 1965), pp. 26–78, DOI 10.1155/2010/150606.
206. P. McCullagh, J. A. Nelder, *Generalized linear models* (CRC Press, Boca Raton, Florida, 1989), vol. 37, ISBN: 0412317606.
207. B. McCune *et al.*, *Analysis of ecological communities* (MjM Software Design, Gleneden Beach, Oregon, 2002), vol. 28, ISBN: 0972129008.
208. B. McCune, *Nonparametric multiplicative regression for habitat modeling* (Oregon State University, Corvallis, Oregon, 2009).
209. B. McCune, *Journal of Vegetation Science* **17**, 819–830, ISSN: 1654-1103, DOI 10.1111/j.1654-1103.2006.tb02505.x (2006).
210. S. Mclafferty, *Annals of GIS* **21**, 127–133, ISSN: 1947-5683, 1947-5691, DOI 10.1080/19475683.2015.1008572 (2015).
211. J. E. Kelsall, P. J. Diggle, *Statistics in Medicine* **14**, 2335–2342, ISSN: 1097-0258, DOI 10.1002/sim.4780142106 (1995).
212. J. E. Kelsall, P. J. Diggle, *Bernoulli* **1**, 3–16, ISSN: 1573-9759, 1350-7265, DOI 10.2307/3318678 (1995).
213. D. C. Wheeler, *International Journal of Health Geographics* **6**, 13, ISSN: 1476-072X, DOI 10.1186/1476-072X-6-13 (2007).
214. H. E. Clough *et al.*, *Preventive Veterinary Medicine* **89**, 67–74, ISSN: 0167-5877, DOI 10.1016/j.prevetmed.2009.01.008 (2009).
215. J. F. Bithell, *Statistics in Medicine* **10**, 1745–1751, ISSN: 1097-0258, DOI 10.1002/sim.4780101112 (1991).
216. J. F. Bithell, *Statistics in Medicine* **9**, 691–701, ISSN: 1097-0258, DOI 10.1002/sim.4780090616 (1990).
217. B. W. Silverman, *Density estimation for statistics and data analysis* (Routledge, London, United Kingdom, 2018), ISBN: 0412246201.
218. L. Waller, in *Handbook in spatial statistics*, ed. by A. E. Gelfand *et al.* (CRC Press, Boca Raton, Florida, 2010), pp. 403–423, ISBN: 1420072877.
219. M. L. Hazelton, T. M. Davies, *Biometrical Journal* **51**, 98–109, ISSN: 1521-4036, 0323-3847, DOI 10.1002/bimj.200810495 (2009).
220. E. Parzen, *The Annals of Mathematical Statistics* **33**, 1065–1076, ISSN: 2168-8990, DOI 10.1214/aoms/1177704472 (1962).
221. T. M. Davies, M. L. Hazelton, *Statistics in Medicine* **29**, 2423–2437, ISSN: 1097-0258, DOI 10.1002/sim.3995. (2010).
222. A. T. Peterson *et al.*, *Ecological Modelling* **213**, 63–72, ISSN: 0304-3800, DOI 10.1016/j.ecolmodel.2007.11.008 (2008).
223. J. Davis, M. Goadrich, in *Proceedings of the 23rd International Conference on Machine Learning* (Association for Computing Machinery, 2006), pp. 233–240, ISBN: 1595933832, DOI 10.1145/1143844.1143874.
224. T. J. Hastie, in *Statistical models in S* (Routledge, London, United Kingdom, 2017), pp. 249–307, ISBN: 0412830402.
225. M. P. Wand, M. C. Jones, *Kernel smoothing* (CRC Press, Boca Raton, Florida, 1994), ISBN: 0412552701.
226. I. S. Abramson, *The Annals of Statistics*, 1217–1223, ISSN: 2168-8966, DOI 10.1214/aos/1176345986 (1982).

227. M. L. Hazelton, *Statistics in Medicine* **27**, 2269–2272, ISSN: 1097-0258, DOI 10.1002/sim.3047 (2008).
228. T. M. Davies *et al.*, *Statistics in Medicine* **32**, 4240–4258, ISSN: 1097-0258, DOI 10.1002/sim.5806 (2013).
229. G. R. Terrell, *Journal of the American Statistical Association* **85**, 470–477, ISSN: 0162-1459, 1537-274X, DOI 10.2307/2289786 (1990).
230. J. C. Marshall, M. L. Hazelton, *Journal of Multivariate Analysis* **101**, 949–963, ISSN: 0047-259X, DOI 10.1016/j.jmva.2009.09.003 (2010).
231. S. K. Collinge *et al.*, *EcoHealth* **2**, 102–112, ISSN: 1612-9202, 1612-9210, DOI 10.1007/s10393-005-3877-5 (2005).
232. R. R. Parmenter *et al.*, *American Journal of Tropical Medicine and Hygiene* **61**, 814–821, ISSN: 0002-9637, 1476-1645, DOI 10.4269/ajtmh.1999.61.814 (1999).
233. R. R. Parmenter *et al.*, *A report for the Federal Centers for Disease Control and Prevention*, DOI 10.13140/RG.2.1.2623.1841 (1993).
234. R. R. Parmenter, R. Vigil, *The HARDS epidemic in the Southwest: an assessment of autumn rodent densities and population demographics in central and northern New Mexico, October 1993* (Sevilleta Long-Term Ecological Research Program, Department of Biology, University of New Mexico, Albuquerque, New Mexico, 1993).
235. S. K. M. Ernest *et al.*, *Oikos* **88**, 470–482, ISSN: 0030-1299, DOI 10.1034/j.1600-0706.2000.880302.x (2000).
236. B McCune, M. J. Mefford, *HyperNiche. Non-parametric multiplicative habitat modeling*, version 1, Glenden Beach, Oregon, 2004.
237. S. J. Phillips *et al.*, *Ecological Modelling* **190**, 231–259, ISSN: 0304-3800, DOI 10.1016/j.ecolmodel.2005.03.026 (2006).
238. C. B. Yackulic *et al.*, *Methods in Ecology and Evolution* **4**, 236–243, ISSN: 2041-210X, DOI 10.1111/2041-210x.12004 (2013).
239. C. Merow *et al.*, *Ecography* **36**, 1058–1069, ISSN: 0906-7590, DOI 10.1111/j.1600-0587.2013.07872.x (2013).
240. M. B. Araujo, A. Guisan, *Journal of Biogeography* **33**, 1677–1688, ISSN: 0305-0270, DOI 10.1111/j.1365-2699.2006.01584.x (2006).
241. J. Stolar, S. E. Nielsen, *Diversity and Distributions* **21**, 595–608, ISSN: 1366-9516, DOI 10.1111/ddi.12279 (2015).
242. A. El-Gabbas, C. F. Dormann, *Ecology and Evolution* **8**, 2196–2206, ISSN: 2045-7758, DOI 10.1002/ece3.3834 (2018).
243. V. Lecours *et al.*, *PLoS One* **11**, e0167128, ISSN: 1932-6203, DOI 10.1371/journal.pone.0167128 (2016).
244. S. A. Levin, *Ecology* **73**, 1943–1967, ISSN: 0012-9658, DOI 10.2307/1941447 (1992).
245. J. J. Wiens, *Philosophical Transactions of the Royal Society B: Biological Sciences* **366**, 2336–2350, ISSN: 0962-8452, 1471-2954, DOI 10.1098/rstb.2011.0059 (2011).
246. A. T. Peterson *et al.*, *Science* **285**, 1265–1267, ISSN: 0036-8075, DOI 10.1126/science.285.5431.1265 (1999).
247. A. T. Peterson, J. Soberón, *Natureza & Conservação* **10**, 102–107, ISSN: 1679-0073, DOI 10.4322/natcon.2012.019 (2012).

3.6 Appendices

3.6.1 Appendix A: Panels

Panel 4: Definition of key terms

The following are definitions used for key terms in this chapters:

Ecological niche model: A type of statistical method that estimates the distribution of a species in a space comprised of environmental dimensions. The spatial distribution of an estimated niche can then be predicted in geographic space. See (*12*, *247*).

Enzootic: Period of low incidence, but maintenance, of a disease in an animal population.

Epizootic: Period of high incidence of a disease in an animal population.

Fundamental ecological niche: The response of a species to the combination of abiotic conditions (*201*).

Geographic space: Area relative to its position on earth (i.e., longitude and latitude).

Habitat: The combination of environmental characteristics that describe a particular location (*199*).

Predictor space: Area relative to its position among dimensions of variables. If the dimensions are environmental variables, this can be referred to as "environmental space."

Realized ecological niche: The response of a species to a combination of abiotic and biotic conditions (*201*).

Suitable habitat: Area in which a species (or disease) can potentially or does occur.

Sylvatic: The occurrence and transmission of a disease within wildlife populations.

Unsuitable habitat: Area in which a species (or disease) can not potentially or does not occur.

Zoonotic: The occurrence and transmission of a disease from animal populations to human populations.

Panel 5: Considerations for coyotes as a sentinel species for plague activity

Coyotes (*Canis latrans*) have been used as a sentinel species of plague (*Yersinia pestis*) activity in rodent populations of North America (*95*, *112–116*), but there are limitations for focusing on coyote-based plague surveillance data. I outline some notable limitations below:

- Coyotes are a wide-ranging species (*144*, *145*, *171*) and effectively sample a broad variety of habitats in which plague occurs and thus sample a considerable extent of the ecological niche of plague (*85*). However, coyotes are not observed homogeneously across North America (Chapter 5) and other species may be better indicators of plague activity in some habitats (e.g., rodent species for alpine regions of California). Therefore, without rodent data, I am estimating a the fundamental ecological niche of plague in coyotes, which I am considering as an approximation for the entire fundamental ecological niche of plague.
- The location where an individual coyote was truly exposed to *Y. pestis* is unknown and challenging to discern (*91*, *117*). Coyotes are mobile, encountering many rodents across their expansive individual home range (*144*, *145*, *184*) and whose plague antibodies can last for many months (*91*, *112*, *118*). Reinfection of coyotes is also probable and so a coyote can only indicate its most recent plague exposure. Rodent species have smaller individual home range sizes (*185*) and are the gold standard indicator of current or recent plague activity (*108*).
- Coyote specimens are not collected evenly across the United States and instead are collected opportunistically (*107*). Coyotes tested for plague exposure are primarily collected in conjunction with ongoing livestock/wildlife damage management operations conducted by the U.S. Department of Agriculture Animal and Plant Health Inspection Service Wildlife Services. Agricultural and urban areas may be more sampled for plague than other land-use types and their respective habitats. Including other species monitored for plague activity may help overcome this potential spatial sampling effort bias.

3.6.2 Appendix B: Tables

Table 6: Oregon State University Parameter elevation Regression on Independent Slopes Model (PRISM) 30-Year Average Annual Normals (1981–2010). Modeled using a combination of a digital elevation model (DEM) and climatologically-aided interpolation (CAI)*. Variables were modeled at 30 arcseconds (∼800 m) resolution and aggregated to 2.5 arcminutes (∼4 km). See (*125*) for more details.

Variable	Units	Derivation
Precipitation	millimeters (mm)	Modeled; Summing monthly averages (rain + melted snow)
Maximum Temperature	°Celsius (°C)	Modeled; Averaging over all months using a DEM as the predictor grid
Mean Temperature	°Celsius (°C)	Derived; Average of Maximum Temperature and Minimum Temperature
Minimum Temperature	°Celsius (°C)	Modeled; Averaging over all months using a DEM as the predictor grid
Mean Dewpoint Temperature	°Celsius (°C)	Modeled; CAI used minimum temperature as the predictor grid
Maximum Vapor Pressure Deficit	hectopascal (hPA)	Modeled; CAI used mean dewpoint temperature and maximum temperature as the predictor grids
Minimum Vapor Pressure Deficit	hectopascal (hPA)	Modeled; CAI used mean dewpoint temperature and minimum temperature as the predictor grids

*Accuracy of these data is based on the original specification of the Defense Mapping Agency one-degree DEMs. The stated accuracy of the original DEMs is 130-meter circular error with 90% probability. Data sets use all weather stations, regardless of time of observation.

Table 7: Summary of principal component analysis of Oregon State University Parameter elevation Regression on Independent Slopes Model (PRISM) 30-Year Average Annual Normals (1981–2010) at a 2.5 arcminute (∼4 km) resolution. Variables were range standardized using the Gower metric (*197*).

Statistic	**Variable**	**Principal Component**		
		PC1	PC2	PC3
Standard Deviation		0.326	0.170	0.058
Proportion of Variance		0.756	0.205	0.024
Cumulative Proportion of Variance		0.756	0.962	0.985
Loadings	Precipitation	0.063	0.342	0.549
	Maximum Temperature	0.485	-0.208	-0.290
	Mean Temperature	0.519	-0.064	0.015
	Minimum Temperature	0.477	0.088	0.314
	Mean Dewpoint Temperature	0.446	0.532	-0.101
	Maximum Vapor Pressure Deficit	0.243	-0.533	-0.229
	Minimum Vapor Pressure Deficit	0.073	-0.510	0.673

3.6.3 Appendix C: Figures

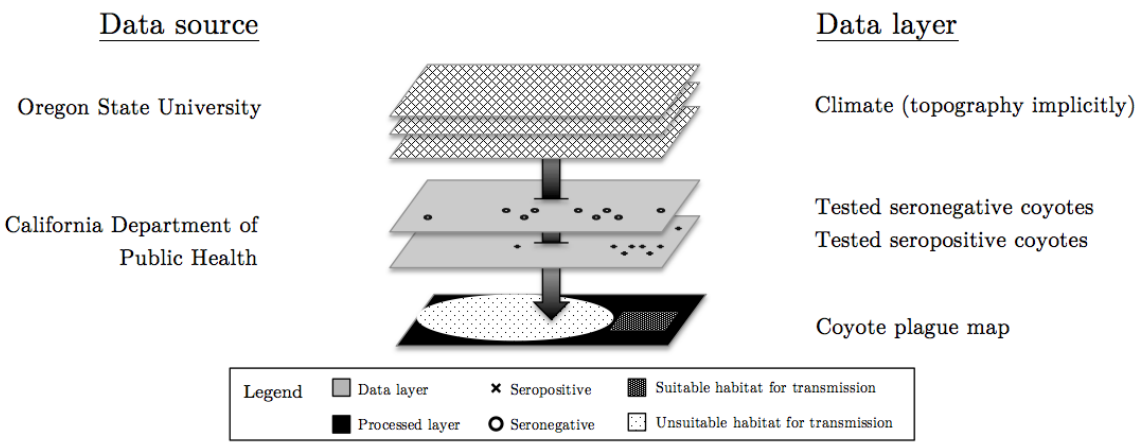
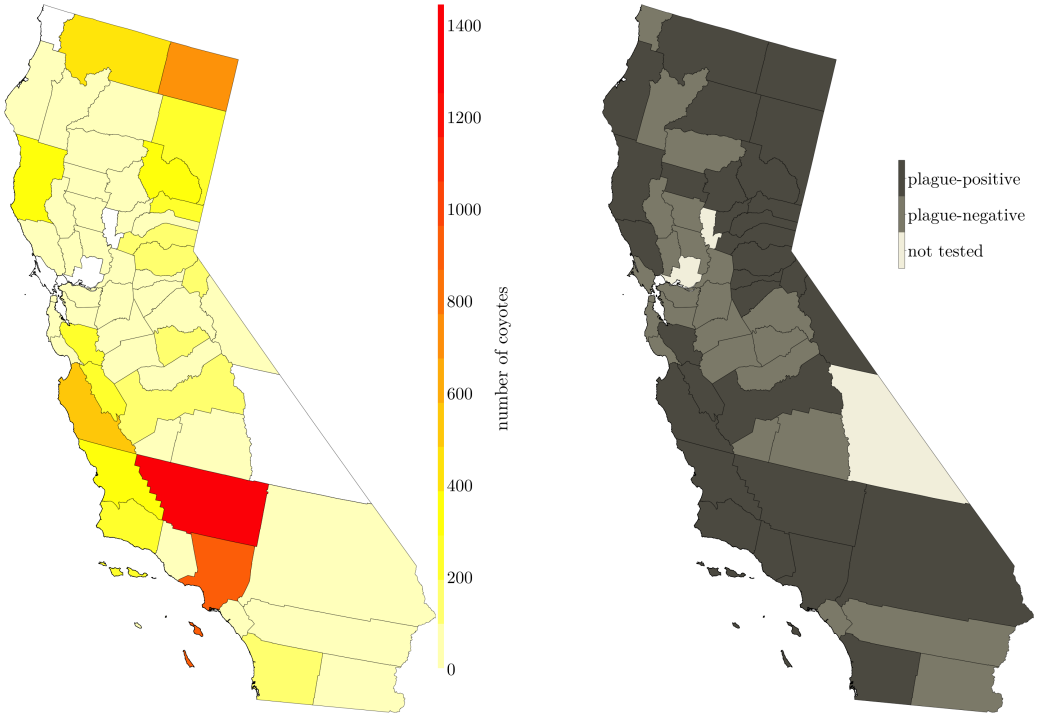


Figure 11: Data layers of proposed method and their sources. Data are structured as point locations and raster grids. Data layers (grey color) are used as inputs for the proposed method to predict the spatial distribution of *Yersinia pestis* in California (black color). The arrow represents the layers involved for the predicted layer.



(a) **(b)**

Figure 12: **(a)** Number of coyotes tested by California Department of Public Health (CDPH) for antibodies against *Yersinia pestis* (1983–2015). Sampling is heterogeneous across the state of California, primarily in historic plague endemic areas and areas of high concern. Limited sampling in the Mojave and Sonoran Deserts. **(b)** Results of coyotes tested by the CDPH for antibodies against *Yersinia pestis* (1983–2015). A coyote with plague antibodies was not observed in every county sampled. Only 2% of observations were unable to be geolocated and ignored in subsequent analysis.

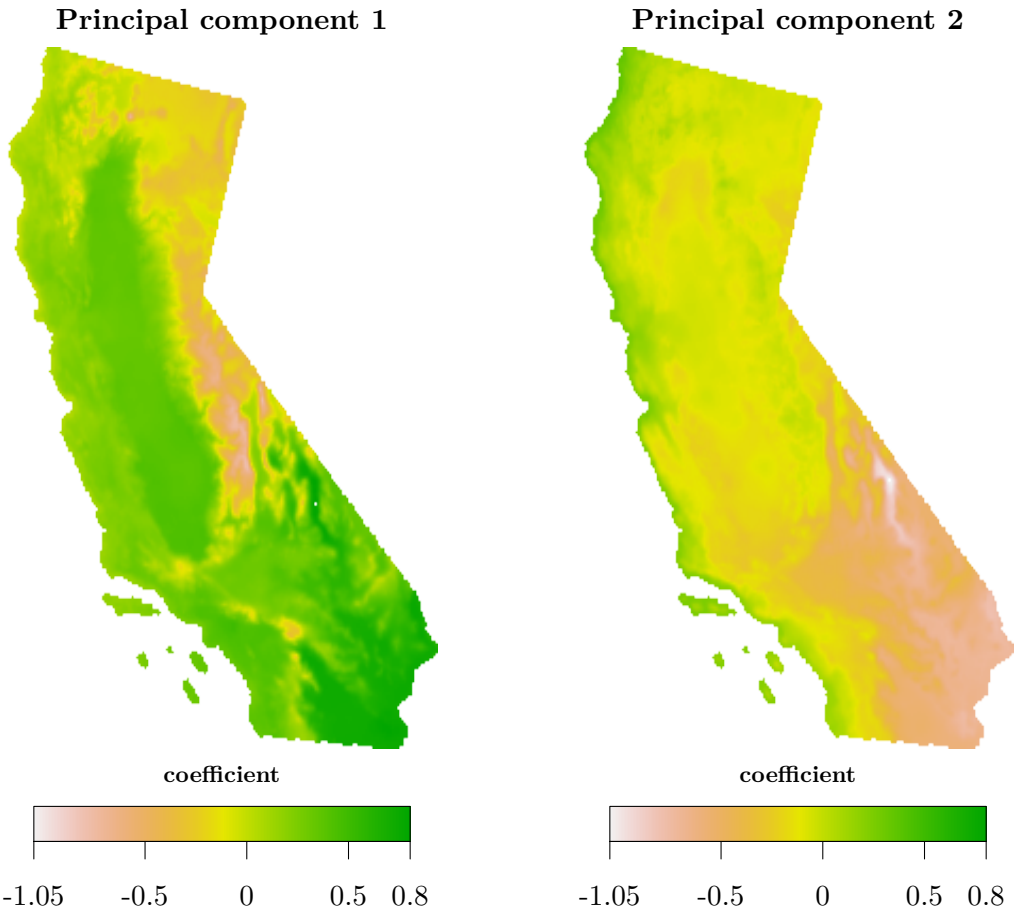


Figure 13: First two principal components of a principal component analysis of Oregon State University Parameter Elevation Regression on Independent Slopes Model 30-year average annual normals (1981–2010) at a 2.5 arcminute (∼4 km) resolution of the Western United States and presented for California. Variables were range standardized using the Gower metric (*197*).

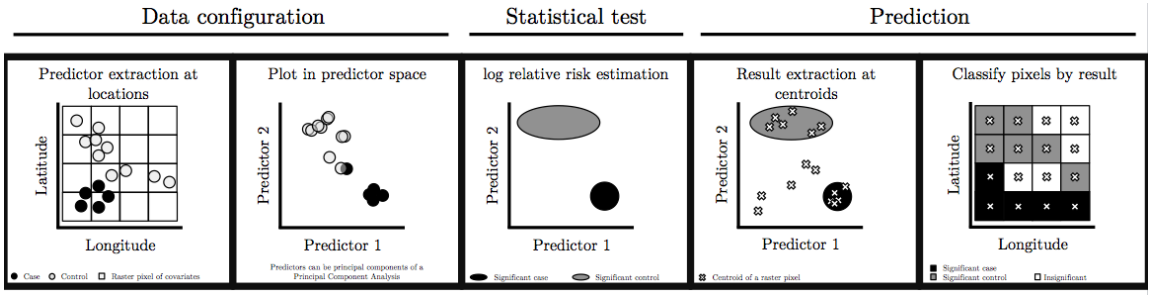


Figure 14: Ecological niche modeling approach adapting the Kelsall & Diggle (*211*, *212*) method into a non-parametric multiplicative regression (*208*, *209*). The Kelsall & Diggle (*211*, *212*) method is employed in "predictor space" instead of geographic space (i.e., latitude and longitude) to detect statistically significant clustering of environmental variables for cases and control information. The significant clusters of cases (i.e., ecological niche of disease) can be back-transformed to geographic space to identify areas that have similar environmental values within the significant clusters – a version of environmental interpolation (*19*).

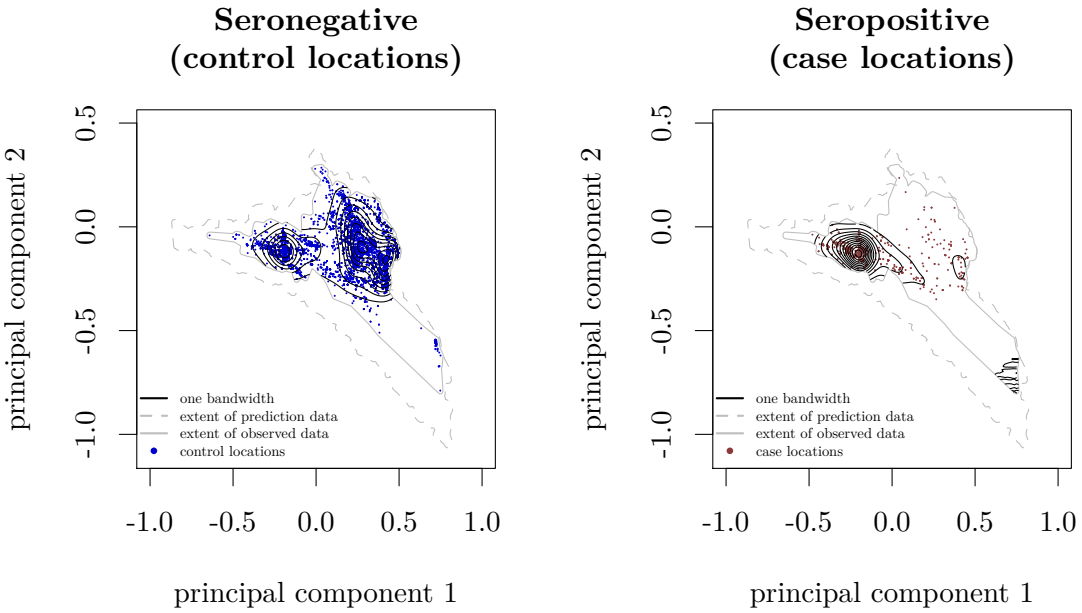


Figure 15: Estimated intensity surface in predictor space of coyotes that tested negative for plague antibodies (controls) and tested positive for plague antibodies (cases) by the California Department of Public Health (CDPH; 1983–2015). Predictor space is comprised of the first two principal components of a principal component analysis of seven range-standardized (*197*) Oregon State University Parameter Elevation Regression on Independent Slopes Model 30-year average annual normals (1981–2010) at a 2.5 arcminute (∼4 km) resolution of the Western United States. The bandwidth (0.052 coefficient units) was chosen using the maximal smoothing principle (*229*). The dashed grey line is the entire extent of predictor space of California and the solid grey line is the extent of predictor space that CDPH has sampled with coyote specimens. The area outside of the solid grey line is habitat that has not been sampled by CDPH.

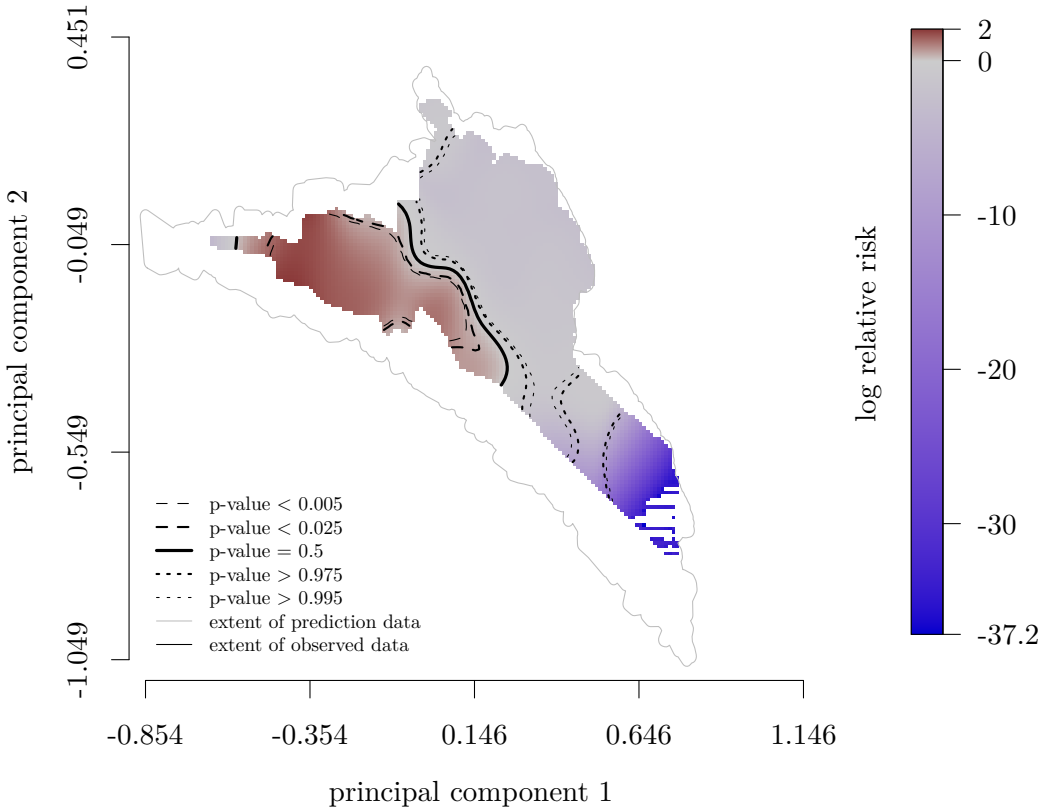


Figure 16: Estimated log relative risk surface in predictor space of coyotes comparing areas that tested negative for plague antibodies (controls) and tested positive for plague antibodies (cases) by the California Department of Public Health (CDPH; 1983–2015) using the proposed method. Calculated asymptotic tolerances at given two-tailed significance levels ($\alpha = 0.05$ & $\alpha = 0.01$) are included (*219*, *221*). Color pertains to log relative risk values where positive log relative risk (more likely cases) are in red and negative log relative risk (more likely controls) with greyer coloring closer to the null log relative risk value (zero). Predictor space is comprised of the first two principal components of a principal component analysis of seven range-standardized (*197*) Oregon State University Parameter Elevation Regression on Independent Slopes Model 30-year average annual normals (1981–2010) at a 2.5 arcminute (∼4 km) resolution of the Western United States. The bandwidth (0.052 coefficient units) was chosen using the maximal smoothing principle (*229*). The dashed grey line is the entire extent of predictor space of California and the thin solid black line is the extent of predictor space that CDPH has sampled with coyote specimens. The area outside of the solid grey line is habitat that has not been sampled by CDPH.

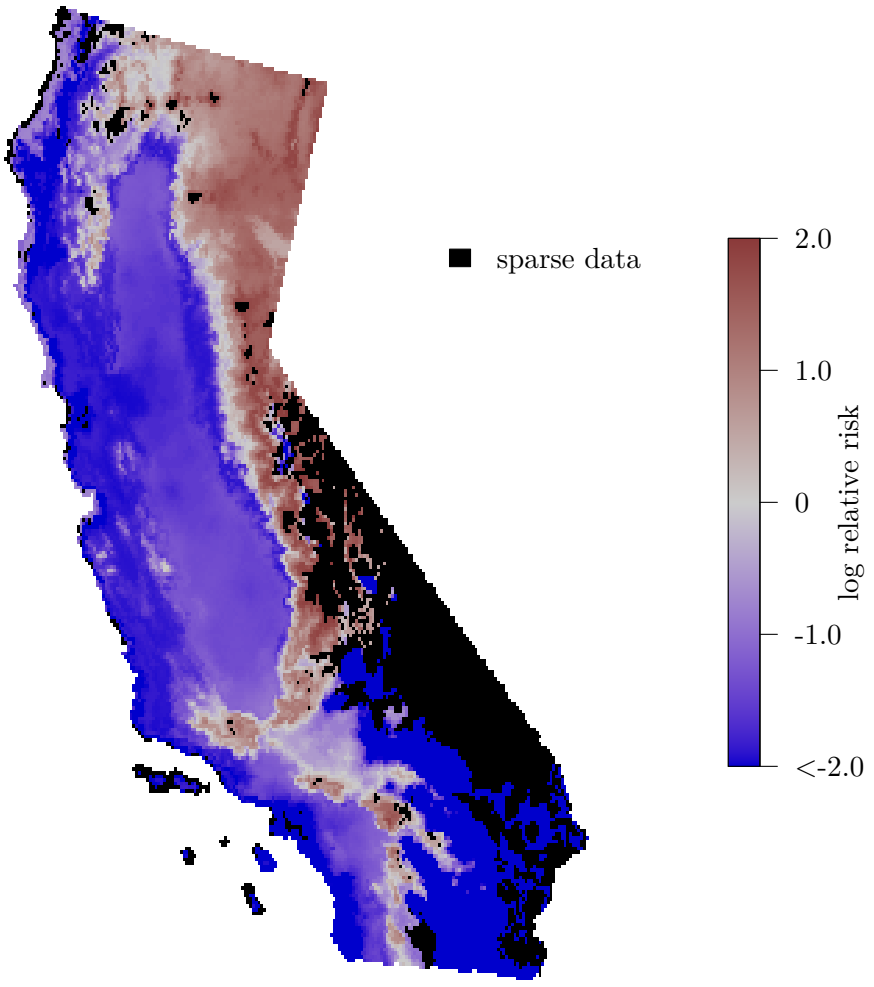


Figure 17: Areas of California predicted with log relative risk surface in predictor space of coyotes that tested negative for plague antibodies (controls) and tested positive for plague antibodies (cases) by the California Department of Public Health (CDPH; 1983–2015) using the proposed method. Color pertains to log relative risk values where positive log relative risk (more likely suitable habitat for plague transmission) are in red and negative log relative risk (more likely unsuitable habitat for plague transmission) with greyer coloring closer to the null log relative risk value (zero). The areas in yellow correspond to habitat that was not sampled by CDPH.

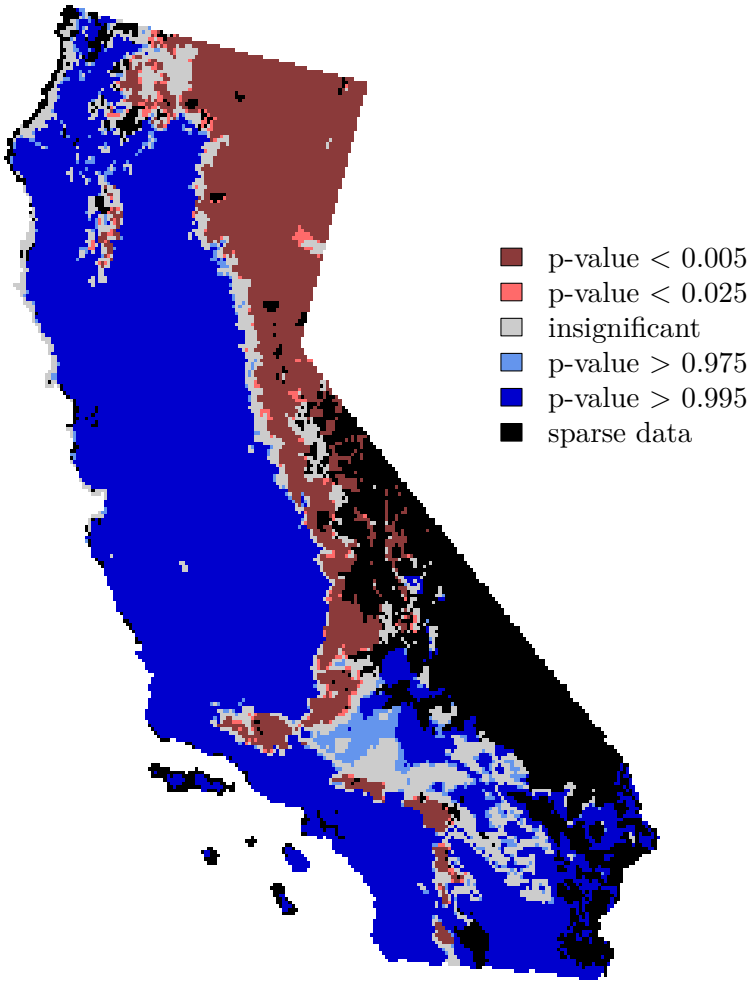


Figure 18: Areas of California predicted with log relative risk surface in predictor space of coyotes that tested negative for plague antibodies (controls) and tested positive for plague antibodies (cases) by the California Department of Public Health (CDPH; 1983–2015) using the proposed method. Color pertains to calculated asymptotic tolerances (*219*, *221*) at given two-tailed significance levels ($\alpha = 0.05$ & $\alpha = 0.01$). Warmer-colored areas are more likely suitable habitat for plague transmission and cooler-colored areas are more likely unsuitable habitat for plague transmission with greyer-colored areas statistically indistinguishable. The areas in yellow correspond to habitat that was not sampled by CDPH.

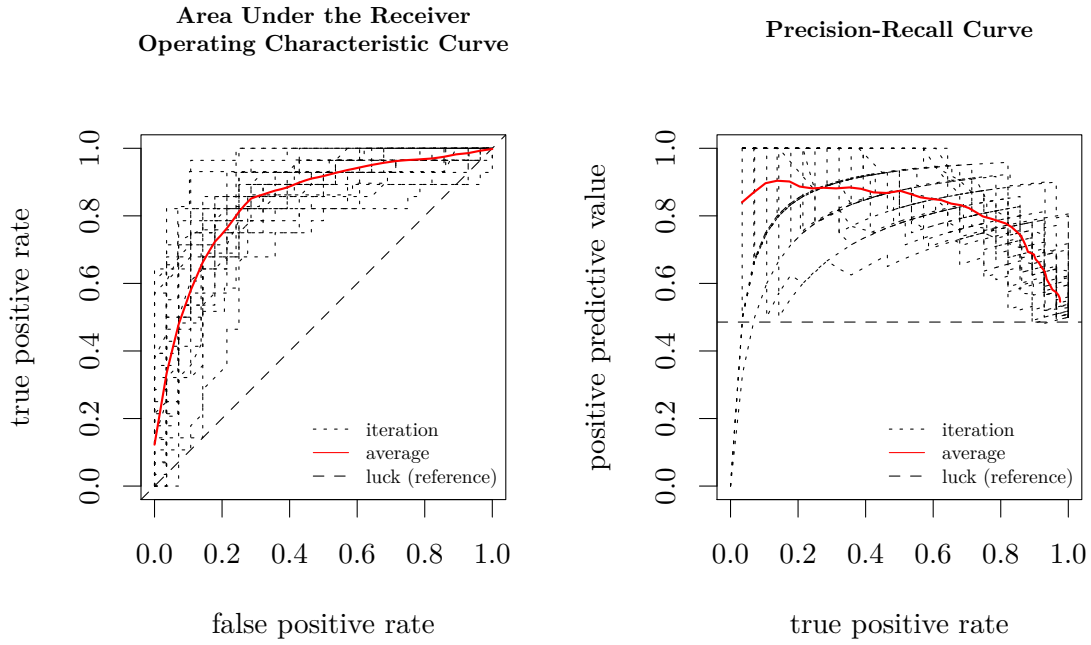


Figure 19: Results of 25-fold cross-validation of the estimated log relative risk surface in predictor space of coyotes that tested negative for plague antibodies (controls) and tested positive for plague antibodies (cases) by the California Department of Public Health (1983–2015) using my proposed method. Iterations were balanced (prevalence = 0.5) by randomly undersampling control locations used for in each fold for cross-validation. Results are robust with a high average Area Under the Receiver Operating Characteristic Curve (AUC) of 0.846 (95% CI: 0.825–0.867) and an acceptable precision-recall curve.

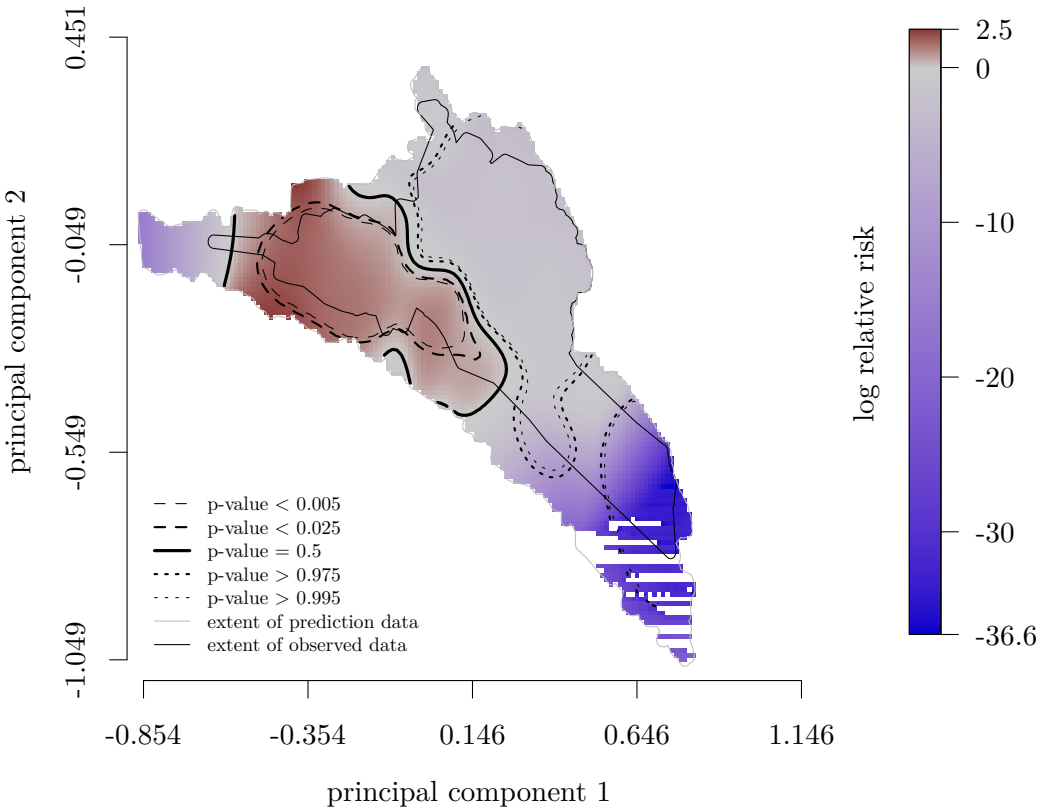
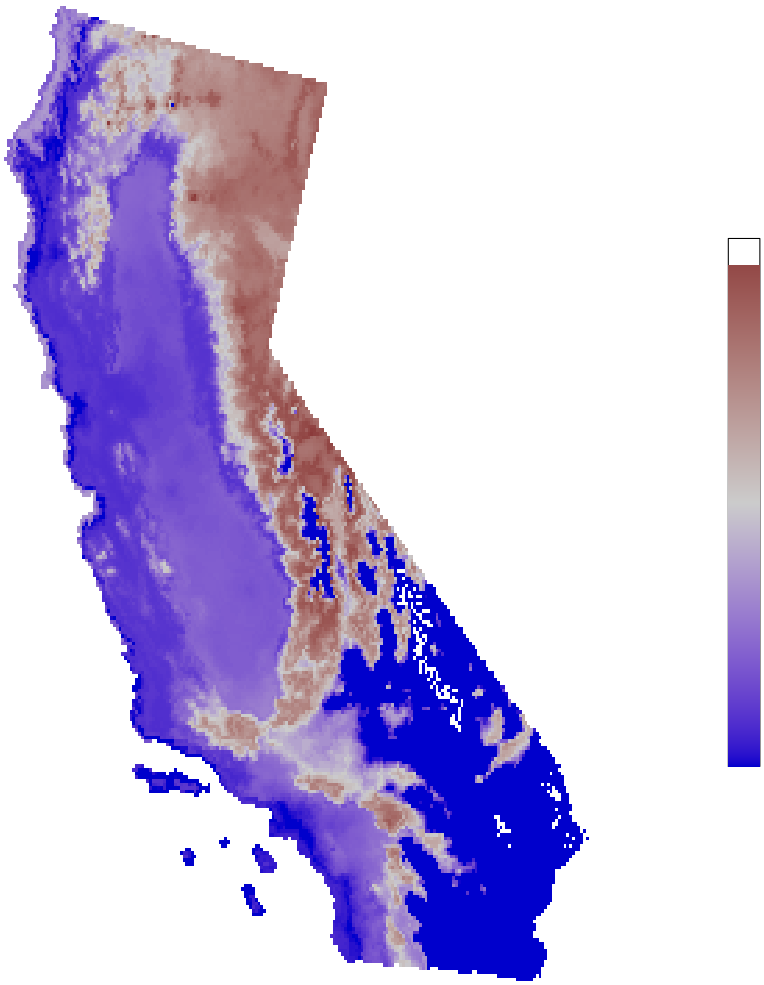


Figure 20: Estimated log relative risk surface in predictor space of coyotes comparing areas that tested negative for plague antibodies (controls) and tested positive for plague antibodies (cases) by the California Department of Public Health (CDPH; 1983–2015) using the proposed method. The log relative risk surface was predicted into habitat of California that was not sampled by CDPH. Calculated asymptotic tolerances at given two-tailed significance levels ($\alpha = 0.05$ & $\alpha = 0.01$) are included (*219*, *221*). Color pertains to log relative risk values where positive log relative risk (more likely cases) are in red and negative log relative risk (more likely controls) with greyer coloring closer to the null log relative risk value (zero). The areas in yellow coloring correspond to habitat with sparse data and statistically dissimilar from sampled habitat. Predictor space is comprised of the first two principal components of a principal component analysis of seven range-standardized (*197*) Oregon State University Parameter Elevation Regression on Independent Slopes Model 30-year average annual normals (1981–2010) at a 2.5 arcminute (∼4 km) resolution of the Western United States. The bandwidth (0.052 coefficient units) was chosen using the maximal smoothing principle (*229*). The dashed grey line is the entire extent of predictor space of California and the thin solid black line is the extent of predictor space that CDPH has sampled with coyote specimens. The area outside of the solid grey line is habitat that has not been sampled by CDPH.



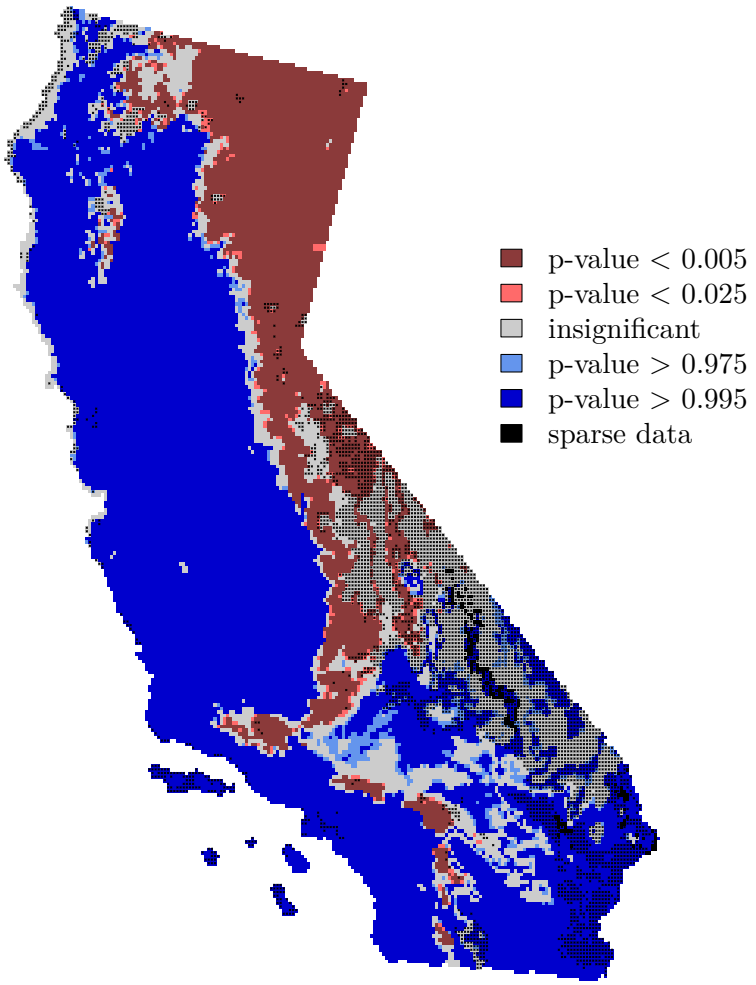
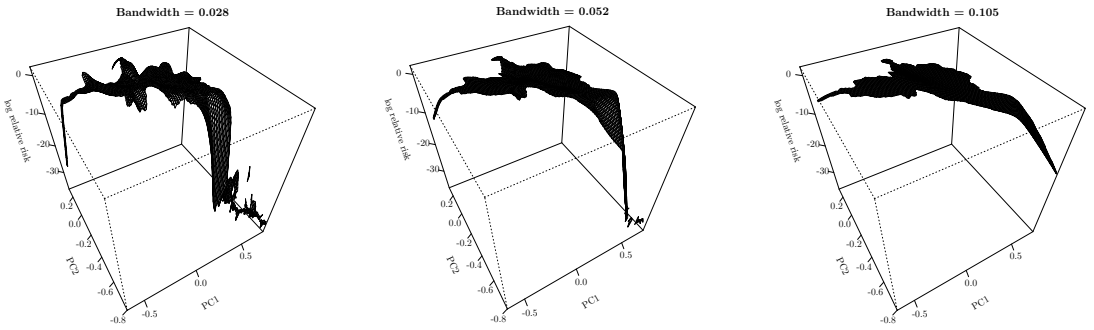


Figure 22: Areas of California predicted with log relative risk surface in predictor space of coyotes that tested negative for plague antibodies (controls) and tested positive for plague antibodies (cases) by the California Department of Public Health (CDPH; 1983–2015) using the proposed method. The prediction was extended to areas of California that CDPH had not been previously sampled. Color pertains to calculated asymptotic tolerances (*219*, *221*) at given two-tailed significance levels ($\alpha = 0.05$ & $\alpha = 0.01$). Warmer-colored areas are more likely suitable habitat for plague transmission and cooler-colored areas are more likely unsuitable habitat for plague transmission with greyer-colored areas statistically indistinguishable. The areas in yellow coloring correspond to habitat with sparse data and statistically dissimilar from sampled habitat.



(a) **(b)** **(c)**

Figure 23: Comparison of bandwidth selection on the log relative risk surface in predictor space of coyotes that tested negative for plague antibodies (controls) and tested positive for plague antibodies (cases) by the California Department of Public Health (1983–2015) using the proposed method. Predictor space was comprised of the first two principal components (PC1 & PC2) of a principal component analysis of seven range-standardized (*197*) Oregon State University Parameter Elevation Regression on Independent Slopes Model 30-year average annual normals (1981-2010) at a 2.5 arcminute (∼4 km) resolution of the Western United States. Choice of bandwidth is highly influential of the log relative risk surface with larger bandwidths smoothing the log relative risk surface, which reduces the detection of local maximums (minimums), but smaller bandwidths cannot estimate the log relative risk surface in areas with sparse data.

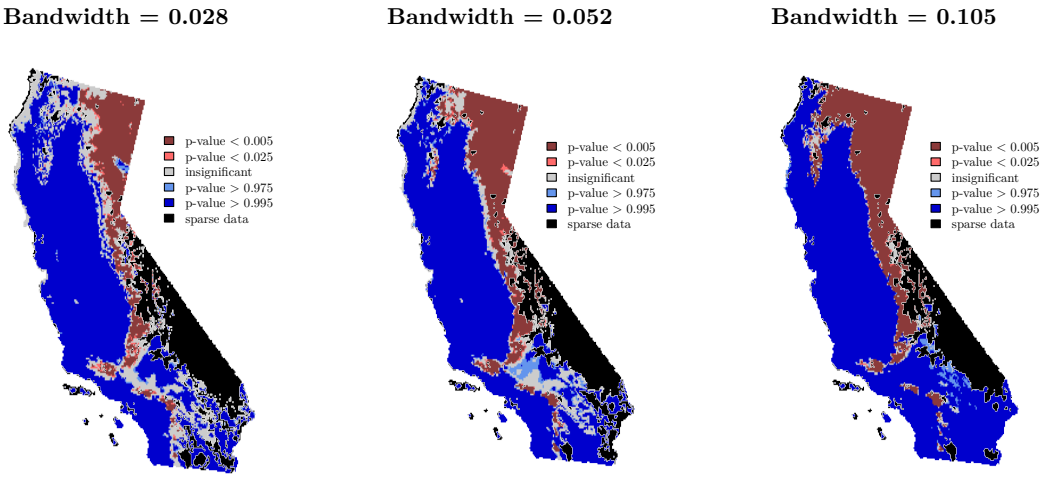


Figure 24: Comparison of bandwidth selection on the predicted area of California with a log relative risk surface in predictor space of coyotes that tested negative for plague antibodies (controls) and tested positive for plague antibodies (cases) by the California Department of Public Health (1983–2015) using the proposed method. Color pertains to calculated asymptotic tolerances (*219*, *221*) at given two-tailed significance levels ($\alpha = 0.05$ & $\alpha = 0.01$). Warmer-colored areas are more likely suitable habitat for plague transmission and cooler-colored areas are more likely unsuitable habitat for plague transmission with greyer-colored areas statistically indistinguishable. The areas in yellow coloring correspond to habitat with sparse data. The prediction is robust across bandwidth selection with a notable reduction in habitats the log relative risk surface cannot statistically distinguish as suitable (unsuitable) for plague transmission. The Sierra Valley changes designation with choice of bandwidth suggesting rodent plague surveillance is required in this area and not rely solely on coyote plague surveillance.

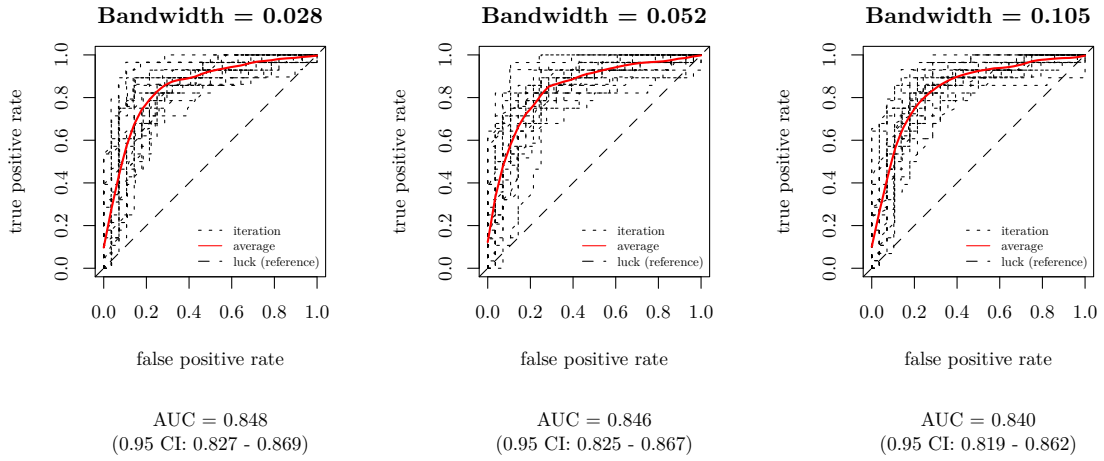


Figure 25: Results of 25-fold cross-validation of a bandwidth selection comparison for the log relative risk surface in predictor space of coyotes that tested negative for plague antibodies (controls) and tested positive for plague antibodies (cases) by the California Department of Public Health (1983–2015) using the proposed method. Iterations were balanced (prevalence = 0.5) by randomly undersampling control locations used for in each fold for cross-validation. The prediction is robust across bandwidth selection with similar Area Under the Receiver Operating Characteristic Curve (AUC) across bandwidths.

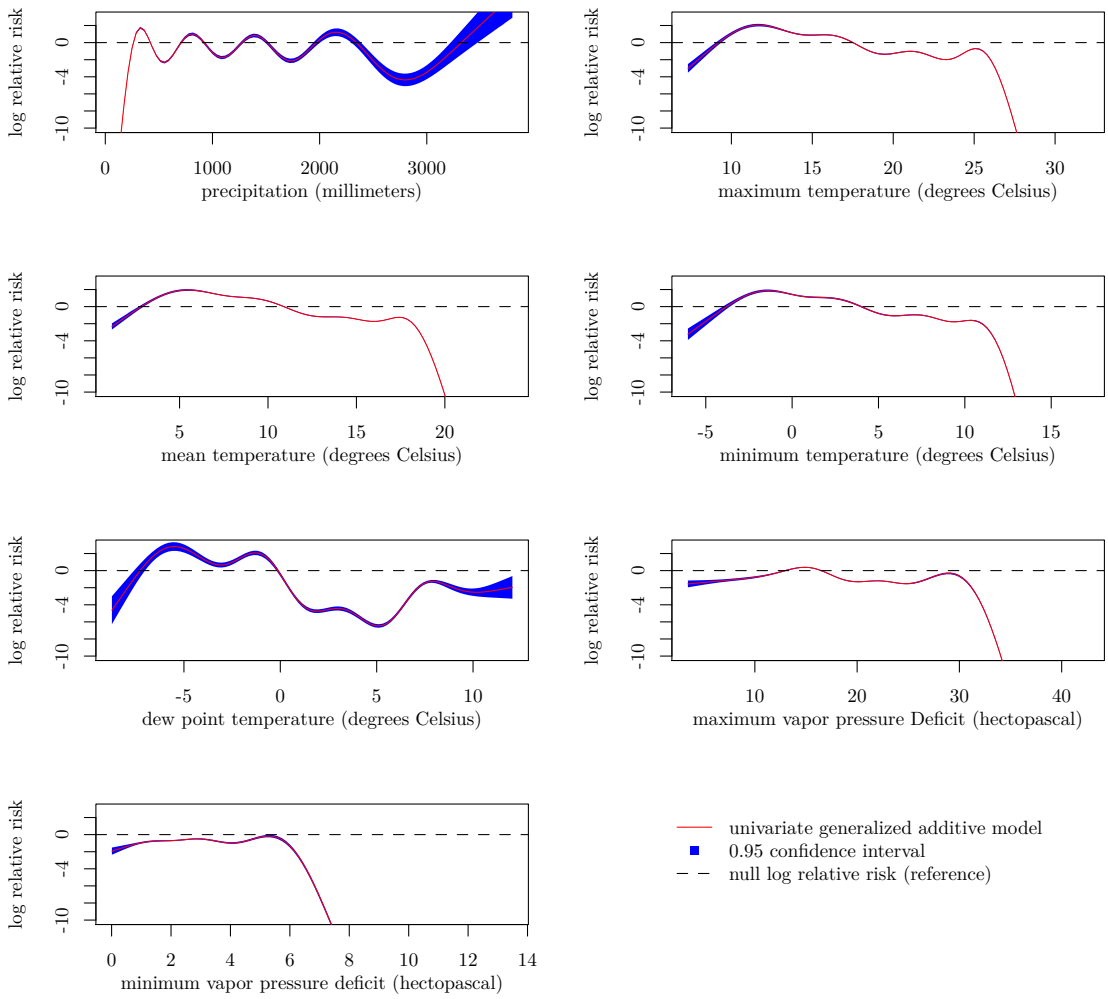


Figure 26: Univariate associations between each Oregon State University Parameter Elevation Regression on Independent Slopes Model 30-year average annual normal variable (1981–2010) at a 2.5 arcminute (∼4 km) resolution and the predicted log relative risk value for California. Values were resampled from each grid cell in California and a smoothed generalized additive model (*224*) trend was estimated (red-colored) with a 95% confidence interval (blue-colored). The null log relative risk value (zero) is included as a black dashed line for reference with trend values above the null value suggest plague-suitable conditions while trend values below the null value suggest plague-unsuitable conditions. Beyond a cyclical pattern of precipitation, plague-suitable habitat appears to be associated with cooler temperatures (average annual maximum, minimum and mean temperature as well as dew point temperature) and mid-range vapor pressure deficit values (i.e., not too dry and not too humid).

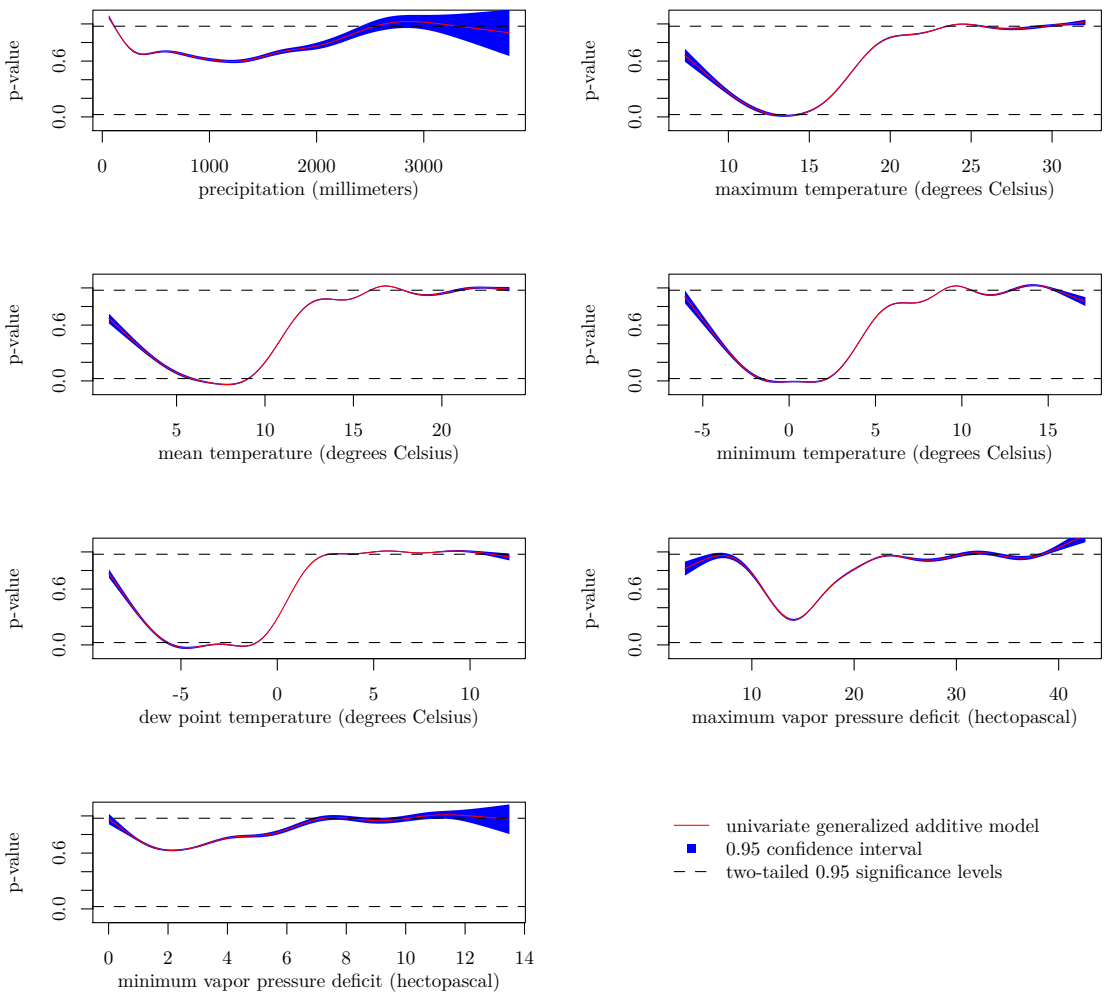


Figure 27: Univariate associations between each Oregon State University Parameter Elevation Regression on Independent Slopes Model 30-year average annual normal variable (1981–2010) at a 2.5 arcminute (∼4 km) resolution and the predicted two-tailed asymptotic p-values (*219*, *221*) of the log relative risk surface for California. Values were resampled from each grid cell in California and a smoothed generalized additive model (*224*) trend was estimated (red-colored) with a 95% confidence interval (blue-colored). The two-tailed 95% significance levels are included as a black dashed line for reference with trend values below the lower significance level value (0.025) value suggest plague-suitable conditions while trend values above the upper significance level value (0.975) value suggest plague-unsuitable conditions. Plague-suitable habitat appears to be associated with cooler temperatures (average annual maximum, minimum and mean temperature as well as dew point temperature), but not cold temperatures. Plague-unsuitable habitat appears to be associated with higher average annual temperature values as well as precipitation and humidity values at both extremes, but not as strong of an association as the temperature variables.

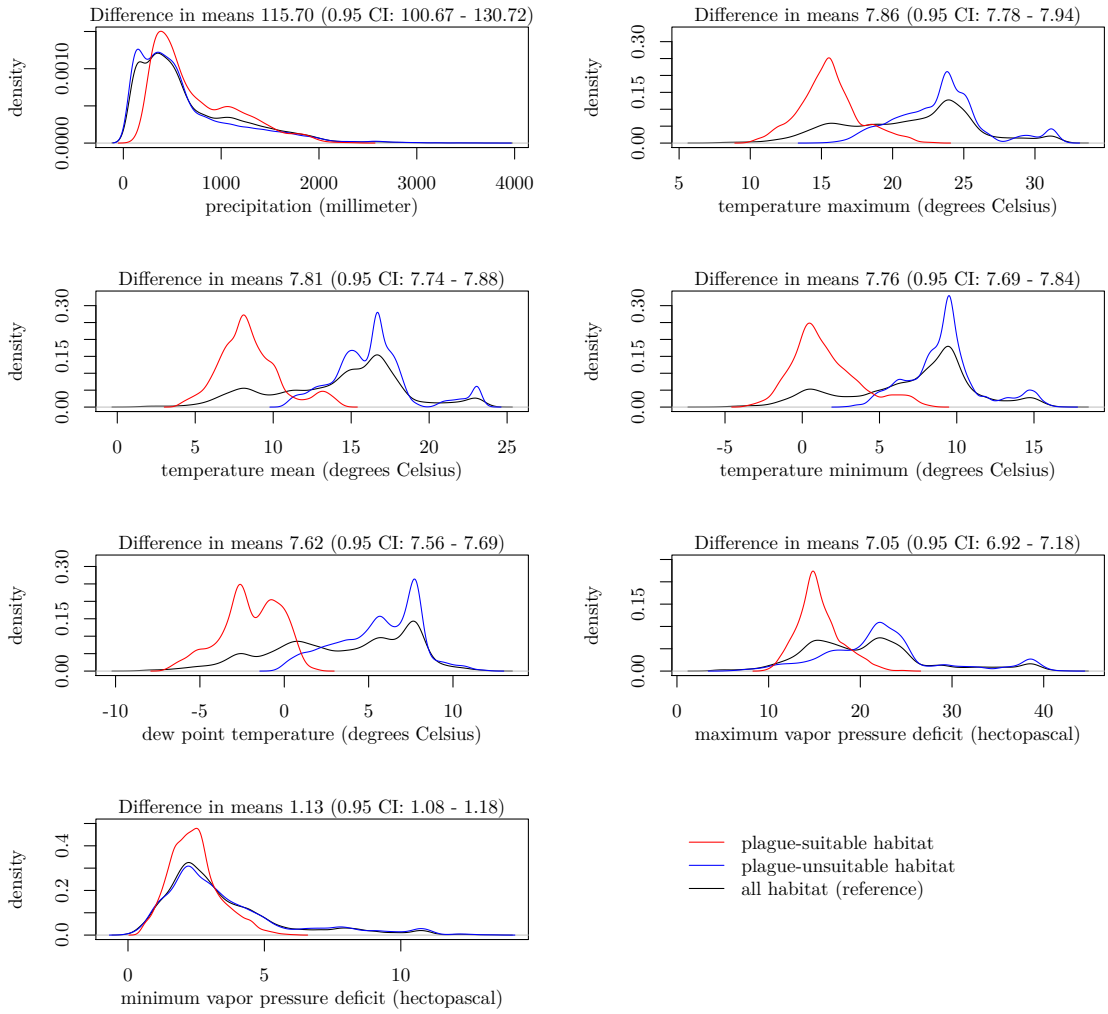


Figure 28: Univariate densities of areas in California predicted plague-suitable and plague-unsuitable with accompanying t-test difference in mean calculations. Plague-suitable habitat is defined as areas with a predicted asymptotic p-value (*219*, *221*) of the log relative risk surface below the two-tailed 0.95 significance level value (red-colored). Plague-unsuitable habitat is defined as areas with a predicted asymptotic p-value (*219*, *221*) of the log relative risk surface above the two-tailed 0.95 significance level value (blue-colored). Values were resampled from each grid cell in California. All areas of California (black-colored) are provided for reference. All environmental variables had a statistically significant (two-tailed t-test $\alpha = 0.05$) difference in means between plague-suitable and plague-unsuitable habitat. Plague-suitable habitat appears to be, on average, associated with more precipitation, cooler temperatures, and more humid conditions than plague-unsuitable habitat across California.

4 Monte Carlo assessment of the effect of positional uncertainty in animal-based plague surveillance: A case study in California, U.S.A.

"Thus we never see the true state of our condition till it is illustrated to us by its contraries, nor know how to value what we enjoy, but by the want of it."

- Robinson Crusoe in *The Life and Adventures of Robinson Crusoe* by Daniel Defoe (*248*)

4.1 Introduction

Knowledge of locations of animal hosts that can act as potential sources of disease may enhance zoonotic disease surveillance by identifying areas of potential human contact and exposure. Effective surveillance for diseases in wildlife relies on rich data from observational studies, but a common problem with these data is any uncertainty about where sampling occurred. Uncertainty can arise from error associated with data collection, collation, and digitization (*249*). Positional uncertainty in the sampling location of a specimen can also arise during the process of geocoding whereby textual descriptions of locations recorded in the field are interpreted retrospectively as an XY coordinate pair that may differ substantially from the true sampling location (*250*). During geocoding, positional uncertainty can occur for many reasons including, for example, nonspecific textual information (i.e., named places with large geographical extents), mistakes during data digitization, or poor reference data (*251*, *252*). No consensus has been reached about the degree to which positional uncertainty affects biogeographical analyses. Ecological niche models (ENMs) and species distribution models (SDMs) are common tools in biogeography used to predict the spatial distribution of a species by associating species occurrences and environmental variables (*12*). Some studies have found that positional uncertainty has a substantial impact on ENM predictions and performance (*250*, *253–257*). Other studies found that the impact was negligible or was mediated by other ecological factors or statistical artifacts (*249*, *258–262*).

In the present study, positional uncertainty in sampling location is defined as the discrepancy between the true sampling location and geocoded sampling location based on textual

descriptions from field collections. In an ecological niche modeling framework, positional uncertainty in sampling location manifests as covariate misclassification because the associations between presences (cases) and absences (controls) are dependent on the associated covariates in the analysis. In the research setting defined in the preceding chapters, variation in an observation's geocoded sampling location can result in an observation being assigned different climate covariate values than those associated with the true sampling location if the two locations are positioned in different raster cells (*251*, *263*). In geographic space, this error can be represented as an area of uncertainty with the geocoded sampling location at its center (Figure 29). Translated into predictor space, the degree of this error can vary from negligible to substantial dependent on the spatial autocorrelation of predictor values (Figure 29). The impact of positional uncertainty in sampling location on ENM predictions is likely greater in a pixel with a spatial neighborhood (i.e., the queen first-order surrounding pixels) with heterogeneous environmental variables (e.g., Sierra Nevada) than in areas with homogeneous environmental variables (e.g., Central Valley, California; *253*, *254*, *263*).

Positional uncertainty in sampling location can be detected and quantified using various techniques. Data cleaning methods can identify observations with positional uncertainty resulting from erroneous digitization (*264*). During the geocoding process, the positional uncertainty of each sampling location can be quantified and then used to categorize observations by their degree of error (*252*). Once categorized, some studies omitted highly uncertain observations from ENM analysis (*252*, *265*), but this practice can lead to small sample sizes that in turn can reduce model performance and introduce statistical bias into the analysis (*249*, *266–268*). Other techniques used a Monte Carlo simulation framework that perturbs XY coordinate pairs by defined probability densities based on an understanding of the spatial nature of the error (*269*). Variation in results based on the perturbed data sets provides assessments of the impact of positional uncertainty on predictions (*253*, *254*). However, errors in spatial data sets, including positional uncertainty, often cannot be characterized by a single accuracy metric because they are dependent on the spatial scale of analysis (i.e., more variation when pixels become smaller; *270*).

While the impact of positional uncertainty in sampling location has been compared for ENMs using real and simulated species occurrence data (*249*, *253*, *254*, *259*), the present

study aims to assess its impact on the proposed ENM method developed in Chapter 3 for a zoonotic disease. The proposed method uses case (presence) and control (absence) locations of coyotes (*Canis latrans* Say, 1823) tested for antibodies against *Yersinia pestis*, a gram-negative bacterium that causes plague. While human cases are rare in the United States, humans are highly susceptible to *Y. pestis* infection and can experience severe symptoms (*33*). The most common route of transmission of *Y. pestis* to humans in the United States is via the bite of an infected rodent flea (*33*) and human plague cases are frequently linked to recent or ongoing epizootics (i.e., disease outbreaks in animal populations; *57*). Plague surveillance systems in the United States commonly monitor rodents and rodent predators, namely coyotes, to detect regions where plague is active or recently cycling (*107*). Coyotes are not a reservoir host but function as a wide-ranging sentinel species for plague activity because coyotes are mobile rodent predators that tolerate a *Y. pestis* infection and maintain long-lasting antibodies (*91*, *109*, *112*, *118*, *146*, and see Panel 6). This leads to another source of positional uncertainty. The plague exposure location for an individual coyote (Figure 29) is largely unknown and likely different from the sampling location, which can lead to misclassification error in covariate values. However, I do not address this source of error in the present analysis and assume the exposure location and sampling location of each coyote are the same.

Here, I assess the effect of the positional uncertainty in sampling location on the prediction of the ecological niche and spatial distribution of enzootic plague in California using coyotes tested for *Y. pestis* exposure. Enzootic plague foci are areas where a sylvatic cycle (i.e., rodent-to-rodent transmission) is persistently maintaining *Y. pestis* in the environment at low-levels (*31*). Some reported coyote locations are more precise than others, especially locations with nonspecific textual information or older observations without readily available Global Positioning System technology. I chose to assess this level of uncertainty in sampling location by categorizing coyotes by their level of positional uncertainty in sampling location using an adaptation of Wieczorek and colleagues (*252*) guidelines. The effect of positional uncertainty in sampling location on estimates of the ecological niche of plague in coyotes is tested by stratifying the analysis this categorization as well as perturbing each coyote by the size of their positional uncertainty in a Monte Carlo simulation based

analysis (Figure 30). My approach assesses a source of data uncertainty that is of concern to a state-level plague surveillance system and demonstrates that positional uncertainty of coyote sampling locations may have spatially varying impact on my prediction of the spatial distribution of enzootic plague.

4.2 Data and methods

4.2.1 Surveillance data and classification of positional uncertainty

The California Department of Public Health (CDPH) Vector-Borne Disease Section has digitized coyote-based plague surveillance data across California beginning in 1983. The CDPH conducts passive surveillance of coyotes as sentinels for rodent plague activity to fill surveillance gaps in in regions of California having fewer resources as well as corroborate regional increases in plague activity indicated by rodent plague surveillance data (*108*). The CDPH tests coyote specimens for plague in conjunction with the U.S. Department of Agriculture Animal and Plant Health Inspection Service Wildlife Services depredation operations in predominantly agricultural/rural areas of suspected plague risk (i.e., plague enzootic regions) (*107*). If a direct florescent antibody test (*182*, *183*) concluded the presence of *Y. pestis* antibodies (seropositive), a coyote was considered a case, and if the test concluded the absence of *Y. pestis* antibodies (seronegative), a coyote was considered a control. Between 1983 and 2015, the CDPH screened 8,119 coyote blood samples for *Y. pestis* antibodies.

The precise sampling location was not recorded for the vast majority of coyote observations. I geocoded specimens without geographic coordinates using location descriptions provided by the collector, county data, and base maps. For example (toy examples), some coyotes were sampled near roads some distance from a nearby town (i.e., "U.S. Interstate - 5, 2 miles south of Redding, CA") or sampled on private property described by a certain distance away from a town with a cardinal heading (i.e., "5 miles northwest of Redding, CA") or sampled at exact addresses (i.e., a U.S. Forest Service campground). I was unable to locate two percent of tested coyotes (n = 164 total; n = 1 case) below the county-level and I ignore these specimens in my analysis. Of the 7,955 coyotes with geolocation informa-

tion, about 8.8% ($n = 704$ total) were seropositive. Figure 31a presents county-level CDPH coyote sampling and Figure 31b the counties with at least one coyote observed with plague antibodies (data are presented at the county-level to protect landowner privacy).

To assess the impact of uncertainty in recorded geolocation of coyote observations on the prediction of the spatial distribution of enzootic *Y. pestis*, each coyote is categorized by my confidence level in identifying its described sampling location ("Geocode Confidence Level" or GCL; Table 8). Reference towns varied by size and distances varied. Greater uncertainty is introduced with larger distances and the increasing size of the reference points (e.g., campground, small town, large city) recorded by collectors. Therefore, I developed the criteria in Table 9 to estimate the degree of positional uncertainty in sampling location of each coyote. These criteria are an adaptation of guidelines developed by Wieczorek and colleagues (*252*). Instead of a continuous metric of sampling uncertainty, I categorize coyotes based on the environmental data used in the analysis (2.5 arcminutes) because the analysis is conducted in predictor space and the positional uncertainty in sampling location here is a type of mismatch error of assigning the incorrect PC values to a coyote (see *251*, *263*). If the uncertainty around the sampling location of a coyote fell within a ~ 12.5 km^2 area (or within one PRISM raster cell), then a coyote is classified under GCL 1 or "Best Data." If the uncertainty around the sampling location of a coyote fell within a ~ 113 km^2 area (or within a queen neighborhood of a PRISM raster cell), then a coyote is classified under GCL 2 or "Good Data." If the uncertainty around the sampling location of a coyote fell outside of a ~ 113 km^2 area (or outside of a queen neighborhood of a PRISM raster cell), then a coyote is classified under GCL 3 or "Poor Data." If a coyote did not have a discernible sampling location, then it was classified under GCL 4 or "No Data."

Additionally, my categorization penalizes coyotes geocoded in areas of California with excessive variation in elevation. Areas with large changes in elevation likely have multiple habitat types compared to areas with less variation in elevation, which may lead to a mismatch in environmental variables between my estimated geocode and the true sampling location. Therefore, it is more important to accurately estimate the precise sampling location of these coyotes for the analyses. However, because the exact sampling location is not known for every coyote, the location of coyotes geocoded in areas with excessive variation

in elevation are categorized at one lower confidence level than before the penalty, except penalized coyotes in GCL 3 coyotes remained in the analysis as GCL 3. Using the Digital Terrain Elevation Data from the National Aeronautics and Space Administration Shuttle Radar Topology Mission (NASA-SRTM) (*128*), I aggregate elevation data of California from a resolution of 90 meters to 2.5 arcminutes to match the resolution of the PRISM variables. Any pixels with a standard deviation larger than 125 meters were considered to have high elevation variability (Figure 32). The NASA-SRTM data are accessed via the **raster** package (*129*) in the statistical software **R** (*127*). Over 12% of coyotes ($n = 977$ of 7,919) are located in areas of California with excessive variation in elevation (GCL 1: 288, GCL 2: 354, GCL 3: 335; Table 8).

4.2.2 Environmental data processing

Plague occurrence has been associated with climate factors (*58*, *70*, *85*, *89*, *90*). The analysis in preceding chapters employed the Oregon State University Parameter elevation Regression on Independent Slopes Model (PRISM) statistical mapping system, which follows a weighted regression framework that relies on digital elevation models built from a network of ground measurements (*125*). The 30-year average normals (1981–2010) at a 2.5 arcminute (∼4 kilometer by ∼4 kilometer; ∼16 square kilometers) resolution were selected because the temporal range of the data mostly overlapped with CDPH surveillance (1983–2015). PRISM data were accessed via the **prism** package (*126*) in the statistical software **R** (*127*) and appear in Table 10.

Environmental data sets often contain many different measurements, some highly correlated (e.g., elevation and precipitation), so in order to avoid collinearity I reduce the number of variables used in my developed approach (Chapter 3) by conducting a principal component analysis (PCA). PRISM variables are range standardized using the Gower metric (*197*) because PRISM variables are measured on different scales. A PCA is conducted using the **RStoolbox** package (*198*) in the statistical software **R** (*127*). The first two principal components (PC1 and PC2) of my environmental variables accounts for over 96% of the variance (Table 11) and are used as environmental predictors for my analysis (Figure 33).

4.2.3 Statistical methods

I use the ENM proposed in Chapter 3 for my analyses. The proposed method adapts a spatial cluster detection algorithm originally designed for spatial cancer epidemiology to identify spatial clusters of climate signatures within a predictor space. Predictor space comprises two dimensions that are the first and second principal component of the PCA described above. These signatures are then predicted across California, even into areas that are not historically sampled or where data are unavailable as a mechanism for environmental interpolation (*19*). My proposed method is based on a kernel density estimation of spatial relative risk (*211*, *212*) that treats the observed point locations of cases and controls as realizations of spatial Poisson processes. An output of the model is a log relative risk surface (i.e., a smooth surface linking the ratio of estimated spatial density of cases to that of controls; *211–216*). Plague is more likely to be observed in areas with a log relative risk value above zero than areas with a log relative risk value below zero. Another output of the model is a significance level value of a local two-tailed asymptotic tolerance calculation (*219*, *221*) at two significance levels ($\alpha = 0.05$ & $\alpha = 0.01$). An asymptotic tolerance significance level value below 0.025 is considered "plague-suitable habitat" where plague is likely active in a sylvatic cycle and an asymptotic tolerance significance level value above 0.975 is considered "plague-unsuitable habitat" where plague is likely not active in a sylvatic cycle. The proposed method also identifies areas within the predictions with sparse observation data.

To examine the impact of positional uncertainty in sampling location on my prediction of the spatial distribution of enzootic plague, I use my proposed ENM in two separate frameworks. First, coyote observations are stratified by geocode confidence level (i.e., "Best Data," "Good Data," and "Poor Data" levels, separately). I also conduct a cumulative analysis (i.e., "Best Data," "Best Data" and "Good Data," and "All Data"). The "All Data" model is comparable to the model conducted in Chapter 3. Models are compared quantitatively via the average Area Under the Receiver Operating Characteristic Curve (AUC) of the predicted log relative risk values from a 5-fold cross-validation within stratum (*222*). Second, I conduct a Monte Carlo simulation-based assessment of the impact of

location uncertainty on estimates of the ecological niche of plague in coyotes. "Best Data," "Good Data," and "Poor Data" levels are spatially displaced uniformly within an area with a radius of two kilometers, six kilometers, and ten kilometers, respectively, roughly based on the observed location uncertainty observed in each class. I compare these findings to a model where all coyotes are spatially displaced uniformly within an area with a radius of two kilometers. Each spatially displaced analysis has 1,000 iterations. Any observation spatially displaced beyond the extent of the study space (i.e., California state boundary) is assigned the PC values of the nearest raster pixel. Average log relative risk values and average asymptotic tolerance significance level values are calculated for each pixel within a predictor space with its dimensions comprised of the two chosen PCs. Measures of variation are also calculated including the standard deviation of log relative risk and the proportion of iterations each pixel falls outside of the 95% tolerance interval for null relative risk. These statistics are then transformed to geographic space within California.

I use a fixed-bandwidth kernel density estimator to estimate the log relative risk surface with a smoothing parameter (i.e., bandwidth) chosen with the maximal smoothing principle (*229*). The bandwidth size is neither restricted across stratified models nor across permutations in the spatially displaced analysis. The smoothing parameters chosen via the maximal smoothing principle (*229*) for each model are similar. I do not conduct a sensitively analysis of the smoothing parameter because the prediction was found to be robust across various bandwidth sizes in a previous ENM analysis using the same data (see Chapter 3). All analyses are conducted in the statistical software **R** (*127*).

4.3 Results

Statistically significant relative clustering of seropositive coyotes (cases; $n = 704$) and seronegative coyotes (controls; $n = 7251$) by the CDPH (1983–2015) was detected in predictor space for all kernel density estimation-based models. After penalization, there were 1800 observations of the highest geocode confidence level (i.e., "Best Data") with a plague prevalence of 9%, 3509 observations of the moderate geocode confidence level (i.e., "Good Data") with a plague prevalence of 8.1%, and 2645 observations of the lowest geocode confidence level (i.e., "Poor Data") with a plague prevalence of 9.7% (Table 8).

4.3.1 Stratified by geocode confidence level

In general, predicted suitable habitat areas of plague in California were similar across models using observations stratified by geocode confidence level (Figure 34). Differences between models were most evident in Siskiyou County, the Northern Coastal Ranges, and the foothills on the the western slope of the Northern Sierra Nevada. The "Poor Data" model predicted more plague-unsuitable habitat in the Mojave Desert than the other models but did not predict the northern California coast as plague-unsuitable habitat unlike the other models. The "Best Data" model predicted the least amount of plague-suitable habitat in California compared to other models, especially in the Peninsular Ranges and high elevation areas of the Sierra Nevada. The spatial coverage of coyote observations by the CDPH was similar across stratum (Figure 35). The highest frequency of "Best Data" were recorded in the northeastern region of California while the highest frequency of "Good Data" and "All Data" were recoded in the southwestern and south central regions of the state, respectively (Figure 35). No one model performed statistically superior to others as each average AUC value within stratum were captured by the 95% confidence intervals of the three models (Figure 36).

Results were similar when the analysis was conducted cumulatively by geocode confidence level (Figure 37) with an increase in spatial coverage of coyote observations (Figure 38). The analytically chosen bandwidths of each model were similar but smaller in models with larger sample sizes, as one might expect. No one model performed statistically superior to the others but the confidence intervals of AUC within stratum were smaller in models with larger sample sizes (Figure 39). The "All Data" model predicted more plague-suitable habitat in the high elevation areas of the northern Sierra Nevada and more plague-unsuitable habitat in the Mojave Desert than other models. The model using "Best Data" and "Good Data" predicted plague-unsuitable habitat in some high alpine areas of the southern Sierra Nevada. Additionally, the model with the largest sample size ("All Data") was able to predict the most area of California as either plague-suitable or plague-unsuitable. The models could not predict into areas such as Death Valley and parts of the Mojave Desert where there were scarce observations.

4.3.2 Spatially displaced permutations

Across 1,000 permutations where observations were spatially displaced uniformly within an area with a radius based on each class of geocode confidence (i.e., two kilometers, six kilometers, and ten kilometers), statistically significant relative clustering of seropositive coyotes and seronegative coyotes was detected in predictor space. The average log relative risk appears in Figure 40a and the standard deviation of the log relative risk appears in Figure 40b. The average asymptotic tolerance significance level value appears in Figure 41a and the proportion of iterations that were statistically significant appears in Figure 41b. The highest variation in log relative risk and significance level values were located in and around areas of predictor space with few or no observation data. The regions of predictor space with variation in predicted values were also located at the edges of statistically significant relative clusters. The log relative risk values were unstable in data sparse areas of predictor space (i.e., intersection of high PC1 values and low PC2 values).

The estimated values in predictor space were transformed to geographic space within California (Figures 42 and 43). Plague-suitable habitat was primarily located in mountainous regions of California, but not in high alpine zones. Plague-unsuitable habitat was predicted in the Central Valley as well as coastal and desert regions of California. Areas in the state with a higher variation in the significance level values were located in subalpine regions and foothills such as the northern foothills of the Transverse Ranges. Results were similar to the prediction of the "All Data" model in the stratified analysis (Figure 37).

Estimates of average log relative risk and average significance level values were similar when observations were spatially displaced uniformly within a circular area with a two-kilometer radius (Figures 44 and 45) or the other radii under consideration (Figures 40 and 41). While the shape of the significant seronegative clustering was similar, more areas of the significant seronegative clustering fell under the higher significance level ($\alpha = 0.05$ versus $\alpha = 0.01$) in the varying radii analysis than in the fixed radius analysis. This translated to more desert regions of California being predicted plague-unsuitable at a higher significance level in the varying radii analysis (Figure 43a) versus at a lower significance level in the same radius analysis (Figure 47a). There was higher variation in the log relative risk (Figure 42)

and asymptotic tolerance significance level values (Figure 43) in the desert regions and the northern California coastline in the varying distance analysis than in the same distance analysis. Percent change in log relative risk between the same distance analysis and the varying distance analysis appear in Figure 48a. Areas sensitive to higher levels of positional uncertainty in the sampling location appear in Figure 48b.

4.4 Discussion

Predictions of areas in California with climatological conditions associated with the occurrence of a zoonotic pathogen (*Y. pestis*) using an ecological niche modeling framework were robust to positional uncertainty in sampling location of observations of a sentinel species (*C. latrans*) captured by an animal plague surveillance system. Models stratified by the degree of geocode confidence of coyote observations performed similarly, but some areas predicted plague-suitable or plague-unsuitable shifted between models, locally. Predictions across models were similar, in general, likely due to the large sample size of the study, even when stratified. However, local sample size varied among stratum, so positional uncertainty in the sampling location was introduced into an ENM by randomly displacing all geocoded observations iteratively in a Monte Carlo simulation based analysis. The average predictions of the Monte Carlo simulation based ENM analysis closely matched the predictions of the ENM that used all data without spatially displacing observations. The negligible, global effect yet drastic, local effect of positional uncertainty in sampling location on predictions from my proposed method (Chapter 3) is similar to that found in various studies (*249*, *259–261*). Mitchell and colleagues (*261*) found areas sensitive to positional uncertainty in sampling location were at the fringe of optimal habitat.

The analysis identified some areas of California where the predicted spatial distribution of enzootic plague was sensitive to positional uncertainty in sampling location of observations, namely the edges of the seropositive clustering and the seronegative clustering in predictor space. This was expected because the relative risk ratio along the edges of the clustering are closer to the null value (zero) and therefore more likely non-significant than the center of the clustering. As coyote observations were spatially displaced, in some iterations, they may have been located in a neighboring pixel that has principal component values further away

from the center of the clustering and the iteration may have considered that combination of PC values to be in the significant clustering. For example, few coyotes were tested in areas with PC values in between the two seronegative clusters (see Figure 41a). and when observations were spatially displaced, they may have been randomly located in between these two clusters for some iterations (see Figure 41b). In California, areas sensitive to the positional uncertainty in sampling location were zones with climates in between plague-suitable and plague-unsuitable habitat, especially at the higher significance level ($\alpha = 0.05$). For example, the foothills of the Peninsular Ranges were sensitive to the positional uncertainty in sampling location. The peaks of these ranges as well as dry, low elevation areas in the southern Central Valley were strongly predicted as plague-suitable, while the low-lying deserts were strongly predicted as plague-unsuitable. It is imperative that any plague observation in these zones has a precise sampling location (or exposure location), that is, the results suggest that positional uncertainty may have spatially varying impact; an interesting area for future research.

The magnitude of the effect of positional uncertainty in sampling location on predictions from raster-based analyses can be mediated by the degree of spatial autocorrelation in predictors (*253*, *254*, *263*). The present study did not directly test the interaction between spatial autocorrelation in predictors and positional uncertainty in sampling location. However, areas sensitive to positional uncertainty in sampling location were found in areas of elevation gradients that likely have different climates within close geographical proximity (i.e., low spatial autocorrelation). More specifically, these areas were located along elevation gradients where the ENM was unable to statistically distinguish between plague-suitable and plague-unsuitable habitat because seropositive and seronegative coyotes occurred relatively equal in these types of areas.

Restricting an ENM for enzootic plague only to the highest quality location data may well overlook areas of suitable habitat for plague transmission. Coyotes located in areas of California with excessive variation in elevation were penalized in the geocode confidence criteria which resulted in more "Good Data" and "Poor Data" observations located in mountainous regions of California than "Best Data." This may account for why more area was considered plague-suitable in the "Good Data" and "Poor Data" models than the "Best

Data" model (e.g., the Klamath Mountains, namely around Mt. Shasta and Mt. Eddy). Using only "Poor Data" in an ENM had less area predicted as significantly plague-suitable habitat but had more area predicted as plague-unsuitable habitat in the Mojave Desert than a model of only "Good Data" because more "Poor Data" were observed in the desert regions than "Good Data." Omitting the highest quality location data in future analyses would be inconceivable in real-world disease surveillance; thus, a cumulative approach using both "Best Data" and "Good Data" is more realistic. Indeed, a model using "Best and Good Data" predicted areas as plague-suitable similar to a model using all observations regardless of geocoding confidence (i.e. "All Data"). The addition of "Poor Data" in an ENM resulted in more areas to be predicted as plague-unsuitable habitat, except along the northern Pacific Coast. More "Poor Data" were located in Humboldt county, but sample size was low overall. This suggests the analysis was not sensitive to areas or climatological conditions with low sampling or low seroprevalence. The spatially displaced analysis using various radii based on the geocode confidence determining the northern Pacific coastline was sensitive to the positional uncertainty in sampling location. Given historically sparse rodent data (see *89*) and relatively poor coyote data, the northern Pacific coastline might be an area of interest for improved surveillance to determine where and how plague is being maintained.

There were several strengths of the study. The approach was computationally inexpensive even for a Monte Carlo based simulation and thus can provide results readily for ongoing disease surveillance systems. I was able to identify what types of uncertainty exist in the coyote plague system, categorize my observations by the degree of positional uncertainty in sampling location, and then used a Monte Carlo simulation based approach to assess the impact of positional uncertainty in sampling location on my predictions. Monte Carlo simulation based approaches are a general tool to assess the degree of sensitivity to location uncertainty (*253*, *254*, *269*). Additionally, by stratifying the observation data by geocode confidence level the sample size varied between models. Small sample sizes can reduce ENM performance and introduce statistical bias (*249*, *266–268*), even though the seroprevalence of *Y. pestis* antibodies was similar across stratified models (8%–10%). Incorporating all data of various geocode confidences did not negatively affect the prediction of the spatial distribution of enzootic plague in California, even when including observations

with uncertain sampling locations (i.e., "Poor Data"). Instead, incorporating all data of various geocode confidence levels allowed for the selection of a smaller smoothing parameter that could better detect the edge of an ecological niche in predictor space than larger smoothing parameters. With the addition of reservoir host data, namely rodent species, this precision can greatly enhance an ENMs ability to identify disease enzootic refugia (*271*) that maintain plague in the environment during inter-epizootic cycles (*17*).

The problem of scale (*244*) was a notable limitation for the study. The spatial scale of the observational process (i.e., sampling and testing coyotes) was not the same as the spatial scale of the climate variables used in the analysis. Coyotes were sampled in a precise location albeit not recorded at that level of detail. The analysis was conducted at the 2.5 arcminute resolution (∼16 square kilometers) to capture the average climate at the location a coyote was observed. The geocode confidence criteria were designed to align these two processes by assigning coyotes based on their uncertainty within a pixel or its neighborhood at the 2.5 arcminute resolution. This spatial resolution balanced statistical tractability, the precision of annual average climate values at each coyote sampling location, and the representation of the unknown true plague exposure location of each coyote. The location where a coyote was exposed to *Y. pestis* is challenging to discern (*91*, *117*) because coyotes are mobile and encounter many rodents across their expansive home range (*144*, *145*, *184*, and see Panel 6). Although the observed location of a coyote is often considered the true location of plague exposure in plague surveillance systems (*47*, *85*, *89*, *95*, *113–116*), the inaccuracy between the sampling location of a coyote and it true disease exposure location is of concern (Figure 29). A "typical" home range of a coyote is about 25 km^2 (up to 80 km^2; *144*) and an average dispersal range of 40 km (*145*). However, coyote home range, activity, and density are influenced by human development (*153–156*, *272*), latitude (*273*), and sex (*150*), but see (*157*). Here, the PRISM variables at a 2.5 arcminute resolution are locally smoothed (averaged) values that more likely captured both the sampling and exposure location in the same or nearby pixel than higher resolution climate data. Conducting the analysis at a higher resolution would provide more precise climate values at each sampling location but may create a larger discrepancy between the sampling location and true exposure location than a lower resolution analysis. In addition, one individual coyote often cannot indicate

plague activity where it was sampled, but on the aggregate, coyotes indicate plague activity. With a large sample size ($\sim n = 8{,}000$) the central tendency of the ecological niche of plague in coyotes is elucidated as long they coyotes are sampled evenly and thoroughly across the study region. Chapter 5 shows coyotes are not sampled evenly across the study region, including mountainous regions where rodents are better detectors for plague activity and the ecological niche of plague than coyotes. Once the ecological niche of plague is well defined using a myriad of host species observations, future studies can assess the effect of the positional uncertainty of the exposure location for individual coyotes on the predicted ecological niche of plague.

My penalty for coyotes located in pixels with highly variable elevation (i.e., pixels more likely to have multiple habitat types) can be improved. These coyotes were penalized because their true observation location may be, for example, located in a riparian zone in a valley floor (*274*) that may be distinct from the climate values averaged at a 2.5 arcminute resolution. These penalized observations were classified as having a lower geocode confidence level and, in the spatially displaced analysis, were more likely to be assigned the PC values of a neighboring gridded pixel than if the observation was a higher geocode confidence level. While this approach may have accounted for the variability of climate across a neighborhood of gridded pixels, the approach did not account for the variability of climate within a gridded pixel. The PC values of a neighboring pixel were likely not the same values as a unique habitat within a pixel because neighboring pixels in geographic space likely have similar PC value, unless they are in regions of California with large changes in elevation (e.g., mountains). Conducting the analysis at a higher resolution would have more precise PC values for the sampling location and fewer coyotes would have been penalized as lower quality observations because the higher resolution PRISM variables have smaller elevation variability within each cell than lower resolution variables. However, as previously mentioned, a higher resolution analysis may contain a larger discrepancy between the sampling location and true exposure location than a lower resolution analysis. A future sensitivity analysis can assess the combined effects of spatial scale on the prediction of the ecological niche in coyotes and spatial distribution of enzootic plague in California.

The circular, uniformly random spatial displacing may not be realistic of the posi-

tional uncertainty in the sampling location. A circular shape was chosen for computational tractability. Coyotes observed along linear topographical features (e.g., roads or streams) were treated similarly to coyotes not observed along linear topographical features (e.g., campground, park, or private ranch). For observations along linear topographical features, a circle was not the best representation of their positional uncertainty and the analysis potentially heightened the positional uncertainty in sampling location of these observations. A more informed shape of positional uncertainty can be used in future analyses. Spatial displacement was assumed to be a random uniform distribution, which assesses the effect of uninformed positional uncertainty in the sampling location on the predictions. A future investigation can assess other distributions such as, for example, a bivariate normal distribution centered at the geocoded sampling location (*269*) or an informed spatial displacing process that uses additional covariates (e.g., presence of a riparian corridor).

Effective zoonotic disease surveillance relies on observational studies collecting rich data, but real-world disease surveillance systems often have gaps or uncertainties. Data quality is an important consideration for wildlife disease surveillance, especially for passive surveillance systems (*159*) such as coyote plague surveillance. Certain analytical measures can be taken to overcome types of data uncertainties that balance statistical rigor and biological realities. Here, poorer quality data were not an impediment to the prediction of the ecological niche of *Y. pestis* in coyotes or its spatial distribution in California. However, omitting data based on quality issues can lead to smaller samples sizes and local sampling effect biases for predictive models. The approach can help reform ongoing CDPH plague surveillance protocols to record precise sampling location information in statistically-sensitive areas of California. The methodological framework defining and assessing the positional uncertainty in sampling location has great promise for other wildlife and zoonotic disease systems that have data quality concerns from field observations.

4.5 References

12. A. T. Peterson *et al.*, *Ecological niches and geographic distributions (MPB-49)* (Princeton University Press, Princeton, New Jersey, 2011), vol. 56, ISBN: 0691136882.

17. R. J. Eisen, K. L. Gage, *Veterinary Research* **40**, 1, ISSN: 1993-5412, DOI 10.1051/vetres:2008039 (2009).

19. A. T. Peterson, *Mapping disease transmission risk: enriching models using biogeography and ecology* (Johns Hopkins University Press, Baltimore, Maryland, 2014), ISBN: 1421414737.

31. K. L. Gage, M. Y. Kosoy, *Annual Review of Entomology* **50**, 505–528, ISSN: 0066-4170, DOI 10.1146/annurev.ento.50.071803.130337 (2005).

33. K. J. Kugeler *et al.*, *Emerging Infectious Diseases* **21**, 16, ISSN: 1080-6059, 1080-6040, DOI 10.3201/eid2101.140564 (2015).

47. J. L. Lowell *et al.*, *Journal of Vector Ecology* **34**, 22–31, ISSN: 1081-1710, DOI 10.1111/j.1948-7134.2009.00004.x (2009).

57. H. E. Brown *et al.*, *American Journal of Tropical Medicine and Hygiene* **82**, 95–102, ISSN: 0002-9637, 1476-1645, DOI 10.4269/ajtmh.2010.09-0247 (2010).

58. T. Ben-Ari *et al.*, *PLoS Pathogens* **7**, e1002160, ISSN: 1553-7374, 1553-7366, DOI 10.1371/journal.ppat.1002160 (2011).

70. Y. Nakazawa *et al.*, *Vector-Borne and Zoonotic Diseases* **7**, 529–540, ISSN: 1557-7759, 1530-3667, DOI 10.1089/vbz.2007.0125 (2007).

85. S. P. Maher *et al.*, *American Journal of Tropical Medicine and Hygiene* **83**, 736–742, ISSN: 0002-9637, 1476-1645, DOI 10.4269/ajtmh.2010.10-0042 (2010).

89. A. C. Holt *et al.*, *International Journal of Health Geographics* **8**, 38, ISSN: 1476-072X, DOI 10.1186/1476-072X-8-38 (2009).

90. M. Walsh, M. A. Haseeb, *PeerJ* **3**, e1493, ISSN: 2167-8359, DOI 10.7717/peerj.1493 (2015).

91. D. J. Salkeld, P. Stapp, *Vector-Borne and Zoonotic Diseases* **6**, 231–239, ISSN: 1557-7759, 1530-3667, DOI 10.1089/vbz.2006.6.231 (2006).

95. H. E. Brown *et al.*, *Vector-Borne and Zoonotic Diseases* **11**, 1439–1446, ISSN: 1557-7759, 1530-3667, DOI 10.1089/vbz.2010.0196 (2011).

107. S. N. Bevins *et al.*, *Integrative Zoology* **7**, 99–109, ISSN: 1749-4877, DOI 10.1111/j.1749-4877.2011.00277.x (2012).

108. J. R. Tucker *et al.*, *California compendium of plague control*, 2015, (2018; https://www.cdph.ca.gov/Programs/CID/DCDC/CDPH%20Document%20Library/CAPlagueCompendium.pdf).

109. K. L. Gage *et al.*, in *Proceedings of the Sixteenth Vertebrate Pest Conference*, pp. 200–206.

112. P. W. Willeberg *et al.*, *American Journal of Epidemiology* **110**, 328–334, ISSN: 0002-9262, DOI 10.1093/oxfordjournals.aje.a112818 (1979).

113. C. U. Thomas, P. E. Hughes, *Journal of Wildlife Diseases* **28**, 610–613, ISSN: 0090-3558, DOI 10.7589/0090-3558-28.4.610 (1992).

114. N. W. Dyer, L. E. Huffman, *Journal of Wildlife Diseases* **35**, 600–602, ISSN: 0090-3558, DOI 10.7589/0090-3558-35.3.600 (1999).

115. A. Malmlov *et al.*, *Journal of Wildlife Diseases* **50**, 946–950, ISSN: 1943-3700, DOI 10.7589/2014-03-065 (2014).

116. B. R. Hoar *et al.*, *Preventive Veterinary Medicine* **56**, 299–311, ISSN: 0167-5877, DOI 10.1016/S0167-5877(02)00194-0 (2003).

117. R. J. Brinkerhoff *et al.*, *Vector-Borne and Zoonotic Diseases* **9**, 491–497, ISSN: 1557-7759, 1530-3667, DOI 10.1089/vbz.2008.0075 (2009).
118. L. A. Baeten *et al.*, *Journal of Wildlife Diseases* **49**, 932–939, ISSN: 0090-3558, DOI 10.7589/2013-02-040 (2013).
125. C. Daly *et al.*, *International Journal of Climatology* **28**, 2031–2064, ISSN: 1097-0088, DOI 10.1002/joc.1688 (2008).
126. E. M. Hart, K. Bell, *prism: download data from the Oregon PRISM project*, R package version 0.0.6, DOI 10.5281/zenodo.33663, (http://github.com/ropensci/prism).
127. R Core Team, *R: a language and environment for statistical computing*, R Foundation for Statistical Computing (Vienna, Austria, 2018), (https://www.R-project.org/).
128. J. J. van Zyl, *Acta Astronautica* **48**, 559–565, ISSN: 0094-5765, DOI 10.1016/S0094-5765(01)00020-0 (2001).
129. R. J. Hijmans, *raster: geographic data analysis and modeling*, R package version 2.6-7, (https://CRAN.R-project.org/package=raster).
144. R. M. Nowak, *Walker's mammals of the world* (Johns Hopkins University Press, Baltimore, Maryland, 1999), vol. 1, ISBN: 0801857899.
145. V. Howard Jr, G. G. DelFrate, in *Great Plains Wildlife Damage Control Workshop Proceedings* (University of Nebraska-Lincoln, Lincoln, Nebraska, 1991), pp. 39–49.
146. A. M. Barnes, *Symposia of the Zoological Society of London* **50**, 237–270 (1982).
150. M. Bekoff, *Mammalian Species*, 1–9, ISSN: 0076-3519, DOI 10.2307/3503817 (1977).
153. N. McClennon *et al.*, *American Midland Naturalist* **146**, 27–36, ISSN: 0003-0031, 1938-4238, DOI 10.1674/0003-0031(2001)146[0027:TEOSAA]2.0.CO;2 (2001).
154. S. D. Gehrt *et al.*, *Journal of Mammalogy* **90**, 1045–1057, ISSN: 0022-2372, DOI 10.1644/08-MAMM-A-277.1 (2009).
155. V. M. Lukasik, S. M. Alexander, *Cities and the Environment (CATE)* **4**, 8, ISSN: 1932-7048, DOI 10.15365/cate.4182011 (2011).
156. J. M. Fedriani *et al.*, *Ecography* **24**, 325–331, ISSN: 0906-7590, DOI 10.1034/j.1600-0587.2001.240310.x (2001).
157. M. Murray *et al.*, *Proceedings of the Royal Society B: Biological Sciences* **282**, 20150009, ISSN: 0962-8452, DOI 10.1098/rspb.2015.0009 (2015).
159. S. M. Nusser *et al.*, *Journal of Wildlife Management* **72**, 52–60, ISSN: 0022-541X, DOI 10.2193/2007-317 (2008).
171. R. Kays, *Canis latrans*, 2018, (10.2305/IUCN.UK.2018-2.RLTS.T3745A103893556.en).
182. T. J. Quan *et al.*, in *Diagnostic procedures for bacterial, mycotic, and parasitic infections*, ed. by A. Barlows, W. J. Hausler (American Public Health Association, Washington, D.C., ed. 6, 1981), pp. 723–745, ISBN: 0875530869.
183. B. W. Hudson *et al.*, *Epidemiology & Infection* **60**, 443–450, ISSN: 0950-2688, DOI 10.1017/S002217240002057X (1962).
184. J. C. C. Neale, B. N. Sacks, *Oikos* **94**, 236–249, ISSN: 0030-1299, DOI 10.1034/j.1600-0706.2001.940204.x (2001).
185. G. A. Feldhamer *et al.*, Eds., *Wild mammals of North America: biology, management, and conservation* (Johns Hopkins University Press, Baltimore, Maryland, ed. 2, 2003), ISBN: 0801874165.
197. J. C. Gower, *Biometrics*, 857–871, ISSN: 1541-0420, 0006-341X, DOI 10.2307/2528823 (1971).
198. B. Leutner *et al.*, *RStoolbox: tools for remote sensing data analysis*, R package version 0.2.3, (https://CRAN.R-project.org/package=RStoolbox).

211. J. E. Kelsall, P. J. Diggle, *Statistics in Medicine* **14**, 2335–2342, ISSN: 1097-0258, DOI 10.1002/sim.4780142106 (1995).
212. J. E. Kelsall, P. J. Diggle, *Bernoulli* **1**, 3–16, ISSN: 1573-9759, 1350-7265, DOI 10.2307/3318678 (1995).
213. D. C. Wheeler, *International Journal of Health Geographics* **6**, 13, ISSN: 1476-072X, DOI 10.1186/1476-072X-6-13 (2007).
214. H. E. Clough *et al.*, *Preventive Veterinary Medicine* **89**, 67–74, ISSN: 0167-5877, DOI 10.1016/j.prevetmed.2009.01.008 (2009).
215. J. F. Bithell, *Statistics in Medicine* **10**, 1745–1751, ISSN: 1097-0258, DOI 10.1002/sim.4780101112 (1991).
216. J. F. Bithell, *Statistics in Medicine* **9**, 691–701, ISSN: 1097-0258, DOI 10.1002/sim.4780090616 (1990).
219. M. L. Hazelton, T. M. Davies, *Biometrical Journal* **51**, 98–109, ISSN: 1521-4036, 0323-3847, DOI 10.1002/bimj.200810495 (2009).
221. T. M. Davies, M. L. Hazelton, *Statistics in Medicine* **29**, 2423–2437, ISSN: 1097-0258, DOI 10.1002/sim.3995. (2010).
222. A. T. Peterson *et al.*, *Ecological Modelling* **213**, 63–72, ISSN: 0304-3800, DOI 10.1016/j.ecolmodel.2007.11.008 (2008).
229. G. R. Terrell, *Journal of the American Statistical Association* **85**, 470–477, ISSN: 0162-1459, 1537-274X, DOI 10.2307/2289786 (1990).
244. S. A. Levin, *Ecology* **73**, 1943–1967, ISSN: 0012-9658, DOI 10.2307/1941447 (1992).
248. D. Defoe, *The life and adventures of robinson crusoe* (Project Gutenberg, Urbana, Illinois, 1996), (2019; http://www.gutenberg.org/ebooks/521).
249. C. H. Graham *et al.*, *Journal of Applied Ecology* **45**, 239–247, ISSN: 1365-2664, DOI 10.1111/j.1365-2664.2007.01408.x (2008).
250. K. J. Feeley, M. R. Silman, *Journal of Biogeography* **37**, 733–740, ISSN: 0305-0270, DOI 10.1111/j.1365-2699.2009.02240.x (2010).
251. P. A. Zandbergen *et al.*, *Spatial and Spatio-temporal Epidemiology* **3**, 69–82, ISSN: 1877-5845, DOI 10.1016/j.sste.2012.02.007 (2012).
252. J. Wieczorek *et al.*, *International Journal of Geographical Information Science* **18**, 745–767, ISSN: 1365-8824, 1365-8816, DOI 10.1080/13658810412331280211 (2004).
253. B. Naimi *et al.*, *Journal of Biogeography* **38**, 1497–1509, ISSN: 0305-0270, DOI 10.1111/j.1365-2699.2011.02523.x (2011).
254. B. Naimi *et al.*, *Ecography* **37**, 191–203, ISSN: 0906-7590, DOI 10.1111/j.1600-0587.2013.00205.x (2014).
255. C. J. Johnson, M. P. Gillingham, *Ecological Modelling* **213**, 143–155, ISSN: 0304-3800, DOI 10.1016/j.ecolmodel.2007.11.013 (2008).
256. V. Moudrỳ, P. Šímová, *International Journal of Geographical Information Science* **26**, 2083–2095, ISSN: 1365-8824, 1365-8816, DOI 10.1080/13658816.2012.721553 (2012).
257. S. J. Tulowiecki *et al.*, *Plant Ecology* **216**, 67–85, ISSN: 1385-0237, 1573-5052, DOI 10.1007/s11258-014-0417-9 (2015).
258. M. Fernandez *et al.*, *Biodiversity Informatics* **6**, ISSN: 1546-9735, DOI 10.17161/bi.v6i1.3314 (2009).
259. P. E. Osborne, P. J. Leitão, *Diversity and Distributions* **15**, 671–681, ISSN: 1366-9516, DOI 10.1111/j.1472-4642.2009.00572.x (2009).
260. M. A. Hayes *et al.*, *Acta Chiropterologica* **17**, 159–169, ISSN: 1508-1109, 1733-5329, DOI 10.3161/15081109ACC2015.17.1.013 (2015).

261. P. J. Mitchell *et al.*, *Methods in Ecology and Evolution* **8**, 12–21, ISSN: 2041-210X, DOI 10.1111/2041-210X.12645 (2017).
262. A. Soultan, K. Safi, *PLoS One* **12**, e0187906, ISSN: 1932-6203, DOI 10.1371/journal.pone.0187906 (2017).
263. C. Perez-Heydrich *et al.*, *Spatial Demography* **4**, 135–153, ISSN: 2364-2289, 2164-7070, DOI 10.1007/s40980-015-0013-1 (2016).
264. R. J. Hijmans *et al.*, *Genetic resources and crop evolution* **46**, 291–296, ISSN: 0925-9864, 1573-5109, DOI 10.1023/A:1008628005016 (1999).
265. Q. Guo *et al.*, *International Journal of Geographical Information Science* **22**, 1067–1090, ISSN: 1365-8824, 1365-8816, DOI 10.1080/13658810701851420 (2008).
266. P. A. Hernandez *et al.*, *Ecography* **29**, 773–785, ISSN: 0906-7590, DOI 10.1111/j.0906-7590.2006.04700.x (2006).
267. A. S. van Proosdij *et al.*, *Ecography* **39**, 542–552, ISSN: 0906-7590, DOI 10.1111/ecog.01509 (2016).
268. J. M. McPherson *et al.*, *Journal of Applied Ecology* **41**, 811–823, ISSN: 1365-2664, DOI 10.1111/j.0021-8901.2004.00943.x (2004).
269. G. B. M. Heuvelink *et al.*, *International Journal of Geographical Information Science* **21**, 497–513, ISSN: 1365-8824, 1365-8816, DOI 10.1080/13658810601063951 (2007).
270. G. B. Heuvelink, in *Uncertainty in remote sensing and GIS*, ed. by P. Atkinson, G. Foody (John Wiley & Sons, Inc., Hoboken, New Jersey, 2002), pp. 155–165, ISBN: 0470844083, 0470035269, DOI 10.1002/0470035269.
271. G. Keppel *et al.*, *Global Ecology and Biogeography* **21**, 393–404, ISSN: 1466-8238, DOI 10.1111/j.1466-8238.2011.00686.x (2012).
272. S. P. Riley *et al.*, *Conservation Biology* **17**, 566–576, ISSN: 0888-8892, DOI 10.1046/j.1523-1739.2003.01458.x (2003).
273. M. E. Gompper, J. L. Gittleman, *Oecologia* **87**, 343–348, ISSN: 0029-8549, 1432-1939, DOI 10.1007/BF00634589 (1991).
274. J. A. Hilty, A. M. Merenlender, *Conservation Biology* **18**, 126–135, ISSN: 0888-8892, DOI 10.1111/j.1523-1739.2004.00225.x (2004).

4.6 Appendices

4.6.1 Appendix A: Panels

Panel 6: Considerations for coyotes as a sentinel species for plague activity

Coyotes (*Canis latrans*) have been used as a sentinel species of plague (*Yersinia pestis*) activity in rodent populations of North America (*95*, *112–116*), but there are limitations for focusing on coyote-based plague surveillance data. I outline some notable limitations below:

- Coyotes are a wide-ranging species (*144*, *145*, *171*) and effectively sample a broad variety of habitats in which plague occurs and thus sample a considerable extent of the ecological niche of plague (*85*). However, coyotes are not observed homogeneously across North America (Chapter 5) and other species may be better indicators of plague activity in some habitats (e.g., rodent species for alpine regions of California). Therefore, without rodent data, I am estimating a the fundamental ecological niche of plague in coyotes, which I am considering as an approximation for the entire fundamental ecological niche of plague.
- The location where an individual coyote was truly exposed to *Y. pestis* is unknown and challenging to discern (*91*, *117*). Coyotes are mobile, encountering many rodents across their expansive individual home range (*144*, *145*, *184*) and whose plague antibodies can last for many months (*91*, *112*, *118*). Reinfection of coyotes is also probable and so a coyote can only indicate its most recent plague exposure. Rodent species have smaller individual home range sizes (*185*) and are the gold standard indicator of current or recent plague activity (*108*).
- Coyote specimens are not collected evenly across the United States and instead are collected opportunistically (*107*). Coyotes tested for plague exposure are primarily collected in conjunction with ongoing livestock/wildlife damage management operations conducted by the U.S. Department of Agriculture Animal and Plant Health Inspection Service Wildlife Services. Agricultural and urban areas may be more sampled for plague than other land-use types and their respective habitats. Including other species monitored for plague activity may help overcome this potential spatial sampling effort bias.

4.6.2 Appendix B: Tables

Table 8: Sample size of observations across geocode confidence levels following the criteria found in Table 9

Geocode Confidence Level	Plague Result	Count before penalty	Count after penalty
1	Positive	224	162
"Best Data"	Negative	1864	1638
2	Positive	273	285
"Good Data"	Negative	3302	3224
3	Positive	207	257
"Poor Data"	Negative	2085	2388
4	Positive	1	1
"No Data"	Negative	163	163
Total	Positive	705	705
	Negative	7414	7414
	Total	8119	8119

Table 9: Geocode criteria for animal-based plague surveillance in California

Geocode Confidence Level*	Criterion†
1	A specific location identifier ($\leq$ 16 km^2 area), such as:
	Precise location, small town, or medium town without a directional
	Precise location with a $\leq$ 6.8 km directional
	Small town with a $\leq$ 3.4 km directional
	Road section $\leq$ 5.7 km long
	Other defined location $\leq$ 16 km^2 in area (e.g., small recreational area)
2	A less specific location identifier (Between 16 km^2 and 144 km^2 area), such as:
	Precise location with a directional between 6.8 km and 20.5 km
	Small town with a directional between 4.3 km and 17.1 km
	Medium town with a directional $\leq$ 13.7 km
	Large town without a directional
	Road section between 5.7 km and 17 km long
	Other defined location between 16 km^2 and 144 km^2 (e.g., large reservoir)
3	An unspecific location identifier ($>$ 144 km^2 area), such as:
	Precise location with a directional $>$ 20.5 km
	Small town with a directional $>$ 17.1 km
	Medium town with a directional $>$ 13.7 km
	Large town with a directional
	Extra large town with or without a directional
	Road section $>$ 17 km long
	Other defined location $>$ 144 km^2 (e.g., Lake Tahoe)
4	No location identifier, such as:
	Uninterpretable location description

*If a geocoded specimen is located in a grid cell (4 km x 4 km) with variable elevation (standard deviation of elevation greater than 125 meters calculated using the Shuttle Radar Topography Mission digital elevation model aggregated from 90 m x 90 m resolution to 4 km x 4 km resolution), then the specimen is penalized and dropped one geocode confidence level (e.g., from Geocode Confidence Level 1 to Geocode Confidence Level 2). Geocode Confidence Level 3 is the lowest classification (i.e., a specimen cannot drop from Geocode Confidence Level 3 to Geocode Confidence Level 4). Refer to Figure 32 for areas of California that are penalized for subsequent analysis.
†Precise location: $\leq$ 0.4 km^2; Small town: $\leq$ 4 km^2; Medium town: 4 km^2–16 km^2; Large town: 16 km^2–144 km^2; Extra large town: $>$ 144 km^2

Table 10: Oregon State University Parameter elevation Regression on Independent Slopes Model (PRISM) 30-Year Average Annual Normals (1981–2010). Modeled using a combination of a digital elevation model (DEM) and climatologically-aided interpolation (CAI)*. Variables were modeled at 30 arcseconds (∼800 m) resolution and aggregated to 2.5 arcminutes (∼4 km). See (*125*) for more details.

Variable	Units	Derivation
Precipitation	millimeters (mm)	Modeled; Summing monthly averages (rain + melted snow)
Maximum Temperature	°Celsius (°C)	Modeled; Averaging over all months using a DEM as the predictor grid
Mean Temperature	°Celsius (°C)	Derived; Average of Maximum Temperature and Minimum Temperature
Minimum Temperature	°Celsius (°C)	Modeled; Averaging over all months using a DEM as the predictor grid
Mean Dewpoint Temperature	°Celsius (°C)	Modeled; CAI used minimum temperature as the predictor grid
Maximum Vapor Pressure Deficit	hectopascal (hPA)	Modeled; CAI used mean dewpoint temperature and maximum temperature as the predictor grids
Minimum Vapor Pressure Deficit	hectopascal (hPA)	Modeled; CAI used mean dewpoint temperature and minimum temperature as the predictor grids

*Accuracy of these data is based on the original specification of the Defense Mapping Agency one-degree DEMs. The stated accuracy of the original DEMs is 130-meter circular error with 90% probability. Data sets use all weather stations, regardless of time of observation.

Table 11: Summary of principal component analysis of Oregon State University Parameter elevation Regression on Independent Slopes Model (PRISM) 30-Year Average Annual Normals (1981–2010) at a 2.5 arcminute (∼4 km) resolution. Variables were range standardized using the Gower metric (*197*).

Statistic	**Variable**	**Principal Component**		
		PC1	PC2	PC3
Standard Deviation		0.326	0.170	0.058
Proportion of Variance		0.756	0.205	0.024
Cumulative Proportion of Variance		0.756	0.962	0.985
Loadings	Precipitation	0.063	0.342	0.549
	Maximum Temperature	0.485	-0.208	-0.290
	Mean Temperature	0.519	-0.064	0.015
	Minimum Temperature	0.477	0.088	0.314
	Mean Dewpoint Temperature	0.446	0.532	-0.101
	Maximum Vapor Pressure Deficit	0.243	-0.533	-0.229
	Minimum Vapor Pressure Deficit	0.073	-0.510	0.673

4.6.3 Appendix C: Figures

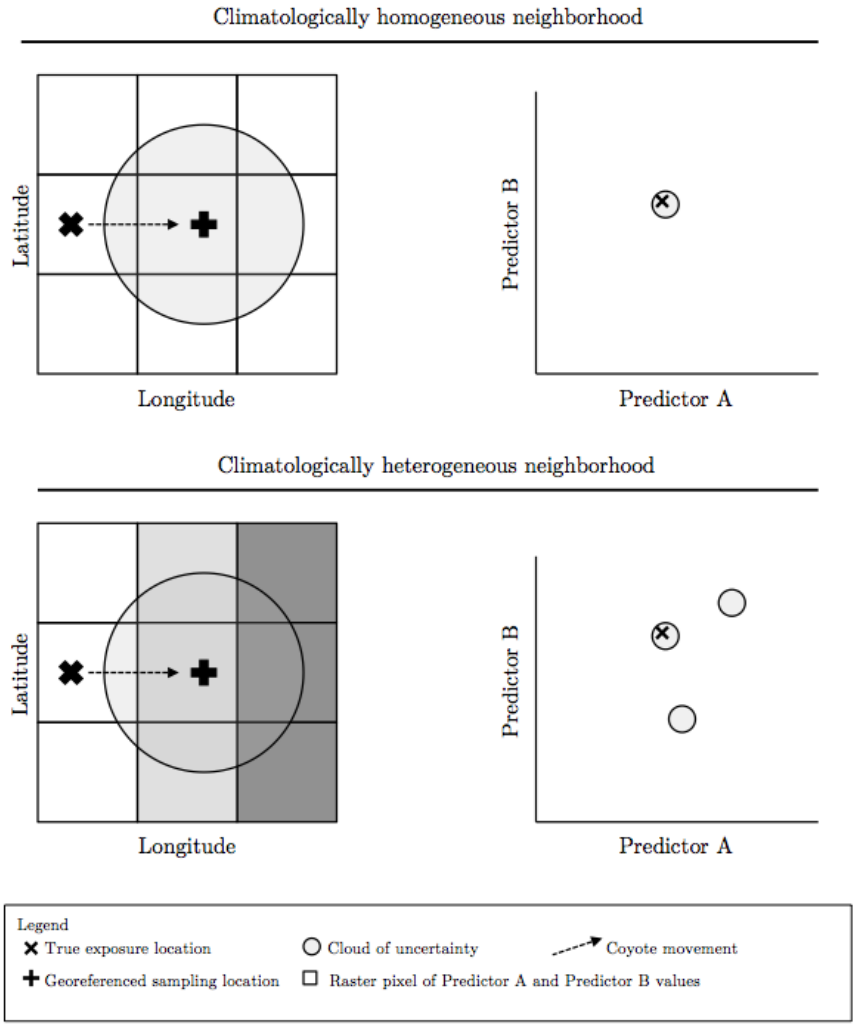


Figure 29: The two types of positional uncertainty for *Yersinia pestis* surveillance using coyote specimen locations without precise geographical information in two types of environments. Geocoding efforts using location descriptions from collectors in the field have uncertainty for the true location of the specimen sampling location represented by the "cloud of uncertainty." Areas sensitive to this uncertainty are climatologically heterogeneous (e.g., foothills) because the cloud of uncertainty contains dissimilar climates and the geocoding of a specimen could greatly impact the climate values assigned to the observation in the ecological niche modeling approach. Areas with homogeneous climates (e.g., floor of the Central Valley, California) are less sensitive to this type of uncertainty because the assigned climate value for a geocoded coyote does not vary greatly within the cloud of uncertainty. The present analysis examined the effect of positional uncertainty in the sampling location on the predicted spatial distribution of enzootic *Y. pestis* in California. The other type of observation uncertainty in this system is the true location where a coyote was exposed to *Y. pestis* and the true location where a coyote was observed may not be in the same location due to coyote movement. Future studies can examine the effect of positional uncertainty in the exposure location on the predicted spatial distribution of enzootic *Y. pestis*.

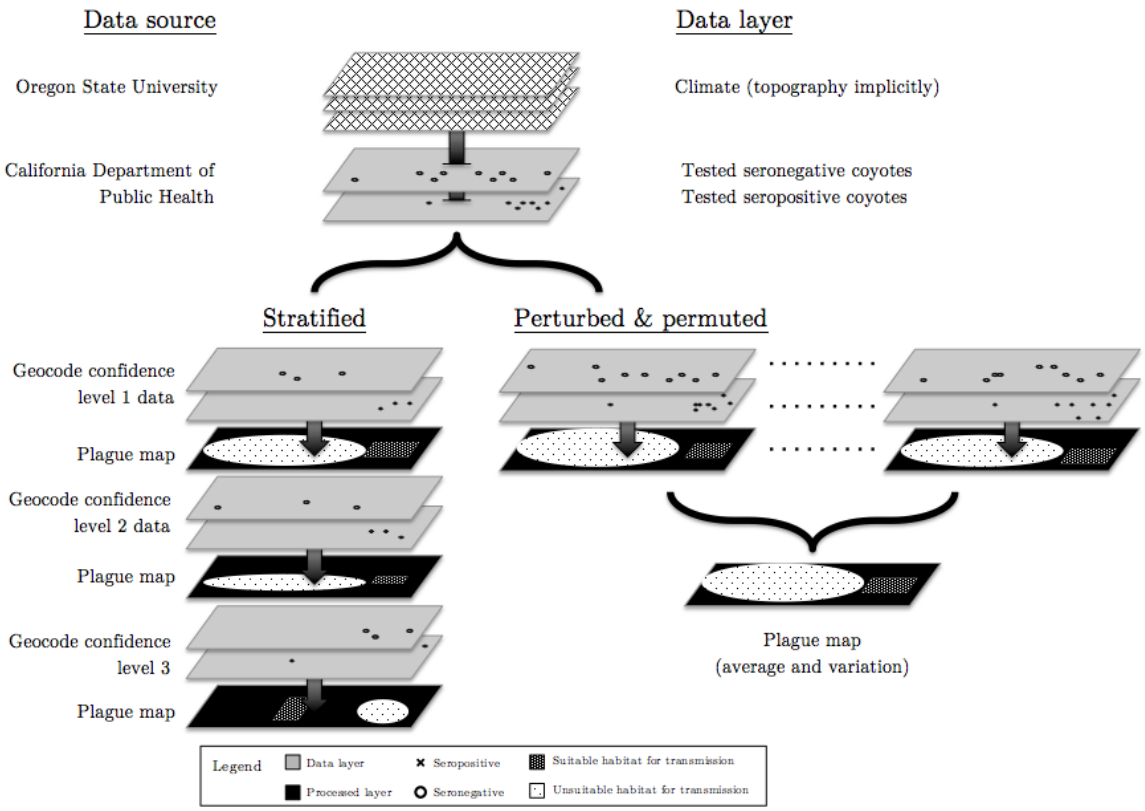


Figure 30: Data layers of proposed method and their sources. Data are structured as point locations and raster grids. Data layers (grey color) are used as inputs for the proposed method to predict the spatial distribution of enzootic *Yersinia pestis* in California (black color). The arrow represents the layers involved for the predicted layer.

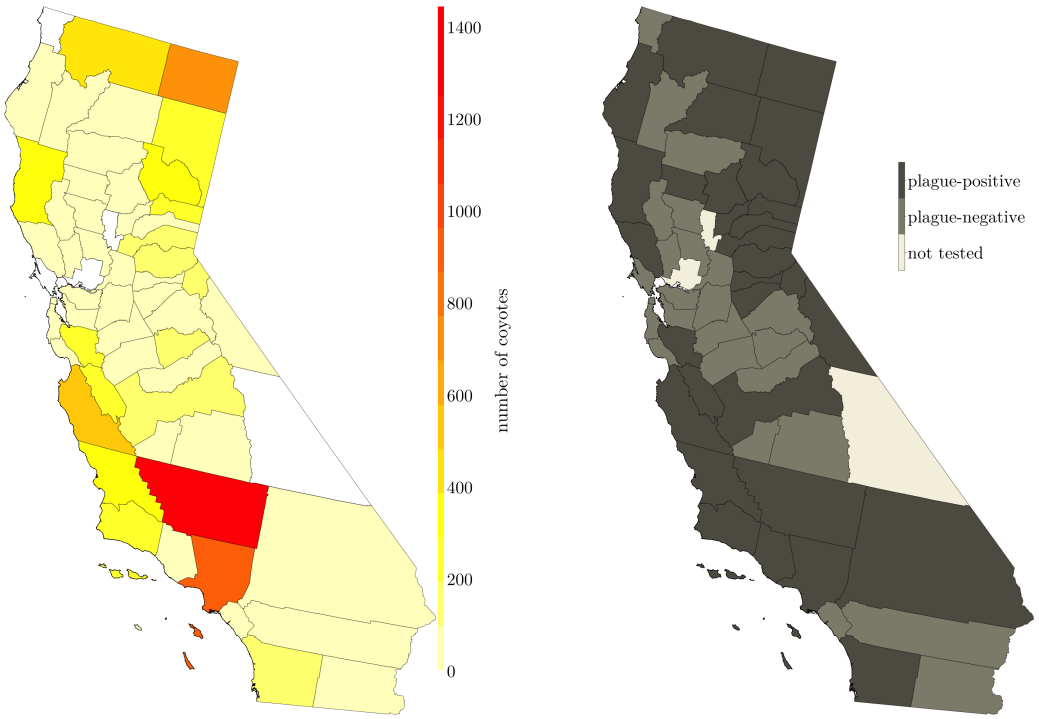


Figure 31: **(a)** Number of coyotes tested by California Department of Public Health (CDPH) for antibodies against *Yersinia pestis* (1983–2015). Sampling is heterogeneous across the state of California, primarily in historic plague enzootic areas and areas of high concern. Limited sampling in the Mojave and Sonoran Deserts. **(b)** Results of coyotes tested by the CDPH for antibodies against *Y. pestis* (1983–2015). A coyote with plague antibodies was not observed in every county sampled. Only 2% of observations were unable to be geocoded and ignored in subsequent analysis.

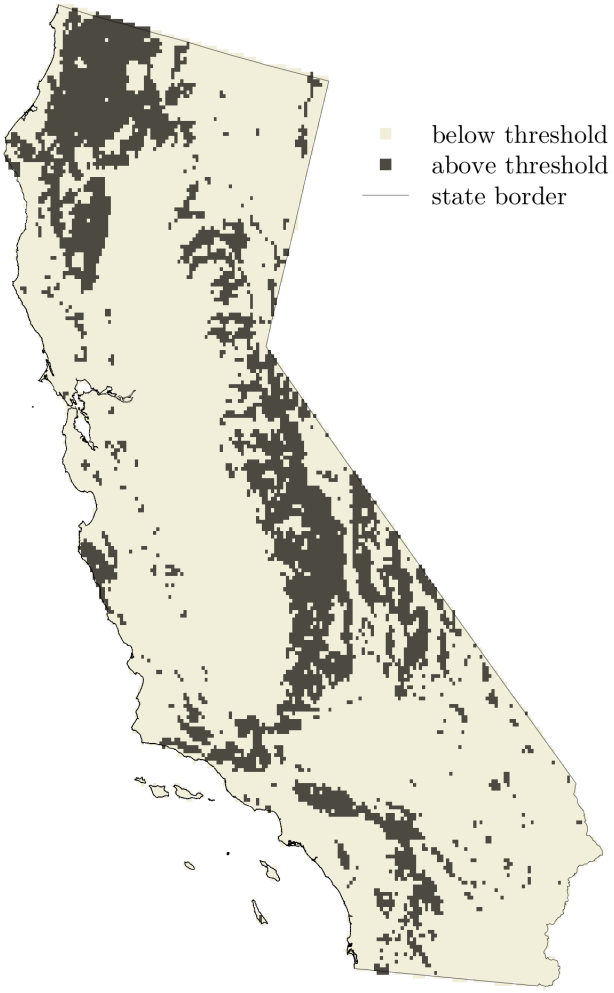


Figure 32: Internal elevation heterogeneity of California when aggregated to 4 km resolution. Areas with excessive elevation heterogeneity have a standard deviation larger than 125 meters (colored black). These areas are mountainous regions, steep coastline, or marked by canyons and likely have multiple habitat types within the 16 km^2 area. Any coyote observed in these areas have penalized geocode confidence levels.

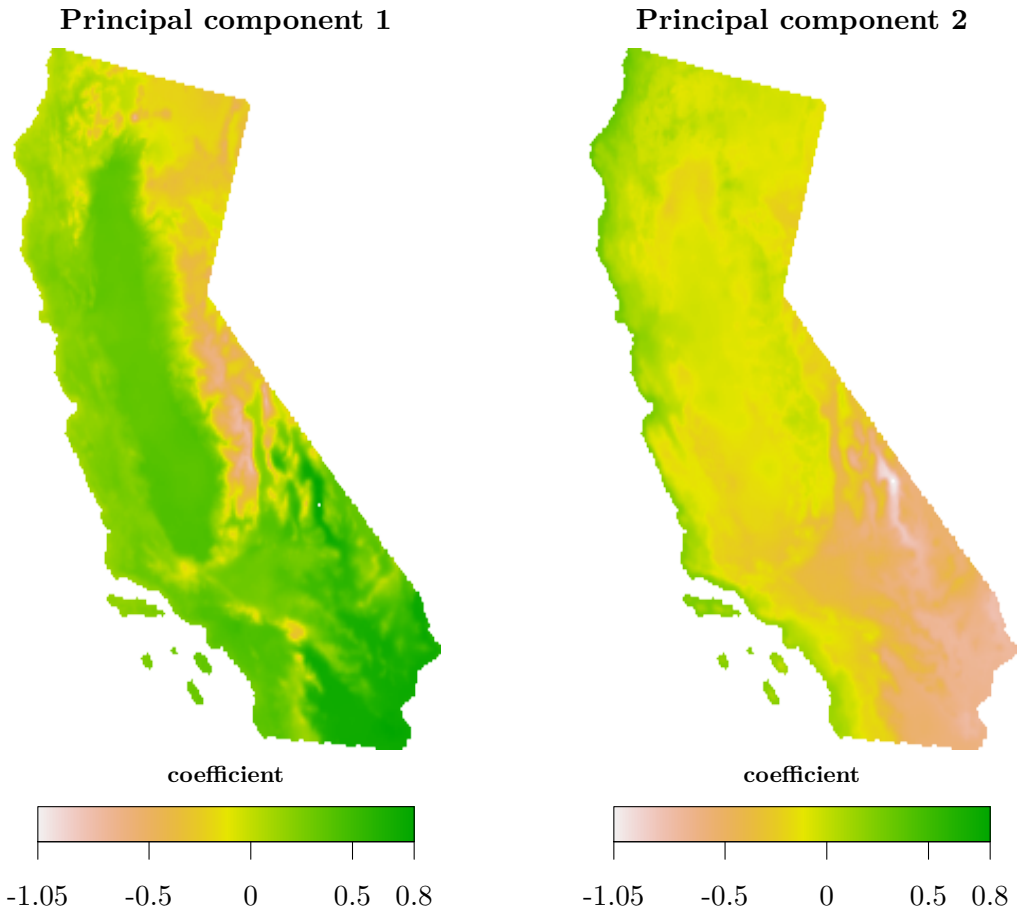


Figure 33: First two principal components of a principal component analysis of Oregon State University Parameter elevation Regression on Independent Slopes Model (PRISM) 30-year average annual normals (1981–2010) at a 2.5 arcminute (∼4 km) resolution and presented for California. Variables were range standardized using the Gower metric (*197*).

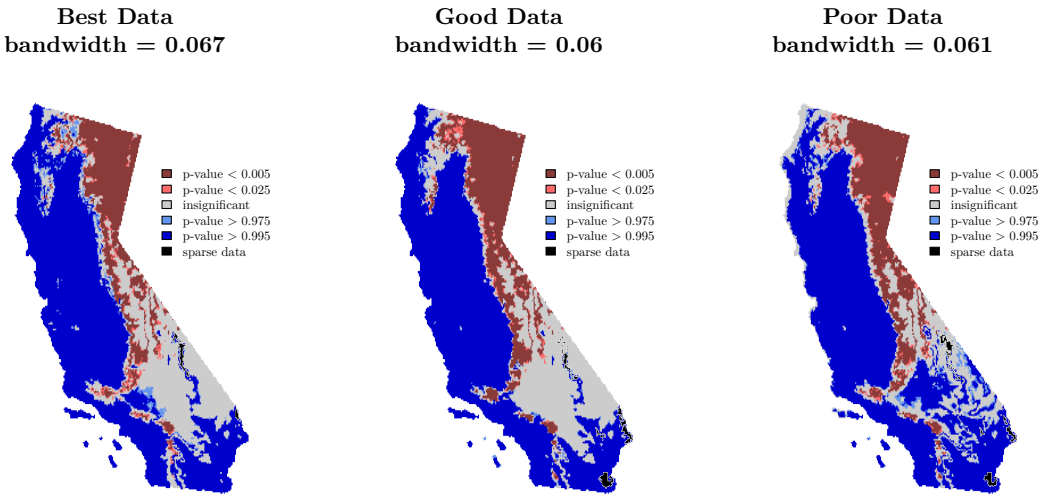


Figure 34: Comparison of observational data quality (Best v. Good v. Poor) on the predicted area of California with a log relative risk surface in predictor space of coyotes that tested negative for plague antibodies (controls) and tested positive for plague antibodies (cases) by the California Department of Public Health (1983–2015) using the developed approach. Color pertains to calculated asymptotic tolerance significance level (*219*, *221*) at given two-tailed significance levels ($\alpha = 0.05$ & $\alpha = 0.01$). Warmer-colored areas are more likely suitable for plague transmission and cooler-colored areas are more likely habitat for plague transmission with grayer-colored areas statistically indistinguishable. The areas in black coloring correspond to habitat with sparse data.

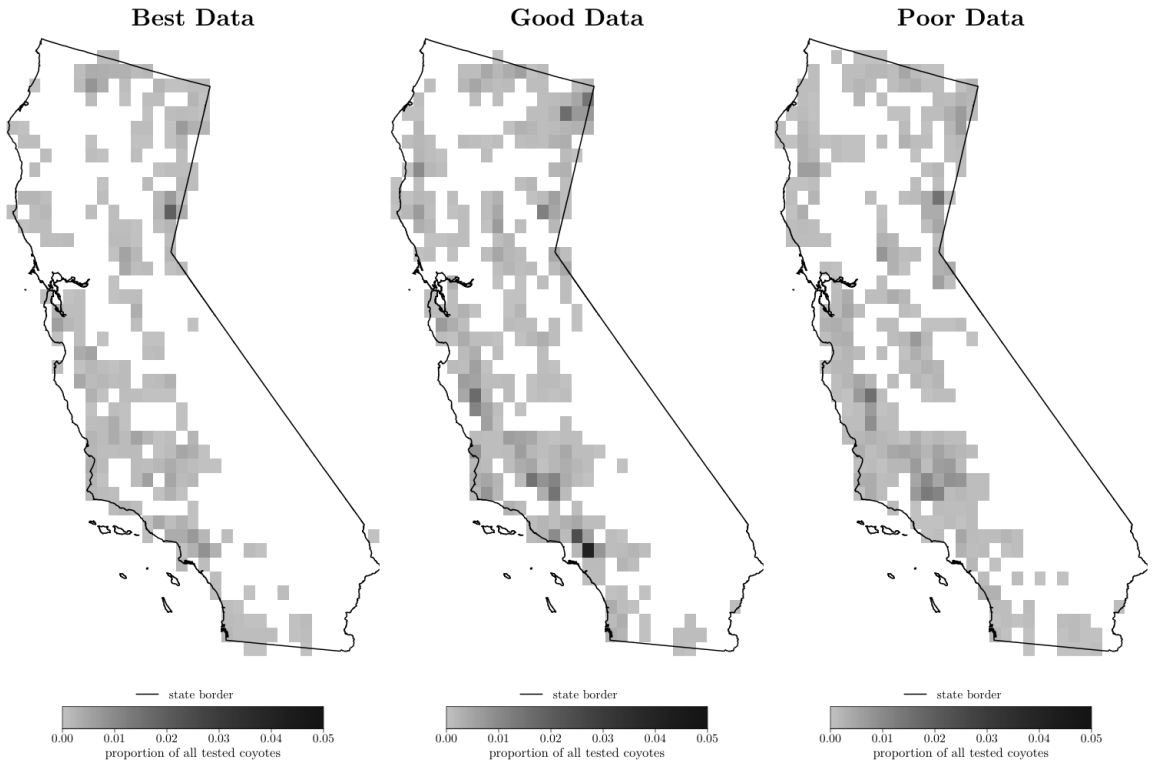


Figure 35: Local sampling effort of coyotes tested by the California Department of Public Health (1983–2015) for the presence of *Yersinia pestis* antibodies stratified by geocode confidence level (separately) following the criteria found in Table 9. Data are aggregated to protect landowner privacy.

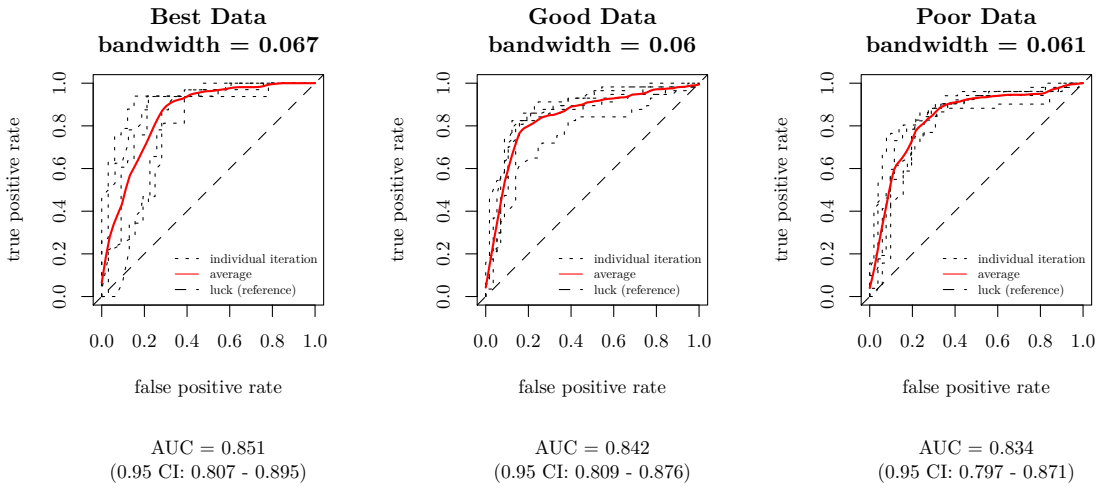


Figure 36: Results of 5-fold cross-validation of data quality (Best v. Good v. Poor) comparison for the log relative risk surface in predictor space of coyotes that tested negative for plague antibodies (controls) and tested positive for plague antibodies (cases) by the California Department of Public Health (1983–2015) using the developed approach. Iterations were balanced (prevalence = 0.5) by randomly undersampling control locations used for in each fold for cross-validation. The prediction is robust across bandwidth selection with similar Area Under the Receiver Operating Characteristic Curve (AUC) across bandwidths.

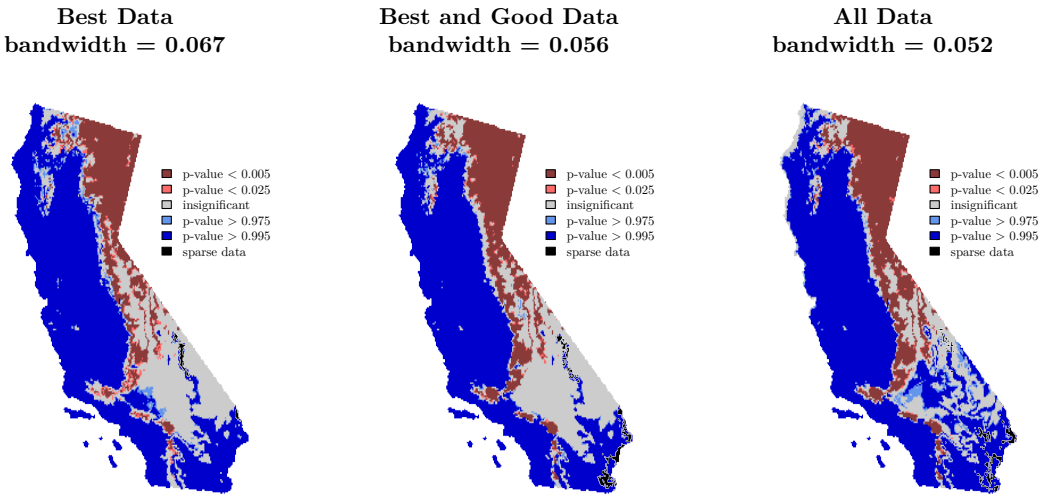


Figure 37: Comparison of observational data quality (Best v. Best & Good v. All) on the predicted area of California with a log relative risk surface in predictor space of coyotes that tested negative for plague antibodies (controls) and tested positive for plague antibodies (cases) by the California Department of Public Health (1983–2015) using the developed approach. Color pertains to calculated asymptotic tolerance significance level value (*219*, *221*) at given two-tailed significance levels ($\alpha = 0.05$ & $\alpha = 0.01$). Warmer-colored areas are more likely suitable for plague transmission and cooler-colored areas are more likely unsuitable for plague transmission with grayer-colored areas statistically indistinguishable. The areas in black coloring correspond to habitat with sparse data.

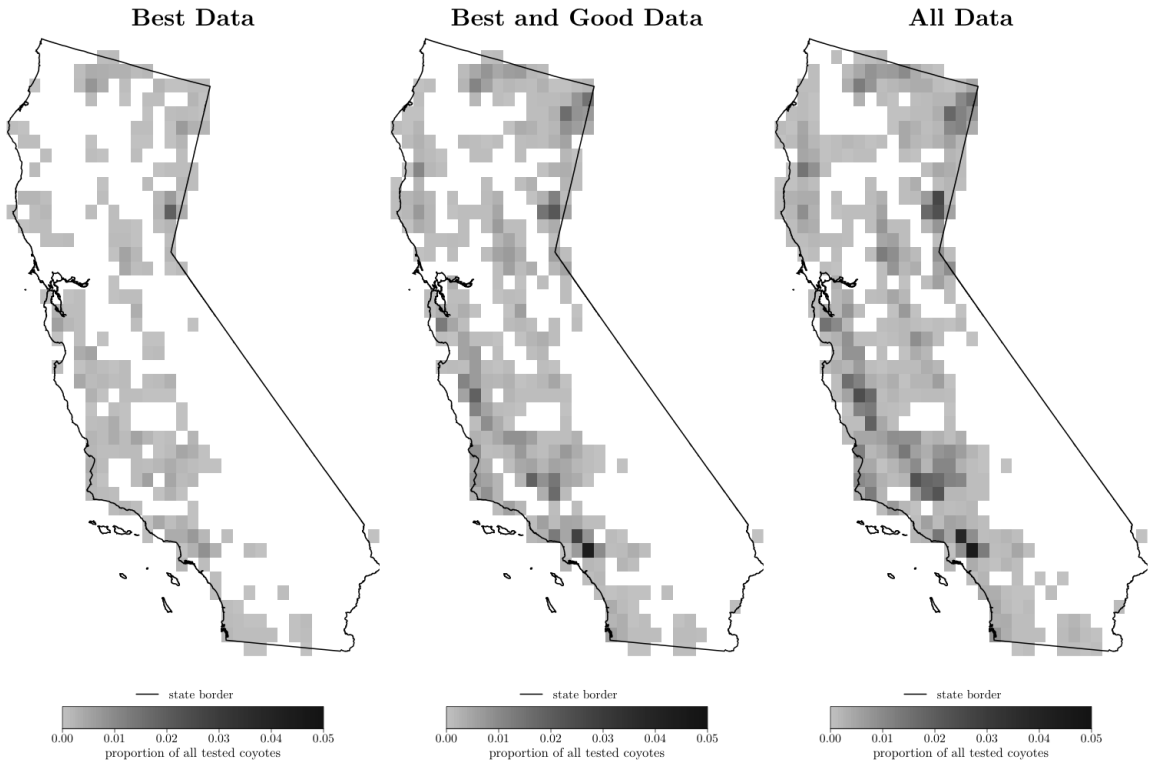


Figure 38: Local sampling effort of tested by the California Department of Public Health (1983–2015) for the presence of *Yersinia pestis* antibodies stratified by geocode confidence level (cumulatively) following the criteria found in Table 9. Data are aggregated to protect landowner privacy.

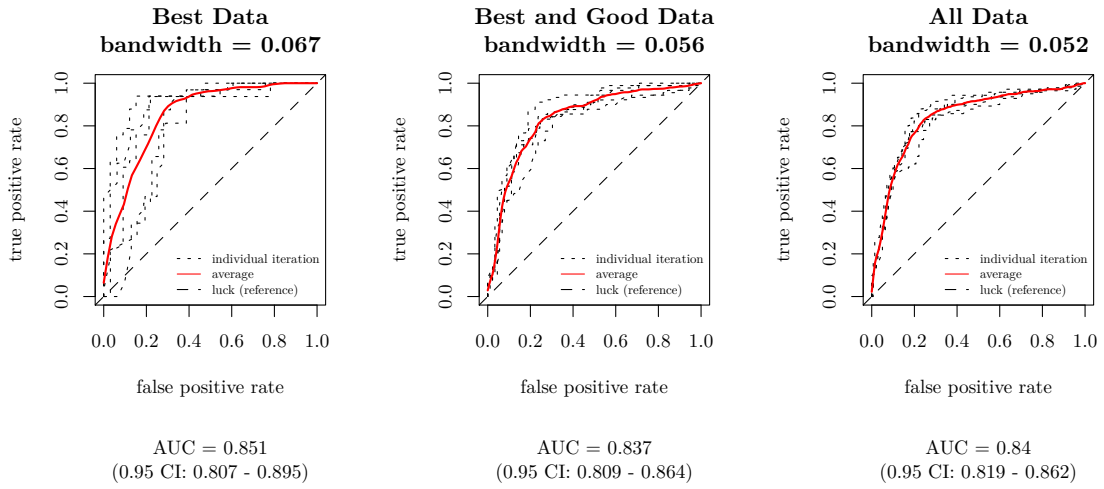
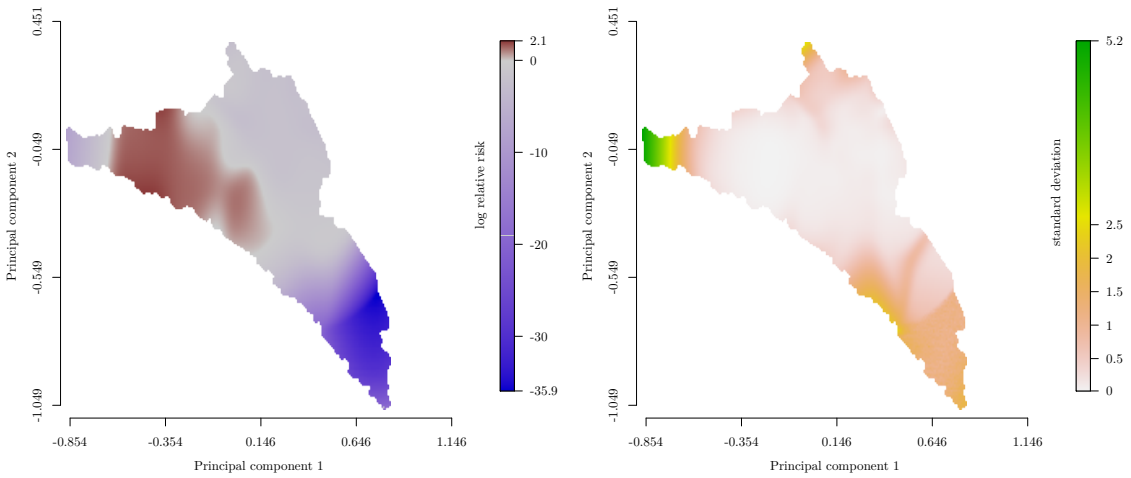


Figure 39: Results of 5-fold cross-validation of data quality (Best v. Best & Good v. All) comparison for the log relative risk surface in predictor space of coyotes that tested negative for plague antibodies (controls) and tested positive for plague antibodies (cases) by the California Department of Public Health (1983–2015) using the developed approach. Iterations were balanced (prevalence = 0.5) by randomly undersampling control locations used for in each fold for cross-validation. The prediction is robust across bandwidth selection with similar Area Under the Receiver Operating Characteristic Curve (AUC) across bandwidths.



(a) **(b)**

Figure 40: **(a)** The average log relative risk surface in predictor space comparing coyotes that tested negative for plague antibodies (controls) and tested positive for plague antibodies (cases) by the California Department of Public Health (1983–2015) using the developed approach, which randomly displaced observation locations uniformly within an area specified by the criteria in Table 9 (n = 1000 iterations). Predictor space is comprised of the first two principal components of a principal component analysis of seven range-standardized (*197*) Oregon State University Parameter Elevation Regression on Independent Slopes Model 30-year average annual normals (1981–2010) at a 2.5 arcminute (∼4 km) resolution. Each bandwidth was chosen using the maximal smoothing principal (*229*). Color pertains to average log relative risk values where positive log relative risk (more likely suitable for plague transmission) are in red and negative log relative risk (more likely unsuitable for plague transmission) with grayer coloring closer to the null log relative risk value (zero). **(b)** Color pertains to the standard deviation of the log relative risk values across random iterations.

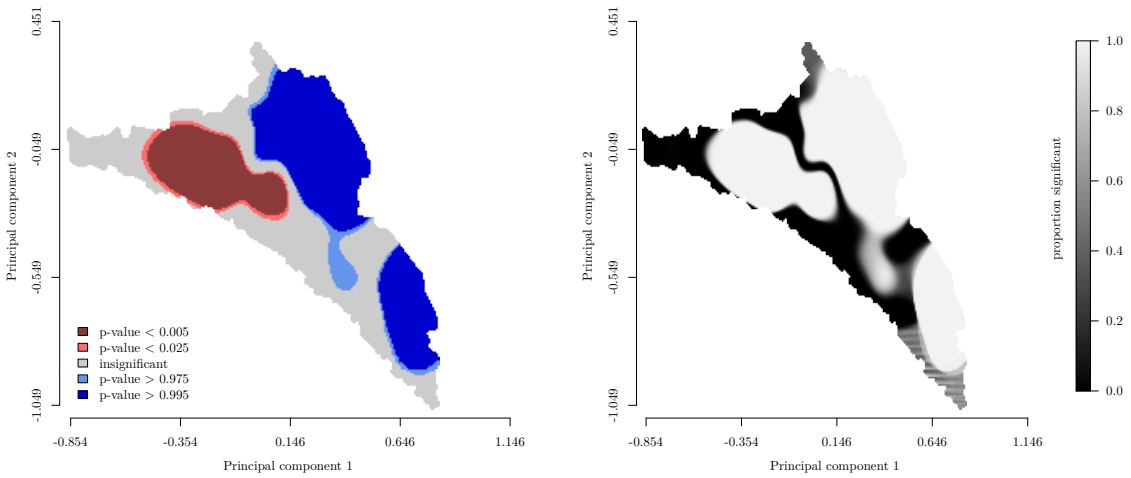
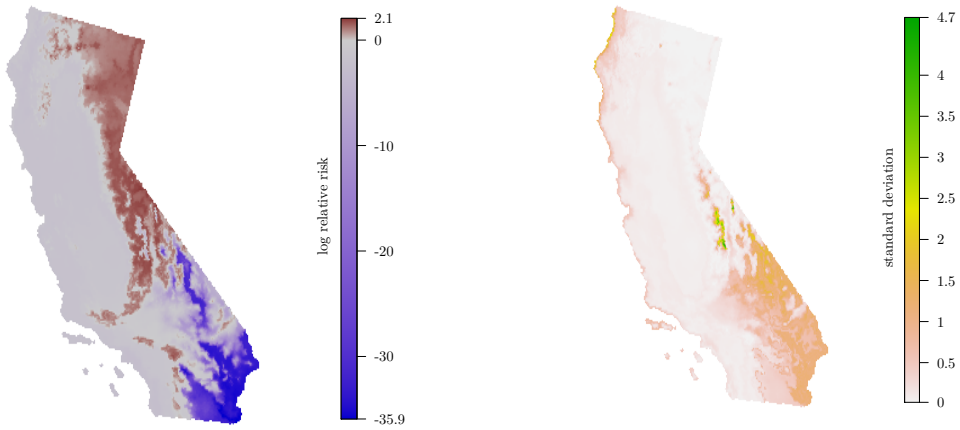


Figure 41: **(a)** The average asymptotic tolerances (*219*, *221*) at given two-tailed significance levels ($\alpha = 0.05$ & $\alpha = 0.01$) of the log relative risk surface in predictor space comparing coyotes that tested negative for plague antibodies (controls) and tested positive for plague antibodies (cases) by the California Department of Public Health (1983–2015) using the developed approach, which randomly displaced observation locations uniformly within an area specified by the criteria in Table 9 (n = 1000 iterations). Predictor space is comprised of the first two principal components of a principal component analysis of seven range-standardized (*197*) Oregon State University Parameter Elevation Regression on Independent Slopes Model 30-year average annual normals (1981–2010) at a 2.5 arcminute (∼4 km) resolution. The each bandwidth was chosen using the maximal smoothing principal (*229*). Warmer colors indicate the ecological niche of plague in coyotes and cooler colors indicate the absence of plague with grayer-colored areas statistically indistinguishable. **(b)** Color pertains to the proportion of iterations a location falls outside of the 95% tolerance interval for null relative risk.

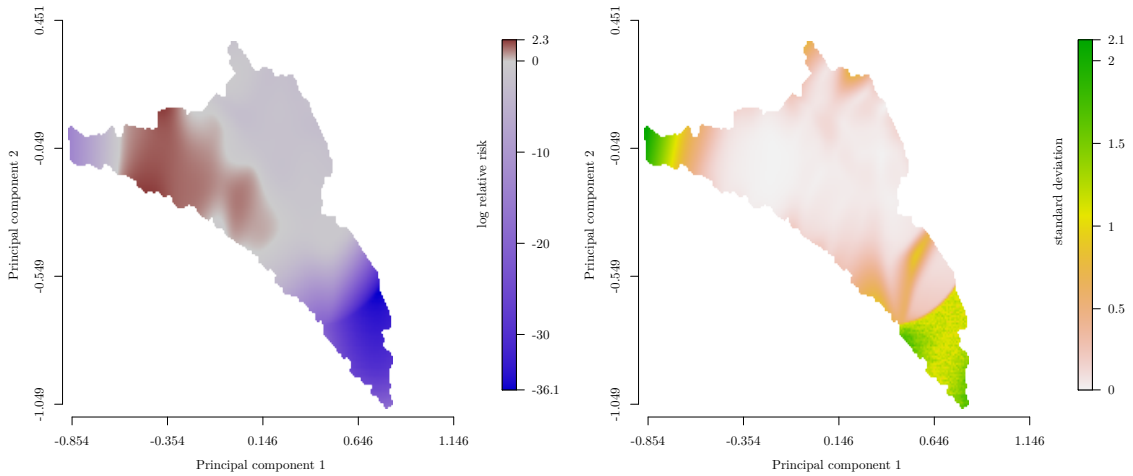


(a) **(b)**

Figure 42: **(a)** Areas of California predicted with average log relative risk surface in predictor space comparing coyotes that tested negative for plague antibodies (controls) and tested positive for plague antibodies (cases) by the California Department of Public Health (1983–2015) using the developed approach, which randomly displaced observation locations uniformly within an area specified by the criteria in Table 9 (n = 1000 iterations). Color pertains to average log relative risk values where positive log relative risk (more likely suitable for plague transmission) are in red and negative log relative risk (more likely unsuitable for plague transmission) are in blue with grayer coloring closer to the null log relative risk value (zero). **(b)** Color pertains to the proportion of iterations fell outside of the 95% standard deviation of the log relative risk values across random iterations in California.

(a) **(b)**

Figure 43: **(a)** Areas of California predicted with average asymptotic tolerances (*219*, *221*) at given two-tailed significance levels ($\alpha = 0.05$ & $\alpha = 0.01$) of the log relative risk surface in predictor space comparing coyotes that tested negative for plague antibodies (controls) and tested positive for plague antibodies (cases) by the California Department of Public Health (1983–2015) using the developed approach, which randomly displaced observation locations uniformly within an area specified by the criteria in Table 9 (n = 1000 iterations). Warmer-colored areas are more likely suitable habitat for plague transmission and cooler-colored areas are more likely unsuitable habitat for plague transmission with grayer-colored areas statistically indistinguishable. **(b)** Color pertains to the standard deviation of the asymptotic tolerances across random iterations. in California.



(a) **(b)**

Figure 44: **(a)** The average log relative risk surface in predictor space comparing coyotes that tested negative for plague antibodies (controls) and tested positive for plague antibodies (cases) by the California Department of Public Health (1983–2015) using the developed approach, which randomly displaced all observation locations uniformly within an area with a radius of 2 km (n = 1000 iterations). Predictor space is comprised of the first two principal components of a principal component analysis of seven range-standardized (*197*) Oregon State University Parameter Elevation Regression on Independent Slopes Model 30-year average annual normals (1981–2010) at a 2.5 arcminute (∼4 km) resolution. Each bandwidth was chosen using the maximal smoothing principal (*229*). Color pertains to average log relative risk values where positive log relative risk (more likely suitable for plague transmission) are in red and negative log relative risk (more likely unsuitable for plague transmission) are in blue with grayer coloring closer to the null log relative risk value (zero). **(b)** Color pertains to the standard deviation of the log relative risk values across random iterations.

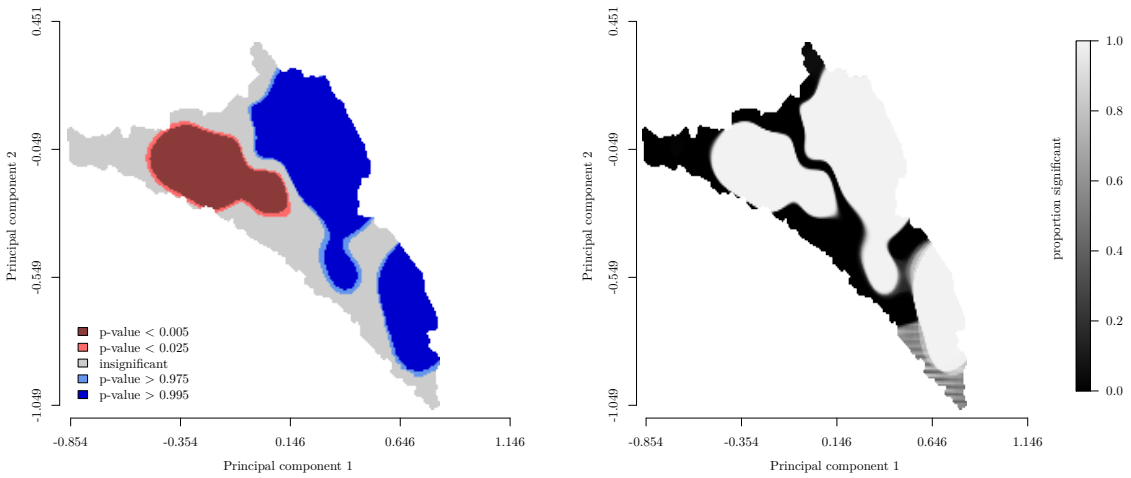
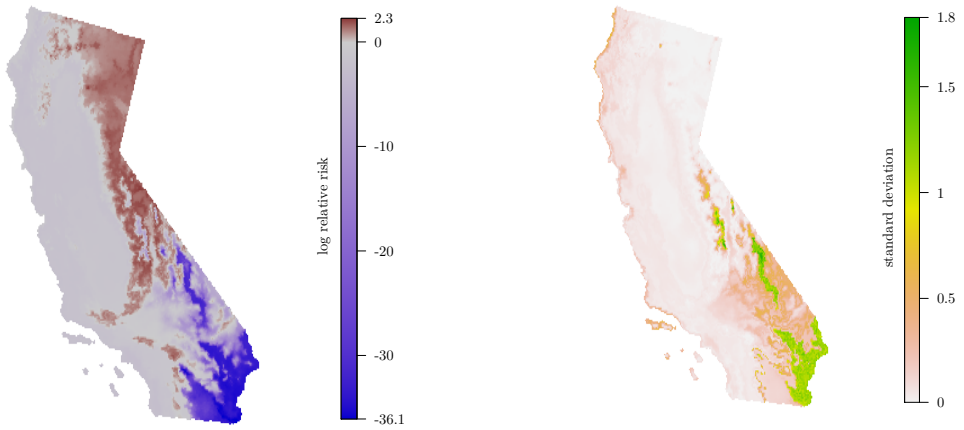


Figure 45: (a) The average asymptotic tolerances (*219*, *221*) at given two-tailed significance levels ($\alpha = 0.05$ & $\alpha = 0.01$) of the log relative risk surface in predictor space comparing coyotes that tested negative for plague antibodies (controls) and tested positive for plague antibodies (cases) by the California Department of Public Health (1983–2015) using the developed approach, which randomly displaced all observation locations uniformly within an area with a radius of 2 km (n = 1000 iterations). Predictor space is comprised of the first two principal components of a principal component analysis of seven range-standardized (*197*) Oregon State University Parameter Elevation Regression on Independent Slopes Model 30-year average annual normals (1981–2010) at a 2.5 arcminute (∼4 km) resolution. The each bandwidth was chosen using the maximal smoothing principal (*229*). Warmer colors indicate the ecological niche of plague in coyotes and cooler colors indicate the absence of plague with grayer-colored areas statistically indistinguishable. **(b)** Color pertains to the proportion of iterations a location falls outside of the 95% tolerance interval for null relative risk.

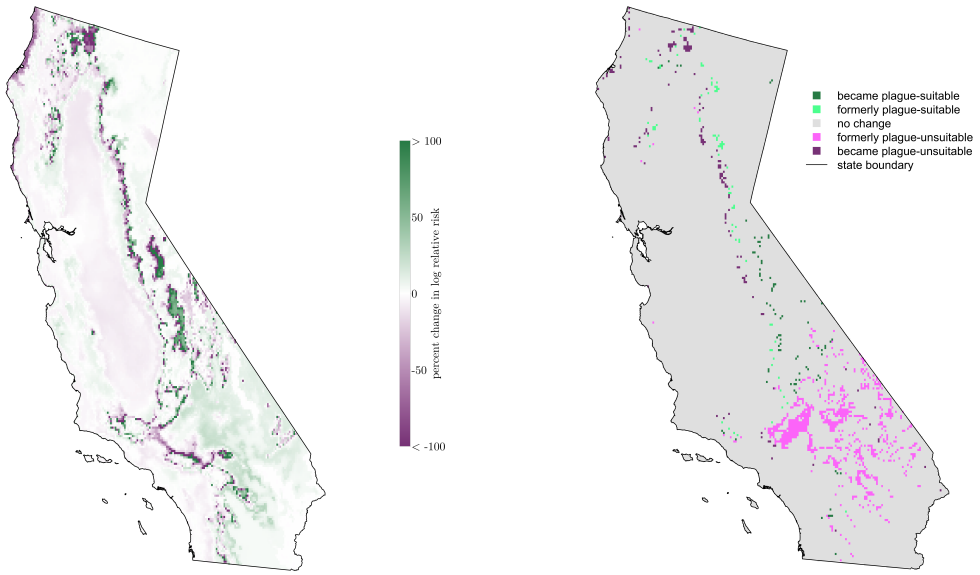


(a) **(b)**

Figure 46: **(a)** Areas of California predicted with average log relative risk surface in predictor space comparing coyotes that tested negative for plague antibodies (controls) and tested positive for plague antibodies (cases) by the California Department of Public Health (1983–2015) using the developed approach, which randomly displaced all observation locations uniformly within an area with a radius of 2 km (n = 1000 iterations). Color pertains to average log relative risk values where positive log relative risk (more likely suitable for plague transmission) are in red and negative log relative risk (more likely unsuitable for plague transmission) are in blue with grayer coloring closer to the null log relative risk value (zero). **(b)** Color pertains to the standard deviation of the log relative risk values across random iterations in California.

(a) **(b)**

Figure 47: (a) Areas of California predicted with average asymptotic tolerances (*219*, *221*) at given two-tailed significance levels ($\alpha = 0.05$ & $\alpha = 0.01$) of the log relative risk surface in predictor space comparing coyotes that tested negative for plague antibodies (controls) and tested positive for plague antibodies (cases) by the California Department of Public Health (1983–2015) using the developed approach, which randomly displaced all observation locations uniformly within an area with a radius of 2 km (n = 1000 iterations). Warmer-colored areas are more likely suitable habitat for plague transmission and cooler-colored areas are more likely unsuitable habitat for plague transmission with grayer-colored areas statistically indistinguishable. **(b)** Color pertains to the proportion of iterations a location falls outside of the 95% tolerance interval for null relative risk.



(a) **(b)**

Figure 48: Change in results after adding more uncertainty in the perturbation analysis (i.e., from the same distance analysis to the varying distance analysis). Same distance analysis randomly displaced all observation locations uniformly within an area with a radius of 2 km (n = 1000 iterations). Varying distance analysis randomly displaced observation locations uniformly within an area specified by the criteria in Table 9. **(a)** Percent change in log relative risk surface in predictor space comparing coyotes that tested negative for plague antibodies (controls) and tested positive for plague antibodies (cases) by the California Department of Public Health (1983–2015) using the developed approach. **(b)** Change in average asymptotic tolerances (*219*, *221*) at given two-tailed significance level ($\alpha = 0.05$) of the log relative risk surface described above.

5 Combining multiple animal-based surveillance systems to predict the spatial distribution of enzootic plague in the western United States

"On the 10th of November [1846] I left Chihuahua, bound for the capital of New Mexico. Passing the Rancho del Sacramento... we entered a large plain well covered with grass, on which were immense flocks of sheep. A coyote lazily crossed the road, and, stopping within a few yards, sat down upon its haunches, and coolly regarded us as we passed."

- *Adventures in Mexico and the Rocky Mountains*
by George Frederick Augustus Ruxton (*275*)

5.1 Introduction

Plague is an infectious disease found throughout the globe that is a concern for conservation and public health, but the precise locations wherein the disease is transmitted and maintained with in the United States remain uncertain. The gram-negative bacterium *Yersinia pestis* causes plague and is primarily transmitted between mammal hosts via bites of infected fleas (*31*). After arriving in California in 1900, plague spread eastward across the United States and reached Kansas by the 1940s where it appeared to stop (*102*). Plague is often found in small mammal species (e.g., rodents) that maintain the bacterium within their populations (*31*, *43*). Many small mammal species are highly susceptible to a *Y. pestis* infection, some dramatically such as prairie dogs (*Cynomys* spp.) that can experience die-off events of up to 98% mortality of an afflicted colony (*105*, *106*). In North America, prairie dogs are keystone species that greatly influence their ecosystem and their losses are a major concern for conservationists (*276*). Humans are also susceptible to a *Y. pestis* infection and experience high fatality (40-95%) without prompt antibiotic treatment (16% mortality with treatment) in the United States (*33*). Human plague cases have been associated with recent or ongoing nearby outbreaks (i.e., epizootics) in wildlife (*57*, *146*) but seldom do these cases result from close contact with a prairie dog or their fleas (*277*, *278*). Identifying regions of the United States where *Y. pestis* is persistent would greatly facilitate public health surveillance and conservation programs because these areas are likely the location of

human and animal exposure to *Y. pestis*-carrying hosts and vectors (*94*, *196*).

Plague surveillance in the United States relies on the sampling of animal populations. Various public health agencies conduct rodent and flea surveillance in response to a human plague case to ascertain the potential source of human plague exposure while some agencies proactively test animals for plague exposure in order to better understand plague occurrence (*107*, *108*). Passive surveillance for plague is not random sampling. Rather, these specimens are often tested for plague exposure: 1) opportunistically because they were originally not collected for the purpose of plague detection and 2) preferentially because specimens were often more likely tested if they were located near historical plague occurrences or suspected plague enzootic regions (*107*). Non-random sampling can lead to statistical biases and is a known challenge of wildlife disease surveillance, in general (*159*). Coyotes (*Canis latrans* Say, 1823) have been used as a sentinel species to monitor plague activity (*95*, *112–116*, but see Panel 8) because coyotes are carnivores that scavenge carcasses and prey on potentially infectious rodents that typically survive plague infection (*109*). In response to exposure to *Y. pestis*, surviving coyotes develop long-lasting antibodies (*91*, *112*). Coyotes are found in numerous types of environments across North America (*171*) and can act as spatially wide-ranging and ecologically comprehensive detectors for plague that is cycling in rodent populations (i.e., sylvatic plague). Plague-positive coyote locations are associated ecologically with plague-positive rodent species locations in California (*89*) and coyote plague cases are temporally associated with human plague cases between 1974 and 1998 in Arizona (*95*).

Using an ecological niche modeling (ENM) framework that identified correlations between environmental variables and the occurrence of coyote plague, Maher and colleagues (*85*) suggest coyote plague cases indicate the fullest extent of plague distribution in North America. Few plague cases have occurred east of the "plague line" in the Great Plains (∼100th meridian) albeit for yet unknown reasons (*102*, *279*). However, Maher and colleagues (*85*) predict plague-suitable habitat in areas with no known plague occurrence such as, for example, the Great Lakes region, likely because the authors were limited by unavailable plague absence data. Furthermore, a pathogen first needs to be established in host populations in order for "disease-suitable habitat" to become an enzootic area (*12*);

therefore, areas with plague-suitable habitat can still be at zero risk of infection if the bacterium has not yet been introduced or established within host populations. This represents a challenge in niche identification, particularly for communicating disease risk. Just because *Y. pestis* could be maintained in an area does not mean that the area is currently at high risk of infection. I avoid this challenge in risk communication by focusing on where risk currently exist: the western United States.

Using an ecological niche modeling framework, I aim to predict the spatial distribution of enzootic *Y. pestis* in the United States found in coyotes that accounted for spatial sampling bias. Maher and colleagues (*85*) found evidence to suggest plague has its own distinct ecological niche that is independent of the distribution of any of its host species ("Plague Niche Hypothesis") and is influenced more by transmission dynamics such as flea vector ecology (Figure 49a). Here, I incorporate an assessment of how much of the host niche is tested for plague exposure and thus how much of the plague niche is observed (Figure 49b). Sampling effort can be accounted in ENMs in various ways from including the host distribution into the estimation of the niche as an explanatory variable (see *90*) to affecting the probability of an ENM algorithm selecting background points for comparison with known occurrences (see *280*). These methods are primarily employed for ENMs that are used in the event only occurrence data are available and absence data are not. I use an extensive collation of coyote plague case (presence) and control (absence) data to estimate the plague niche and, when compared against a separate collation of untested coyote observations, I can estimate how much of the coyote niche was tested for plague. This step-wise process allows me to identify areas of the plague distribution that are sufficiently tested by coyote plague surveillance programs and conservatively predict enzootic areas of plague in North America. My study is conducted in collaboration with federal and state agencies that participate in plague monitoring and the results will identify gaps in monitoring as well as areas of conservation and public health concern.

5.2 Data and methods

I employ an ecological niche modeling framework using a variety of data sources. The data source and design of the analysis can be found in Figure 50.

5.2.1 Coyote location and plague exposure data

Coyotes samples are collected by various agencies. Two sources are used to train an ENM while a third source is used to test the subsequent ENM. Coyote whole blood is collected in the field on strips of filter paper, called "Nobuto strips," allowed to dry, and then transported to a laboratory where the strips are eluted and tested for plague antibodies (*119*). Diagnostic tests for *Y. pestis* exposure vary by agency and are found in Panel 7. A seropositive test result indicates that the coyote was exposed to and developed an antibody response to *Y. pestis*, while a seronegative test result indicates there were no detectable antibodies to *Y. pestis*.

United States Department of Agriculture The U.S. Department of Agriculture (USDA) Animal & Plant Health Inspection Services (APHIS) Wildlife Services (WS) conducts livestock and wildlife damage management (i.e., depredation) operations throughout the United States. Blood samples are taken from a subset of coyotes that are removed during these operations (*107*). Samples are sent to the USDA APHIS National Wildlife Disease Program in Fort Collins, Colorado and are tested for plague at either the USDA National Wildlife Disease Laboratory in Fort Collins, Colorado or at the U.S. Centers for Disease Control and Prevention (CDC) laboratory in Fort Collins, Colorado. Over 28,000 coyotes have been screened for *Y. pestis* exposure (2005–2017). Plague antibodies are detected in 12.7% of coyotes in 15 states ($n = 3,665$) across all 13 years of surveillance (Table 12). Fewer than 2.5% of specimens are missing geographic location information and are ignored in subsequent analysis (Table 13). Of the 28,029 coyotes with geolocation information, about 13.0% ($n = 3,648$ total) are seropositive. In order to protect landowner privacy, county-level USDA coyote sampling appear in Figure 51 and counties with at least one seropositive coyote appear in Figure 52.

California Department of Public Health Blood samples from coyotes managed by USDA APHIS WS in California are sent to the California Department of Public Health (CDPH) Vector-borne Disease Laboratories for plague diagnostic testing. Beginning in 1983 the CDPH Vector-Borne Disease Section digitized coyote-based plague surveillance

data across California. The CDPH conducts passive surveillance of coyotes as sentinels for rodent plague activity to corroborate regional increases in plague activity indicated by rodent plague surveillance data as well as fill surveillance gaps in regions of California with fewer resources or lower access (*108*). Between 1983 and 2015, the CDPH screened 8,119 coyote blood samples for *Y. pestis* antibodies. However, the precise sampling location is not recorded for the vast majority of coyotes. I geocode specimens without geographic coordinates using location descriptions provided by the collector, county data, and base maps. Two percent of tested coyotes lack geographic reference information and are ignored in subsequent analysis (Table 13). Of the 7,955 coyotes with geolocation information, about 8.8% ($n = 704$ total) were seropositive. In order to protect landowner privacy, county-level CDPH coyote sampling appear in Figure 53a and counties with at least one seropositive coyote appear in Figure 53b.

5.2.2 Independently observed coyotes

I collate locations of coyotes observed by other agencies and individuals in order to determine the background coyote population in the United States. Observers include university principal investigators, biologists, citizen scientists, and conservation agencies. These observations are recorded in databases maintained by various museums and government agencies. Here, I assemble coyote observations across the United States from three main databases: 1) United States Geological Survey–Biodiversity Information Serving Our Nation (*121*), 2) Global Biodiversity Information Facility (*122*), and 3) VertNet (*123*). Across the United States 13,972 unique coyote occurrences were reported (Table 14; Figure 55). Taken together, these coyotes are not necessarily tested for plague antibodies, but are included in subsequent analysis because they have accompanying location information. These data are an approximation of coyote habitat preference, which is a better indicator of where coyotes reside than a species home range. Coyotes are found across the majority of North America (*171*) and coyotes are not homogeneously distributed within their species home range. Individual coyote home range information would be favorable for this analysis, but was unavailable.

Other state health departments I use an independent source of animal plague surveillance data to validate my prediction from an ENM using USDA and CDPH data. The CDC partners with state health departments that require the laboratory resources of the CDC to monitor plague in wildlife. Animal specimens are sampled using active and passive surveillance in response to a suspicious animal or human plague case. State health departments send blood samples to the CDC in Fort Collins, Colorado for laboratory testing (Panel 7). Test results are recorded in a massive, paper-based archive of animal-based plague surveillance in North America maintained by the CDC National Center for Emerging and Zoonotic Infectious Disease Division of Vector-Borne Diseases Bacterial Diseases Branch. A digitization effort by Maher and colleagues (*85*) georeferenced 3,777 occurrence points between 1998 and 2008 that included generic or specific reference to host species for 75 mammal taxa, primarily *C. latrans* ($n = 2,516$; 90% of samples). Plague-positive coyote location data were provided by Maher and colleagues (*85*) that contained 93.8% of coyote locations from the original analysis ($n = 2,360$) and non-coyote locations ($n = 929$).

There are duplicates between the CDC and USDA coyote data because the CDC tested coyotes for the USDA until 2012. Complete metadata are not available from the CDC so identifying duplicates proves challenging. Duplicate coyote records between the CDC and the USDA are detected by exact spatial coordinates using the **sp** package (*165*, *166*) in the statistical software **R** (*127*). This is not a perfect duplicate detection method because there is a risk of excluding coyotes that may not truly be duplicates (false-positive error). However, the method is conservative because the CDC data set is being used for external validation and was sought to reduce the risk of false-negative error. Duplicate records in the CDC data set ($n = 559$ or $\sim$23.7% of CDC coyote plague specimens) are excluded from subsequent analysis with the resulting data containing $1,801$ coyote records used for external validation of my ENM. County-level state health department coyote sampling appear in Figure 54.

Environmental data processing Although the relationship between plague and climate is complex (*58*), past studies have established links between *Y. pestis* occurrence and climate factors (*70*, *85*, *89*, *90*). I employ the Oregon State University PRISM (Param-

eter elevation Regression on Independent Slopes Model) statistical mapping system as my predictor variables for the subsequent analysis. PRISM follows a weighted regression framework that relies on digital elevation models built from a network of ground measurements (*125*). The 30-year average normals (1981–2010) at a 2.5 arcminute (∼4 kilometer by ∼4 kilometer; ∼16 square kilometers) resolution for the contiguous United States are selected because the USDA has a long-term partnership with the PRISM Climate Group at Oregon State University and is the official climatological data for the USDA. The variables used in the analysis are reviewed in Table 15. Data are available from the PRISM portal (http://prism.oregonstate.edu/) or via the **prism** package (*126*) in the statistical software **R** (*127*).

Environmental data sets often contain many different measurements, some highly correlated (e.g., elevation and precipitation) so collinearity is avoided by reducing the number of variables used in the approach. I conduct a principal component analysis (PCA), which transforms multiple predictors into a set of linearly uncorrelated variables called principal components. PRISM variables for the contiguous United States range standardized using the Gower metric (*197*) because PRISM variables are measured on different scales. A PCA is conducted using the **RStoolbox** package (*198*) in the statistical software **R** (*127*). The first two principal components of the PRISM variables account for over 96% of the variance (Table 16) and are used as the environmental covariates for the subsequent analysis (Figure 56).

5.2.3 Statistical methods

Ecological niche model construction An ecological niche modeling approach is used for my analyses. Developed in Chapter 3, the approach adapted a spatial cluster detection algorithm originally designed for spatial cancer epidemiology. Here, I use the method as a mechanism for environmental interpolation (*19*) to identify spatial clusters of climate signatures within the PC space. These signatures are then predicted across California even into areas that are not historically sampled or where plague surveillance data are unavailable. The approach is based on a kernel density estimation of spatial relative risk (*211*, *212*) that treats the observed point locations of cases and controls as realizations of spatial Poisson

processes. An output of the model is a log relative risk value (i.e., the log ratio of estimated spatial density of cases to that of controls; *211–216*) where plague is more likely to be observed in areas with a log relative risk value above zero than areas with a log relative risk value below zero. Another output of the model is a local two-tailed asymptotic tolerance calculation (*219*, *221*) at two significance levels ($\alpha = 0.05$ & $\alpha = 0.01$). An asymptotic tolerance level value below 0.025 is considered "plague-suitable habitat" where plague is likely sylvatically cycling and an asymptotic tolerance level value above 0.975 is considered "plague-unsuitable habitat" where plague is likely not sylvatically cycling. The approach also identifies areas within the predictions with sparse observation data. All analyses were conducted in the statistical software **R** (*127*).

Here, USDA and CDPH coyote data are used in an ENM to predict the spatial distribution of enzootic plague in the western United States (i.e., the "disease map"). Seropositive cases and seronegative controls of both agencies are used to estimate a log relative risk surface in predictor space and then are predicted across geographic space to map risk across the western United States. A fixed-bandwidth kernel density estimator provides an estimate of the log relative risk surface with a smoothing parameter (i.e., bandwidth) chosen with the maximal smoothing principle (*229*). I conduct a sensitivity analysis for the choice of smoothing parameter (i.e., bandwidth) in predictor space. Unlike my findings in Chapter 3 my prediction is not robust across increasing bandwidth sizes. A smaller bandwidth (i.e., half the size chosen by the maximal smoothing principle; *229*) performed similarly to the initial model, but a larger bandwidth (i.e., double the size chosen by the maximal smoothing principle; *229*) performed worse than the initial model according to 95% confidence intervals of the average Area Under the Receiver Operating Characteristic Curve (AUC; *222*) across iterations from a 50-fold cross-validation. This suggest that the scale of observation associated with twice the selected bandwidth is not sensitive enough to detect the same pattern as detected by the two smaller bandwidths. Prediction results and validation statistics of the bandwidth sensitivity analysis appear in Figure 57 and Figure 58, respectively. An ENM with a bandwidth chosen by the maximal smoothing principle (*229*) is used for subsequent analysis and the predicted disease map is saved for the next step.

Accounting for spatial sampling effort The analysis accounts for the spatial sampling effort bias of coyote plague surveillance by the USDA and the CPDH in two ways. First, because the ENM uses both cases (presences) and controls (absences) the method implicitly accounts for sampling effort, provided the same limitations apply to both cases and controls (*212*). Absence data improves ENM predictions and can identify low habitat suitability areas that may have been considered high suitability if only presence data are used in an ENM (*281*). For example, Brotons and colleagues, (*161*) suggests absence data helps to remove bias in model predictions for uncommon habitat of a study region if presence and absences are located in them. Without absence data, an ENM can associate uncommon habitat with occurrences. Second, I conduct a separate ENM to detect the area of predictor space that is tested for plague as an indicator of the sampling effort of coyote plague surveillance (i.e., the "bias layer"). I then combine the disease map and bias layer together to indicate zones of the spatial distribution of enzootic plague in the western United States that are sufficiently tested for the presence of *Y. pestis* in coyotes.

For the bias layer, USDA and CDPH coyote data are combined, regardless of plague exposure, and compared against the independently observed coyote observations from museum repositories. In this step, coyotes that are tested for plague (USDA and CDPH) are considered "cases" and coyotes that are not tested for plague (museum data) are considered "controls" in an ENM. A fixed-bandwidth kernel density estimator is used for the log relative risk surface with a smoothing parameter chosen with the maximal smoothing principle (*229*). The predicted spatial distribution from this ENM indicates areas that are more likely sampled for plague ("sufficiently sampled") and less likely sampled for plague ("insufficiently sampled") as an indicator of sampling effort bias. Insufficiently sampled areas are zones with habitat types where coyotes have been historically observed, but are likely not observed by the USDA or the CDPH and tested for plague. These areas indicate gaps in plague surveillance. As a comparison, I calculate a spatial relative risk surface of the bias layer in geographic space using the **sparr** package (*282*) in the statistical software **R** (*127*) with a fixed bandwidth of one degree latitude/longitude. A bias layer constructed from the spatial density of observations is a traditional method of accounting for spatial sampling bias in ENMs (*280*).

I combine the disease map and the bias layer in two ways. First, I weight the disease map by the bias layer to create a weighted continuous layer. The predicted log relative risk of the bias layer is standardized by:

$$r(x)_{standardized} = \frac{r(x)}{|min(r(x))|} + 1$$

where $r(x)$ is the log density relative ratio of the bias layer that is centered at the null value (zero) by:

$$r(x) \leq -max(r(x)) \rightarrow r(x) = -max(r(x)).$$

The standardized bias layer ranges from a maximum of two (sufficiently sampled areas) to a minimum of zero (insufficiently sampled areas) and is centered at a null value of one (i.e., no change in weight). I weight log relative risk of coyote plague in the western United States by multiplying the predicted log relative risk of the disease map by the standardized bias layer. Second, a composite of the disease map and bias layer is created based on each asymptotic tolerance significance levels. Both layers are categorized at a local two-tailed significance level of $\alpha = 0.05$ and overlaid to identify areas of the disease map with, for example, statistically significant suitable plague habitat that are insufficiently tested for plague when compared to historical coyote occurrence data. The output of this composite is categorical in nature.

Model performance Model performance and statistical inference are conducted on the predicted spatial distribution of enzootic plague in the western United States (disease map). Predictive model performance is evaluated by 50-fold cross-validation. Controls are randomly undersampled to balance plague prevalence (0.5) in each iteration. The average AUC and a 95% confidence interval are calculated across iterations (*222*). A precision-recall curve (*223*) is also used to evaluate the prediction. Additionally, the positive predictive value (PPV) of the predicted spatial distribution of enzootic plague in the western United States is calculated for the CDC-tested animal plague surveillance data. The performance of the disease map is assessed for coyotes and other non-coyote species within the

CDC data.

5.3 Results

Statistically significant relative clustering of seropositive coyotes (cases; $n = 4,352$) and seronegative coyotes (controls; $n = 31,632$) by the USDA (2005–2017) and the CDPH (1983–2015) was detected in predictor space using the kernel-density estimation-based approach in predictor space (bandwidth = 0.054 PCA coefficient). The estimated intensity surfaces of seropositive cases and seronegative controls in predictor space appear in Figure 59 and calculated asymptotic tolerances appear in Figure 60. Additionally, statistically significant relative clustering of coyotes tested by the USDA (2005–2017) and the CDPH (1983–2015) for *Y. pestis* antibodies (cases; $n = 35,984$) and historical coyote observations (controls; $n = 13,972$) from museum repositories was detected in predictor space using the kernel-density estimation-based approach in predictor space (bandwidth = 0.05 PCA coefficient). The estimated intensity surfaces of plague-tested cases and plague-untested controls in predictor space appear in Figure 61 and calculated asymptotic tolerances appear in Figure 62.

5.3.1 Spatial distribution of enzootic plague in the western United States

The ecological niche of plague in coyotes was predicted for the western United States. The log relative risk and asymptotic tolerance significance level values appear in Figure 63 and Figure 64, respectively. Plague-suitable habitat was predicted in every western state but not east of the 100th meridian. Plague-suitable habitat was predicted in mountainous and plateau regions of the western United States. Areas of the western United States with high log relative risk occur in the southwestern region compared to the northern states. Plague-unsuitable habitat occurs in the eastern Great Plains, the Pacific coastline, and desert regions. The ENM could not determine the habitat suitability for plague transmission in alpine regions or elevation transition zones between plague-suitable and plague-unsuitable areas. There were sparse coyote observations in areas such as, for example, the northwestern region, high alpine habitat, and regions of the Pacific Ocean and Gulf of Mexico coastline. The prediction of log relative risk across the western United States had an average Area

Under the Receiver Operating Characteristic Curve (AUC) of 0.757 (95% CI: 0.747–0.767) and an acceptable precision-recall curve after a 50-fold cross-validation (Figure 65). The PPV for all CDC-tested animals (1998–2008) was high (88.1), especially for non-coyote specimens (91.6), but also for coyotes (86.3). Of the CDC-tested coyotes that were observed outside of plague-suitable areas ($n = 423$) about a third (30.3%) were located within the circular diameter of an average coyote home range size of 25 km^2 (*144*) and the vast majority (85.8%) were within the observed dispersal range (∼40 km) of a coyote (*145*).

5.3.2 Spatial sampling effort of plague surveillance in the United States

The observed niche of plague in coyotes was predicted for the contiguous United States. The log relative risk and asymptotic tolerance significance level values appear in Figure 66 and Figure 67, respectively. Sufficient plague testing occurred in every western state. Mountainous and plateaued regions of the western United States were relatively more likely tested for plague occurrence than the eastern United States. Mid-elevation levels of mountainous regions of the western United States were relatively more likely tested for plague occurrence than high alpine regions. For example, the highest elevation of Utah, the Unita Mountains (∼3,900 meters), was relatively more likely untested for plague occurrence than the Wasatch Range (∼3,300 meters). In the southwestern United States the trend switches where, for example, the high and low elevation areas of Arizona were relatively more likely tested for plague occurrence than the transitional zone between these regions. Desert regions were relatively more likely tested for plague occurrence than plains. For example, the Chihuahuan and Sonoran Deserts were relatively more likely tested for plague occurrence than the Central Valley in California or southern Great Plains. The ENM could not determine if areas of the Great Plains in Kansas and Oklahoma were sampled sufficiently or insufficiently tested for plague, as well as the area surrounding Las Vegas, Nevada. There were sparse coyote observations in areas such as, for example, the northwestern region, extremely high alpine habitat, and small unique climate areas of the eastern United States.

The environmentally interpolated bias layer using an ENM (Figure 67) was similar to a spatially interpolated bias layer using a spatial relative risk function (Figure 68). The areas predicted more likely untested for plague occurrence in Arizona and Wyoming are

similar in the two models. However, the ENM predicted habitat in southern California and eastern Washington, for example, as relatively more likely tested for plague than the spatial interpolation.

5.3.3 Spatial distribution of enzootic plague in western United States accounting for spatial sampling effort

When the disease map and bias layer are combined the predicted spatial distribution of enzootic plague is moderated. The majority of predicted plague-suitable habitat was sufficiently tested for plague occurrence (Figure 69). Areas predicted plague-suitable but were undersampled for plague occur in eastern and central New Mexico, the low elevation levels of the Colorado Plateau, and the foothills of ranges within the Rocky Mountains. The majority of predicted plague-unsuitable habitat was undersampled for plague occurrence except in, for example, desert regions the northern Great Plains, and the Columbia Basin. Areas of high weighted log relative risk were predicted in the Great Salt Lake area in Utah and eastern foothills of the Sangre de Cristo Mountains in Colorado and New Mexico (Figure 70) similar to the unweighted log relative risk disease map (Figure 63). However, there were more areas closer to the null log relative risk value (zero). The transitional zone and Colorado Plateau in Arizona as well as the Black Range in New Mexico, for example, that had elevated log relative risk values in the disease map were down weighted when accounting for the bias layer in the predicted disease map. In addition, central Texas as well as the Central Valley in California, Willamette Valley in Oregon, and western slope of the Cascade Mountains in Washington that had low predicted log relative risk values in the disease map were up weighted when accounting for the bias layer in the predicted disease map.

5.4 Discussion

Sampling effort bias was evident in the surveillance of a zoonotic pathogen (*Y. pestis*) in the western United States using a sentinel species (*C. latrans*) monitored by federal and state animal-based surveillance systems. Accounting for sampling effort bias in an ecological niche modeling framework, I predicted the spatial distribution of enzootic plague across the western United States. My prediction using case (presence) and control (absence) coyote

location data was more constrained than the extent of the spatial distribution predicted by Maher and colleagues (*85*) who used only cases in their ENM. By accounting for spatial sampling effort of my prediction the eastern most extent of the predicted distribution of enzootic plague was restricted to west of the 100th meridian. Few animal plague cases have occurred east of this "plague line" even though susceptible hosts (e.g., prairie dogs, coyotes, and humans) occur on either side (*102*, *276*, *279*). My results suggest climate may be a contributing factor to this phenomenon based on an ENM comprised of 30-year annual average climatological variables.

USDA coyote-based plague surveillance is preferentially opportunistic sampling whereby The USDA collects blood samples from coyotes typically on agricultural/rural land (livestock and wildlife damage management; *107*). Coyotes provided to the CDPH for plague testing were from USDA APHIS WS operations in California (*283*). Preferential sampling explains why sampling effort was sufficient for the majority of the western United States predicted suitable for plague transmission and insufficient for areas predicted unsuitable for plague transmission. In Oregon, for example, the Columbia Plateau has the type of environment that was sufficiently tested and predicted suitable for plague transmission, but the Willamette Valley has the type of environment that was predicted unsuitable for plague transmission and insufficiently sampled when compared to historical coyote records. Additionally, preferential plague sampling is evident on a regional scale where the vast majority of USDA plague-tested coyotes were located west of the "plague line" although the USDA has sufficiently tested both sides of the "plague line" in the northern Great Plains. Opportunistic sampling explains why certain sampling effort was insufficient in some areas and environments. With the exception of extremely high alpine environments, coyotes in the historical record were observed in high elevation regions of the western United States but were infrequently tested for plague because agricultural land is rare at high altitude and fewer requests for USDA APHIS WS occurred in these areas. The USDA APHIS WS operates only by request for assistance and thus has jurisdiction restrictions that creates spatial gaps in sampling, for example, the Four Corners regions of the southwestern United States.

Some spatial limitations in sampling were overcome by using an ENM as an environ-

mental interpolation technique (*19*). USDA sampling was sparse in northeastern Arizona and northwestern New Mexico where many Native American Tribal Nations are located. A spatial interpolation of plague relative risk (*282*) in coyotes would be challenged by sparse data in this region. Much of the environment in this region is located in other parts of the western United States and was tested for plague, which allowed me to predict plague-suitable habitat into data sparse areas. Domestic dogs are commonly used as sentinel species on the Navajo Reservation (*146*, *284*) and the CDC tested their blood samples. The portion of those data were digitized between 1998 and 2008 (∼300 dogs) and were used in the Maher and colleagues study (*85*). My ENM had high prediction performance, especially for non-coyote specimens (PPV = 91.6; 31% were domestic dogs in Arizona and New Mexico) suggesting my predictions are accurate in this region of the western United States. Additionally, seropositive coyotes were located in Yakima County, Washington but were excluded for lack of sub-county geographical identification. I predicted plague-suitable area in Yakima County without these data in my ENM demonstrating the utility of environmental interpolation. While the USDA found no seropositive coyotes in North Dakota a separate survey found one *Y. pestis* exposed coyote (*114*) in McKenzie County, North Dakota and I predicted parts of western North Dakota as suitable for plague transmission. seropositive

My results are similar to other studies. Three studies by Eisen and colleagues predicted high-risk plague areas in the southwestern United States (*71*, *72*) and only for New Mexico (*73*) by conducting a logistic land-use regression approach using human plague cases ranging from 1957 to 2004. My prediction captured all areas predicted by Eisen and colleagues (see Figure 2 in *71*, Figure 1 in *72*, and Figure 1 in *73*) as high plague risk. However, because the analysis was not limited to certain land-use types (e.g., Rocky Mountain Ponderosa pine) I predicted more areas of southwestern states (Arizona, Colorado, New Mexico, and Utah) were suitable for plague transmission. My results also suggests that my prediction was more sensitive but not as specific as for human-plague risk by Eisen and colleagues (*71*–*73*). Areas with higher log relative risk in my prediction were in elevations immediately below areas predicted by Eisen and colleagues (*71*–*73*) as high-plague risk. Seropositive coyotes may be more commonly found in habitats at these lower elevations and either exposed there or

immigrated to lower elevations after being exposed at higher elevations. Using an ENM and a small sample of deer mice (*Peromyscus maniculatus*) Walsh and Haseeb (*90*) predicted plague-suitable habitat in patchy, high elevation areas across the western United States and further east than my prediction (see Figure 3 in *90*). Even though deer mice are monitored for plague activity (*108*), deer mice are more an indicator of a recent epizootic event (*92*, *93*, *285*) than a predictor of enzootic areas. My prediction was more similar to Holt and colleagues (*89*) predictions than my predictions in Chapter 3 and Chapter 4 using only CDPH coyote serology data. Holt and colleagues (*89*) used rodent serology data from the CDPH to predict plague-suitable areas of California, including high elevation areas of the Santa Lucia Range and Diablo Range. My predictions in Chapter 3 and Chapter 4 did not predict plague-suitable area in the central Coastal Ranges and I concluded that was for lack of detection in CDPH coyote data. Here, I was able to predict plague-suitable area in the central Coastal Ranges because I leveraged regional data from the USDA and seropositive coyotes were found in habitats similar to the central Coastal Ranges, thus demonstrating the utility of a regional analysis.

My study was limited to the western United States. Plague-infected animals have been found in southern parts of British Columbia, Alberta, and Saskatchewan, Canada (*286*). Maher and colleagues (*85*) predicted plague-suitable habitat in Canada using climate data from WorldClim (*287*) for Canada and the United States. Here, I focused on interpolation within the United States where I had plague surveillance data and predicted plague in Montana and the Columbia Basin of Washington that have analogs in Canada. A future study could predict sample bias corrected plague-suitable habitat in Canada using Canadian plague surveillance data with or without United States plague surveillance data. While I predicted plague-suitable habitat in southern New Mexico and western Texas, no plague infected flea vectors have been found in Mexico (*288*, *289*) and no human plague case has occurred since 1923 (*290*). Hot, dry climates like the Sonoran Desert were predicted unsuitable for plague transmission potentially due to the lower efficiency of flea vector transmission at higher temperatures (*82*, *291* but see *81*). Therefore, there may be less plague activity in hotter, dryer climates of Mexico, which could explain the lack of animal cases. But transmission can still occur in suboptimal conditions (*41*, *46*) and plague may

be enzootic in Mexico at low levels. A future study could project plague-suitable habitat into Mexico using United States plague surveillance data.

There are limitations to using only coyotes to predict the spatial distribution of enzootic plague in the United States (Panel 8). While coyotes are wide-ranging and sample the fullest extent of the plague niche (*85*), the bacterium is a generalist pathogen found in numerous mammal species (*43*) and other mammals may be better suited as detectors of plague, especially if their species range exists in areas where coyotes do not. Maher and colleagues (*85*) conclude plague has its own niche distinct from any one of its mammal hosts; therefore, while I accounted for how well I observed the coyote plague niche (Figure 49b) by excluding non-coyote species, I did not sample the entire plague niche. My ENM using coyotes primarily collected on agricultural land could not determine if some areas were plague-suitable or plague-unsuitable. These areas included alpine habitats that was likely due to a lack of sampling and other mammals, namely rodent species, can and are used to detect plague in these environments (*89*). For example, the ENM could not determine if the environment in and around Yellowstone National Park and Grand Teton National Park was plague-suitable or plague-unsuitable. Plague was found absent or at low prevalence (<1%) in surveys of small mammals (*292*, *293*) but at varying prevalence (0%-57%) in coyotes (*294*) in Teton County, Wyoming. Because coyotes have large home range sizes (*144*, *145*), the true location of *Y. pestis* exposure is not known for an individual coyote. Coyotes could be exposed in one area and emigrate to another area where they are sampled which could be problematic for prediction, especially for areas on the boundary of plague-suitable and plague-unsuitable. Indeed, almost half (47.8%) of the CDC-tested coyotes that were not observed in a plague-suitable raster pixel were within the raster neighborhood (12 km x 12 km) of a plague-suitable pixel. Using small mammals with smaller home ranges along these boundaries or in unique environments like Teton County, Wyoming would provide precise measurements of plague activity instead of only relying on coyotes that may have immigrated. This source of uncertainty was not addressed in the present analysis and assessment of its impact on the prediction of the spatial distribution of enzootic plague is recommended for future investigation.

My estimate of the sampling effort bias of United States federal and state agencies that

participate in plague monitoring is imperfect because coyote occurrences in museum repositories are an imperfect proxy for the coyote niche. Biodiversity databases improve data access and availability for investigations of species distributions; however, sampling effort, spatial scale, type of data collected, and data storage protocols distort large-scale biodiversity patterns (*167–169*). Not all coyotes were observed and of those that were observed not all were recorded in museum repositories. Further, coyote observations in museum repositories that did not have geographical location information were ignored in the present analysis. There were fewer coyote observations with geographical location information in museum repositories than those collected in coyote-based United States plague surveillance systems, but I aimed to better match sampling effort. Observations from museum repositories were not restricted to the present study period (1983-2017) and include historical coyote occurrences (early 1900s) in order to capture as many records as possible. Inclusion of pre-1983 records may result in a temporal bias in the analysis but the species range of coyotes has not drastically changed in the western United States since 1900 (see Figure 3 in *295*). A future study could account for sampling effort by using the spatial density of coyote occurrences (from museums and coyote-based plague surveillance) as a covariate in the estimation of the plague niche (see *90*), either in the principal component analysis or as a separate variable.

The ENM model used in the analysis is analytically limited to pairwise comparisons in two dimensions (variables). Other types of this model have been extended to many dimensions (*208*). Future studies can extend this ENM method to more than two dimensions for applications where more than two variables are necessary, such as a variable for sampling effort or the spatial distribution of the host. The ENM model is designed to compare cases (presences) and controls (absences), but cases for two separate diseases could be compared. The ENM model is an adaptation of a spatial relative risk (*211*, *212*) that originally assessed cancer cases clustered within the population at risk in the United Kingdom. Plague cases of two different species could be compared to assess where in predictor space they relatively cluster together (niche overlap) or to assess how the niche of a particular host overlaps with the plague niche, which can facilitate the identification of reservoir, spillover, and sentinel hosts – a central focus of animal plague investigations in the United States (*47*, *57*, *85*, *91*,

107, *117*, *147*, *196*, *285*, *296*, *297*). Sin Nombre Virus (Bunyaviridae), another zoonotic pathogen that causes the deadly hantavirus pulmonary syndrome (*298*), is found in plague enzootic regions (*71*, *299*). The relative clustering of plague and hantavirus cases could be assessed using the developed ENM approach to predict the type of environment and areas that both diseases are active for unified surveillance operations. Other methods can detect niche overlap for more than two groups (*300*, *301*) and are recommended if that is of interest.

A map of enzootic areas of plague can help inform conservation and public health programs, such as surveillance and vaccine delivery. My results could identify zones to monitor for plague in threatened species and to deploy a vaccine currently developed for prairie dogs (*302*) and, potentially, a human vaccine that is in development (*303*). The utility of the predicted spatial distribution of enzootic *Y. pestis* in the western United States to describe the spatial patterns in human cases is tested in Chapter 6. When accounting for sampling effort, I identified areas where additional testing is warranted to determine the spatial distribution of enzootic plague. My results will likely not impact policy decisions about where a coyote is sampled because that is the purview of USDA APHIS WS livestock and wildlife damage management operations and not intended for plague surveillance. However, laboratory testing of coyote blood samples for plague surveillance could be prioritized to fill gaps in surveillance such as, for example, the Llano Estacado in New Mexico and the Little Colorado River basin in Arizona. Even though the majority of USDA-tested coyotes were located in New Mexico and Arizona, local-level shifts in testing can optimize laboratory resources. Compared to historical coyote records, USDA plague surveillance has sufficiently sampled the "plague line" in the northern Great Plains, but further testing in the Texas Panhandle is recommended to validate plague is not enzootic east of the 100th meridian. Surveillance in these areas may become more important in the future because of climate change. Nakazawa and colleagues (*70*) predicted plague would shift northward and the "plague line" may shift eastward as the Great Plains becomes more arid (*304*, *305*) further threatening prairie dog colonies that have previously been unaffected by plague.

5.5 References

12. A. T. Peterson *et al.*, *Ecological niches and geographic distributions (MPB-49)* (Princeton University Press, Princeton, New Jersey, 2011), vol. 56, ISBN: 0691136882.

19. A. T. Peterson, *Mapping disease transmission risk: enriching models using biogeography and ecology* (Johns Hopkins University Press, Baltimore, Maryland, 2014), ISBN: 1421414737.

31. K. L. Gage, M. Y. Kosoy, *Annual Review of Entomology* **50**, 505–528, ISSN: 0066-4170, DOI 10.1146/annurev.ento.50.071803.130337 (2005).

33. K. J. Kugeler *et al.*, *Emerging Infectious Diseases* **21**, 16, ISSN: 1080-6059, 1080-6040, DOI 10.3201/eid2101.140564 (2015).

41. R. J. Eisen *et al.*, *Proceedings of the National Academy of Sciences* **103**, 15380–15385, ISSN: 0027-8424, DOI 10.1073/pnas.0606831103 (2006).

43. R. Pollitzer, *Bulletin of the World Health Organization* **23**, 313, ISSN: 0042-9686 (1960).

46. R. J. Eisen *et al.*, *Journal of Medical Entomology* **44**, 672–677, ISSN: 0022-2585, DOI 10.1093/jmedent/44.4.672 (2007).

47. J. L. Lowell *et al.*, *Journal of Vector Ecology* **34**, 22–31, ISSN: 1081-1710, DOI 10.1111/j.1948-7134.2009.00004.x (2009).

57. H. E. Brown *et al.*, *American Journal of Tropical Medicine and Hygiene* **82**, 95–102, ISSN: 0002-9637, 1476-1645, DOI 10.4269/ajtmh.2010.09-0247 (2010).

58. T. Ben-Ari *et al.*, *PLoS Pathogens* **7**, e1002160, ISSN: 1553-7374, 1553-7366, DOI 10.1371/journal.ppat.1002160 (2011).

70. Y. Nakazawa *et al.*, *Vector-Borne and Zoonotic Diseases* **7**, 529–540, ISSN: 1557-7759, 1530-3667, DOI 10.1089/vbz.2007.0125 (2007).

71. R. J. Eisen *et al.*, *American Journal of Tropical Medicine and Hygiene* **77**, 999–1004, ISSN: 0002-9637, 1476-1645, DOI 10.4269/ajtmh.2007.77.999 (2007).

72. R. J. Eisen *et al.*, *Journal of Medical Entomology* **44**, 530–537, ISSN: 0022-2585, DOI 10.1603/0022-2585(2007)44 (2007).

73. R. J. Eisen *et al.*, *American Journal of Tropical Medicine and Hygiene* **77**, 121–125, ISSN: 0002-9637, 1476-1645, DOI 10.4269/ajtmh.2007.77.121 (2007).

81. A. M. Schotthoefer *et al.*, *Journal of Medical Entomology* **48**, 411–417, ISSN: 0022-2585, DOI 10.1603/ME10155 (2011).

82. D. C. Cavanaugh, *American Journal of Tropical Medicine and Hygiene* **20**, 264–273, ISSN: 0002-9637, 1476-1645, DOI 10.4269/ajtmh.1971.20.264 (1971).

85. S. P. Maher *et al.*, *American Journal of Tropical Medicine and Hygiene* **83**, 736–742, ISSN: 0002-9637, 1476-1645, DOI 10.4269/ajtmh.2010.10-0042 (2010).

89. A. C. Holt *et al.*, *International Journal of Health Geographics* **8**, 38, ISSN: 1476-072X, DOI 10.1186/1476-072X-8-38 (2009).

90. M. Walsh, M. A. Haseeb, *PeerJ* **3**, e1493, ISSN: 2167-8359, DOI 10.7717/peerj.1493 (2015).

91. D. J. Salkeld, P. Stapp, *Vector-Borne and Zoonotic Diseases* **6**, 231–239, ISSN: 1557-7759, 1530-3667, DOI 10.1089/vbz.2006.6.231 (2006).

92. R. J. Eisen *et al.*, *Journal of Medical Entomology* **45**, 1160–1164, ISSN: 0022-2585, DOI 10.1093/jmedent/45.6.1160 (2008).

93. J. D. Lang, *Journal of Vector Ecology* **29**, 236–247, ISSN: 1081-1710 (2004).

94. K. L. Gage, in *Advances in Yersinia research* (Springer, New York, New York, 2012), pp. 79–94, ISBN: 1461435600.

95. H. E. Brown *et al.*, *Vector-Borne and Zoonotic Diseases* **11**, 1439–1446, ISSN: 1557-7759, 1530-3667, DOI 10.1089/vbz.2010.0196 (2011).
102. J. Z. Adjemian *et al.*, *American Journal of Tropical Medicine and Hygiene* **76**, 365–375, ISSN: 0002-9637, 1476-1645, DOI 10.4269/ajtmh.2007.76.365 (2007).
105. D. E. Biggins, M. Y. Kosoy, *Journal of Mammalogy* **82**, 906–916, ISSN: 0022-2372, DOI 10.1644/1545-1542(2001)082<0906:IOIPON>2.0.CO;2 (2001).
106. D. E. Biggins, M. Y. Kosoy, *Journal of the Idaho Academy of Science and Engineering* **37**, 62–66 (2001).
107. S. N. Bevins *et al.*, *Integrative Zoology* **7**, 99–109, ISSN: 1749-4877, DOI 10.1111/j.1749-4877.2011.00277.x (2012).
108. J. R. Tucker *et al.*, *California compendium of plague control*, 2015, (2018; https://www.cdph.ca.gov/Programs/CID/DCDC/CDPH%20Document%20Library/CAPlagueCompendium.pdf).
109. K. L. Gage *et al.*, in *Proceedings of the Sixteenth Vertebrate Pest Conference*, pp. 200–206.
112. P. W. Willeberg *et al.*, *American Journal of Epidemiology* **110**, 328–334, ISSN: 0002-9262, DOI 10.1093/oxfordjournals.aje.a112818 (1979).
113. C. U. Thomas, P. E. Hughes, *Journal of Wildlife Diseases* **28**, 610–613, ISSN: 0090-3558, DOI 10.7589/0090-3558-28.4.610 (1992).
114. N. W. Dyer, L. E. Huffman, *Journal of Wildlife Diseases* **35**, 600–602, ISSN: 0090-3558, DOI 10.7589/0090-3558-35.3.600 (1999).
115. A. Malmlov *et al.*, *Journal of Wildlife Diseases* **50**, 946–950, ISSN: 1943-3700, DOI 10.7589/2014-03-065 (2014).
116. B. R. Hoar *et al.*, *Preventive Veterinary Medicine* **56**, 299–311, ISSN: 0167-5877, DOI 10.1016/S0167-5877(02)00194-0 (2003).
117. R. J. Brinkerhoff *et al.*, *Vector-Borne and Zoonotic Diseases* **9**, 491–497, ISSN: 1557-7759, 1530-3667, DOI 10.1089/vbz.2008.0075 (2009).
118. L. A. Baeten *et al.*, *Journal of Wildlife Diseases* **49**, 932–939, ISSN: 0090-3558, DOI 10.7589/2013-02-040 (2013).
119. K. L. Wolff, B. W. Hudson, *Applied and Environmental Microbiology* **28**, 323–325, ISSN: 0003-6919 (1974).
120. J. C. Chandler *et al.*, *Journal of Clinical Microbiology* **56**, e00273–18, ISSN: 0095-1137, 1098-660X, DOI 10.1128/JCM.00273-18 (2018).
121. Core Science Analytics, Synthesis, & Libraries Program of the U.S. Geological Survey (USGS), *Biodiversity Information Serving Our Nation (BISON)*, Washington, D.C., 2012.
122. GBIF.org, *GBIF occurrence download*, Accessed: 1 February 2018, DOI 10.15468/dl.7dnihd.
123. VertNet, *Query: specificepithet:latrans genus:Canis mappable:1*, Accessed: 30 January 2018, (http://vertnet.org).
125. C. Daly *et al.*, *International Journal of Climatology* **28**, 2031–2064, ISSN: 1097-0088, DOI 10.1002/joc.1688 (2008).
126. E. M. Hart, K. Bell, *prism: download data from the Oregon PRISM project*, R package version 0.0.6, DOI 10.5281/zenodo.33663, (http://github.com/ropensci/prism).
127. R Core Team, *R: a language and environment for statistical computing*, R Foundation for Statistical Computing (Vienna, Austria, 2018), (https://www.R-project.org/).
144. R. M. Nowak, *Walker's mammals of the world* (Johns Hopkins University Press, Baltimore, Maryland, 1999), vol. 1, ISBN: 0801857899.

145. V. Howard Jr, G. G. DelFrate, in *Great Plains Wildlife Damage Control Workshop Proceedings* (University of Nebraska-Lincoln, Lincoln, Nebraska, 1991), pp. 39–49.
146. A. M. Barnes, *Symposia of the Zoological Society of London* **50**, 237–270 (1982).
147. C. R. Smith *et al.*, *Journal of Vector Ecology* **35**, 1–12, ISSN: 1081-1710, DOI 10.1111/j.1948-7134.2010.00051.x (2010).
159. S. M. Nusser *et al.*, *Journal of Wildlife Management* **72**, 52–60, ISSN: 0022-541X, DOI 10.2193/2007-317 (2008).
161. L. Brotons *et al.*, *Ecography* **27**, 437–448, ISSN: 0906-7590, DOI 10.1111/j.0906-7590.2004.03764.x (2004).
165. E. J. Pebesma, R. S. Bivand, *R News* **5**, 9–13, ISSN: 1609-3631, (https://CRAN.R-project.org/doc/Rnews/) (2005).
166. R. S. Bivand *et al.*, *Applied spatial data analysis with R* (Springer, New York, New York, ed. 2, 2013), ISBN: 1461476177, 1461476184, DOI 10.1007/978-1-4614-7618-4, (http://www.asdar-book.org/).
167. E. H. Boakes *et al.*, *PLoS Biology* **8**, e1000385, ISSN: 1545-7885, DOI 10.1371/journal.pbio.1000385 (2010).
168. W. Yang *et al.*, *Journal of Biogeography* **40**, 1415–1426, ISSN: 0305-0270, DOI 10.1111/jbi.12108 (2013).
169. J. Beck *et al.*, *Ecological Informatics* **19**, 10–15, ISSN: 1574-9541, DOI 10.1016/j.ecoinf.2013.11.002 (2014).
171. R. Kays, *Canis latrans*, 2018, (10.2305/IUCN.UK.2018-2.RLTS.T3745A103893556.en).
181. M. C. Chu, *Laboratory manual of plague diagnostic tests* (World Health Organization, Geneva, Switzerland, 2000).
182. T. J. Quan *et al.*, in *Diagnostic procedures for bacterial, mycotic, and parasitic infections*, ed. by A. Barlows, W. J. Hausler (American Public Health Association, Washington, D.C., ed. 6, 1981), pp. 723–745, ISBN: 0875530869.
183. B. W. Hudson *et al.*, *Epidemiology & Infection* **60**, 443–450, ISSN: 0950-2688, DOI 10.1017/S002217240002057X (1962).
184. J. C. C. Neale, B. N. Sacks, *Oikos* **94**, 236–249, ISSN: 0030-1299, DOI 10.1034/j.1600-0706.2001.940204.x (2001).
185. G. A. Feldhamer *et al.*, Eds., *Wild mammals of North America: biology, management, and conservation* (Johns Hopkins University Press, Baltimore, Maryland, ed. 2, 2003), ISBN: 0801874165.
196. J. L. Lowell *et al.*, *Journal of Clinical Microbiology* **43**, 650–656, DOI 10.1128/JCM.43.2.650-656.2005 (2005).
197. J. C. Gower, *Biometrics*, 857–871, ISSN: 1541-0420, 0006-341X, DOI 10.2307/2528823 (1971).
198. B. Leutner *et al.*, *RStoolbox: tools for remote sensing data analysis*, R package version 0.2.3, (https://CRAN.R-project.org/package=RStoolbox).
208. B. McCune, *Nonparametric multiplicative regression for habitat modeling* (Oregon State University, Corvallis, Oregon, 2009).
211. J. E. Kelsall, P. J. Diggle, *Statistics in Medicine* **14**, 2335–2342, ISSN: 1097-0258, DOI 10.1002/sim.4780142106 (1995).
212. J. E. Kelsall, P. J. Diggle, *Bernoulli* **1**, 3–16, ISSN: 1573-9759, 1350-7265, DOI 10.2307/3318678 (1995).
213. D. C. Wheeler, *International Journal of Health Geographics* **6**, 13, ISSN: 1476-072X, DOI 10.1186/1476-072X-6-13 (2007).

214. H. E. Clough *et al.*, *Preventive Veterinary Medicine* **89**, 67–74, ISSN: 0167-5877, DOI 10.1016/j.prevetmed.2009.01.008 (2009).
215. J. F. Bithell, *Statistics in Medicine* **10**, 1745–1751, ISSN: 1097-0258, DOI 10.1002/sim.4780101112 (1991).
216. J. F. Bithell, *Statistics in Medicine* **9**, 691–701, ISSN: 1097-0258, DOI 10.1002/sim.4780090616 (1990).
219. M. L. Hazelton, T. M. Davies, *Biometrical Journal* **51**, 98–109, ISSN: 1521-4036, 0323-3847, DOI 10.1002/bimj.200810495 (2009).
221. T. M. Davies, M. L. Hazelton, *Statistics in Medicine* **29**, 2423–2437, ISSN: 1097-0258, DOI 10.1002/sim.3995. (2010).
222. A. T. Peterson *et al.*, *Ecological Modelling* **213**, 63–72, ISSN: 0304-3800, DOI 10.1016/j.ecolmodel.2007.11.008 (2008).
223. J. Davis, M. Goadrich, in *Proceedings of the 23rd International Conference on Machine Learning* (Association for Computing Machinery, 2006), pp. 233–240, ISBN: 1595933832, DOI 10.1145/1143844.1143874.
229. G. R. Terrell, *Journal of the American Statistical Association* **85**, 470–477, ISSN: 0162-1459, 1537-274X, DOI 10.2307/2289786 (1990).
275. G. F. A. Ruxton, *Adventures in mexico and the rocky mountains* (John Murray, 1847), vol. 36, (2019; https://catalog.hathitrust.org/Record/000277011).
276. J. Hoogland, *Conservation of the black-tailed prairie dog: saving North America's western grasslands* (Island Press, Washington, D.C., 2013), ISBN: 1559634984, 1597268523.
277. S. D. Melman *et al.*, *Zoonoses and Public Health* **65**, e254–e258, ISSN: 1863-1959, DOI 10.1111/zph.12419 (2018).
278. A. M. Barnes, in *Proceedings of the Symposium on the Management of Prairie Dog Complexes for the Reintroduction of the Black-Footed Ferret* (Biological Report. U.S. Fish and Wildlife Service, Washington, D.C., 1993), vol. 13, pp. 28–38.
279. M. F. Antolin *et al.*, in *Transactions of the 67th North American Wildlife and Natural Resources Conference* (Wildlife Management Institute, Washington, D.C., 2002).
280. D. L. Warren *et al.*, *Diversity and Distributions* **20**, 334–343, ISSN: 1366-9516, DOI 10.1111/ddi.12160 (2014).
281. A. H. Hirzel *et al.*, *Ecological Modelling* **145**, 111–121, ISSN: 0304-3800, DOI 10.1016/S0304-3800(01)00396-9 (2001).
282. T. M. Davies *et al.*, *Statistics in Medicine* **37**, 1191–1221, ISSN: 1097-0258, DOI 10.1002/sim.7577 (2018).
283. D. Orthmeyer, *California state report fiscal year 2012*, 2015, (2019; https://www.aphis.usda.gov/wildlife_damage/informational_notebooks/2015/WS%20State%20Operations/California.pdf).
284. A. M. Barnes *et al.*, *Morbidity and Mortality Weekly Report: Surveillance Summaries*, 11–16, ISSN: 1546-0738, 1545-8636 (1988).
285. D. J. Salkeld, P. Stapp, *Vector-Borne and Zoonotic Diseases* **8**, 331–338, ISSN: 1557-7759, 1530-3667, DOI 10.1089/vbz.2007.0199 (2008).
286. G. Wobeser *et al.*, *The Canadian Veterinary Journal* **50**, 1251, ISSN: 0008-5286, 0008-5286 (2009).
287. R. J. Hijmans *et al.*, *International Journal of Climatology* **25**, 1965–1978, ISSN: 1097-0088, DOI 10.1002/joc.1276 (2005).
288. A. M. Fernández-González *et al.*, *Journal of Medical Entomology* **53**, 199–205, ISSN: 0022-2585, DOI 10.1093/jme/tjv181 (2015).

289. C. Zapata-Valdés *et al.*, *Journal of Wildlife Diseases* **54**, 26–33, ISSN: 0090-3558, DOI 10.7589/2016-09-214 (2018).
290. M. C. Schneider *et al.*, *PLoS Neglected Tropical Diseases* **8**, ISSN: 1935-2735, 1935-2727, DOI 10.1371/journal.pntd.0002680 (2014).
291. L. Kartman, F. M. Prince, *American Journal of Tropical Medicine and Hygiene* **5**, 1058–1070, ISSN: 0002-9637, 1476-1645, DOI 10.4269/ajtmh.1956.5.1058 (1956).
292. A. Dagan *et al.*, *University of Wyoming National Park Service Research Center Annual Report* **24**, 34–36 (2000).
293. F. J. Jannett Jr, *University of Wyoming National Park Service Research Center Annual Report* **20**, 52–56 (1996).
294. E. M. Gese *et al.*, *Journal of Wildlife Diseases* **33**, 47–56, ISSN: 0090-3558, DOI 10.7589/0090-3558-33.1.47 (1997).
295. J. W. Hody, R. Kays, *ZooKeys*, 81, ISSN: 1313-2970, 1313-2970, DOI 10.3897/zookeys.759.15149 (2018).
296. D. J. Salkeld *et al.*, *Journal of Wildlife Diseases* **43**, 425–431, ISSN: 0090-3558, DOI 10.7589/0090-3558-43.3.425 (2007).
297. P. Stapp *et al.*, *Journal of Animal Ecology* **78**, 807–817, ISSN: 0021-8790, DOI 10.1111/j.1365-2656.2009.01541.x (2009).
298. S. T. Nichol *et al.*, *Science* **262**, 914–917, ISSN: 0036-8075, DOI 10.1126/science.8235615 (1993).
299. G. E. Glass *et al.*, *Emerging Infectious Diseases* **6**, 238, ISSN: 1080-6059, 1080-6040, DOI 10.3201/eid0603.000303 (2000).
300. H. Qiao *et al.*, *Ecography* **39**, 805–813, ISSN: 0906-7590, DOI 10.1111/ecog.01961 (2016).
301. D. L. Warren *et al.*, *Ecography* **33**, 607–611, ISSN: 0906-7590, DOI 10.1111/j.1600-0587.2009.06142.x (2010).
302. R. C. Abbott *et al.*, *EcoHealth* **9**, 243–250, ISSN: 1612-9202, 1612-9210, DOI 10.1007/s10393-012-0783-5 (2012).
303. J. Hu *et al.*, *Human Vaccines & Immunotherapeutics* **14**, 2701–2705, ISSN: 2164-554X, DOI 10.1080/21645515.2018.1486154 (2018).
304. R. Seager *et al.*, *Earth Interactions* **22**, 1–22, ISSN: 1087-3562, DOI 10.1175/EI-D-17-0011.1 (2018).
305. R. Seager *et al.*, *Earth Interactions* **22**, 1–24, ISSN: 1087-3562, DOI 10.1175/EI-D-17-0012.1 (2018).

5.6 Appendices

5.6.1 Appendix A: Panels

Panel 7: Diagnostic tests for *Yersinia pestis* in wildlife
The following are four of the recognized diagnostic tests for *Yersinia pestis* infection:
Passive hemagglutination Assay (PHA) & F1-inhibitation (PHI) test
- Antigen-antibody reactivity is visible qualitatively using F1 antigen coated on glutaraldehyde-fixed sheep red blood cells as a sensitizing antigen. If serum reacts positively, a PHI test is run to verify the specificity of the PHA test that calculates the titer.
- Positive Test: At least fourfold change in antibody titer, specific for F1 antigen of *Y. pestis*.
- Strength: Sensitive and generally reproducible.
- Limitations: Relies on unstable reagent, interpretation is fairly subjective, and prone to nonspecific reactivity of natural antibodies.
- More detailed procedures found in (*181*).

F1 Luminex Plague Assay (F1-LPA)
- Semi-automated bead-based flow cytometric assay, specific for the F1 antigen of *Y. pestis*.
- Positive Test: With a baseline background noise of 250 mean fluorescent intensity (MFI), a Signal to Noise Ratio (S/N) $\geq$ 10 (or $\geq$ 2,500 MFI) is considered a positive test.
- Strength: More sensitive (x64) than PHA-PHI, fewer false negative results.
- Limitation: Has not been assessed for human tissue.
- More detailed procedures found in (*120*).

Fluorescent Antibody (FA)
- Smears of suspected tissue are prepared with plague antiserum and examined via fluorescent microscopy.
- Positive Test: Smear brightly fluoresce as "apple green-colored hollow rods" (Or less brightly if conducting a fluorescence inhibition test).
- Strength: Quick assay and requires a small amount of test material.
- Limitations: Sensitive with fresh tissue samples, not optimized for field collection.
- More detailed procedures found in (*182*, *183*).

Bacterial Culture
- Suspected tissue samples are suspended in blood agar and examined for growth daily for at least 7-10 days.
- Positive Test: Bacterial growth appears, typically a "stalactite"-type pattern or colonies about 4-7 mm in diameter.
- Strength: The gold-standard assay.
- Limitation: Time intensive. *Y. pestis* is slow growing.
- More detailed procedures found in (*182*).

Panel 8: Considerations for coyotes as a sentinel species for plague activity

Coyotes (*Canis latrans*) have been used as a sentinel species of plague (*Yersinia pestis*) activity in rodent populations of North America (*95*, *112–116*), but there are limitations for focusing on coyote-based plague surveillance data. I outline some notable limitations below:

- Coyotes are a wide-ranging species (*144*, *145*, *171*) and effectively sample a broad variety of habitats in which plague occurs and thus sample a considerable extent of the ecological niche of plague (*85*). However, coyotes are not observed homogeneously across North America (Chapter 5) and other species may be better indicators of plague activity in some habitats (e.g., rodent species for alpine regions of California). Therefore, without rodent data, I am estimating a the fundamental ecological niche of plague in coyotes, which I am considering as an approximation for the entire fundamental ecological niche of plague.
- The location where an individual coyote was truly exposed to *Y. pestis* is unknown and challenging to discern (*91*, *117*). Coyotes are mobile, encountering many rodents across their expansive individual home range (*144*, *145*, *184*) and whose plague antibodies can last for many months (*91*, *112*, *118*). Reinfection of coyotes is also probable and so a coyote can only indicate its most recent plague exposure. Rodent species have smaller individual home range sizes (*185*) and are the gold standard indicator of current or recent plague activity (*108*).
- Coyote specimens are not collected evenly across the United States and instead are collected opportunistically (*107*). Coyotes tested for plague exposure are primarily collected in conjunction with ongoing livestock/wildlife damage management operations conducted by the U.S. Department of Agriculture Animal and Plant Health Inspection Service Wildlife Services. Agricultural and urban areas may be more sampled for plague than other land-use types and their respective habitats. Including other species monitored for plague activity may help overcome this potential spatial sampling effort bias.

5.6.2 Appendix B: Tables

Table 12: Summary of *Canis latrans* testing for antibodies against *Yersinia pestis* by California Department of Public Health (CPDH; 1983–2015), U.S. Department of Agriculture (USDA; 2005–2017) and U.S. Centers for Disease Control and Prevention (CDC; 1998–2008).

Agency	Plague Result	Count Tested	Count With Location	Prevalence
CDPH	Positive	705	704	8.8%
	Negative	7,414	7,251	
USDA	Positive	3,665	3,648	13.0%
	Negative	25,082	24,381	
CDC	Positive	2,516	2,360	
	Negative	Not digitized	Not digitized	

Table 13: Missingness of coyote specimens by U.S. Department of Agriculture (2005–2017) and California Department of Public Health (CDPH; 1983–2015). Specimens were considered missing if textual geographic location information was unavailable or indiscernible to geocode. Fewer than 2.5% of specimens were missing and ignored in subsequent analysis.

State	Missing	Total
Arizona	33	3,097
California*	164	8,199
Colorado	3	2,161
Idaho	1	576
Kansas	2	240
Montana	43	6,272
Nebraska	5	586
New Mexico	42	5,998
Nevada	9	3,269
Oklahoma	12	1,177
Oregon	3	543
South Dakota	5	521
Texas	16	1,084
Utah	3	487
Washington†	540	590
Wyoming	0	1,224

*Missing specimens were located in 28 counties (∼48% of counties in California). Over 30% of missing specimens were located in Santa Clara County, California and San Luis Obispo County, California.
†All missing specimens were located in Yakima County, Washington.

Table 14: Sources of Independent Coyote Observations in the United States

Database	Coyote Observations
Angelo State Natural History Collections	100
California Academy of Sciences	59
California State University, Long Beach	2
Chicago Academy of Sciences	11
Cornell University Lab of Ornithology	42
Cornell University Museum of Vertebrates	36
Denver Museum of Nature & Science	83
Fort Hays Sternberg Museum of Natural History	70
Florida Museum of Natural History	120
Humboldt State University	1
iNaturalist.org	4,485
Illinois State University	3
James R. Slater Museum of Natural History	78
Louisiana State University Museum of Natural Science	43
Michigan State University Museum	91
Museum of Cultural and Natural History–Central Michigan University	4
Museum of Comparative Zoology, Harvard University	4
Museum of Southwestern Biology	458
Museum of Texas Tech University	250
Museum of Vertebrate Zoology	987
Natural History Museum of Geneva	3
Natural History Museum of Los Angeles County	166
naturgucker.de	11
New Mexico Museum of Natural History and Science	1
National Museum of Natural History, Smithsonian Institution	1
National Parks Service	207
Natural History Museum of Utah	186
New York State Museum	528
North Carolina Museum of Natural Sciences	86
Northern Michigan University	5
The Ohio State University Borror Lab of Bioacoustics	139
The Ohio State University Museum of Biological Diversity	363
Royal Ontario Museum	46
Sam Noble Oklahoma Museum of Natural History	556
Santa Barbara Museum of Natural History	37
Tall Timbers Research Station and Land Conservancy	1
Texas A&M University Biodiversity Research and Teaching Collections	44
United States Geological Survey (USGS)	30
USGS Western Ecological Research Center San Diego Field Station	2,899
University of Alaska Museum of the North	93
University of Arizona Museum of Natural History	31
University of California–Los Angeles Dickey Collection	48
University of California–Santa Barbara Marine Science Institute	62
University of Colorado Museum of Natural History	1
University of Kansas Biodiversity Institute	676
University of Michigan Museum of Zoology	257
University of Texas at El Paso Biodiversity Collections	409
University of Washington Burke Museum	90
University of Wyoming Museum of Vertebrates	9
Yale University Peabody Museum	59

Table 15: Oregon State University Parameter elevation Regression on Independent Slopes Model (PRISM) 30-Year Average Annual Normals (1981-2010). Modeled using a combination of a digital elevation model (DEM) and climatologically-aided interpolation (CAI)*. Variables were modeled at 30 arcseconds (∼800 m) resolution and aggregated to 2.5 arcminutes (∼4 km). See (*125*) for more details.

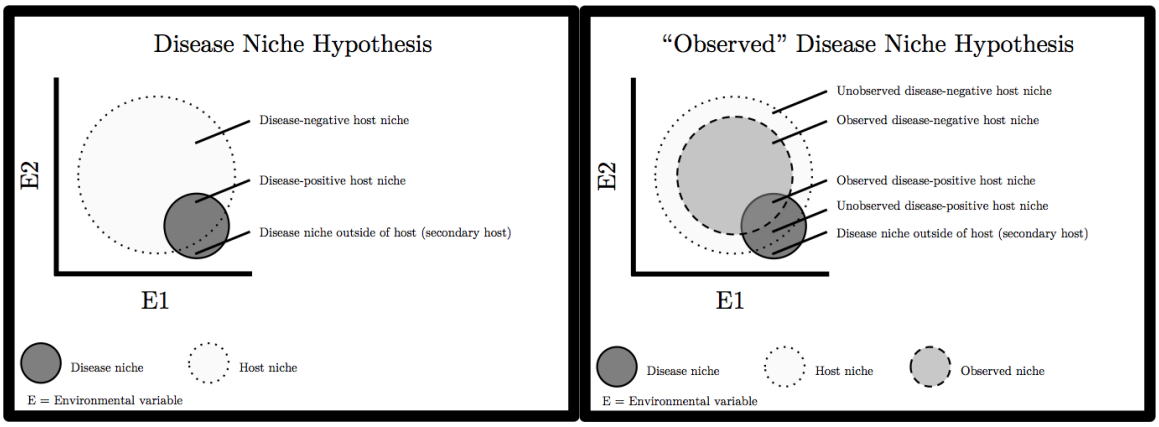
Variable	Units	Derivation
Precipitation	millimeters (mm)	Modeled; Summing monthly averages (rain + melted snow)
Maximum Temperature	°Celsius (°C)	Modeled; Averaging over all months using a DEM as the predictor grid
Mean Temperature	°Celsius (°C)	Derived; Average of Maximum Temperature and Minimum Temperature
Minimum Temperature	°Celsius (°C)	Modeled; Averaging over all months using a DEM as the predictor grid
Mean Dewpoint Temperature	°Celsius (°C)	Modeled; CAI used minimum temperature as the predictor grid
Maximum Vapor Pressure Deficit	hectopascal (hPA)	Modeled; CAI used mean dewpoint temperature and maximum temperature as the predictor grids
Minimum Vapor Pressure Deficit	hectopascal (hPA)	Modeled; CAI used mean dewpoint temperature and minimum temperature as the predictor grids

*Accuracy of these data is based on the original specification of the Defense Mapping Agency one-degree DEMs. The stated accuracy of the original DEMs is 130-meter circular error with 90% probability. Data sets use all weather stations, regardless of time of observation.

Table 16: Summary of principal component analysis of Oregon State University Parameter elevation Regression on Independent Slopes Model (PRISM) 30-Year Average Annual Normals (1981-2010) at a 2.5 arcminute (∼4 km) resolution. Variables were masked to the western United States and were range standardized using the Gower metric (*197*).

Statistic	**Variable**	**Principal Component**		
		PC1	PC2	PC3
Standard Deviation		0.326	0.170	0.058
Proportion of Variance		0.756	0.205	0.024
Cumulative Proportion of Variance		0.756	0.962	0.985
Loadings	Precipitation	0.063	0.342	0.549
	Maximum Temperature	0.485	-0.208	-0.290
	Mean Temperature	0.519	-0.064	0.015
	Minimum Temperature	0.477	0.088	0.314
	Mean Dewpoint Temperature	0.446	0.532	-0.101
	Maximum Vapor Pressure Deficit	0.243	-0.533	-0.229
	Minimum Vapor Pressure Deficit	0.073	-0.510	0.673

5.6.3 Appendix C: Figures



(a) **(b)**

Figure 49: **(a)** The disease niche hypothesis postulated for *Yersinia pestis* by Maher and colleagues (*85*) where the ecological niche of *Y. pestis* is independent of its host niches and plague occurs in hosts only in areas with suitable conditions for *Y. pestis* transmission. **(b)** The proposed "observed" disease niche hypothesis that augments the disease niche hypothesis for *Y. pestis* by including the region of predictor space that is sampled and tested for plague. The bacterium and its hosts may occur in areas not sampled and tested for plague.

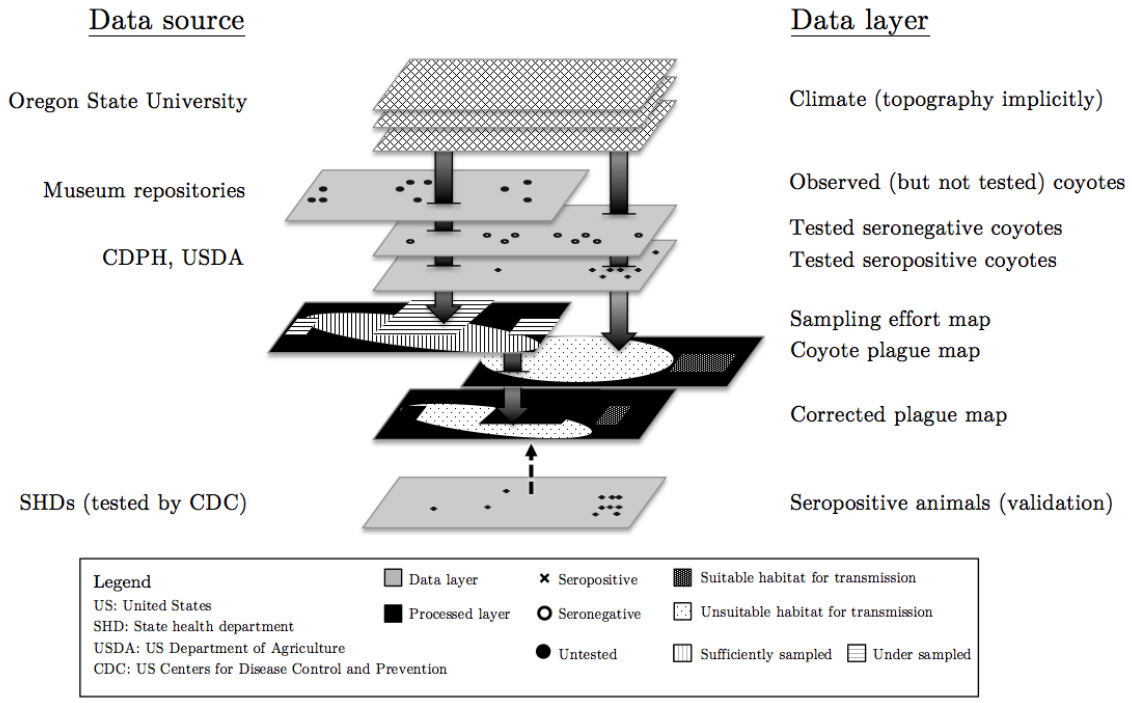


Figure 50: Data layers of the present analysis and their sources. Data are structured as point locations or raster grids. Three layers are predicted (black color) using data layers (grey color). Arrows represent the layers involved for each predicted layer.

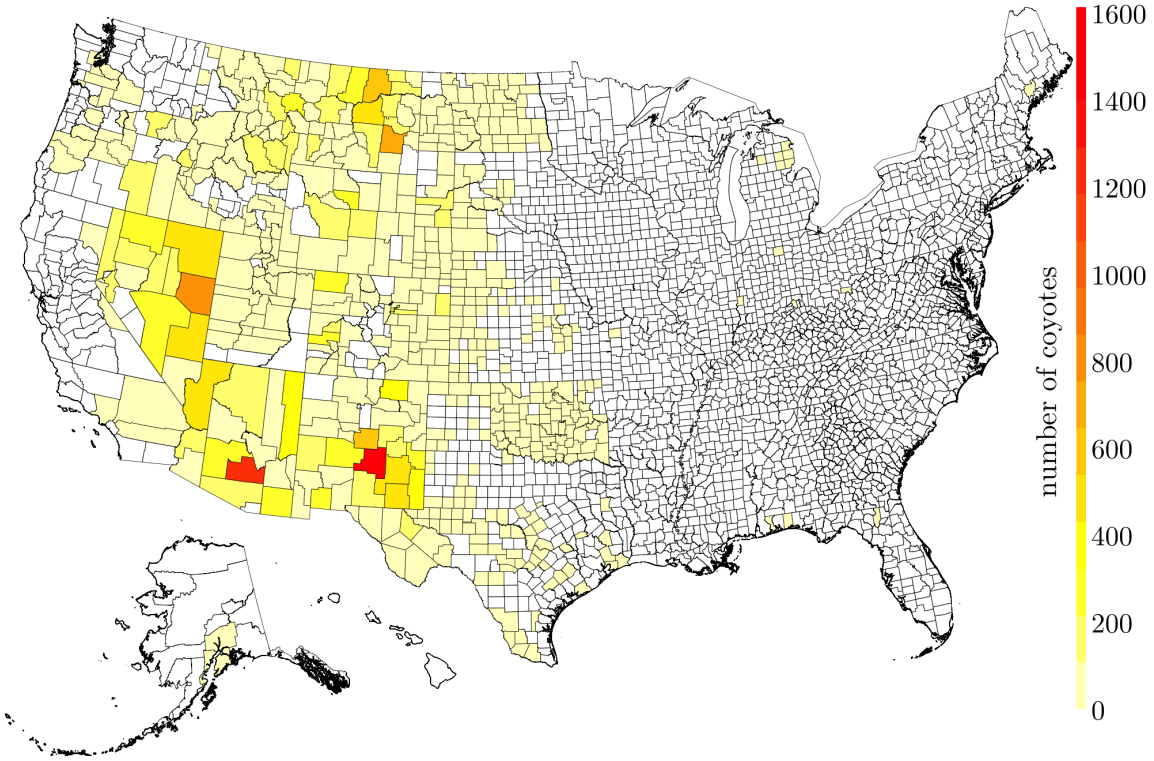


Figure 51: Number of coyotes tested by U.S. Department of Agriculture Animal and Plant Health Inspection Service National Wildlife Disease Program for antibodies against *Yersinia pestis* (2005–2017). Sampling is heterogeneous across the United States, primarily in historic plague enzootic areas. Data for coyotes from California are managed by the California Department of Public Health and were not included with these data.

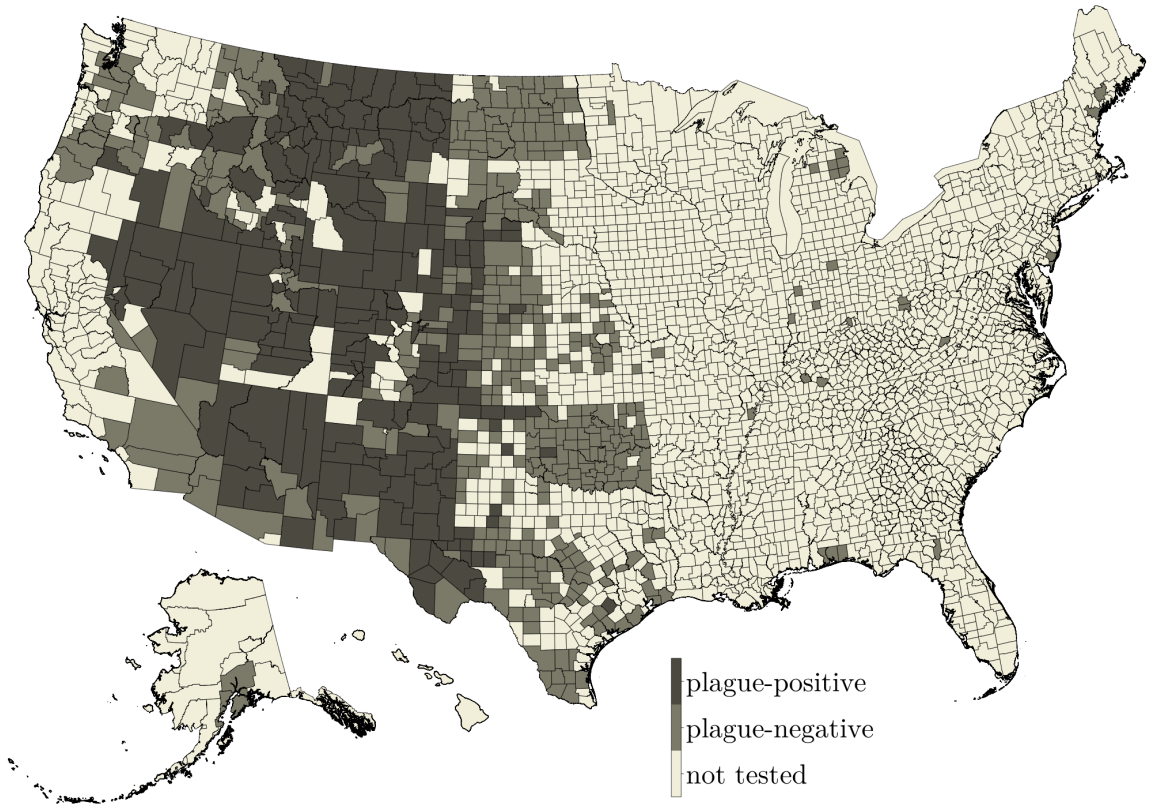
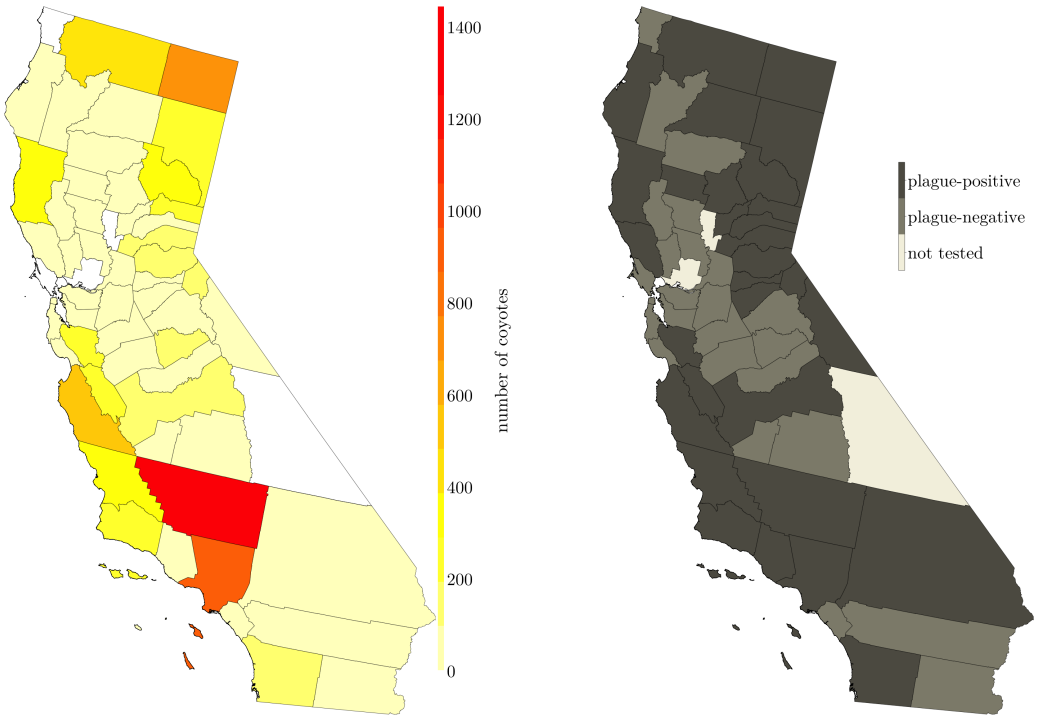


Figure 52: Results of coyotes tested by U.S. Department of Agriculture Animal and Plant Health Inspection Service National Wildlife Disease Program for antibodies against *Yersinia pestis* (2005–2017). A coyote with plague antibodies was not observed in every county sampled. Counties with a coyote that tested plague-positive are found throughout the western region of the United States. A figure of crude rates is not included



(a) **(b)**

Figure 53: (a) Number of coyotes tested by the California Department of Public Health (CDPH) for antibodies against *Yersinia pestis* (1983–2015). Sampling is heterogeneous across the state of California, primarily in historic plague enzootic areas and areas of high concern. Limited sampling in the Mojave and Sonoran Deserts. **(b)** Results of coyotes tested by CDPH for antibodies against *Y. pestis* (1983–2015). A coyote with plague antibodies was not observed in every county sampled. Only 2% of observations were unable to be geolocated and ignored in subsequent analysis.

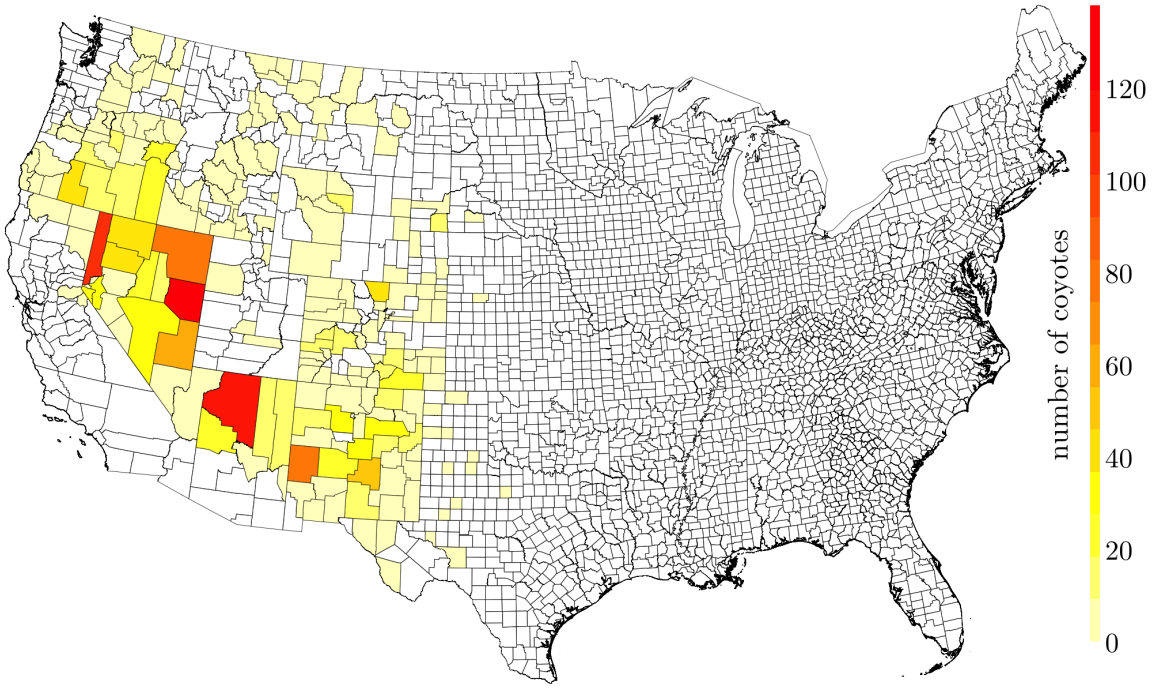


Figure 54: Number of coyotes sampled by state health departments and tested by the U.S. Centers for Disease Control and Prevention (CDC) for antibodies against *Yersinia pestis* (1998–2008). All coyotes in this data set tested positive for plague antibodies because coyotes that tested negative were not electronically available. See (*85*) for more details. Possible duplicate coyote specimens between the U.S. Department of Agriculture and the CDC were removed from the CDC data and are not presented here.

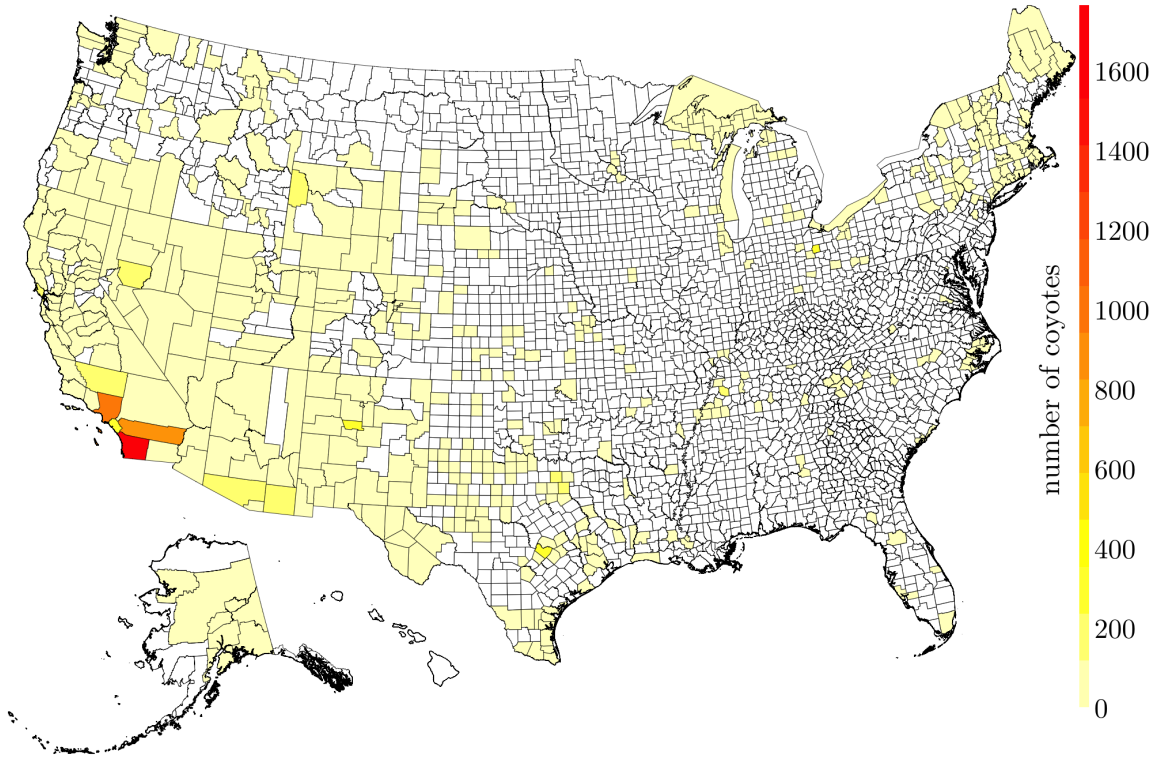


Figure 55: Number of coyote observations with geolocation information collated by various biodiversity repositories (1900-2017). These coyotes were not tested for antibodies against *Yersinia pestis*. Coyotes are found throughout the United States, including Alaska, with higher observations recorded in Southern California and the western United States, in general. See (*121–123*) for more details.

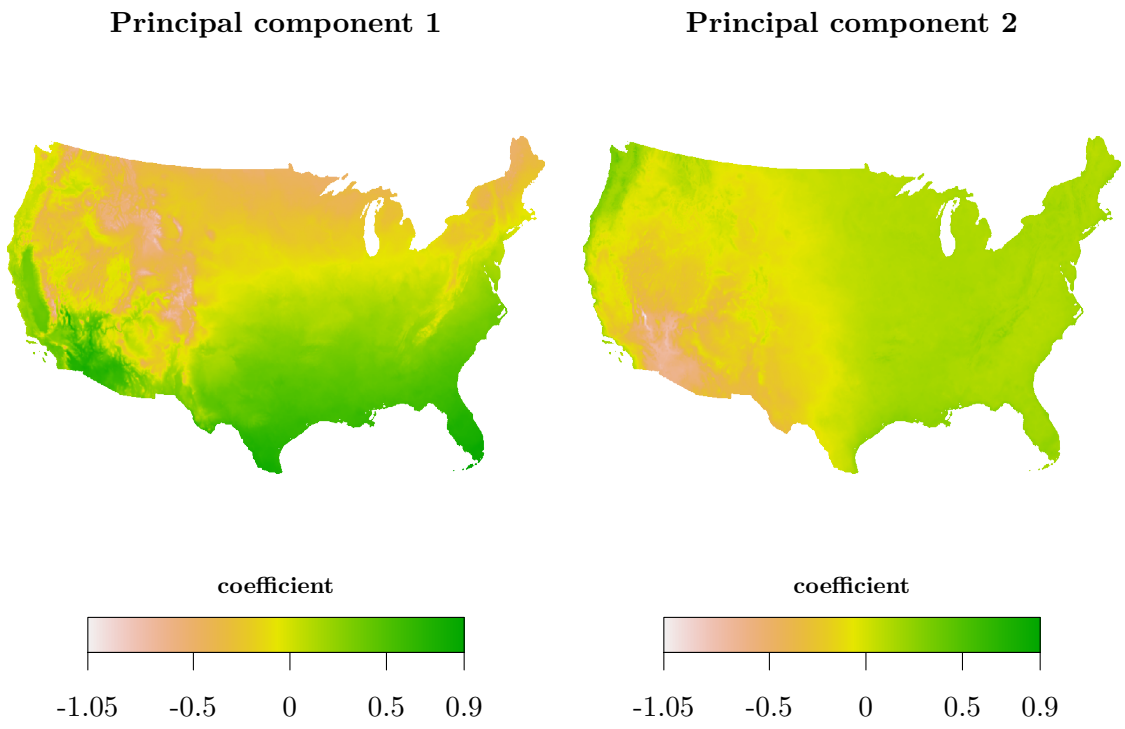


Figure 56: First two principal components of a principal component analysis of Oregon State University Parameter Elevation Regression on Independent Slopes Model 30-year average annual normals (1981-2010) at a 2.5 arcminute (∼4 km) resolution of the contiguous United States. Variables were range standardized using the Gower metric (*197*).

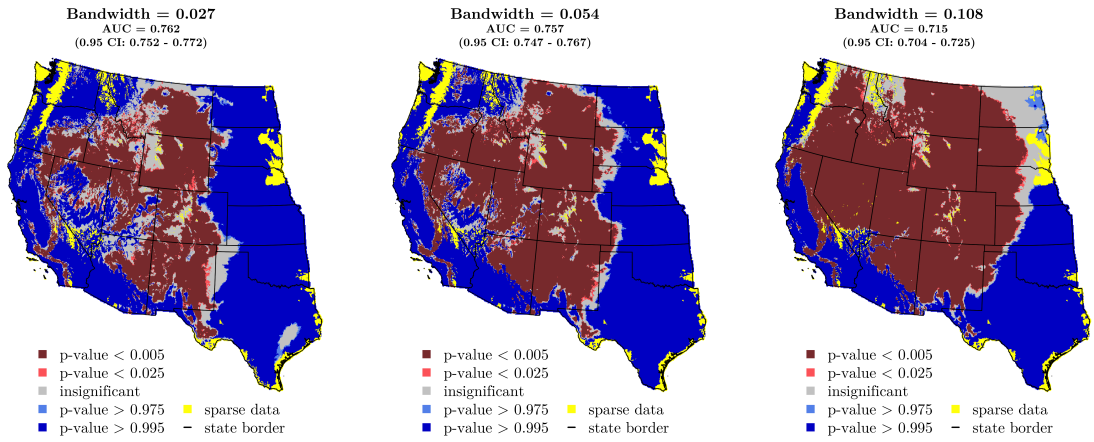


Figure 57: Comparison of bandwidth selection on the predicted area of California with a log relative risk surface in predictor space of coyotes of coyotes that were observed but not tested for plague antibodies (controls) and coyotes that were tested the presence of plague antibodies (cases) by the California Department of Public Health (CDPH; 1983–2015) and U.S. Department of Agriculture (USDA; 2005–2017) using the developed approach. Color pertains to calculated asymptotic tolerances (*219*, *221*) at given local two-tailed significance levels ($\alpha = 0.05$ & $\alpha = 0.01$). Warmer-colored areas are more likely suitable habitat for plague transmission and cooler-colored areas are more likely unsuitable habitat for plague transmission with greyer-colored areas statistically indistinguishable. The areas in yellow coloring correspond to habitat with sparse data. The prediction is robust across smaller bandwidth selection with a notable reduction in habitats the log relative risk surface cannot statistically distinguish as suitable (unsuitable) for plague transmission as bandwidth increases.

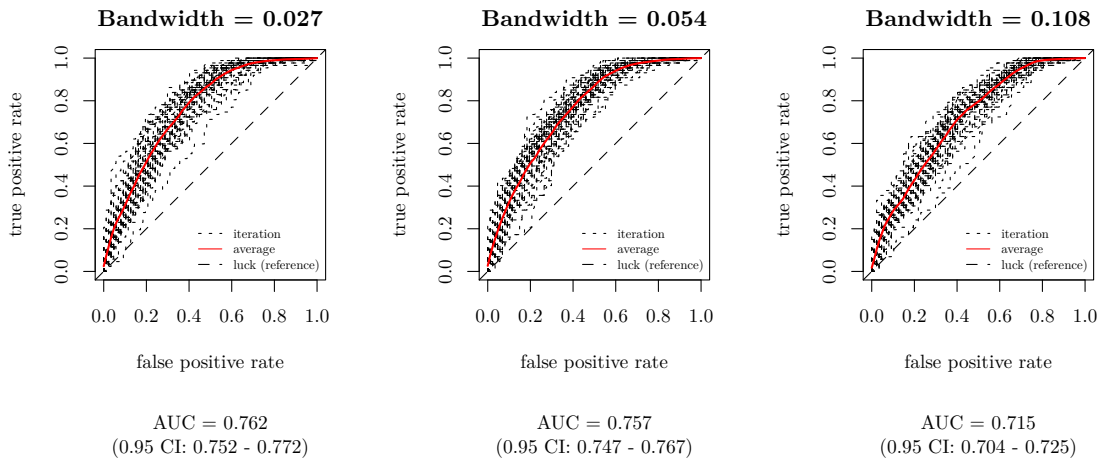


Figure 58: Results of 50-fold cross-validation of a bandwidth selection comparison for the log relative risk surface in predictor space of coyotes that were observed but not tested for plague antibodies (controls) and coyotes that were tested for the presence of plague antibodies (cases) by the California Department of Public Health (CDPH; 1983–2015) and U.S. Department of Agriculture (USDA; 2005–2017) using the developed approach. Iterations were balanced (prevalence = 0.5) by randomly undersampling control locations used for in each fold for cross-validation. The prediction was robust across small bandwidths, but not a larger bandwidth. Area Under the Receiver Operating Characteristic Curve (AUC) was similar between the bandwidth chosen using the maximal smoothing principle (*229*) and a bandwidth half the size. A bandwidth double the size chosen by the maximal smoothing principle (*229*) resulted in poorer cross-validation than other bandwidths.

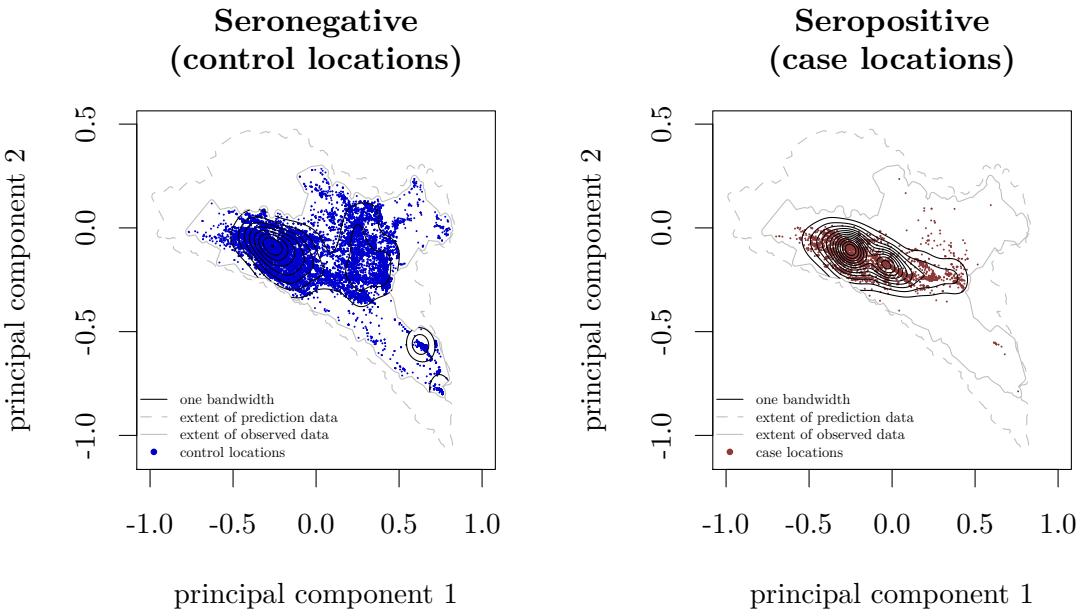


Figure 59: Estimated intensity surface in predictor space of seronegative coyotes (controls) and seropositive coyotes (cases) tested by the California Department of Public Health (CDPH; 1983–2015) and U.S. Department of Agriculture (USDA; 2005–2017) using the developed approach. Predictor space is comprised of the first two principal components of a principal component analysis of seven range-standardized (*197*) Oregon State University Parameter Elevation Regression on Independent Slopes Model 30-year average annual normals (1981-2010) at a 2.5 arcminute (∼4 km) resolution. The bandwidth (0.054) was chosen using the maximal smoothing principle (*229*). The dashed grey line is the entire extent of predictor space of the western United States and the solid grey line is the extent of predictor space that the CDPH and the USDA sampled with coyote specimens. The area outside of the solid grey line is habitat that was neither sampled by the CDPH nor the USDA.

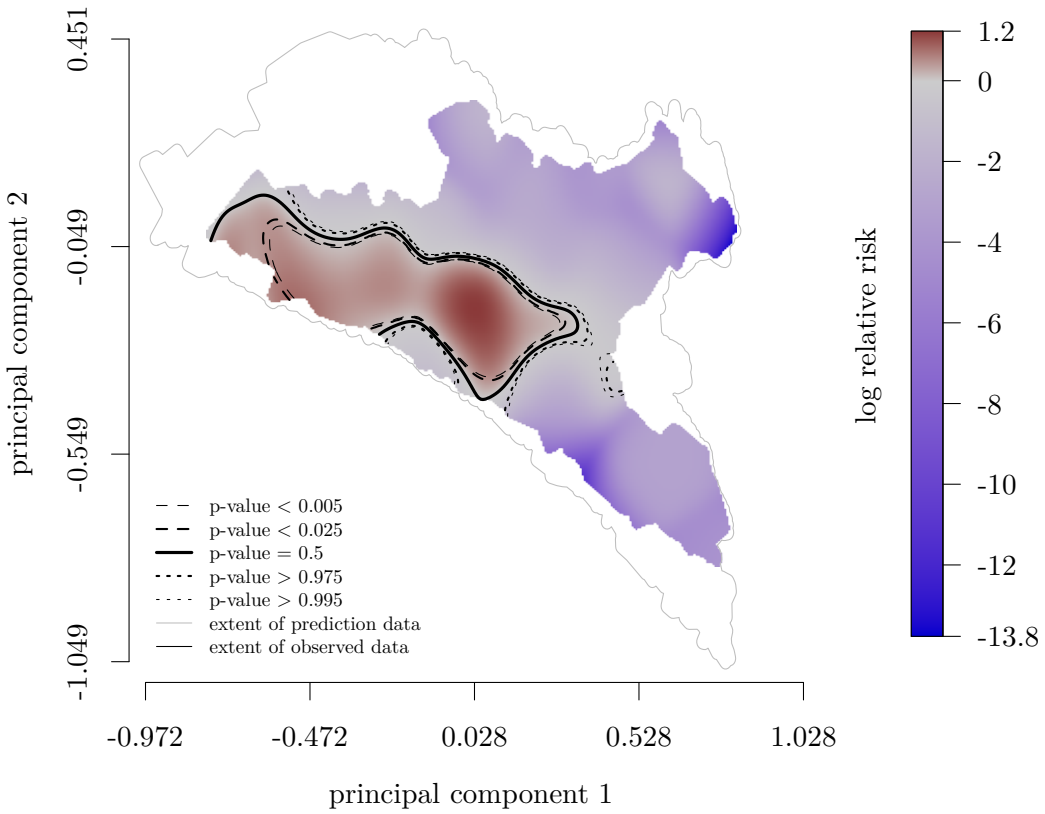


Figure 60: The estimated log relative risk surface in predictor space of seronegative coyotes (controls) and seropositive coyotes (cases) tested by the California Department of Public Health (CDPH; 1983–2015) and the U.S. Department of Agriculture (USDA; 2005–2017) using the developed approach. Calculated asymptotic tolerances at given local two-tailed significance levels ($\alpha = 0.05$ & $\alpha = 0.01$) are included (*219*, *221*). Color pertains to log relative risk values where positive log relative risk (more likely cases) are in red and negative log relative risk (more likely controls) in blue with greyer coloring closer to the 0 null log relative risk value. Predictor space is comprised of the first two principal components of a principal component analysis of seven range-standardized (*197*) Oregon State University Parameter Elevation Regression on Independent Slopes Model 30-year average annual normals (1981-2010) at a 2.5 arcminute (∼4 km) resolution. The bandwidth (0.054) was chosen using the maximal smoothing principle (*229*). The dashed grey line is the entire extent of predictor space of the western United States and the solid grey line is the extent of predictor space that the CDPH and the USDA sampled with coyote specimens. The area outside of the solid grey line is habitat that was neither sampled by the CDPH nor the USDA.

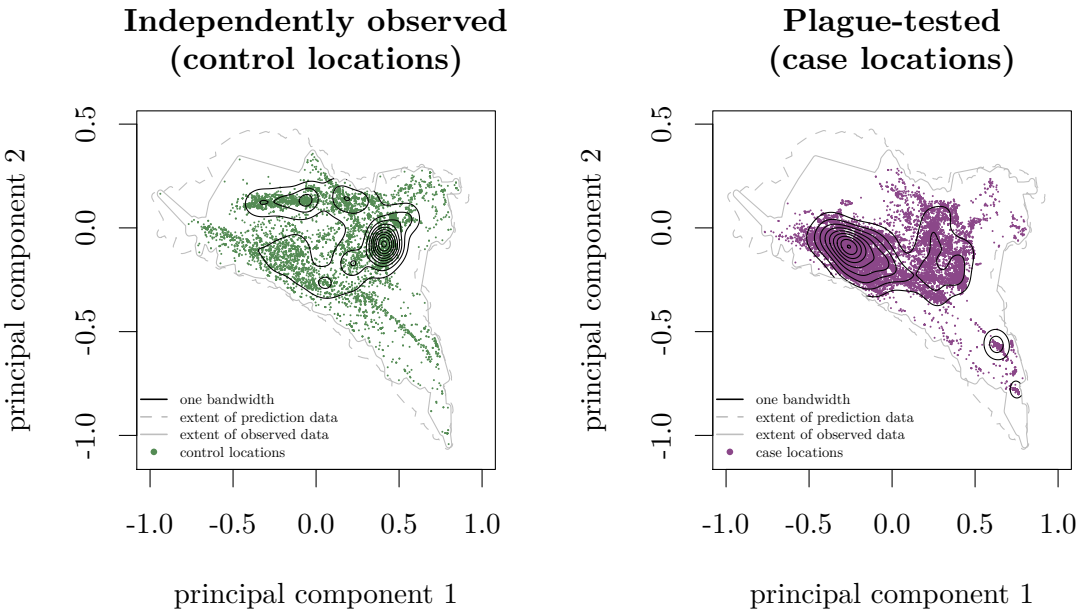


Figure 61: Estimated intensity surface in predictor space of coyotes that were observed but not tested for plague antibodies (controls) and coyotes that were tested the presence of plague antibodies (cases) by the California Department of Public Health (CDPH; 1983–2015) and the U.S. Department of Agriculture (USDA; 2005–2017). Predictor space is comprised of the first two principal components of a principal component analysis of seven range-standardized (*197*) Oregon State University Parameter Elevation Regression on Independent Slopes Model 30-year average annual normals (1981-2010) at a 2.5 arcminute (∼4 km) resolution. The bandwidth (0.054) was chosen using the maximal smoothing principle (*229*). The dashed grey line is the entire extent of predictor space the contiguous United States and the thin solid black line is the extent of predictor space that the CDPH, USDA, and independent investigators observed coyotes. The area outside of the solid grey line is habitat that did not have a coyote observation.

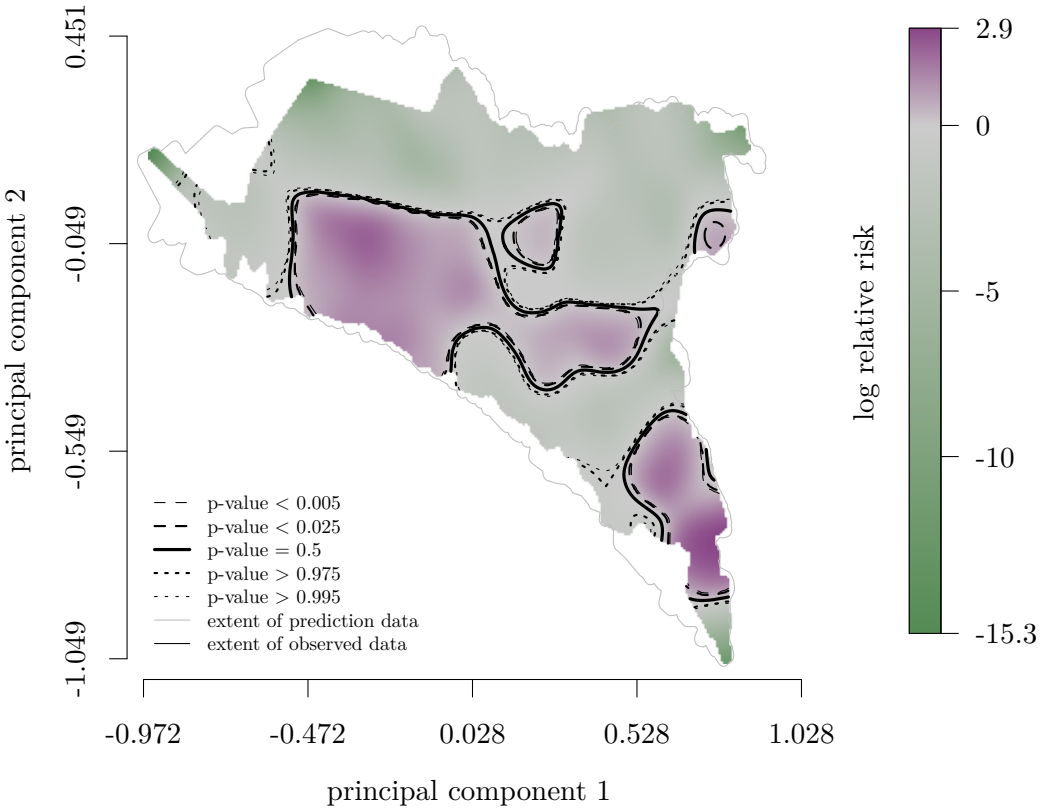


Figure 62: The estimated log relative risk surface in predictor space of coyotes that were observed but not tested for plague antibodies (controls) and coyotes that were tested the presence of plague antibodies (cases) by the California Department of Public Health (CDPH; 1983–2015) and the U.S. Department of Agriculture (USDA; 2005–2017) using the developed approach. Calculated asymptotic tolerances at given local two-tailed significance levels ($\alpha = 0.05$ & $\alpha = 0.01$) are included (*219*, *221*). Color pertains to log relative risk values where positive log relative risk (more likely cases) are in purple and negative log relative risk (more likely controls) are in green with greyer coloring closer to the 0 null log relative risk value. Predictor space is comprised of the first two principal components of a principal component analysis of seven range-standardized (*197*) Oregon State University Parameter Elevation Regression on Independent Slopes Model 30-year average annual normals (1981-2010) at a 2.5 arcminute (∼4 km) resolution. The bandwidth (0.05) was chosen using the maximal smoothing principle (*229*). The dashed grey line is the entire extent of predictor space of the contiguous United States and the thin solid black line is the extent of predictor space that the CDPH, USDA, and independent investigators observed coyotes. The area outside of the solid grey line is habitat that did not have a coyote observation.

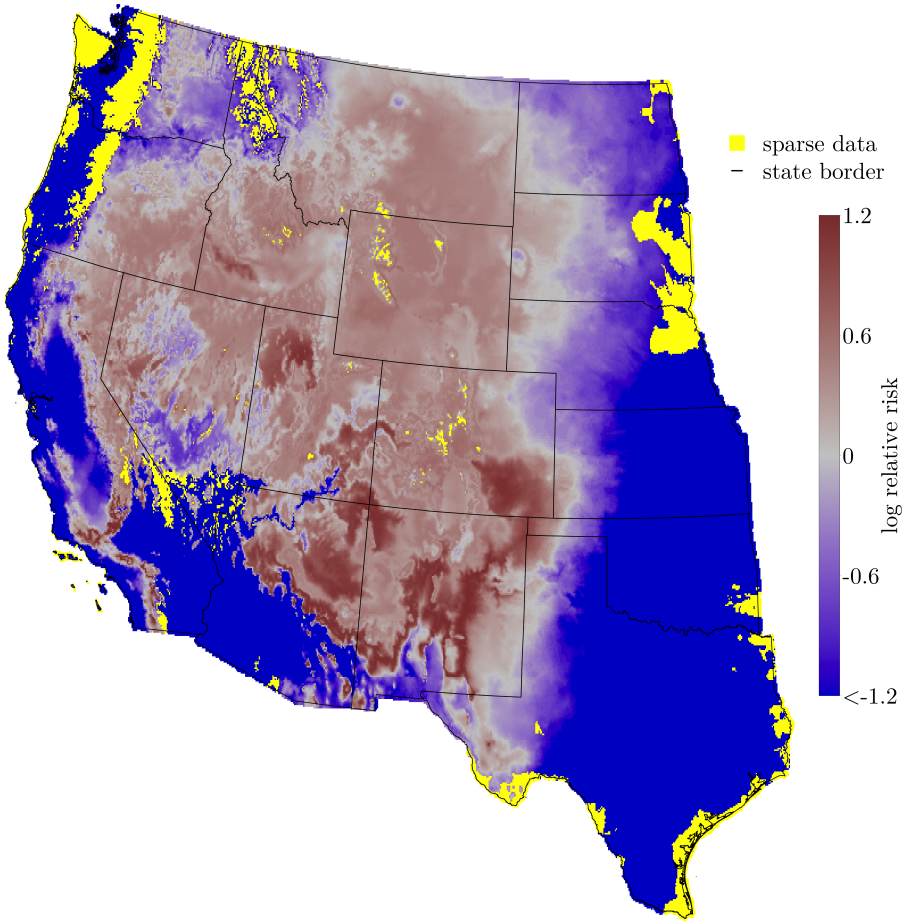


Figure 63: Areas of the western United States predicted with log relative risk surface in predictor space of seronegative coyotes (controls) and seropositive coyotes (cases) tested by the California Department of Public Health (CDPH; 1983–2015) and the U.S. Department of Agriculture (USDA; 2005–2017) using the developed approach. Color pertains to log relative risk values where positive log relative risk (more likely suitable habitat for plague transmission) are in red and negative log relative risk (more likely unsuitable habitat for plague transmission) are in blue with greyer coloring closer to the 0 null log relative risk value. The areas in yellow coloring correspond to habitat that was neither sampled by the CDPH nor the USDA.

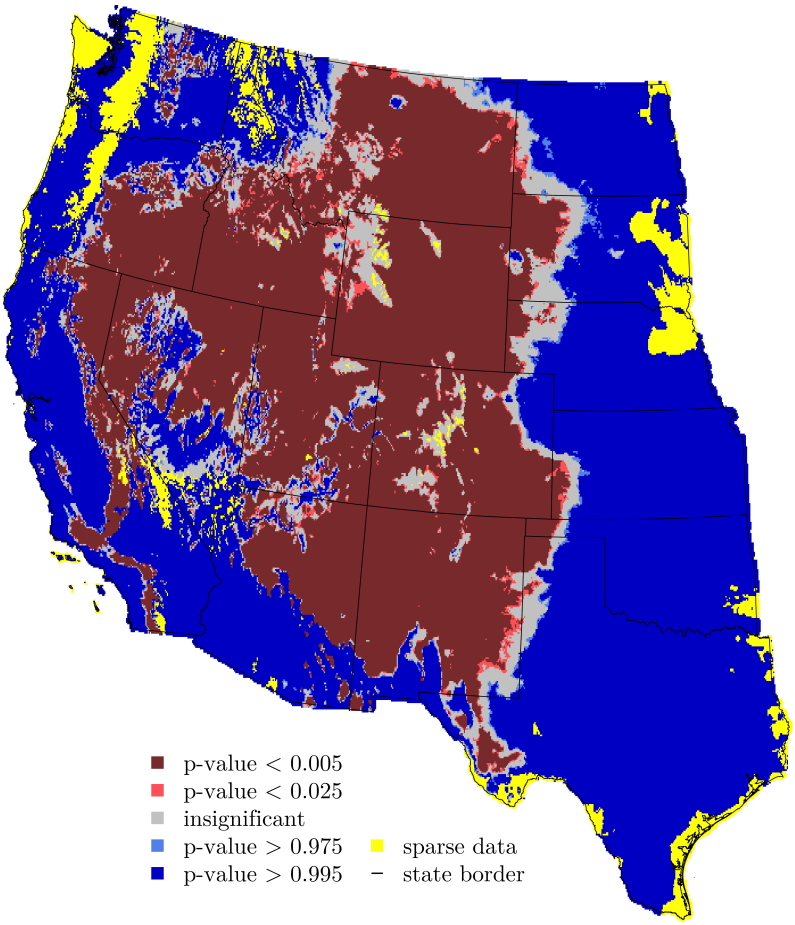


Figure 64: Areas of the western United States predicted with log relative risk surface in predictor space of seronegative coyotes (controls) and seropositive coyotes (cases) tested by the California Department of Public Health (CDPH; 1983–2015) and the U.S. Department of Agriculture (USDA; 2005–2017) using the developed approach. Color pertains to calculated asymptotic tolerances (*219*, *221*) at given local two-tailed significance levels ($\alpha = 0.05$ & $\alpha = 0.01$). Warmer-colored areas are more likely suitable habitat for plague transmission and cooler-colored areas are more likely unsuitable habitat for plague transmission with greyer-colored areas statistically indistinguishable. The areas in yellow coloring correspond to habitat that was neither sampled by the CDPH nor the USDA.

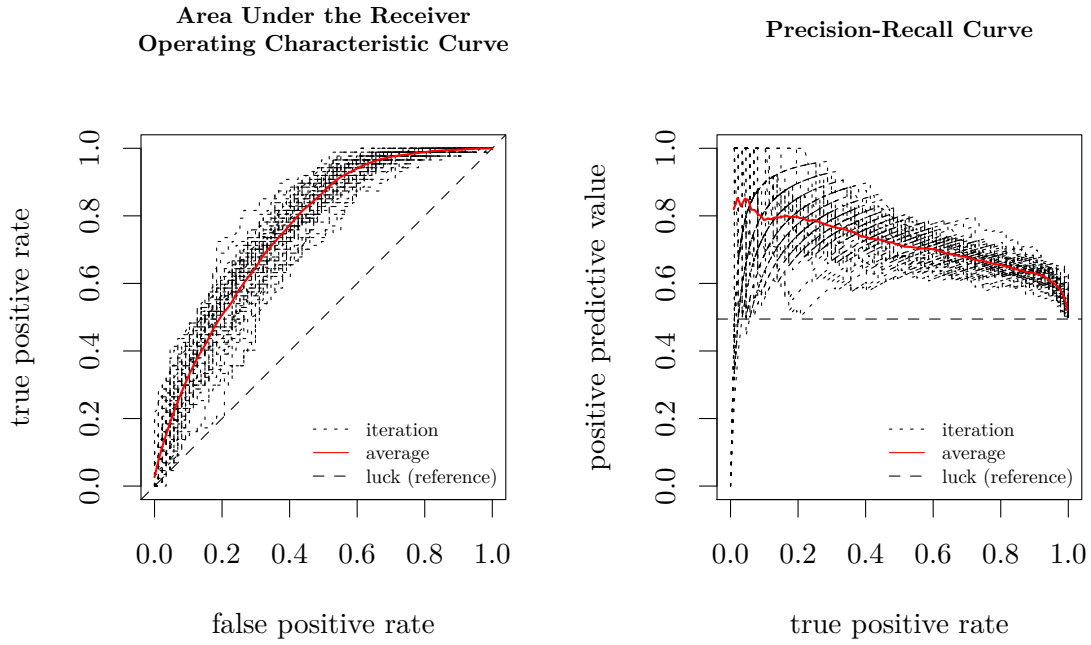


Figure 65: Results of 50-fold cross-validation of the estimated log relative risk surface in predictor space of seronegative coyotes (controls) and seropositive coyotes (cases) tested by the California Department of Public Health (CDPH; 1983–2015) and the U.S. Department of Agriculture (USDA; 2005–2017) using the developed approach. Iterations were balanced (prevalence = 0.5) by randomly undersampling control locations used for in each fold for cross-validation. Results are fairly robust with a high average Area Under the Receiver Operating Characteristic Curve (AUC) of 0.757 (95% CI: 0.747–0.767) and an fairly acceptable precision-recall curve.

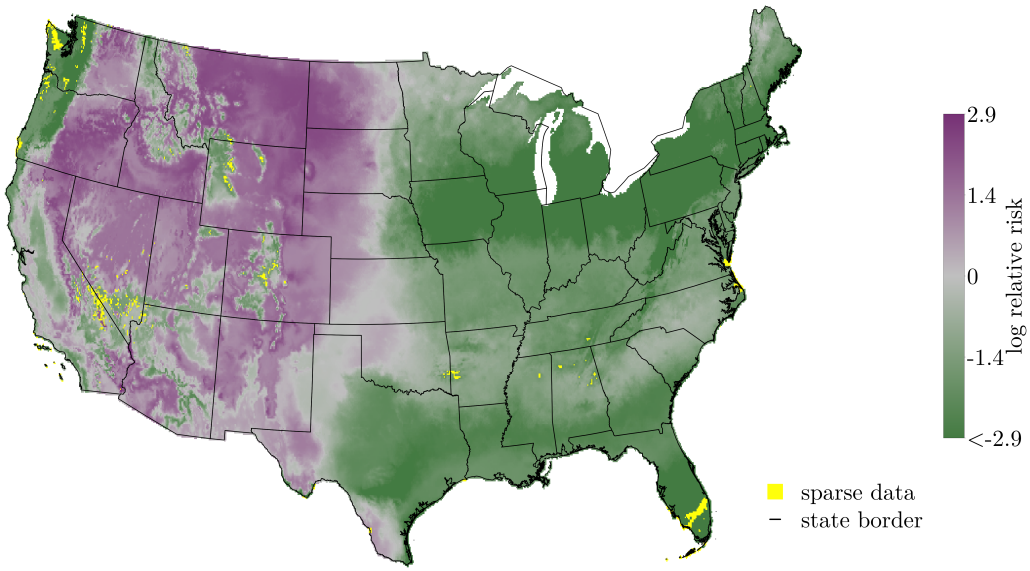


Figure 66: Areas of contiguous United States predicted with log relative risk surface in predictor space of coyotes that were observed but not tested for plague antibodies (controls) and coyotes that were tested the presence of plague antibodies (cases) by the California Department of Public Health (CDPH; 1983–2015) and the U.S. Department of Agriculture (USDA; 2005–2017) using the developed approach. Color pertains to log relative risk values where positive log relative risk (more likely tested for plague antibodies) are in purple and negative log relative risk (more likely not tested for plague antibodies) are in green with greyer coloring closer to the 0 null log relative risk value. The areas in yellow coloring correspond to habitat with sparse data and statistically dissimilar from sampled habitat.

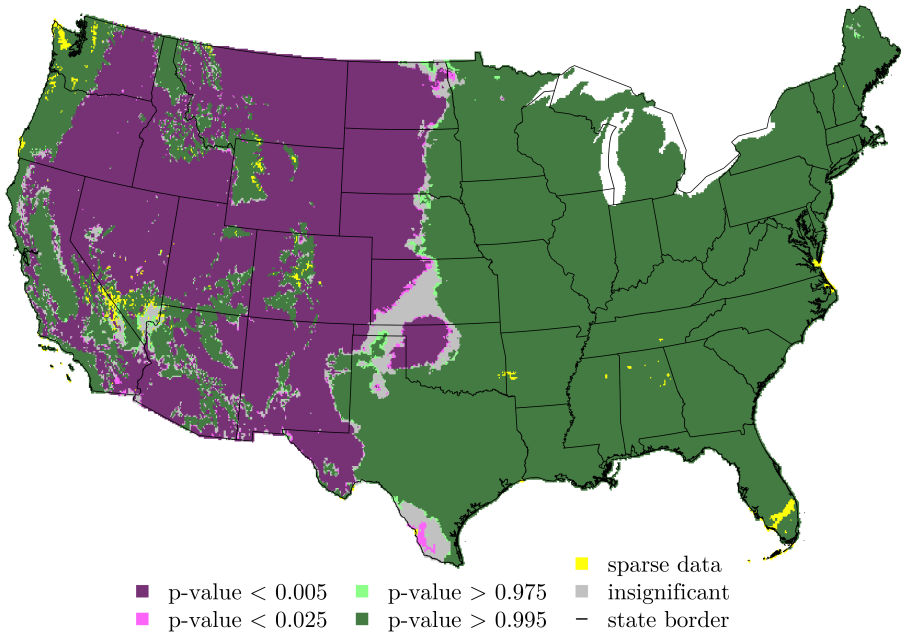


Figure 67: Areas of contiguous United States predicted with log relative risk surface in predictor space of coyotes that were observed but not tested for plague antibodies (controls) and coyotes that were tested the presence of plague antibodies (cases) by the California Department of Public Health (CDPH; 1983–2015) and the U.S. Department of Agriculture (USDA; 2005–2017) using the developed approach. Color pertains to calculated asymptotic tolerances (*219*, *221*) at given local two-tailed significance levels ($\alpha = 0.05$ & $\alpha = 0.01$). Purple-colored areas were more likely tested for the presence of plague antibodies in coyotes and green-colored areas were more likely not tested for the presence of plague antibodies in coyotes with greyer-colored areas statistically indistinguishable. The areas in yellow coloring correspond to habitat with sparse data and statistically dissimilar from sampled habitat.

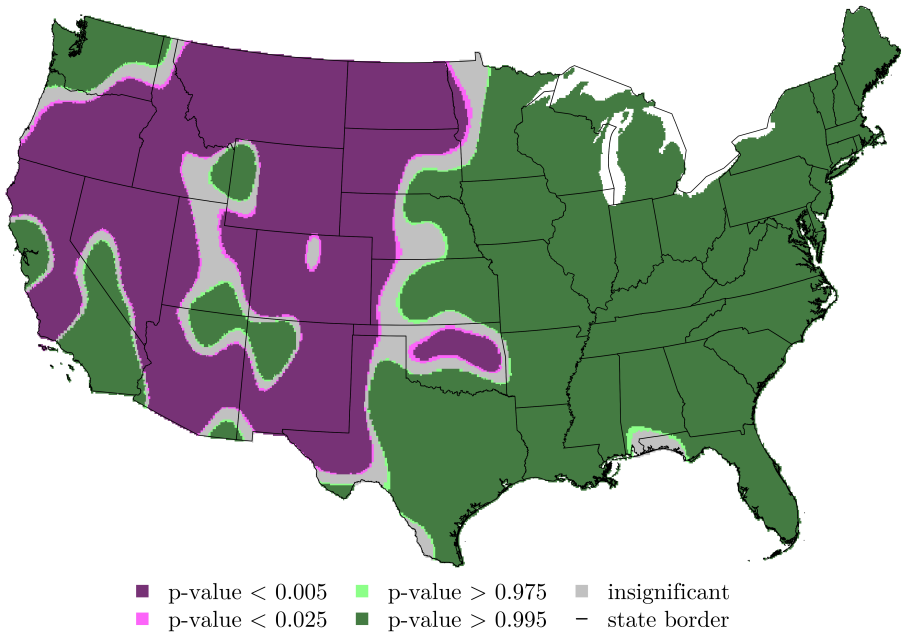


Figure 68: Areas of contiguous United States predicted with log relative risk surface in geographic space (*221*) of coyotes that were observed but not tested for plague antibodies (controls) and coyotes that were tested the presence of plague antibodies (cases) by the California Department of Public Health (CDPH; 1983–2015) and U.S. Department of Agriculture (USDA; 2005–2017). A one-degree bandwidth was chosen. Color pertains to calculated asymptotic tolerances (*219*, *221*) at given local two-tailed significance levels ($\alpha = 0.05$ & $\alpha = 0.01$). Purple-colored areas were more likely tested for the presence of plague antibodies in coyotes and green-colored areas were more likely not tested for the presence of plague antibodies in coyotes with greyer-colored areas statistically indistinguishable.

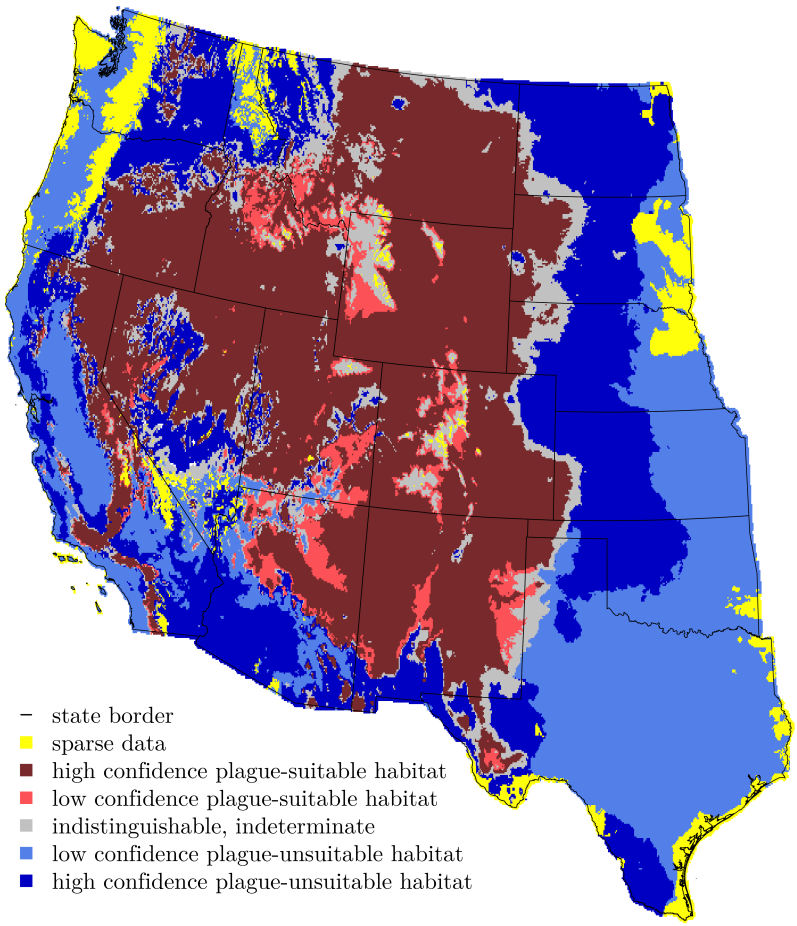


Figure 69: Composite of the disease map (Figure 64) and bias layer (Figure 67) categorized by calculated asymptotic tolerances (*219*, *221*) at given local two-tailed significance levels ($\alpha = 0.05$ & $\alpha = 0.01$) of a predicted log relative risk surface in predictor space. Seronegative coyotes (controls) and seropositive coyotes (cases) tested by the California Department of Public Health (CDPH; 1983–2015) and the U.S. Department of Agriculture (USDA; 2005–2017) were compared using the developed approach. Warmer-colored areas are more likely suitable habitat for plague transmission and cooler-colored areas are more likely unsuitable habitat for plague transmission with greyer-colored areas statistically indistinguishable. I am more confident in the predicted areas that were sufficiently sampled by the USDA and the CDPH, indicated in darker colors, than undersampled areas when compared to historical coyote occurrences from museum repositories. The areas in yellow coloring correspond to habitat that did not have a coyote occurrence either from the USDA, CDPH, or museum repositories.

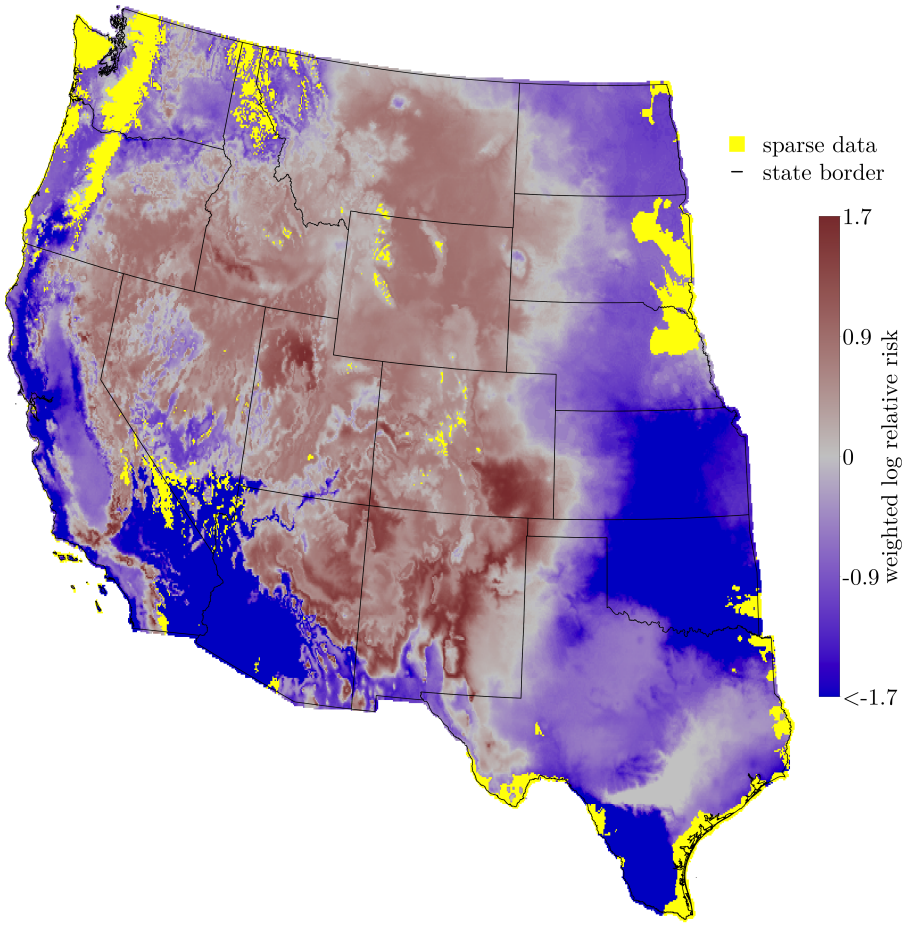


Figure 70: Composite of the disease map (Figure 63) and bias layer (Figure 66) where log relative risk values of the disease map are scaled by the bias layer. In the disease map seronegative coyotes (controls) and seropositive coyotes (cases) tested by the California Department of Public Health (CDPH; 1983–2015) and the U.S. Department of Agriculture (USDA; 2005–2017) were compared using the developed approach. Warmer-colored areas are more likely suitable habitat for plague transmission and cooler-colored areas are more likely unsuitable habitat for plague transmission with greyer-colored areas statistically indistinguishable. In the bias layer coyote observations from museum repositories (controls) were compared to coyotes tested for plague antibodies by the UDSA and the CDPH (cases) using the developed approach. Bias layer values were standardized by the minimum bias layer value. The areas in yellow coloring correspond to habitat that did not have a coyote occurrence either from the USDA, CDPH, or museum repositories.

6 Associating human risk with the spatial distribution of enzootic plague in the western United States

"We may brave human laws, but we cannot resist natural ones."

- Captain Nemo in *Twenty-Thousand Leagues Under the Sea* by Jules Verne (*306*)

6.1 Introduction

Disease mapping can detect irregular spatial clustering of disease cases and predict disease risk in areas not historically monitored for a disease (*307*). There is a separate category of disease mapping called "ecological regression" in which the goal of the analysis is to elucidate associations between disease incidence and risk factors. With the increasing availability of secondary data sources, ecological regression has been used to study the relationship between environmental risk factors and disease risk including, for example, cancer (*308*), cardiovascular disease (*309*), and non-alcoholic liver disease (*310*) as well as infectious diseases such as malaria (*311*), lyme disease (*312*), and hand, foot, and mouth disease (*313*). Ecological regression has also been used to assess veterinary diseases (e.g., canine heartworm; *314*) and wildlife diseases (e.g., coral diseases; *315*). However, a review by Hay and colleagues (*316*) determines only two percent (seven of 355) infectious diseases of clinical importance for humans have been comprehensively charted. Here, I perform an ecological regression on a vector-borne zoonotic disease that originates in rodents and determine the risk of human infection in the western United States.

Plague is a Category A infectious disease caused by the bacterium *Yersinia pestis* and human cases are reportable to the World Health Organization (WHO), U.S. Centers for Disease Control and Prevention (CDC), and state health authorities. The WHO case definition of plague is found in Panel 9. While rare in the United States, plague is a high consequence disease that can cause severe pathology, including death (*33*). Plague is also a security concern as it has the potential to be used as a biological weapon (*101*). Individual risk factors include living in close proximity to or handling infected animals and their flea

vectors (*31*, *38*, *55*).

Individual-level risk factors are heightened in areas where plague is active within a sylvatic cycle (i.e., rodent-to-rodent transmission). Human plague cases have been associated with epizootic (i.e., outbreak within animals) events (*57*), but active transmission also occurs during enzootic (i.e., maintenance within animal populations) periods. Therefore, humans are at risk of plague infection in areas with persistent plague activity (i.e., enzootic areas). Eisen and colleagues (*71*–*73*) found an association between human plague risk and landscape factors such as, for example, elevation and distance from habitat types associated with plague activity. However, the authors restricted most of their analyses (*72*, *73*) to human cases with a known exposure occurring within 2 kilometers of their place of residence resulting in a smaller sample size (e.g., $n = 266$ of 346 human plague cases, 1957–2004, in *72*). The authors also restricted their analyses to the southwestern United States (*71*–*73*) where the majority of human plague cases have occurred since its introduction in 1900 (*33*). However, cases have occurred in 13 western states with no estimation of high risk areas for human plague infection.

Here, I estimate the relative risk of human plague cases across the western United States in relation to environmental factors. I test the utility of predicting the spatial distribution of enzootic plague as a risk factor of human plague. In Chapter 5, I predicted the spatial distribution of enzootic plague in coyotes across the western United States using coyote-based plague surveillance data and climatological variables. Climatological and landscape variables are readily available, but the indices of enzootic plague required additional data and computational resources to predict. Therefore, in this chapter, I test whether the coyote-based plague data helped predict historical human plague risk in the western United States (1950 – 2017). I compare indices of enzootic plague and the same climatological variables used to predict the spatial distribution of enzootic plague on how well they explain the spatial variation in historic human plague cases. I conduct my analysis at the county level due to data availability and capture the exposure location of all human cases instead of restricting the analysis to residence-linked cases. Human plague cases are rare, especially outside of the southwestern United States, so I use a spatial statistical method to borrow information from nearby areas to stabilize risk estimates even in areas without a detected

human plague case.

6.2 Data and methods

6.2.1 Human data collation

I aim to predict county-level counts of human plague cases between 1950 and 2017 in the western United States. I restrict my analysis to cases after 1950 because they were likely acquired in a plague enzootic area during an epizootic event (i.e., disease outbreak animal populations; *57*) and to assure that they were not part of urban-rat-transmission phenomena. The CDC receives aggregated human plague case data from state health departments on an annual basis. Data are reported monthly at the county level to protect the privacy of cases. Since its introduction in 1900, *Y. pestis* has infected 1,045 people in the United States (1900–2017; Table 17 and Figure 71a). Plague arrived in California in 1900 and quickly jumped from commensal rats into native rodent populations where it spread eastward and reached western Kansas by the 1940s (*102*). Human cases before the 1930s were marked by outbreaks in port cities (e.g., San Francisco, CA and Los Angeles, CA), but by 1950 human cases occurred in the country interior only. Since 1950, human plague cases have occurred in 13 western states (521 total cases) an over 80% of these cases have occurred in the Four Corner States (Arizona, Colorado, New Mexico, and Utah; Figure 71b). There have been an average of six cases per year since 2000 (*32*, *33*, *103*). At least one plague case has occurred in 118 of the 1,062 counties of the western United States (Figure 72a) and only one plague case has occurred in 58 (∼49%) of those counties (Figure 72b).

I consider background human population at risk of plague to be the western United States population in 2010 based on the 2010 United States census (*132*) at the county-level. The human population density of the western United States in 2010 appears in Figure 73a. Human plague counts are aggregated across the study period (1950–2017) and I create a variable to account for the change in human population over the study period. The percent change in human population size between 1950 and 2017 is calculated using the population of each western United States county recorded in the 1950 United States census (*317*) and the 2010 United States census (*132*). The percent change in human population

size of the western United States (1950–2010) appears in Figure 73b. Three counties were created during the study period and are assigned their earliest population value. Broomfield County, Colorado was a city in 1961 and is assigned its population according to the 1961 United States census (*187*). La Paz County, Arizona and Cibola County, New Mexico were created in the 1980s and are assigned their population according to the 1990 United States census (*188*). While the United States Census Bureau collects socio-economic and socio-demographic information associated with a sample of households and plague cases have been associated with socio-demographic factors (*318*), I do not include these factors here as they are not the focus of the study.

6.2.2 Environmental data processing

Plague occurrence involves the interaction of a vector (fleas) and mammalian hosts; hence, occurrence typically relates to the intersections of each in the environment. Occurrence has been associated with climate factors (*58*, *59*, *70*, *85*, *89*, *90*) and found at certain elevations (*71–73*). To examine the utility of my prediction of the spatial distribution of enzootic plague to explain the spatial distribution of human plague cases (1950–2017), I first assess how climatological and topographical variables may explain the spatial distribution of human plague cases (1950–2017). I use the same topographical and climatological variables that were used to predict the spatial distribution of enzootic plague (See Chapter 5 for more details). I aggregate these environmental variables to the county-level for my analysis by using the county-level means. I use elevation data from the the National Aeronautics and Space Administration Shuttle Radar Topology Mission (NASA-SRTM) Digital Terrain Elevation Data (*128*). I obtain NASA-SRTM data via the **raster** package (*129*) in the statistical software **R** (*127*). County average elevation appears in Figure 74. I employ climatological data from the Oregon State University Parameter elevation Regression on Independent Slopes Model (PRISM) statistical mapping system. PRISM uses a weighted regression framework that relies on digital elevation models built from a network of ground measurements (*125*). I select the 30-year average normals (1981–2010) at a 2.5 arcminute ($\sim$4 kilometer by $\sim$4 kilometer; $\sim$16 square kilometers) resolution because the temporal range of the data mostly overlap with the animal-based plague surveillance I use in previous

chapters (1983–2017). I obtain PRISM data via the **prism** package (*126*) in the statistical software **R** (*127*) and review each variable in Table 18. The county average PRISM variables appear in Figure 74 and Figure 75.

To predict the spatial distribution of enzootic plague in the western United States, I use the first two principal components (PCs) of a principal component analysis (PCA). Inputs in the PCA are seven range-standardized PRISM variables. I use the Gower metric (*197*) to standardize the variables because they are measured on different scales. I use the **RStoolbox** package (*198*) in the statistical software **R** (*127*) to conduct the PCA. The first two principal components (PC1 and PC2) accounted for over 96% of the variance (Table 19) of the PRISM variables across the contiguous United States. Here, I also use PC1 and PC2 to explain the spatial distribution of human plague cases in the western United States (1950–2017) in the model comparison. The county average principal component coefficients appear in Figure 76.

County-level indices of enzootic plague The utility of the predicted plague niche to explain the spatial distribution of human plague cases (1950–2017) is investigated using various indices. Both raw outputs and outputs accounting for sampling effort bias of state and federal agencies that monitor plague activity are used in the analysis. These agencies included the U.S. Department of Agriculture (USDA; 2005–2017) and the California Department of Public Health (CDPH; 1983–2015). Raw outputs include 1) the log relative risk of enzootic plague and 2) the area predicted significantly plague-suitable. The log relative risk values are predicted comparing seropositive coyotes and seronegative coyotes in predictor space (see Chapter 3 and Chapter 5). Predictor space is comprised of PC1 and PC2. Plague-suitable areas are predicted as locations falling outside of the 95% tolerance interval for null log relative risk of observing a seropositive coyote to that of observing a seronegative coyote (*219*, *221*).

Sampling effort bias is accounted for in both Chapter 5 outputs. The log relative risk of *enzootic plague* is weighted by a standardized log relative risk of *plague testing* by the USDA and CDPH. The log relative risk of plague testing is predicted using the same ecological niche model (see Chapter 3) comparing coyotes tested for *Y. pestis* antibodies to coyotes not

tested for *Y. pestis* antibodies in predictor space. Coyotes not tested for plague antibodies are collated from historical coyote observations in museum repositories (see Chapter 5). Predictor space is the same as above. The log relative risk of plague testing is standardized by its minimum value and null value is centered at one. Sampling effort variations between state and federal agencies that monitor plague activity are accounted for in predicted plague-suitable areas by considering only plague-suitable areas that are also sufficiently tested by state and federal agencies that monitor plague activity (USDA and CDPH). I define areas sufficiently-tested for plague as locations that fall outside of the 95% tolerance interval for null log relative risk of observing a seropositive coyote to that of observing a seronegative coyote (*219*, *221*).

The indices of enzootic plague represent a county-level aggregation of Chapter 5 outputs. First, the log relative risk of enzootic plague, both raw and bias corrected values, are averaged within each county (Figure 78). Any counties with missing values ($n = 10$) are assigned the average value of its neighboring counties. Second, plague-suitable areas, both raw and bias corrected values, are assigned several separate indices. The proportion of plague-suitable area within each county is calculated (Figure 79). In order to account for the various sizes of counties, I also calculate the total area predicted plague-suitable in each county. To test if there is a threshold effect for the amount of area within a county predicted plague-suitable on the spatial distribution of human plague cases (1950–2017), I only consider counties with at least a certain proportion of plague-suitable area. Here, I consider counties where at least 50%, 75%, or 90% of their total area is predicted plague-suitable. The raw plague-suitable area indices appear in Figure 80 and bias corrected plague-suitable area indices appear in Figure 81.

Finally, I compare these derived indices of enzootic plague to an indicator of county-level coyote plague sampling by state and federal agencies that monitor plague activity. Counties with at least one seropositive coyote sampled by the USDA and tested by the USDA, CDC, or CDPH were considered "plague-positive," counties with no seropositive coyotes are considered "plague-negative," and counties without any tested coyotes are considered "not tested." This categorical variable appears in Figure 77. Serological results are often presented at an aggregated spatial scale to protect land-owner privacy (see *107*)). These

data are absent of any environmental interpolation (*19*) unlike the derived plague indices that use climate variables to predict where plague may be cycling in the environment.

6.2.3 Statistical methods

I employ a county-level spatial model comparison approach to evaluate the utility of enzootic plague to explain the spatial distribution of human cases in western United States counties (1950–2017). The modeling framework is atemporal because cases are aggregated for the 67-year study period. I compare the performance of models using univariate plague indices to univariate models of environmental variables. Descriptions of each model appear in Table 20 and the model comparison approach is described below.

The western United States is my region of interest S and is split into n contiguous small-areas (counties). I aim to detect which counties exhibit elevated disease risks as well as case counts based on their 2010 population. Let $y_i = (y_1, ..., y_n)$ denote the number of observed cases in county i and $E_i = (E_1, ..., E_n)$ contain offsets, defined here as the population of each county i based on the 2010 census (*132*). Because human plague cases are rare I use a Bayesian hierarchical model-based approach, which estimates disease risk in each county using county-level covariates and a set of random effects. The random effects borrow information from neighboring counties, resulting in spatial smoothing of estimated rates and stabilization of their variances (due to the very small observed numbers of human cases).

I use a hierarchical model defined as:

$$y_i|\lambda_i \sim Poisson(\lambda_i)$$

$$\lambda_i = E_i\rho_i$$

$$\eta_i = log(\rho_i) = x_i^T\beta_i + \psi_i.$$

Where η_i is the local relative risk of disease in county i and a set of p county-level fixed-effects $x_i^T = (1, x_{i1}, ..., x_{ip})$ with regression coefficients $\beta_i = (\beta_1, ..., \beta_p)$. County-level fixed-effects depend on the model specification found in Table 20. To model any spatial autocorrelation

that remain in the data beyond the covariate effects I have a set of random effects $\psi_i \in (\psi_1, ..., \psi_n)$ where

$$\psi_i = \gamma_i + \phi_i.$$

A conditional autoregressive (CAR) model assigns spatial structure to the random effects. Here, I use a convolution or Besag-York-Mollié (BYM) CAR model outlined in Besag and colleagues 1991 (*319*) that contains a convolution of non-spatially autocorrelated random effects $\gamma_i = (\gamma_1, ..., \gamma_n)$ and spatially autocorrelated random effects $\phi_i = (\phi_1, ..., \phi_n)$:

$$\gamma_i \overset{ind}{\sim} N(0, \sigma_\gamma), 1, \ldots, n^2$$

$$\phi_i | \phi_{-i}, \boldsymbol{W}, \sigma_\phi^2 \sim N\left(\frac{\sum_{i=1}^{n} \omega_{ni}\phi_i}{\sum_{i=1}^{n} \omega_{ni}}, \frac{\sigma_\phi^2}{\sum_{i=1}^{n} \omega_{ni}}\right)$$

$$log(\tau_\phi), log(\tau_\gamma) = log(1/\sigma_\phi^2), log(1/\sigma_\gamma^2) \sim \text{logGamma}(a, b).$$

A non-negative symmetric $n \times n$ neighborhood matrix $\boldsymbol{W}$ containing positive elements ω_{ni} is specified based on geographical contiguity between counties where $\omega_{ni} = 1$ if counties share a common border, and is zero otherwise ($\omega_{ii} = 0$ by convention), forcing geographically adjacent counties to be autocorrelated and non-adjacent counties to be conditionally independent given the remaining random effects. Hyperpriors (a, b) for the variance parameters $(\sigma_\phi^2, \sigma_\gamma^2)$ are set at vague default levels ($a = 1$ and $b = 0.0005$). Future models can investigate other CAR model specifications, including other globally and locally smooth CAR models.

The vast majority of counties in the western United States did not have an observed human plague case (Figure 72a). Therefore, I control for overdispersion by considering two zero-inflated Poisson (ZIP) modeling approaches (*320–322*). One ZIP model assumes all zero counts result from a county having no current plague circulation (i.e., a *structural* zero where zero is the only allowed observation); the other model assumes that a fraction of the observed zeros are structural (no plague circulation), and the other zeros are *observed* zeros (i.e., human plague could be observed but was not).

The first zero-inflated Poisson modeling approach (ZIP0) assumes the zero human plague case counts in the western United States (1950–2017) are all structural zeros (i.e. zero is the only observable value). Given n counties, the probability function for $\boldsymbol{y} = (y_1, ..., y_n)$ is:

$$p(y_i|\lambda_i, \pi_0) = \pi_0 I(y_i = 0) + (1 - \pi_0) I(y_i > 0) \frac{\exp(\lambda_i)\lambda_i^{y_i}}{y_i!}$$

where $I(y_i = 0)$ is the indicator variable and the zero-probability hyperparameter π_0 is parameterized with a default Normal(-1,0.2) prior distribution (*323–328*) and is transformed with the inverse logit transformation $\frac{\exp(\pi_0)}{1+\exp(\pi_0)}$ (see *329*). The probability of observing a zero in the ith county is $\pi_0 + (1 - \pi_0)\exp(-\lambda_i)$ with a mean and variance is:

$$E(y_i) = (1 - \pi_0\lambda_i)$$

$$Var(y_i) = (1 - \pi_0\lambda_i) + \frac{\pi_0}{1 + \pi_0}((1 - \pi_0)\lambda_i).$$

The second zero-inflated Poisson modeling approach (ZIP1) assumes the zero human plague case counts in the western United States (1950–2017) are a combination of structural zeros and sampling zeros (i.e., zero is observed but may be another value). Given n counties, the probability function for $\boldsymbol{y} = (y_1, ..., y_n)$ is:

$$p(y_i|\lambda_i, \pi_0) = \pi_0 I(y_i = 0) + (1 - \pi_0) \frac{\exp(\lambda_i)\lambda_i^{y_i}}{y_i!}.$$

Poisson, ZIP0, and ZIP1 models are run for each model specification (Table 20) and compared using the Deviance Information Criteria (DIC; *330*) and the Watanabe-Akaike information criterion (WAIC; *331*, *332*). The best-performing model has the lowest DIC and WAIC score. I first assess if including the percent change in population (1950–2010) improves upon a null (Poisson, ZIP0, and ZIP1) model with only an intercept. The best-performing Poisson and zero-inflated Poisson models are the subsequent focus of the analysis. I compare the proportion of variance explained by the spatially-structured component of each model (*329*):

$$frac_{\phi} = \frac{s_{\phi}^2}{s_{\phi}^2 + \sigma_{\gamma}^2}$$

where σ_{γ}^2 is the variance of the spatially-unstructured effect and s_{ϕ}^2 is the estimate of the posterior marginal variance of the spatially-structured effect:

$$s_{\phi}^2 = \frac{\Sigma_{i=1}^{n}(\phi_i - \bar{\phi})^2}{n-1}$$

where $\bar{\phi}$ is the average of the spatially-structured component ($\boldsymbol{\phi}$). s_{ϕ}^2 is estimated empirically using a simulation-based approach where I calculate the variance from 100,000 samples of the marginal posterior distribution of γ_i. I also examine the spatially-structured residuals (ϕ_i) and spatially-unstructured residuals (γ_i) of the best-performing model as well as the residual relative risk not explained by the fixed-effects ($\exp(\psi_i)$) and the excess residual relative risk ($p((\psi_i) > 0|\boldsymbol{y})$).

All models are performed in the **R-INLA** package (*323–328*) in the statistical software **R** (*127*). The integrated nested Laplace approximation (INLA) method is an alternative to the commonly used Markov Chain Monte Carlo (MCMC) simulations developed by Rue and colleagues (*323*) that approximates Bayesian inference.

6.3 Results

Of the 81 candidate models (Table 20), the best-performing model was a Poisson model comprised of the percent change in population (1950–2010), the two principal components (PC1 and PC2), and the interaction term between PC1 and PC2 (DIC: 714.17; WAIC: 715.67). However, there was evidence of overdispersion in the human plague data (1950–2017; Figure 72) and the best-performing zero-inflated Poisson model was comprised of the log relative risk of enzootic plague that did not account for sampling effort of state and federal agencies that monitor plague activity (DIC: 758.45; WAIC: 764.89). Model performance goodness of fit results appear in Table 21. The subsequent analysis focused on the comparison and interpretation of these models.

All risk factors were statistically significant in the best fitting Poisson model, including the interaction term between PC1 and PC2 (Table 22). For every one percentage point

increase in human population (1950–2010) within a county, relative risk of a human plague case (1950–2017) decreased by an average of 0.13% (95% credibility interval: 0.19%–0.05% decrease). If PC2 is zero, for every one hundredth of one unit coefficient value increase in PC1, the relative risk of a human plague case (1950–2017) decreased by an average of 4.7% (95% credibility interval: 6.5%–2.8% decrease). If PC1 is zero, for every one hundredth of one unit coefficient value increase in PC2, the relative risk of a human plague case (1950–2017) decreased by an average of 13.4% (95% credibility interval: 16.9%–10.0% decrease). The interaction term is positive and large (23.6% with a 95% credibility interval: 12.8%–36.4% increase). Mean and standard deviation of predicted human plague case counts (1950–2017) from the best-performing Poisson model appear in Figure 82. Predicted mean case counts were high in the southwestern United States, primarily in Arizona and New Mexico. Counties predicted with at least one plague case occurred in 15 of the 17 western states, including in highly-populated areas around cities in the Great Plains such as Dallas, Texas and Oklahoma City, Oklahoma. The model did not predict more than one human plague case (1950–2017) in any county in southern Oregon, however standard deviation of the predicted human plague cases (1950–2017) was higher in this region. Higher standard deviations in predicted case counts were located in counties of the southwestern United States and California.

All risk factors were statistically significant in the best-fitting ZIP model (Table 22). For every one percent point increase in human population (1950–2010) within a county, relative risk of a human plague case (1950–2017) decreased by an average of 0.09% (95% credibility interval: 0.17%–0.03% decrease). For every one tenth of one log relative risk value increase in enzootic plague, the relative risk of a human plague case (1950–2017) increased by an average of 20% (95% credibility interval: 15.9%–23.8% increase). Mean and standard deviation of predicted human plague case counts (1950–2017) from the best-performing ZIP model appear in Figure 83. Predicted mean case counts were high in the southwestern United States, primarily in Arizona and New Mexico but also in populated areas of California, Colorado, and western Texas. Counties predicted with at least one plague case occurred in all western United States states except North Dakota, including in highly-populated areas around cities in the Great Plains such as San Antonio, Texas and

Omaha, Nebraska. The model did not predict a human plague case (1950–2017) in counties surrounding Las Vegas, Nevada. Clark County, Nevada had low standard deviation in predicted human plague case counts (1950–2017). Higher standard deviation in predicted case counts were located in the southwestern United States and highest in El Paso County, Texas.

Both best-fitting models predicted counties with high human plague case counts (1950–2017) better than counties with low or no human plague case (Figure 84), as expected. On average, both best-fitting models underestimated the number of counties with zero recorded human plague cases (1950–2017) and overestimated the number of counties with one to three human plague cases (1950–2017). The best-fitting ZIP model underestimates predicted case counts more than the Poisson model (Figure 85), as expected, especially with decreasing human plague case counts (1950–2017). The proportion of variance explained by the structured spatial component in the best-fitting Poisson model is about 100% and about 84% in the best-fitting ZIP model (Table 22). For reference, the proportion of variance explained by the structured spatial component in the null Poisson model and null ZIP model (percent change in population only; see Table 20) are both almost 100%. Also for reference, the ZIP model comprised of PC1 and PC2 with their interaction had a lower proportion of variance explained by the structured spatial component (55%) and the Poisson model comprised of the log relative risk of enzootic plague (not accounting for sampling effort bias) was almost 100%. All risk factors in reference models were statistically significant. ZIP models had a lower mean precision of the spatially-uncorrelated random effect and their 95% credibility interval than Poisson models by orders of magnitude, as expected.

I examined the best-fitting ZIP model. The spatially-correlated and spatially-uncorrelated residuals of the best-fitting ZIP model appear in Figure 86. The spatial distribution of the residuals was similar with high values in counties at the California-Oregon border and northern counties of Arizona and New Mexico. Low residual values (both spatially-correlated and spatially-uncorrelated) were located in the northern counties of the western United States as well as Salt Lake City, Utah and west Texas. Five counties had high spatially-uncorrelated residual values: Humboldt County, California; Coos County and Lake County in Oregon; and McKinley County and Rio Arriba County in New Mexico. Residual relative risk was

high in counties of northern New Mexico, the California-Oregon border, and Mariposa County, California (Figure 87a). Excess residual relative risk was high in the southern regions of the western United States and Oregon (Figure 88). The mean zero-probability parameter of the best-fitting ZIP model was 0.20 (95% credibility interval: 0.197%–0.206% decrease; Table 22) and varied across the western United States (Figure 88). Counties with a low mean zero-probability were predicted in the southwestern United States except in Clark County, Nevada and nearby counties. Counties with a high mean zero-probability were predicted in western Washington State and in the Great Plains, except highly-populated areas such as Houston, Texas.

6.4 Discussion

The spatial variation and distribution of historical human plague cases in the United States (1950–2017) are explained by environmental risk factors, including the spatial distribution of enzootic plague, and random effects, especially spatially-correlated random effects. The best-fitting hierarchical model is a Poisson model comprised of the percent change in population (1950–2010), the two principal components (PC1 and PC2), and the interaction term between PC1 and PC2. However, this model does not account for overdispersion in historical human plague cases in the United States (1950–2017) and its statistically significant risk factors are difficult to interpret for human plague risk inference, especially the interaction term. The vast majority of the variation in human plague risk is explained by the spatially-correlated random effect; thus, the fixed effects, while statistically significant, do not explain much of the spatial variation and distribution of historical human plague cases in the United States (1950–2017). The exceedingly high precision for the spatially-unstructured component of the best-fitting Poisson model further suggests the Poisson model is not sufficient.

A zero-inflated Poisson model is a more appropriate model for historical human plague cases (1950–2017) because it allows for the existence of some structural zeros from counties with zero underlying risk. The best ZIP model is comprised of the percent change in population (1950–2010) and the average log relative risk of enzootic plague (not accounting for sampling effort by state and federal agencies that monitor plague activity). The precision for the spatially-unstructured component of the best-fitting ZIP model is stable suggesting

the ZIP model is more appropriate than a Poisson model. The average log relative risk of enzootic plague has a statistically significant positive association with human plague risk. Counties with an increase in one tenth of one log relative risk value see an average 20% (95% credibility interval: 15.9%–23.8%) increase in human plague risk. Other county-level indices of enzootic plague (e.g., proportion of a county with plague-suitable habitat or total area of a county considered plague-suitable) did not explain the spatial variation and distribution of historical human plague cases in the United States (1950–2017) as well as average log relative risk of enzootic plague. The ecological niche of enzootic plague is more likely located in counties with a higher log relative risk of enzootic plague, on average, than counties with a lower log relative risk of enzootic plague. Counties with a higher average log relative risk of enzootic plague can be considered highly-persistent plague counties where climatological conditions are prime for sylvatic plague transmission.

The log relative risk of enzootic plague captures some of the interaction between PC1 and PC2 and may explain why these models perform similarly. I estimated the log relative risk of enzootic plague in Chapter 5 using PC1 and PC2. The second best-performing Poisson model is comprised of the percent change in population (1950–2010) and the average log relative risk of enzootic plague (not accounting for sampling effort by state and federal agencies that monitor plague activity). The second best-performing ZIP model is comprised of the percent change in population (1950–2010), the two principal components (PC1 and PC2), and the interaction term between PC1 and PC2. Predicting the spatial distribution of enzootic plague in the United States using coyote-based plague surveillance has utility for predicting human plague risk both for explaining some of the variation and being more interpretable than the individual principal components and their interaction. This relationship breaks down somewhat when spatially aggregated to the county-level, so the county average log relative risk of enzootic plague cannot account for all the interaction between PC1 and PC2. A future study can develop a framework to comprehensively model the three levels of the ecological regression (human plague risk layer estimated using an animal plague risk layer that is predicted using animal-based plague data and environmental variables) in a multi-stage model instead of step by step as was conducted in the present analysis. A hierarchical model can account for the variability in each level and the relationship between

the three levels can be examined.

Spatial scale is a limitation of the analysis. Human plague cases were only available at a county-level for privacy considerations. A county may have areas of high log relative risk of enzootic plague, but these areas can be hidden when averaged across an entire county. For example, the Transverse and Peninsular Ranges in southern California are suitable for plague transmission (see Chapters 3-5), but the majority of area within southern California counties are plague-unsuitable. Therefore, at a county-level average, the counties in southern California have low log relative risk of enzootic plague (Figure 78a) even though historical human plague cases (1950–2017) have occurred in southern California. I observed the same data aggregation effect for other indices of enzootic plague (e.g., the proportion of a county predicted plague-suitable) where smaller counties are more likely to have a higher plague values. Aggregation at the county-level more accurately captures the location of exposure for every case if the location of exposure is known, which is not always the case (see *133*). For example, Coos County in southern Oregon has high model residuals, which suggests the average log relative risk of enzootic plague fixed effect does not explain the human plague risk in this county. Two possible explanations are that either cases in this county resulted from epizootic events occurring outside of the predicted enzootic areas or the true locations of exposure for these human plague cases were not in Coos County, Oregon. McKinley County and Rio Arriba County in New Mexico have moderate average log relative risk of enzootic plague values but are nearby counties with high log relative risk of enzootic plague values (i.e., highly-persistent plague areas). Epizootic conditions may have spilled over from nearby counties into McKinley County and Rio Arriba County, New Mexico that are not captured in the analysis, resulting in high residuals for these counties.

Human plague is rare in the United States, especially outside of the southwestern United States (*33*), and the best-fitting ZIP model underestimates counties with no historical human plague cases (1950–2017). Human plague cases are rare because of the low risk of human infection but also the low human population levels outside major cities in the western United States. Plague is actively cycling in many regions of the western United States, but low human interaction within the natural transmission cycle results in few human cases. While fortunate for public health, the rarity of human plague cases presents a challenge for

disease surveillance. Eisen and colleagues (*72*, *73*) focused on human plague risk in the southwestern United States where the majority of historical human cases have occurred in the United States. I extend human plague risk estimation throughout the western United States where human plague cases have also occurred, but in doing so, I incorporate counties with zero human plague cases. While I account for such structural zeros by using a ZIP model, a future study can restrict the analysis to counties in the southwestern United States. A more parsimonious Poisson model may be appropriate for this smaller spatial extent because there will be fewer counties with zero cases in the analysis. A future study can assess other models that may be more appropriate to account for overdispersion of historical human plague cases (1950–2017) such as, for example, a zero-inflated negative binomial distribution.

The best-fitting models predict human plague cases in highly-populated areas of the Great Plains such as, for example, San Antonio, Texas and Omaha, Nebraska. The large baseline population at risk in these urban areas increase the chance of a human plague case occurring in these counties. However, the individual relative risk of plague infection in these counties is extremely low. Ben-Ari and colleagues (*59*) found no correlation between human plague outbreak frequency (1950–2005) and population density at the county level in the United States. Population growth (1950–2010) was weakly significantly negatively associated with human plague risk. Fast-growing urban populations likely drive this relationship such as, for example, Las Vegas in Clark County, Nevada that grew over 8,000% within the study period. This is consistent with other findings from Schotthoefer and colleagues (*318*) who found human plague cases do not occur in urban areas of New Mexico (1976–2007) and human plague cases in rural or lightly suburbanized areas were associated with human population growth. A future study could limit human plague risk estimation to rural and suburban areas across the western United States to examine the effect of changing socio-demographics. Humans at risk of plague infection also include visitors and non-county residents. For instance, Kwit and colleagues (*32*) linked two human plague cases from Georgia to exposure in Yosemite National Park in California. Popular destinations within plague enzootic areas are high-risk locations of exposure. A future study can more accurately estimate human plague risk by incorporating tourism in the background

population at-risk.

Human plague risk maps should ideally incorporate environmental and social factors that influence human plague risk. Plague in humans and wildlife is associated with environmental factors, such as climate patterns and landscape features (*58*, *59*, *70–73*, *85*, *89*, *90*). Human plague is associated with socio-economic factors (*104*, *318*). These factors interact and a human plague case likely occurs under the perfect combination of conditions in time and location. Therefore, a spatio-temporal human plague analysis is desirable for public health action, but is intractable due to the rarity of human plague cases in the United States. Here, I did not conduct a spatio-temporal analysis because the sample size of human cases was even smaller within shorter time periods and the problem of overdispersion was worse within smaller study periods. Some studies have conducted a spatial analysis of temporally-aggregated human plague cases to predict high human plague risk areas (*70*, *72*, *73*). Ben-Ari and colleagues (*59*) compared human plague cases ($n = 105$ counties with at least one case) in three regions of the western United States (1950–2005) but did not predict human plague risk. Schotthoefer and colleagues (*318*) compared three periods of human plague cases ($n = 123$ of 162) in New Mexico (1976–2007) but did not conduct a spatio-temporal ecological regression. To assess the interaction of environmental and social risk factors, a future spatio-temporal ecological regression could focus on New Mexico where the majority of historical human cases in the United States (1950–2017) has occurred.

The results have public health relevance. I identified an environmental risk factor for human plague infection that is measurable at the county level. Human plague cases have been associated with epizootic conditions (*57*) and inter-annual seasonal patterns (*59*), but I demonstrated that historical human plague cases (1950–2017) are also associated with enzootic plague. The bacterium *Y. pestis* is transmitting within rodent populations at low levels and, while low-risk, a human can contract plague if they interact with this sylvatic cycle. A human is more likely to interact with sylvatically-cycling plague in plague enzootic areas. Therefore, identifying areas where plague is persistent can benefit public health action by informing healthcare providers or directing preventive programs. Figure 88 shows counties within the western United States that were predicted to have at least one human plague case (1950–2017). Public health action can be targeted to counties with a high

probability of observing at least one case (low probability of zero cases). The results also provide an example of a spatial model that accounts for overdispersion and structural zero case counts, which may be useful for other disease systems. In particular, surveillance systems for rare events such as, for example, zoonoses or antibiotic-resistant pathogen strains can benefit from the methodology utilized in this study.

6.5 References

19. A. T. Peterson, *Mapping disease transmission risk: enriching models using biogeography and ecology* (Johns Hopkins University Press, Baltimore, Maryland, 2014), ISBN: 1421414737.

31. K. L. Gage, M. Y. Kosoy, *Annual Review of Entomology* **50**, 505–528, ISSN: 0066-4170, DOI 10.1146/annurev.ento.50.071803.130337 (2005).

32. N. Kwit *et al.*, *Morbidity and Mortality Weekly Report* **64**, 918–919, ISSN: 0149-2195, 1545-861X (2015).

33. K. J. Kugeler *et al.*, *Emerging Infectious Diseases* **21**, 16, ISSN: 1080-6059, 1080-6040, DOI 10.3201/eid2101.140564 (2015).

38. H. L. Gould *et al.*, *Zoonoses and Public Health* **55**, 448–454, ISSN: 1863-1959, DOI 10.1111/j.1863-2378.2008.01132.x (2008).

55. C. F. von Reyn *et al.*, *American Journal of Tropical Medicine and Hygiene* **25**, 626–629, ISSN: 0002-9637, 1476-1645, DOI 10.4269/ajtmh.1976.25.626 (1976).

57. H. E. Brown *et al.*, *American Journal of Tropical Medicine and Hygiene* **82**, 95–102, ISSN: 0002-9637, 1476-1645, DOI 10.4269/ajtmh.2010.09-0247 (2010).

58. T. Ben-Ari *et al.*, *PLoS Pathogens* **7**, e1002160, ISSN: 1553-7374, 1553-7366, DOI 10.1371/journal.ppat.1002160 (2011).

59. T. Ben-Ari *et al.*, *American Journal of Tropical Medicine and Hygiene* **83**, 624–632, ISSN: 0002-9637, 1476-1645, DOI 10.4269/ajtmh.2010.09-0775 (2010).

70. Y. Nakazawa *et al.*, *Vector-Borne and Zoonotic Diseases* **7**, 529–540, ISSN: 1557-7759, 1530-3667, DOI 10.1089/vbz.2007.0125 (2007).

71. R. J. Eisen *et al.*, *American Journal of Tropical Medicine and Hygiene* **77**, 999–1004, ISSN: 0002-9637, 1476-1645, DOI 10.4269/ajtmh.2007.77.999 (2007).

72. R. J. Eisen *et al.*, *Journal of Medical Entomology* **44**, 530–537, ISSN: 0022-2585, DOI 10.1603/0022-2585(2007)44 (2007).

73. R. J. Eisen *et al.*, *American Journal of Tropical Medicine and Hygiene* **77**, 121–125, ISSN: 0002-9637, 1476-1645, DOI 10.4269/ajtmh.2007.77.121 (2007).

85. S. P. Maher *et al.*, *American Journal of Tropical Medicine and Hygiene* **83**, 736–742, ISSN: 0002-9637, 1476-1645, DOI 10.4269/ajtmh.2010.10-0042 (2010).

86. J. C. Z. Adjemian *et al.*, *Journal of Medical Entomology* **43**, 93–103, ISSN: 0022-2585, DOI 10.1093/jmedent/43.1.93 (2006).

89. A. C. Holt *et al.*, *International Journal of Health Geographics* **8**, 38, ISSN: 1476-072X, DOI 10.1186/1476-072X-8-38 (2009).

90. M. Walsh, M. A. Haseeb, *PeerJ* **3**, e1493, ISSN: 2167-8359, DOI 10.7717/peerj.1493 (2015).

101. T. V. Inglesby *et al.*, *Journal of the American Medical Association* **283**, 2281–2290, ISSN: 0098-7484, DOI 10.1001/jama.283.17.2281 (2000).

102. J. Z. Adjemian *et al.*, *American Journal of Tropical Medicine and Hygiene* **76**, 365–375, ISSN: 0002-9637, 1476-1645, DOI 10.4269/ajtmh.2007.76.365 (2007).

103. R. B. Craven *et al.*, *Journal of Medical Entomology* **30**, 758–761, ISSN: 0022-2585, DOI 10.1093/jmedent/30.4.758 (1993).

104. U.S. Centers for Disease Control and Prevention, *Morbidity and Mortality Weekly Report* **33**, 145, ISSN: 0149-2195, 1545-861X (1984).

107. S. N. Bevins *et al.*, *Integrative Zoology* **7**, 99–109, ISSN: 1749-4877, DOI 10.1111/j.1749-4877.2011.00277.x (2012).

125. C. Daly *et al.*, *International Journal of Climatology* **28**, 2031–2064, ISSN: 1097-0088, DOI 10.1002/joc.1688 (2008).
126. E. M. Hart, K. Bell, *prism: download data from the Oregon PRISM project*, R package version 0.0.6, DOI 10.5281/zenodo.33663, (http://github.com/ropensci/prism).
127. R Core Team, *R: a language and environment for statistical computing*, R Foundation for Statistical Computing (Vienna, Austria, 2018), (https://www.R-project.org/).
128. J. J. van Zyl, *Acta Astronautica* **48**, 559–565, ISSN: 0094-5765, DOI 10.1016/S0094-5765(01)00020-0 (2001).
129. R. J. Hijmans, *raster: geographic data analysis and modeling*, R package version 2.6-7, (https://CRAN.R-project.org/package=raster).
132. U.S. Census Bureau, *2010 census*, U.S. Department of Commerce, 2011.
133. C. Myron, *Elmore County child recovering after plague infection*, Central District Health Department (2018; http://www.cdhd.idaho.gov/pdfs/News/2018/06-12-18-human-plague-case.pdf).
186. S. Manson *et al.*, *IPUMS National Historical Geographic Information System: Version 13.0 nhgis0001*, Minneapolis, Minnesota, 2018, DOI 10.18128/D050.V13.0.
187. U.S. Census Bureau, *1960 census*, U.S. Department of Commerce, 1961.
188. U.S. Census Bureau, *1990 census*, U.S. Department of Commerce, 1991.
197. J. C. Gower, *Biometrics*, 857–871, ISSN: 1541-0420, 0006-341X, DOI 10.2307/2528823 (1971).
198. B. Leutner *et al.*, *RStoolbox: tools for remote sensing data analysis*, R package version 0.2.3, (https://CRAN.R-project.org/package=RStoolbox).
219. M. L. Hazelton, T. M. Davies, *Biometrical Journal* **51**, 98–109, ISSN: 1521-4036, 0323-3847, DOI 10.1002/bimj.200810495 (2009).
221. T. M. Davies, M. L. Hazelton, *Statistics in Medicine* **29**, 2423–2437, ISSN: 1097-0258, DOI 10.1002/sim.3995. (2010).
306. J. Verne, *Twenty thousand leagues under the sea (slightly abridged)* (Project Gutenberg, Urbana, Illinois, 1994), (2019; http://www.gutenberg.org/ebooks/164).
307. L. A. Waller, C. A. Gotway, *Applied spatial statistics for public health data* (John Wiley & Sons, Hoboken, New Jersey, 2004), vol. 368, ISBN: 0471387711.
308. R. Carroll *et al.*, *Annals of Epidemiology* **27**, 42–51, ISSN: 1047-2797, DOI 10.1016/j.annepidem.2016.08.014 (2017).
309. I. H. d. Silveira, W. L. Junger, *Revista de Saude Publica* **52**, 49, ISSN: Revista de Saude Publica, DOI 10.11606/s1518-8787.2018052000290 (2018).
310. F. Zhang *et al.*, *Environmental Health* **14**, 41, ISSN: 1476-069X, DOI 10.1186/s12940-015-0026-7 (2015).
311. M. Santos-Vega *et al.*, *PLoS Neglected Tropical Diseases* **10**, e0005155, ISSN: 1935-2735, 1935-2727, DOI 10.1371/journal.pntd.0005155 (2016).
312. L. A. Waller *et al.*, *Environmental and Ecological Statistics* **14**, 83, ISSN: 1352-8505, 1573-3009, DOI 10.1007/s10651-006-0002-z (2007).
313. Y. Liu *et al.*, *BMC Infectious Diseases* **15**, 146, ISSN: 1471-2334, 1471-2334, DOI 10.1186/s12879-015-0901-4 (2015).
314. D. D. Bowman *et al.*, *Parasites & Vectors* **9**, 540, ISSN: 1756-3305, DOI 10.1186/s13071-016-1804-y (2016).
315. R. van Woesik, C. J. Randall, *Ecosphere* **8**, ISSN: 2150-8925, DOI 10.1002/ecs2.1814 (2017).
316. S. I. Hay *et al.*, *Philosophical Transactions of the Royal Society B: Biological Sciences* **368**, 20120250, ISSN: 0962-8452, 1471-2954, DOI 10.1098/rstb.2012.0250 (2013).

317. U.S. Census Bureau, *1950 census*, U.S. Department of Commerce, 1951.
318. A. M. Schotthoefer *et al.*, *Emerging Infectious Diseases* **18**, 1151–1154, ISSN: 1080-6059, 1080-6040, DOI 10.3201/eid1807.120121 (2012).
319. J. Besag *et al.*, *Annals of the Institute of Statistical Mathematics* **43**, 1–20, ISSN: 0020-3157, 1572-9052, DOI 10.1007/BF00116466 (1991).
320. D. Lambert, *Technometrics* **34**, 1–14, ISSN: 0040-1706, DOI 10.2307/1269547 (1992).
321. D. K. Agarwal *et al.*, *Environmental and Ecological Statistics* **9**, 341–355, ISSN: 1352-8505, DOI 10.1023/A:1020910605990 (2002).
322. S. Gschlößl, C. Czado, *Statistical Papers* **49**, 531, ISSN: 0932-5026, 1613-9798, DOI 10.1007/s00362-006-0031-6 (2008).
323. H. Rue *et al.*, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **71**, 319–392, ISSN: 1369-7412, DOI 10.1111/j.1467-9868.2008.00700.x (2009).
324. T. G. Martins *et al.*, *Computational Statistics & Data Analysis* **67**, 68–83, ISSN: 1872-7352, DOI 10.1016/j.csda.2013.04.014 (2013).
325. F. Lindgren *et al.*, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **73**, 423–498, ISSN: 1369-7412, DOI 10.1111/j.1467-9868.2011.00777.x (2011).
326. F. Lindgren, H. Rue, *Journal of Statistical Software* **63**, 1–25, ISSN: 1548-7660, DOI 10.18637/jss.v063.i19 (2015).
327. H. Rue *et al.*, *Annual Review of Statistics and its Applications* **4**, 395–421, ISSN: 2326-831X, DOI 10.1146/annurev-statistics-060116-054045 (2017).
328. H. Bakka *et al.*, *Wiley Interdisciplinary Reviews: Computational Statistics* **10**, e1443, ISSN: 1939-5108, 1939-0068, DOI 10.1002/wics.1443 (2018).
329. M. Blangiardo, M. Cameletti, *Spatial and spatio-temporal Bayesian models with R-INLA* (John Wiley & Sons, Hoboken, New York, 2015), ISBN: 1118326558.
330. D. J. Spiegelhalter *et al.*, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**, 583–639, ISSN: 1369-7412, DOI 10.1111/1467-9868.00353 (2002).
331. S. Watanabe, *Journal of Machine Learning Research* **14**, 867–897, ISSN: 1533-7928, 1532-4435 (2013).
332. S. Watanabe, *Journal of Machine Learning Research* **11**, 3571–3594, ISSN: 1533-7928, 1532-4435 (2010).

6.6 Appendices

6.6.1 Appendix A: Panels

Panel 9: World Health Organization case definition for plague surveillance

Disease characterized by rapid onset of fever, chills, headache, severe malaise, prostration, with

- Bubonic form: extreme painful swelling of lymph nodes (buboes)
- Pneumonic form: cough with blood-stained sputum, chest pain, difficult breathing

Note: Both forms can progress to a septicemic form with toxemia. Sepsis without evident buboes rarely occurs.

Laboratory criteria for diagnosis

- Isolation of *Yersinia pestis* in cultures from buboes, blood, cerebrospinal fluid or sputum or
- Passive hemagglutination (PHA) test, demonstrating an at least fourfold change in antibody titer, specific for F1 antigen of *Y. pestis*, as determined by the hemagglutination inhibition test (HI) in paired sera.

Case classification

- Suspected: A case compatible with the clinical description. May or may not be supported by laboratory finding of Gram stain negative bipolar coccobacilli in clinical material (bubo aspirate, sputum, tissue, blood).
- Probable: A suspected case with Positive direct fluorescent antibody (FA) test for *Y. pestis* in clinical specimen; or passive hemagglutination test, with antibody titer of at least 1:10, specific for the F1 antigen of *Y. pestis* as determined by the hemagglutination inhibition test (HI); or epidemiological link with a confirmed case.
- Confirmed: A suspected or probable case that is laboratory-confirmed.

6.6.2 Appendix B: Tables

Table 17: Human plague cases in United States by state of exposure (1900–2017). Unpublished data courtesy of Ken Gage, Ph.D. at the U.S. Centers for Disease Control and Prevention in Fort Collins, Colorado.

State	Number of Human Cases			
	Total	Pre-1950	Post-1949	Post-1983
Arizona	68	0	68	27
California	492	441	48	26
Colorado	67	0	67	50
Florida	10	10	0	0
Idaho	5	1	4	3
Illinois*	1	0	1	1
Louisiana	51	51	0	0
Maryland*	1	0	1	0
Michigan	1	1	0	0
Montana	2	0	2	2
Nevada	6	0	6	3
New Mexico	283	3	280	140
Oklahoma	2	0	2	2
Oregon	19	1	18	9
Texas	35	31	4	3
Utah	16	1	15	10
Washington	9	8	1	1
Wyoming	5	0	5	3
Unknown	2	1	1	1
Total	1,045	522	523	281

*Laboratory Acquired

Table 18: Oregon State University Parameter elevation Regression on Independent Slopes Model (PRISM) 30-Year Average Annual Normals (1981–2010). Modeled using a combination of a digital elevation model (DEM) and climatologically-aided interpolation (CAI)*. Variables were modeled at 30 arcseconds (∼800 meters) resolution and aggregated to 2.5 arcminutes (∼4 kilometers). See (*125*) for more details.

Variable	Units	Derivation
Precipitation	millimeters (mm)	Modeled; Summing monthly averages (rain + melted snow)
Maximum Temperature	°Celsius (°C)	Modeled; Averaging over all months using a DEM as the predictor grid
Mean Temperature	°Celsius (°C)	Derived; Average of Maximum Temperature and Minimum Temperature
Minimum Temperature	°Celsius (°C)	Modeled; Averaging over all months using a DEM as the predictor grid
Mean Dewpoint Temperature	°Celsius (°C)	Modeled; CAI used minimum temperature as the predictor grid
Maximum Vapor Pressure Deficit	hectopascal (hPA)	Modeled; CAI used mean dewpoint temperature and maximum temperature as the predictor grids
Minimum Vapor Pressure Deficit	hectopascal (hPA)	Modeled; CAI used mean dewpoint temperature and minimum temperature as the predictor grids

*Accuracy of these data is based on the original specification of the Defense Mapping Agency (DMA) 1 degree digital elevation models (DEM). The stated accuracy of the original DEMs is 130 meter circular error with 90% probability. Datasets use all weather stations, regardless of time of observation.

Table 19: Summary of principal component analysis of Oregon State University Parameter elevation Regression on Independent Slopes Model (PRISM) 30-Year Average Annual Normals (1981–2010) at a 2.5 arcminute (∼4 km) resolution. Variables were masked to the western United States and were range standardized using the Gower metric (*197*).

Statistic	**Variable**	**Principal Component**		
		PC1	PC2	PC3
Standard Deviation		0.326	0.170	0.058
Proportion of Variance		0.756	0.205	0.024
Cumulative Proportion of Variance		0.756	0.962	0.985
Loadings	Precipitation	0.063	0.342	0.549
	Maximum Temperature	0.485	-0.208	-0.290
	Mean Temperature	0.519	-0.064	0.015
	Minimum Temperature	0.477	0.088	0.314
	Mean Dewpoint Temperature	0.446	0.532	-0.101
	Maximum Vapor Pressure Deficit	0.243	-0.533	-0.229
	Minimum Vapor Pressure Deficit	0.073	-0.510	0.673

Table 20: Specification of performed models. Two types of models were performed: Poisson and zero-inflated Poisson. Poisson model assumes count data distributed with equivalent mean and variance. A zero-inflated Poisson model assumes count data with many zero observations (*320*). Two zero-inflated Poisson models were used. Zero-inflated Poisson 0 assumes all zeros are structural zeros where zero was the only observable value. Zero-inflated Poisson 1 assumes some zeros are structural and others are sampling zeros where zeros were observed but may have been a different value. Models were compared using Deviance Information Criteria (DIC; *330*) and the Watanabe-Akaike information criterion (WAIC; *331*, *332*).

Number	Formula	Description	Visual Representation
$NULL_0$	1	Null model with offset of human population in 2010	Figure 73a
$NULL_1$	$1 + P_\Delta$	Null model with percent change in human population size (1950–2010) and offset	Figure 73
1	$NULL_i + ELEV$	PRISM variables are interpolations of a digital elevation model, county average	Figure 74
2	$NULL_i + PPT$	PRISM 30-year average annual total precipitation, county average	Figure 74
3	$NULL_i + T_{dew}$	PRISM 30-year average annual mean dewpoint temperature, county average	Figure 74
4	$NULL_i + T_{max}$	PRISM 30-year average annual maximum temperature, county average	Figure 74
5	$NULL_i + T_{mean}$	PRISM 30-year average annual mean temperature, county average	Figure 75
6	$NULL_i + T_{min}$	PRISM 30-year average annual minimum temperature, county average	Figure 75
7	$NULL_i + VPD_{max}$	PRISM 30-year average annual maximum vapor pressure deficit, county average	Figure 75
8	$NULL_i + VPD_{min}$	PRISM 30-year average annual minimum vapor pressure, county average	Figure 75
9	$NULL_i + PC_1$	First principal component (75.6% of variation), county average	Figure 76a
10	$NULL_i + PC_2$	Second principal component (20.5% of variation), county average	Figure 76b
11	$NULL_i + PC_1 + PC_2$	First and second principal components, county averages	Figure 76
12	$NULL_i + PC_1 + PC_2 + PC_1 * PC_2$	First and second principal components, county averages, with interaction	Figure 76
13	$NULL_i + C$	Categorical indicator of county-level results of coyote-based plague surveillance	Figure 77
14	$NULL_i + LRR_{raw}$	Log relative risk of plague (not accounting for sample bias), county average	Figure 78a
15	$NULL_i + PROP_{raw}$	Proportion of county significantly plague-suitable (not accounting for sample bias)	Figure 79a
16	$NULL_i + A_{raw,total}$	Total area of county plague-suitable (not accounting for sample bias)	Figure 80
17	$NULL_i + A_{raw,50}$	Area of county plague-suitable (not accounting for sample bias) if at least 50%	Figure 80
18	$NULL_i + A_{raw,75}$	Area of county plague-suitable (not accounting for sample bias) if at least 75%	Figure 80
19	$NULL_i + A_{raw,90}$	Area of county plague-suitable (not accounting for sample bias) if at least 90%	Figure 80
20	$NULL_i + LRR_{unbias}$	log relative risk of plague (accounting for sample bias), county average	Figure 78b
21	$NULL_i + PROP_{unbias}$	Proportion of county significantly plague-suitable (accounting for sample bias)	Figure 79b
22	$NULL_i + A_{unbias,total}$	Total area of county plague-suitable (accounting for sample bias)	Figure 81
23	$NULL_i + A_{unbias,50}$	Area of county plague-suitable (accounting for sample bias) if at least 50%	Figure 81
24	$NULL_i + A_{unbias,75}$	Area of county plague-suitable (accounting for sample bias) if at least 75%	Figure 81
25	$NULL_i + A_{unbias,90}$	Area of county plague-suitable (accounting for sample bias) if at least 90%	Figure 81

Total models: 27 variable combinations x 3 types of models = 81

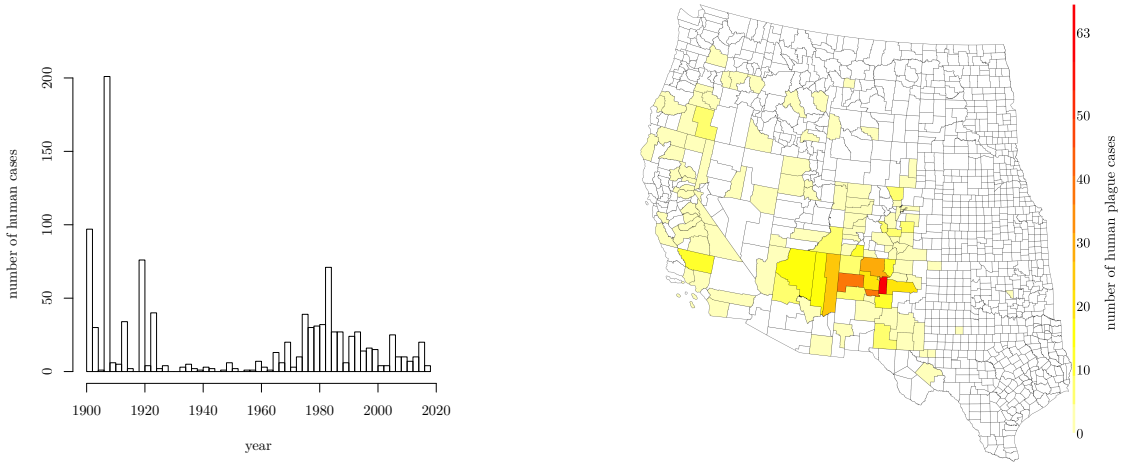
Table 21: Performance of all models. Models were compared using Deviance Information Criteria (DIC; *330*) and the Watanabe-Akaike information criterion (WAIC; *331*, *332*). Poisson model assumes count data distributed with equivalent mean and variance. A zero-inflated Poisson model assumes count data with many zero observations (*320*). Two zero-inflated Poisson models were used. Zero-inflated Poisson 0 (ZIP0) assumes all zeros are structural zeros where zero was the only observable value. Zero-inflated Poisson 1 (ZIP1) assumes some zeros are structural and others are sampling zeros where zeros were observed but may have been a different value. See Table 20 for a description of each model. The best-performing Poisson model and best-performing zero-inflated model are highlighted in yellow.

		Poisson		**ZIP0**		**ZIP1**	
Number	**Formula**	**DIC**	**WAIC**	**DIC**	**WAIC**	**DIC**	**WAIC**
$NULL_0$	1	732.21	1544.15	1108.13	1113.96	1310.09	1329.48
$NULL_1$	$1 + P_\Delta$	732.40	1085.42	1108.82	1113.00	1307.87	1328.72
1	$NULL_1 + ELEV$	729.58	736.60	1102.45	1105.38	815.57	836.59
2	$NULL_1 + PPT$	731.11	922.02	1107.74	1111.86	842.39	848.95
3	$NULL_1 + T_{dew}$	724.16	795.85	1098.78	1099.14	877.97	888.99
4	$NULL_1 + T_{max}$	733.76	858.38	1107.23	1110.09	882.73	876.50
5	$NULL_1 + T_{mean}$	734.40	813.31	1106.17	1108.87	870.99	871.94
6	$NULL_1 + T_{min}$	734.93	798.03	1105.05	1107.78	1317.32	1340.12
7	$NULL_1 + VPD_{max}$	733.23	1297.16	1109.77	1114.84	1301.55	1320.97
8	$NULL_1 + VPD_{min}$	732.86	1758.75	1109.70	1115.05	1300.03	1317.36
9	$NULL_1 + PC_1$	733.71	785.21	1105.25	1107.61	838.90	836.03
10	$NULL_1 + PC_1$	728.80	926.53	1107.35	1111.49	1301.62	1320.27
11	$NULL_1 + PC_1 + PC_2$	722.89	840.42	1099.26	1099.58	1193.00	1216.86
12	$NULL_1 + PC_1 + PC_2 + PC_1 * PC_2$	714.17	715.67	1099.13	1100.24	822.20	833.77
13	$NULL_1 + C$	727.04	1294.41	1106.32	1108.98	1298.82	1312.94
14	$NULL_1 + LRR_{raw}$	720.84	2251.52	1104.51	1107.92	758.45	764.89
15	$NULL_1 + PROP_{raw}$	721.78	820.26	1103.43	1106.56	1302.63	1319.78
16	$NULL_1 + A_{raw,total}$	730.95	891.64	1103.83	1106.16	1307.34	1324.76
17	$NULL_1 + A_{raw,50}$	732.17	927.85	1104.96	1107.70	1306.57	1326.01
18	$NULL_1 + A_{raw,75}$	732.23	887.48	1104.15	1106.74	1310.09	1326.31
19	$NULL_1 + A_{raw,90}$	733.09	960.59	1306.55	1326.12	1306.55	1326.12
20	$NULL_1 + LRR_{unbias}$	728.72	3647.21	1104.16	1108.35	1614.25	1643.29
21	$NULL_1 + PROP_{unbias}$	725.33	890.37	1107.39	1112.09	1302.21	1317.88
22	$NULL_1 + A_{unbias,total}$	732.07	962.29	1107.57	1111.38	1307.84	1325.11
23	$NULL_1 + A_{unbias,50}$	732.07	999.52	1108.23	1112.08	1307.95	1325.30
24	$NULL_1 + A_{unbias,75}$	733.02	1014.93	1109.02	1113.40	1309.00	1327.03
25	$NULL_1 + A_{unbias,90}$	732.84	1012.56	1109.30	1113.66	1305.73	1326.03

Table 22: Summary statistics of the best-performing Poisson model and zero-inflated Poisson model appear here with their reference models. Posterior estimates of fixed effects (β_i) and their 95% credibility interval (CI) of each model given their formula also appear here as well as posterior estimates for hyperparameters, including the precision for the spatially-unstructured random effect (τ_ϕ), precision for the spatially-structured random effect (τ_γ), and the zero-probability parameter (π_0). The proportion of variance explained by the structured spatial component $frac_\phi$ is also included. See Table 20 for a description of each model.

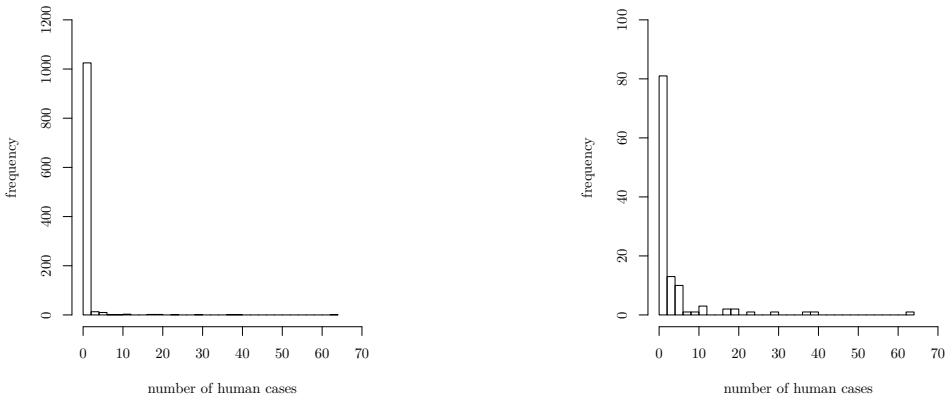
Type	**Variable**	**Estimate**	**Mean**	**Standard Deviation**	**Lower 95% CI**	**Upper 95% CI**	$frac_\phi$
Poisson	Intercept	β_0	-14.822	0.367	-15.599	-14.159	0.9996
	P_Δ	β_1	-0.0013	4e-04	-0.0020	-6e-04	
	PC_1	β_2	-4.818	1.013	-6.810	-2.826	
	PC_2	β_3	-14.437	2.015	-18.499	-10.577	
	$PC_1 * PC_2$	β_4	21.177	4.827	12.043	31.024	
		τ_ϕ	1889.3	1850.1	127.4	6779.1	
		τ_γ	0.159	0.030	0.108	0.226	
Zero-inflated	Intercept	β_0	-11.948	0.245	-12.440	-11.477	0.8439
Poisson 1	P_Δ	β_1	-9e-04	4e-04	-0.0017	-0.00028	
	LRR_{raw}	β_2	1.791	0.170	1.472	2.139	
		π_0	0.200	0.002	0.197	0.206	
		τ_ϕ	3.124	0.019	3.089	3.163	
		τ_γ	0.423	0.005	0.417	0.435	
References							
Poisson	Intercept	β_0	-15.065	0.394	-15.901	-14.359	0.9999
	P_Δ	β_1	-0.0012	4e-04	-0.0020	-4e-04	
		τ_ϕ	1855.6	1835.1	128.8	6700.6	
		τ_γ	0.082	0.013	0.059	0.111	
Zero-inflated	Intercept	β_0	-2.467	0.168	-2.793	-2.134	0.9995
Poisson 1	P_Δ	β_1	-0.0024	4e-04	-0.0031	-0.00165	
		π_0	0.886	0.010	0.866	0.905	
		τ_ϕ	470.233	458.628	25.026	1682.908	
		τ_γ	0.042	0.005	0.032	0.051	
Zero-inflated	Intercept	β_0	-12.948	0.189	-13.330	-12.589	0.5535
Poisson 1	P_Δ	β_1	-0.0017	3e-04	-0.0023	-0.0012	
	PC_1	β_2	-4.381	0.582	-5.541	-3.256	
	PC_2	β_3	-13.295	1.154	-15.600	-11.069	
	$PC_1 * PC_2$	β_3	15.448	3.411	8.544	21.944	
		π_0	0.389	0.008	0.374	0.406	
		τ_ϕ	2.732	0.196	2.386	3.153	
		τ_γ	2.961	0.184	2.589	3.308	
Poisson	Intercept	β_0	-12.841	0.383	-13.640	-12.136	0.9995
	P_Δ	β_1	-8e-04	4e-04	-0.0016	-4e-05	
	LRR_{raw}	β_2	1.707	0.208	1.312	2.129	
		τ_ϕ	1466.2	1329.2	86.3	4964.6	
		τ_γ	0.142	0.027	0.097	0.202	

6.6.3 Appendix C: Figures



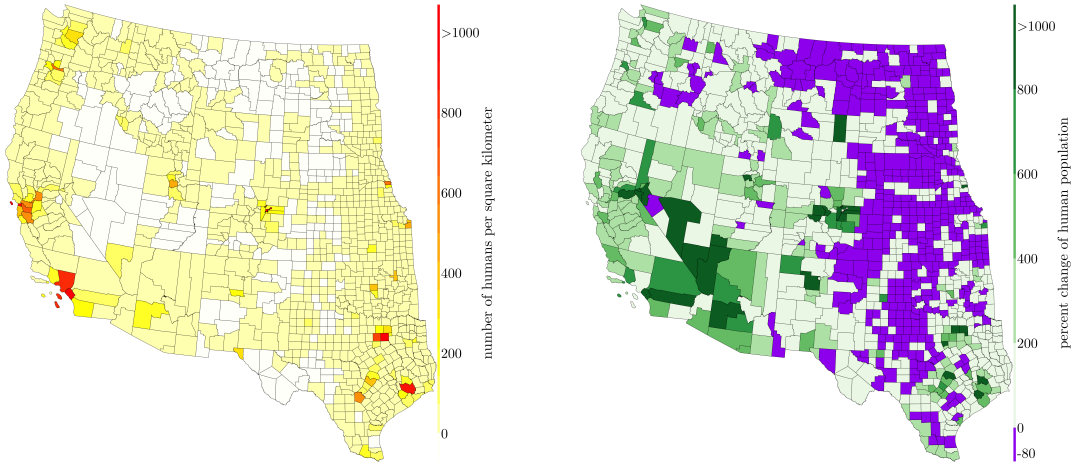
(a) **(b)**

Figure 71: (a) Human plague cases in contiguous United States since introduction (1900–2017). The early period (1900–1930) contained outbreaks in port cities (e.g., San Francisco, CA and Los Angeles, CA), but by 1950 plague became locally endemic in the western United States and human cases occurred in the country interior only. **(b)** Human plague cases in western United States (1950–2017) by county of exposure. Plague spread from California to Kansas by the 1940s (*86*). Since 1950, human plague cases have occurred in 13 western states (521 total cases) where over 80% of these cases have occurred in the Four Corner States (Arizona, Colorado, New Mexico, and Utah; *32*, *33*, *103*). Data are unpublished and courtesy of Kenneth Gage, Ph.D. at the U.S. Centers for Disease Control and Prevention in Fort Collins, Colorado.



(a) **(b)**

Figure 72: **(a)** Frequency of human plague case counts in western United States (1950–2017; $n = 1,062$ counties, or 89%). The vast majority of western United States counties have not had a human case ($n = 944$). **(b)** Frequency of human plague case counts in western United States (1950–2017) for counties with at least one case ($n = 118$ counties). The almost half of western United States counties with at least one human plague case have only one case ($n = 58$ counties or 49%).



(a) **(b)**

Figure 73: **(a)** Human population density of the western United States (2010) per square kilometer by county (*132*). **(b)** Percent change of human population in the western United States 1950 to 2010 by county (*132*, *186*). Counties without 1950 census data ($n = 3$) were given earliest population estimate. Broomfield County, Colorado was created in 2001 from Boulder County, Colorado but existed as a town since 1961 ($n = 4,535$ people; *187*). Cibola County, New Mexico was created in 1981 from Valencia County, New Mexico and its first census was in 1990 ($n = 23,794$ people; *188*). La Paz County, Arizona was created in 1983 from Yuma County, Arizona and its first census was in 1990 ($n = 13,844$ people; *188*).

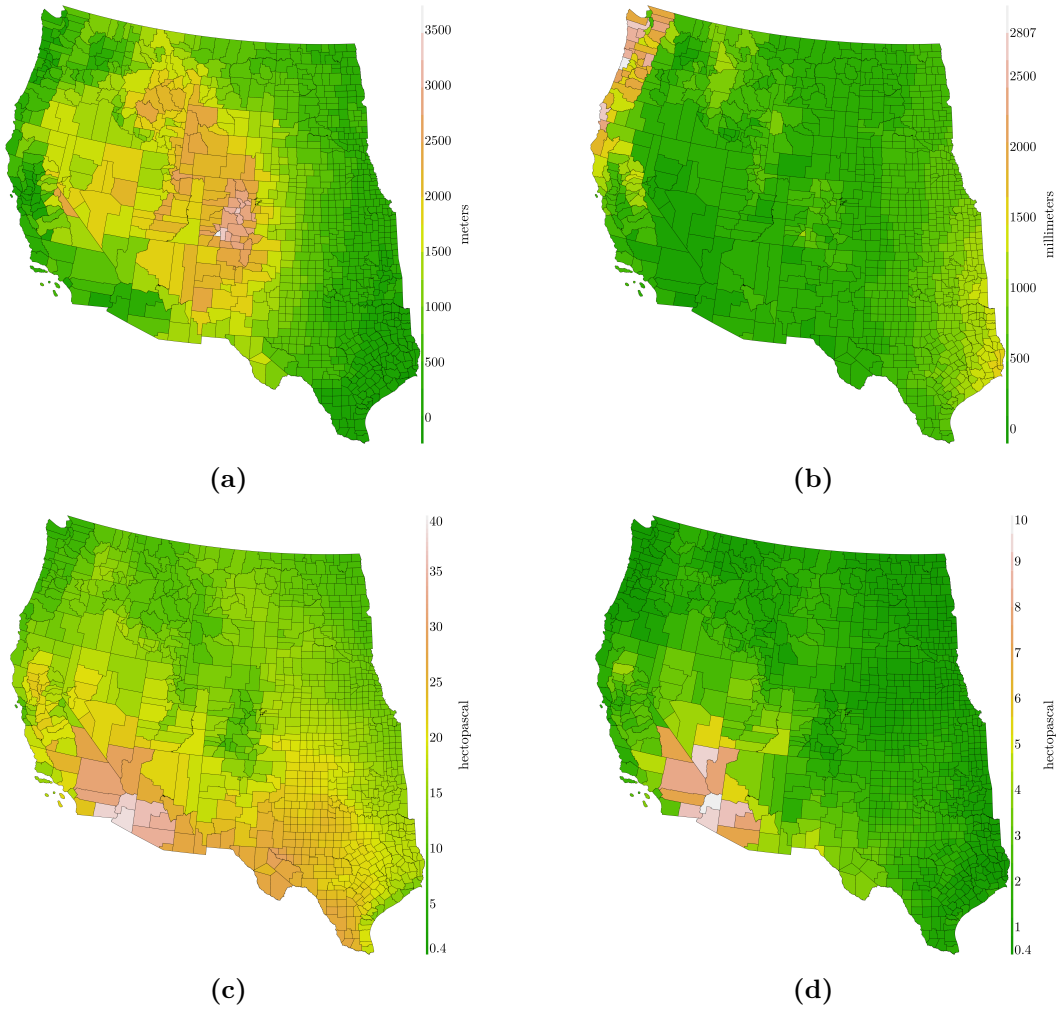


Figure 74: A subset of environmental variables used in the study, averaged within counties. Climatological variables were aggregated from the Oregon State University Parameter Elevation Regression on Independent Slopes Model (PRISM) 30-year average annual normals (1981–2010) at a 2.5 arcminute (∼4 km) resolution (*125*). **(a)** Elevation aggregated from the Digital Terrain Elevation Data provided by the National Aeronautics and Space Administration Shuttle Radar Topology Mission (*128*). **(b)** Total annual precipitation. **(c)** Annual average maximum vapor pressure deficit. **(d)** Annual average minimum vapor pressure deficit.

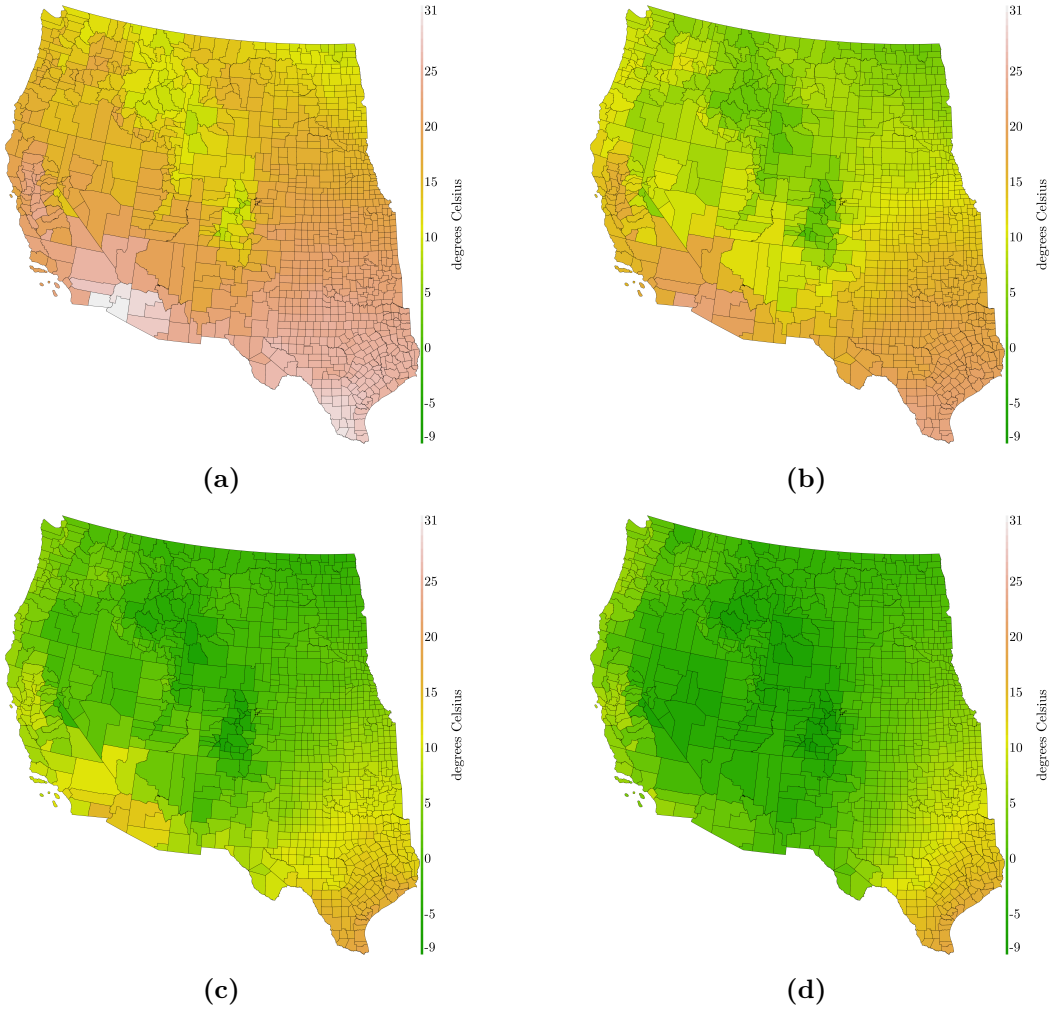
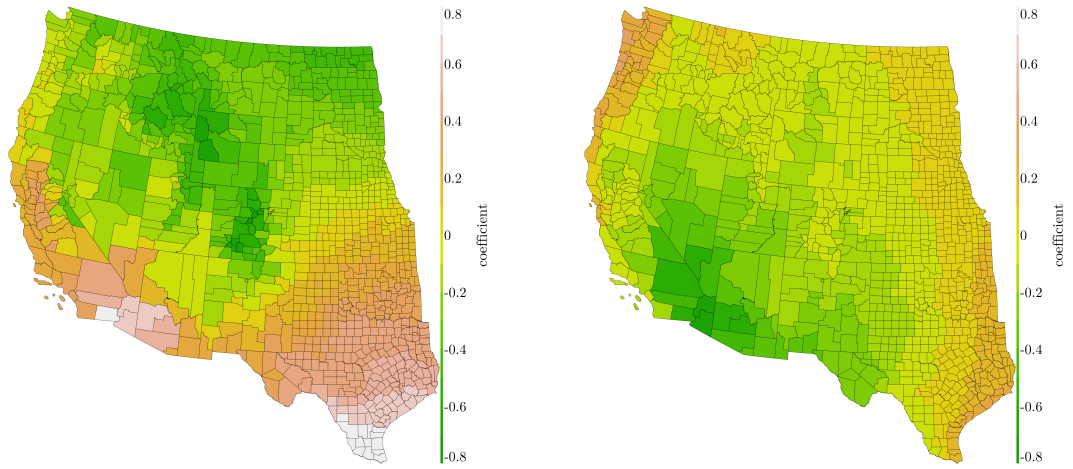


Figure 75: A subset of environmental variables used in the study, averaged within counties. Climatological variables were aggregated from the Oregon State University Parameter Elevation Regression on Independent Slopes Model (PRISM) 30-year average annual normals (1981–2010) at a 2.5 arcminute (∼4 km) resolution (*125*). **(a)** Annual average maximum temperature. **(b)** Annual average mean temperature. **(c)** Annual average minimum temperature. **(d)** Annual average mean dewpoint temperature.



(a) **(b)**

Figure 76: County average coefficient values of the top two principal components from a principal component analysis of seven range-standardized (*197*) Oregon State University Parameter Elevation Regression on Independent Slopes Model 30-year average annual normals (1981–2010) at a 2.5 arcminute (∼4 km) resolution of the contiguous United States (*125*). **(a)** First principal component accounted for 75.6% of the variation. **(b)** Second principal component accounted for 20.5% of the variation.

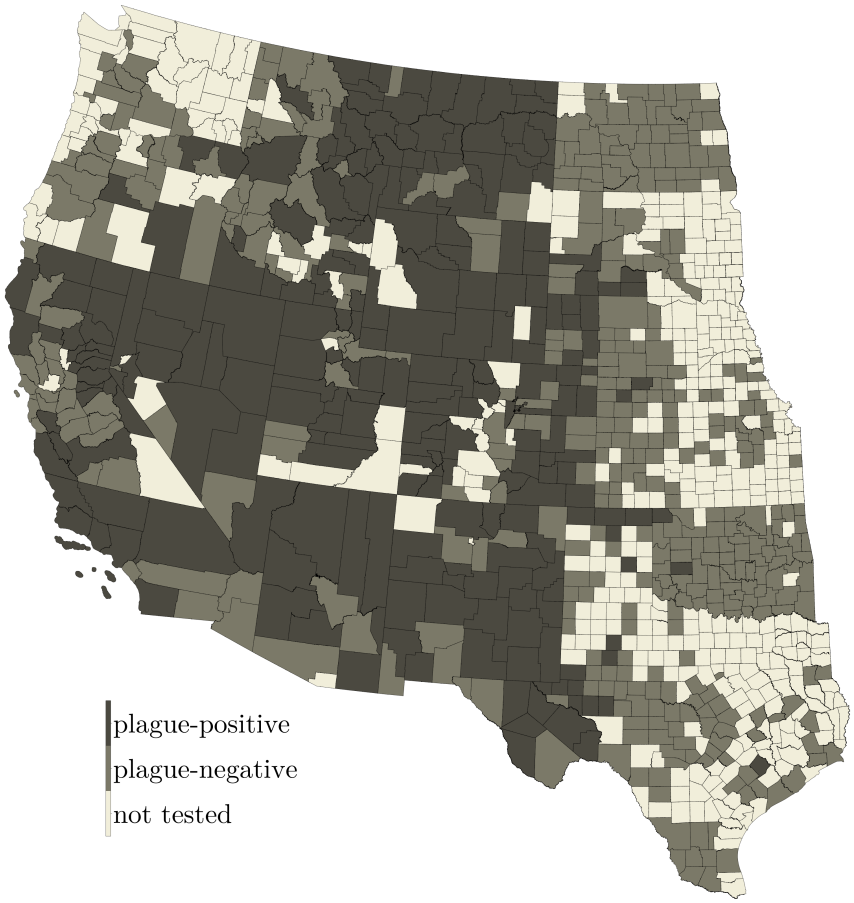
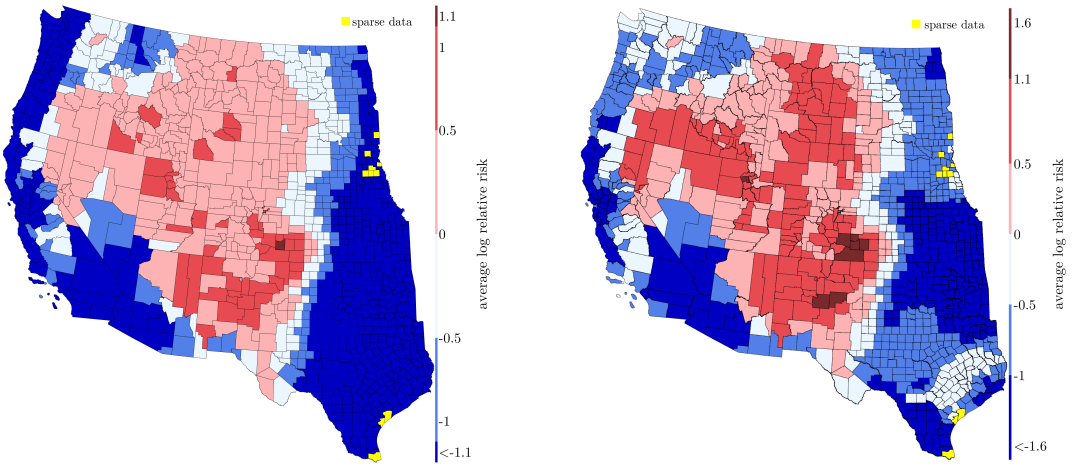
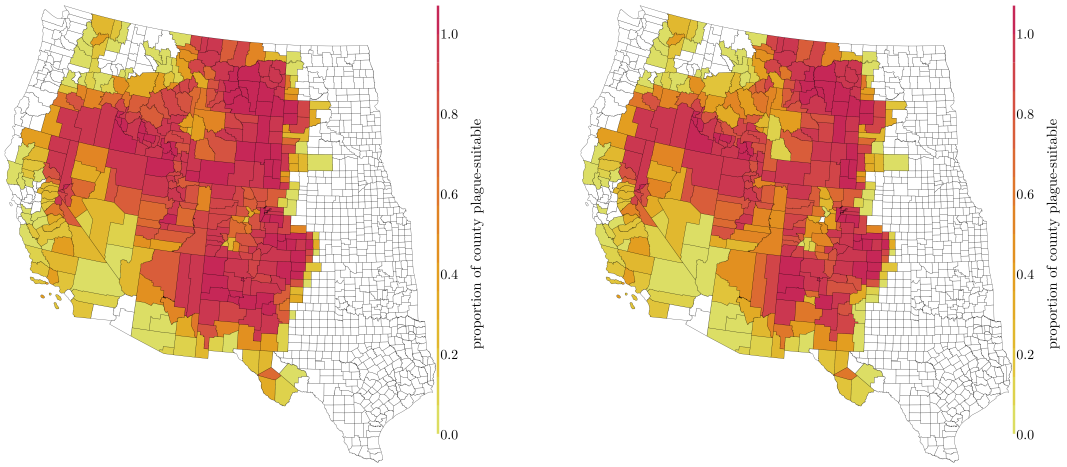


Figure 77: County-level detection of *Yersinia pestis* in coyote (*Canis latrans*) specimen across the western United States. Coyotes were collected and sampled by the U.S. Department of Agriculture (USDA) Animal and Plant Health Inspection Service Wildlife Services. The USDA, U.S. Centers for Disease Control and Prevention, and California Department of Public Health tested a large portion of these coyotes for antibodies against *Y. pestis* (1983–2017) in their respective laboratories. Some tested specimen were missing or had indiscernible geographic location information, primarily in Yakima County, Washington.



(a) **(b)**

Figure 78: County average log relative risk of plague presence in the environment. Values were predicted using a developed ecological niche model comparing seropositive coyotes and seronegative coyotes tested for *Y. pestis* antibodies by the U.S. Department of Agriculture (USDA; 2005–2017) and the California Department of Public Health (CDPH; 1983–2015) in predictor space. Predictor space was comprised of the first two principal components of a principal component analysis of seven range-standardized (*197*) Oregon State University Parameter Elevation Regression on Independent Slopes Model 30-year average annual normals (1981–2010) at a 2.5 arcminute (∼4 km) resolution of the contiguous United States (*125*). **(a)** County average log relative risk value (not accounting for sampling effort bias). **(b)** County average log relative risk value accounting for sampling effort bias. Sampling effort bias was accounted for by weighting raw log relative risk values by a standardized log relative risk of plague testing by the USDA and the CDPH. The log relative risk of plague testing was predicted using a developed ecological niche model comparing coyotes tested for *Y. pestis* antibodies to coyotes not tested for *Y. pestis* antibodies in predictor space. Coyotes not tested for plague antibodies were collated from historical coyote observations in museum repositories. Predictor space was the same as above. The log relative risk of plague testing was standardized by its minimum value and centered the null value at one.



(a) **(b)**

Figure 79: Proportion of county predicted suitable for *Yersinia pestis* transmission in the environment. Area within a county was considered suitable for plague transmission if its environment was within the estimated ecological niche of plague in coyotes. The ecological niche of plague in coyotes was estimated as the asymptotic tolerance (*219*, *221*) of the log relative risk surface at a two-tailed significance level ($\alpha = 0.05$). The log relative risk was estimated using a developed ecological niche model comparing seropositive coyotes and seronegative coyotes tested for *Y. pestis* antibodies by the U.S. Department of Agriculture (USDA; 2005–2017) and the California Department of Public Health (CDPH; 1983–2015) in predictor space. Predictor space was comprised of the first two principal components of a principal component analysis of seven range-standardized (*197*) Oregon State University Parameter Elevation Regression on Independent Slopes Model (PRISM) 30-year average annual normals (1981–2010) at a 2.5 arcminute (∼4 km) resolution of the contiguous United States (*125*). **(a)** Proportion of county predicted suitable for plague transmission (not accounting for sampling effort bias). **(b)** Proportion of county predicted suitable for plague transmission accounting for sampling effort bias. Sampling effort bias was accounted for by considering only plague-suitable area that was also sufficiently tested by state and federal agencies that monitor plague activity (USDA and CDPH). Areas were considered sufficiently tested for plague if its environment was within the observed ecological niche of coyotes that were more likely tested for plague than not tested for plague. The observed ecological niche of coyotes was estimated as the asymptotic tolerance (*219*, *221*) of the log relative risk surface at a two-tailed significance level ($\alpha = 0.05$). The log relative risk of plague testing was predicted using a developed ecological niche model comparing coyotes tested for *Y. pestis* antibodies to coyotes not tested for plague antibodies in predictor space. Coyotes not tested for *Y. pestis* antibodies were collated from historical coyote observations in museum repositories. Predictor space was the same as above.

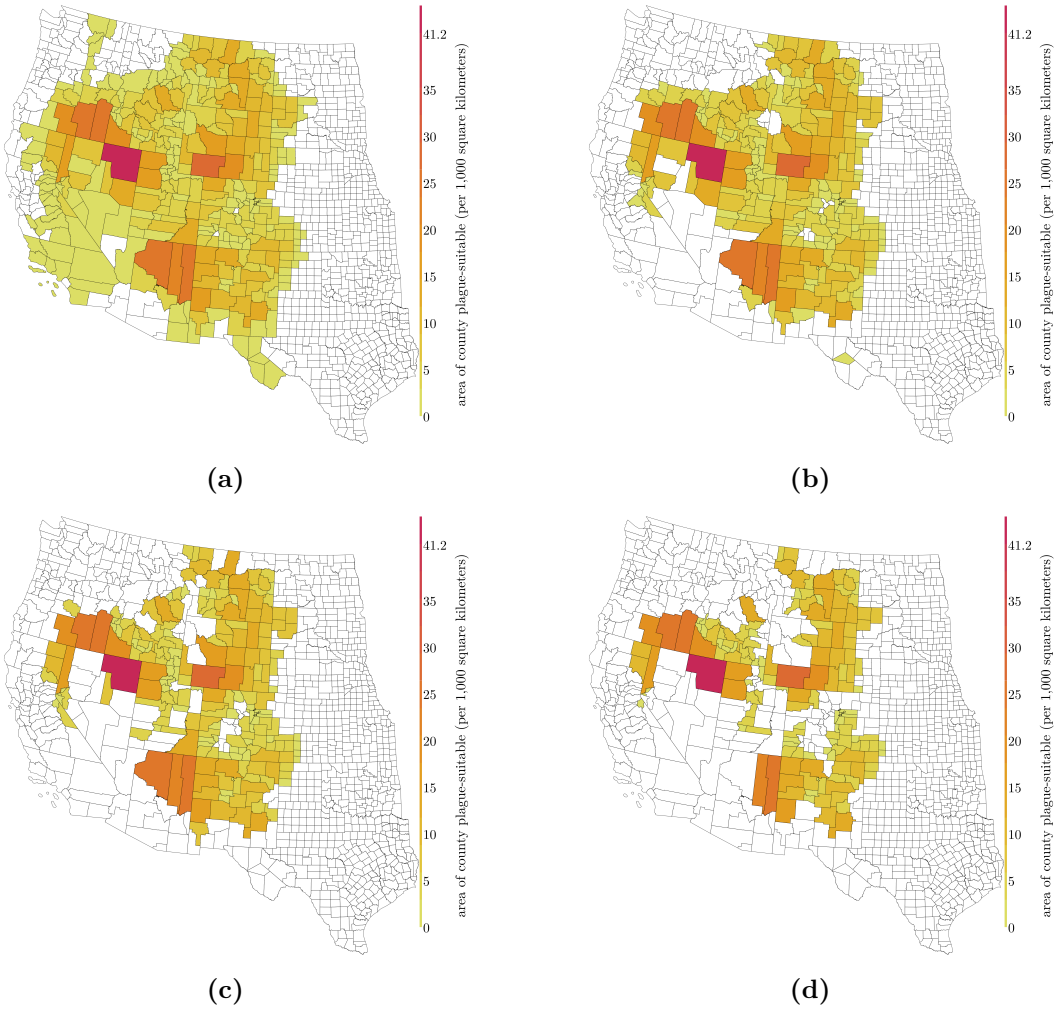


Figure 80: Area within a county predicted suitable for plague transmission in the environment at various proportion thresholds. Area within a county was considered suitable for plague transmission if its environment was within the estimated ecological niche of plague in coyotes. The ecological niche of plague in coyotes was estimated as the asymptotic tolerance (*219*, *221*) of the log relative risk surface at a two-tailed significance level ($\alpha = 0.05$). The log relative risk was estimated using a developed ecological niche model comparing seropositive coyotes and seronegative coyotes tested for *Y. pestis* antibodies by the U.S. Department of Agriculture (USDA; 2005–2017) and the California Department of Public Health (CDPH; 1983–2015) in predictor space. Predictor space was comprised of the first two principal components of a principal component analysis of seven range-standardized (*197*) Oregon State University Parameter Elevation Regression on Independent Slopes Model 30-year average annual normals (1981–2010) at a 2.5 arcminute (∼4 km) resolution of the contiguous United States (*125*). **(a)** All areas (no threshold). **(b)** Only counties with at least 50% of its area predicted plague-suitable. **(c)** Only counties with at least 75% of its area predicted plague-suitable. **(d)** Only counties with at least 90% of its area predicted plague-suitable.

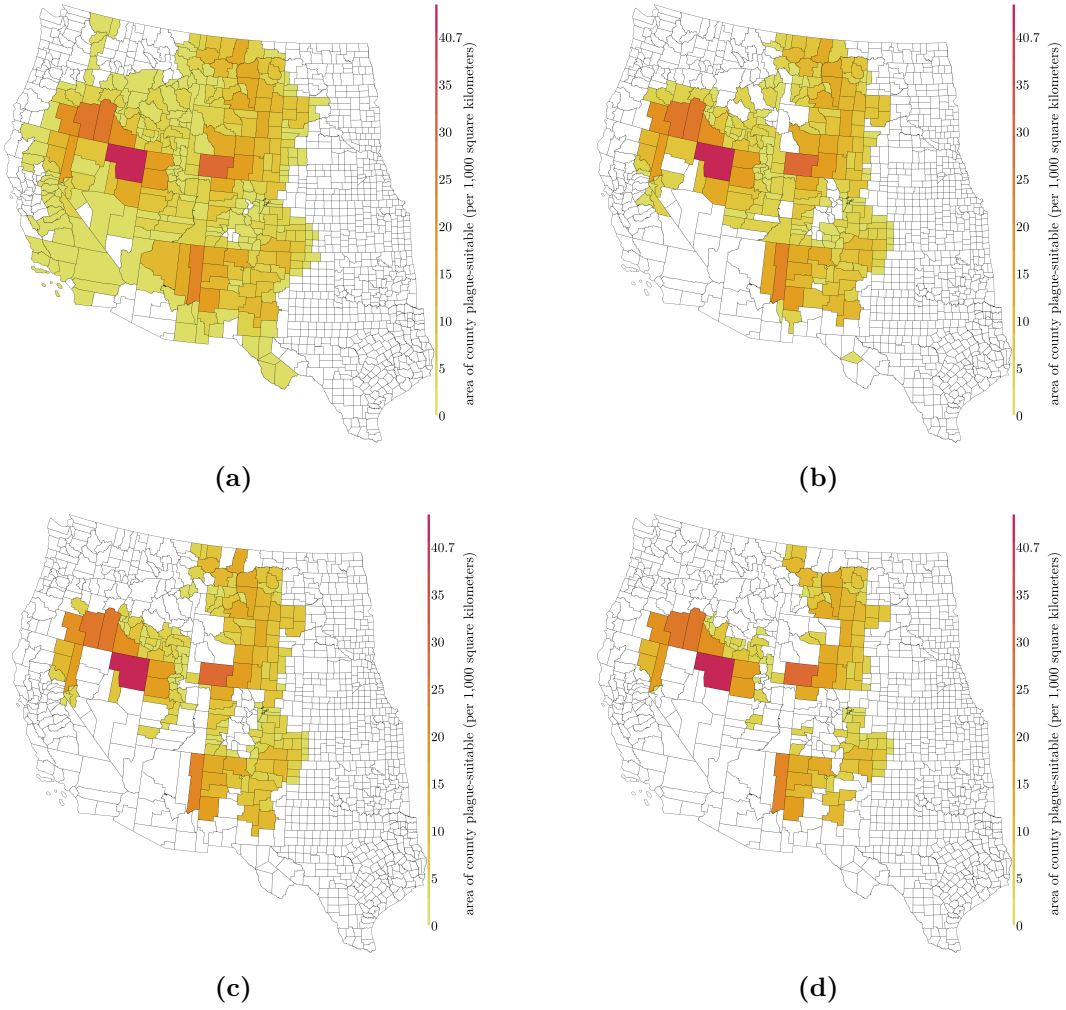
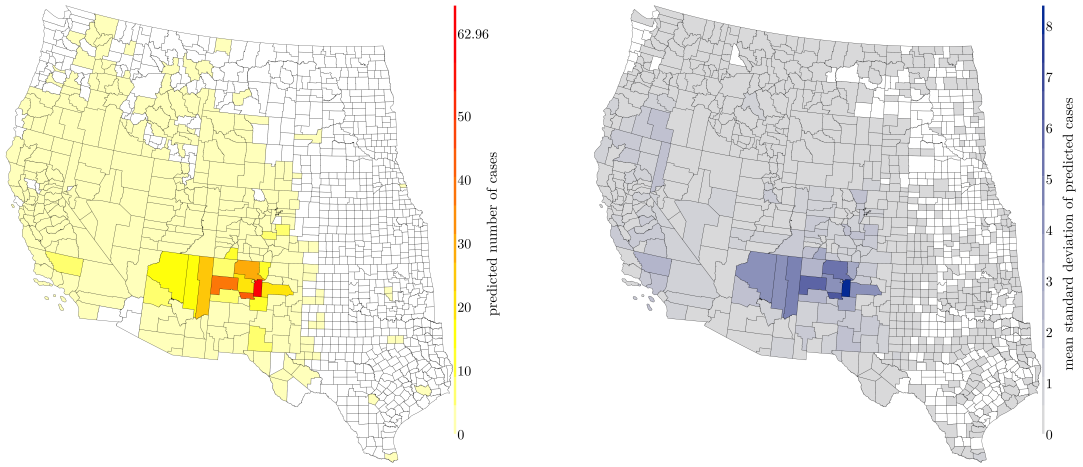
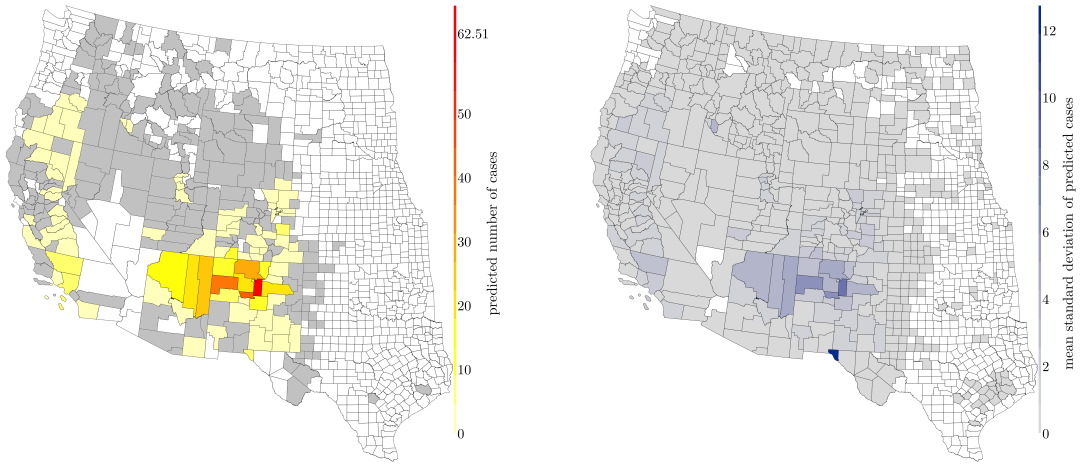


Figure 81: Area within a county predicted suitable for plague transmission in the environment while accounting for sampling effort bias at various proportion thresholds. Area within a county was considered suitable for plague transmission if its environment was within the estimated ecological niche of plague in coyotes. The ecological niche of plague in coyotes was estimated as the asymptotic tolerance (*219*, *221*) of the log relative risk surface at a two-tailed significance level ($\alpha = 0.05$). The log relative risk was estimated using a developed ecological niche model comparing seropositive coyotes and seronegative coyotes tested for *Y. pestis* antibodies by the U.S. Department of Agriculture (USDA; 2005–2017) and the California Department of Public Health (CDPH; 1983–2015) in predictor space. Predictor space was comprised of the first two principal components of a principal component analysis of seven range-standardized (*197*) Oregon State University Parameter Elevation Regression on Independent Slopes Model 30-year average annual normals (1981–2010) at a 2.5 arcminute (∼4 km) resolution of the contiguous United States (*125*). Sampling effort bias was accounted for by considering only plague-suitable area that was also sufficiently tested by state and federal agencies that monitor plague activity (USDA and CDPH). Areas were considered sufficiently tested for plague if its environment was within the observed ecological niche of coyotes that were more likely tested for plague than not tested for plague. The observed ecological niche of coyotes was estimated as the asymptotic tolerance (*219*, *221*) of the log relative risk surface at a two-tailed significance level ($\alpha = 0.05$). The log relative risk of plague testing was predicted using a developed ecological niche model comparing coyotes tested for *Y. pestis* antibodies to coyotes not tested for plague antibodies in predictor space. Coyotes not tested for *Y. pestis* antibodies were collated from historical coyote observations in museum repositories. Predictor space was the same as above. **(a)** All areas (no threshold). **(b)** Only counties with at least 50% of its area predicted plague-suitable. **(c)** Only counties with at least 75% of its area predicted plague-suitable. **(d)** Only counties with at least 90% of its area predicted plague-suitable. 235



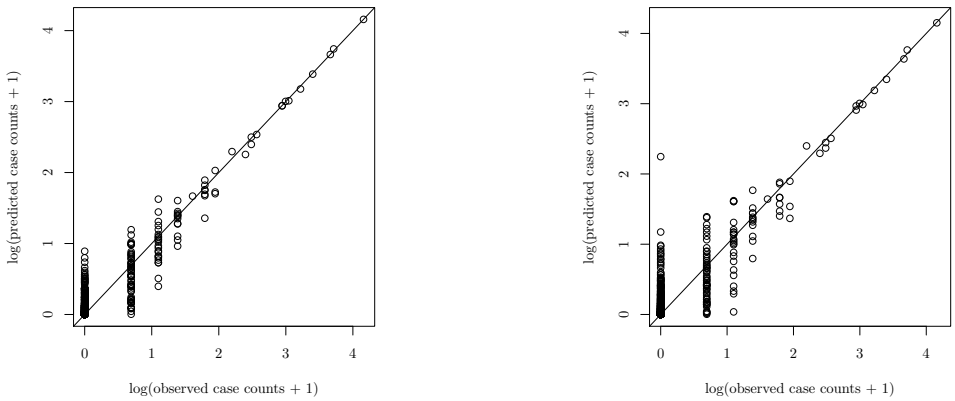
(a) **(b)**

Figure 82: Prediction of the best-performing Poisson model that used the first and second principal components of a principal component analysis of seven range-standardized (*197*) Oregon State University Parameter Elevation Regression on Independent Slopes Model 30-year average annual normals (1981–2010) at a 2.5 arcminute (∼4 km) resolution of the contiguous United States (*125*) and their interaction term. **(a)** County-specific posterior mean number of predicted human plague cases in the western United States (1950–2017). **(b)** County-specific posterior mean standard deviation of predicted human plague cases in the western United States (1950–2017).



(a) **(b)**

Figure 83: Prediction of the best-performing zero-inflated Poisson model (not assuming all zeros are structural) that used the log relative risk of plague averaged within county (not accounting for sampling effort bias). Log relative risk values were predicted using a developed ecological niche model comparing seropositive coyotes and seronegative coyotes tested for *Yersinia pestis* antibodies by the U.S. Department of Agriculture (2005–2017) and the California Department of Public Health (1983–2015) in predictor space. Predictor space was comprised of the first two principal components of a principal component analysis of seven range-standardized (*197*) Oregon State University Parameter Elevation Regression on Independent Slopes Model 30-year average annual normals (1981–2010) at a 2.5 arcminute (∼4 km) resolution of the contiguous United States (*125*). **(a)** County-specific posterior mean number of predicted human plague cases in the western United States (1950–2017). The grey-colored counties have a posterior mean between 0.1 and 1 predicted predicted human plague cases. **(b)** County-specific posterior mean standard deviation of predicted human plague cases in the western United States (1950–2017).



(a) **(b)**

Figure 84: Comparison of observed human plague case count in the western United States (1950–2017) and the predicted case counts. **(a)** Predicted counts of the best-performing Poisson model using the first and second principal components of a principal component analysis of seven range-standardized (*197*) Oregon State University Parameter Elevation Regression on Independent Slopes Model 30-year average annual normals (1981–2010) at a 2.5 arcminute (∼4 km) resolution of the contiguous United States (*125*) and their interaction term. **(b)** Predicted counts of the best-performing zero-inflated Poisson model (not assuming all zeros are structural) using the log relative risk of plague averaged within county (not accounting for sampling effort bias). Log relative risk values were predicted using a developed ecological niche model comparing seropositive coyotes and seronegative coyotes tested for *Yersinia pestis* antibodies by the U.S. Department of Agriculture (2005–2017) and the California Department of Public Health (1983–2015) in predictor space. Predictor space was comprised of the first two principal components of a principal component analysis of seven range-standardized (*197*) Oregon State University Parameter Elevation Regression on Independent Slopes Model 30-year average annual normals (1981–2010) at a 2.5 arcminute (∼4 km) resolution of the contiguous United States (*125*).

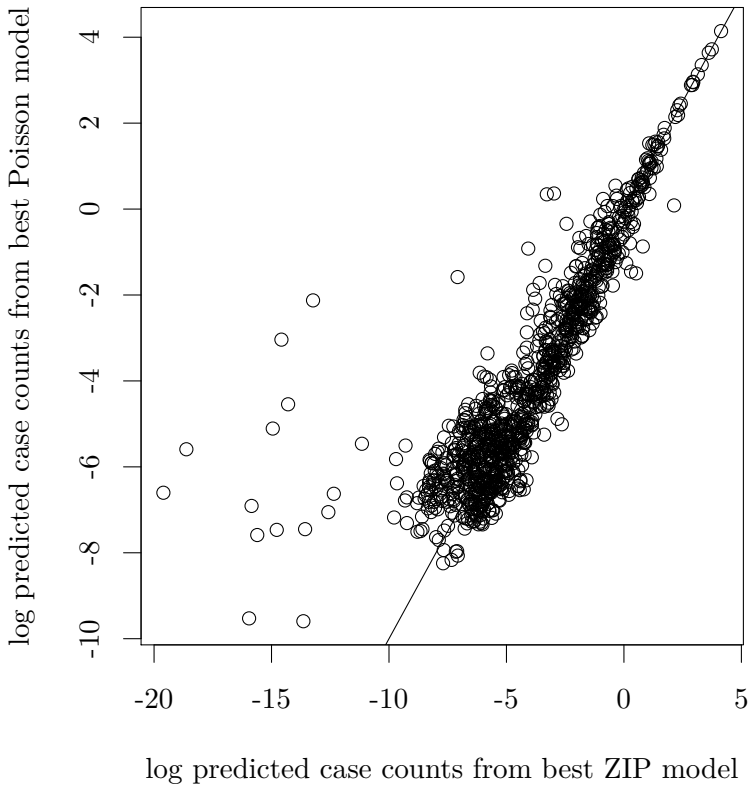
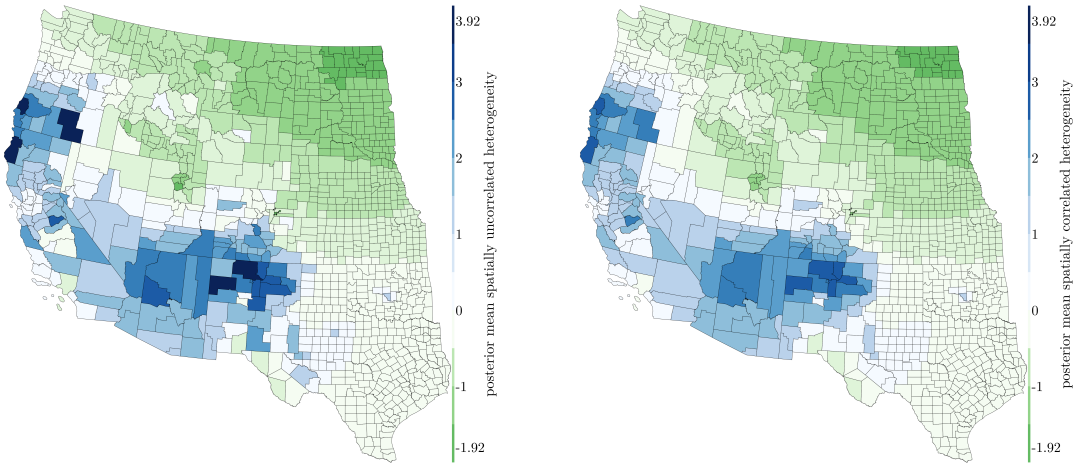
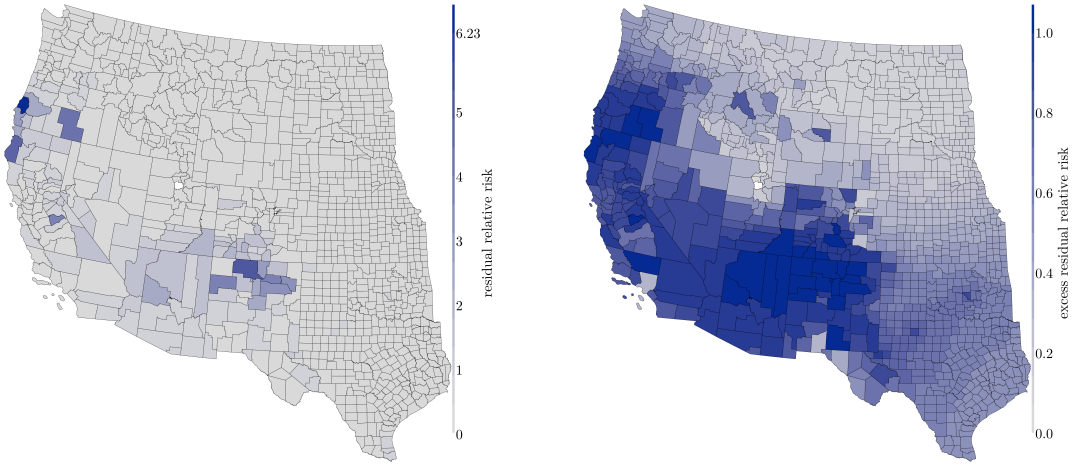


Figure 85: Comparison of predicted human plague case counts in the western United States (1950–2017) between the best Poisson model and the best zero-inflated Poisson (ZIP) model that did not assume all zeros were structural. The best-fitting Poisson model used the county average first and second principal components of a of a principal component analysis of seven range-standardized (*197*) Oregon State University Parameter Elevation Regression on Independent Slopes Model 30-year average annual normals (1981–2010) at a 2.5 arcminute (∼4 km) resolution of the contiguous United States (*125*) and their interaction. The best-performing ZIP model used the county average log relative risk of plague value (not accounting for sampling effort bias). Values were predicted using a developed ecological niche model comparing seropositive coyotes and seronegative coyotes tested for *Yersinia pestis* antibodies by the U.S. Department of Agriculture (2005–2017) and the California Department of Public Health (1983–2015) in predictor space. Predictor space was the same as above.



(a) **(b)**

Figure 86: County-specific residuals of the best-performing zero-inflated Poisson model. **(a)** Spatially-unstructured component. **(b)** Spatially-structured component. The fraction of variation explained by the spatially-structured component was high (about 0.84), which is shown in the similarity of the residuals especially in regions of the western United States with few to no human plague cases.



(a) **(b)**

Figure 87: County-specific relative risk the best-performing zero-inflated Poisson model after fixed-effects (percent change in population (1950–2010) and the log relative risk of plague not accounting for sampling effort bias) were taken into account. **(a)** Residual relative risk. The fraction of variation explained by the spatially-structured component was high (about 0.84) and counties with few to no cases near a county with many cases show high residual variability. **(b)** Excess residual relative risk (probability the residual relative risk is above 1). The fixed-effects are explaining the variability in the northern region of the western United States where few to no human plague cases have occurred. However, the fixed-effects are not explaining the variability in the southwestern United States where many human plague cases have occurred in a select few counties.

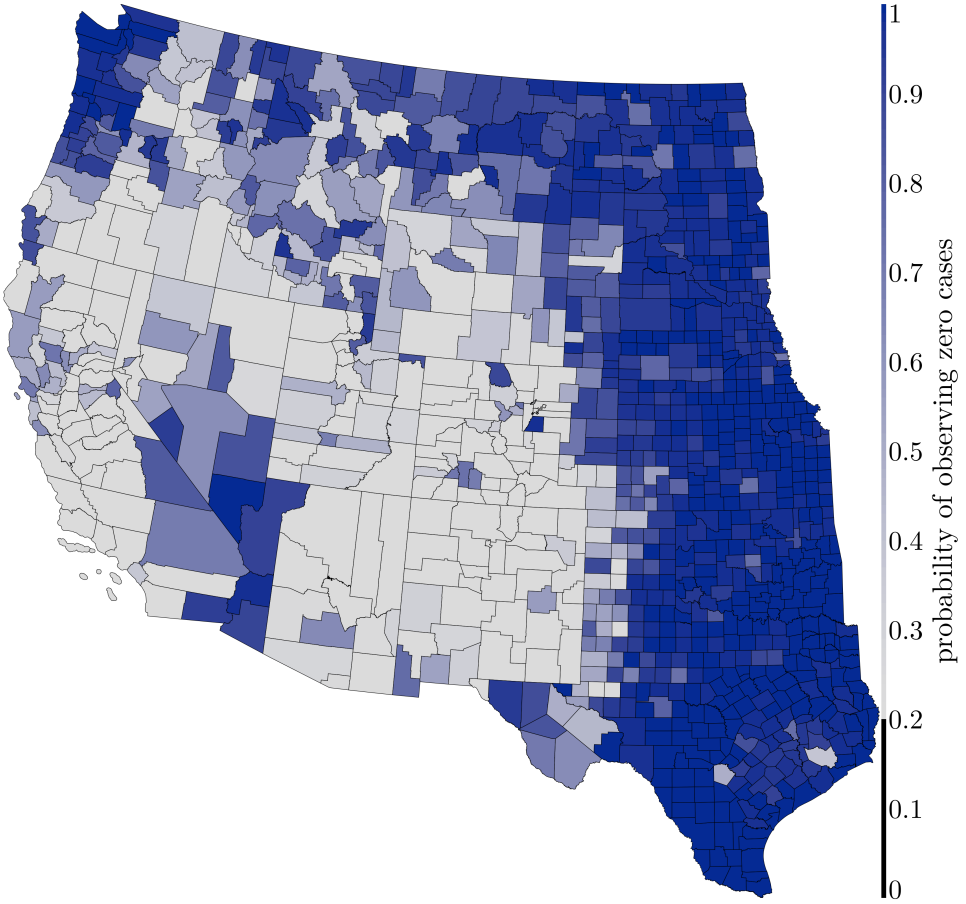


Figure 88: County-specific posterior mean zero-probability of a human plague case (1950–2017) based on the performing zero-inflated Poisson model. The minimum zero-probability was about 20% ($\pi_0 = 0.200$). Counties with a high probability of observing no human plague cases (1950–2017) were located in the northern and northwestern regions of the United States and the eastern Great Plains. The desert regions of California, Nevada, and Arizona also have a high probability of observing no human plague cases (1950–2017).

7 Discussion and Future Directions

"For much as men differ with regard to places in which they live, or in the law of their daily life, or in natural bent, or in active pursuits, or in whatever else man differs from man, in the case of this disease alone the difference availed naught. And it attacked some in the summer season, others in the winter, and still others at the other times of the year... spreading in either direction right out to the ends of the world, as if fearing lest some corner of the earth might escape it. For it left neither island nor cave nor mountain ridge which had human inhabitants... And this disease always took its start from the coast, and from there went up to the interior."

- Book 1, XXII of *History of the Wars* by Procopius
translated by H.B. Dewing (*333*)

7.1 Summary

Plague is a deadly disease of conservation and public health concern caused by the bacterium *Yersinia pestis*. Plague is dynamically active across the western United States, primarily within rodent species and their fleas (*31*). Surveillance is challenging and it is intractable to systematically monitor rodents across a region of over 4.7 million square kilometers in size where plague is suspected of being enzootic (*13*, *75*). Instead, I take advantage of an extensive ongoing national program that collects blood samples of a wide-ranging sentinel species for plague. Coyotes (*Canis latrans*) are carnivores that scavenge carcasses and prey on potentially infectious rodents, typically survive plague infection, and develop long-lasting antibodies against *Y. pestis* that can be detected in serological laboratory tests (*91*, *109*, *112*). In this dissertation, I collaborate with a state health department (California Department of Public Health) and a national agency (U.S. Department of Agriculture), both of whom routinely use coyote data in plague surveillance.

Along with coyote data, I define spatial statistical methods to predict areas of high risk for plague infection in animals and humans. In Chapter 3, I propose an ecological niche modeling approach to interpolate the spatial distribution of enzootic plague even into areas not historically sampled by large-scale administrative plague surveillance systems. In Chapter 5, I extend my prediction to the western United States and use this outcome

as a risk factor for human plague cases in Chapter 6. Human plague cases likely occur during epizootic events (*57*), but enzootic areas explain some of the spatial pattern of historical human plague cases (1950–2017) at a county level. I find that the spatial pattern in human plague cases is explained better by a county-level indicator for enzootic plague than climatological or topographical information alone. Plague epizootic events occur in enzootic areas under special climatologic or demographic conditions (*17*), so understanding enzootic spread of *Y. pestis* is a first step in forecasting future plague outbreaks.

I also find that my prediction of the spatial distribution of enzootic plague is conservative and robust. In Chapter 2, I recognize potential sources of data uncertainty and biases due to real-world plague surveillance limitations and prioritization of obligations by large-scale administrative surveillance systems. My method proposed in Chapter 3 accounts for a source of sampling effort bias by utilizing both case (presence) and control (absence) information. In Chapter 4, I use my proposed method to examine the effect of a source of sampling uncertainty in coyote data collected by the California Department of Public Health. I identify areas that are sensitive to this particular data uncertainty and develop a method to categorize coyotes by the degree of uncertainty in their sampling location. In Chapter 5, I further account for sampling effort bias by determining areas of the western United States that have not been sampled sufficiently by large-scale administrative plague surveillance systems. I adjust my predicted spatial distribution of enzootic plague based on sampling effort.

7.2 Future Directions

Coyotes are sentinel species for plague but are not a perfect proxy for sylvatic plague. The *Y. pestis* bacterium is a generalist pathogen found in numerous mammal species (*43*) and the plague niche is distinct from any mammal host niche (*85*). Therefore, including additional animal-based plague surveillance data would enhance my predictions of the spatial distribution of enzootic plague because one would sample more of the plague niche. Rodents are routinely monitored by various public health agencies and some flea information is recorded (*107*, *108*). Active rodent and flea plague surveillance is costly, frequently yields few positive detections unless sampled during an epizootic event (*107*, *109–111*, *147*, *296*,

334–336), and is not implemented systematically across or within states, due to limitations in personnel effort and budgetary constraints. In addition, flea data are spatially limited (*86*) and are unavailable for this dissertation. Pet surveillance has been and continues to be conducted by veterinary and state health agencies (*57*, *284*, *334*, *337*, *338*) and is a useful indicator for human risk, but pets are accidental hosts and are likely only indicators of epizootic events. Plague has been recorded in many other mammal species such as, for example, black bears (*Ursus americanus*) and mountain lions (*Felis concolor*), but these are not frequently used for surveillance because of low plague detection and limited number of observations (*85*, *107*, *339–343*).

A species' ecological niche is defined by abiotic factors, biotic interactions between hosts and pathogen, and limitations to dispersal (*344*). Disease transmission occurs in areas abiotically and biotically suitable and accessible for a pathogen (*12*). However, biotic interactions may be ignored for generalist pathogens when estimating its ecological niche at a coarse spatial scale (*12*). A generalist pathogen can maintain itself within a community of hosts and vectors (i.e. polyhostal and polyvectored; (*345*)) and abiotic conditions are more crucial for defining its ecological niche than biotic interactions (*20*, *85*). I use the Grinnellian definition of an ecological niche (*11*) because *Y. pestis* is a generalist pathogen. Plague dispersal appeared to halt at the 100th Meridian (*102*) and my results demonstrate this may be due to climatological factors. Biotic interactions between hosts and vectors can be assessed by examining the overlap in their ecological niches (*12*, *20*, *346*) or by using biotic factors in the ecological niche model (ENM). My proposed method in Chapter 3 is restricted to pairwise comparisons so other ENMs methods are suggested (e.g., NicheA; *300*) to assess niche overlaps. While Maher and colleagues (*85*) examined biotic interactions in the ecological niche of plague at a coarse spatial scale, future studies at finer spatial scales should incorporate biotic factors because they influence plague transmission at local scales (*149*, *231*).

Spatial scale is a considerable theme throughout the dissertation. I conduct an ENM at a 2.5 arcminute resolution (∼4 km). This spatial scale was chosen because of data availability and sampling uncertainty compromises (see Chapter 4). Future studies can investigate sensitivity of my results at various spatial scales, especially if rodent plague surveillance

data are included with coyotes. Rodents are precise indicators of plague, spatially and temporally; therefore, finer resolution climate and weather information can be used in an ENM leading to precise prediction of enzootic areas. Eisen and colleagues (*71–73*) were able to use human plague case data at a point level, which enabled the use of an approach similar to an ENM to predict areas of high-risk for human plague in the southwestern United States. I pursued finer resolution human plague case information for the Chapter 6 analysis but was restricted to a county-level analysis due to data privacy concerns. Finer resolution human plague data (i.e., at a census tract or census block) may be desirable for more precise risk estimation, but human plague cases are rare and at finer spatial scales contain many areal units with zero cases, further challenging statistical methods. A future human plague risk map could focus on the southwestern United States where the majority of human cases have occurred and then extrapolate to the greater western United States region.

My proposed method in Chapter 3 is a version of environmental interpolation (*19*) but future work could extrapolate the ecological niche of enzootic plague outside of the United States. Extrapolation comes with its own challenges, considerations, and interpretations. For example, ENMs find correlative matches between observed areas and predicted areas, but dispersal to or biotic interactions in the extrapolated areas limit the extrapolated spatial distribution of a species or disease (*12*, *20*, *344*, *347*). Plague is a global disease and *Y. pestis* has multiple enzootic foci across all continents except Oceania and Antarctica (*13*, *75*). Maher and colleagues (*85*) predicted plague in Canada where animal plague cases have been recorded (reviewed by *286*). There is little evidence of plague in Mexico (*348*), but an initial extrapolation study could extend my prediction to Canada and Mexico. Hay and colleagues (*316*) determined only two percent (seven of 355) of infectious diseases of clinical importance have been comprehensively charted. A global plague map is possible just like efforts for malaria (*29*, *30*), dengue (*349*), or Zika (*350*) because global plague data are available. There are records of plague occurring in South America (reviewed by *290*), Central Asia (*351*), the Middle East (*53*), Vietnam (*352*), and Madagascar (*353*). Neerinchx and colleagues (*68*) used an ENM to predict the spatial distribution of plague on the African continent. Other investigators have assessed the plague foci in China (*61*, *64*, *354*, *355*). A collaborative effort can compile plague records and estimate the global

ecological niche of plague. There may be more than one niche given plague has enzootic cycles driven by native rodent populations and epizootic episodes drive by rats in urban environments (*31*). Combined with genetic information, a future study could also assess the conservatism of the plague niche (*356*) by comparing Old World and New World plague niches.

Of interest to plague surveillance systems is how plague will be affected by global climate change and demographic shifts. Nakazawa and colleagues (*70*) used human plague cases and predicted a slight northern shift of the spatial distribution of plague in 2055. My prediction of the ecological niche of plague in coyotes is a cross-sectional study that temporally aggregates coyote plague data (1983–2017) due to data limitations. The U.S. Department of Agriculture has a large sample of coyote observations but only for a 12-year period. A future study could predict a future spatial distribution of plague in California using the California Department of Public Health coyote plague data, but an assessment of spatial and temporal sampling effort will need to accompany the analysis. Shifts in the ecological niche of plague in coyotes would be interesting if the spatial distribution of plague varied between decades or years. Persistent locations could potentially be refugia (i.e., ecological source) of *Y. pestis*, but see (*357*) for important considerations. Identifying refugia can inform plague surveillance systems by focusing resources to heavily monitor the activity within and dispersal outside of these areas. Human plague cases (1950–2017) were temporally aggregated because of their small sample size. Human plague cases have been associated with climatological oscillations and epizootic events (*57*, *59*, *95*), but a spatio-temporal human plague analysis may only be feasible in the southwestern United States where the majority of human plague cases have occurred.

Environmental factors alone cannot explain human plague risk. Risk factors include living in close proximity to rodents (*31*), available sources of rodent food and harborage near a domicile (*358*), and handling, skinning, butchering or consuming an infected animal (*38*, *55*). Pet ownership is another risk factor (*31*), especially indoor-outdoor cat ownership (*359*), sleeping with a dog on the bed (*38*), or low-levels of flea control for pets (*358*). Human plague cases in the United States are associated with sociodemographic factors and have changed over time. Schotthoefer and colleagues (*318*) determined plague

became a disease of affluent areas of New Mexico in the 2000s with cases shifting away from impoverished areas since the 1990s. The shift may be the result of public health programs in impoverished areas or the overall decline in small game hunting (*360*). However, population-level information about specific plague risk factors are not available and the rarity of human cases poses statistical challenges.

7.3 Broader Impact

My results provide both applied and theoretical contributions. I work closely with government scientists within agencies that monitor plague activity and their ongoing constructive input ensures that my results have immediate impact. My results pinpoint locations for further plague surveillance and can inform public health policy. A map of enzootic areas of plague can help inform conservation and public health programs such as, for example, by notifying doctors in high relative risk counties or strategically delivering plague vaccines for prairie dogs (*302*) and, potentially, humans at particular high risk (*303*). The U.S. Department of Agriculture (USDA), one of my collaborating agencies, collects coyote blood samples in concurrence with a separate, ongoing USDA program of livestock and wildlife harassment management. Therefore, my results will not influence future coyote blood sample collection, but my results identify undersampled regions of the western US, which can help prioritize laboratory testing of coyote samples.

My proposed method (Chapter 3) is a useful tool to predict the occurrence of a disease (or species) using case (presence) and control (absence) location information. The method can be used to predict the spatial distribution of other zoonotic disease or wildlife species that may be of importance to public health or environmental conservation. Motivation for this dissertation arose from Nusser and colleagues' (*159*) work with a prion disease called chronic wasting disease (CWD), which is enzootic in North America and causes population declines in cervid species (i.e., deer and elk; (*361*, *362*)). While CWD has not been found in humans, transmission is possible under experimental conditions and surveillance is warranted (reviewed by *363*). The USDA also monitors wildlife diseases (e.g., rabies) and invasive species (e.g., feral swine, *Sus scrofa*). My proposed method (Chapter 3) can be used in a future ensemble of ecological niche modeling approaches to

predict the potential spatial distribution of these diseases and species. Ensemble ecological niche modeling is an exciting frontier because it controls for variation between individual ENM approaches (see *364*).

Beyond its use as an ENM, my proposed method (Chapter 3) can be applied to other scientific fields. It is a type of nonparametric multiplicative regression (*207*) that can detect clusters in two-dimensional predictor space (i.e., "environmental space" or "state space"). These predictors can be comprised of any variable combinations. In this dissertation, I conduct my proposed method within a predictor space comprised of the first two principal components of a principal component analysis. Other fields also conduct analyses within principal component space with examples found in genetics (*365*, *366*), economics (*367*), immunology (*368*), neuroscience (*369*), water treatment (*370*), and bioinformatics (*371*).

In closing, I have shown how spatial statistical and data science methods can extend the value of administrative surveillance data, providing input on how to gain information from existing data and where and how often to look next.

7.4 References

11. J. Grinnell, *The Auk* **34**, 427–433, ISSN: 938-4254, DOI 10.2307/4072271 (1917).

12. A. T. Peterson *et al.*, *Ecological niches and geographic distributions (MPB-49)* (Princeton University Press, Princeton, New Jersey, 2011), vol. 56, ISBN: 0691136882.

13. A. T. Peterson, *Naturwissenschaften* **95**, 483–491, ISSN: 0028-1042, 1432-1904, DOI 10.1007/s00114-008-0352-5 (2008).

17. R. J. Eisen, K. L. Gage, *Veterinary Research* **40**, 1, ISSN: 1993-5412, DOI 10.1051/vetres:2008039 (2009).

19. A. T. Peterson, *Mapping disease transmission risk: enriching models using biogeography and ecology* (Johns Hopkins University Press, Baltimore, Maryland, 2014), ISBN: 1421414737.

20. L. E. Escobar, M. E. Craft, *Frontiers in Microbiology* **7**, 1174, ISSN: 1664-302X, DOI 10.3389/fmicb.2016.01174 (2016).

29. P. W. Gething *et al.*, *Malaria Journal* **10**, 378, ISSN: 1475-2875, DOI 10.1186/1475-2875-10-378 (2011).

30. P. W. Gething *et al.*, *PLoS Neglected Tropical Diseases* **6**, e1814, ISSN: 1935-2735, 1935-2727, DOI 10.1371/journal.pntd.0001814 (2012).

31. K. L. Gage, M. Y. Kosoy, *Annual Review of Entomology* **50**, 505–528, ISSN: 0066-4170, DOI 10.1146/annurev.ento.50.071803.130337 (2005).

38. H. L. Gould *et al.*, *Zoonoses and Public Health* **55**, 448–454, ISSN: 1863-1959, DOI 10.1111/j.1863-2378.2008.01132.x (2008).

43. R. Pollitzer, *Bulletin of the World Health Organization* **23**, 313, ISSN: 0042-9686 (1960).

53. A. Arbaji *et al.*, *Annals of Tropical Medicine and Parasitology* **99**, 789–793, ISSN: 0003-4983, DOI 10.1179/136485905X65161 (2005).

55. C. F. von Reyn *et al.*, *American Journal of Tropical Medicine and Hygiene* **25**, 626–629, ISSN: 0002-9637, 1476-1645, DOI 10.4269/ajtmh.1976.25.626 (1976).

57. H. E. Brown *et al.*, *American Journal of Tropical Medicine and Hygiene* **82**, 95–102, ISSN: 0002-9637, 1476-1645, DOI 10.4269/ajtmh.2010.09-0247 (2010).

59. T. Ben-Ari *et al.*, *American Journal of Tropical Medicine and Hygiene* **83**, 624–632, ISSN: 0002-9637, 1476-1645, DOI 10.4269/ajtmh.2010.09-0775 (2010).

61. T. Ben-Ari *et al.*, *Proceedings of the National Academy of Sciences* **109**, 8196–8201, ISSN: 1091-6490, DOI 10.1073/pnas.1110585109 (2012).

64. L. Xu *et al.*, *Proceedings of the National Academy of Sciences of the United States of America* **108**, 10214–10219, ISSN: 1091-6490, 0027-8424, DOI 10.1073/pnas.1019486108 (2011).

68. S. B. Neerinckx *et al.*, *International Journal of Health Geographics* **7**, 54, ISSN: 1476-072X, DOI 10.1186/1476-072X-7-54 (2008).

70. Y. Nakazawa *et al.*, *Vector-Borne and Zoonotic Diseases* **7**, 529–540, ISSN: 1557-7759, 1530-3667, DOI 10.1089/vbz.2007.0125 (2007).

71. R. J. Eisen *et al.*, *American Journal of Tropical Medicine and Hygiene* **77**, 999–1004, ISSN: 0002-9637, 1476-1645, DOI 10.4269/ajtmh.2007.77.999 (2007).

72. R. J. Eisen *et al.*, *Journal of Medical Entomology* **44**, 530–537, ISSN: 0022-2585, DOI 10.1603/0022-2585(2007)44 (2007).

73. R. J. Eisen *et al.*, *American Journal of Tropical Medicine and Hygiene* **77**, 121–125, ISSN: 0002-9637, 1476-1645, DOI 10.4269/ajtmh.2007.77.121 (2007).

75. M. Anker, D. Schaaf, in *WHO report on global surveillance of epidemic-prone infectious diseases* (World Health Organization, Geneva, Switzerland, 2000), chap. 3, pp. 25–37.
85. S. P. Maher *et al.*, *American Journal of Tropical Medicine and Hygiene* **83**, 736–742, ISSN: 0002-9637, 1476-1645, DOI 10.4269/ajtmh.2010.10-0042 (2010).
86. J. C. Z. Adjemian *et al.*, *Journal of Medical Entomology* **43**, 93–103, ISSN: 0022-2585, DOI 10.1093/jmedent/43.1.93 (2006).
91. D. J. Salkeld, P. Stapp, *Vector-Borne and Zoonotic Diseases* **6**, 231–239, ISSN: 1557-7759, 1530-3667, DOI 10.1089/vbz.2006.6.231 (2006).
95. H. E. Brown *et al.*, *Vector-Borne and Zoonotic Diseases* **11**, 1439–1446, ISSN: 1557-7759, 1530-3667, DOI 10.1089/vbz.2010.0196 (2011).
102. J. Z. Adjemian *et al.*, *American Journal of Tropical Medicine and Hygiene* **76**, 365–375, ISSN: 0002-9637, 1476-1645, DOI 10.4269/ajtmh.2007.76.365 (2007).
107. S. N. Bevins *et al.*, *Integrative Zoology* **7**, 99–109, ISSN: 1749-4877, DOI 10.1111/j.1749-4877.2011.00277.x (2012).
108. J. R. Tucker *et al.*, *California compendium of plague control*, 2015, (2018; https://www.cdph.ca.gov/Programs/CID/DCDC/CDPH%20Document%20Library/CAPlagueCompendium.pdf).
109. K. L. Gage *et al.*, in *Proceedings of the Sixteenth Vertebrate Pest Conference*, pp. 200–206.
110. B. Thiagarajan *et al.*, *Journal of Wildlife Diseases* **44**, 731–736, ISSN: 0090-3558, DOI 10.7589/0090-3558-44.3.731 (2008).
111. R. D. Holt, M. Barfield, in *Spatial ecology*, ed. by S. Cantrell *et al.* (CRC Press, Boca Raton, Florida, 2010), pp. 189–211, ISBN: 1420059858.
112. P. W. Willeberg *et al.*, *American Journal of Epidemiology* **110**, 328–334, ISSN: 0002-9262, DOI 10.1093/oxfordjournals.aje.a112818 (1979).
147. C. R. Smith *et al.*, *Journal of Vector Ecology* **35**, 1–12, ISSN: 1081-1710, DOI 10.1111/j.1948-7134.2010.00051.x (2010).
149. P. Stapp *et al.*, *Frontiers in Ecology and the Environment* **2**, 235–240, ISSN: 1540-9295, DOI 10.1890/1540-9295(2004)002[0235:POEIPD]2.0.CO;2 (2004).
159. S. M. Nusser *et al.*, *Journal of Wildlife Management* **72**, 52–60, ISSN: 0022-541X, DOI 10.2193/2007-317 (2008).
207. B. McCune *et al.*, *Analysis of ecological communities* (MjM Software Design, Gleneden Beach, Oregon, 2002), vol. 28, ISBN: 0972129008.
231. S. K. Collinge *et al.*, *EcoHealth* **2**, 102–112, ISSN: 1612-9202, 1612-9210, DOI 10.1007/s10393-005-3877-5 (2005).
284. A. M. Barnes *et al.*, *Morbidity and Mortality Weekly Report: Surveillance Summaries*, 11–16, ISSN: 1546-0738, 1545-8636 (1988).
286. G. Wobeser *et al.*, *The Canadian Veterinary Journal* **50**, 1251, ISSN: 0008-5286, 0008-5286 (2009).
290. M. C. Schneider *et al.*, *PLoS Neglected Tropical Diseases* **8**, ISSN: 1935-2735, 1935-2727, DOI 10.1371/journal.pntd.0002680 (2014).
296. D. J. Salkeld *et al.*, *Journal of Wildlife Diseases* **43**, 425–431, ISSN: 0090-3558, DOI 10.7589/0090-3558-43.3.425 (2007).
300. H. Qiao *et al.*, *Ecography* **39**, 805–813, ISSN: 0906-7590, DOI 10.1111/ecog.01961 (2016).
302. R. C. Abbott *et al.*, *EcoHealth* **9**, 243–250, ISSN: 1612-9202, 1612-9210, DOI 10.1007/s10393-012-0783-5 (2012).

303. J. Hu *et al.*, *Human Vaccines & Immunotherapeutics* **14**, 2701–2705, ISSN: 2164-554X, DOI 10.1080/21645515.2018.1486154 (2018).
316. S. I. Hay *et al.*, *Philosophical Transactions of the Royal Society B: Biological Sciences* **368**, 20120250, ISSN: 0962-8452, 1471-2954, DOI 10.1098/rstb.2012.0250 (2013).
318. A. M. Schotthoefer *et al.*, *Emerging Infectious Diseases* **18**, 1151–1154, ISSN: 1080-6059, 1080-6040, DOI 10.3201/eid1807.120121 (2012).
333. Procopius, *History of the wars* (Gutenberg Project, 2005), vol. I and II (of VIII), (2019; http://www.gutenberg.org/ebooks/16764).
334. B. C. Nelson *et al.*, in *Proceedings of the Twelfth Vertebrate Pest Conference*, p. 46.
335. J. F. Cully Jr *et al.*, *Journal of Wildlife Diseases* **36**, 389–392, ISSN: 0090-3558, DOI 10.7589/0090-3558-36.2.389 (2000).
336. S. R. Ubico *et al.*, *Journal of Wildlife Diseases* **24**, 399–406, ISSN: 0090-3558, DOI 10.7589/0090-3558-24.3.399 (1988).
337. B. B. Chomel *et al.*, *Comparative Immunology, Microbiology and Infectious Diseases* **17**, 111–123, ISSN: 0147-9571, DOI 10.1016/0147-9571(94)90036-1 (1994).
338. D. M. Gubernot *et al.*, *Public Health Reports* **123**, 300–315, ISSN: 0033-3549, 1468-2877, DOI 10.1177/003335490812300310 (2008).
339. J. R. Clover *et al.*, *Journal of Wildlife Diseases* **25**, 52–60, ISSN: 0090-3558, DOI 10.7589/0090-3558-25.1.52 (1989).
340. S. N. Bevins *et al.*, *Emerging Infectious Diseases* **15**, 2021, ISSN: 1080-6059, 1080-6040, DOI 10.3201/eid1512.090526 (2009).
341. Y. A. Girard *et al.*, *Vector-Borne and Zoonotic Diseases* **12**, 913–921, ISSN: 1557-7759, 1530-3667, DOI 10.1089/vbz.2011.0858 (2012).
342. M. A. Wild *et al.*, *Journal of Wildlife Diseases* **42**, 646–650, ISSN: 0090-3558, DOI 10.7589/0090-3558-42.3.646 (2006).
343. R. Biek *et al.*, *Journal of Wildlife Diseases* **38**, 840–845, ISSN: 0090-3558, DOI 10.7589/0090-3558-38.4.840 (2002).
344. J. Soberon, A. T. Peterson, *Biodiversity Informatics* **2**, ISSN: 1546-9735, DOI 10.17161/bi.v2i0.4 (2005).
345. E. N. Pavlovsky, *Natural nidality of transmissible diseases with special reference to the landscape epidemiology of zooanthroponoses* (University of Illinois Press, Urbana, Illinois, 1966), ISBN: 0252727266.
346. R. M. May, *Theoretical Population Biology* **5**, 297–332, ISSN: 0040-5809, DOI 10.1016/0040-5809(74)90055-0 (1974).
347. L. E. Escobar *et al.*, *Ecology and Evolution* **8**, 4757–4770, ISSN: 2045-7758, DOI 10.1002/ece3.4014 (2018).
348. G. Varela, A. Vazquez, *Revista del Instituto de Salubridad y Enfermedades Tropicales* **14**, 219–223, ISSN: 0370-5781, 0370-5781 (1954).
349. S. Bhatt *et al.*, *Nature* **496**, 504, ISSN: 0028-0836, DOI 10.1038/nature12060 (2013).
350. C. J. Carlson *et al.*, *PLoS Neglected Tropical Diseases* **10**, e0004968, ISSN: 1935-2735, 1935-2727, DOI 10.1371/journal.pntd.0004968 (2016).
351. A. Aikimbayew *et al.*, *Przeglad Epidemiologiczny* **57**, 593–598, ISSN: 0033-2100 (2003).
352. H. V. Pham *et al.*, *International Journal of Epidemiology* **38**, 1634–1641, ISSN: 0300-5771, DOI 10.1093/ije/dyp244 (2009).
353. E. Giorgi *et al.*, *Spatial and Spatio-temporal Epidemiology* **19**, 125–135, ISSN: 1877-5845, DOI 10.1016/j.sste.2016.10.001 (2016).
354. M. X. Gao *et al.*, *Science China Earth Sciences* **53**, 8–15, ISSN: 1674-7313, DOI 10.1007/s11430-010-4122-9 (2010).

355. Q. Qian *et al.*, *BMC Infectious Disease* **14**, 382, ISSN: 1471-2334, 1471-2334, DOI 10.1186/1471-2334-14-382 (2014).
356. J. J. Wiens, C. H. Graham, *Annual Review of Ecology, Evolution, and Systematics* **36**, 519–539, ISSN: 1545-2069, 1545-2069, DOI 10.1146/annurev.ecolsys.36.102803.095431 (2005).
357. M. B. Ashcroft, *Journal of Biogeography* **37**, 1407–1413, ISSN: 0305-0270, DOI 10.1111/j.1365-2699.2010.02300.x (2010).
358. J. M. Mann *et al.*, *The Journal of Infectious Diseases* **140**, 397–401, ISSN: 0022-1899, DOI 10.1093/infdis/140.3.397 (1979).
359. M. Eidson *et al.*, *American Journal of Public Health* **78**, 1333–1335, ISSN: 0090-0036, 1541-0048, DOI 10.2105/AJPH.78.10.1333 (1988).
360. R. Aiken, *Fishing and hunting 1991-2001: avid, casual, and intermediate participation trends*, 2004, (https://digitalmedia.fws.gov/digital/collection/document/id/303/rec/4).
361. M. T. DeVivo *et al.*, *PLoS One* **12**, e0186512, ISSN: 1932-6203, DOI 10.1371/journal.pone.0186512 (2017).
362. D. R. Edmunds *et al.*, *PLoS One* **11**, e0161127, ISSN: 1932-6203, DOI 10.1371/journal.pone.0161127 (2016).
363. L. Waddell *et al.*, *Transboundary and Emerging Diseases* **65**, 37–49, ISSN: 1865-1674, DOI 10.1111/tbed.12612 (2018).
364. R. J. Eisen *et al.*, *Journal of Medical Entomology* **55**, 1133–1142, ISSN: 0022-2585, DOI 10.1093/jme/tjy060 (2018).
365. W. Bodmer, *Genetics* **199**, 267–279, ISSN: 0016-6731, 1943-2631, DOI 10.1534/genetics.114.173062 (2015).
366. Z. B. Abrams *et al.*, *BMC Genomics* **19**, 738, ISSN: 1471-2156, 1471-2156, DOI 10.1186/s12864-018-5093-z (2018).
367. C. G. Tsangarides, M. S. Qureshi, *World Development* **36**, 1261–1279, ISSN: 0305-750X, DOI 10.1016/j.worlddev.2007.06.019 (2008).
368. Y. Tian *et al.*, *Nature Communications* **8**, 1473, ISSN: 2041-1723, DOI 10.1038/s41467-017-01728-5 (2017).
369. T. P. Gilmour *et al.*, *Neuroscience Letters* **469**, 97–101, ISSN: 0304-3940, DOI 10.1016/j.neulet.2009.11.052 (2010).
370. P. Praus, *Applied Water Science* **8**, 167, ISSN: 2190-5495, 2190-5495, DOI 10.1007/s13201-018-0794-7 (2018).
371. S. Kim *et al.*, *Bioinformatics* **34**, 1321–1328, ISSN: 1367-4811, 1367-4803, DOI 0.1093/bioinformatics/btx765 (2017).