Received: 14 July 2022 | Revised: 21 November 2022 | Accepted: 10 January 2023

DOI: 10.1002/eap.2827

ARTICLE

ECOLOGICAL APPLICATIONS
ECOLOGICAL SOCIETY OF AMERICA

A big data–model integration approach for predicting epizootics and population recovery in a keystone species

Gabriel M. Barrile[1,2] | **David J. Augustine**[3] | **Lauren M. Porensky**[3] | **Courtney J. Duchardt**[4] | **Kevin T. Shoemaker**[5] | **Cynthia R. Hartway**[6] | **Justin D. Derner**[7] | **Elizabeth A. Hunter**[8] | **Ana D. Davidson**[1,2]

[1]Colorado Natural Heritage Program, Colorado State University, Fort Collins, Colorado, USA

[2]Department of Fish, Wildlife, and Conservation Biology, Colorado State University, Fort Collins, Colorado, USA

[3]USDA-Agricultural Research Service, Fort Collins, Colorado, USA

[4]Department of Natural Resource Ecology and Management, Oklahoma State University, Stillwater, Oklahoma, USA

[5]Department of Natural Resources and Environmental Science, University of Nevada, Reno, Nevada, USA

[6]Wildlife Conservation Society, Bronx, New York, USA

[7]USDA-Agricultural Research Service, Cheyenne, Wyoming, USA

[8]U.S. Geological Survey, Virginia Cooperative Fish and Wildlife Research Unit, Department of Fisheries and Wildlife Conservation, Virginia Tech, Blacksburg, Virginia, USA

Correspondence
Gabriel M. Barrile
Email: gbarrile15@gmail.com

Funding information
National Institute of Food and Agriculture, Grant/Award Number: 2020-67019-31153

Handling Editor: Sarah P. Saunders

Abstract

Infectious diseases pose a significant threat to global health and biodiversity. Yet, predicting the spatiotemporal dynamics of wildlife epizootics remains challenging. Disease outbreaks result from complex nonlinear interactions among a large collection of variables that rarely adhere to the assumptions of parametric regression modeling. We adopted a nonparametric machine learning approach to model wildlife epizootics and population recovery, using the disease system of colonial black-tailed prairie dogs (BTPD, *Cynomys ludovicianus*) and sylvatic plague as an example. We synthesized colony data between 2001 and 2020 from eight USDA Forest Service National Grasslands across the range of BTPDs in central North America. We then modeled extinctions due to plague and colony recovery of BTPDs in relation to complex interactions among climate, topoedaphic variables, colony characteristics, and disease history. Extinctions due to plague occurred more frequently when BTPD colonies were spatially clustered, in closer proximity to colonies decimated by plague during the previous year, following cooler than average temperatures the previous summer, and when wetter winter/springs were preceded by drier summers/falls. Rigorous cross-validations and spatial predictions indicated that our final models predicted plague outbreaks and colony recovery in BTPD with high accuracy (e.g., AUC generally >0.80). Thus, these spatially explicit models can reliably predict the spatial and temporal dynamics of wildlife epizootics and subsequent population recovery in a highly complex host–pathogen system. Our models can be used to support strategic management planning (e.g., plague mitigation) to optimize benefits of this keystone species to associated wildlife communities and ecosystem functioning. This optimization can reduce conflicts among different landowners and resource managers, as well as economic losses to the ranching industry. More broadly, our big data–model integration approach provides a general framework for

spatially explicit forecasting of disease-induced population fluctuations for use in natural resource management decision-making.

KEYWORDS
climate, disease, population dynamics, prairie dog, random forest, spatial connectivity, western Great Plains, *Yersinia pestis*

INTRODUCTION

The introduction of infectious diseases to novel environments threatens the conservation of global biodiversity (Daszak et al., 2000; Jones et al., 2008). In recent decades, taxonomic groups from bees to bats to amphibians have suffered catastrophic declines and species extinctions due to novel diseases (Goulson et al., 2015; Hoyt et al., 2021; Scheele et al., 2019). Consequently, there is growing recognition that forecasting the timing and location of disease outbreaks could improve management of species challenged by novel diseases. Yet, predicting the spatiotemporal dynamics of outbreaks remains challenging as wildlife epizootics generally involve complex interactions and nonlinear relationships among a large collection of variables that rarely adhere to the assumptions of traditional parametric regression modeling (Fountain-Jones et al., 2019; Han et al., 2020).

To overcome these limitations, ecologists have increasingly adopted more flexible, nonparametric approaches such as machine learning algorithms (Han & Drake, 2016; Olden et al., 2008). This "rise of machines" in disease ecology (Pandit & Han, 2020) has included predicting zoonotic reservoir status for disease surveillance (Plowright et al., 2019; Walsh et al., 2019) and spillover risk (Basinski et al., 2021), identifying potential zoonotic vectors (Babayan et al., 2018; Evans et al., 2017), quantifying associations between outbreaks and animal movements (Machado et al., 2019), aiding diagnostic laboratory tests (Romero et al., 2022), and predicting habitats conducive to the co-occurrence of pathogens and their hosts (Khalil et al., 2017). For instance, machine learning models captured strong nonlinear patterns and complex interactions among variables shaping the exposure risk of African lions (*Panthera leo*) to canine distemper virus and feline parvovirus (e.g., higher risk in young lions, but only during low rainfall; Fountain-Jones et al., 2019). Recently, Peters et al. (2018) coupled human-guided machine learning strategies with fine-scale, highly resolved data sets to develop a big data–model integration approach for improving understanding, prediction, and management of infectious diseases. Whereas big data–model integration has been applied to analyzing the drivers of disease spread in animals (Peters et al., 2018, 2020) and detecting early warning signs of an outbreak (Peters et al., 2020), such an approach has not yet been used to predict the timing and location of epizootics.

Due to their importance as zoonotic reservoirs, rodent disease systems are remarkably well studied (Han et al., 2015), with host data available for a large proportion of species (Jones et al., 2009). Here we use a well-studied rodent disease system—black-tailed prairie dogs (BTPDs, *Cynomys ludovicianus*) and sylvatic plague in the western Great Plains of North America—to apply a big data–model integration approach for predicting spatiotemporal dynamics of wildlife epizootics. The plague bacterium *Yersinia pestis*, introduced into western North America in the early 1900s (Link, 1955), is a zoonotic pathogen affecting wildlife populations in this region (Antolin et al., 2002; Gage & Kosoy, 2005). Plague often causes >99% mortality in BTPD colonies during epizootics (Augustine et al., 2008; Cully & Williams, 2001). In areas affected by epizootic plague, BTPD population dynamics have shifted from relatively stable colony complexes that existed prior to European settlement of North America (Keuler et al., 2020; Knowles et al., 2002) to extreme boom-and-bust cycles (Davidson, Augustine, Jacobsen, et al., 2022). Colony dynamics are now more commonly characterized by catastrophic collapse due to plague ("busts"), followed by population recovery ("booms") leading up to the next epizootic (e.g., Augustine et al., 2008; Cully et al., 2010; Hartley et al., 2009; Figure 1a).

BTPDs serve as a keystone species within the central grasslands of North America (Davidson et al., 2012; Kotliar et al., 1999). Extreme fluctuations in colony size and abundance negatively affect grassland ecosystems as plague-driven crashes result in subsequent declines in species heavily reliant on BTPDs as prey, such as large predatory birds (Duchardt et al., 2022; Seery & Matiatos, 2000), or as creators of critical habitat through their herbivory and burrowing (Augustine & Skagen, 2014; Eads & Biggins, 2015). Conversely, colony expansion can conflict with livestock production as BTPDs compete with cattle for available forage (Augustine & Derner, 2021; Crow et al., 2022), particularly during drier periods (Connell et al., 2019).

Several management tactics exist that, if employed strategically spatially and temporally, may help reduce the extreme fluctuations in BTPD populations due to plague epizootics. Applications of insecticide can reduce

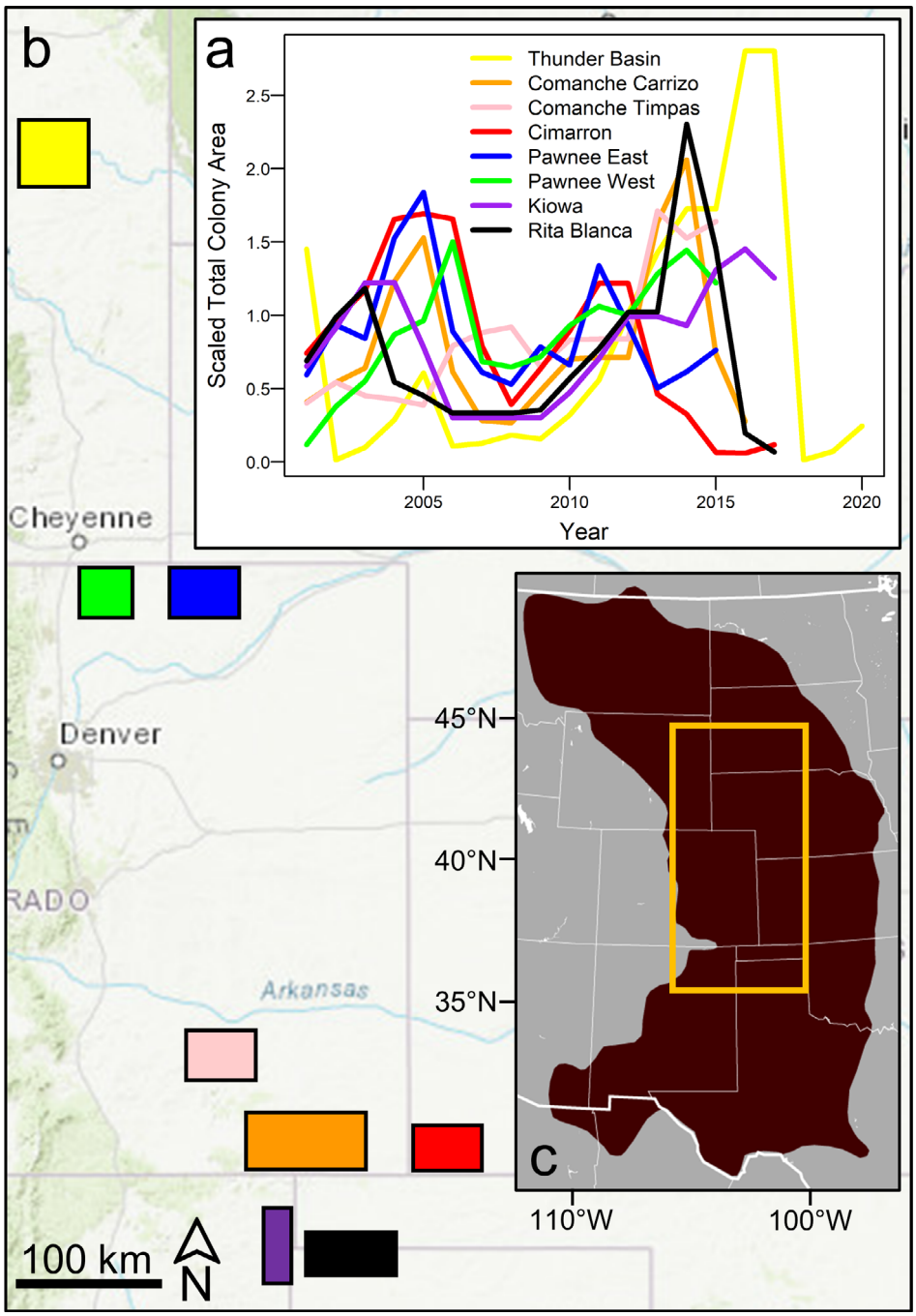

FIGURE 1 (a) Black-tailed prairie dogs (BTPDs, *Cynomys ludovicianus*) exhibit extreme population fluctuations in USDA Forest Service National Grasslands (NGs) where plague is known to be present. The total area of BTPD colonies in (a) was scaled (divided by root mean square) such that colony data for each NG could be visualized. Flat lines across years in (a) indicate when surveys did not occur or when partial surveys were averaged. (b) We synthesized colony data between 2001 and 2020 from eight NGs (colored boxes) across the western Great Plains of North America. Note that each NG in (b) matches the line colors in (a). Inset (c) displays the historical range of BTPDs in brown and the extent of the main map (b) in the gold box.

flea abundance—the main vector of plague—and plague transmission (Eads & Biggins, 2019; Tripp et al., 2017), whereas rodenticide can control colony expansion (Knowles et al., 2002). Employing these tactics in a manner that optimizes ecological benefits of BTPD colonies while minimizing livestock production costs requires accurate spatiotemporal predictions of colony dynamics to identify when and where to deploy mitigation.

Despite extensive research on plague dynamics in BTPDs (Biggins & Eads, 2019; Cully & Williams, 2001), a reliable predictive model for plague outbreaks in BTPD colonies remains elusive. Epizootics depend on how the

plague bacterium, flea vector, and prairie dog host respond to variations in temperature (Collinge, Johnson, Ray, Matchett, Grensten, Cully, Gage, et al., 2005; Snäll et al., 2008), precipitation (Eads & Biggins, 2017; Stapp et al., 2004), soil type (Eads, 2014), landscape features (Collinge, Johnson, Ray, Matchett, Grensten, Cully, & Martin, 2005; Snäll et al., 2008), and the size and connectedness of host populations (George et al., 2013; Johnson et al., 2011). Previous efforts to build a predictive model of plague epizootics in BTPDs employed detailed, mechanistic approaches, which improved understanding of processes (e.g., Collinge, Johnson, Ray, Matchett, Grensten, Cully, & Martin, 2005) but were unable to incorporate complex nonlinear interactions among the variables listed earlier. Furthermore, restricted spatial (e.g., Stapp et al., 2004) and temporal (e.g., Cully et al., 2010) coverage of input data limited the scope of inference and precluded the development of a more generalizable model. The recent integration of machine learning in wildlife disease ecology (Pandit & Han, 2020), in conjunction with increasing availability of large host and environmental data sets, provides an excellent opportunity to build a predictive model to guide the management of BTPDs in areas affected by plague.

We synthesized long-term, high-resolution data sets collected during annual mapping of BTPD colonies on federally managed grasslands to forecast host population dynamics using a big data–model integration approach. We assembled colony data collected during 2001–2020 from eight USDA Forest Service National Grasslands (NGs) across the range of BTPD in central North America (Figure 1b,c). We then adopted a machine learning approach via a random forest algorithm (Breiman, 2001) to model extinctions due to plague and colony recovery in relation to complex interactions among climate, topoedaphic variables, colony characteristics, and disease history. To evaluate predictive accuracy, we subjected our model to a series of rigorous cross-validations and compared the metrics of spatial overlap between model predictions and on-the-ground mapping of BTPD colonies. We discuss several applications of our model, including its use for planning and targeting plague mitigation. Although applied to the prairie dog–plague system in this study, our modeling approach could be applied to any wildlife disease system wherein spatially explicit predictions could inform management decisions.

METHODS

Study area

We focused on eight study sites with long-term (~2001–2020), high-resolution data sets for BTPD colonies (latitudinal range = 36–44° N; longitudinal range = 101–106° W; Figure 1b). Study sites consisted of USDA NGs or geographically distinct subunits of NGs across a climate gradient. For instance, the eastern and western units of Pawnee NG encompass a precipitation gradient, and colonies within the two units exhibit different patterns in plague epizootics (Stapp et al., 2004); we therefore treated these two units as separate sites. Similarly, the Kiowa and Rita Blanca NGs also encompass a precipitation gradient, with colonies in New Mexico (Kiowa) exhibiting different plague epizootic patterns than colonies in Texas and Oklahoma (Rita Blanca; Cully et al., 2010). In addition to an east–west precipitation gradient (eastern sites received more precipitation on average than western sites), our study area also was characterized by a north–south temperature gradient, with southern sites experiencing warmer temperatures on average than northern sites. Finally, our northernmost site (Thunder Basin NG) was situated primarily in northern mixed-grass prairie—a predominantly needlegrass, western wheatgrass (*Pascopyrum smithii*), and blue grama (*Bouteloua gracilis*) community—whereas all other sites were situated in shortgrass steppe (grama/buffalograss community), with some overlap in southern mixed-grass prairie (sandsage/bluestem community; Augustine et al., 2021).

We focused only on federal land within administrative boundaries of our eight study sites. We included all federal land within each NG (totaling 373,556 ha; Table 1), with the exception of Thunder Basin, where we focused on a 17,143-ha subsection of the overall NG that received consistent mapping of BTPD colonies during 2001–2020 (Davidson, Augustine, Jacobsen, et al., 2022).

Prairie dog colony data

BTPD colony boundaries were mapped annually, with some minor exceptions described in what follows. Mapping involved biologists from multiple agencies and organizations (e.g., USDA Forest Service, U.S. Geological Survey, university partners, private contractors) identifying the outermost boundaries of colonies (usually during summer months) and using GPS devices to record the outermost perimeter of each colony. Colony boundaries were demarcated based on the locations of entrances to active burrows (evidenced by recent digging, soil disturbance, BTPD trails, fresh feces) and stark transitions between recently cropped vegetation versus taller vegetation without signs of cropping by BTPDs (Augustine et al., 2008; Cully et al., 2010). Mapping the perimeter of BTPD colonies produced shapefiles with polygons (vector data) delineating colony boundaries for a given NG and year. We refer to shapefiles with polygons of colony

TABLE 1 Total area (hectares) and years of black-tailed prairie dog (BTPD, *Cynomys ludovicianus*) colony mapping data from eight national grasslands across central North America (Figure 1). Area under the curve (AUC) values represent the prediction accuracy of plague and colony growth models from our most rigorous cross-validation routine, wherein data from each respective National Grassland were withheld during model training.

National grassland	Years	Hectares	AUC (plague)	AUC (colony growth)
Thunder Basin	2001–2020	17,143	0.81	0.88
Pawnee West	2001–2015	46,057	0.62	0.88
Pawnee East	2001–2015	38,205	0.71	0.92
Comanche Timpas	2001–2015	71,796	0.59	0.93
Comanche Carrizo	2001–2016	94,718	0.70	0.94
Cimarron	2001–2017	44,232	0.73	0.94
Kiowa	2001–2017	23,580	0.72	0.91
Rita Blanca	2001–2017	37,825	0.75	0.92

Note: To ensure that random forest models made accurate predictions, we subjected them to a rigorous series of cross-validations. Here we display the results of the cross-validation wherein all colonies and years from an entire region (i.e., national grassland) were withheld from the data set when training the model. We then used the trained model to make predictions on the withheld data. As displayed in the table, we evaluated models based on the AUC of a receiver operator characteristic, where AUC values closer to 1 indicate better prediction and those closer to 0.5 indicate predictions closer to random. More specifically, for the plague model, prediction accuracy refers to the ability of the model to correctly classify between raster cells that experienced extinctions due to plague between years (coded as 1) versus cells that remained occupied (coded as 0). For the colony growth model, prediction accuracy refers to the ability of the model to correctly classify between raster cells that were colonized by BTPDs between years (coded as 1) versus cells that remain unoccupied (coded as 0).

boundaries as "colony data." For more detail on colony mapping and site description, see Davidson, Augustine, Jacobsen, et al. (2022) for Thunder Basin NG, Stapp et al. (2004) and Savage et al. (2011) for Pawnee NG, and Cully et al. (2010) and Johnson et al. (2011) for Cimarron, Kiowa, Rita Blanca, and Comanche NGs (including both Timpas and Carrizo units, hereafter referred to as Comanche Timpas and Comanche Carrizo, or simply Comanche when referring to both units).

We initially requested and received colony data from agency biologists working in multiple NGs, including several sites not included in the final analysis. To determine whether to include or exclude data sets, we vetted each data set based on several criteria. First, sites must have been mapped in consecutive years (at least three consecutive years), with complete mapping of the entire grassland in each year. The only exceptions were 2014–2017 at Thunder Basin, where expansive colonies required 2 years (2014–2015 and 2016–2017) to map their full extent, Comanche and Cimarron NGs, which were not mapped in 2012, and Kiowa and Rita Blanca NGs, which were not mapped during 2007–2008 and 2013. We accounted for this data deficiency in our modeling by averaging covariate values for those years (see *Model covariates* section in what follows). Second, various management efforts were implemented across NGs to control BTPD colony expansion (e.g., shooting, poisoning), increase prairie dog abundance (e.g., translocation, mowing), and reduce plague transmission (e.g., administering deltamethrin dust and/or fipronil bait to reduce flea abundance). We excluded grasslands where management efforts affected more than 20% of total colony area. For instance, we excluded Oglala NG because most colonies were poisoned during the years for which we were able to obtain data.

Site-years with colony data that met the standard for inclusion in our analyses are listed in Table 1. Prior to analysis, we converted colony shapefiles for these site-years into raster layers with 1-ha resolution using functions within the raster package in R version 4.1.1 (R Core Team, 2021). We chose a cell size of 1 ha because we considered that to be the minimum area necessary to constitute an active BTPD colony (Davidson et al., 2014; Hoogland, 2013).

Model covariates

We reviewed the BTPD and plague literature to identify variables that could influence plague epizootics and colony recovery. We then derived model covariates from geospatial data layers available across our entire study area as we aimed to generate spatial predictions of BTPD colony dynamics. Therefore, some biotic factors, such as flea and BTPD densities, which play roles in plague dynamics (Biggins & Eads, 2019; Tripp et al., 2009), were not included in our suite of covariates because values were not available for each 1-ha pixel. Rather, we included surrogate geospatial variables that could dictate biotic factors (e.g., flea density, alternate host abundance; Salkeld et al., 2010) such as climate and soil characteristics. We ultimately selected 61 covariates, including variables related to temperature, precipitation, soil type,

landscape features, colony characteristics, and disease history at a given site. Refer to Appendix S1: Tables S1 and S2 for the full list of covariates and their hypothesized relationships to plague epizootics and colony growth, respectively.

We derived time-varying temperature and precipitation variables from the Daymet (Sadoti et al., 2020; Thornton et al., 1997) data product using the "get_daymet()" function within the FedData package in R (Bocinsky, 2021). Daymet provides gridded estimates of daily climate data at a 1-km spatial resolution. Given that climatic effects on disease outbreaks can manifest with 1- to 2-year time lags (Ben Ari et al., 2008; Stapp et al., 2004), we considered all temperature and precipitation variables in our models at both time t and $t-1$ when predicting the probability of extinction due to plague between t and $t+1$. Winter/spring precipitation also was considered at time $t+1$, directly before colony mapping during the summer at $t+1$. Furthermore, abnormal weather conditions can increase host susceptibility to disease (e.g., extreme events or longer-term occurrences such as prolonged drought; Cazelles et al., 2005; Pascual et al., 2008). Thus, in addition to their absolute values, we also considered the extent to which each weather variable in our model deviated from the long-term average (during 2000–2020) at that site.

We obtained soil characteristics from the POLARIS (Chaney et al., 2019) data set using the "ximages()" function within the XPolaris package (Moro Rosso et al., 2021a, 2021b). POLARIS builds upon the National Soil Survey's SSURRGO database by providing soil texture and other attributes at a finer resolution (30-m pixel size). We derived landscape features by first importing digital elevation models (10-m resolution) from the National Elevation Dataset (U.S. Geological Survey, 2022) using the "get_ned()" function in the FedData package. We then used the elevation layers to calculate various topographic metrics (e.g., slope, terrain ruggedness index) using functions within the spatialEco package (Evans & Ram, 2021). We also derived a Topographic Wetness Index from the elevation layer using the "upslope.area()" function within the dynatopmodel package (Metcalfe et al., 2018) and created distance to stream and road layers using the "roads()" and "linear_water()" functions within the tigris package (Walker, 2021). We used colony rasters and functions within the landscapemetrics package (Hesselbarth et al., 2019) to calculate patch metrics and landscape configuration (e.g., colony area, mean nearest neighbor) for each site-year. We also used colony rasters and functions within the raster package to calculate neighborhood features and disease history variables for each site-year, such as the cost-weighted distance to sites impacted by epizootic plague during the previous year.

Lastly, we included a BTPD habitat suitability layer from a study that modeled BTPD occurrence as a function of soil, climate, topography, and land-cover data across the entire species range in the United States (Davidson, Augustine, Menefee, et al., 2022). We expected this static habitat suitability layer would be important in our colony growth model, as we anticipated that BTPDs would colonize the most suitable habitat following epizootics (whereas temporally dynamic climate variables could predict *when* colonization of suitable habitat could be, more or less rapid). Finally, we resampled all covariate layers to a 1-ha spatial resolution, stacked them with the colony rasters, and extracted the corresponding values to create our data frame for analysis.

Inferring plague epizootics and colony growth

We inferred plague epizootics and population recovery from annual changes in colony data. To distinguish between plague-induced die-offs and mortality due to predation, drought, or other natural processes, we defined plague epizootics as catastrophic declines of greater than 50% loss in colony area between consecutive years (Colman et al., 2021; Johnson et al., 2011). Thus, within colonies that suffered greater than 50% loss in area between years, we classified raster cells that transitioned from occupied by BTPDs to unoccupied as extinction events due to plague. We were confident that these large-scale extinction events were primarily attributable to plague because plague is the only disease known to cause extensive die-offs among BTPDs over short time periods (Barnes, 1993), and no other mechanism (e.g., predation, drought) is likely to cause the extreme population reductions that could be confused with the effects of plague (e.g., Colman et al., 2021; Cully et al., 2010; Johnson et al., 2011; Matchett et al., 2021). Indeed, serological testing of BTPD carcasses and fleas collected in and around burrows confirmed the presence of the plague bacterium at all eight of our study sites (Cully et al., 2010; Johnson et al., 2011; Savage et al., 2011; Thiagarajan et al., 2008). Nonetheless, without serologic evidence for each site-year, we acknowledge that a few extinctions, particularly of smaller colonies, may have been due to predation, drought, or other natural mortality. We thus recognize that annual changes in colony area served as a proxy for plague epizootics and that population fluctuations did not equate to disease dynamics.

We classified raster cells that transitioned from unoccupied to occupied between consecutive years as colonization during population recovery. We assumed that colony area was a reasonable surrogate for

population abundance (Johnson et al., 2011). However, we acknowledge that colony expansion may not always equate to population growth; BTPDs may increase their search radius for food resources during drought conditions, such that colony area increases without concurrent increases in population abundance (Bruggeman & Licht, 2020). Because our study sites were affected by plague, however, we were confident that most colonization events by BTPDs represented population expansion after epizootic events.

Random forest modeling

We conducted two random forest (Breiman, 2001) modeling procedures, one for extinctions due to plague and one for colony recovery. We trained our first random forest model (plague model) to distinguish between raster cells that experienced extinctions due to plague (coded as 1) versus cells that remained occupied (coded as 0). We trained our second model (colony growth model) to distinguish between cells that were colonized by BTPDs (coded as 1) versus cells that remain unoccupied (coded as 0). In both procedures, we first removed highly correlated (Pearson's correlation coefficients > |0.65|) and multicollinear variables. We used the "multi.collinear()" function within the rfUtilities package (Evans et al., 2011), which provides a test for multicollinearity using QR-matrix decomposition (Roozbeh & Najarian, 2018). Within the function, we specified a multicollinearity threshold of 0.05, which is recommended when the number of variables (i.e., covariates) is greater than 20. The function returns a vector of collinear variables, all of which we removed prior to analysis.

We next subset our data to reduce autocorrelation and hasten computation time by randomly sampling 10% of each data set (final data sets: $n = 22{,}442$ grid cells for plague model and $n = 23{,}508$ grid cells for colony growth model). Then, to reduce the number of parameters in each model, we employed a model selection routine using the "rf.modelSel()" function within the rfUtilities package (Murphy et al., 2010). We specified 1001 bootstrap replicates (i.e., number of trees) and used the model improvement ratio to rank and select parameters to include in our final models. This routine ultimately selected six covariates to include in our final plague model and six covariates to include in our final colony growth model (Appendix S1: Tables S3 and S4).

Random forest modeling can be sensitive to class imbalances in the response data, whereby bootstrap samples overrepresent the majority class, resulting in underprediction of the minority class (Chen et al., 2004; Freeman et al., 2012). Both of our data sets included approximately 33% 1s and 67% 0s, so we considered both data sets moderately imbalanced. To address this issue, we downsampled the majority class (Kubat et al., 1998) such that each tree in the ensemble was built by drawing a bootstrap sample with approximately the same number of cases from the majority and minority classes (Evans & Cushman, 2009). We then inserted the selected variables from our model selection routines into final models, which we achieved using the "randomForest()" function in the randomForest package (Liaw & Wiener, 2002). When covariates vary in their scale of measurement and number of categories, however, which many of ours did, classical approaches, such as "randomForest()," tend to bias in favor of variables with many potential cut points. Unbiased tree algorithms, by contrast, do not artificially favor splits in variables with many categories or continuous variables (Strobl et al., 2007). To obtain unbiased variable importance estimates and corroborate variable selection, we employed the unbiased tree algorithm available with the "ctree()" function for conditional inference trees in the party package (Hothorn et al., 2006). We then implemented random forest models based on these unbiased trees using the "cforest()" function in the party package, which enables learning unbiased forests (Strobl et al., 2007).

For both implementations, "randomForest()" and "cforest()," we built models with 1001 trees and specified five as the number of variables randomly selected at each node. Further, although the party package is slower and computationally more expensive, "cforest()" can yield unbiased variable importance and more reliable predictions due to its use of unbiased trees (Strobl et al., 2008). We therefore used "cforest()" models to evaluate variable importance, univariate relationships, covariate interactions, and predictive accuracy via a series of rigorous cross-validation routines. Due to computational limitations, however, we used "randomForest()" models for spatial predictions as each prediction back to the landscape (i.e., prediction for each site-year) ran for days to weeks with "cforest()" models.

Model validation

The random forest provides an internal evaluation of model performance, whereby model error is assessed against the out-of-bag data—the portion of data not contained in the bootstrap sample used to build an individual tree—in each bootstrap replicate, providing an error distribution from which measures of model performance can be computed (Evans et al., 2011; Fox et al., 2017). While this approach to testing model sensitivity to sample distribution is robust, additional assessments

of model performance (e.g., data-withholding cross-validation techniques) typically are recommended when prediction accuracy is the main focus (Ambroise & McLachlan, 2002; Evans et al., 2011). We therefore subjected our models to a rigorous series of cross-validations. We trained each model on a subset of available sites and cross-validated in three ways: (1) withholding all years of data for a randomly selected subset of colonies from the entire training data set, (2) withholding each year of the training data set, and (3) withholding all colonies and years of an entire NG. Whereas the first cross-validation (withholding colonies) is characteristic of other BTPD-plague models, the other two cross-validations are much more rigorous, which we deemed necessary because of our focus on prediction. For each cross-validation, we used final random forest models to make predictions on the withheld data, then accumulated these predictions to compute performance measures. We evaluated models based on the area under the curve (AUC) of a receiver operator characteristic (Fawcett, 2006), where AUC values closer to 1 indicate better prediction and those closer to 0.5 indicate predictions closer to random.

Parameter evaluation

After validating the final plague and colony growth models, we first examined the relative importance of predictor variables (all six covariates in each model) in their ability to discriminate extinctions due to plague and colony expansion using the "varimp()" function within the party package (Strobl et al., 2007, 2008). We then plotted univariate relationships of predictor variables included in final models to explore the effect of each covariate on corresponding response metrics. Finally, we identified and plotted the most important covariate interactions affecting extinctions due to plague and colony recovery using custom functions within the *RF_Extensions.R* script, which can be accessed via the public repository "Random-Forest-Functions" (https://github.com/kevintshoemaker/Random-Forest-Functions/blob/master/RF_Extensions.R). Specifically, we used the custom functions "RF_FindInteractions()" and "RF_InteractionPlots()" to identify and visualize covariate interactions (Appendix S1: Tables S5 and S6).

Spatial predictions

We applied our final plague and colony growth models to predict BTPD distribution for each site-year in our data set ($n = 124$ site-years). To achieve this, we first obtained predictions from our random forest models using covariate data at time t, which generated probability surfaces for extinctions due to plague and BTPD colonization between t and $t + 1$ (Figure 2a,b). Next, we converted extinction rasters into the probability of persistence by applying the equation $1 - \text{extinction} = \text{persistence}$, or the probability that a raster cell will remain occupied between t and $t + 1$ (Figure 2c). We then masked persistence rasters to existing colonies at time t, because only occupied cells can remain occupied (Figure 2d), and masked colonization rasters to exclude existing colonies at time t, because only unoccupied cells can become colonized (Figure 2e). Next, we reclassified masked colonization and persistence layers into binary rasters based on three different probability thresholds for later comparison. For persistence rasters, probabilities greater than 0.15, 0.10, and 0.05 received a 1 and probabilities less than those threshold values received a 0 (i.e., raster cells were assumed to experience extinctions due to plague only when extinction probabilities were greater than threshold values of 0.85, 0.90, and 0.95, respectively; Figure 2f). For colonization rasters, probabilities greater than 0.85, 0.90, and 0.95 received a 1, and probabilities less than those threshold values received a 0 (Figure 2g). We then combined binary layers of persistence and colonization to produce rasters of predicted colonies at time $t + 1$ (Figure 2h). Finally, we converted these rasters to shapefiles and compared various spatial metrics between colonies predicted by our model at $t + 1$ versus actual colonies that were mapped by biologists during $t + 1$ (Figure 2i).

We compared three metrics between actual and predicted colonies (Figure 3a,b): (1) total colony area, calculated as the percentage difference in total area between predicted and actual colonies; (2) percentage of area of actual colonies nonoverlapping with model predictions (i.e., false negatives; Figure 3c); and (3) percentage of area of model predictions nonoverlapping with actual colonies (i.e., false positives; Figure 3c). We calculated these metrics at each probability threshold for all site-years in our data set ($n = 124$) and summarized the results to determine how well our models forecast BTPD colony distribution in time and space.

RESULTS

Models achieved high prediction accuracy

Random forest provided excellent model fit for both extinctions due to plague and colony growth. The out-of-bag error for the final plague model was 5.95%, with classification error equally balanced between extinction and

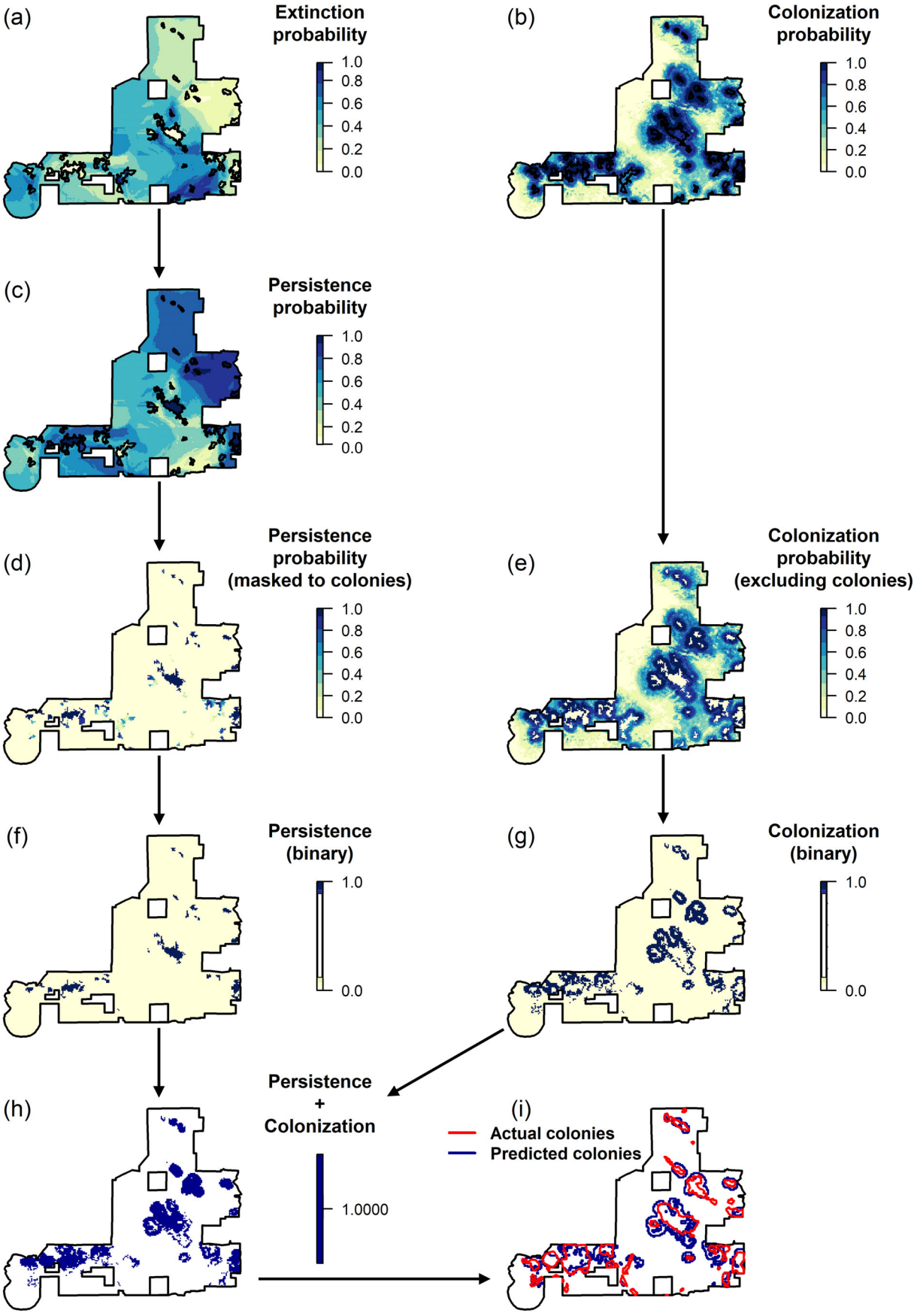

FIGURE 2 Legend on next page.

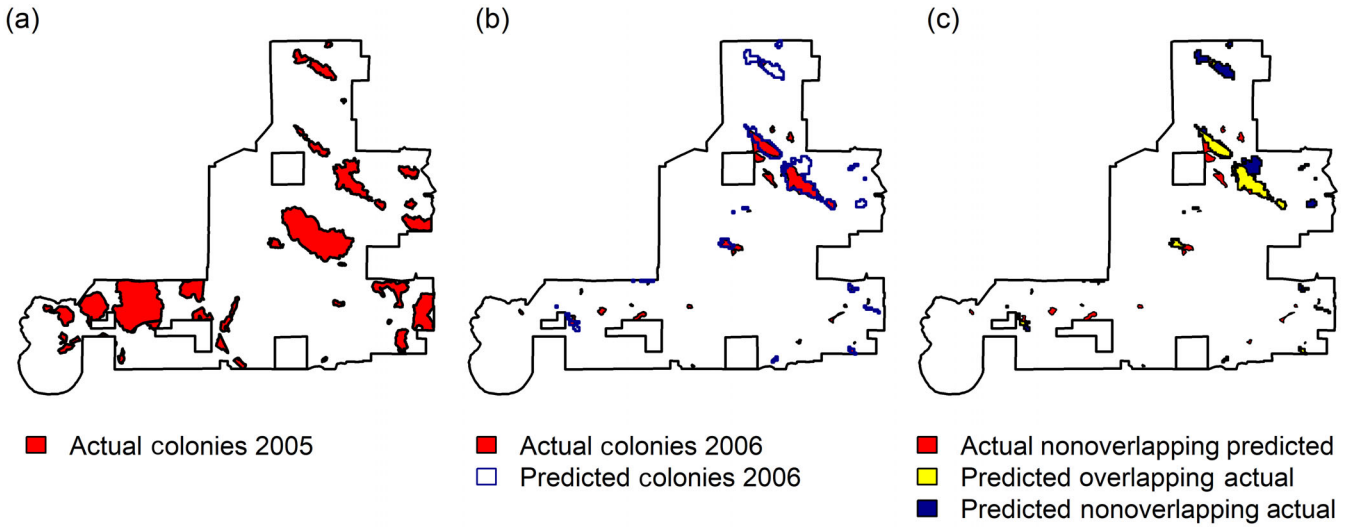

FIGURE 3 Actual mapped black-tailed prairie dog (BTPD, *Cynomys ludovicianus*) colonies at Thunder Basin National Grassland (Figure 1) in (a) 2005 and (b) 2006 in comparison to colony distribution in 2006 as predicted by our final random forest models (using a 0.95 probability threshold). Many colonies suffered plague die-offs between 2005 and 2006, and our model was able to predict with high accuracy which colonies would experience plague and which would persist. Plot (c) serves an example of how we compared the spatial distribution of colonies predicted by our model at time $t + 1$ versus colonies that were actually mapped during time $t + 1$. We compared three metrics between actual and predicted colonies: (1) total colony area, (2) percentage of area of actual colonies nonoverlapping with model predictions (red polygons in [c]), and (3) percentage of area of model predictions nonoverlapping with actual colonies (blue polygons in [c]). The latter two metrics evaluate error in the exact spatial overlap between actual and predicted colonies. We calculated these metrics at each probability threshold for all site-years in our data set ($n = 124$) and summarized the results (Table 2) to determine how well our models could forecast BTPD colony distribution in time and space.

persistence classes. Back-prediction to the data demonstrated an almost perfect fit, with an AUC of 0.994. The final colony growth model exhibited a lower out-of-bag error (5.84%), and back-prediction provided an AUC of 0.998. Like the simple back-predictions, the colony growth model performed slightly better than the plague model when subjected to rigorous cross-validation. The first cross-validation of withholding randomly selected colonies achieved an AUC of 0.96 in the colony growth model and a 0.87 in the plague model. The second cross-validation of withholding each year of data produced an AUC of 0.94 (weighted mean, weighted by number of observations in each year) in the colony growth model and a 0.80 in the plague model. Finally, the third cross-validation of withholding data from each NG achieved an AUC of 0.91 (weighted mean, weighted by number of observations at each NG) in the colony growth model and a 0.78 in the plague model, with high variability among NGs in cross-validation success (Table 1). For instance, the plague model did not predict Comanche Timpas particularly well (AUC = 0.59), likely because BTPDs did not experience large-scale extinctions due to plague at that NG. By contrast, the plague model predicted best for Thunder Basin (AUC = 0.81), where BTPDs suffered three large-scale extinctions due to plague during 2001–2020 (Figure 1a). Overall, although

FIGURE 2 Workflow for achieving spatial predictions from our random forest models using 2004–2005 data from our study area within Thunder Basin National Grassland (Figure 1) to illustrate the process; black polygons in (a–c) show actual colonies of black-tailed prairie dogs (*Cynomys ludovicianus*) mapped in 2004, and red polygons in (i) show actual colonies mapped in 2005. We first converted the probability of extinction due to plague (a) between time t and $t + 1$ (i.e., 2004 and 2005 in this example) to the probability of persistence (c), which is simply 1 minus the probability of extinction. We then masked colonization rasters (b) to exclude colonies that existed at time t (e) and masked persistence rasters (c) to only include colonies that existed at time t (d). Next, we reclassified persistence (d) and colonization (e) layers into binary rasters based on three different probability thresholds (we used 0.95 for colonization and extinction in this example; see text for details). We then combined binary layers of peristence (f) and colonization (g) to produce rasters of predicted colonies at time $t + 1$ (h). Finally, we converted colony distribution rasters (h) to shapefiles (i) and compared various spatial metrics between colonies predicted by our model at time $t + 1$ versus colonies that were actually mapped during time $t + 1$. Note that blue polygons denoting predicted colony boundaries in (i) are of the same extent as the predicted colonies in (h); blue polygons in (i) are simply not filled in to better visualize the overlap between actual and predicted colonies.

the plague model generally underperformed relative to the colony growth model, both models maintained very good to excellent predictive accuracy even when subjected to rigorous cross-validation.

Population fluctuations involved complex interactions among covariates

BTPD colonies suffered extinctions due to plague more often when colonies were spatially clustered (Figure 4c,f), following cooler than average temperatures during the previous summer (Figure 4a), and when wetter than average winter/springs were preceded by drier than average summer/falls (Figure 4d). Extinctions also were more frequent at lower latitudes (Figure 4e) and within colonies near sites decimated by epizootic plague during the previous year (Figure 4b). Furthermore, highly connected colonies exhibited a greater risk of extinction due to plague only after temperatures during the previous summer were cooler than average (Figure 5a). By contrast, the probability of extinction remained relatively high following cooler summers irrespective of proximity to previously plagued sites

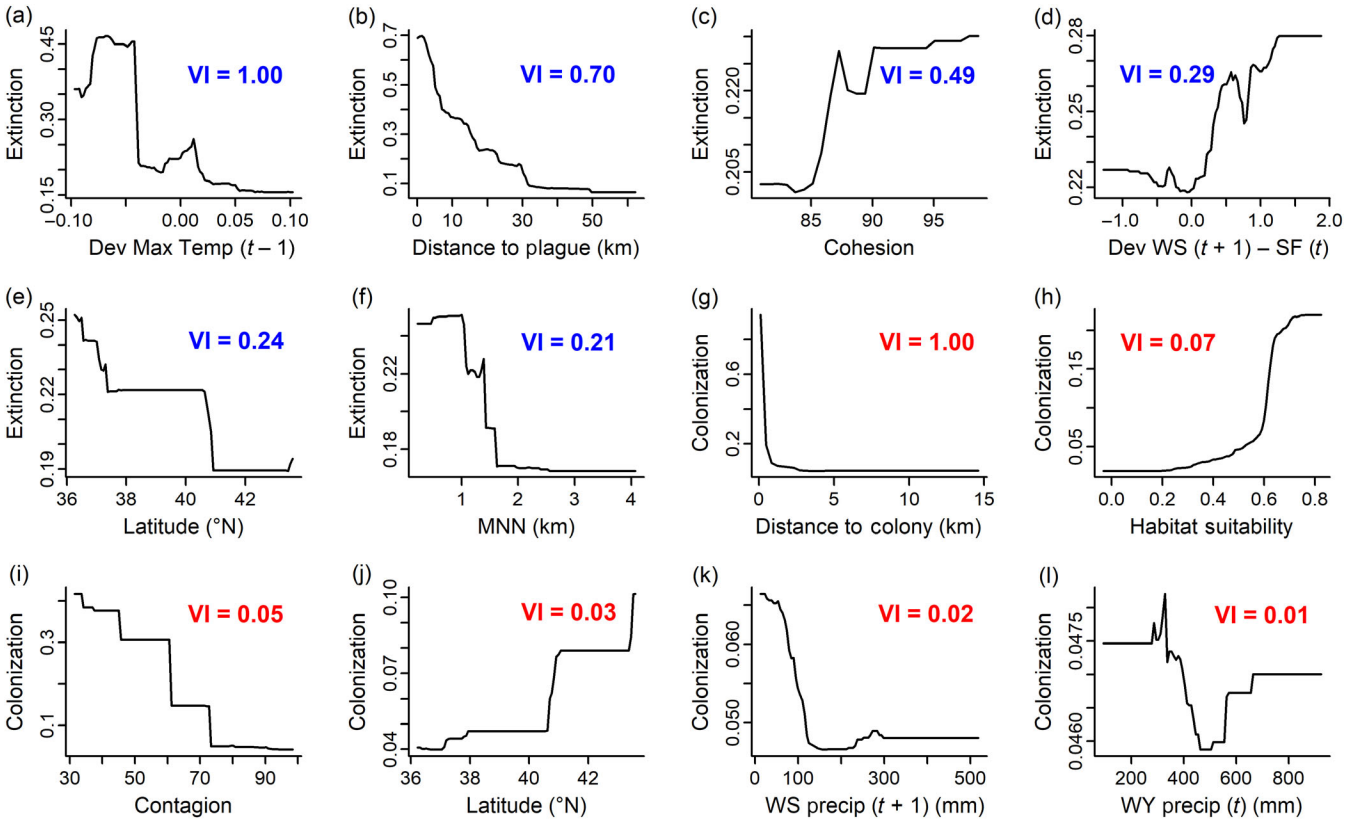

FIGURE 4 Univariate relationships between all covariates included in our final random forest models and (a–f) the probability of extinction due to sylvatic plague or (g–l) colonization of unoccupied habitat during population recovery by black-tailed prairie dogs (BTPDs, *Cynomys ludovicianus*) across our eight study sites between 2001 and 2020. Plotted relationships appear in order of variable importance (VI), with scaled VI values appearing in blue (extinction plots) or red text (colonization plots) in each plot. *Dev max temp* represents the deviation in the average maximum temperature during the hottest months (June–August) of the previous year ($t - 1$) from the long-term average during 2000–2020 (e.g., a negative value denotes a cooler than average summer). *Distance to plague* and *distance to colony* represent the distance to plagued sites from the previous year and the distance to existing colonies from the previous year, respectively. *Cohesion*, *MNN*, and *contagion* describe the spatial configuration of colonies, where *cohesion* measures the connectedness of colonies (higher values signify more aggregated colonies), *MNN* indicates the mean nearest-neighbor distance between colonies, and *contagion* assesses the distribution of colonies (higher values signify more dispersed colonies). *Habitat suitability* indicates the quality of habitat for BTPDs (higher values signify higher-quality habitat). *WS*, *SF*, and *WY* refer to periods over which total precipitation was summarized, where *WS* denotes winter/spring (January–May), *SF* denotes summer/fall (June–September), and *WY* denotes whole year (January–December) for current (t) and previous years ($t - 1$). *WS* was considered also at time $t + 1$, right before colony mapping during the summer in $t + 1$. *Dev WS–SF* therefore represents the extent to which *WS* precipitation at time $t + 1$ deviated from the long-term average *WS* precipitation, compared to the extent to which *SF* precipitation at time t deviated from the long-term average *SF* precipitation. In other words, positive values for *Dev WS–SF* indicate that a drier than average summer/fall was followed by a wetter than average winter/spring. For a more detailed description of covariates (definition and derivation), please refer to Appendix S1: Tables S1 and S2.

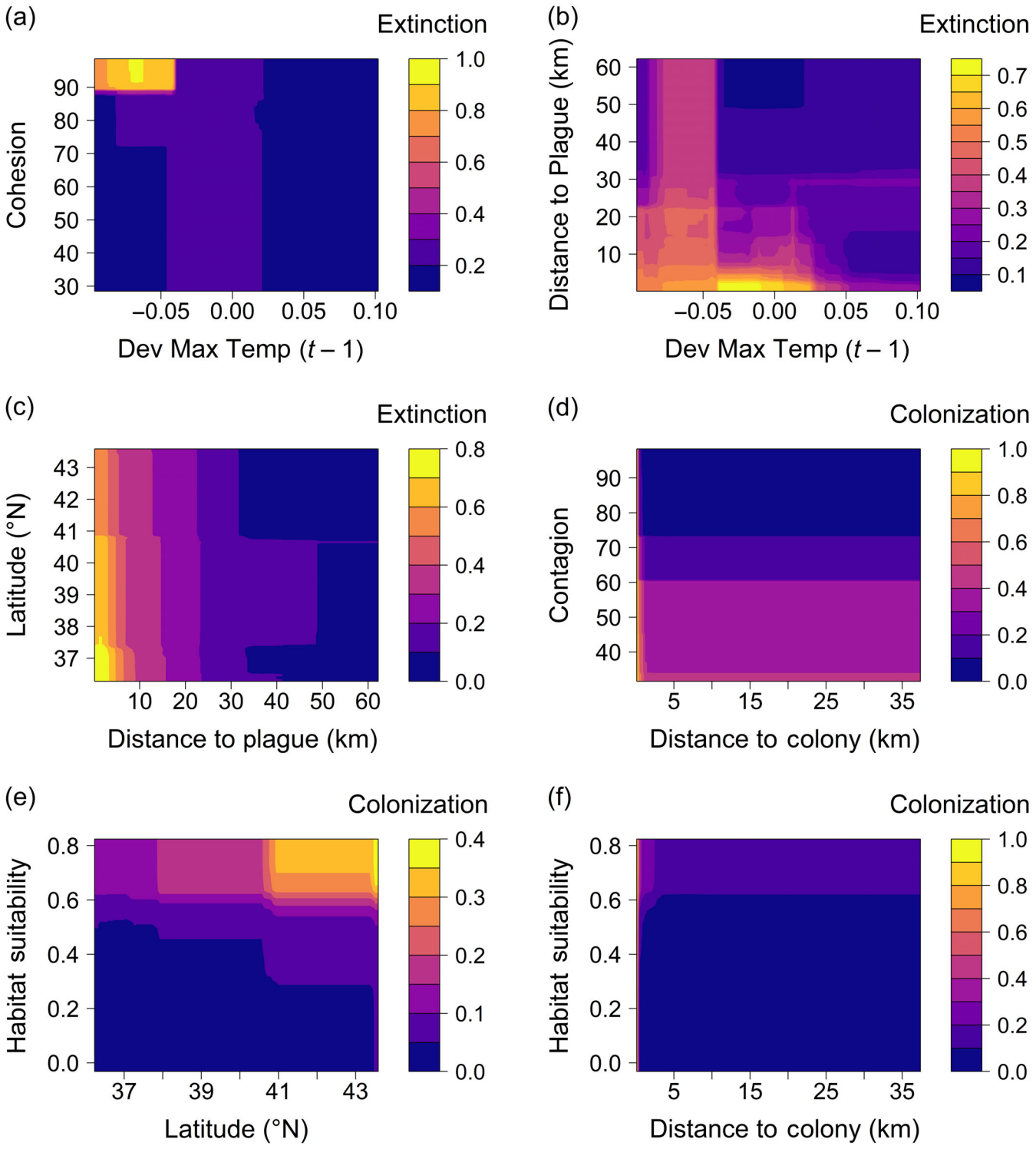

FIGURE 5 Important covariate interactions affecting prediction of (a–c) extinction due to sylvatic plague and (d–f) colonization of unoccupied habitat by black-tailed prairie dogs (BTPDs, *Cynomys ludovicianus*) across our eight study sites between 2001 and 2020. We identified the most important interactions among predictor variables using custom functions developed by coauthor Kevin Shoemaker, which can be accessed via a permanent archive on Zenodo (Shoemaker, 2023). For a detailed description of covariates, please refer to the Figure 4 caption and Appendix S1: Tables S1 and S2.

(Figure 5b). Finally, proximity to previously plagued sites influenced extinctions to a greater extent at lower latitudes (Figure 5c).

Following population crashes, BTPDs expanded into suitable habitat in close proximity to existing colonies (Figure 4g,h). Colonies grew more rapidly at higher

latitudes (Figure 4j), after both wetter and drier years and directly following drier winter/springs (Figure 4k,l), and when colonies were more aggregated in space (Figure 4i). Notably, BTPDs colonized areas separated from existing colonies only when colonies were highly aggregated (denoted by lower contagion values; Figure 5d). Furthermore, the establishment of new colonies isolated from existing colonies was extremely rare and only occurred in high-quality habitat (Figure 5f). Lastly, BTPDs colonized high-quality habitat more readily at higher latitudes (Figure 5e).

Spatial predictions were highly accurate in grasslands with large-scale extinctions due to plague

We compared spatial metrics between actual and predicted colonies at each probability threshold (0.85, 0.90, and 0.95) for all site-years (Table 2; Figure 3). The mean percentage difference (with SEs in parentheses) in total area between predicted and actual colonies was 56.90% (4.03), 31.56% (2.99), and 7.47% (2.83) for thresholds 0.85, 0.90, and 0.95, respectively. In other words, whereas the probability thresholds 0.85 and 0.90 tended to overpredict total colony area, threshold 0.95 produced predictions closest to actual colony area at all NGs (i.e., 0.95 overpredicted the least; Table 2, Figure 6). Furthermore, despite the tendency to overpredict, our model was able to predict which colonies would suffer extinctions due to plague and which would persist with a high degree of accuracy (i.e., overprediction occurred in colonies that persisted on the landscape, not those that experienced extinction; Figure 3).

With respect to exact spatial overlap, the mean percentage of area of actual colonies that did not overlap with model predictions (i.e., false negatives) was 22.91% (1.45), 26.43% (1.53), and 32.52% (1.60) for thresholds 0.85, 0.90, and 0.95, respectively (averaged across all sites and years). Conversely, the mean percentage of area of model predictions that did not overlap spatially with actual colonies (i.e., false positives) was 48.24% (1.36), 42.74% (1.30), and 35.66% (1.35) for thresholds 0.85, 0.90, and 0.95, respectively (averaged across all sites and years). Thus, compared with less conservative thresholds (0.85 and 0.90), more conservative thresholds (0.95) produced more false negatives and fewer false positives (Table 2). Similar to the cross-validation results given earlier, spatial predictions generally were more accurate at NGs that experienced large-scale extinctions due to plague (i.e., Thunder Basin, Pawnee East, Comanche Carrizo, Cimarron, Kiowa, and Rita Blanca) compared with those that experienced smaller-scale die-offs during the study period (i.e., Comanche Timpas, Pawnee West) (Table 2; Figure 6).

DISCUSSION

We successfully developed spatially explicit models that reliably predict local extinctions due to disease and subsequent population recovery in a highly complex host–pathogen system. BTPD colonies exhibited extreme boom-and-bust cycles during our study period (2001–2020) in areas affected by sylvatic plague. The timing and magnitude of extinctions due to plague varied across study sites, which was expected as NGs in our data set encompassed a significant portion of the BTPD range and differed considerably in climatic regime, ecological community composition, ownership pattern, and management planning (Augustine et al., 2021). Despite the uniqueness of each NG, our model retained high predictive accuracy throughout the entire study area where large-scale extinctions occurred, resulting in a generalizable model to forecast the timing and distribution of extreme population fluctuations in areas affected by plague.

Factors driving extinctions due to plague

Consistent with the literature, temporal changes in colony complex area demonstrated that plague likely spreads across the landscape in a wavelike pattern (Johnson et al., 2011; Savage et al., 2011). Our model corroborated this observation, as proximity to previously plagued sites played a large role in determining whether a colony experienced a die-off. The only exception involved extinctions at Thunder Basin NG, where most colonies collapsed during a single event (Davidson, Augustine, Jacobsen, et al., 2022). Covariate interactions supported this outlier, as proximity to previously plagued sites influenced extinctions to a lesser extent at higher latitudes; Thunder Basin represented our highest latitude site. The study area at Thunder Basin was considerably smaller than that at the other NGs in our analysis, however, such that spatial scale may have confounded our interpretation of disease spread at this location.

Extinctions due to plague occurred when colonies were spatially clustered, consistent with the findings of previous work (Johnson et al., 2011; Stapp et al., 2004). However, highly clustered colonies experienced high risk of die-off only after temperatures during the previous summer were unseasonably cool. The relationship observed between temperature and extinctions due to plague may be generalizable in the context of the thermal

TABLE 2 Spatial metrics derived from comparing actual and predicted black-tailed prairie dog (BTPD, *Cynomys ludovicianus*) colonies at three probability thresholds (0.85, 0.90, and 0.95) for all sites (Figure 1) and years. Metrics included the mean percentage difference in total area between predicted and actual colonies (total area), mean percentage of area of actual colonies that did not overlap spatially with model predictions (ANP), and the mean percentage of area of model predictions that did not overlap spatially with actual colonies (PNA). SEs appear in parentheses.

National grassland	Years	Threshold	Total area	ANP	PNA
Thunder Basin	2001–2020	0.85	53.00 (12)	27.00 (7)	51.10 (4)
Thunder Basin	2001–2020	0.90	23.05 (10)	31.72 (7)	46.46 (4)
Thunder Basin	2001–2020	0.95	−0.04 (08)	38.84 (7)	39.05 (5)
Pawnee West	2001–2015	0.85	70.40 (12)	22.06 (2)	51.98 (3)
Pawnee West	2001–2015	0.90	45.88 (10)	25.45 (2)	46.64 (3)
Pawnee West	2001–2015	0.95	15.34 (07)	31.91 (2)	39.04 (3)
Pawnee East	2001–2015	0.85	48.69 (11)	23.67 (4)	45.97 (4)
Pawnee East	2001–2015	0.90	27.95 (08)	26.51 (4)	40.90 (4)
Pawnee East	2001–2015	0.95	−2.16 (06)	33.45 (4)	31.41 (3)
Comanche Timpas	2001–2015	0.85	55.62 (12)	29.77 (3)	52.74 (3)
Comanche Timpas	2001–2015	0.90	25.56 (10)	33.84 (3)	45.20 (3)
Comanche Timpas	2001–2015	0.95	−5.37 (07)	39.54 (3)	34.95 (2)
Comanche Carrizo	2001–2016	0.85	53.74 (09)	20.27 (3)	45.84 (4)
Comanche Carrizo	2001–2016	0.90	30.78 (06)	23.86 (3)	40.43 (4)
Comanche Carrizo	2001–2016	0.95	8.98 (05)	29.93 (4)	34.69 (4)
Cimarron	2001–2017	0.85	59.08 (12)	24.83 (4)	48.33 (5)
Cimarron	2001–2017	0.90	31.17 (08)	27.70 (5)	42.39 (5)
Cimarron	2001–2017	0.95	18.85 (13)	32.18 (5)	38.66 (5)
Kiowa	2001–2017	0.85	51.45 (10)	12.88 (2)	40.21 (3)
Kiowa	2001–2017	0.90	33.45 (08)	16.00 (2)	35.23 (3)
Kiowa	2001–2017	0.95	10.88 (05)	21.91 (3)	28.21 (3)
Rita Blanca	2001–2017	0.85	63.94 (14)	21.35 (3)	48.85 (4)
Rita Blanca	2001–2017	0.90	36.73 (06)	24.69 (3)	43.55 (3)
Rita Blanca	2001–2017	0.95	13.93 (07)	30.54 (3)	37.81 (3)

Note: Spatial metrics (total area, ANP, PNA) evaluated error between predicted and actual colonies to determine how well our models forecast colony distribution in time and space.

mismatch hypothesis, which is well supported in other wildlife disease systems (e.g., Cohen et al., 2017, 2020). This hypothesis posits that cold- and warm-adapted hosts are at greatest risk of infection under abnormally warm and cool conditions, respectively, because smaller-bodied pathogens generally have broader thermal performance curves than larger-bodied hosts (Cohen et al., 2019; Nowakowski et al., 2016). Indeed, the introduced plague bacterium is tolerant of a wide range of temperatures (Gage & Kosoy, 2005) and perhaps performs more optimally under abnormal conditions than locally adapted BTPD hosts. Furthermore, fleas tend to be more abundant during milder weather conditions and may transmit *Y. pestis* more efficiently at lower temperatures (Stenseth et al., 2006; Williams et al., 2013). Our finding that BTPD colonies experienced plague die-offs following abnormally cool summers is, therefore, consistent with the thermal mismatch hypothesis. However, in the BTPD–plague system, temperature relations between pathogen and vector may be more influential than interactions between pathogen and host. For instance, fleas are less able to clear gut-blocked plague infection at temperatures below 28°C (Gage & Kosoy, 2005), which suggests that plague outbreaks are more likely below that temperature (Collinge, Johnson, Ray, Matchett, Grensten, Cully, Gage, et al., 2005; Savage et al., 2011).

Our results support the hypothesis (Eads & Biggins, 2017) that transitions from dry to wet years increase the probability of plague outbreaks (see also Davidson, Augustine, Jacobsen, et al., 2022; Eads & Hoogland, 2016). We

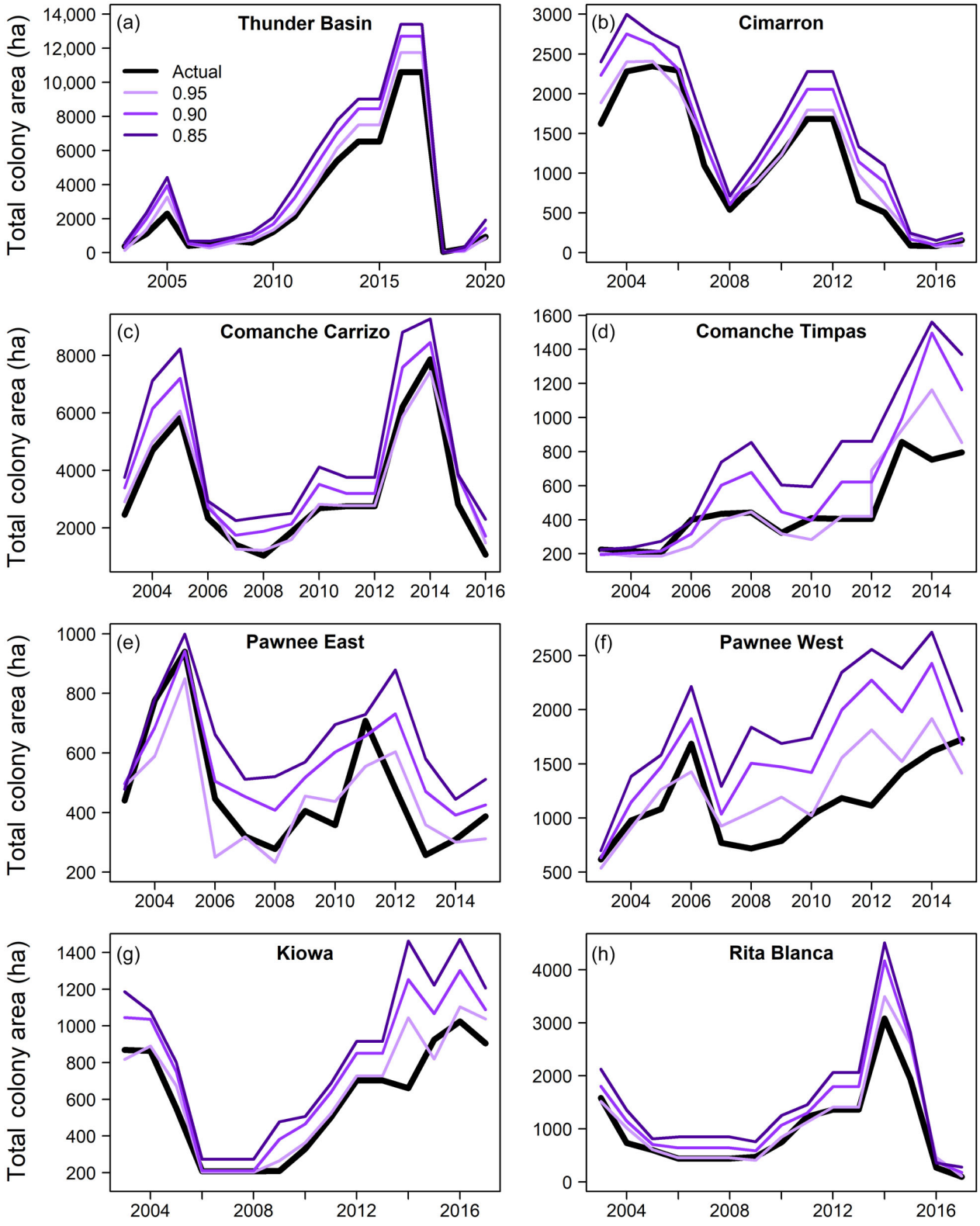

FIGURE 6 Total area (ha) of actual black-tailed prairie dog (BTPD; *Cynomys ludovicianus*) colonies mapped over time at each National Grassland (thick black lines) compared with the total area predicted by our final models at each probability threshold: 0.85 (dark purple lines), 0.90 (medium purple lines), and 0.95 (light purple lines). Probability thresholds were used to reclassify colonization and persistence layers into binary rasters prior to merging them into one raster of predicted colony distribution at time $t + 1$ (Figure 2f–h). For example, 0.95 refers to both colonization and extinction probabilities, above which values were classified as 1 (i.e., those raster cells were assumed to be colonized or to go extinct, respectively) when predicting colony distribution and ultimately calculating total colony area. Flat lines indicate where surveys did not occur or partial surveys were averaged.

found that dry summers/falls in year t followed by wet winter/springs in year $t + 1$ increased the probability of extinction due to plague between year t and $t + 1$. Although drought can suppress both fleas and *Y. pestis*, dry conditions also stress BTPDs, reducing their body condition and creating the potential for future increases in flea loads (Eads et al., 2016). Then, when moisture returns during the following year, conditions favor greater flea densities and aboveground activity of hosts, both of which contribute to increased plague transmission (Eads & Biggins, 2017). Milder summer temperatures in year $t - 1$ may exacerbate moisture conditions by promoting flea abundance, which may result in hosts becoming ridden with fleas during drier summer/falls in year t when BTPDs are stressed and reduce grooming behavior (Eads, 2014; Eads & Biggins, 2017). Covariate interactions support the idea that cooler summer temperatures in year $t - 1$ trigger epizootics between t and $t + 1$ since, in the present study, the effect of this temperature variable on extinction did not depend greatly on proximity to previously plagued sites. Following the initiation of an epizootic, distance to previously plagued sites then likely plays a key role in colony die-offs as plague spreads across the landscape.

Extinctions due to plague responded to relative deviations from average weather conditions more than the absolute values of weather variables. Indeed, abnormal weather conditions can increase host susceptibility to disease, whether that be extreme weather-related events or longer-term climatic changes like prolonged drought (Bruno et al., 2007; Cazelles et al., 2005; Koelle et al., 2005; Pascual et al., 2008). During meteorological anomalies, for instance, pathogens and vectors can exploit the disruption of environmental conditions and their consequent impact on host biology (McMichael, 2015; Rohr et al., 2011). Our study sites encompassed a large geographical area, including a north–south temperature gradient and an east–west precipitation gradient, such that extreme conditions at one site may constitute average conditions at another. Relative variables therefore were important in our plague modeling, for both practical and ecological reasons. In future studies, we encourage researchers to consider deviations from typical weather conditions when modeling host–pathogen dynamics, particularly when predicting outbreaks over a large geographical area with considerable variation in climatic regimes.

Factors driving colony growth

Seasonal precipitation patterns influenced colony recovery after epizootic events. Colonies expanded into suitable habitat more rapidly immediately following drier winter/springs. Drier conditions likely limit plant growth, such that BTPDs can visually detect potential predators, leading to greater population size and area occupied by each colony (Augustine et al., 2008). Alternatively, under drought stress, BTPDs may require larger home ranges to obtain sufficient food resources, so densities can decline even when their colonies expand (Bruggeman & Licht, 2020). When considering the total precipitation during year t, colonies grew more after both dry and wet years, which may reflect a latitudinal effect. For instance, BTPD reproduction improved in the Northern Plains following wetter years (Grassel et al., 2016; Stephens et al., 2018), whereas the same amount of precipitation in the Central Plains may diminish the ability of BTPDs to effectively detect predators (e.g., high plant growth; Augustine et al., 2008).

Colonies expanded into nearby suitable habitat more rapidly once established following an epizootic event, which was indicative of exponential growth during the recovery phase. Establishment of new colonies isolated from existing colonies was relatively rare and predicted to occur only in high-quality habitat. Interestingly, BTPDs colonized high-quality habitat more frequently at higher latitudes, namely the latitude represented by Thunder Basin NG (i.e., 43–44° N). Perhaps Thunder Basin offers greater availability of high-quality habitat compared with other grasslands in our data set, which may explain why colonies expanded to much larger areas at Thunder Basin before experiencing plague die-offs.

Spatial prediction accuracy

Rigorous cross-validations and spatially explicit predictions indicated that our colony growth model predicted dynamic growth patterns with high accuracy at all eight NGs, albeit with slight overpredictions of total colony area in each system (e.g., possibly because we did not incorporate predator effects into growth models). Conversely, our plague model predicted more accurately at those NGs that experienced large-scale extinctions due to plague compared with those that experienced smaller-scale die-offs. The use of our plague model, therefore, is more appropriate for areas experiencing regular boom-and-bust cycles due to plague, and we recommend proceeding with caution when using this model to predict to sites without large-scale epizootics.

Caveats

As stated earlier in the *Methods* section, we acknowledge that population fluctuations in BTPDs did not necessarily equate to disease dynamics, as we did not have

concomitant disease data during each colony die-off. Although we were confident that large-scale extinctions were due to plague in our study, this confidence may be restricted to our study area (i.e., central Great Plains), where much work has confirmed the pattern of spread and effects of plague on BTPD colonies (e.g., Cully et al., 2010; Johnson et al., 2011; Matchett et al., 2021). Thus, researchers should exercise caution when applying our model to grasslands outside of our study area, such as more southern grasslands where drought can result in colony contractions (e.g., Ceballos et al., 2010; Facka et al., 2010). We also caution researchers when applying our model to systems wherein natural population cycles involve extreme fluctuations (unrelated to disease), such as in lemmings (e.g., *Lemmus lemmus*) and snowshoe hares (*Lepus americanus*) (Ims et al., 2011; Krebs et al., 2001).

Applications

Our final models and overall approach offer many practical applications. First, researchers could produce spatial predictions as demonstrated in this study, wherein we organized population data and georefenced covariate values from time t and used each model to create predictive raster surfaces of extinction risk and population growth during $t + 1$. Although our population data consisted of colony shapefiles (with polygons to designate colony boundaries), one could easily incorporate other population data types such as host occurrence (i.e., presence/absence) or abundance. If desired, users could define probability thresholds to combine persistence (1 − extinction risk) and colonization (population growth) into a single prediction of population distribution at time $t + 1$. A threshold of 0.95 for colonization and extinction produced predictions closest to actual colony distribution, though managers could adopt different threshold values depending on their tolerance for false positives/negatives.

Our model could be used to decide when and where to deploy available mitigation options. Digitally removing colonies or portions of colonies mimics poisoning with rodenticide (Knowles et al., 2002), which could alter the values of colony connectedness variables in the model (*cohesion* and *mean nearest neighbor*) and recalculate the probability of extinction. This process of digitally "poisoning" colonies (or treating them with insecticide, which effectively prevents them from spreading disease; Poché et al., 2017; Tripp et al., 2017) could be repeated until the probability of extinction falls below a desired threshold. These types of exercises could assist in allocating limited and costly mitigation resources for BTPD management. Ultimately, the model could be used to support strategic management planning that reduces the volatility of boom-and-bust cycles in BTPD. This may help maximize the benefits of this keystone species to associated wildlife communities and ecosystem functioning, while also minimizing costs to the ranching industry (Augustine & Derner, 2021; Crow et al., 2022; Sierra–Corona et al., 2015).

Conclusions

Introduced pathogens cause extreme fluctuations in their host populations across many disease systems (e.g., cowpox virus in rodents: Smith et al., 2008; chytrid fungus in frogs: Newell et al., 2013; pneumonia pathogens in bighorn sheep: Plowright et al., 2013). Using the BTPD–plague system, we demonstrated how a big data–model integration approach could reliably forecast the timing and location of local extinctions due to disease and the rate at which populations are likely to recover. Our approach can be applied to any wildlife disease system, particularly when spatially explicit predictions could guide management decisions. By employing machine learning algorithms to forecast host population fluctuations, our study leveraged data science techniques to bridge the gap between population demography and disease ecology, resulting in a framework for future study of infectious diseases in wildlife populations.

AUTHOR CONTRIBUTIONS

Gabriel M. Barrile wrote the original draft, reviewed and edited subsequent drafts, and contributed to conceptualization, data analysis, and data visualization. Ana D. Davidson, David J. Augustine, Lauren M. Porensky, Courtney J. Duchardt, Kevin T. Shoemaker, Cynthia R. Hartway, Justin D. Derner, and Elizabeth A. Hunter contributed to conceptualization, funding acquisition, and data collection and analysis and co-authored the manuscript. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. government.

ACKNOWLEDGMENTS

We are grateful to Paul Stapp, Eddie Childers, Buck Cully, Jeff Abeglen, Kristin Philbrook, Janice Naylor, Erin Considine, Andy Chappell, Jennifer Avising, Ruben Mares, Gwyn McKee, Peter McDonald, Travis Livieri, Lynne Deibel, Jessica Alexander, Daniel Kinka, Kristy Bly, Nicole Rosmarino, Tara Stephens, and numerous other biologists and field technicians for their data collection and sharing of prairie dog colony boundaries and other associated data across the Great Plains region. We thank David Eads, David Pellatz, and two anonymous

reviewers for providing comments that helped to greatly improve the manuscript. This work is part of the Thunder Basin Research Initiative. Funding was provided by the National Institute of Food and Agriculture (Award no.: 2020-67019-31153).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

We used novel code to identify and visualize the most important covariate interactions influencing the collapse and recovery of BTPD (*C. ludovicianus*) colonies challenged by sylvatic plague (*Y. pestis*). Novel code consisted of custom functions developed by coauthor Kevin Shoemaker. All novel code is archived permanently on Zenodo (Shoemaker, 2023) and available at https://doi.org/10.5281/zenodo.7523381. Data (Barrile et al., 2023) are available on Dryad at https://doi.org/10.5061/dryad.63xsj3v6g. These data include the final data sets used in model fitting. The full data sets (i.e., colony shapefiles) supporting this research are sensitive and not available publicly. Data are sensitive insofar as they contain detailed spatial information on a species of conservation concern. Colony shapefiles at Thunder Basin NG are available to qualified researchers by contacting Tim Byer (Wildlife Biologist, United States Department of Agriculture US Forest Service; e-mail: timothy.byer@usda.gov). Colony shapefiles from all other NGs used in this study are available to qualified researchers by contacting Kevin Shoemaker (Associate Professor, University of Nevada, Reno, Department of Natural Resources and Environmental Science; e-mail: kevinshoemaker@unr.edu). Additional data sets utilized for this research are as follows: the Daymet (Sadoti et al., 2020) data product, which provides gridded estimates of daily climate data at a 1-km spatial resolution; the POLARIS (Moro Rosso et al., 2021a, 2021b) data product, which provides soil texture and other attributes at a 30-m resolution; the National Elevation Dataset (U.S. Geological Survey, 2022), which provides elevation data at a one-third arc-second resolution. Finally, we used a habitat suitability model that contains detailed spatial information on a sensitive species and thus is not available publicly; these data are available to qualified researchers from Ana Davidson (Colorado State University, Colorado Natural Heritage Program; e-mail: ana.davidson@colostate.edu).

ORCID

Gabriel M. Barrile https://orcid.org/0000-0003-2632-9034
David J. Augustine https://orcid.org/0000-0003-3144-0466
Lauren M. Porensky https://orcid.org/0000-0001-6883-2442
Courtney J. Duchardt https://orcid.org/0000-0003-4563-0199
Kevin T. Shoemaker https://orcid.org/0000-0002-3789-3856
Cynthia R. Hartway https://orcid.org/0000-0002-3806-278X
Justin D. Derner https://orcid.org/0000-0001-8076-0736
Elizabeth A. Hunter https://orcid.org/0000-0003-4710-167X
Ana D. Davidson https://orcid.org/0000-0003-4701-5923

REFERENCES

Ambroise, C., and G. J. McLachlan. 2002. "Selection Bias in Gene Extraction on the Basis of Microarray Gene-Expression Data." *Proceedings of the National Academy of Sciences* 99(10): 6562–6. https://doi.org/10.1073/pnas.102102699.

Antolin, M. F., P. Gober, B. Luce, D. E. Biggins, W. E. Van Pelt, D. B. Seery, M. Lockhart, et al. 2002. "The Influence of Sylvatic Plague on North American Wildlife at the Landscape Level, with Special Emphasis on Black-Footed Ferret and Prairie Dog Conservation." *Transactions of the North American Wildlife and Natural Resources Conference* 67: 104–27.

Augustine, D., A. Davidson, K. Dickinson, and B. Van Pelt. 2021. "Thinking like a Grassland: Challenges and Opportunities for Biodiversity Conservation in the Great Plains of North America." *Rangeland Ecology and Management* 78: 281–95. https://doi.org/10.1016/j.rama.2019.09.001.

Augustine, D. J., and J. D. Derner. 2021. "Long-Term Effects of Black-Tailed Prairie Dogs on Livestock Grazing Distribution and Mass Gain." *The Journal of Wildlife Management* 85(7): 1332–43. https://doi.org/10.1002/jwmg.22103.

Augustine, D. J., M. R. Matchett, T. P. Toombs, J. F. Cully, T. L. Johnson, and J. G. Sidle. 2008. "Spatiotemporal Dynamics of Black-Tailed Prairie Dog Colonies Affected by Plague." *Landscape Ecology* 23(3): 255–67. https://doi.org/10.1007/s10980-007-9175-6.

Augustine, D. J., and S. K. Skagen. 2014. "Mountain Plover Nest Survival in Relation to Prairie Dog and Fire Dynamics in Shortgrass Steppe." *The Journal of Wildlife Management* 78(4): 595–602. https://doi.org/10.1002/jwmg.700.

Babayan, S. A., R. J. Orton, and D. G. Streicker. 2018. "Predicting Reservoir Hosts and Arthropod Vectors from Evolutionary Signatures in RNA Virus Genomes." *Science* 362(6414): 577–80. https://doi.org/10.1126/science.aap9072.

Barnes, A. M. 1993. "A Review of Plague and its Relevance to Prairie Dog Populations and the Black-Footed Ferret." In *Management of Prairie Dog Complexes for the Reintroduction of the Black-Footed Ferret*, 28–38. Washington, DC: US Fish and Wildlife Service Biological Report, 13.

Barrile, G., D. Augustine, L. Porensky, C. Duchardt, K. Shoemaker, C. Hartway, J. Derner, E. Hunter, and A. Davidson. 2023. "A Big Data–Model Integration Approach for Predicting Epizootics and Population Recovery in a Keystone Species." Dryad, Dataset. https://doi.org/10.5061/dryad.63xsj3v6g.

Basinski, A. J., E. Fichet-Calvet, A. R. Sjodin, T. J. Varrelman, C. H. Remien, N. C. Layman, and B. H. Bird. 2021. "Bridging the Gap: Using Reservoir Ecology and Human Serosurveys to Estimate Lassa Virus Spillover in West Africa." *PLoS Computational Biology* 17(3): e1008811. https://doi.org/10.1371/journal.pcbi.1008811.

Ben Ari, T., A. Gershunov, K. L. Gage, T. Snäll, P. Ettestad, K. L. Kausrud, and N. C. Stenseth. 2008. "Human Plague in the USA: The Importance of Regional and Local Climate." *Biology Letters* 4(6): 737–40. https://doi.org/10.1098/rsbl.2008.0363.

Biggins, D. E., and D. A. Eads. 2019. "Prairie Dogs, Persistent Plague, Flocking Fleas, and Pernicious Positive Feedback." *Frontiers in Veterinary Science* 6: 75. https://doi.org/10.3389/fvets.2019.00075.

Bocinsky, R. K. 2021. "FedData: Functions to Automate Downloading Geospatial Data Available from Several Federated Data Sources." R Package Version 3.0.1.9000. https://CRAN.R-project.org/package=FedData.

Breiman, L. 2001. "Random forests." *Machine Learning* 45(1): 5–32. https://doi.org/10.1023/A:1010933404324.

Bruggeman, J. E., and D. S. Licht. 2020. "Drought-Mediated Changes in Black-Tailed Prairie Dog Colonies in the Northern Great Plains." *Journal of Mammalogy* 101(4): 1189–204. https://doi.org/10.1093/jmammal/gyaa070.

Bruno, J. F., E. R. Selig, K. S. Casey, C. A. Page, B. L. Willis, C. D. Harvell, H. Sweatman, and A. M. Melendy. 2007. "Thermal Stress and Coral Cover as Drivers of Coral Disease Outbreaks." *PLoS Biology* 5(6): e124. https://doi.org/10.1371/journal.pbio.0050124.

Cazelles, B., M. Chavez, A. J. McMichael, and S. Hales. 2005. "Nonstationary Influence of El Nino on the Synchronous Dengue Epidemics in Thailand." *PLoS Medicine* 2(4): e106. https://doi.org/10.1371/journal.pmed.0020106.

Ceballos, G., A. Davidson, R. List, J. Pacheco, P. Manzano-Fischer, G. Santos-Barrera, and J. Cruzado. 2010. "Rapid Decline of a Grassland System and its Ecological and Conservation Implications." *PLoS One* 5(1): e8562. https://doi.org/10.1371/journal.pone.0008562.

Chaney, N. W., B. Minasny, J. D. Herman, T. W. Nauman, C. W. Brungard, C. L. Morgan, A. B. McBratney, E. F. Wood, and Y. Yimam. 2019. "POLARIS Soil Properties: 30-m Probabilistic Maps of Soil Properties over the Contiguous United States." *Water Resources Research* 55(4): 2916–38. https://doi.org/10.1029/2018WR022797.

Chen, C., A. Liaw, and L. Breiman. 2004. *Using Random Forest to Learn Imbalanced Data* 24. Berkeley, CA: University of California.

Cohen, J. M., D. J. Civitello, M. D. Venesky, T. A. McMahon, and J. R. Rohr. 2019. "An Interaction between Climate Change and Infectious Disease Drove Widespread Amphibian Declines." *Global Change Biology* 25(3): 927–37. https://doi.org/10.1111/gcb.14489.

Cohen, J. M., E. L. Sauer, O. Santiago, S. Spencer, and J. R. Rohr. 2020. "Divergent Impacts of Warming Weather on Wildlife Disease Risk across Climates." *Science* 370(6519): eabb1702. https://doi.org/10.1126/science.abb1702.

Cohen, J. M., M. D. Venesky, E. L. Sauer, D. J. Civitello, T. A. McMahon, E. A. Roznik, and J. R. Rohr. 2017. "The Thermal Mismatch Hypothesis Explains Host Susceptibility to an Emerging Infectious Disease." *Ecology Letters* 20(2): 184–93. https://doi.org/10.1111/ele.12720.

Collinge, S. K., W. C. Johnson, C. Ray, R. Matchett, J. Grensten, J. F. Cully, K. Gage, M. Kosoy, J. Loye, and A. P. Martin. 2005a. "Testing the Generality of a Trophic-Cascade Model for Plague." *EcoHealth* 2(2): 102–12. https://doi.org/10.1007/s10393-005-3877-5.

Collinge, S. K., W. C. Johnson, C. Ray, R. Matchett, J. Grensten, J. F. Cully, Jr., and A. P. Martin. 2005b. "Landscape Structure and Plague Occurrence in Black-Tailed Prairie Dogs on Grasslands of the Western USA." *Landscape Ecology* 20(8): 941–55. https://doi.org/10.1007/s10980-005-4617-5.

Colman, R. E., R. J. Brinkerhoff, J. D. Busch, C. Ray, A. Doyle, J. W. Sahl, P. Keim, S. K. Collinge, and D. M. Wagner. 2021. "No Evidence for Enzootic Plague within Black-Tailed Prairie Dog (*Cynomys ludovicianus*) Populations." *Integrative Zoology* 16(6): 834–51. https://doi.org/10.1111/1749-4877.12546.

Connell, L. C., L. M. Porensky, and J. D. Scasta. 2019. "Prairie Dog (*Cynomys ludovicianus*) Influence on Forage Quantity and Quality in a Grazed Grassland-Shrubland Ecotone." *Rangeland Ecology and Management* 72(2): 360–73. https://doi.org/10.1016/j.rama.2018.10.004.

Crow, L., L. M. Porensky, D. Augustine, J. Ritten, C. T. Bastian, and S. I. Paisley. 2022. "Evaluating Prairie Dog–Cattle Competition from the Perspective of a Ranching Enterprise in the Western Great Plains: Economic Analysis of Potential Effects on Long-Term Profitability." *Rangeland Ecology & Management* 85: 56–65. https://doi.org/10.1016/j.rama.2022.09.003.

Cully, J. F., T. L. Johnson, S. K. Collinge, and C. Ray. 2010. "Disease Limits Populations: Plague and Black-Tailed Prairie Dogs." *Vector Borne and Zoonotic Diseases* 10(1): 7–15. https://doi.org/10.1089/vbz.2009.0045.

Cully, J. F., and E. S. Williams. 2001. "Interspecific Comparisons of Sylvatic Plague in Prairie Dogs." *Journal of Mammalogy* 82(4): 894–905. https://doi.org/10.1644/1545-1542(2001)082<0894:ICOSPI>2.0.CO;2.

Daszak, P., A. A. Cunningham, and A. D. Hyatt. 2000. "Emerging Infectious Diseases of Wildlife–Threats to Biodiversity and Human Health." *Science* 287(5452): 443–9. https://doi.org/10.1126/science.287.5452.443.

Davidson, A. D., D. Augustine, H. Jacobsen, D. Pellatz, L. M. Porensky, G. McKee, and C. Duchardt. 2022. "Boom and Bust Cycles of Black-Tailed Prairie Dog Populations in the Thunder Basin Grassland Ecosystem." *Journal Of Mammology* 103(5): 1112–26. https://doi.org/10.1093/jmammal/gyac035.

Davidson, A. D., D. Augustine, M. Menefee, M. Fink, L. Sterling-Krank, and B. Van Pelt. 2022. "Part I: Habitat Suitability Model for the Black-Tailed Prairie Dog Ecosystem." In *Final Report: Identifying Potential Landscapes for Conservation across the Grasslands of North America: Integrating Keystone Species, Land Use Patterns, and Climate Change to Enhance Current and Future Grassland Restoration Efforts*, edited by A. D. Davidson, D. Augustine, M. Menefee, L. Sterling-Krank, and B. Van Pelt. Fort Collins, CO: Colorado State University.

Davidson, A. D., J. K. Detling, and J. H. Brown. 2012. "Ecological Roles and Conservation Challenges of Social, Burrowing, Herbivorous Mammals in the world's Grasslands." *Frontiers in*

Ecology and the Environment 10(9): 477–86. https://doi.org/10.1890/110054.

Davidson, A. D., M. T. Friggens, K. T. Shoemaker, C. L. Hayes, J. Erz, and R. Duran. 2014. "Population Dynamics of Reintroduced Gunnison's Prairie Dogs in the Southern Portion of their Range." *The Journal of Wildlife Management* 78(3): 429–39. https://doi.org/10.1002/jwmg.681.

Duchardt, C. J., D. J. Augustine, L. M. Porensky, J. L. Beck, J. D. Hennig, D. W. Pellatz, J. D. Scasta, L. C. Connell, and A. D. Davidson. 2022. "Disease and Weather Induce Rapid Shifts in a Rangeland Ecosystem Mediated by a Keystone Species (*Cynomys ludovicianus*)." *Ecological Applications* 33: e2712. https://doi.org/10.1002/eap.2712.

Eads, D. A. 2014. *Factors Affecting Flea Densities in Prairie Dog Colonies: Implications for the Maintenance and Spread of Plague*. Fort Collins, CO: Colorado State University.

Eads, D. A., and D. E. Biggins. 2015. "Plague Bacterium as a Transformer Species in Prairie Dogs and the Grasslands of Western North America." *Conservation Biology* 29(4): 1086–93. https://doi.org/10.1111/cobi.12498.

Eads, D. A., and D. E. Biggins. 2017. "Paltry Past-Precipitation: Predisposing Prairie Dogs to Plague?" *The Journal of Wildlife Management* 81(6): 990–8. https://doi.org/10.1002/jwmg.21281.

Eads, D. A., and D. E. Biggins. 2019. "Plague Management of Prairie Dog Colonies: Degree and Duration of Deltamethrin Flea Control." *Journal of Vector Ecology* 44(1): 40–7. https://doi.org/10.1111/jvec.12327.

Eads, D. A., D. E. Biggins, D. H. Long, K. L. Gage, and M. F. Antolin. 2016. "Droughts May Increase Susceptibility of Prairie Dogs to Fleas: Incongruity with Hypothesized Mechanisms of Plague Cycles in Rodents." *Journal of Mammalogy* 97(4): 1044–53. https://doi.org/10.1093/jmammal/gyw035.

Eads, D. A., and J. L. Hoogland. 2016. "Factors that Affect Parasitism of Black-Tailed Prairie Dogs by Fleas." *Ecosphere* 7(7): e01372. https://doi.org/10.1002/ecs2.1372.

Evans, J. S., and S. A. Cushman. 2009. "Gradient Modeling of Conifer Species Using Random Forests." *Landscape Ecology* 24(5): 673–83. https://doi.org/10.1007/s10980-009-9341-0.

Evans, J. S., M. A. Murphy, Z. A. Holden, and S. A. Cushman. 2011. "Modeling Species Distribution and Change Using Random Forest." In *Predictive Species and Habitat Modeling in Landscape Ecology*, edited by C. Drew, Y. Wiersma, and F. Huettmann, 139–59. New York, NY: Springer. https://doi.org/10.1007/978-1-4419-7390-0_8.

Evans, J. S., and K. Ram. 2021. "Package 'spatialEco'." R CRAN Project.

Evans, M. V., T. A. Dallas, B. A. Han, C. C. Murdock, and J. M. Drake. 2017. "Data-Driven Identification of Potential Zika Virus Vectors." *eLife* 6: e22053. https://doi.org/10.7554/eLife.22053.

Facka, A. N., G. W. Roemer, V. L. Mathis, M. Kam, and E. Geffen. 2010. "Drought Leads to Collapse of Black-Tailed Prairie Dog Populations Reintroduced to the Chihuahuan Desert." *The Journal of Wildlife Management* 74(8): 1752–62. https://doi.org/10.2193/2009-208.

Fawcett, T. 2006. "An Introduction to ROC Analysis." *Pattern Recognition Letters* 27(8): 861–74. https://doi.org/10.1016/j.patrec.2005.10.010.

Fountain-Jones, N. M., G. Machado, S. Carver, C. Packer, M. Recamonde-Mendoza, and M. E. Craft. 2019. "How to Make More from Exposure Data? An Integrated Machine Learning Pipeline to Predict Pathogen Exposure." *Journal of Animal Ecology* 88(10): 1447–61. https://doi.org/10.1111/1365-2656.13076.

Fox, E. W., R. A. Hill, S. G. Leibowitz, A. R. Olsen, D. J. Thornbrugh, and M. H. Weber. 2017. "Assessing the Accuracy and Stability of Variable Selection Methods for Random Forest Modeling in Ecology." *Environmental Monitoring and Assessment* 189(7): 1–20. https://doi.org/10.1007/s10661-017-6025-0.

Freeman, E. A., G. G. Moisen, and T. S. Frescino. 2012. "Evaluating Effectiveness of Down-Sampling for Stratified Designs and Unbalanced Prevalence in Random Forest Models of Tree Species Distributions in Nevada." *Ecological Modelling* 233: 1–10. https://doi.org/10.1016/j.ecolmodel.2012.03.007.

Gage, K. L., and M. Y. Kosoy. 2005. "Natural History of Plague: Perspectives from More than a Century of Research." *Annual Review of Entomology* 50: 505–28. https://doi.org/10.1146/annurev.ento.50.071803.130337.

George, D. B., C. T. Webb, K. M. Pepin, L. T. Savage, and M. F. Antolin. 2013. "Persistence of Black-Tailed Prairie-Dog Populations Affected by Plague in Northern Colorado, USA." *Ecology* 94(7): 1572–83. https://doi.org/10.1890/12-0719.1.

Goulson, D., E. Nicholls, C. Botías, and E. L. Rotheray. 2015. "Bee Declines Driven by Combined Stress from Parasites, Pesticides, and Lack of Flowers." *Science* 347(6229): 1255957. https://doi.org/10.1126/science.12559.

Grassel, S. M., J. L. Rachlow, and C. J. Williams. 2016. "Reproduction by Black-Tailed Prairie Dogs and Black-Footed Ferrets: Effects of Weather and Food Availability." *Western North American Naturalist* 76(4): 405–16. https://doi.org/10.3398/064.076.0403.

Han, B. A., and J. M. Drake. 2016. "Future Directions in Analytics for Infectious Disease Intelligence: Toward an Integrated Warning System for Emerging Pathogens." *EMBO Reports* 17(6): 785–9. https://doi.org/10.15252/embr.201642534.

Han, B. A., S. M. O'Regan, J. Paul Schmidt, and J. M. Drake. 2020. "Integrating Data Mining and Transmission Theory in the Ecology of Infectious Diseases." *Ecology Letters* 23(8): 1178–88. https://doi.org/10.1111/ele.13520.

Han, B. A., J. P. Schmidt, S. E. Bowden, and J. M. Drake. 2015. "Rodent Reservoirs of Future Zoonotic Diseases." *Proceedings of the National Academy of Sciences* 112(22): 7039–44. https://doi.org/10.1073/pnas.1501598112.

Hartley, L. M., J. K. Detling, and L. T. Savage. 2009. "Introduced Plague Lessens the Effects of an Herbivorous Rodent on Grassland Vegetation." *Journal of Applied Ecology* 46(4): 861–9. https://doi.org/10.1111/j.1365-2664.2009.01660.x.

Hesselbarth, M. H., M. Sciaini, K. A. With, K. Wiegand, and J. Nowosad. 2019. "*landscapemetrics*: An Open-Source R Tool to Calculate Landscape Metrics." *Ecography* 42(10): 1648–57. https://doi.org/10.1111/ecog.04617.

Hoogland, J., ed. 2013. *Conservation of the Black-Tailed Prairie Dog: Saving North America's Western Grasslands*. Washington, DC: Island Press.

Hothorn, T., K. Hornik, and A. Zeileis. 2006. "Unbiased Recursive Partitioning: A Conditional Inference Framework." *Journal of*

Computational and Graphical Statistics 15(3): 651–74. https://doi.org/10.1198/106186006X133933.

Hoyt, J. R., A. M. Kilpatrick, and K. E. Langwig. 2021. "Ecology and Impacts of White-Nose Syndrome on Bats." *Nature Reviews Microbiology* 19(3): 196–210. https://doi.org/10.1038/s41579-020-00493-5.

Ims, R. A., N. G. Yoccoz, and S. T. Killengreen. 2011. "Determinants of Lemming Outbreaks." *Proceedings of the National Academy of Sciences* 108(5): 1970–4. https://doi.org/10.1073/pnas.1012714108.

Johnson, T. L., J. F. Cully, S. K. Collinge, C. Ray, C. M. Frey, and B. K. Sandercock. 2011. "Spread of Plague among Black-Tailed Prairie Dogs Is Associated with Colony Spatial Characteristics." *The Journal of Wildlife Management* 75(2): 357–68. https://doi.org/10.1002/jwmg.40.

Jones, K. E., J. Bielby, M. Cardillo, S. A. Fritz, J. O'Dell, C. D. L. Orme, and K. Safi. 2009. "PanTHERIA: A Species-Level Database of Life History, Ecology, and Geography of Extant and Recently Extinct Mammals: Ecological Archives E090-184." *Ecology* 90(9): 2648–8. https://doi.org/10.1890/08-1494.1.

Jones, K. E., N. G. Patel, M. A. Levy, A. Storeygard, D. Balk, J. L. Gittleman, and P. Daszak. 2008. "Global Trends in Emerging Infectious Diseases." *Nature* 451(7181): 990–3. https://doi.org/10.1038/nature06536.

Keuler, K. M., G. M. Bron, R. Griebel, and K. L. Richgels. 2020. "An Invasive Disease, Sylvatic Plague, Increases Fragmentation of Black-Tailed Prairie Dog (*Cynomys ludovicianus*) Colonies." *PLoS One* 15(7): e0235907. https://doi.org/10.1371/journal.pone.0235907.

Khalil, H., G. Olsson, M. Magnusson, M. Evander, B. Hörnfeldt, and F. Ecke. 2017. "Spatial Prediction and Validation of Zoonotic Hazard through Micro-Habitat Properties: Where Does Puumala Hantavirus Hole–Up?" *BMC Infectious Diseases* 17(1): 1–13. https://doi.org/10.1186/s12879-017-2618-z.

Knowles, C. J., J. D. Proctor, and S. C. Forrest. 2002. "Black-Tailed Prairie Dog Abundance and Distribution in the Great Plains Based on Historic and Contemporary Information." *Great Plains Research* 12: 219–54. https://www.jstor.org/stable/23779568.

Koelle, K., X. Rodó, M. Pascual, M. Yunus, and G. Mostafa. 2005. "Refractory Periods and Climate Forcing in Cholera Dynamics." *Nature* 436(7051): 696–700. https://doi.org/10.1038/nature03820.

Kotliar, N. B., B. W. Baker, A. D. Whicker, and G. Plumb. 1999. "A Critical Review of Assumptions about the Prairie Dog as a Keystone Species." *Environmental Management* 24(2): 177–92. https://doi.org/10.1007/s002679900225.

Krebs, C. J., R. Boonstra, S. Boutin, and A. R. Sinclair. 2001. "What Drives the 10-Year Cycle of Snowshoe Hares?" *Bioscience* 51(1): 25–35. https://doi.org/10.1641/0006-3568(2001)051[0025:WDTYCO]2.0.CO;2.

Kubat, M., R. C. Holte, and S. Matwin. 1998. "Machine Learning for the Detection of Oil Spills in Satellite Radar Images." *Machine Learning* 30(2): 195–215. https://doi.org/10.1023/A:1007452223027.

Liaw, A., and M. Wiener. 2002. "Classification and Regression by randomForest." *R News* 2(3): 18–22.

Link, V. B. 1955. "A History of Plague in the United States." *Public Health Monographs* 70: 1–120.

Machado, G., C. Vilalta, M. Recamonde-Mendoza, C. Corzo, M. Torremorell, A. Perez, and K. VanderWaal. 2019. "Identifying Outbreaks of Porcine Epidemic Diarrhea Virus through Animal Movements and Spatial Neighborhoods." *Scientific Reports* 9(1): 1–12. https://doi.org/10.1038/s41598-018-36934-8.

Matchett, M. R., T. R. Stanley, M. F. Mccollister, D. A. Eads, J. T. Boulerice, and D. E. Biggins. 2021. "Oral Sylvatic Plague Vaccine Does Not Adequately Protect Prairie Dogs (*Cynomys* spp.) for Endangered Black-Footed Ferret (*Mustela nigripes*) Conservation." *Vector-Borne and Zoonotic Diseases* 21(12): 921–40. https://doi.org/10.1089/vbz.2021.0049.

McMichael, A. J. 2015. "Extreme Weather Events and Infectious Disease Outbreaks." *Virulence* 6(6): 543–7. https://doi.org/10.4161/21505594.2014.975022.

Metcalfe, P., K. Beven, J. Freer, M. P. Metcalfe, and I. deSolve. 2018. "*dynatopmodel*: Implementation of the Dynamic TOPMODEL Hydrological Model." R Package Version 1.2.1. https://CRAN.R-project.org/package=dynatopmodel.

Moro Rosso, L. H., A. F. de Borja Reis, A. A. Correndo, and I. A. Ciampitti. 2021a. "Retrieving POLARIS Data Using R-Software." Harvard Dataverse, V2. https://doi.org/10.7910/DVN/DCZ0N3.

Moro Rosso, L. H., A. F. de Borja Reis, A. A. Correndo, and I. A. Ciampitti. 2021b. "XPolaris: An R-Package to Retrieve United States Soil Data at 30-Meter Resolution." *BMC Research Notes* 14(1): 1–3. https://doi.org/10.1186/s13104-021-05729-y.

Murphy, M. A., J. S. Evans, and A. Storfer. 2010. "Quantifying *Bufo boreas* Connectivity in Yellowstone National Park with Landscape Genetics." *Ecology* 91(1): 252–61. https://doi.org/10.1890/08-0879.1.

Newell, D. A., R. L. Goldingay, and L. O. Brooks. 2013. "Population Recovery Following Decline in an Endangered Stream-Breeding Frog (*Mixophyes fleayi*) from Subtropical Australia." *PLoS One* 8(3): e58559. https://doi.org/10.1371/journal.pone.0058559.

Nowakowski, A. J., S. M. Whitfield, E. A. Eskew, M. E. Thompson, J. P. Rose, B. L. Caraballo, J. L. Kerby, et al. 2016. "Infection Risk Decreases with Increasing Mismatch in Host and Pathogen Environmental Tolerances." *Ecology Letters* 19(9): 1051–61. https://doi-org.libproxy.uwyo.edu/10.1111/ele.12641.

Olden, J. D., J. J. Lawler, and N. L. Poff. 2008. "Machine Learning Methods without Tears: A Primer for Ecologists." *The Quarterly Review of Biology* 83(2): 171–93. https://doi.org/10.1086/587826.

Pandit, P., and B. A. Han. 2020. "Rise of Machines in Disease Ecology." *Bulletin of the Ecological Society of America* 101(1): 1–4. https://doi.org/10.1002/bes2.1625.

Pascual, M., B. Cazelles, M. J. Bouma, L. F. Chaves, and K. Koelle. 2008. "Shifting Patterns: Malaria Dynamics and Rainfall Variability in an African Highland." *Proceedings of the Royal Society B: Biological Sciences* 275(1631): 123–32. https://doi.org/10.1098/rspb.2007.1068.

Peters, D. P., N. D. Burruss, L. L. Rodriguez, D. S. McVey, E. H. Elias, A. M. Pelzel-McCluskey, J. Derner, et al. 2018. "An Integrated View of Complex Landscapes: A Big Data-Model Integration Approach to Transdisciplinary Science." *Bioscience* 68(9): 653–69. https://doi.org/10.1093/biosci/biy069.

Peters, D. P., D. S. McVey, E. H. Elias, A. M. Pelzel-McCluskey, J. D. Derner, N. D. Burruss, and T. S. Schrader. 2020. "Big

Data–Model Integration and AI for Vector-Borne Disease Prediction." *Ecosphere* 11(6): e03157. https://doi.org/10.1002/ecs2.3157.

Plowright, R. K., D. J. Becker, D. E. Crowley, A. D. Washburne, T. Huang, P. O. Nameer, and E. S. Gurley. 2019. "Prioritizing Surveillance of Nipah Virus in India." *PLoS Neglected Tropical Diseases* 13(6): e0007393. https://doi.org/10.1371/journal.pntd.0007393.

Plowright, R. K., K. Manlove, E. F. Cassirer, P. C. Cross, T. E. Besser, and P. J. Hudson. 2013. "Use of Exposure History to Identify Patterns of Immunity to Pneumonia in Bighorn Sheep (*Ovis canadensis*)." *PLoS One* 8(4): e61919. https://doi.org/10.1371/journal.pone.0061919.

Poché, D. M., D. Hartman, L. Polyakova, and R. M. Poché. 2017. "Efficacy of a Fipronil Bait in Reducing the Number of Fleas (*Oropsylla* spp.) Infesting Wild Black-Tailed Prairie Dogs." *Journal of Vector Ecology* 42(1): 171–7. https://doi.org/10.1111/jvec.12252.

R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. www.R-project.org.

Rohr, J. R., A. P. Dobson, P. T. Johnson, A. M. Kilpatrick, S. H. Paull, T. R. Raffel, D. Ruiz-Moreno, and M. B. Thomas. 2011. "Frontiers in Climate Change–Disease Research." *Trends in Ecology & Evolution* 26(6): 270–7. https://doi.org/10.1016/j.tree.2011.03.002.

Romero, M. P., Y. M. Chang, L. A. Brunton, J. Parry, A. Prosser, P. Upton, and J. A. Drewe. 2022. "Machine Learning Classification Methods Informing the Management of Inconclusive Reactors at Bovine Tuberculosis Surveillance Tests in England." *Preventive Veterinary Medicine* 199: 105565. https://doi.org/10.1016/j.prevetmed.2021.105565.

Roozbeh, M., and M. Najarian. 2018. "Efficiency of the QR Class Estimator in Semiparametric Regression Models to Combat Multicollinearity." *Journal of Statistical Computation and Simulation* 88(9): 1804–25. https://doi.org/10.1080/00949655.2018.1448088.

Sadoti, G., S. McAfee, E. Nicklen, P. Sousanes, and C. Roland. 2020. "Evaluating Multiple Historical Climate Products in Ecological Models under Current and Projected Temperatures." Dryad. Dataset. https://doi.org/10.5061/dryad.fttdz08qt.

Salkeld, D. J., M. Salathé, P. Stapp, and J. H. Jones. 2010. "Plague Outbreaks in Prairie Dog Populations Explained by Percolation Thresholds of Alternate Host Abundance." *Proceedings of the National Academy of Sciences* 107(32): 14247–50. https://doi.org/10.1073/pnas.1002826107.

Savage, L. T., R. M. Reich, L. M. Hartley, P. Stapp, and M. F. Antolin. 2011. "Climate, Soils, and Connectivity Predict Plague Epizootics in Black-Tailed Prairie Dogs (*Cynomys ludovicianus*)." *Ecological Applications* 21(8): 2933–43. https://doi.org/10.1890/10-1946.1.

Scheele, B. C., F. Pasmans, L. F. Skerratt, L. Berger, A. N. Martel, W. Beukema, and A. A. Acevedo. 2019. "Amphibian Fungal Panzootic Causes Catastrophic and Ongoing Loss of Biodiversity." *Science* 363(6434): 1459–63. https://doi.org/10.1126/science.aav0379.

Seery, D. B., and D. J. Matiatos. 2000. "Response of Wintering Buteos to Plague Epizootics in Prairie Dogs." *Western North American Naturalist* 60: 420–5. https://www.jstor.org/stable/41717059.

Shoemaker, K. 2023. "kevintshoemaker/Random-Forest-Functions: Random Forest Helper Functions (random_forest_functions)." Zenodo. https://doi.org/10.5281/zenodo.7523381.

Sierra–Corona, R., A. Davidson, E. L. Fredrickson, H. Luna-Soria, H. Suzan-Azpiri, E. Ponce-Guevara, and G. Ceballos. 2015. "Black-Tailed Prairie Dogs, Cattle, and the Conservation of North America's Arid Grasslands." *PLoS One* 10(3): e0118602. https://doi.org/10.1371/journal.pone.0118602.

Smith, M. J., A. White, J. A. Sherratt, S. Telfer, M. Begon, and X. Lambin. 2008. "Disease Effects on Reproduction Can Cause Population Cycles in Seasonal Environments." *Journal of Animal Ecology* 77(2): 378–89. https://doi.org/10.1111/j.1365-2656.2007.01328.x.

Snäll, T., R. B. O'Hara, C. Ray, and S. K. Collinge. 2008. "Climate-Driven Spatial Dynamics of Plague among Prairie Dog Colonies." *The American Naturalist* 171(2): 238–48. https://doi.org/10.1086/525051.

Stapp, P., M. F. Antolin, and M. Ball. 2004. "Patterns of Extinction in Prairie Dog Metapopulations: Plague Outbreaks Follow El Niño Events." *Frontiers in Ecology and the Environment* 2(5): 235–40. https://doi.org/10.1890/1540-9295(2004)002[0235:POEIPD]2.0.CO;2.

Stenseth, N. C., N. I. Samia, H. Viljugrein, K. L. Kausrud, M. Begon, S. Davis, and H. Leirs. 2006. "Plague Dynamics Are Driven by Climate Variation." *Proceedings of the National Academy of Sciences* 103(35): 13110–5. https://doi.org/10.1073/pnas.0602447103.

Stephens, T., S. C. Wilson, F. Cassidy, D. Bender, D. Gummer, D. H. Smith, N. Lloyd, J. M. McPherson, and A. Moehrenschlager. 2018. "Climate Change Impacts on the Conservation Outlook of Populations on the Poleward Periphery of Species Ranges: A Case Study of Canadian Black-Tailed Prairie Dogs (*Cynomys ludovicianus*)." *Global Change Biology* 24(2): 836–47. https://doi.org/10.1111/gcb.13922.

Strobl, C., A. L. Boulesteix, T. Kneib, T. Augustin, and A. Zeileis. 2008. "Conditional Variable Importance for Random Forests." *BMC Bioinformatics* 9(1): 1–11. https://doi.org/10.1186/1471-2105-9-307.

Strobl, C., A. L. Boulesteix, A. Zeileis, and T. Hothorn. 2007. "Bias in Random Forest Variable Importance Measures: Illustrations, Sources and a Solution." *BMC Bioinformatics* 8(1): 1–21. https://doi.org/10.1186/1471-2105-8-25.

Thiagarajan, B., Y. Bai, K. L. Gage, and J. F. Cully, Jr. 2008. "Prevalence of *Yersinia pestis* in Rodents and Fleas Associated with Black-Tailed Prairie Dogs (*Cynomys ludovicianus*) at Thunder Basin National Grassland, Wyoming." *Journal of Wildlife Diseases* 44(3): 731–6. https://doi.org/10.7589/0090-3558-44.3.731.

Thornton, P. E., S. W. Running, and M. A. White. 1997. "Generating Surfaces of Daily Meteorological Variables over Large Regions of Complex Terrain." *Journal of Hydrology* 190(3–4): 214–51. https://doi.org/10.1016/S0022-1694(96)03128-9.

Tripp, D. W., K. L. Gage, J. A. Montenieri, and M. F. Antolin. 2009. "Flea Abundance on Black-Tailed Prairie Dogs (*Cynomys ludovicianus*) Increases during Plague Epizootics."

Vector-Borne and Zoonotic Diseases 9(3): 313–21. https://doi.org/10.1089/vbz.2008.0194.

Tripp, D. W., T. E. Rocke, J. P. Runge, R. C. Abbott, and M. W. Miller. 2017. "Burrow Dusting or Oral Vaccination Prevents Plague-Associated Prairie Dog Colony Collapse." *EcoHealth* 14(3): 451–62. https://doi.org/10.1007/s10393-017-1236-y.

U.S. Geological Survey. 2022. "3D Elevation Program 1-Meter Resolution Digital Elevation Model (published 20200606)," May 17, 2022. https://www.usgs.gov/the-national-map-data-delivery.

Walker, K. E. 2021. "*tigris*: An R Package to Access and Work with Geographic Data from the US Census Bureau." https://doi.org/10.32614/RJ-2016-043.

Walsh, D. P., T. F. Ma, H. S. Ip, and J. Zhu. 2019. "Artificial Intelligence and Avian Influenza: Using Machine Learning to Enhance Active Surveillance for Avian Influenza Viruses." *Transboundary and Emerging Diseases* 66(6): 2537–45. https://doi.org/10.1111/tbed.13318.

Williams, S. K., A. M. Schotthoefer, J. A. Montenieri, J. L. Holmes, S. M. Vetter, K. L. Gage, and S. W. Bearden. 2013. "Effects of Low-Temperature Flea Maintenance on the Transmission of *Yersinia pestis* by *Oropsylla montana*." *Vector-Borne and Zoonotic Diseases* 13(7): 468–78. https://doi.org/10.1089/vbz.2012.1017.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Barrile, Gabriel M., David J. Augustine, Lauren M. Porensky, Courtney J. Duchardt, Kevin T. Shoemaker, Cynthia R. Hartway, Justin D. Derner, Elizabeth A. Hunter, and Ana D. Davidson. 2023. "A Big Data–Model Integration Approach for Predicting Epizootics and Population Recovery in a Keystone Species." *Ecological Applications* e2827. https://doi.org/10.1002/eap.2827