DISSERTATION

OMINOUS MOUSE: TRACING ZOONOTIC DISEASE RISK IN THE SOUTH CAUCASUS

Submitted by

Kendra E. Gilbertson

Graduate Degree Program in Ecology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Spring 2025

Doctoral Committee:

Advisor: Colleen Webb

Michael Kosoy
Helen Sofaer
Rebekah Kading

Copyright by Kendra E. Gilbertson 2025

All Rights Reserved

ABSTRACT

OMINOUS MOUSE: TRACING ZOONOTIC DISEASE RISK IN THE SOUTH CAUCASUS

Most emerging infectious diseases that affect humans originate in animal species. In order to accurately predict disease risk, we must understand the spatial distribution of the various environmental, ecological, and host factors that allow pathogens to spillover into human populations. The most foundational of these requirements is the presence of one or more host species that allow the pathogen to persist in a given location. We accomplished this by using a Bayesian Additive Regression Trees (BART) model to predict the distributions of 15 medically-relevant small mammals species across the South Caucasus while also exploring different methodological strategies. Our flexible machine learning approach allowed us to create predictions that had high ecological accuracy and strong discrimination ability crucial for discerning the likelihood of species presence. We also found that models trained on random background points were well suited for our goal of maximizing discrimination ability without sacrificing biological realism. We recommend the consideration of this approach to those using species distributions to predict disease risk. Next we incorporated these predictions as predictive layers in a risk prediction model of tularemia in order to understand the spatial risk and ecological drivers of this complex pathogen. We found that tularemia was primarily a rural disease in humans, partially a result of the increased contact between humans and animals in agricultural settings. While we observed different distributions between clinical manifestations of tularemia, the causes of these diverging patterns warrant further inquiry. We extended our examination of zoonotic risk to include six additional pathogens. To paint a picture of overall disease risk in the South Caucasus, we combined the predicted distributions of all diseases to create a disease richness map. This is of public health import as it identified areas of elevated transmission risk across pathogens. We also found that zoonoses in the South Caucasus typically occur in rural areas, and those with occupational exposures were at the most risk. These

areas are more likely to support higher diversity of host and vector species, allowing for circulation of pathogens in the environment and greater contact rates between humans and disease-carrying animals. Our research adds to a broader understanding of zoonotic risk and its ecological factors in the South Caucasus and we provided methodological recommendations relevant to the field of disease ecology.

ACKNOWLEDGEMENTS

Thank you, Colleen, for taking a chance on a student with little statistical and even less ecological background; I'm sure you've had second thoughts over the years. I appreciate your guidance, support, and, most importantly, challenging me to be the scientist I knew I could be.

I would like to thank my committee members for their time and feedback throughout this process. Helen, thank you for your insight on SDM methodology, and Michael, your depth of knowledge is something I aspire to. Thank you to the entire Atlas Project team, including Lile Malania, Andrei Kandaurov, Ioseb Natradze, Nana Bolashvili, Irma Burjanadze, Tengiz Chaligava, Zviad Asanishvili, and many others.

I am deeply grateful to the Webb lab for their encouragement and support, both in and out of the lab. Lindsay, Clint, Ben, Brooke, Catherine, Sam, Jonathan, and Christopher I cannot imagine a better group to spend the past five and a half years with. Lindsay, thank you especially for your patience and mentorship.

Thank you, most of all, to my family, who first taught me the value of education-perhaps too well. To my parents, thank you for always being proud of me and thinking whatever I was working on was the most interesting topic in the world. And special thanks to my dad for the coding help. Brett, I can't imagine this is what you had in mind when we met! Thank you for sticking by me through the hardest thing I've ever done, for always believing in me, and for agreeing to get another puppy. I promise you all I'm done with school.

The work was made possible by support provided by the US Defense Threat Reduction Agency (CBR/GG-27 project) through the Cooperative Biological Engagement Program in Georgia and Azerbaijan. The findings, opinions, and views expressed herein belong to the author and do not reflect an official position of the Department of the Army, Department of Defense of the US Government or any other organization listed.

DEDICATION

For Doe and Kit, who were with me from beginning to end.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

Zoonotic diseases, which spread between animals and humans, are a significant threat to global health and economic stability, with approximately 75% of emerging infectious diseases originating in animal species [1]. These diseases impact society through direct healthcare costs, lost productivity, and reduced livestock production, but the true global burden is unknown [2,3]. The One Health framework recognizes this web of interactions, emphasizing that human well-being is intertwined with both animal and environmental health. This perspective is crucial for understanding and preventing zoonotic spillover events. Despite extensive research across multiple disciplines, identifying the precise determinants of spillover remains challenging due to the complex and changing nature of multi-host pathogen systems and the varying influences of landscape characteristics and human activities across space and time [4]. Critical questions persist regarding the specific pathways and environmental conditions that enable pathogens to move from wildlife and domestic animals into human populations, and how the complex interplay of ecological and environmental factors shapes this transmission process.

Successful transmission of zoonotic pathogens from animals to humans requires alignment of multiple conditions: the concurrent presence of the pathogen and its reservoir host, human exposure to the infectious agent and human susceptibility to infection [5]. Each of these components varies across landscapes, creating a complex mosaic of disease risk that operates at multiple biological and spatial scales. This heterogeneity manifests from the individual level—where variations in behavior and immune response affect exposure and susceptibility—to the species level, where different hosts and vectors show varying competence for onward transmission [6]. The physical landscape itself plays a role in shaping the risk of disease through habitat configuration, connectivity between populations, and patterns of land use [7]. Understanding how these factors interact and vary across space is crucial for identifying potential transmission hotspots and developing targeted intervention strategies. By mapping and analyzing these heterogeneity patterns, we can

better predict where and when spillover events are most likely to occur, enabling more effective disease surveillance and prevention efforts.

The South Caucasus region, which includes Georgia, Armenia and Azerbaijan, represents a unique system for studying the transmission of zoonotic diseases. As one of the world's few temperate biodiversity hotspots, its mosaic landscape supports an array of wildlife, including small mammals and wild ungulates, alongside substantial livestock populations that have great cultural and financial importance [8, 9]. This ecological richness, combined with the region's heavily agricultural nature, creates numerous interfaces for pathogen transmission between wildlife, domestic animals, and humans. The region is endemic for multiple zoonoses of significant public health concern, including anthrax, brucellosis, Crimean-Congo hemorrhagic fever (CCHF), leptospirosis, tick-borne encephalitis (TBE), and tularemia [9]. These pathogens persist through complex transmission cycles that involve various wildlife reservoirs and vectors, including small mammals, ticks, and fleas. While previous research has often focused on individual countries [10–15], examining the South Caucasus as a region is advantageous for understanding disease patterns, as these countries share geographic features, habitat types, and ecological characteristics. This regional approach provides a more comprehensive picture of disease dynamics and enhances our ability to predict and manage zoonotic risk across this ecologically complex landscape.

Understanding disease dynamics in the South Caucasus' complex ecological landscape requires us to first examine the foundation of zoonotic transmission systems: the presence of reservoir host species. These species form the foundation of disease transmission systems. In this context, small mammals—particularly rodents—act as important reservoir hosts worldwide. Rodents harbor the highest diversity of zoonotic viruses among all mammalian orders [16], with approximately 11% of rodent species known to host zoonotic pathogens and 85 distinct zoonotic diseases identified within this order [17, 18]. These species inhabit both built and natural environments, interacting with wild and domestic animals and sharing landscapes with human populations. In Chapter 2, we build the first block in understanding disease risk heterogeneity by developing predictive spatial distribution models for 15 medically-relevant small mammal species across the South Caucasus. To

our knowledge, this is the first time host species distribution maps have been created for the region. Through this analysis, we not only map host distributions but also advance species distribution modeling methodology specifically for applications in disease ecology.

While understanding host distributions provides the foundation, we must also consider how environmental factors, host dynamics, and human behavior intersect to create spillover opportunities. This complexity is particularly evident in tularemia, a zoonotic disease caused by the highly infectious bacteria *Francisella tularensis* [19]. In the South Caucasus, Type B (*F. tularensis* subsp. *holarctica*) and *F. tularensis* subsp. *mediasiatica* persist through complex transmission cycles involving both terrestrial and aquatic systems [20]. The pathogen's ecology is intricate as it can infect numerous animal species, with *Apodemus* sp., *Microtus socialis*, and *Microtus arvalis* serving as primary reservoirs in the region [9]. Ticks and fleas act as vectors, with ticks capable of maintaining the bacteria for up to three years [21]. Human infection occurs through multiple pathways, including direct contact with infected animals, vector bites, contaminated water exposure, and inhalation of contaminated particles [21–23]. These diverse transmission routes result in various clinical manifestations, from ulceroglandular and glandular forms typically caused by physical contact, to oropharyngeal infections from ingesting contaminated water, to pneumonic cases from inhaled particles [24, 25].

To unravel the complex spatial patterns of tularemia in the South Caucasus, we developed Bayesian risk models for human cases and host-vector occurrence (Chapter 3). Our modeling framework combined two key sets of predictors: biological covariates that captured the ecological dynamics of hosts and vectors, and geographic covariates that represented the underlying landscape features and bioclimatic conditions. This dual approach allowed us to examine how environmental and ecological factors shape disease risk across the region. We further explored the spatial patterns of different clinical manifestations, providing insight into how various transmission pathways are influenced by specific environmental and biological conditions. By synthesizing these spatial predictions with patterns of human infection across different demographic groups, we added to the understanding of tularemia risk in the South Caucasus.

While tularemia illustrates the complex interplay of ecological factors in disease transmission, we expanded our scope to examine patterns of risk across seven zoonoses simultaneously. The spatial heterogeneity of zoonotic disease risk arises from complex interactions between hosts, vectors, and human populations across landscapes. Of particular interest are transitional zones between ecological regions, or ecotones, which can have elevated pathogen transmission due to increased tick densities and higher rates of human-animal contact [7, 26]. In the South Caucasus, these dynamics may be especially pronounced in transition zones between built environments and agricultural lands, where occupational exposure through farming and other similar activities facilitates contact with infectious animals. However, human hotspots may occur both in agricultural areas with high occupational exposure and in urban areas where better healthcare access increases the likelihood of detection and diagnosis. Understanding these patterns across multiple diseases can reveal both pathogen-specific transmission drivers and common factors that facilitate spillover more broadly.

To investigate these concepts, we developed Bayesian models predicting both human and animal risk for seven zoonotic diseases across the South Caucasus (Chapter 4). By comparing human and animal risk predictions, we identified areas of increased disease richness and examined why some areas with high animal risk show limited human cases. We also examined the role of occupational exposure in agricultural settings and the influence of species distributions across different habitat types, specifically examining how transitional areas affected transmission patterns. These findings have important implications for public health efforts, enabling more targeted surveillance and control measures while optimizing resource allocation for shared risk factors across diseases.

In summary, our work provides an additional step in understanding the spatial dynamics of zoonotic disease risk in the South Caucasus, a region of unique ecological, cultural, and public health significance. These efforts enhance our understanding of the connection between hosts, vectors, and human populations across diverse landscapes. This not only advances disease ecology research for several neglected diseases, but also offers practical tools for targeted surveillance, prevention, and control efforts. By identifying high-risk areas and highlighting the shared drivers

of zoonotic spillover, our work supports a One Health approach to managing zoonotic diseases, ultimately contributing to the broader goal of mitigating their impact on human and animal health worldwide.

Chapter 2

Predicting Species Distributions for Small Mammal Host Species in the South Caucasus

2.1 Introduction

Small mammals, particularly rodents, commonly act as hosts whose presence is the foundational requirement for spillover of zoonotic pathogens from animals to humans [5, 27]. Of all mammalian orders, rodents host the highest number zoonotic viruses [16]; approximately 11% of rodent species are hosts for zoonotic disease, and 85 diseases have been identified in this order [17, 18]. As reservoir species are required to maintain and spread pathogens in the environment, understanding their geographic distribution is a key aspect to predicting where transmission to human populations may occur. These species interact with other animals and humans across shared landscapes, linking environment, animal, and human health in the One Health triad.

Accurate species distributions are essential, but methodologically complex [28]. Most species distribution models (SDMs) estimate a species' niche from its geographical distribution and project these estimates across the area of interest [29]. The reliance on observed data is inevitably subject to multiple sources of uncertainty from species misidentification, sampling bias, and imperfect detection. Since many datasets contain only presence points, models commonly incorporate background points for training. While random background points reflect the full environmental range, they can capture sampling bias instead of true habitat suitability. Newer methods focus on reducing sampling bias to avoid this problem [30–32]. One approach uses presence points of similar species as target-group background points, assuming consistent sampling bias across groups [33]. The choice of method should align with research objectives. For landscape risk assessment, models must accurately predict across unsampled areas and the full environmental range. It also requires accurate model discrimination, which may be of less importance in other fields.

These methodological considerations are particularly relevant in regions like the South Caucasus, where ecological and zoonotic complexity converge. The South Caucasus, consisting of Georgia, Armenia, and Azerbaijan, forms a unique ecological zone with a number of endemic zoonoses. Zoonotic pathogens are maintained in the environment and spread to humans through diverse transmission pathways involving small mammals, wild ungulates, fleas, ticks, and livestock [22,34,35]. It is one of the world's few temperate biodiversity hotspots, with a highly mosaic landscape [8,9]. This diversity of habitats supports a large number of small mammal species, several of which are unique to the region, which may increase the potential for spillover. Understanding the unique composition and spatial distribution of the South Caucasus' small mammal species will assist in evaluating spatial disease risk across the region.

The high species diversity in the South Caucasus also allows us to examine the affect of biodiversity on pathogen transmission. Community composition may impact transmission by either amplifying or diluting pathogens within small mammal populations. The dilution effect hypothesis suggests that increased biodiversity reduces transmission through mechanisms such as decreased host density, reduced encounter rates between competent hosts, and the presence of dead-end hosts [36]. Conversely, a greater number of host species may amplify prevalence by increasing the likelihood that a competent host will come into contact with a carrier of said pathogen. However, empirical evidence for this effect is mixed, with some studies suggesting that the specific combination of species in an environment, rather than species richness alone, drives disease spillover [37]. The heterogeneity of the South Caucasus landscape allows us to examine these questions as it supports a large number of small mammal species, many of which are known hosts, and the landscape heterogeneity results in areas of both high and low species richness.

We estimated the distribution of 15 medically relevant small mammal species across the South Caucasus and examined their spatial patterns to identify areas at risk for zoonotic pathogens. Given the competing advantages of random and target-group background points, we tested both to find the prediction with the best discrimination ability. Model training allowed us to identify how environmental variables correlated with the spatial distribution of our species of interest. This

can provide information as to what climate or biological covariates are useful predictors across species. To examine the relationship between community diversity and pathogen prevalence, we compared species richness to richness of mammals positive for *Yersinia pestis* and *Francisella tularensis*—two pathogens strongly linked to small mammals reservoirs. This allowed us to assess how variations in small mammal community composition may influence the spatial distribution of pathogens in the South Caucasus.

2.2 Methods

Our species distribution models, trained on a biologically-appropriate set of geographic and climatic covariates, predicted the relative likelihood of species occurrence across the Caucasus. The covariates were chosen to capture key environmental factors that support or limit species presence, enabling us to explore how specific ecological conditions shape the spatial distribution of small mammal reservoirs. By modeling each species' distribution across space, we identified and compared regions where they overlap, illustrating areas of high small mammal biodiversity. These biodiversity hotspots not only highlight areas of increased host richness, but also provided insight into spatial patterns that may influence pathogen spread via amplification or dilution. These risk predictions act as the foundational layers for estimating pathogen risk in space by satisfying the spillover requirement of host species presence. Our maps, combined with spatial representations of additional spillover requirements, can be used to illustrate heterogeneity in disease risk across space and inform public health intervention measures.

2.2.1 Species distribution models

In this study, we employed Bayesian Additive Regression Trees (BART) to explore the likelihood of species presence across the landscape. BART is a powerful technique that combines machine learning and Bayesian inference into a nonparametric regression ensemble [38]. Unlike traditional regression approaches, BART does not assume linear relationships between predictors

and outcomes, making it particularly suitable for modeling ecological relationships where linearity is rare.

The fundamental idea of BART is to decompose the relationship between predictors and outcomes into multiple simple decision trees. Each tree in the ensemble acts as a "weak learner," explaining a portion of the relationship while remaining relatively shallow. The model incorporates three key prior distributions for the probability of a tree terminating at a node of given depth, the likelihood of selecting specific variables for splitting rules, and the probability of splitting a variable at particular values [39]. The tree depth prior follows a negative power law, which penalizes deep trees and overfitting.

Model construction begins with a set of single node decision trees. Through subsequent iterations of Markov Chain Monte Carlo (MCMC) sampling, the trees add or eliminate splits and restructure decision paths. The iterations are combined into a sum-of-trees model, with each tree adapted to the remaining residuals. All sum-of-trees models retained after burn-in are combined into the posterior distribution. Variable selection is performed automatically using 100 models per iteration of stepwise reduction with ten trees per model to maximize the difference between informative and uninformative covariates. The importance of each variable is measured as the proportion of total branches each variable used. After prediction, the Bayesian approach allows us to explicitly measure uncertainty by extracting the 95% credible interval from the prediction posterior distribution, providing spatially explicit confidence measures for our predictions. This direct quantification of uncertainty is an improvement over traditional methods like MaxEnt, which approximate uncertainty through repeated model runs. [38, 39]

The model's ability to handle complex, nonlinear relationships while avoiding overfitting makes it particularly valuable for our study region, where we predict to unsampled areas due to geographic and political constraints. The explicit uncertainty quantification is important for both interpreting results and planning future surveillance efforts, allowing us to identify areas where additional sampling would be most beneficial. We predicted the species distribution of each medically-relevant small mammal species using the `embarcadero` package.

We also assessed both random background points, with their stronger ability to predict across unsampled space, and target-group background points, which better account for sampling bias. The target-group background points were selected randomly from the remaining 14 species. We selected the same number of background points as presence points for our species of interest for both model types.

To evaluate model fit and select the most accurate predictions, we used both the Boyce index and the area under the curve (AUC) [40]. AUC measures the model's ability to correctly discriminate between locations where the species is present and where it is absent (or where background points are used as a proxy for absence). The AUC is derived from the Receiver Operating Characteristic (ROC) curve, which plots sensitivity and specificity at each threshold value [41]. A value of 0.5 indicated performance equal to random chance, and a value of 1.0 indicated perfect discrimination. Strong discrimination is crucial in disease ecology, as the primary goal is to determine whether a species is present to serve as a suitable host for a pathogen.

Though it is widely used and easy to interpret, AUC has several limitations. A model may have a strong AUC score, while vastly under or overestimating species presence, as long as the presence had a higher estimated probability than the background point [41]. This is an important drawback, as we are interested in the likelihood of species presence as well as accurate discrimination, and discrimination is typically uninformative or misleading with regards to the biological accuracy [42]. Additionally, the AUC score is affected by the type of background points used. Random background points are usually easier to discern from presence points as they occur in areas unsuitable for a given species' survival. Target-group background points are typically drawn from areas that are environmentally similar to the presence data. This makes discrimination inherently more difficult for target-group models compared to random background point models [33].

To fairly compare the predictive performance of models trained using different background point types, we also calculated the out-of-sample AUC using an independent set of random background points for evaluation. This approach ensured that the evaluation was not biased by the background data used during model training. Specifically, we extracted predicted values for pres-

ence points and a new, randomly generated, set of background points from both models. AUC was calculated for each model based on its ability to discriminate between presence and the new random background points, providing an objective comparison of their predictive performance. We used DeLong's test to measure whether the difference between AUC values was significant.

We also included a presence-only evaluation metric that used real data only and provided a measure of biological accuracy from our predictions. The Boyce index, which was developed to assess model predictions using presence data, complements AUC by evaluating the model's ability to correctly rank habitat suitability over space [40, 43]. This measure is calculated using the ratio of the predicted frequency of points within each habitat suitability bin to the expected frequency in that bin. The Boyce index is the Spearman rank correlation between this ratio and the suitability class, so as suitability increases so too should the predicted-to-expected ratio. This method evaluates how well the model ranks suitable areas across the species' sampled occurrence range, making it useful for assessing ecological realism in habitat suitability predictions.

Using these metrics in tandem provided a more comprehensive assessment of model performance by balancing overall discrimination accuracy (AUC) with the model's ecological fit (Boyce index). By employing AUC, we assessed the ability of each model to discern presence from background points, which is a key aspect of our study question. We balanced the shortcoming of AUC, namely the lack of biological accuracy, by including a ranking metric for continuous estimates. By doing so, we also addressed the bias toward random background points that has been observed using AUC [33].

Moving beyond individual model fit, assessed how similar the predictions resulting from models trained on different absence points were. While model fit is informative, we were ultimately interested in how the use of these different point types affected the final product: the species prediction. To do this, we employed Schoener's D, which measures niche overlap between two probability distributions in geographic space [44]. This metric ranges from 0 (no overlap) to 1 (complete overlap), allowing us to quantify how similar different SDM predictions are. While the performance metrics measure how well each model performs, assessing the degree of predic-

tion overlap gives us a clearer picture of the impact our different background points have on the resulting predicted distributions.

2.2.2 Small mammal data

Transect sampling of all small mammal species was conducted several times per decade and combined with data from publications, museum and university collections, and other sources. A large portion of our data was the result of fieldwork where species and location of all animals found at a given site were recorded. Coordinates were sourced from either the literature, maps, or GPS coordinates taken at the field sampling sites. Data from 1927 to 2021 were available in Armenia. Of the 6012 observations from all small mammals species collected there, 28.2% were from field data, and 71.8% came from published literature. A total of 1637 records were available from Azerbaijan from a period of 1902 to 2021. Their data were 37.5% field data, 50.5% publications, and 11.9% collections or other sources. In Georgia, there were 5767 observations from all small mammal species from years 1855 to 2021. Sixteen percent of the observations were collected from field data, 49.9% were collected from published literature, 34.4% from museum and other collections.

Of the 61 total small mammal species collected during the study period, 15 had a documented or suspected role in zoonotic disease systems. The association between each of these species and zoonotic pathogens is summarized in Table 2.1. The full list of species collected is located in Appendix A, and additional information on the medically-relevant small mammal species can be found in Kosoy (2024).

2.2.3 Covariate layers

To ensure consistency temporally and in measurement across countries, we used global, publicly available covariate layers for model prediction. We selected 19 bioclimatic layers and elevation from WorldClim Version 2 (Appendix B) that provided globally standardized grid data averaged from 1970 to 2000 at 30 second spatial resolution [45]. These layers represent tempera-

Table 2.1: Summary of South Caucasus medically-relevant small mammal species and their associated pathogens or diseases.

Species	Select Pathogens	Source
Apodemus uralensis	*Brucella*, Crimean-Congo hemorrhagic fever (CCHF), *Francisella tularensis*, *Leptospira*, tick-borne encephalitis (TBE)	[9, 48, 49]
Apodemus witherbyi	*Brucella*, *F. tularensis*, *Leptospira*, TBE	[9, 48, 49]
Arvicola amphibius	*Brucella*, *F. tularensis*, *Yersinia pestis*, *Leptospira*, TBE	[9, 14]
Erinaceus concolor	CCHF, *F. tularensis*, *Leptospira*	[35, 50]
Lepus europaeus	*Bacillus anthracis*, *F. tularensis*	[35, 49, 51]
Meriones libycus	*B. anthracis*, *F. tularensis*, *Leptospira*, *Y. pestis*	[9, 35, 48, 52–54]
Meriones tristrami	*F. tularensis*	[9]
Microtus arvalis	*B. anthracis*, *F. tularensis*, *Leptospira*, *Y. pestis*	[35, 55]
Microtus socialis	*F. tularensis*, *Leptospira*, *Y. pestis*	[35, 35, 48]
Mus musculus	CCHF, *F. tularensis*, *Leptospira*, *Mycobacterium tuberculosis*	[48, 49]
Nothocricetulus migratorius	*F. tularensis*, *Leptospira*, *Y. pestis*	[35, 56]
Rattus norvegicus	*Leishmania*, *Leptospira*, *Toxoplasmosis gondii*	[48, 57, 58]
Rattus rattus	*Brucella*, *F. tularensis*, *Leptospira*, *Y. pestis*	[35]
Scarturus elater	*Leptospira*, *Y. pestis*	[9, 35]
Scarturus williamsi	CCHF	[59]

ture and precipitation patterns that can affect species range [46,47] and dimensions of fundamental niche space.

The remaining 14 layers were chosen from the `geodata` package that aggregates sources for global spatial datasets (Appenix B). These layers were selected to represent key environmental and anthropogenic factors likely to impact small mammal habitats. All layers were masked to the South Caucasus region, and those not already within a 0-1 range were scaled.

In order to identify useful predictors of species presence, we examined how often each covariate was selected across all species distribution models. We also compared the variable importance of the covariates generated by the BART model. This may provide information as to which covariates are useful predictors in the South Caucasus or in small mammal SDMs.

2.2.4 Species and disease richness

To assess whether there was support for the dilution effect in our system, we measured the correlation between both medically-relevant species richness and all small mammal species richness with richness of mammal infections. This gave us a measure of how correlated areas of high and low biodiversity were to areas with high and low disease occurrence. A strong negative correlation between species richness and disease richness lends support to the dilution effect, particularly if the correlation for all small mammal species richness is lower than the correlation between medically-relevant species richness. A strong positive correlation would support the amplification effect by indicating higher disease presence in areas of high biodiversity. This also allowed us to evaluate the impact of non-host species on disease prevalence.

We calculated species richness for medically relevant small mammal species and all small mammal species by counting the number of unique species that occurred in each grid cell imposed across the region. To reduce noise between cells, we smoothed the map using a queen matrix. In this method, each cell takes the value of the average of each of its boundary and corner neighbors, generating a more gradual and realistic species richness gradient.

Using the same richness calculation method, we identified pathogen richness using separate datasets of positive mammal samples for *Y. pestis* and *F. tularensis*; two pathogens at least partially maintained in the environment by small mammals. This broader dataset included small mammals, larger wildlife, livestock, and domestic animals.

We measured the correlation between species richness and disease richness by rasterizing the estimates onto a common grid and computing the Pearsons's correlation coefficient between them. We used Dutilleul's modified t-test to account for spatial autocorrelation when assessing the significance of these correlations [60].

2.3 Results

Nearly all models performed well with both background types on both AUC and Boyce measures of fit (Table 2.2). Using the AUC cutoff of 0.8, all random background models had good discrimination, while two target-group models had fair to poor performance [61]. While in-sample AUC values were overall very similar, the random background models slimly outperformed target-group background models in all but one species. This pattern was more apparent in the out-of-sample AUC values, where the differences were larger. Twelve random background models had significantly higher out-of-sample AUC values compared to one target-group (p-value < 0.01). As expected, in-sample AUC values were nearly always higher than out-of-sample AUC values.

The models performed well using the Boyce metric as well, with the worst performing model scoring 0.67, but all others measuring 0.94 to 1 [62]. Of our 15 models, random background performed marginally better 12 times based on the Boyce index. The overlap between predicted species distributions from different background point models, measured by Schoener's D, ranged from 0.69 to 0.82. The majority of values fell between 0.70 and 0.75, indicating substantial overlap in predicted distributions between models. Notably, the areas of elevated likelihood were larger in the target-group predictions than the random background predictions (not pictured).

To select our final models for each species, we balanced the fit from both metrics. As the out-of-sample AUC measured each prediction on the same background points, we deferred to

this method of measuring discrimination ability over in-sample AUC values. In cases were AUC differed greatly between models (≥ 0.5), we selected the model that performed best on that metric. When AUC was similar between between models we used the model with the highest Boyce index to preserve biological accuracy in our predictions. However, we expect the final selection to have limited impact for most species given the high correlation between spatial predictions.

Building on this selection process, we examined the covariates associated with species prediction and identified key environmental characteristics associated with species presence. No one covariate appeared in all of our final models. The most commonly selected covariates were wind speed (ten models), and elevation, built environment, and human footprint (nine models each). Isothermality, temperature annual range, and precipitation of warmest quarter were selected eight times, and all other covariates were selected less often. Only four covariates; wetlands, snow, shrubs, and moss were not selected in any model. The full results are available in Appendix C.

The patterns in covariate selection provided important insights into the predicted distributions of individual species. Based on our predictions in Figure 2.1 and their selected covariates, we found that *Apodemus witherbyi* (n=360), *Erinaceus concolor* (n=166), *Lepus europeaus* (n=770), *Mus musculus* (n=642), *Rattus norvegicus* (n=378), and *Rattus rattus* (n=270) exhibited the strongest affinities for built environments. Among these, *E.concolor* and *R. rattus* also appeared to inhabit forests and forest edges. In contrast, *Microtus arvalis* (n=1369) and *Meriones tristrami* (n=122) had dissimilar distributions but both appeared to avoid forested areas. *Arvicola amphibius* (n=432) also avoided forests and was the only species whose model selected water as a covariate and showed a preference for habitats along lakes and rivers. The habitats of the remaining species were more difficult to characterize. *Meriones libycus* (n=302) and *Nothocricetulus migratorius* (n=809) showed a preference for bare areas alongside other landscape types, and *Apodemus uralensis* (n=654) was more likely to occur in forested areas. Finally, *Scarturus elater* (n=184), *Microtus socialis* (n=717), and *Scarturus williamsi* (n=163) were more idosynchratic. Multiple equally important covariates were combined, making it difficult to infer simplified relationships between covariates and outcome.

Table 2.2: Boyce and AUC model evaluation metrics for species distribution models using random and target-group background points. Schoener's distance measures pairwise spatial overlap between species predictions using each model type (0 = no overlap, 1 = complete overlap). Light gray indicates worse performance using natural cut points in the Boyce and AUC values across models.

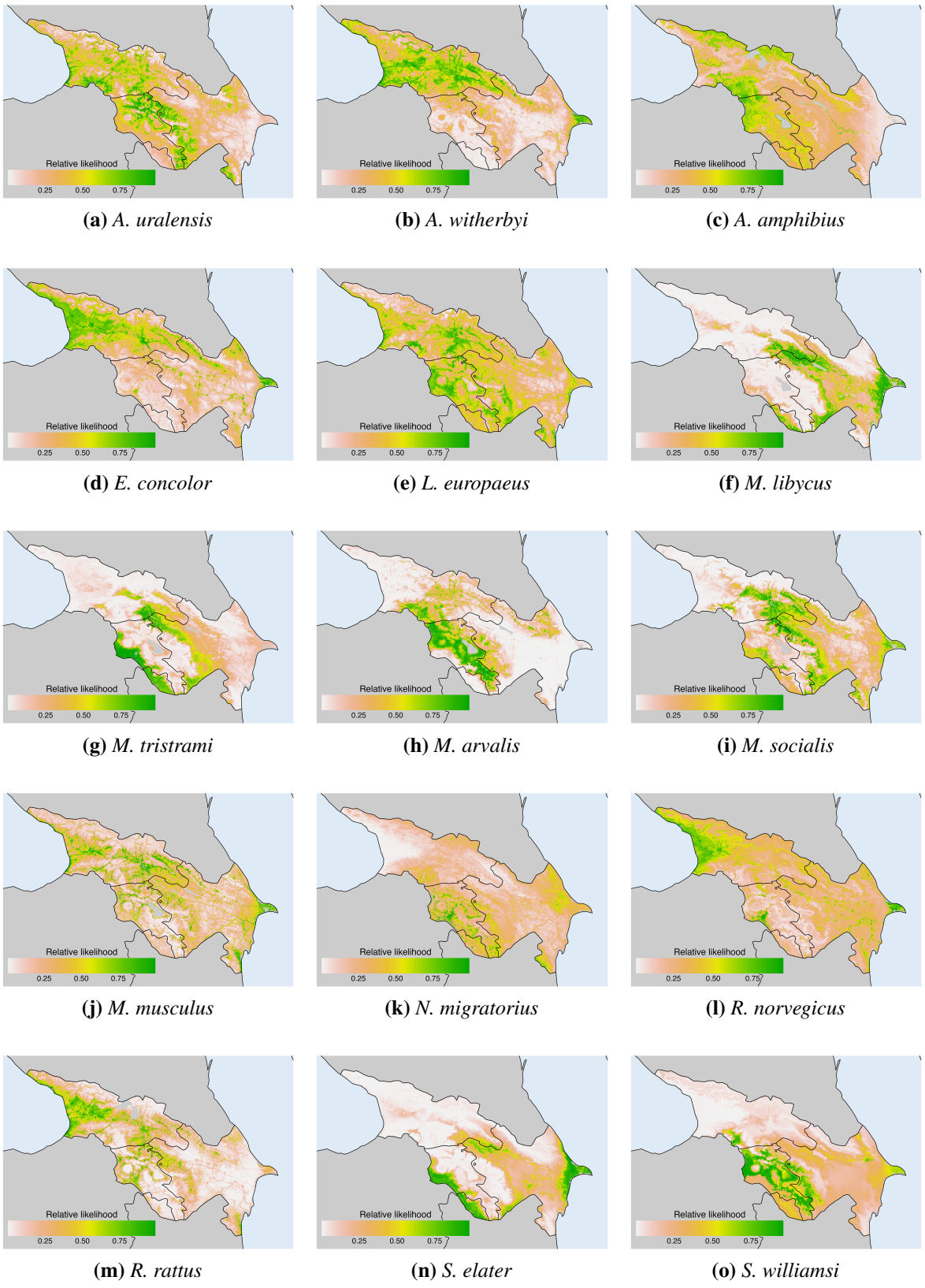
	Random background			Target-group background			Schoener's D
Species	Boyce	In-sample AUC	Out-of-sample AUC	Boyce	In-sample AUC	Out-of-sample AUC	
A. amphibius	0.99	0.95	0.93	1.00	0.85	0.86	0.75
A. uralensis	1.00	0.89	0.86	0.96	0.85	0.72	0.73
A. witherbyi	1.00	0.94	0.92	0.94	0.91	0.80	0.81
E. concolor	0.99	0.95	0.92	0.98	0.94	0.85	0.78
L. europaeus	1.00	0.84	0.81	0.67	0.78	0.57	0.76
M. libycus	0.96	0.95	0.95	0.95	0.99	0.95	0.82
M. tristrami	0.98	0.99	0.95	0.97	0.98	0.95	0.73
M. arvalis	1.00	0.96	0.95	0.96	0.93	0.89	0.72
M. musculus	1.00	0.92	0.90	0.98	0.83	0.76	0.72
M. socialis	1.00	0.93	0.93	1.00	0.90	0.88	0.76
N. migratorius	0.99	0.96	0.95	1.00	0.91	0.93	0.70
R. norvegicus	0.99	0.96	0.93	1.00	0.87	0.89	0.69
R. rattus	0.99	0.98	0.96	0.97	0.95	0.92	0.73
S. elater	0.99	0.98	0.97	0.96	0.97	0.94	0.74
S. williamsi	0.98	0.98	0.95	0.97	0.96	0.97	0.75

Understanding these habitat associations provides valuable context for exploring broader ecological dynamics, including the potential relationship between small mammal species richness and disease risk. To evaluate evidence of the the dilution effect, we calculated both species richness (Figure 2.2a and 2.2b) and disease richness (Figure 2.2c) across the region. Small mammal species richness was greater in coastal Azerbaijan and in central and eastern Georgia for both medically-relevant small mammals species and all small mammal species. However, there did appear to be a greater concentration of medically-relevant small mammals in and around Baku, Azerbaijan than small mammals overall. Richness of *Y. pestis* and *F. tularensis* infection in all mammals appeared sporadically in Azerbaijan, southeast Georgia, and through central Armenia. When accounting for spatial autocorrelation, we found a positive trend between medically-relevant species richness and disease richness (r = 0.49, $p = 0.08$), and a marginally weaker correlation when all small mammal species were included (r = 0.43, $p = 0.21$).

2.4 Discussion

Using a novel multi-decade dataset from three countries, we generated regional species distribution predictions to examine both disease-biodiversity relationships and methodological questions in species distribution modeling (SDM). Our findings show that methodological choices affect species predictions, suggesting approaches should align with specific research objectives. Though focused on prediction, our analysis revealed correlations between small mammal distributions, species richness, and disease presence.

Contrary to our expectations, model predictions remained relatively stable across different background point types. This stability likely stems from our large dataset, which, despite not being a random sample, reduced uncertainty compared to smaller sample sizes. Random background points produced predictions with marginally but significantly better discrimination ability than target-group points, as measured by out-of-sample AUC. We attribute this to the target-group model's training on a limited subset of environmental conditions within the study range, constraining its ability to generalize to unsampled areas. Target-group models predicted broader areas of high likelihood, likely because the smaller environmental contrast between presence and back-



(a) *A. uralensis* **(b)** *A. witherbyi* **(c)** *A. amphibius*

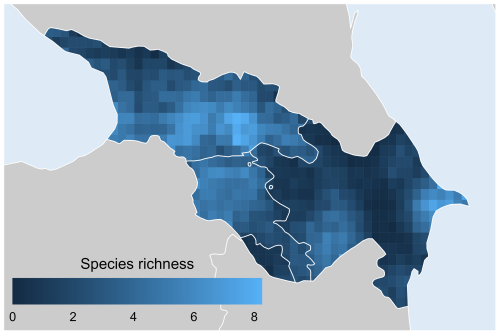
(d) *E. concolor* **(e)** *L. europaeus* **(f)** *M. libycus*

(g) *M. tristrami* **(h)** *M. arvalis* **(i)** *M. socialis*

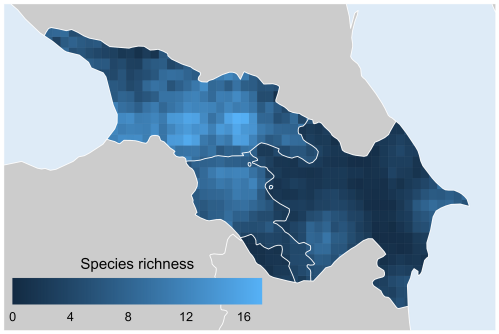
(j) *M. musculus* **(k)** *N. migratorius* **(l)** *R. norvegicus*

(m) *R. rattus* **(n)** *S. elater* **(o)** *S. williamsi*

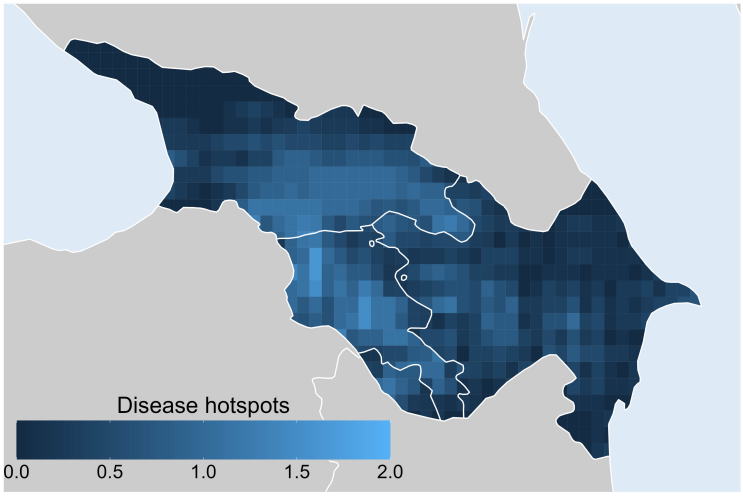
Figure 2.1: Predicted medically-relevant small mammal species distributions across the South Caucasus.



(a) Medically relevant small mammal species



(b) All small mammal species



(c) Disease richness in all mammals

Figure 2.2: Small mammal species richness and disease richness in all mammals across the South Caucasus.

ground points made the model less selective. In contrast, random background models, trained on a wider range of environmental conditions, appeared more conservative in their predictions by better identifying unsuitable areas. This difference in selectivity may explain the target-group models' lower out-of-sample AUC values. Random background models also performed better on the Boyce index, addressing concerns about AUC's ability to evaluate ecological relevance.

Given these findings, we recommend considering a random background approach for use in disease ecology. While both methods produced similar predictions based on their overlap, random background points offered superior discrimination ability—a priority when determining host presence. In our case, this approach demonstrated better generalizability to unsampled areas while maintaining biological accuracy. Beyond its predictive advantages, the random background method is practical: it requires no additional species data, is easily implemented, and encompasses the study area's full environmental range.

We also recommend using out-of-sample AUC over traditional in-sample AUC to evaluate discrimination ability. In-sample AUC consistently inflated performance metrics, as expected when testing models on their training data. Out-of-sample AUC, which tests predictions against new background points, provides a more reliable assessment and ensures fair comparison between models of different background types. Surprisingly, target-group models showed a larger increase from out-of-sample to in-sample AUC than random background models. This suggests that the reported advantage of random background models for in-sample AUC [33] was outweighed by the difficulty target-group models had discerning random background points from presence points during out-of-sample AUC. Out-of-sample AUC also better reflects real-world disease ecology applications, where it may be necessary to predict host presence from disease case locations. For our research questions, this discrimination-focused metric proved more relevant than the Boyce index's habitat suitability measure.

For evaluating host and vector distribution predictions in disease ecology, we recommend a dual-metric approach. First, use out-of-sample AUC to assess the model's ability to differentiate between species presence and absence. This should be used in conjunction with a measure

of biological accuracy, such as Boyce index, to verify ecological realism in predicted gradients. Although our models showed similar performance across both metrics, this consistency is not universal [42]. This combined evaluation approach provides a comprehensive assessment of SDM performance while aligning with research priorities in the field of disease ecology.

Effective SDMs require both robust evaluation metrics and ecologically relevant covariates. Our models highlighted that predictive accuracy depends on capturing both large-scale and fine-scale processes, emphasizing the role of scale and heterogeneity in covariate selection. Key large-scale variables included wind speed and elevation, while the built environment and human footprint represented fine-scale mosaic patterns. This emphasizes the importance of considering heterogeneity, as well as appropriate scale and biological relevance, in choosing environmental covariates. This mix of scales reflects ecological reality, where observed patterns may result from processes operating at different scales than the patterns themselves [63]. We recommend selecting covariates based on biological relevance across multiple scales and levels of heterogeneity. When combined in flexible models, these variables can interact as they do in natural systems [63], producing geographically accurate predictions. The BART model, though relatively new to ecology, proved particularly effective at handling these complex, non-linear relationships across heterogeneous landscapes. Its tree-based approach, combined with explicit spatial uncertainty measures, makes it well-suited for predictive SDMs, especially when creating fine-scale predictions across varied environments.

Our SDMs provide spatial detail previously unavailable in species distribution maps for South Caucasus' small mammals. While IUCN Red List maps show broad Extent of Occurrence using continuous boundaries around known occurrences [64], our predictions reveal more nuanced distribution patterns. For instance, we identified suitable habitat for *L. europaeus* and *M. tristrami* adjacent to their IUCN-reported ranges [65, 66], and found evidence that *M. libycus* and *M. schidlovskii* may occur outside their documented ranges [67, 68]. Though broad-scale ranges may suit certain conservation purposes, they are inadequate for disease ecology, where infection processes operate at fine scales [69].

The results of our species predictions reveal interesting patterns, particularly regarding habitat preferences and human associations. Most notably, we identified a distinct group of species strongly associated with human-modified landscapes, including cosmopolitan commensal rodents (*M. musculus*, *R. norvegicus*, *R. rattus*) as well as *A. witherbyi*, *E. concolor*, and *L. europaeus*. *A. witherbyi* and *L. europaeus* had broader distributions extending further beyond human settlements compared to *M. musculus*, suggesting a wider variety of suitable habitat. These findings reinforced the well-documented adaptive success of commensal rodents in human environments and highlighted other species known to exploit these habitats in the South Caucasus [9]. The preference of *R. rattus* for both built environments and forests aligns with previous research showing their ecological flexibility and ability to succeed in edge habitats between natural and anthropogenic landscapes [70].

We also observed a dichotomy in species' relationships with forested areas. While some species avoided forested areas (*M. arvalis*, *M. tristrami*) others appeared to prefer it (*A. uralensis*). This pattern likely reflects the evolutionary history and niche partitioning of these species, with grassland specialists like *Microtus* species maintaining their historical habitat preferences. The unique association of *A. amphibius* with waterways is consistent with its semi-aquatic lifestyle and specialized niche within this group. The idiosyncratic distributions of *S. elater*, *M. socialis*, and *S. williamsi*) suggests more complex habitat requirements indirectly captured by our covariates.

The observed habitat preferences and ecological niches of individual species provide context for understanding broader patterns in host community composition and their implications for disease dynamics. Species like *M. arvalis* and *A. uralensis*, which occupy distinct habitats, contribute to the spatial heterogeneity of host communities, potentially influencing pathogen distribution and transmission. These habitat-driven patterns underscore the importance of examining community composition alongside individual distributions to assess how host diversity influences disease dynamics.

In addition to individual species' distributions, host community composition influences disease dynamics, particularly in regions where multiple pathogens circulate in wildlife. In our

system, we found a positive trend between medically-relevant species richness and disease richness, suggesting a potential amplification effect. Our control group of all small mammal species showed a marginally weaker association. However, the disease samples were primarily made up of medically-relevant small mammal species, which may explain the similarity in association. While our results trend toward disease amplification rather than dilution, the correlation between host richness and disease suggests the possibility of amplification, not what such a mechanism would be [37].

The practical significance of this work extends beyond ecological understanding to critical public health applications. We successfully generated distribution models for 15 medically-relevant small mammal species across the South Caucasus that not only characterize individual species distributions but identify areas of high species richness that may influence zoonotic spillover risk. Given the accuracy and discrimination ability of our models, we recommend their use for understanding zoonotic spillover risk.

Chapter 3

Spatial Modeling of Tularemia Risk and Uncertainty in the South Caucasus

3.1 Introduction

The transmission dynamics of zoonotic diseases are shaped by a complex interplay of environmental factors, within and between-host dynamics, and opportunities for human contact [71]. A critical aspect of these dynamics is how the spatial distribution of reservoir species directly influences pathogen heterogeneity across landscapes. While both pathogens and their reservoir hosts exhibit distinct spatial patterns, the relationship between these patterns requires careful consideration when determining appropriate scales of study and data collection [4, 72]. This complexity is further amplified in systems with multiple susceptible species, varied lifecycles, and multiple transmission modes. Therefore, a multi-faceted approach is necessary to predict spatio-temporal variation in disease risk and understand spillover drivers into human populations, particularly when studying pathogens with intricate ecological relationships.

One such pathogen is tularemia, caused by the highly infectious bacteria *Fancisella tularensis* and requiring only a small number of bacteria to cause infection [19]. It affects numerous animals and arthropod species in both terrestrial and aquatic cycles across the northern hemisphere [19,22]. In the aquatic cycle, pathogen survival appears to be affected by water temperature and salinity [73]. Occurrence in the environment is likely insufficient for disease to occur in hosts in the area, which may explain why it is geographically and temporally sporadic with up to decades between outbreaks [22, 74] *F. tularensis* currently has four recognized subspecies, of which *F. tularensis* subsp. *holarctica* (Type B) and *F. tularensis* subsp. *mediasiatica* are found in the Caucasus [20]. The subspecies are distinct in geographic distribution, pathogenicity, as well as phylogeny.

Type B is typically associated with aquatic systems and has a wide range in North America, Europe, and Asia [20, 75]. It affects hares, ticks, small rodents, and mosquitoes proximate to water systems. In the South Caucasus *Apodemus* sp., *Microtus socialis*, and *Microtus arvalis* are considered the main reservoirs [9]. Important vectors include *Ixodes* sp., *Haemaphysalis* sp., ticks, and fleas [9]. Ticks retain *F. tularensis* throughout their lifecycle of up to two to three years, which allows *F. tularensis* to spread to rodent hosts over time while maintaining the pathogen in the environment [21]. In humans, Type B can cause severe disease depending on the route of infection [20, 76].

Tularemia has several clinical manifestations in humans that result from different infection routes and levels of exposure to infected hosts. The two most common forms of tularemia—ulceroglandular and glandular—occur when *F. tularensis* enters through broken skin or mucous membranes via cuts; bites; contaminated water exposure; or contact with infected animals, meat, or vectors [21–23]. Water contamination can result from small mammals entering settlements and infecting wells or other sources of drinking water [73]. While both forms affect the lymph nodes, ulceroglandular tularemia is distinguished by an ulcer at the infection site [24]. Oropharyngeal tularemia develops following ingestion of contaminated food or water with particular risk from spring water, communal water sources, and well water that may be contaminated by infected carcasses or feces [73]. Pneumonic tularemia results from inhaling F. tularensis in dust or aerosols [21]. This presents an occupational hazard primarily affecting farmers and hunters exposed to contaminated hay, infected carcasses, or sheep shearing activities [22, 24]. Oculoglandular tularemia occurs when bacteria enter the body directly through the eye and has similar risk factors to pneumonic infections [24, 25]. The final form, typhoidal, is a difficult to diagnose systemic infection with a non-specific infection route [24, 25].

Symptomology and mortality of tularemia vary with infection type, with an untreated mortality rate of up to 60% [24]. The severity of disease and its potential for weaponization make this a disease of public health concern [22]. Tularemia has been re-emerging in the South Caucasus, and

the variety of infection types and species involved in the system make this an ecologically relevant and practically important region for study [77, 78].

Given the endemic nature of tularemia in the South Caucasus, we used tree-based Bayesian models to predict tularemia occurrence in humans and host-vectors, as well as the spatial distribution of each clinical manifestation in humans. By analyzing the spatial patterns of predicted risk, and their associations with biological and geographic covariates, we identified potential hotspots of spillover and their correlates. Comparing the overlap between human and host-vector risk predictions helped distinguish areas where direct versus indirect exposure may be more common, reflecting different environmental conditions or risk behaviors. The models also allowed us to examine how environmental variables and the presence of vertebrate hosts and arthropod vectors correlate with the spatial distribution of different clinical manifestations, each representing distinct transmission pathways. We expected each clinical manifestation to exhibit distinct spatial patterns of predicted risk due to the combination of environment, host, and behavior patterns that influence the type of exposure. In addition, understanding how different clinical forms of tularemia are distributed across demographic groups and countries can be useful in directing intervention strategies and educational initiatives towards those at increased risk. This integrated approach provides insights into the spatial epidemiology of tularemia in the region and can inform public health surveillance and intervention strategies.

3.2 Methods

Using selected biological and geographic covariates, we developed predictive models of tularemia risk in both humans and host-vectors across the Caucasus to identify patterns of risk and potential hotspots. The biological covariates captured information on host and vector ecology, while geographic covariates represented both bioclimatic conditions and landscape features, allowing us to explore how specific conditions affect the spatial distribution of risk. We also modeled the spatial distribution of each clinical manifestation separately to examine how these biological and geographic factors correlate with different transmission pathways. By integrating spatial

predictions with demographic patterns of infection, we characterized the multi-faceted nature of tularemia risk in the South Caucasus.

3.2.1 Disease risk models

We used a Bayesian Additive Regression Trees (BART) model to quantify the likelihood of tularemia risk across the landscape. This approach applies Bayesian inference to a powerful non-parametric machine learning ensemble [38]. As this regression approach does not assume linear relationships between predictors and outcomes, it is well suited for use in ecology, where linearity is rare. The ability to handle nuanced, nonlinear relationships while avoiding overfitting makes this model type particularly valuable for our study region, where we predict to unsampled areas due to geographic and political constraints. The explicit uncertainty quantification is important for both interpretation and planning, allowing us to identify areas where additional sampling would be most beneficial. [38, 39]

We predicted the risk of tularemia across the region for humans, each clinical manifestation of disease, and host-vectors using the `embarcadero` package. Available tularemia data were only positives, so, based on the model fit from an exploratory analysis and recommendations from (Chapter 2), we used random background points as a proxy for absences in equal number to presence points to train the model. To evaluate model fit for discrimination ability and biological accuracy we used both the in-sample and out-of-sample area under the curve (AUC) and the Boyce index (Chapter 2).

3.2.2 Human and host-vector data

Human case data were collected through two systems: historical records (pre-2006) compiled from quarterly and annual reports, and modern records (post-2006) gathered through passive surveillance via the Electronic Integrated Disease Surveillance System (EIDSS). The National Center for Disease Control and Public Health (NCDC) in Georgia, National Center for Disease Control and Prevention (NCDCP) in Armenia, and Republican Anti-plague Station (RAPS) in Azerbaijan provided these records. The dataset (N=1053, 1956-2020) included infection location

(latitude and longitude), year, patient gender, clinical manifestation, and occupation for laboratory-confirmed tularemia cases. Occupational risk was categorized as a binary variable, with veterinary staff, farmers, livestock workers, and others classified as high-risk occupations.

Host and vector sampling for *F. tularensis* primarily occurred in response to suspected or confirmed human cases. Additional samples were occasionally collected during routine surveillance and epizootic investigations without human spillover. The resulting dataset included 1060 positive host-vector samples made up of hosts only, vectors only, or host-vector pairs. Each record included species or family identification for both hosts and vectors. Human cases and animal positives were confirmed using either culture, PCR, or serology testing.

3.2.3 Covariate layers

To maintain consistency temporally and in measurement across countries, we used global, publicly available covariate layers for model prediction. We selected 19 bioclimatic layers and elevation from WorldClim Version 2 (Appendix B) that provided globally standardized grid data averaged from 1970 to 2000 at 30 second spatial resolution [45]. These layers represent temperature and precipitation patterns that may affect pathogen, host, or vector distributions and influence human exposure and risk behaviors in space.

Fourteen layers were chosen from the `geodata` package that aggregates global spatial datasets (Appendix B). These layers were selected to represent environmental and anthropogenic factors that may influence the pathogen lifecycle and contact rates between infectious hosts, vectors, and humans. All layers were masked to the cover the South Caucasus region, and those not already within a 0-1 range were scaled.

Lastly, because small mammal hosts and tick vectors play a role in tularemia persistence and transmission, we include predicted species distributions for 15 small mammal (Chapter 2) and 21 tick species (Appendix D), with a known or suspected connection to zoonotic disease as covariate layers (Appendix E, methods described in Chapter 2). As the range for each layer was between 0 and 1, the layers were not scaled.

We analyzed both the pattern of covariate selection across models and the variable importance metrics to identify biological and geographic covariates most strongly associated with predicted tularemia occurrence. This assessment helps characterize variables that may influence tularemia distribution patterns across the region.

3.2.4 Spatial overlap between risk patterns

To quantify the spatial consistency between predicted human and host-vector occurrence patterns, and to assess which clinical manifestations of tularemia showed the strongest spatial overlap with host-vector risk, we calculated Schoener's D [44]. Strong spatial overlap between predicted host-vector and human occurrence (D values closer to 1) may suggest transmission pathways involving direct contact, while weaker spatial associations (D values closer to 0) might indicate indirect environmental exposure pathways.

3.3 Results

Of the 1053 human cases, 61% were male, with similar median ages between men (25 years) and women (26 years). Twenty-three percent had a direct occupational risk, 43% had an indirect occupational risk, and 34% did not report occupation. Men were more likely to contract oculoglandular, typhoidal, and ulceroglandular tularemia compared to women. For all clinical manifestations except oculoglandular, the majority of cases did not report a risky occupation, though this information had high missingness.

Clinical presentations varied in frequency, with glandular tularemia being most common, followed by typhoidal, oculoglandular, oropharyngeal, ulceroglandular, and pneumonic forms (Appendix F). The distribution of clinical manifestations differed markedly between countries. In Armenia, glandular cases predominated (79%), while typhoidal cases were most frequent in Azerbaijan (79%). Cases in Georgia were the most evenly distributed between oropharyngeal (34%), glandular (18%), oculoglandular (17%), and ulceroglandular (17%) manifestations.

Table 3.1: Tularemia clinical manifestation demographics across the South Caucasus.

	Glandular (N = 342)	**Oculoglandular (N = 166)**	**Oropharyngeal (N = 160)**	**Pneumonic (N = 42)**	**Typhoidal (N = 260)**	**Ulceroglandular (N = 83)**
Gender						
Male	187 (55%)	108 (65%)	86 (54%)	20 (48%)	192 (74%)	53 (64%)
Female	155 (45%)	60 (36%)	74 (46%)	22 (52%)	68 (26%)	30 (36%)
Age (year)						
Median	22	24	26	15	29	26
Range	0-74	4-70	3-81	8-51	6-74	5-74
Occupational risk						
Direct	109 (32%)	65 (39%)	35 (22%)	8 (19%)	12 (5%)	17 (20%)
Indirect	172 (50%)	64 (39%)	55 (34%)	24 (57%)	101 (39%)	32 (39%)
Missing	61 (18%)	37 (22%)	70 (44%)	10 (24%)	147 (57%)	32 (39%)

Forty-seven percent of the host-vector dataset were host-vector pairs, 21% were host only, and 26% were vector only. Host positives included livestock (*Ovis aries* and *Bos taurus*), wildlife (Ranidae family, *Vulpes vulpes*, *Sus scrofa*, *Vormela peregusna*, *Oryctolagus cuniculus*, *Meles meles*, *Crocidura armenica* and *Canis lupus*), and *Felis catus*. The data included sixteen rodent species including mice, rats, voles, hamster, gerbil, and jird. *Meriones libycus* and *M. arvalis* were by far the most common species observed. Of the host only and host-vector pairs, 51% were *M. libycus*, and 36% were *M. arvalis*. The vectors were a combination of tick, flea, and mite species. Of the positive samples involving a vector, *Dermacentor marginatus* made up 22%, *Hyalomma marginatum* and *Rhipicephalus annulatus* were each 14%, and *Ixodes laguri* was 11%. Unspecified species of the *Ixodes* genus made up an additional 11%.

We predicted the relative risk of tularemia for overall human cases, host-vectors, and each clinical manifestation of tularemia (Figure 3.2). Model performance was strong across all predictions, with AUC values ≥ 0.98 for in-sample AUC and ≥ 0.96 for out-of-sample. The Boyce index was ≥ 0.85 for most models, with lower values for the oropharyngeal (0.8) and pneumonic risk (0.71) predictions. Prediction uncertainty showed a consistent pattern across all models: high uncertainty in areas of high predicted risk and low uncertainty in areas of low predicted risk (Figure 3.1).

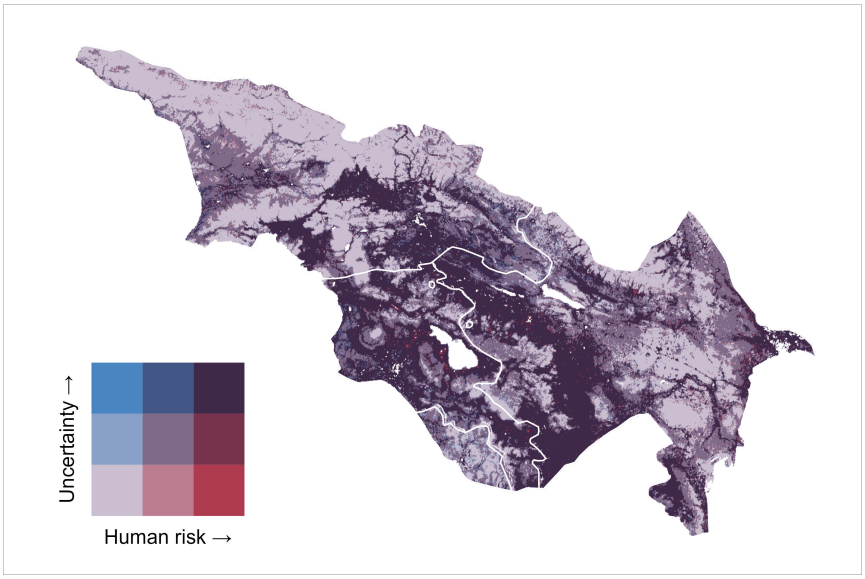
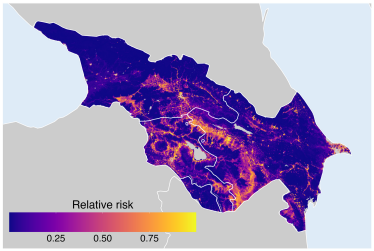
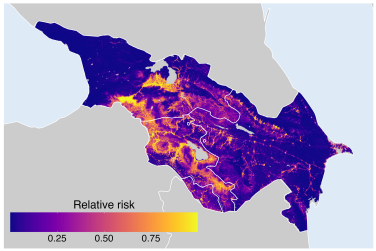


Figure 3.1: Uncertainty and predicted risk of human tularemia across the South Caucasus.

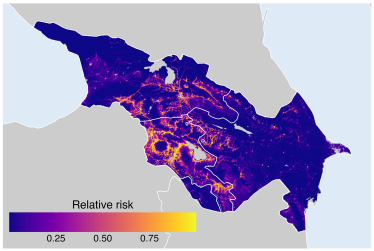
As expected, the predicted risk distributions varied by species and clinical manifestation (Figure 3.2). Overall human risk was relatively mosaic with large areas of near null risk. Host-vector risk followed generally the same pattern, but appeared to surround areas of high human risk with a 58% niche overlap (Table 3.2). While there were distinct distribution patterns between predictions, most combinations had a 60 to 65% niche overlap. The predicted host-vector risk had the most overlap with glandular tularemia (65%) and the lowest with typhoidal tularemia (44%). Glandular, oculoglandular, and oropharyngeal tularemia showed similar spatial distributions, with highest predicted risk along the north-south corridor, closely matching the pattern of overall human risk. Ulceroglandular and pneumonic forms exhibited more diffuse risk patterns. Pneumonic tularemia showed relatively low but widespread risk across the region, notably absent only in forested areas. Ulceroglandular risk was also moderately low and spread throughout the central and western part of the Caucasus. Typhoidal tularemia had the most extensive area of high risk among all clinical manifestations, with elevated risk concentrated along the foothills of the Lesser Caucasus range. Oculoglandular and oropharyngeal forms had the highest overlap between clinical forms (69%) while typhoidal and ulceroglandular forms had the least (41%).



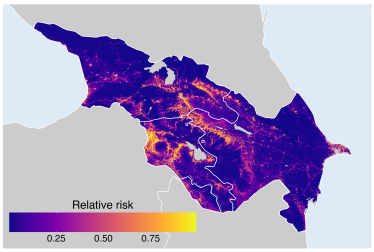
(a) Human risk



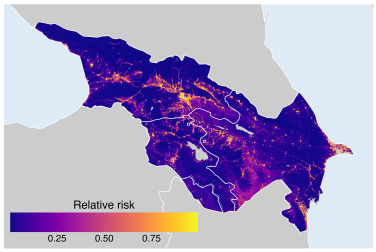
(b) Host-vector risk



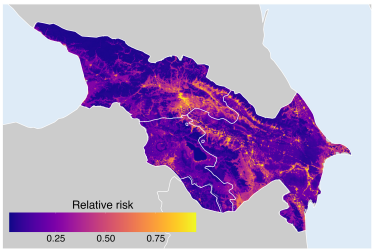
(c) Glandular tularemia



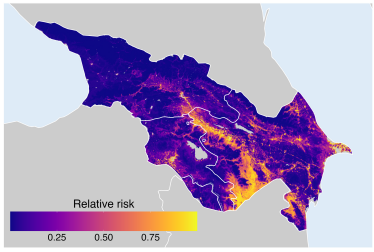
(d) Oculoglandular tularemia



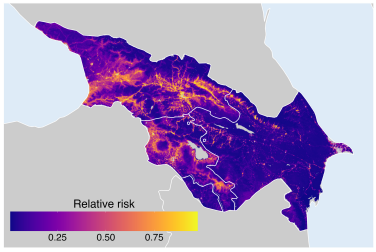
(e) Oropharyngeal tularemia



(f) Pneumonic tularemia



(g) Typhoidal tularemia



(h) Ulceroglandular tularemia

Figure 3.2: Predicted tularemia risk across the South Caucasus.

Table 3.2: Pairwise spatial overlap (Schoener's D) between predicted tularemia risk patterns. Note: Values represent Schoener's D statistic (0 = no overlap, 1 = complete overlap).

	Host-vector	Human	Gland.	Typh.	Oculo.	Oroph.	Ulcero.
Human	0.58	–					
Glandular	0.65	0.62	–				
Typhoidal	0.44	0.69	0.42	–			
Oculoglandular	0.63	0.66	0.64	0.54	–		
Oropharyngeal	0.56	0.67	0.57	0.61	0.69	–	
Ulceroglandular	0.63	0.54	0.62	0.41	0.65	0.59	–
Pneumonic	0.56	0.61	0.50	0.65	0.63	0.67	0.63

Across all models, host and vector species prediction layers were selected more often than geographic and climate characteristics. However, the individual predictor selected most often, in all eight models, was proportion of built environment. This was followed by host species *Lepus europaeus*, *M. arvalis*, *Mus musculus*, and temperature seasonality in five, and by vectors *Nothocricetulus migratorius*, *H. marginatum*, and host *Meriones tristrami* in four models. Supplementary analyses (not shown) revealed that areas with higher host-vector likelihood had a higher density of human cases. In contrast, randomly selected absence points showed higher density in areas with lower host-vector likelihood. Variable importance showed minimal variation within each model. Selected variables had almost complete overlap in importance values. No model was dominated by a single predictor; instead, predictions reflected a complex combination of multiple variables with similar contributions.

3.4 Discussion

Using decades of surveillance data on tularemia in humans, vertebrate hosts, and arthropod vectors across the South Caucasus, we characterized spatial patterns of predicted risk and their biological and geographic correlates. We found that clinical manifestations of tularemia showed distinct spatial distributions and varied across counties. Armenia showed predominantly glandular infections, Azerbaijan reported mainly non-specific typhoidal infections, while Georgia exhibited a more diverse pattern with oropharyngeal, glandular, oculoglandular, and ulceroglandular forms all being common. These differences could reflect either distinct natural foci of *F. tularensis*, or

systematic differences in risk behaviors between countries. It has been proposed that in Georgia risk differs seasonally, with winter outbreaks caused by closer contact with rodents moving near homes, and in summer and autumn though agricultural work and recreation [79]. The multiple pathways of infection suggested by this hypothesis (contamination of food and water in winter, and direct contact with hosts and vectors in summer and autumn) are consistent with the multiple clinical manifestations that appear; oropharyngeal infection via ingestion, and glandular and oculoglandular infection via direct contact in the summer months. However, we were unable to assess seasonal patterns directly, and glandular infection may also result from contaminated water [13]. Unlike Georgia, glandular and oculoglandular forms were the most common in Armenia, and were primarily the result of a waterborne outbreak in 2003 in the Kotayk marz with infection occurring cutaneously [78]. Because tularemia is a rare disease, the patterns of large stochastic outbreaks may obscure those of other smaller outbreaks.

While glandular, ulceroglandular, typhoidal and pneumonic tularemia can result from direct or indirect contact with infected animals, their overlap with host-vector risk suggest different patterns. The moderate spatial overlap between host-vector risk and both glandular and ulceroglandular manifestations suggests that spillover from these forms more often results from direct contact. The lower spatial overlap of pneumonic and typhoidal manifestations with host-vector occurrence may suggest that while direct transmission is possible, indirect transmission was more common. However, we also observed that oculoglandular tularemia, which has an indirect transmission route, had an equal amount of overlap to the ulceroglandular form. This overlap between oculoglandular and host-vector risk may be explained by the fact that this form of tularemia often occurs in farmers, and host-vector densities may be higher in agricultural areas [24, 26, 80]. Areas of high human risk and low host-vector risk may suggest that the presence of these species is not necessary to cause disease in humans, with infection occurring though indirect pathways. Factors beyond host or vector presence, such as environmental conditions or human behavior patterns, combine to determine which clinical manifestations predominate in different areas. For example, pneumonic cases may

cluster in areas where occupational activities generate contaminated aerosols, regardless of current animal density [78].

To better understand these spillover patterns, we examined which medically relevant small mammal and vector species showed the strongest associations with predicted tularemia risk patterns. As the source and species of human infection were not available, we examined which medically-relevant small mammal and vector species showed the strongest associations with tularemia risk patterns. Hosts *M. musculus*, *M. arvalis*, *Apodemus witherbyi*, *Apodemus uralensis* and vector *H. marginatum* were selected in both our ulceroglandular and glandular tularemia models— clinical manifestations often resulting from direct contact with infected hosts or vectors. These infection types were more likely to occur in areas with higher predicted species likelihood compared to random pseudo-absence locations. Of these species, only *M. arvalis* is currently recognized as a reservoir species for tularemia [77, 81]. Two other reservoir species, *M. socialis* and *Arvicola amphibious*, were selected in the glandular and ulceroglandular model respectively [9,14]. Interestingly, only one tick vector species was selected for the ulceroglandular model, while five were retained in the glandular model. This may suggest different spatial associations between vector occurrence and these clinical manifestations, though the biological mechanisms for such a difference are unclear as both glandular and ulceroglandular infection occur though the same pathways.

While the variables retained across models were primarily medically relevant host and vector species, the built environment was the most frequently selected predictor. Cases typically occurred in grid cells with approximately 20% built environment, compared to random locations with less than 1%. This likely represents areas of urban-rural edges where human settlements abut natural areas [82]. These are often areas of increased spillover risk due to abundant vectors and livestock, higher contact rates with wildlife, and favorable habitats for rodents [26,80]. These transition zones are particularly relevant for tularemia, which is predominantly a rural disease requiring contact between humans and infected animals or contaminated environmental sources for transmission

[23]. The moderate level of development may also mean people rely on untreated water sources like springs, rather than municipal water systems.

This interface between human infrastructure and natural systems illustrates the complex factors influencing disease transmission that is reflected in our models' uncertainty patterns. Across all models we observed the highest uncertainty in areas with highest predicted tularemia risk and lowest uncertainty at low disease risk. This was expected, as several requirements must be satisfied for spillover to occur [5]. Simply put, the environment must be suitable for the pathogen, reservoir hosts and vectors must maintain said pathogen, and a susceptible human must come into contact with a reservoir or vector species in a way that allows transmission to occur. Each step in this process has a high degree of uncertainty, therefore we observe much less predictive uncertainty where the environment is wholly unsuitable for transmission than areas where a complex series of ecological and epidemiological factors align for spillover into humans.

Given the complex ecology of tularemia, our study leaves room for further exploration. While tularemia is a reportable disease, tularemia type B infections are likely underestimated due to milder symptoms [79]. Regional differences in surveillance and access to healthcare and diagnostics may influence observed spatial patterns. The multitude of infection pathways with different risk patterns may weaken or confuse the association between predictor and outcome. We attempted to mitigate this issue by separately modeling each clinical manifestation. However, multiple sources of infection are possible within each clinical type and it was unclear why clinical manifestations with identical transmission routes, such as the glandular and ulceroglandular, had different distribution patterns. There was also a potential for sampling bias between the human and host-vector datasets. However, comparing the relative strength of spatial overlap across different clinical manifestations remains informative as a preliminary step. All clinical manifestations shared this sampling bias with respect to host-vector occurrence, therefore, differences in their spatial overlap patterns are theoretically equally impacted.

To our knowledge, this study represent the first prediction of tularemia risk in the South Caucasus, providing a foundation for future research and public health interventions. Our findings sug-

gest several directions for enhancing the understanding of tularemia in the region. We recommend active surveillance of human cases to better understand regional heterogeneity of clinical manifestations. Independent sampling of animals, particularly in previously unsampled areas, would address current sampling biases and confirm or correct the current animal risk distributions. We also recommend categorizing cases by transmission pathway (direct contact, environmental contamination, and indirect contact) in order to align with the source of infection rather than the route of infection. Identifying the spatial patterns of these three categories would allow public health officials to target intervention efforts tailored to the risk factors common in the area.

Chapter 4

Disease Richness: Identifying Hotspots Among Seven Zoonoses in the South Caucasus

4.1 Introduction

Zoonotic diseases, transmitted between animals and humans, are a growing threat to global health with approximately 75% of emerging infectious diseases originating in animal species [1]. The one health framework emphasizes the interdependence of human, animal, and environmental health and is essential to understanding and preventing zoonotic spillover. Despite the research into zoonoses across various fields, the complexities of multi host pathogen systems and differences in landscape and human activity mean that the determinates of spillover are difficult to identify and vary over space and time [4]. Key questions remain about the specific pathways and environmental conditions that enable pathogens to spill over from wildlife and domestic animals to humans across space and pathogen type, and how interactions across ecological and environmental factors shape the spillover process.

The South Caucasus is endemic for several zoonoses, including anthrax, brucellosis, Crimean-Congo hemorrhagic fever (CCHF), leptospirosis, tick-borne encephalitis (TBE), and tularemia [8, 9]. The diverse ecosystem of this temperate biodiversity hotspot supports a variety of wild ungulates, small mammals, fleas, and ticks that serve as reservoirs and vectors for these pathogens [22, 34, 35]. The ecological complexity makes the South Caucasus an informative and challenging system for examining zoonotic transmission.

While existing literature has tended to focus on individual countries, [10–15], Georgia, Armenia, and Azerbaijan share geographic features and certain habitat types that warrant examination as a single region [9]. Analyzing disease risk regionally offers a more comprehensive picture of disease patterns and improves the predictive power of our models. Examining how bioclimatic,

landscape, and host and vector species influence zoonotic risk across borders is useful in determining ecological drivers and planning surveillance and response efforts by incorporating neighboring risk.

As zoonotic infections generally result from contact with infected hosts or vectors, we hypothesize that human disease risk is spatially closely related to host and vector disease. Transitional zones between two ecological regions, or ecotones, have been identified as areas of increased pathogen transmission due to higher vector densities, and increased contact rates with humans [7, 26]. We believe that this will manifest in our system in regions of transition between the built environment and surrounding agricultural land. We also expect that medically relevant species will be significant predictors of disease risk for each pathogen, with the species selected by the model as possible hosts or vectors. These species may also be more common in agricultural areas due to increased food and habitat, which increases their contact rates with humans. By examining the regional patterns of risk across pathogens, we will identify potential drivers of spillover unique to each pathogen, as well as common factors contributing to spillover across multiple diseases.

These findings can aid public health efforts by highlighting areas of unexpected low or high risk, guiding targeted sampling and surveillance. Additionally, identifying hotspots across multiple pathogens can direct resource allocation, enabling the efficient use of limited public health resources to mitigate the spread of several diseases simultaneously. We believe that a large portion of transmission is the result of occupational exposure given the prevalence of agricultural activities in this region, and expect human hotspots occur in both agricultural areas where occupational risk of zoonotic exposure is higher and in cities where access to healthcare makes detection and diagnosis more likely than in rural areas. We also expect human hotspots to be a subset of animal hotspots, as pathogens spillover from animal to human populations.

To address these questions, we employed Bayesian modeling to predict both human and animal risk for seven zoonotic diseases across the South Caucasus. By training the model on a variety of ecological covariates, we identified factors influencing high- and low-risk areas, providing insights

into the spatial determinants of zoonotic risk. Additionally, comparing human and animal risk predictions allowed us to identify areas at increase risk of spillover from animals to humans and examine why that was not the case in other areas. We also explored patterns of occupational risk and species likelihood in different habitats to better understand the spatial heterogeneity of disease risk and the importance of transitional zones in our system. Finally, by combining risk maps across multiple diseases, we identified hotspots for zoonotic transmission. These results have implications for targeting prevention and control efforts and guiding resource allocation for future data collection.

4.2 Methods

We used a Bayesian approach to fit prediction models using geographic, climatic, and host and vector distribution covariates that represented the requirements and limitations to pathogen survival and spread to create an estimate of the relative risk across the region. This approach allowed us to examine the ecological drivers that shaped disease risk over space. By modeling both human and animal risk, we were able to examine patterns between them. Areas of high risk for both indicate areas of spillover from hosts or vectors into human populations, while areas of high risk for only one or the other represent a more complex ecological or data collection phenomenon. By exploring these patterns across seven different endemic zoonoses, we were able to highlight hotspots of risk across pathogens, which is useful for public health planning and intervention. To determine whether human infection was largely driven by agricultural risk factors, we examined the distribution of occupational risk and the likelihood of medically-relevant small mammal host species occurring in cropland areas.

4.2.1 Diseases

The diseases chosen for this study were selected for their endemicity in the South Caucasus and their impact on both human and livestock health. A number of these diseases are considered

neglected and are considerably underreported. Questions remain as to their ecological drivers, and, to our knowledge, their distributions have not been predicted in the South Caucasus.

Anthrax, caused by the bacterium *Bacillus anthracis*, primarily infects ungulates [83], but can spread to humans and other mammals [84, 85]. Humans contract anthrax through contact with infected animals or animal products such as meat or hides during processing [3]. Anthrax can be weaponized, and both symptoms and fatality rate vary with exposure route [84, 86]. It is endemic in the Caucasus region, with Georgia mandating annual livestock vaccination until 2008 and Azerbaijan until 2010 [10, 34]. Current voluntary vaccination programs in Georgia cover an estimated 15% to 20% of livestock.

Brucellosis is a common but neglected zoonoses often found in cattle and other livestock species [87]. Spillover to humans is most often the result of ingesting contaminated food, such as unpasteurized dairy or occupational exposure to livestock [88, 89]. Prevention of human disease is primarily achieved by controlling disease in animals. Varying symptoms make differential diagnosis difficult, and the disease is often underreported due to this and a lack of surveillance and reporting.

Crimean-Congo hemorrhagic fever is a tick-borne virus that follows a tick-vertebrate-tick life cycle [90]. Tick are both vector and reservoir, particularly *Hyalomma* species and *Rhipicephalus* species, and a wide variety of domestic and wild vertebrates can become infected [91]. Crimean-Congo hemorrhagic fever frequently causes severe disease in humans, including hemorrhage, and has a lengthy recovery time. Human fatality rates are estimated to be 30 to 50%, but much higher fatality rates have also been reported. Those who have occupational exposure to ticks are particularly at risk, though the pathogen can also be contracted by handling infected meat. Human to human transmission is possible, and healthcare workers and those working with medical samples are at high risk.

Leptospirosis is a bacterial disease carried and transmitted primarily by rats and other rodents, but able to infect other mammals [92, 93]. It is underreported globally because human infection is often asymptomatic or mild with nonspecific symptoms. However, disease may be severe with 5

to 15% fatality rates. Infection typically occurs through contact with lakes or rivers contaminated with the urine of infected hosts, but it can also be contracted via contaminated soil or direct contact with urine, blood, or animal tissue. Leptospirosis is strongly associated with occupational risk (agriculture, slaughterhouses and butchers, veterinary medicine, and sewage maintenance), and outdoor water recreation.

Plague has caused several major pandemics throughout history, leading to a large scale loss of life [94]. Disease is severe and progresses swiftly with a high mortality rate without prompt treatment. It has many natural foci around the world and is typically maintained by a rodent-flea cycle. Humans usually contract plague via flea bites, but infection can also occur by handling infected animals, consuming infected meat, or inhaling respiratory droplets from an infected animal [95].

Tick-borne encephalitis, a viral disease of the central nervous system, is primarily transmitted to humans via tick bite [96]. There are several clinical forms of human disease with varying symptoms and fatality rates. Severity of disease also depends on the diagnostic and treatment options available in a given area [97]. In addition to bites, infection can spread in the unpasteurized milk of infected animals. The virus is most commonly associated with *Ixodes* ticks, though other species carry it as well. Infection typically occurs in rural areas where leisure activities and foraging increase the risk for tick bites [98].

Tularemia is caused by the highly infectious bacterium, *Fancisella tularensis* [19]. It affects numerous animals and arthropod species in both terrestrial and aquatic cycles across the Northern Hemisphere [19, 22]. Humans can become infected with tularemia via broken skin or mucous membrane by drinking or bathing in infected water, inhaling contaminated dust, or exposure to the bacteria through the eye [21–23, 73]. Symptomology and mortality vary with infection type, with an untreated mortality rate of up to 60% [24]. Infection is an occupational hazard for farmers and hunters, who come in contact with contaminated hay, sheep shearing, or have contact with infected carcasses [22, 24].

4.2.2 Disease risk models

We employed Bayesian Additive Regression Trees (BART) to predict the likelihood of species presence across the landscape. This powerful machine-learning tool uses Bayesian inference in a nonparametric regression ensemble framework [38]. Unlike traditional regression models, BART does not assume linear relationships between covariate predictors and the outcome, making it suitable for modeling ecological relationships where linearity is rare. The Bayesian approach offers another advantage: we can measure uncertainty by extracting the 95% credible interval from the predictive posterior distribution, which helps evaluate prediction certainty. [38, 39]. We predicted the risk of each disease for humans, host and vectors, and soil (where appropriate) using the `embarcadero` package.

Available case and sampling data were only positives, so, based on exploratory analysis and recommendations from (Chapter 2), we used random background points as a proxy for absences in equal number to presence points. We evaluated model fit for discrimination ability and biological accuracy using both the in-sample and out-of-sample area under the curve (AUC) and the Boyce index (Chapter 2).

4.2.3 Human and host-vector data

Human case data were collected through two systems: historical records (pre-2006) compiled from quarterly and annual reports, and modern records (post-2006) gathered through passive surveillance via the Electronic Integrated Disease Surveillance System (EIDSS). The National Center for Disease Control and Public Health (NCDC) in Georgia, National Center for Disease Control and Prevention (NCDCP) in Armenia, and Republican Anti-plague Station (RAPS) in Azerbaijan provided these records. The dataset (1946-2020) included infection location (latitude and longitude), year, patient gender, clinical manifestation, and occupation for laboratory-confirmed cases. Occupational risk was categorized as a binary variable, with veterinary staff, farmers, livestock workers, and others classified as high-risk occupations. Human case data for TBE was not available from Armenia.

Presence data for infected host, vector, and soil sampling primarily occurred in response to suspected or confirmed human cases. Additional samples were occasionally collected during routine surveillance and epizootic investigations without human spillover. Each record included species or family identification for hosts, vectors, or both. Negative test records were unavailable, as were animal data for leptospirosis in Georgia. Human cases and animal positives were confirmed used either culture, PCR, or serology testing. The case and sample data are summarized in Table 4.1.

4.2.4 Covariate layers

To ensure consistency temporally and in measurement across countries, we used global, publicly available covariate layers for model prediction. We selected 19 bioclimatic layers and elevation from WorldClim Version 2 (Appendix B) that provided globally standardized grid data averaged from 1970 to 2000 at 30 second spatial resolution [45]. These layers represent temperature and precipitation patterns that may affect pathogen, host, or vector distributions and influence human exposure and risk behaviors in space.

Fourteen layers were chosen from the `geodata` package that aggregates sources for global spatial datasets (Appendix B). These layers were selected to represent environmental and anthropogenic factors that may influence the pathogen lifecycle and contact rates between infectious hosts, vectors, and humans. All layers were masked to the cover the South Caucasus region, and those not already within a 0-1 range were scaled.

Finally, because hosts and vectors play a role in zoonotic persistence and transmission, we included predicted species distributions for 15 small mammal species (Chapter 2), 21 tick species (Appendix D), and eight wild ungulates (Appendix G) with a known or suspected connection to zoonotic disease as covariate layers (see complete list in Appendix E, methods described in Chapter 2). As the range for each layer was between 0 and 1, these were not scaled.

We analyzed both the pattern of covariate selection across models and the variable importance metrics to identify biological and geographic covariates most strongly associated with disease risk.

Table 4.1: Summary of South Caucasus disease data and covariate groups used in modeling.

Disease	Armenia	Azerbaijan	Georgia	Host-vector Covariates
Anthrax				Wild ungulates
Human data	326	1414	639	
Animal data	100	1719	624	
Soil data	114	1027	630	
Brucellosis				Wild ungulates
Human data	15394	19452	7830	
Animal data	2265	645	8639	
CCHF				Ticks
Human data	1	400	98	
Host-vector data	11	0	53	
Leptospirosis				Wild ungulates
Human data	56	49	894	
Animal data	1333	52	Data unavailable	
Plague				Small mammals, ticks
Animal data	169	669	17	
Flea data	548	714	66	
TBE				Ticks
Human data	Data unavailable	1	28	
Host-vector data	20	0	27	
Tularemia				Small mammals, ticks
Human data	323	288	442	
Host-vector data	215	375	470	

This assessment helps characterize variables that may influence spillover patterns across the South Caucasus.

4.2.5 Disease hotspots

To identify disease hotspots for humans and host-vectors, we transformed each continuous disease prediction layer into binary values using threshold cutoffs that maximized the True Skill Statistic (TSS), a metric that evaluates binary classification accuracy by optimizing sensitivity and specificity. This approach categorized each raster into two groups: 1 for areas with disease risk and 0 for areas without risk. For plague, we combined data from both mammal hosts and flea vectors to create a single, binary layer of host-vector risk. To identify human hotspots, the human risk layers for six diseases (no human cases of plague reported) were stacked, and the sum of all disease risk layers was calculated for each cell, resulting in a cumulative disease risk or richness map. This process was repeated for the seven disease predictions to identify host-vector hotspots. This allowed us to map areas where risk of multiple diseases overlap and compare patterns of overall risk between animals and humans.

4.2.6 Risk heterogeneity

Along with the patterns of risk and covariates selected, we compared human and animal risk predictions by dividing each prediction into three levels of risk: low (0 to 0.33), medium (0.34 to 0.66), and high (0.67 to 1). Levels of human and animal risk were then compared spatially using bivariate plots. This allowed us to identify areas of spillover and other more complex patterns. We measured how similar the resulting human and host-vector risk predictions were using Schoener's D, which quantifies spatial overlap between two probability distributions in geographic space [44]. This metric ranges from 0 (no overlap) to 1 (complete overlap).

We also used the predicted disease richness to assess whether occupational risk led to increased spillover. To do this, we calculated the proportion of human cases with occupation risk for each level of disease richness (0 to 6). We expected the proportion of cases with occupational risk, and

the number of diseases present in an area, to increase together. This would imply that increased occupational risk in the population leads to increased spillover of zoonoses across pathogens.

Similarly, we explored whether transitional zones have increased risk of spillover due to increased contact with infectious species. We classified areas into transition zones based on the proportion of the cell that was built environment. From first principles, cells with a built environment between 20% and 40% were classified as transition (transition = 1), while all other locations were classified as non-transition (transition = 0) [82]. To balance the dataset, we included all observations classified as transition zones and selected an equal number of non-transition observations through random sampling. We calculated the proportion of cases with elevated occupation risk for both transition and non-transition and assessed significance using a Pearson's Chi-squared test. We expected transition zones to have a higher proportion of cases with occupational risk than non-transition zones, as this is consistent with higher contact rates in those areas.

We also evaluated whether vector species were more likely in transition zones compared to non-transition zones, as has been the case in other systems. To do so, we created an overall measure of vector presence by adding together the likelihood of each vector species into a single layer. The study region was divided into a grid, and the average value of built environment, overall vector presence, and each vector species individually were calculated for each cell. We then calculated the mean likelihood of all vector layers separately for transition and non-transition zones and measured the difference in species likelihood between them using a Wilcoxon rank-sum test. Elevated overall vector likelihood in transition areas is the consistent with the hypothesis that these regions serve as favorable habitats and potential hotspots for increased interactions between vectors, hosts, and humans.

Using the same methodology, we examined the impact of land use on the likelihood of small mammal hosts using cropland. Using the natural break at the 75th percentile (18.55% cropland) we classified each cell as either "high" or "low" cropland areas. Elevated overall host likelihood in high cropland areas is consistent with the hypothesis that this is favorable habitat, and provides increased contact opportunities between hosts and humans.

4.3 Results

From our human cases across all diseases, 22.90% had an occupational risk and 77.10% did not. Men were more likely to have an occupational risk than women (Chi-square p-value <0.01), and those with occupational risk were more likely to be older, with a median age of 41, compared to 17 for no risk (Wilcoxon p-value < 0.01). We found that the proportion of cases with occupational risk was significantly higher in areas with higher disease richness (Figure 4.1), suggesting that increased risk behaviors leads to greater spillover across pathogens. Occupational risk was also higher in transition (29.42%) compared to non-transition areas (20.23%) (Chi-square p-value < 0.01). This supports our hypothesis that contact rates between humans and host-vector species are higher in those zones.

Using our case data and covariate layers, we predicted the risk of anthrax, brucellosis, CCHF, leptospirosis, plague, TBE, and tularemia (Appendix H). Despite the small sample size for some of the diseases, the models performed well with in-sample and out-of-sample AUC values from 0.94 to 1 and Boyce index from .97 to 1 for all but three models. The models with smaller sample sizes (TBE human and host-vector risk, and CCHF host-vector risk) had slightly lower Boyce values, ranging from 0.82 to 0.87.

Across all fifteen models, covariates representing landscape and land use were selected the most often (Figure 4.2). The built environment was most commonly used, in addition to human footprint; two measures representing the presence and impact of humans on the landscape. Several climatic covariates were included, representing either their influence on disease risk, or as a representation of geographic space. Several covariates (grassland, trees, and cropland) represent both habitat types for host-vector species and proxies for level and type of exposure to infectious hosts for humans. Croplands may be important to spillover, as the likelihood of overall small mammal host presence was significantly higher in areas of high cropland compared to areas of low cropland, despite variations across individual species (Figure 4.3a). We also found that transition zones are favorable habitats to most tick vector species we examined (Figure 4.3b). The likelihood of overall vector species presence was significantly higher in transition compared to non-transition

areas. Host and vector species layers were selected for all but one model-human risk of leptospirosis. *Rhipicephalus bursa* was the most commonly selected vector species, *Capreolus capreolus* the most common wild ungulate species, and *Microtus arvalis* the most common small mammal species.

Upon visual inspection of the spatial patterns, the distributions were a reflection of the selected covariates. Disease risk was high in areas of cropland and grassland around the built environment, and lower in forested areas (Figure 4.4). Human risk was mosaic and variable across diseases, but highest in built environments, particularly around the capital cities of Tbilisi and Yerevan. These areas of elevated localized risk coincided with individual cities or towns surrounded by agricultural land. Animal risk was also variable, but was generally more diffuse than human risk. Animal hotspots were more widespread compared to human hotspots, and the highest risk for animals occurred in agricultural and rangeland (pictured in Kosoy 2024).

When comparing human and animal risk (Figure 4.5), we observed middling distribution overlap for three of our diseases (tularemia Schoener's D = 0.59; anthrax Schoener's D = 0.57; CCHF Schoener's D = 0.51). CCHF in particular predicted high human risk in areas of low host-vector risk, which was surprising as vector-borne diseases are typically spread to humans via direct contact [99]. Our other vector-borne disease, TBE met these expectations, as it had the highest niche overlap (0.71) and the closest approximately linear relationship despite its ability to also be transmitted indirectly via unpasteurized dairy. Brucellosis (0.38) and leptospirosis (0.26) had the least overlap. Lower leptospirosis overlap in Georgia may be because animal records for leptospirosis infection were not available for that country, and the animal model was trained on Armenia and Azerbaijan only.

4.4 Discussion

Using decades of zoonotic disease data from humans, hosts, and vectors, we predicted the risk of seven endemic diseases across the South Caucasus. Examining patterns of risk across individual and multiple diseases together is informative for targeting resources and interventions and directing sampling efforts to gain a clearer understanding of risk. These patterns can also help us address

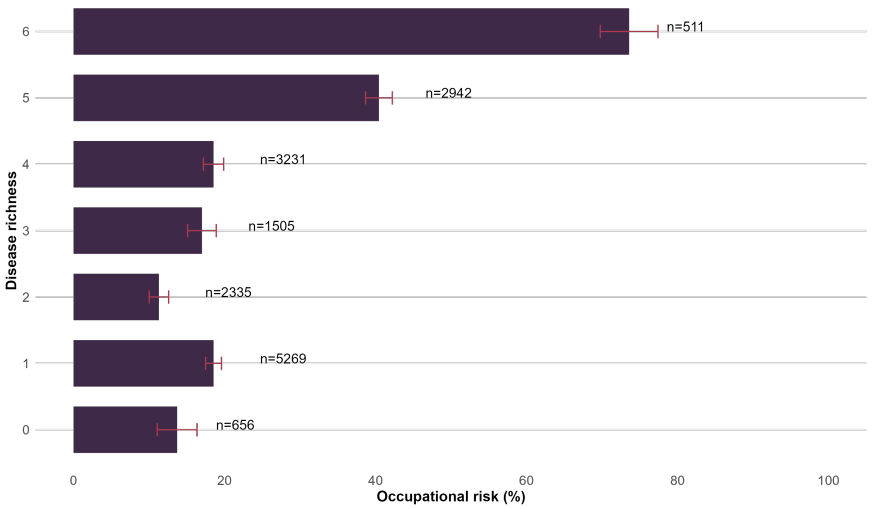


Figure 4.1: Proportions of cases with occupational risk at each level of predicted disease richness.

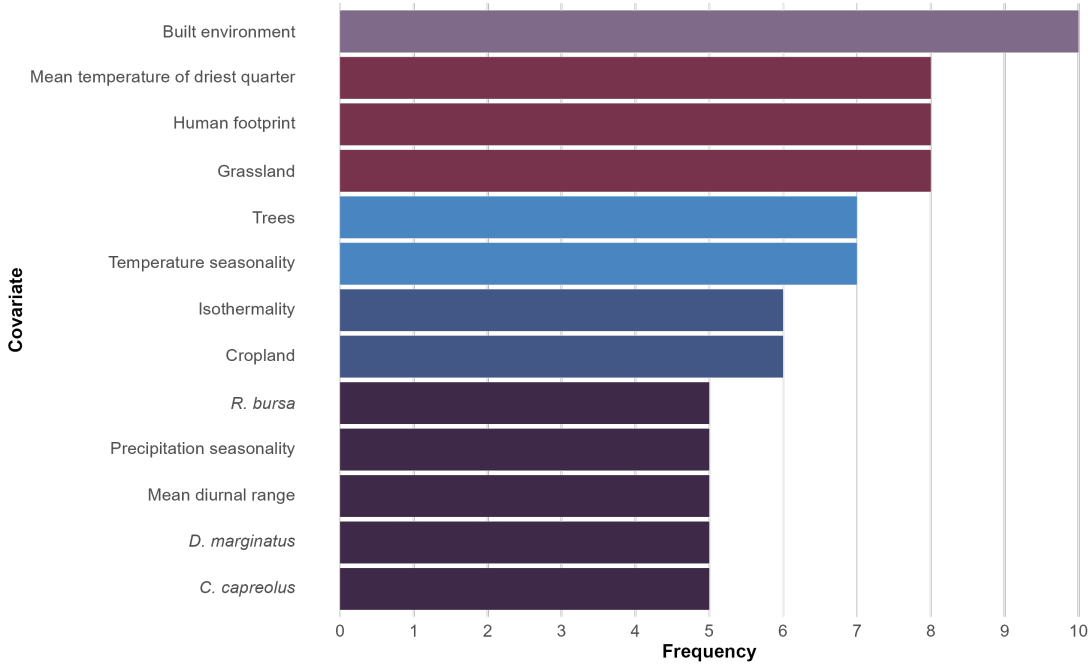
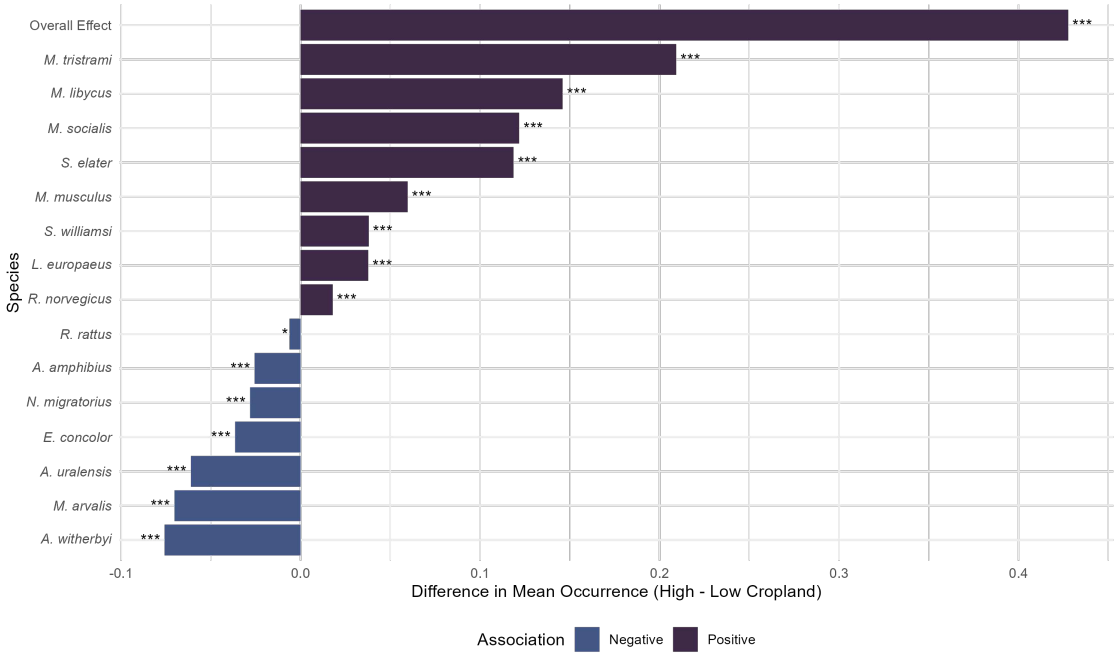
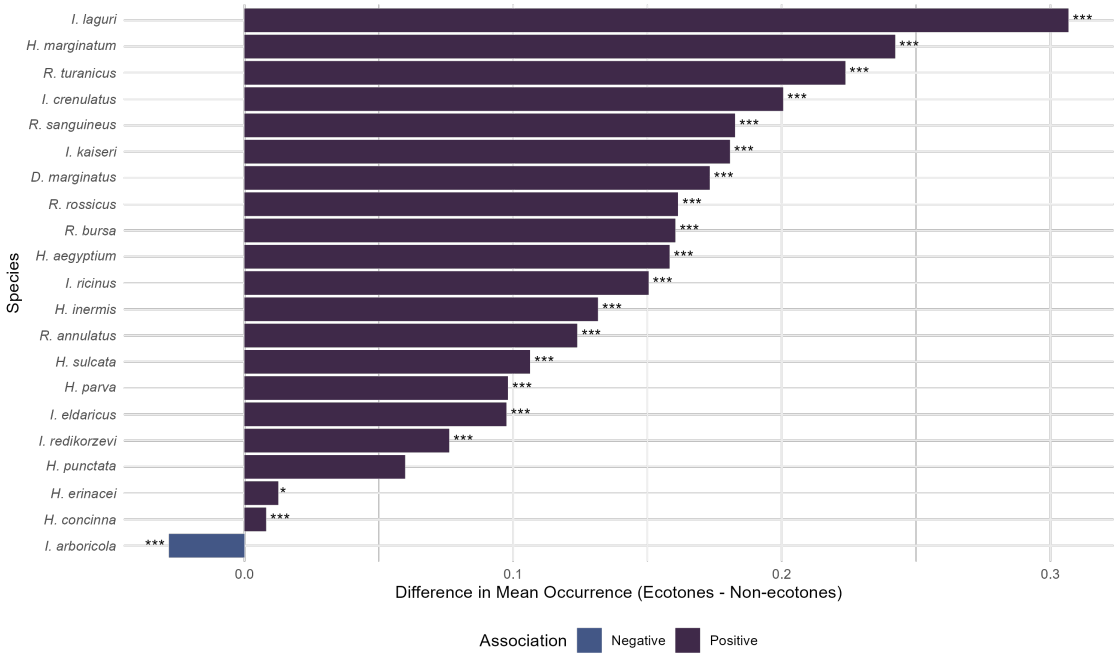


Figure 4.2: Most frequently selected covariates across all disease models. Only covariates occurring ≥ 5 times were included.

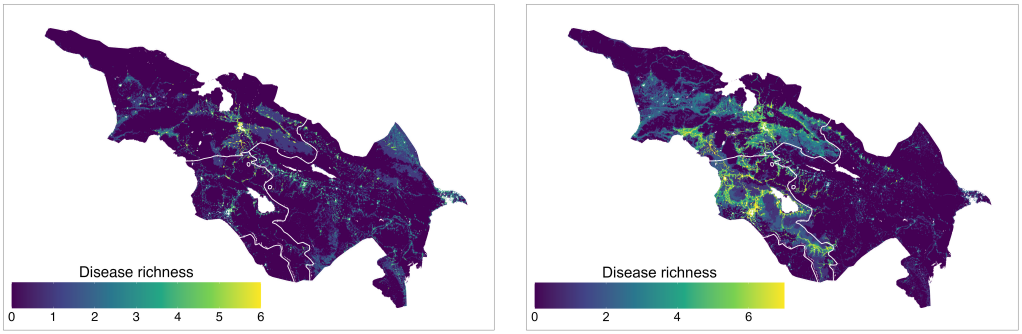


(a) Difference in host species' occurrence between high and low cropland areas. High cropland: ≥18.55% cover, Low cropland: <18.55% cover. * p<0.05, ** p<0.01, *** p<0.001



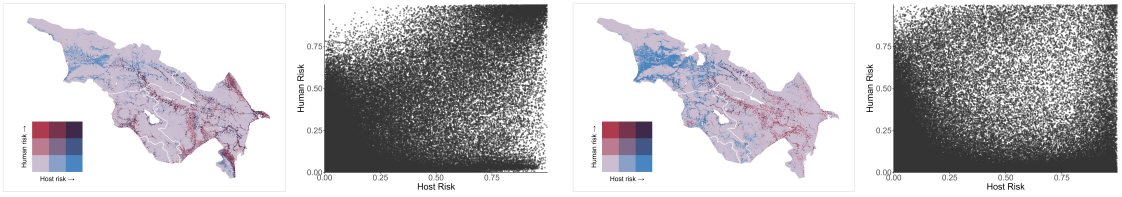
(b) Difference in vector species' occurrence between transition and non-transition zones. Transition: 20-40% built environment. * p<0.05, ** p<0.01, *** p<0.001

Figure 4.3: The effect of landuse on host and vector species presence in the South Caucasus.

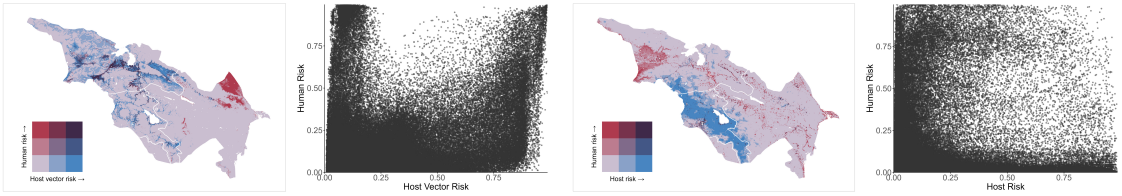


(a) Human hotspots **(b)** Animal hotspots

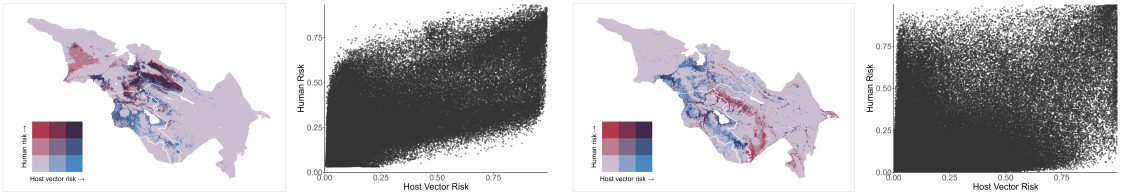
Figure 4.4: Predicted human and animal disease richness across the South Caucasus.



(a) Anthrax risk **(b)** Brucellosis risk



(c) CCHF risk **(d)** Leptospirosis risk



(e) TBE risk **(f)** Tularemia risk

Figure 4.5: Predicted human and animal disease risk in the South Caucasus for anthrax, brucellosis, Congo-Crimean hemorrhagic fever, leptospirosis, tick-borne encephalitis, and tularemia.

ecological questions related to the spatial drivers of spillover, particularly the patterns of contact between host and susceptible human populations.

Much attention in the fields of epidemiology and disease ecology has focused on the concept and impact of disease hotspots. Hotspots have been described as areas of increased prevalence, spatial clustering, pathogen density per host, elevated transmission rates, and increased emergence and the concept has been applied to multiple scales, from individuals to landscapes [6, 7, 27, 100, 101]. The methods used to identify hotspots are varied, including kernel density estimation, spatial autocorrelation, and clustering analyses [102, 103]. Although traditional methods have focused on identifying areas of higher risk for individual pathogens, we advocate for applying species richness to identify existing hotspots, defining them as areas where multiple pathogens are present and transmission to humans is common.

This approach is similar to that used in conservation biology, where biodiversity hotspots are of higher priority and interventions are targeted to address multi-species threats [8]. In disease ecology, areas where multiple pathogens co-occur have met the environmental conditions, host ecology, and human behavior requirements for various zoonotic pathogens to circulate and transmit into the human population. The location of these hotspots is determined by spatial heterogeneity in the requirements for pathogen persistence and transmission, leading to areas where transmission risk across multiple pathogens is high. This "disease richness" approach offers several advantages. When multiple pathogens cause disease in the same area, it suggests shared environmental and behavioral risk patterns that create favorable conditions for zoonotic spillover. This is contrasted with more individual aspects of spillover, such as susceptibility and immune differences. Understanding these shared requirements can inform more efficient intervention strategies that simultaneously address multiple pathogens by intervening on a shared risk factor.

The formation of these multi-pathogen hotspots is shaped by the spatial alignment of heterogeneous ecological and behavioral requirements that result in shared transmission risk factors. One of these foundational ecological requirements is the presence of host and vector species and human exposure to its infected members. Animal host risk over space is influenced by host density

in addition to other factors [6]. Greater host density leads to higher contact rates, increasing the likelihood of human exposure. Rodents and other host species may be more prevalent in agricultural areas, and can experience ecological release in croplands, while human settlements provide nesting, feeding, and breeding opportunities for both rodents and arthropod vectors [26, 80]. Regional variability in host contact rates due to increased density, particularly around farms, and concomitant occupational risk behaviors can create hotspots for disease transmission [6]. Our results supported these findings, as we found that the overall likelihood of small mammal host presence was higher in areas of high cropland, despite variation in individual species' habitats. While previous work has suggested that abundance of host species is a common risk factor that reoccurs across different systems, our work demonstrates its potential as a shared risk factor in multi-pathogen systems when spatial heterogeneity in host abundance aligns across pathogens.

Ecotones have also been identified in individual disease systems as areas of increased zoonotic risk via several proposed mechanisms, including greater vector abundance and increased contact between wildlife and vectors and humans [26]. In the South Caucasus, a similar area of importance is the transition between the built environment and agricultural land. We found that both hotspots of human disease richness and the proportion of cases with occupational risk across disease were higher in transitions than non-transitions. This supports the hypothesis that contact rates between humans and animals is higher in transition areas. We also found agreement in our system that vector species were more abundant in these zones. This suggests that, across pathogens, spillover occurred in transitional regions due to potentially greater animal abundance in nearby agricultural areas and higher occupational risk. By documenting these patterns in areas of high disease richness, we were able to move beyond their contribution to individual diseases and establish their impact as a shared mechanism of risk in a large, holistic, multi-zoonotic system.

While many zoonotic infections are the result of direct contact between humans and infected animals, spillover also occurs indirectly through contaminated water or food [13, 34, 78, 104], raising questions about the spatial scale of indirect transmission. We believe that indirect infection of zoonoses in the South Caucasus occurs on a relatively local scale given the overlap between

human and animal risk, and that we did not observe human risk expanding outward spatially from areas of high animal risk. We also observed moderate overlap between human and host-vector risk for anthrax and tularemia, and high overlap for TBE, all of which can be transmitted indirectly. In addition, for most settlements in the region the host-vector risk was elevated for one or more pathogens, meaning that a local source of infection was typically close by. The proportion of cases with an occupational risk was high in hotspots, suggesting direct infection, and the diversification of labor within households (tending crops, livestock, poultry and bees means) means that those in rural settlements without occupational risk are likely to share a household with someone who does [105]. This is supported by the isolation of rural communities that may prevent infected products from being transported long distances, and the water-borne outbreaks caused by local water sources [13, 106]. This implies that indirect transmission acts on a similar scale as direct transmission, and that the risk landscape of indirect transmission is either similar to, or interconnected with, the risk landscape leading to direct transmission.

The relationship between human and animal disease risk presents an interesting contradiction in our results. We expected human disease risk to represent a subset of animal risk, as spillover requires the presence of the pathogen in reservoir or host species [5]. This pattern should be especially pronounced in vector-borne diseases, where transmission typically occurs through direct contact [99]. However, we observed a more complex spatial relationship when comparing risk overlap. We observed high human risk for CCHF, TBE, leptospirosis, and tularemia in areas with low animal risk. This was particularly reflected in the low Schoener's D values for brucellosis and leptospirosis. These patterns could reflect true differences in transmission dynamics across the landscape. Areas with high human risk but low animal risk may experience increased transmission rates due to local variation in host-human contact rates, environmental conditions favoring pathogen survival not reflected in host infection, landscape patterns that concentrate human interactions with a relatively smaller number of infected animals, or behavior differences that increase indirect transmission or increase human contact with a relatively small number of infected animals. However, the areas of high human risk and low animal risk occurred in transition and agricultural

zones with elevated host and vector likelihood. Hotspots of human disease were also characterized by increased contact through occupational risk. It is therefore more likely that these patterns are the result of undersampling or underdetection of animal positives rather than ecological or epidemiological differences within these areas.

Like animal samples, under reporting of human cases was inevitable as a number of our infections present with non-specific symptoms or can be asymptomatic [107]. This is a particular risk in rural areas where occupational risk is high, access to healthcare and diagnosis is limited, and health seeking behaviors may be impacted by socioeconomic factors. Environmental sampling in the South Caucasus in the post-Soviet era, like many regions, has also typically been reactive, triggered by human cases rather than systematic surveillance. While the majority of our results are consistent with the ecological and epidemiological literature, we believe that a small portion, the areas of elevated risk for humans or host-vectors only, reflect sampling limitations.

Future systematic, representative sampling and surveillance efforts in these areas will clarify these patterns. Focusing sampling efforts on hotspots can also help us learn about multiple pathogens simultaneously using a combined campaign approach that is more efficient than investigating each pathogen individually. In addition to targeting hotspots for efficient sampling and surveillance, we believe that intervention efforts should be similarly focused on the local level, as the majority of indirect transmission is likely occurring at small geographic scales.

We also encountered data gaps, as we did not have access to leptospirosis environmental samples in Georgia and TBE human cases in Armenia. While this could limit our models, our overall risk predictions largely align with known natural foci [9] and previously documented patterns of environmental risk. Similarly, despite the dependence of host-vector and environmental sampling on human cases, we believe that the size and scale of our dataset, and the consistency with our results and previous knowledge, allowed us to create an accurate picture of host-vector risk in the region.

This study provides a foundation for future research and surveillance efforts in the South Caucasus while demonstrating the utility of disease richness as a framework for identifying zoonotic

hotspots. This approach enabled us to look beyond individual risk factors and identify shared mechanisms of risk that provide a holistic picture of zoonotic threats to human and animal health in the region. By taking a multi-pathogen approach to One Health we were able to broaden our understanding of the system as a whole while giving practical guidance for effective use of resources for public health interventions.

Chapter 5

Conclusion

This dissertation advances our understanding of zoonotic disease ecology through a comprehensive analysis of host-pathogen dynamics in the South Caucasus region. By leveraging decades of surveillance data and applying methodological approaches in new ways, we have demonstrated how species distribution modeling and intra- and inter-disease patterns can inform our understanding of disease risk and transmission. Our findings contribute both methodological insights for our field and SDM research and practical implications for public health interventions.

Our methodological contributions begin with recommendations for species distribution modeling specifically within disease ecology contexts. The BART model, though relatively new to ecology, proved effective at handling the complex, non-linear relationships we examined. Its tree-based approach, combined with explicit spatial uncertainty measures, makes it well-suited for predictive SDMs and modeling complex systems. We found that random background points typically generated predictions with better discrimination ability compared to target-group background points, while maintaining biological accuracy. This approach offers practical advantages as it requires no additional species data, is simple to implement, and represents the full environmental space of the study area. We also demonstrated the value of using both out-of-sample AUC and Boyce index as metrics for evaluation, as they provides a more comprehensive assessment of model performance compared to each individually. This is again important for disease ecology where discrimination as well as biological realism are required for our research goals. These methodological insights can help guild the development of species distribution predictions for applications in disease ecology.

The application of these methods to tularemia revealed complex spatial patterns in disease manifestation and transmission. We found distinct geographical distributions of clinical manifestations across countries, suggesting either different natural foci or ecology of *F. tularensis*, or systematic differences in diagnosis and surveillance. The spatial overlap between host-vector risk and different clinical manifestations provided insights into transmission pathways, while the con-

sistent importance of the built environment highlighted the role of urban-rural interfaces in disease transmission. These findings demonstrate how fine-scale species distribution models can reveal nuanced patterns of disease risk that might be overlooked in broader-scale analyses.

Our approach to identifying disease hotspots through pathogen richness provides a framework for understanding multi-pathogen transmission dynamics. Our results reveal several disease spillover mechanisms to be shared drivers of transmission across multiple pathogens. We found that both disease richness and occupational exposure were elevated in transitional areas compared to non-transition, extending beyond single-pathogen frameworks to demonstrate how these zones facilitate spillover across multiple diseases [6]. Similarly, the expected higher diversity of hosts and vectors in ecotones emerges as a system-wide phenomenon [26, 80], suggesting that direct contact with hosts and vectors plays a central role in disease transmission within our system.

Our findings reveal how shared mechanisms drive both direct and indirect transmission of multiple zoonoses at a local scale in the South Caucasus. The spatial correlation between human and animal risk, coupled with the high proportion of occupational exposures in disease hotspots, suggests that direct and indirect transmission pathways operate within the same local risk landscape. In these areas, increased presence of host and vector species creates opportunities for multi-pathogen transmission to both domestic animals and humans through direct contact (bites or exposure to infected wild animals) or indirect contact (contaminated food, water, or fomites). Rural agricultural practices also create shared pathways for multiple zoonotic pathogens, as livestock can transmit diseases to humans through both direct contact and contaminated animal products, highlighting the connection between environmental, animal, and human health. This dynamic exemplifies the One Health principle that environmental, animal, and human health are fundamentally interconnected, but moves beyond the typical consideration of single-pathogen systems. The disease richness approach to One Health offers practical advantages for optimizing both surveillance efforts and interventions. Studying these hotspots helps us better understand shared risk factors which can be targeted to simultaneously reduce the risk of multiple zoonoses.

Under this framework we believe the most effective interventions target two shared threats to human health in the South Caucasus: infections in livestock and infections in rodents. For livestock, we recommend vaccination campaigns against anthrax, brucellosis, and CCHF, while for rodents, we propose control measures such as physical barriers or habitat modification to mitigate leptospirosis, plague, and tularemia. These interventions address shared drivers of disease transmission across multiple pathogens, benefiting both human and animal health. By improving livestock health, these measures not only protect human populations, but also reduce economic losses in this agriculturally dependent region. Unlike daily preventive measures, such as wearing protective clothing or using insect repellents, these periodic, large-scale interventions require less individual compliance and commitment.

The broader implications of this work extend beyond the South Caucasus region to inform approaches to zoonotic disease surveillance and control. Our methodological recommendations for species distribution modeling can be applied to other geographic areas and disease systems. The identification of urban-rural interfaces as areas of elevated disease transmission highlights the importance of considering landscape patterns in disease risk assessment. This is particularly relevant for similar areas where agriculture is widespread.

We recommend several steps to better understand zoonoses in the South Caucasus. First, expanding our understanding of the shared risk mechanism of host and vector species by applying our SDM methodology to medically-relevant fleas, mosquitoes, and bats [108]. This information would contribute to our knowledge of the zoonoses described in this work [76] as well as the heterogeneous risk landscape. Second, including additional zoonotic diseases found in the South Caucasus, such as leishmaniasis [109, 110] and Q fever [111], would continue to develop the picture of disease richness and the shared mechanisms of spillover. Finally, we recommend applying the disease richness approach to other regions. We expect that applying this methodology to other highly agricultural regions would show that the shared risk mechanisms we identified apply to other ecologically similar but geographically disparate areas. It would also be beneficial to apply this framework to regions and systems with entirely different suspected risk factors. For example,

areas where Water, Sanitation, and Hygiene (WASH) diseases are the primary causes of morbidity and mortality such as the flood-prone regions of India, Bangladesh, and Southeast Asia. This would build on the literature of shared risk mechanisms for different disease archetypes than we have examined in our work.

Bibliography

[1] L. H. Taylor, S. M. Latham, and M. E.J. Woolhouse. Risk factors for human disease emergence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 356(1411):983–989, 2001.

[2] Delia Grace, Jeff Gilbert, Thomas Randolph, and Erastus Kang'ethe. The multiple burdens of zoonotic disease and an ecohealth approach to their assessment. *Tropical Animal Health and Production*, 44(SUPPL.1):67–73, 2012.

[3] Ian Maudlin, Mark Charles Eisler, and Susan Christina Welburn. Neglected and endemic zoonoses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364:2777–2787, 2009.

[4] Paul C. Cross, Diann J. Prosser, Andrew M. Ramey, Ephraim M. Hanks, and Kim M. Pepin. Confronting models with data: the challenges of estimating disease spillover. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 374:1–10, 2019.

[5] Raina K. Plowright, Colin R. Parrish, Hamish McCallum, Peter J. Hudson, Albert I. Ko, Andrea L. Graham, and James O. Lloyd-Smith. Pathways to zoonotic spillover. *Nature Reviews Microbiology*, 15:502–510, 2017.

[6] Sara H. Paull, Sejin Song, Katherine M. McClure, Loren C. Sackett, A. Marm Kilpatrick, and Pieter T.J. Johnson. From superspreaders to disease hotspots: Linking transmission across hosts and space. *Frontiers in Ecology and the Environment*, 10(2):75–82, 2012.

[7] Eric F. Lambin, Annelise Tran, Sophie O. Vanwambeke, Catherine Linard, and Valérie Soti. Pathogenic landscapes: Interactions between land, people, disease vectors, and their animal hosts. *International Journal of Health Geographics*, 9(54):1–13, 2010.

[8] Norman Myers, Russell A. Mittermeier, Cristina G Mittermeier, Gustavo A B da Fonseca, and Jennifer Kent. Biodiversity hotspots for conservation priorities. *Nature*, 403:853–858, 2000.

[9] Michael Kosoy, Paata Imnadze, Lile Malania, Nana Bolashvili, Andrei Kandaurov, Colleen T. Webb, and Kendra Gilbertson, editors. *Atlas of Zoonotic Diseases in the South Caucasus*, volume 1. Atlas of Zoonotic Diseases in the South Caucasus, 2024.

[10] Ana Kasradze, Diana Echeverria, Khatuna Zakhashvili, Christian Bautista, Nicholas Heyer, Paata Imnadze, and Veriko Mitrskhulava. Rates and risk factors for human cutaneous anthrax in the country of Georgia: National surveillance data, 2008–2015. *PLoS ONE*, 13(2):e0192031, 2018.

[11] Ian T. Kracalik, Lile Malania, Nikoloz Tsertsvadze, Julietta Manvelyan, Lela Bakanidze, Paata Imnadze, Shota Tsanava, and Jason K. Blackburn. Evidence of Local Persistence of Human Anthrax in the Country of Georgia Associated with Environmental and Anthropogenic Factors. *PLoS Neglected Tropical Diseases*, 7(9):e2388, 2013.

[12] S. Melikjanyan, A. Vanyan, A. Avetisyan, A. Avetisyan, and N. Bakunts. GIS Analysis of Tularemia Outbreaks in Armenia, 1996-2013. *Online Journal of Public Health Informatics*, 6(1), 2014.

[13] N. Chitadze, T. Kuchuloria, D. V. Clark, E. Tsertsvadze, M. Chokheli, N. Tsertsvadze, N. Trapaidze, A. Lane, L. Bakanidze, S. Tsanava, M. J. Hepburn, and P. Imnadze. Waterborne outbreak of oropharyngeal and glandular tularemia in Georgia: Investigation and follow-up. *Infection*, 37(6):514–521, 2009.

[14] Danielle V. Clark, Afrail Ismailov, Esmiralda Seyidova, Ayten Hajiyeva, Sevinj Bakhishova, Huseyn Hajiyev, Tahir Nuriyev, Saleh Piraliyev, Sadigulla Bagirov, Afag Aslanova, Amanda K. Debes, Maqsud Qasimov, and Matthew J. Hepburn. Seroprevalence of Tularemia in Rural Azerbaijan. *Vector-Borne and Zoonotic Diseases*, 12, 2012.

[15] L. Babayan, A. Manucharyan, L. Paronyan, H. Vardanyan, R. Danielyan, G. Melik-Andreasyan, and J.E. Achenbach. Distribution of vectors and arboviruses, and healthcare workers' knowledge of vector-borne diseases in armenia. *Journal of Infection in Developing Countries*, 18:1442–1449, 2024.

[16] Angela D. Luis, David T.S. Hayman, Thomas J. O'Shea, Paul M. Cryan, Amy T. Gilbert, Juliet R.C. Pulliam, James N. Mills, Mary E. Timonin, Craig K.R. Willis, Andrew A. Cunningham, Anthony R. Fooks, Charles E. Rupprecht, James L.N. Wood, and Colleen T. Webb. A comparison of bats and rodents as reservoirs of zoonotic viruses: Are bats special? *Proceedings of the Royal Society B: Biological Sciences*, 280:1–9, 2013.

[17] Barbara A Han, Andrew M Kramer, and John M Drake. Global patterns of zoonotic disease in mammals. *Trends in Parasitology*, 32(7):565–577, jul 2016.

[18] Barbara A. Han, John Paul Schmidt, Sarah E. Bowden, and John M. Drake. Rodent reservoirs of future zoonotic diseases. *PNAS*, 112(22):7039–7044, 2015.

[19] Jeannine M. Petersen and Martin E. Schriefer. Tularemia: emergence/re-emergence. *Veterinary Research*, 36:455–467, 2005.

[20] Paul S. Keim and David M. Wagner. Humans and evolutionary and ecological forces shaped the phylogeography of recently emerged diseases. *Nature Reviews Microbiology*, 7:813–821, 2009.

[21] G. Hestvik, E. Warns-Petit, L. A. Smith, N. J. Fox, H. Uhlhorn, M. Artois, D. Hannant, M. R. Hutchings, R. Mattsson, L. Yon, and D. Gavier-Widen. The status of tularemia in Europe in a one-health context: a review. *Epidemiology and infection*, 143:2137–2160, 2015.

[22] Anders Sjöstedt. Tularemia: History, Epidemiology, Pathogen Physiology, and Clinical Manifestations. *Annals of the New York Academy of Sciences*, 1105:1–29, 2007.

[23] David T. Dennis, Thomas V. Inglesby, Donald A. Henderson, John G. Bartlett, Michael S. Ascher, Edward Eitzen, Anne D. Fine, Arthur M. Friedlander, Jerome Hauer, Marcelle Layton, Scott R. Lillibridge, Joseph E. McDade, Michael T. Osterholm, Tara ÖToole, Gerald Parker, Trish M. Perl, Philip K. Russell, and Kevin Tonat. Tularemia as a Biological Weapon: Medical and Public Health Management. *Journal of the American Medical Association*, 285(21):2763–2773, 2001.

[24] Katherine Anne Feldman. Zoonosis Update - Tularemia. *Journal of the American Veterinary Medical Association*, 222(6):725–730, 2003.

[25] Paul Rega, Michael Guinness, and Christopher McMahon. Tularemia – A Review with Concern for Bioterrorism. *Medical Research Archives*, 5(8):1–15, 2017.

[26] Dickson Despommier, Brett R. Ellis, and Bruce A. Wilcox. The role of ecotones in emerging infectious diseases. *EcoHealth*, 3:281–289, 2006.

[27] James O Lloyd-Smith, Dylan George, Kim M Pepin, Virginia E Pitzer, Juliet R C Pulliam, Andrew P Dobson, Peter J Hudson, and Bryan T Grenfell. Epidemic Dynamics at the Interface, Human-animal. *Science*, 326(December):1362–1368, 2009.

[28] Gurutzeta Guillera-Arroita, José J. Lahoz-Monfort, Jane Elith, Ascelin Gordon, Heini Kujala, Pia E. Lentini, Michael A. Mccarthy, Reid Tingley, and Brendan A. Wintle. Is my species distribution model fit for purpose? Matching data and models to applications. *Global Ecology and Biogeography*, 24(3):276–292, 2015.

[29] Colin M. Beale and Jack J. Lennon. Incorporating uncertainty in predictive species distribution modelling. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367:247–258, 2012.

[30] Robert A. Barber, Stuart G. Ball, Roger K.A. Morris, and Francis Gilbert. Target-group backgrounds prove effective at correcting sampling bias in Maxent models. *Diversity and Distributions*, 28:128–141, 2022.

[31] Julien Vollering, Rune Halvorsen, Inger Auestad, and Knut Rydgren. Bunching up the background betters bias in species distribution models. *Ecography*, 42:1–11, 2019.

[32] Valerie A. Steen, Morgan W. Tingley, Peter W.C. Paton, and Chris S. Elphick. Spatial thinning and class balancing: Key choices lead to variation in the performance of species distribution models with citizen science data. *Methods in Ecology and Evolution*, 12:216–226, 2021.

[33] Steven J. Phillips, Miroslav Dudík, Jane Elith, Catherine H. Graham, Anthony Lehmann, John Leathwick, and Simon Ferrier. Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data. *Ecological Applications*, 19(1):181–197, 2009.

[34] Ian Kracalik, Rakif Abdullayev, Kliment Asadov, Rita Ismayilova, Mehriban Baghirova, Narmin Ustun, Mazahir Shikhiyev, Aydin Talibzade, and Jason K. Blackburn. Changing Patterns of Human Anthrax in Azerbaijan during the Post-Soviet and Preemptive Livestock Vaccination Eras. *PLoS Neglected Tropical Diseases*, 8(7), 2014.

[35] Zdenek Hubálek and Ivo Rudolf. Vertebrates as Hosts and Reservoirs of Zoonotic Microbial Agents. In *Microbial Zoonoses and Sapronoses*, pages 1–457. 2011.

[36] Felicia Keesing, Lisa K. Belden, Peter Daszak, Andrew Dobson, C. Drew Harvell, Robert D. Holt, Peter Hudson, Anna Jolles, Kate E. Jones, Charles E. Mitchell, Samuel S. Myers, Tiffany Bogich, and Richard S. Ostfeld. Impacts of biodiversity on the emergence and transmission of infectious diseases. *Nature*, 468(7324):647–652, 2010.

[37] Daniel J. Salkeld, Kerry A. Padgett, and James Holland Jones. A meta-analysis suggesting that the relationship between biodiversity and risk of zoonotic pathogen transmission is idiosyncratic. *Ecology Letters*, 16(5):679–686, 2013.

[38] Hugh A. Chipman, Edward I. George, and Robert E. McCulloch. BART: Bayesian additive regression trees. *Annals of Applied Statistics*, 4(1):266–298, 2010.

[39] Colin J. Carlson, Ian T. Kracalik, Noam Ross, Kathleen A. Alexander, Martin E. Hugh-Jones, Mark Fegan, Brett T. Elkin, Tasha Epp, Todd K. Shury, Wenyi Zhang, Mehriban Bagirova, Wayne M. Getz, and Jason K. Blackburn. The global distribution of Bacillus anthracis and associated anthrax risk to humans, livestock and wildlife. *Nature Microbiology*, 4:1337–1343, 2019.

[40] M. S. Boyce, P. R. Vernier, S. E. Nielsen, and F. K.A. Schmiegelow. Evaluating resource selection functions. *Ecological Modelling*, 157:281–300, 2002.

[41] Jorge M. Lobo, Alberto Jiménez-valverde, and Raimundo Real. AUC: A misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, 17:145–151, 2008.

[42] Dan L. Warren, Nicholas J. Matzke, and Teresa L. Iglesias. Evaluating presence-only species distribution models with discrimination accuracy is uninformative for many applications. *Journal of Biogeography*, 47:167–180, 2020.

[43] Alexandre H. Hirzel, Gwenaëlle Le Lay, Véronique Helfer, Christophe Randin, and Antoine Guisan. Evaluating the ability of habitat suitability models to predict species presences. *Ecological Modelling*, 199(2):142–152, 2006.

[44] Thomas W. Schoener. The Anolis Lizards of Bimini : Resource Partitioning in a Complex Fauna. *Ecology*, 49(4):704–726, 1968.

[45] S.E. Fick and R.J. Hijmans. Worldclim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 2017.

[46] Craig Moritz, James L. Patton, Chris J. Conroy, Juan L. Parra, Gary C. White, and Steven R. Beissinger. Impact of a century of climate change on small-mammal communities in Yosemite National Park, USA. *Science*, 322:261–264, 2008.

[47] James W. Pearce-Higgins, Nancy Ockendon, David J. Baker, Jamie Carr, Elizabeth C. White, Rosamunde E.A. Almond, Tatsuya Amano, Esther Bertram, Richard B. Bradbury,

Cassie Bradley, Stuart H.M. Butchart, Nathalie Doswald, Wendy Foden, David J.C. Gill, Rhys E. Green, William J. Sutherland, and Edmund V.J. Tanner. Geographical variation in species' population responses to changes in temperature and precipitation. *Proceedings of the Royal Society B: Biological Sciences*, 282, 2015.

[48] Mohammad Hasan Rabiee, Ahmad Mahmoudi, Roohollah Siahsarvie, Boris Kryštufek, and Ehsan Mostafavi. Rodent-borne diseases and their public health importance in Iran. *PLoS Neglected Tropical Diseases*, 12(4):1–20, 2018.

[49] Ehsan Mostafavi, Ahmad Ghasemi, Mahdi Rohani, Leila Molaeipoor, Saber Esmaeili, Zeinolabedin Mohammadi, Ahmad Mahmoudi, Mansour Aliabadian, and Anders Johansson. Molecular Survey of Tularemia and Plague in Small Mammals from Iran. *Frontiers in Cellular and Infection Microbiology*, 8(215), 2018.

[50] Mustafa Ekici, Adem Keskin, Ahmet Bursali, and Saban Tekin. Investigation of Crimean Congo Hemorrhagic Fever Virus in Ticks (Acari: Ixodidae) Infesting on Hedgehogs in Turkey. *Journal of New Results in Science*, 2:31–38, 2013.

[51] Miklós Gyuranecz, Jenő Reiczigel, Katalin Krisztalovics, László Monse, Gabriella Kükedi Szabóné, Andrásné Szilágyi, Bálint Szépe, László Makrai, Tibor Magyar, Mangesh Bhide, and Károly Erdélyi. Factors influencing emergence of tularemia, Hungary, 1984-2010. *Emerging Infectious Diseases*, 18(8):1379–1381, 2012.

[52] Mohammad Reza Yaghoobi-Ershadi, Reza Jafari, and Ahmad Ali Hanafi-Bojd. A new epidemic focus of zoonotic cutaneous leishmaniasis in central Iran. *Annals of Saudi Medicine*, 24(2):98–101, 2004.

[53] B. Pourmohammadi, M. H. Motazedian, and M. Kalantari. Rodent infection with Leishmania in a new focus of human cutaneous leishmaniasis, in northern Iran. *Annals of Tropical Medicine and Parasitology*, 102(2):127–133, 2008.

[54] Majid Hemati, Mohammad Khalili, Mahdi Rohani, Balal Sadeghi, Saber Esmaeili, Ahmad Ghasemi, Ahmad Mahmoudi, Miklós Gyuranecz, and Ehsan Mostafavi. A serological and molecular study on Francisella tularensis in rodents from Hamadan province, Western Iran. *Comparative Immunology, Microbiology and Infectious Diseases*, 68, 2020.

[55] Juan José Luque-Larena, François Mougeot, Dolors Vidal Roig, Xavier Lambin, Ruth Rodríguez-Pastor, Elena Rodríguez-Valín, Pedro Anda, and Raquel Escudero. Tularemia Outbreaks and Common Vole (Microtus arvalis) Irruptive Population Dynamics in Northwestern Spain, 1997-2014. *Vector-Borne and Zoonotic Diseases*, 15(9):568–570, 2015.

[56] E. Fallah, M. Farshchian, A. Mazlomi, J. Majidi, A. Kusha, and A. Mardi. Study on the prevalence of visceral leishmaniasis in rodent's of Azarshahr district (new focus), northwest of Iran. *Archives of Razi Institute*, 61(1):27–33, 2006.

[57] J. Saki and S. Khademvatan. Detection of Toxoplasma gondii by PCR and mouse bioassay in rodents of Ahvaz District, Southwestern Iran. *BioMed Research International*, 2014, 2014.

[58] Mohammad Hossein Motazedian, Masoumeh Parhizkari, Davood Mehrabani, Gholamreza Hatam, and Qasem Asgari. First detection of leishmania major in Rattus norvegicus from Fars province, Southern Iran. *Vector-Borne and Zoonotic Diseases*, 10(10):969–975, 2010.

[59] S Saidi, J Casals, and M A Faghih. Crimean Hemorrhagic Fever-Congo (CHF-C) Virus Antibodies in Man, and in Domestic and Small Mammals, in Iran. *The American Journal of Tropical Medicine and Hygiene*, 24(2), 1975.

[60] Pierre Dutilleul, Peter Clifford, Sylvia Richardson, and Denis Hemon. Modifying the t Test for Assessing the Correlation Between Two Spatial Processes. *Biometrics*, 49(1):305–314, 1993.

[61] Ş. K. Çorbacıoğlu and G. Aksel. Receiver operating characteristic curve analysis in diagnostic accuracy studies: A guide to interpreting the area under the curve value. *Turkish Journal of Emergency Medicine*, 23(4):195–198, 2023.

[62] Haldun Akoglu. User's guide to correlation coefficients. *Turkish Journal of Emergency Medicine*, 18:91–93, 2018.

[63] S. A. Levin. The problem of pattern and scale in ecology. *Ecology*, 73(6):1943–1967, 1992.

[64] Red List Technical Working Group. Mapping standards and data quality for the iucn red list categories and criteria, version 1.16. Prepared by the Red List Technical Working Group, September 2018. Accessed: [Insert Access Date].

[65] K Hacklander and S Schai-Braun. Lepus europaeus. *The IUCN Red List of Threatened Species 2019*, 2019.

[66] M. Sozen, A. Bukhnikashvili, G. Shenbrot, D. Scott, G. Amori, B. Kryštufek, N. Yigit, and G. Mitsainas. Meriones tristrami (amended version of 2016 assessment). *The IUCN Red List of Threatened Species 2021*, 2021.

[67] L Granjon. Meriones libycus. *The IUCN Red List of Threatened Species 2016*, 2016.

[68] G. Shenbrot. Microtus schidlovskii. *The IUCN Red List of Threatened Species 2016*, 2016.

[69] Gregory F. Albery, Amy R. Sweeny, Daniel J. Becker, and Shweta Bansal. Fine-scale spatial patterns of wildlife disease are common and understudied. *Functional Ecology*, 36:214–225, 2022.

[70] Robin Loveridge, Oliver R Wearn, Marcus Vieira, Henry Bernard, and Robert M. Ewers. Movement Behavior of Native and Invasive Small Mammals Shows Logging May Facilitate Invasion in a Tropical Rain Forest. *Biotropica*, 48(3):373–380, 2016.

[71] Kathleen A. Alexander, Colin J. Carlson, Bryan L. Lewis, Wayne M. Getz, Madhav V. Marathe, Stephen G. Eubank, Claire E. Sanderson, and Jason K. Blackburn. The Ecology of

Pathogen Spillover and Disease Emergence at the Human-Wildlife-Environment Interface. In *The Connections Between Ecology and Infectious Disease*, pages 267–298. 2018.

[72] Daniel J. Becker, Alex D. Washburne, Christina L. Faust, Erin A. Mordecai, and Raina K. Plowright. The problem of scale in the prediction and management of pathogen spillover. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 374:1–9, 2019.

[73] Aurélie Hennebique, Sandrine Boisset, and Max Maurin. Tularemia as a waterborne disease: a review. *Emerging Microbes and Infections*, 8(1):1027–1042, 2019.

[74] Henrik Eliasson, Tina Broman, Mats Forsman, and Erik Bäck. Tularemia: Current Epidemiology and Disease Management. *Infectious Disease Clinics of North America*, 20:289–311, 2006.

[75] Halis Akalin, Safiye Helvaci, and Suna Gedikoğlu. Re-emergence of tularemia in Turkey. *International Journal of Infectious Diseases*, 13:547–551, 2009.

[76] Paul Keim, Anders Johansson, and David M. Wagner. Molecular Epidemiology, Evolution, and Ecology of Francisella. *Annals of the New York Academy of Sciences*, 1105:30–66, 2007.

[77] Eka Elashvili, Ian Kracalik, Irma Burjanadze, Sophio Datukishvili, Gvantsa Chanturia, Nikoloz Tsertsvadze, Levan Beridze, Merab Shavishvili, Archil Dzneladze, Marina Grdzelidze, Paata Imnadze, Andrew Pearson, and Jason K. Blackburn. Environmental monitoring and surveillance of rodents and vectors for Francisella tularensis following outbreaks of human tularemia in Georgia. *Vector-Borne and Zoonotic Diseases*, 15(10):633–636, 2015.

[78] Syuzanna Melikjanyan, Karo Palayan, Artavazd Vanyan, Lilit Avetisyan, Nune Bakunts, Marine Kotanyan, and Marta Guerra. Human cases of tularemia in Armenia, 1996-2012. *American Journal of Tropical Medicine and Hygiene*, 97(3):819–825, 2017.

[79] N. Akhvlediani, I. Burjanadze, D. Baliashvili, T. Tushishvili, M. Broladze, A. Navdarashvili, S. Dolbadze, N. Chitadze, M. Topuridze, P. Imnadze, N. Kazakhashvili, T. Tsertsvadze, T. Kuchuloria, T. Akhvlediani, L. A. McNutt, and G. Chanturia. Tularemia transmission to humans: A multifaceted surveillance approach. *Epidemiology and Infection*, 146:2139–2145, 2018.

[80] Gabriel E. García-Peña and André V. Rubio. Unveiling the impacts of land use on the phylogeography of zoonotic New World Hantaviruses. *Ecography*, page e06996, 2024.

[81] Şaban Gürcan. Epidemiology of tularemia. *Balkan Medical Journal*, 31:3–10, 2014.

[82] Michael L. McKinney. Urbanization, Biodiversity, and Conservation. *BioScience*, 52(10):883–890, 2002.

[83] D. C. Dragon and R. P. Rennie. The ecology of anthrax spores: Tough but not invincible. *Canadian Veterinary Journal*, 36:295–301, 1995.

[84] P.C.B. Turnbull. Introduction : Anthrax History, Disease and Ecology. *Current Topics in Microbiology and Immunology*, 271:1–19, 2002.

[85] Colin J. Carlson, Wayne M. Getz, Kyrre L. Kausrud, Carrie A. Cizauskas, Jason K. Blackburn, Fausto A. Bustos Carrillo, Rita Colwell, W. Ryan Easterday, Holly H. Ganz, Pauline L. Kamath, Ole A. Økstad, Wendy C. Turner, Anne Brit Kolstø, and Nils C. Stenseth. Spores and soil from six sides: interdisciplinarity and the environmental biology of anthrax (Bacillus anthracis). *Biological Reviews*, 93:1813–1831, 2018.

[86] R. M. Atlas. Responding to the threat of bioterrorism: A microbial ecology perspective - The case of anthrax. *International Microbiology*, 5(4):161–167, 2002.

[87] Michael J. Corbel. Brucellosis: An Overview. *Emerging Infectious Diseases*, 3(2):213–221, 1997.

[88] Mohamed N. Seleem, Stephen M. Boyle, and Nammalwar Sriranganathan. Brucellosis: A re-emerging zoonosis. *Veterinary Microbiology*, 140:392–398, 2010.

[89] Tamar Akhvlediani, Christian T. Bautista, Natalia Garuchava, Lia Sanodze, Nora Kokaia, Lile Malania, Nazibrola Chitadze, Ketevan Sidamonidze, Robert G. Rivard, Matthew J. Hepburn, Mikeljon P. Nikolich, Paata Imnadze, and Nino Trapaidze. Epidemiological and clinical features of brucellosis in the country of Georgia. *PLoS ONE*, 12(1):1–12, 2017.

[90] Chris A. Whitehouse. Crimean-Congo hemorrhagic fever. *Antiviral Research*, 64:145–160, 2004.

[91] Hasmik Gevorgyan, Gohar G. Grigoryan, Hripsime A. Atoyan, Martin Rukhkyan, Astghik Hakobyan, Hovakim Zakaryan, and Sargis A. Aghayan. Evidence of Crimean-Congo haemorrhagic fever virus occurrence in Ixodidae ticks of Armenia. *Journal of Arthropod-Borne Diseases*, 13(1):9–16, 2019.

[92] Ajay R Bharti, Jarlath E Nally, Jessica N Ricaldi, Michael A Matthias, Monica M Diaz, Michael A Lovett, Paul N Levett, Robert H Gilman, Michael R Willig, Eduardo Gotuzzo, and Joseph M Vinetz. Leptospirosis: a zoonotic disease of global importance. *The Lancet Infectious Diseases*, 3:757–771, 2003.

[93] M. Picardeau. Diagnosis and epidemiology of leptospirosis. *Medecine et Maladies Infectieuses*, 43:1–9, 2013.

[94] Nils Chr Stenseth, Bakyt B. Atshabar, Mike Begon, Steven R. Belmain, Eric Bertherat, Elisabeth Carniel, Kenneth L. Gage, Herwig Leirs, and Lila Rahalison. Plague: Past, present, and future. *PLoS Medicine*, 5(1):0009–0013, 2008.

[95] Didier Raoult, Nadjet Mouffok, Idir Bitam, Renaud Piarroux, and Michel Drancourt. Plague: History and contemporary analysis. *Journal of Infection*, 66:18–26, 2013.

[96] Lars Lindquist and Olli Vapalahti. Tick-borne encephalitis. *The Lancet*, 371:1861–1871, 2008.

[97] T.S. Gritsun, V.A. Lashkevich, and E.A. Gould. Tick-borne encephalitis. *Antiviral Research*, 57:129–146, 2003.

[98] Uga Dumpis, Derrick Crook, and Jarmo Oksi. Tick-Borne Encephalitis. *Clinical Infectious Diseases*, 28(4):882–890, 1999.

[99] Mark E.J. Woolhouse, Louise H. Taylor, and Daniel T. Haydon. Population biology of multihost pathogens. *Science*, 292:1109–1112, 2001.

[100] Justin Lessler, Andrew S. Azman, Heather S. McKay, and Sean M. Moore. Perspective piece: What is a hotspot anyway? *American Journal of Tropical Medicine and Hygiene*, 96(6):1270–1273, 2017.

[101] T. Allen, K.A. Murray, C. Zambrana-Torrelio, S.S. Morse, C. Rondinini, M. Di Marco, N. Breit, K.J. Olival, and P. Daszak. Global hotspots and correlates of emerging zoonotic diseases. *Nature Communications*, 8(1124), 2017.

[102] Sara Mclafferty. Disease cluster detection methods: recent developments and public health implications. *Annals of GIS*, 21(2):127–133, 2015.

[103] Trisalyn A. Nelson and Barry Boots. Detecting spatial hot spots in landscape ecology. *Ecography*, 31:556–566, 2008.

[104] Rita Ismayilova, Natiq Xalilov, and Masud Khatibi. A "One Health" Surveillance and Control of Brucellosis in Azerbaijan. *One Health I& Risk Management*, 2023.

[105] Regina Neudert, Naiba Allahverdiyeva, Niyaz Mammadov, Alexandre Didebulidze, and Volker Beckmann. Diversification of livestock-keeping smallholders in mountainous rural regions of Azerbaijan and Georgia. *Land*, 9(267):1–23, 2020.

[106] George Welton, Armen A. Asatryan, and David Jijelava. COMPARATIVE ANALYSIS OF AGRICULTURE IN THE SOUTH CAUCASUS. Technical report, United Nations Development Programme, 2013.

[107] Cheryl L. Gibbons, Marie Josée J. Mangen, Dietrich Plass, Arie H. Havelaar, Russell John Brooke, Piotr Kramarz, Karen L. Peterson, Anke L. Stuurman, Alessandro Cassini, Eric M. Fèvre, and Mirjam EE Kretzschmar. Measuring underreporting and under-ascertainment in infectious disease datasets: A comparison of methods. *BMC Public Health*, 14(147):1–17, 2014.

[108] Ying Bai, Lela Urushadze, Lynn Osikowicz, Clifton McKee, Ivan Kuzmin, Andrei Kandaurov, Giorgi Babuadze, Ioseb Natradze, Paata Imnadze, and Michael Kosoy. Molecular survey of bacterial zoonotic agents in bats from the country of Georgia (Caucasus). *PLoS ONE*, 12(1):1–12, 2017.

[109] Margarita V. Strelkova, Evgeny N. Ponirovsky, Evgeny N. Morozov, Ekaterina N. Zhirenkina, Shavkat A. Razakov, Dmitriy A. Kovalenko, Lionel F. Schnur, and Gabriele Schönian. A narrative review of visceral leishmaniasis in Armenia, Azerbaijan, Georgia, Kazakhstan, Kyrgyzstan, Tajikistan, Turkmenistan, Uzbekistan, the Crimean Peninsula and Southern Russia. *Parasites and Vectors*, 8(330):1–18, 2015.

[110] Yusuf Özbel, Seray Töz, Clara Muñoz, Maria Ortuño, Zarima Jumakanova, Pedro Pérez-Cutillas, Carla Maia, Cláudia Conceição, Gad Baneth, André Pereira, Yves Van der Stede, Céline M. Gossner, and Eduardo Berriatua. The current epidemiology of leishmaniasis in Turkey, Azerbaijan and Georgia and implications for disease emergence in European countries. *Zoonoses and Public Health*, 69:395–407, 2022.

[111] Lusine Paronyan, Eduard Zardaryan, Vahe Bakunts, Zaruhi Gevorgyan, Vigen Asoyan, Hripsime Apresyan, Alvard Hovhannisyan, Karo Palayan, Christian T. Bautista, Tinatin Kuchuloria, and Robert G. Rivard. A retrospective chart review study to describe selected zoonotic and arboviral etiologies in hospitalized febrile patients in the Republic of Armenia. *BMC Infectious Diseases*, 16:1–8, 2016.

[112] O. Venter, E. W. Sanderson, A. Magrach, J. R. Allan, J. Beher, K. R. Jones, H. P. Possingham, W. F. Laurance, P. Wood, B. M. Fekete, M. A. Levy, and J. E. Watson. Sixteen Years

of Change in the Global Terrestrial Human Footprint and Implications for Biodiversity Conservation. *Nature Communications*, 7:12558, 2016.

[113] D. Zanaga, R. Van De Kerchove, W. De Keersmaecker, N. Souverijns, C. Brockmann, R. Quast, J. Wevers, A. Grosu, A. Paccini, S. Vergnaud, O. Cartus, M. Santoro, S. Fritz, I. Georgieva, M. Lesiv, S. Carter, M. Herold, Linlin Li, N.E. Tsendbazar, F. Ramoino, and O. Arino. ESA WorldCover 10 m 2020 v100, 2021.

[114] CIESIN. Columbia University 2018 Gridded population of the world, version 4 (GPWv4): population density, 2018. Accessed 6 July 2021.

[115] L. Poggio, L. M. de Sousa, N. H. Batjes, G. B. M. Heuvelink, B. Kempen, E. Ribeiro, and D. Rossiter. SoilGrids 2.0: Producing Soil Information for the Globe with Quantified Spatial Uncertainty. *Soil*, 7:217–240, 2021.

[116] Andy Nelson. Travel time to cities and ports in the year 2015, 2019. Dataset.

Appendix A

Small Mammal Data

Table A.1: Summary of South Caucasus medically-relevant small mammal species and their sample size. Species indicated with an * are medically relevant in the South Caucasus.

Scientific Name	**Common Name**	**Sample Size**
Apodemus agrarius	Striped field mouse	32
Apodemus hyrcanicus	Hyrcanian field mouse	20
Apodemus mystacinus	Eastern broad-toothed field mouse	107
Apodemus ponticus	Caucasus field mouse	99
**Apodemus uralensis*	Herb field mouse	654
**Apodemus witherbyi*	Steppe field mouse	360
**Arvicola amphibius*	Eurasian water vole	432
Chionomys gud	Gudaur snow vole	142
Chionomys roberti	Robert's snow vole	151
Chionomys syriacus (formerly *Chionomys nivalis*)	European snow vole	216
Clethrionomys glareolus (formerly *Myodes glareolus*)	Bank vole	6
Crocidura gueldenstaedtii	Gueldenstädt's shrew	908
Crocidura leucodon	Bicolored shrew	225
Crocidura zarudnyi	Zarudny's rock shrew	23
Dryomys nitedula	Forest dormouse	271
Ellobius lutescens	Transcaucasian mole vole	91

Scientific Name	**Common Name**	**Sample Size**
**Erinaceus concolor*	Southern white-breasted hedgehog	166
Glis glis	Edible dormouse	249
Hystrix indica	Indian crested porcupine	38
**Lepus europaeus*	European hare	770
Meriones dahli	Dahl's jird	27
**Meriones libycus*	Libyan jird	302
Meriones persicus	Persian jird	107
**Meriones tristrami*	Tristram's jird	122
Meriones vinogradovi	Vinogradov's jird	174
Mesocricetus brandti	Brandt's Hamster	282
Mesocricetus raddei	Ciscaucasian hamster	6
Micromys minutus	Eurasian harvest mouse	18
**Microtus arvalis*	Common vole	1369
Microtus daghestanicus	Daghestan pine vole	246
Microtus majori	Major's pine vole	416
Microtus mystacinus (formerly *Microtus levis*)	East european vole	16
Microtus nasarovi	Nasarov's vole	7
Microtus schidlovskii	Schidlovsky pine vole	133
**Microtus socialis*	Social vole	717
Mus macedonicus	Macedonian mouse	135
**Mus musculus*	House mouse	642
Myocastor coypus	Nutria or Coypu	72

Scientific Name	**Common Name**	**Sample Size**
Nannospalax nehringi (alternate name *Nannospalax xanthodon*)	Nehring's blind mole rat	185
Neomys teres	Transcaucasian water shrew	170
**Nothocricetulus migratorius* (formerly *Cricetulus migratorius*)	Grey dwarf hamster	809
Ondatra zibethicus	Common muskrat	9
Prometheomys schaposchnikowi	Long-clawed mole vole	100
**Rattus norvegicus*	Brown rat	378
**Rattus rattus*	House rat	270
**Scarturus elater* (formerly *Allactaga elater*)	Small five-toed jerboa	184
**Scarturus williamsi* (formerly *Allactaga williamsi*)	Williams' jerboa	163
Sciurus anomalus	Caucasian squirrel	220
Sciurus vulgaris	Eurasian red squirrel	81
Sicista armenica	Armenian birch mouse	10
Sicista caucasica	Caucasian birch mouse	8
Sicista kazbegica	Kazbeg birch mouse	18
Sicista kluchorica	Kluchor birch mouse	13
Sorex raddei	Radde's shrew	84
Sorex satunini	Caucasian shrew	205
Sorex volnuchini	Caucasian pygmy shrew	191
Spermophilus xanthoprymnus	Anatolian ground squirrel	78
Suncus etruscus	Pygmy white-toothed shrew	78
Talpa caucasica	Caucasian mole	168
Talpa levantis	Levant mole	239

Scientific Name	Common Name	Sample Size
Talpa talyschensis	Talysch mole	4

Appendix B

Covariate Layers

Table B.1: Covariate layers used in species distribution and disease modeling.

Name	**Description**	**Source**
BIO1	Annual mean temperature	[45]
BIO2	Mean Diurnal Range (Mean of monthly (max temp - min temp)).	[45]
BIO3	Isothermality (BIO2/BIO7) (×100)	[45]
BIO4	Temperature Seasonality (standard deviation ×100)	[45]
BIO5	Max Temperature of Warmest Month	[45]
BIO6	Min Temperature of Coldest Month	[45]
BIO7	Temperature Annual Range (BIO5-BIO6)	[45]
BIO8	Mean Temperature of Wettest Quarter	[45]
BIO9	Mean Temperature of Driest Quarter	[45]
BIO10	Mean Temperature of Warmest Quarter	[45]
BIO11	Mean Temperature of Coldest Quarter	[45]
BIO12	Annual Precipitation	[45]
BIO13	Precipitation of Wettest Month	[45]
BIO14	Precipitation of Driest Month	[45]
BIO15	Precipitation Seasonality (Coefficient of Variation)	[45]
BIO16	Precipitation of Wettest Quarter	[45]
BIO17	Precipitation of Driest Quarter (This layer was excluded from our analysis because of its perfect correlation to precipitation of driest month, BIO14)	[45]

Name	Description	Source
BIO18	Precipitation of Warmest Quarter	[45]
BIO19	Precipitation of Coldest Quarter	[45]
Wind	Wind speed 2 meters above the ground	[45]
Elevation	Meters above sea level	[45]
Human footprint	Estimated impact of humans on the environment on a scale from 0 (low) to 50 (high) during during the year 1993 at a 30 second spatial resolution	[112]
Trees	30 second spatial resolution data for the portion (0 to 1) of each grid occupied by trees	[113]
Moss	30 second spatial resolution data for the portion (0 to 1) of each grid occupied by moss	[113]
Snow	30 second spatial resolution data for the portion (0 to 1) of each grid occupied by snow	[113]
Grassland	30 second spatial resolution data for the portion (0 to 1) of each grid occupied by grassland	[113]
Shrubs	30 second spatial resolution data for the portion (0 to 1) of each grid occupied by shrubs	[113]
Cropland	30 second spatial resolution data for the portion (0 to 1) of each grid occupied by cropland	[113]
Built environment	30 second spatial resolution data for the portion (0 to 1) of each grid occupied by the built environment	[113]
Bare area	30 second spatial resolution data for the portion (0 to 1) of each grid occupied by bare area	[113]
Wetlands	30 second spatial resolution data for the portion (0 to 1) of each grid occupied by wetlands	[113]

Name	Description	Source
Water	30 second spatial resolution data for the portion (0 to 1) of each grid occupied by water	[113]
Population density	Population density from the year 2000, at 30 second spatial resolution	[114]
Soil pH	30 second spatial resolution for soil pH (H_2O) at a depth of zero to five centimeters	[115]
Soil nitrogen	30 second spatial resolution for total soil nitrogen at a depth of zero to five centimeters	[115]
Travel time to cities	Estimated travel time in minutes to the nearest city with a population of 20,000 or greater in 2015 at a 30 second spatial resolution	[116]

Appendix C

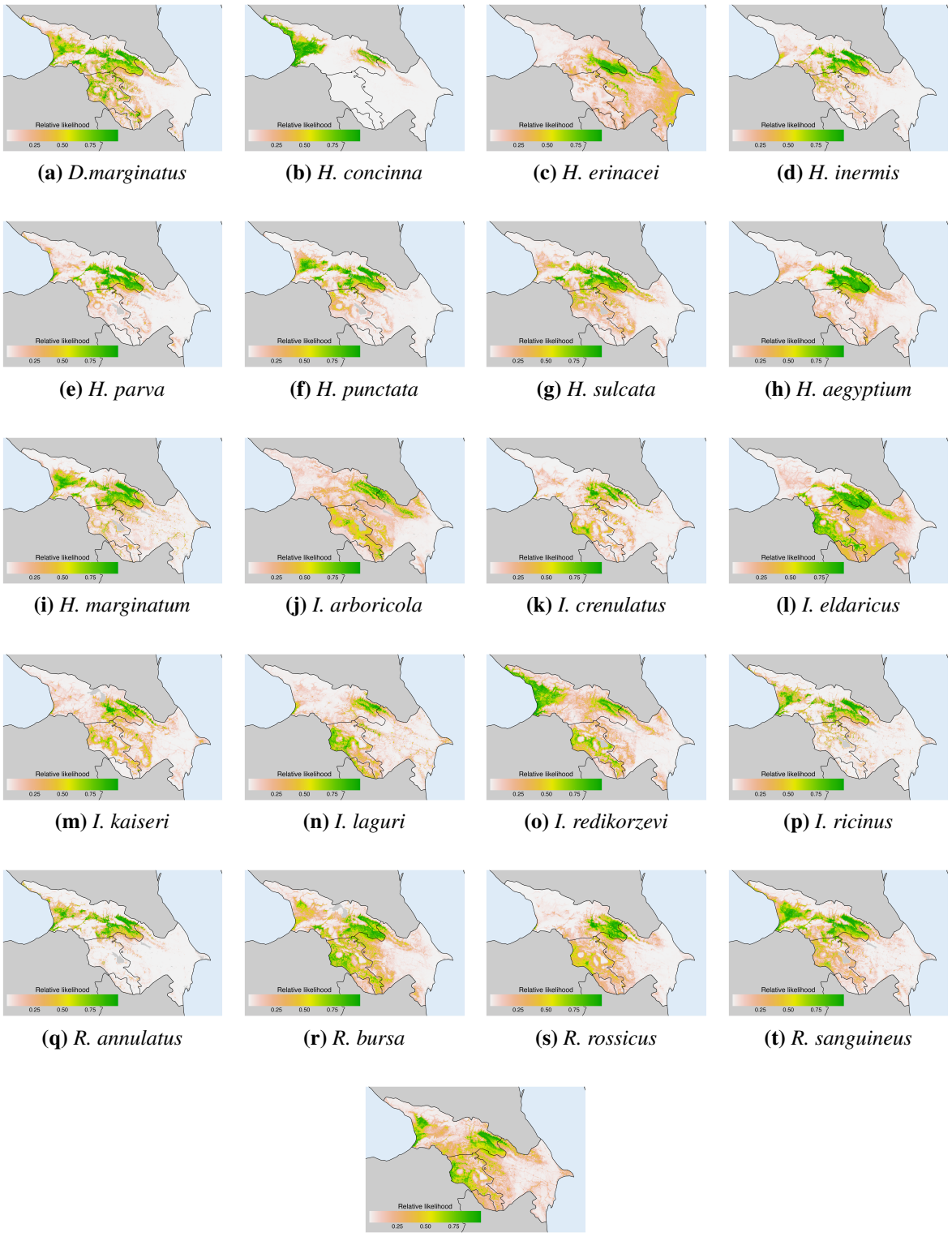
Covariates Selected for Medically-Relevant Small Mammal Models

Table C.1: Frequency with which each covariate was selected for a medically-relevant small mammal species distribution model.

Covariate	Model Count
Wind speed	10
Built environment	9
Human footprint	9
Elevation	9
Isothermality	8
Precipitation of warmest quarter	8
Temperature annual range	8
Travel time to cities	7
Trees	7
Annual mean temperature	6
Precipitation of driest month	6
Precipitation seasonality	6
Grassland	5
Mean temperature of coldest quarter	5
Annual precipitation	5
Precipitation of wettest quarter	5
Mean diurnal range	5
Temperature seasonality	5
Water pH	4
Mean temperature of warmest quarter	4
Minimum temperature of coldest month	4
Bare	3
Maximum temperature of warmest month	3
Mean temperature of driest quarter	3
Soil nitrogen	2
Population density	2
Cropland	1
Water	1
Precipitation of wettest month	1
Precipitation of coldest quarter	1
Mean temperature of wettest quarter	1
Moss	0
Shrubs	0
Snow	0
Wetland	0

Appendix D

Tick Vector Covariate Distributions



(a) *D.marginatus* **(b)** *H. concinna* **(c)** *H. erinacei* **(d)** *H. inermis*

(e) *H. parva* **(f)** *H. punctata* **(g)** *H. sulcata* **(h)** *H. aegyptium*

(i) *H. marginatum* **(j)** *I. arboricola* **(k)** *I. crenulatus* **(l)** *I. eldaricus*

(m) *I. kaiseri* **(n)** *I. laguri* **(o)** *I. redikorzevi* **(p)** *I. ricinus*

(q) *R. annulatus* **(r)** *R. bursa* **(s)** *R. rossicus* **(t)** *R. sanguineus*

(u) *R. turanicus*

Figure D.1: Predicted tick species distributions across the South Caucasus.

Appendix E

Host and Vector Covariates

Table E.1: List of animal covariate layers for disease modeling.

Species Type	**Scientific Name**
Tick vector species	*Dermacentor marginatus*
Tick vector species	*Haemaphysalis concinna*
Tick vector species	*Haemaphysalis erinacei*
Tick vector species	*Haemaphysalis inermis*
Tick vector species	*Haemaphysalis parva*
Tick vector species	*Haemaphysalis punctata*
Tick vector species	*Haemaphysalis sulcata*
Tick vector species	*Hyalomma aegyptium*
Tick vector species	*Hyalomma marginatum*
Tick vector species	*Ixodes arboricola*
Tick vector species	*Ixodes crenulatus*
Tick vector species	*Ixodes eldaricus*
Tick vector species	*Ixodes kaiseri*
Tick vector species	*Ixodes laguri*
Tick vector species	*Ixodes redikorzevi*
Tick vector species	*Ixodes ricinus*
Tick vector species	*Rhipicephalus annulatus*
Tick vector species	*Rhipicephalus bursa*
Tick vector species	*Rhipicephalus rossicus*
Tick vector species	*Rhipicephalus sanguineus*
Tick vector species	*Rhipicephalus turanicus*

Species Type	**Scientific Name**
Small mammal host species	*Apodemus uralensis*
Small mammal host species	*Apodemus witherbyi*
Small mammal host species	*Arvicola amphibius*
Small mammal host species	*Erinaceus concolor*
Small mammal host species	*Lepus europaeus*
Small mammal host species	*Meriones libycus*
Small mammal host species	*Meriones tristrami*
Small mammal host species	*Microtus arvalis*
Small mammal host species	*Msocialis*
Small mammal host species	*Mus musculus*
Small mammal host species	*Nothocricetulus migratorius*
Small mammal host species	*Rattus norvegicus*
Small mammal host species	*Rattus rattus*
Small mammal host species	*Scarturus elater*
Small mammal host species	*Scarturus williamsi*
Wild ungulate host species	*Capra aegagrus*
Wild ungulate host species	*Capra caucasica*
Wild ungulate host species	*Capra cylindricornis*
Wild ungulate host species	*Capreolus capreolus*
Wild ungulate host species	*Cervus elaphus*
Wild ungulate host species	*Gazella subgutturosa*
Wild ungulate host species	*Rupicapra rupicapra*
Wild ungulate host species	*Sus scrofa*

Appendix F

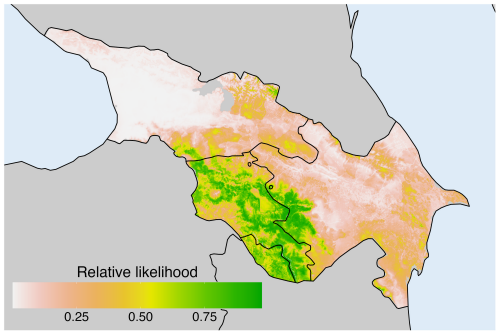
Tularemia Case Demographics by Country

Table F.1: Tularemia case demographics in the South Caucasus.

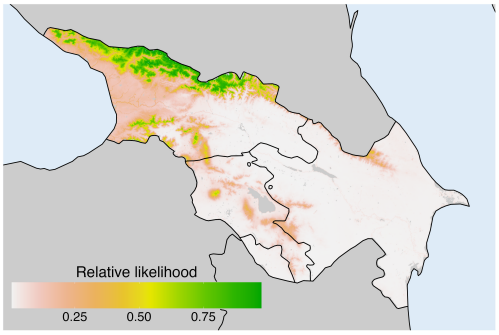
	Armenia (N=323)	**Azerbaijan (N=288)**	**Georgia (N=442)**
Gender			
Male	191 (59%)	224 (78%)	229 (52%)
Female	132 (41%)	64 (22%)	213 (48%)
Age (year)			
Median	22	31	25
Range	0-74	5-74	2-81
Clinical manifestation			
Glandular	255 (79%)	6 (2%)	81 (18%)
Oculoglandular	63 (20%)	26 (9%)	77 (17%)
Oropharyngeal	0 (0%)	11 (4%)	149 (34%)
Pneumonic	0 (0%)	14 (5%)	28 (6%)
Typhoidal	0 (0%)	228 (79%)	32 (7%)
Ulceroglandular	5 (2%)	3 (1%)	75 (17%)
Occupational risk			
Direct	120 (37%)	0 (0%)	126 (29%)
Indirect	152 (47%)	100 (35%)	196 (44%)
Missing	51 (16%)	188 (65%)	120 (27%)

Appendix G

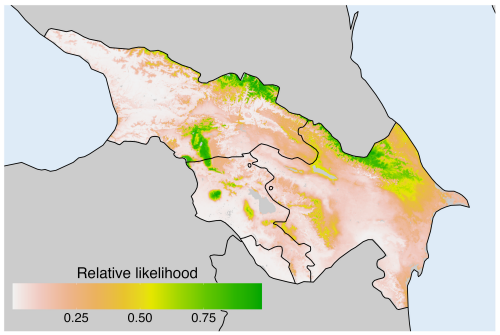
Wild Ungulate Distributions



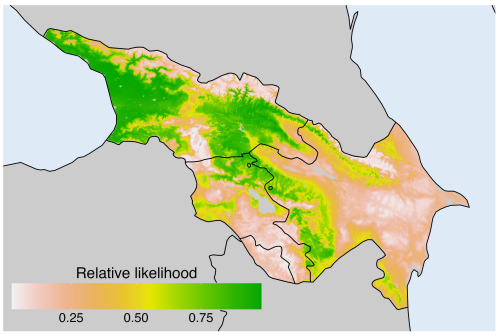
(a) *C. aegagrus*



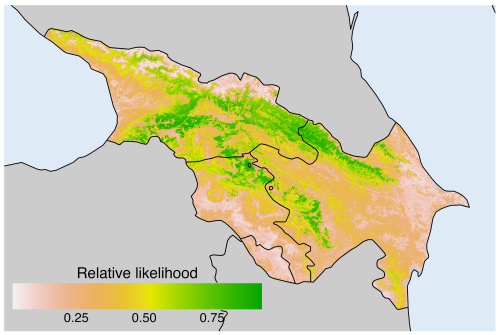
(b) *C. caucasica*



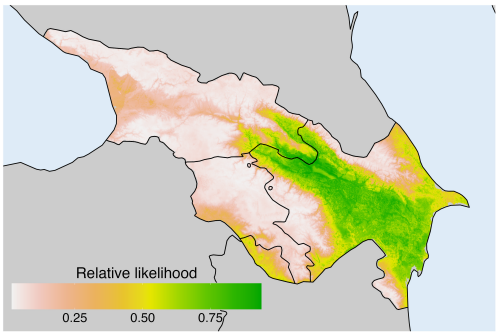
(c) *C. cylindricornis*



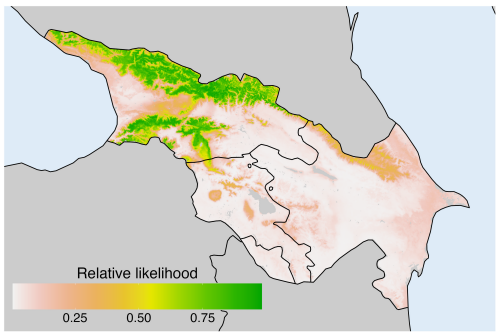
(d) *C. capreolus*



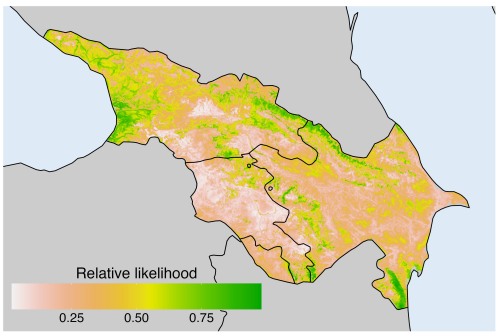
(e) *C. elaphus*



(f) *G. subgutturosa*



(g) *R. rupicapra*

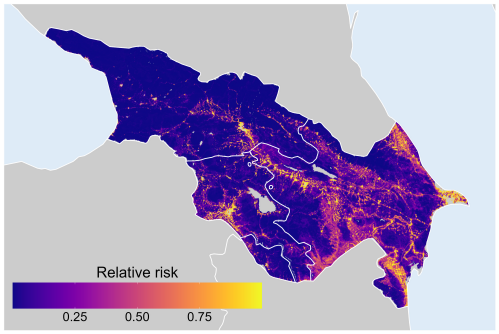


(h) *S. scrofa*

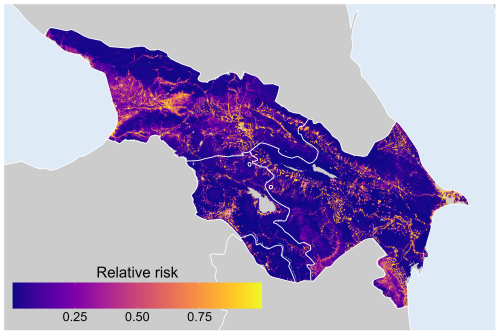
Figure G.1: Predicted wild ungulate species distributions across the South Caucasus.

Appendix H

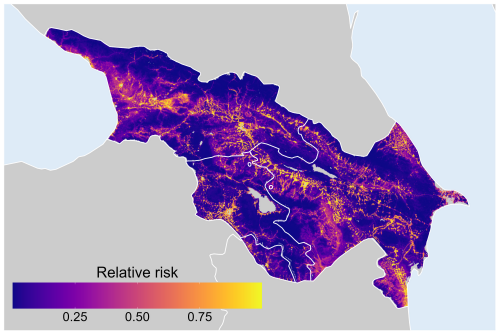
Zoonotic Disease Prediction



(a) Anthrax human risk

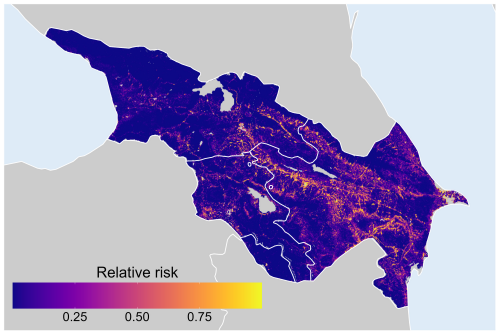


(b) Anthrax animal risk

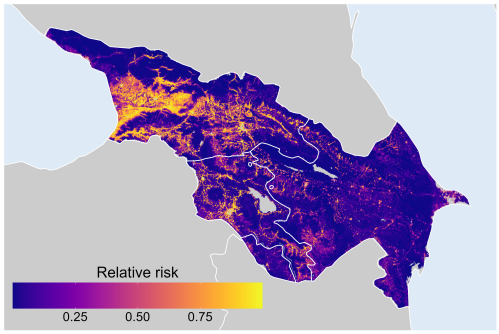


(c) Anthrax soil risk

Figure H.1: Anthrax risk across the South Caucasus.

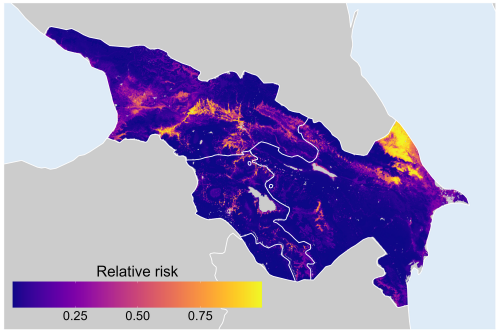


(a) Brucellosis human risk

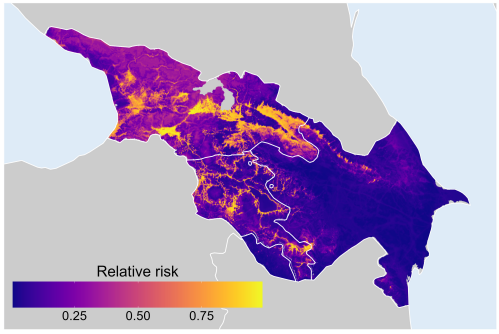


(b) Brucellosis animal risk

Figure H.2: Brucellosis risk across the South Caucasus.

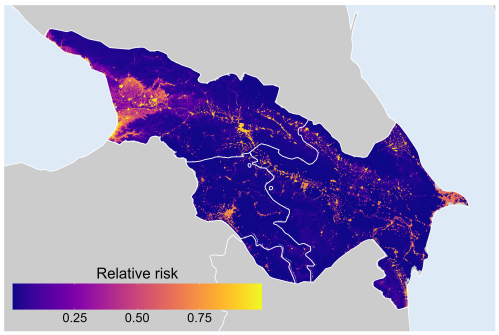


(a) CCHF human risk

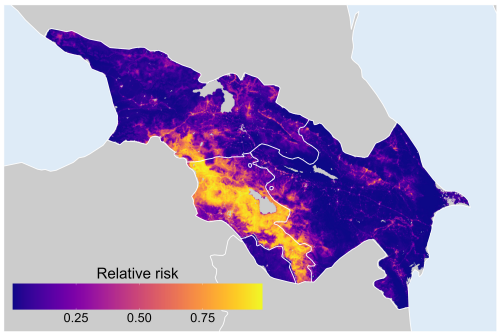


(b) CCHF host and vector risk

Figure H.3: Crimean-Congo Hemorrhagic Fever risk across the South Caucasus.

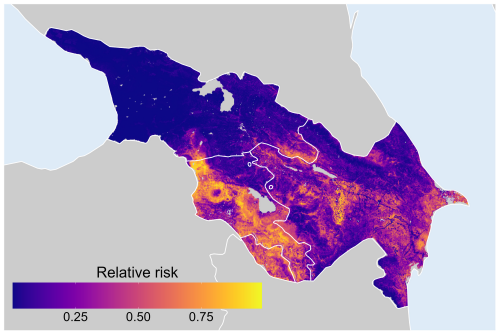


(a) Leptospirosis human risk

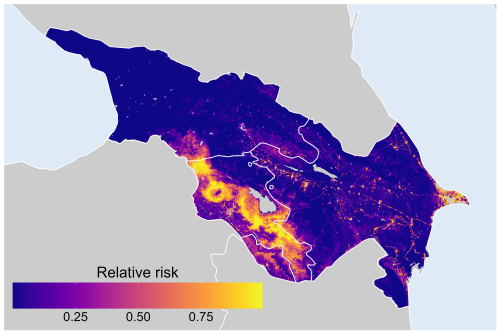


(b) Leptospirosis animal risk

Figure H.4: Leptospirosis risk across the South Caucasus.

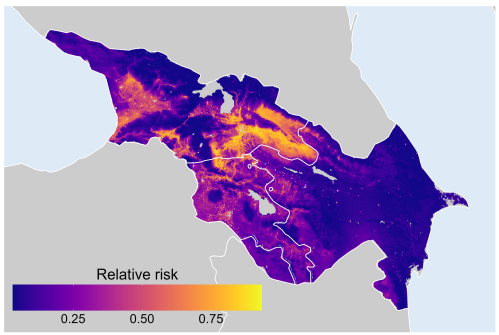


(a) Plague animal risk

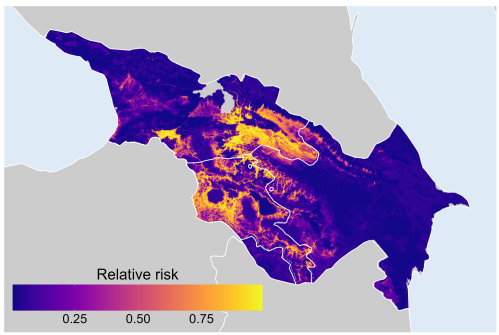


(b) Plague flea risk

Figure H.5: Plague risk across the South Caucasus.

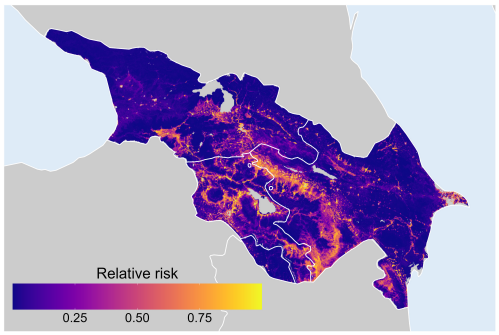


(a) TBE human risk

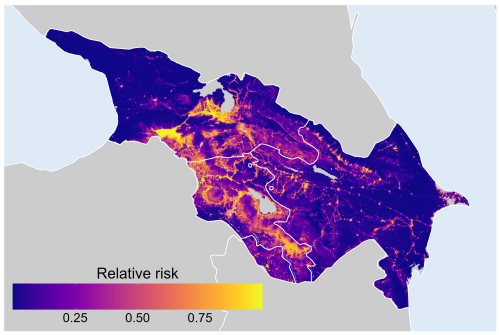


(b) TBE host and vector risk

Figure H.6: Tick-borne Encephalitis risk across the South Caucasus.



(a) Tularemia human risk



(b) Tularemia host and vector risk

Figure H.7: Tularemia risk across the South Caucasus.