

RESEARCH

Open Access



Forecasting United States heartworm *Dirofilaria immitis* prevalence in dogs

Dwight D. Bowman¹, Yan Liu², Christopher S. McMahan², Shila K. Nordone³, Michael J. Yabsley⁴ and Robert B. Lund^{2*}

Abstract

Background: This paper forecasts next year's canine heartworm prevalence in the United States from 16 climate, geographic and societal factors. The forecast's construction and an assessment of its performance are described.

Methods: The forecast is based on a spatial-temporal conditional autoregressive model fitted to over 31 million antigen heartworm tests conducted in the 48 contiguous United States during 2011–2015. The forecast uses county-level data on 16 predictive factors, including temperature, precipitation, median household income, local forest and surface water coverage, and presence/absence of eight mosquito species. Non-static factors are extrapolated into the forthcoming year with various statistical methods. The fitted model and factor extrapolations are used to estimate next year's regional prevalence.

Results: The correlation between the observed and model-estimated county-by-county heartworm prevalence for the 5-year period 2011–2015 is 0.727, demonstrating reasonable model accuracy. The correlation between 2015 observed and forecasted county-by-county heartworm prevalence is 0.940, demonstrating significant skill and showing that heartworm prevalence can be forecasted reasonably accurately.

Conclusions: The forecast presented herein can a priori alert veterinarians to areas expected to see higher than normal heartworm activity. The proposed methods may prove useful for forecasting other diseases.

Keywords: Autoregression, CAR Model, Head-banging, Heartworm, Kriging, Prevalence, Spatio-temporal correlation

Background

Heartworm disease, caused by the mosquito-borne filarial nematode *Dirofilaria immitis*, is arguably the most medically important parasitic infection of domestic dogs in the United States (US), affecting at least 115,000 dogs in 2015. Beyond the US, heartworm disease is a global veterinary healthcare problem, with *D. immitis* affecting dogs in many parts of South America, Europe, Asia, and Australia [1, 2]. Infection is associated with life-threatening complications and significant financial burden, costing millions in veterinary care annually for disease treatment [3–7]. Although less common and less studied, heartworm disease is also a health concern for other mammals such as domestic cats, domestic ferrets, and some wildlife species [8]. Clinical signs of heartworm disease in domestic dogs

include exercise intolerance, coughing, dyspnea, cachexia, anorexia, epistaxis and ascites. Dogs with a high burden of adult heartworms can suffer from pulmonary arterial occlusion and inflammation, leading to pulmonary hypertension and potentially right heart failure. Cats and ferrets may experience similar signs but acute death may occur, even with very low worm burdens. Humans can also be infected with *D. immitis*, but infections are rare, with fewer than 100 cases reported in the US over the last 60 years [9]. Human infection is most commonly asymptomatic, with people considered dead-end hosts for the parasite. While rare, human *D. immitis* infection is highly problematic in that it most often manifests as “coin lesions” in the lungs that may be mistaken for a neoplasm on chest radiographs; surgical excision is necessary to differentiate the two entities [10].

Heartworm disease in dogs is most commonly diagnosed through the detection of circulating *D. immitis* antigen in the blood [3, 11]. The prevalence of heartworm

* Correspondence: lund@clmson.edu

²Department of Mathematical Sciences, Clemson University, Clemson, SC, USA

Full list of author information is available at the end of the article



infection in the US varies considerably by geographical region. Two nationwide surveillance studies of *D. immitis* infection seroprevalence (henceforth prevalence) in domestic dogs found the highest prevalence in the Southeast and the lowest in the Northeast [11]. For unknown reasons, a noted decrease in the prevalence of *D. immitis* occurred between the 2001–2007 and 2010–2012 in these studies. Importantly, regardless of time period and even within areas where heartworm infection is considered common, there can be considerable local variation, with prevalence reaching as high as 13 % [12, 13].

Numerous factors are purported to be associated with regional and local variations in *D. immitis* prevalence in domestic dogs. Highly effective commercially available anthelmintics (e.g. macrocyclic lactones (ML) [3], including the avermectins (ivermectin, selamectin) and the milbemycins (moxidectin, milbemycin oxime) can be administered monthly to prevent the development of immature stages into adult worms. Year-round preventive use is recommended throughout the US, yet the majority of dogs only receive seasonal treatment [14]. Even within highly endemic regions, anthelmintic use varies based on client compliance, knowledge, or dog owner's demographics. In addition, resistance of *D. immitis* to ML has been recently documented and is a growing concern in the Gulf States, but the current extent of resistant phenotypes remains unknown [7, 15]. *Dirofilaria immitis* can be transmitted by over 70 species of mosquitoes, although certain species (e.g. *Aedes trivittatus*, *Aedes sierrensis* and *Culex quinquefasciatus*) are considered more important vectors [16]. Because the density of mosquitoes and community composition of competent vector species is influenced greatly by habitat use and climate, these factors should be considered when investigating factors influencing heartworm disease. In support of this, a previous study found that temperature, median household income, population density, precipitation, elevation, relative humidity, forestation coverage, and surface water coverage all significantly influence *D. immitis* prevalence in dogs [16].

Clearly, it would be advantageous to accurately forecast *D. immitis* prevalence on a local scale, providing an a priori alert to veterinarians in problem areas where immediate remediation measures could be taken. Annual forecasts of emergent infection will also inform veterinary and public health officials to shifting areas of infection, particularly in temperate regions of the US where *D. immitis* is generally absent, rare, or prevalence is highly influenced by annual variation in biotic or abiotic factors.

Methods

Data structure

The data studied here contain 31,345,244 heartworm antigen test results from dogs in the conterminous

United States from 2011 to 2015, and various climate, geographic and socio-economic factors purported to influence heartworm prevalence. The raw tests were obtained from the Antech and IDEXX laboratories [17, 18]. Over all 5 years in the study, 384,905 of the tests were positive (1.23 %). The test data contain the county/parish of the testing clinic and the month when the tests were conducted; however, no measure of uncertainty is given with the individual test results.

The test data were aggregated into the number of positive and negative tests for each year in each conterminous United States county/parish. Table 1 lists 16 explanatory factors that are purportedly related to dog heartworm prevalence, as well as their time period of record and geographic scale of collection. These 16 factors include the climatic variables of annual temperature, precipitation and relative humidity, the geographic variables of county elevation, forestation coverage and surface water coverage, the socio-economic variables of county population density and median household income, and the presence of eight mosquito vectors. For more details on these factors, see [19].

Figure 1 displays county-level raw heartworm prevalences obtained by dividing the number of positive tests by the number of tests over all 5 years in the study. The raw prevalences exhibit a large degree of spatial correlation in that neighboring counties tend to report similar prevalences. Significant temporal dependence is also present in the data: the current prevalence is similar to past prevalence. Therefore, this data set requires a statistical model with both spatial and temporal dependence.

Figure 2 provides a spatially smoothed prevalence map, using a head-banging smoothing procedure, based on the Fig. 1 prevalences. In the head-banging smoothing procedure, 45 triples were employed. The smoothing was also weighted proportionally to the number of tests in each county over the 5-year period. This prevents the map from signaling a high/low prevalence that is more likely attributed to a small sample (one positive out of three tests has the same prevalence as one hundred positive in 300 tests, though the latter is more indicative of infection risk). Details on head-banging smoothing and its uses in disease mapping are contained in [16]. Figure 2 serves as a contemporary depiction of the "baseline" heartworm risk for dogs in the United States.

Statistical modeling

The model and methods used to statistically analyze the heartworm tests are now described. The goal here is to assess the significance of the 16 factors and accurately estimate regional heartworm prevalence.

Let $Y_s(t)$ and $n_s(t)$ denote the number of positive and total tests conducted in county s during year t , respectively, for

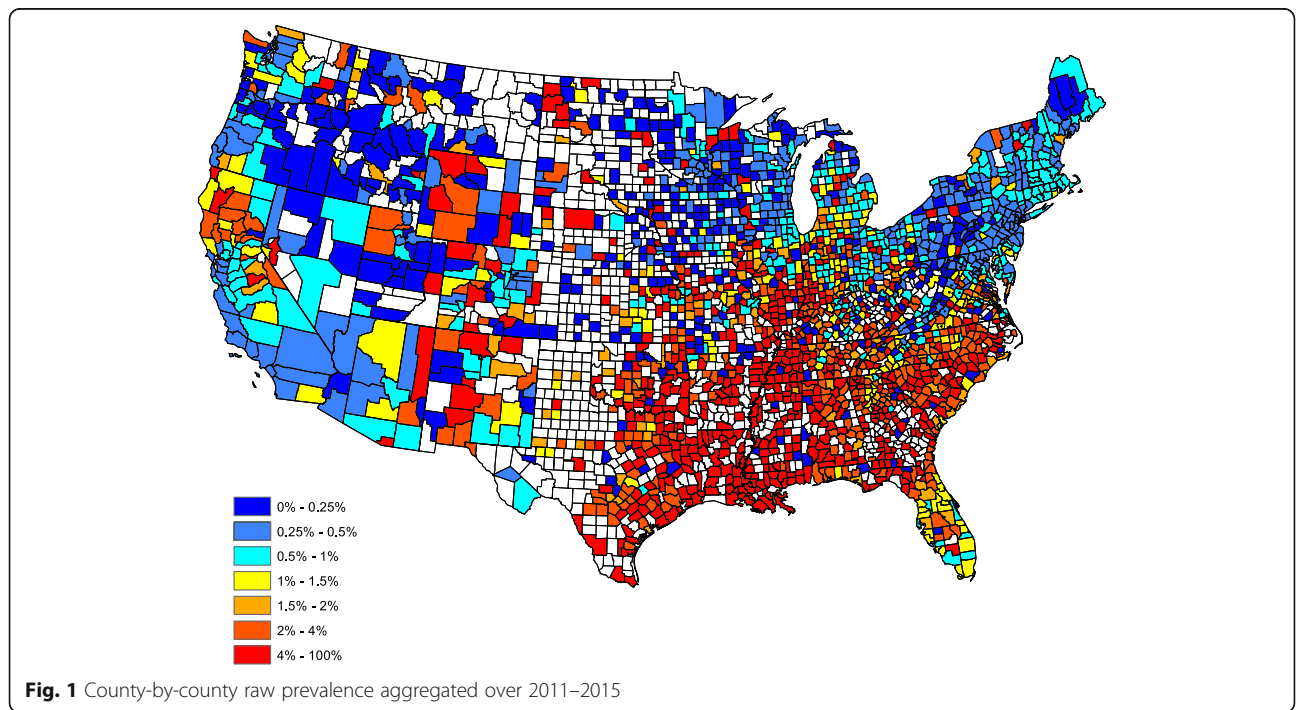
Table 1 Factors purported to influence heartworm prevalence

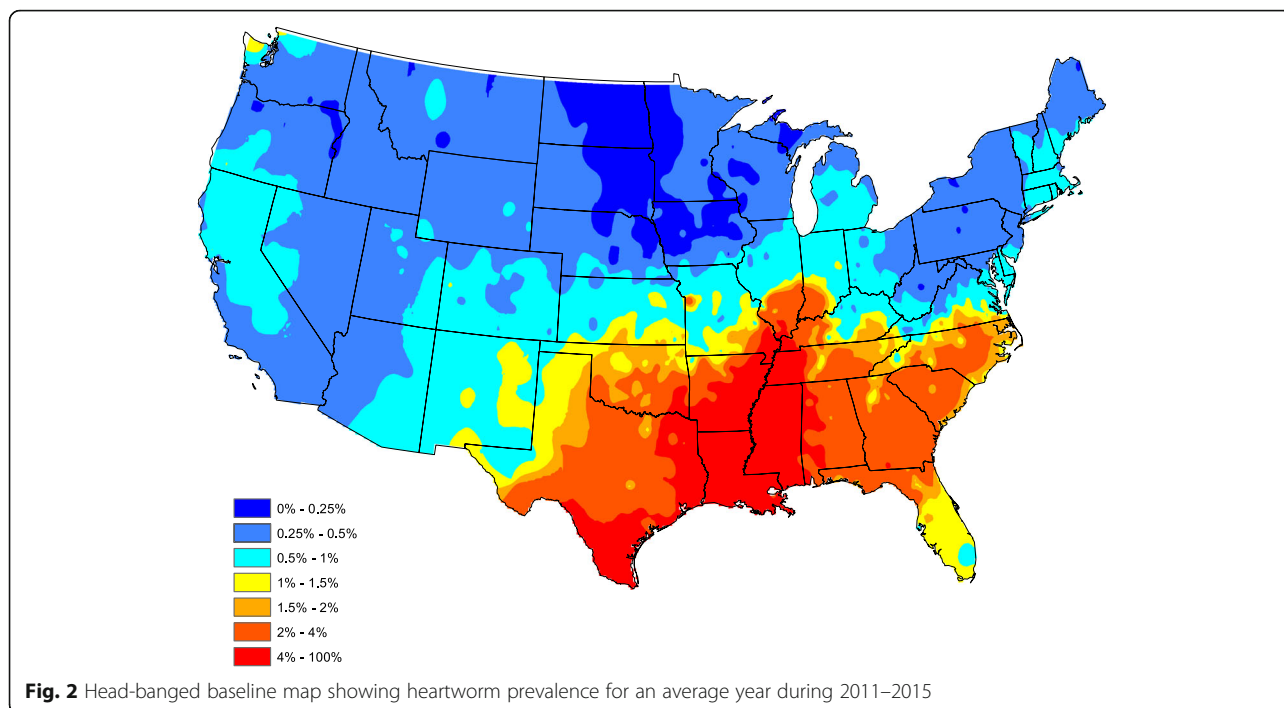
	Factor	Data period	Scale	Notation	Numerical scale of data
Climate factors	Annual temperature	1895–2015	Climate Division	$X_{s,1}(t)$	Continuous
	Annual precipitation	1895–2015	Climate Division	$X_{s,2}(t)$	
	Annual relative humidity	2006–2015	Climate Division	$X_{s,3}(t)$	
Geographic factors	Elevation	2012	County	$X_{s,4}(t)$	Continuous
	Percentage forest coverage	2012	County	$X_{s,5}(t)$	
	Percentage surface water coverage	2010	County	$X_{s,6}(t)$	
Societal factors	Population density	2011–2014	County	$X_{s,7}(t)$	Continuous
	Median household income	1997–2014	County	$X_{s,8}(t)$	
Mosquito species	<i>Aedes aegypti</i>	2008	County	$X_{s,9}(t)$	$X_{s,k} = 1$ if present, and $X_{s,k} = 0$ otherwise
	<i>Aedes albopictus</i>	2012	County	$X_{s,10}(t)$	
	<i>Aedes canadensis</i>	2004	County	$X_{s,11}(t)$	
	<i>Aedes sierrensis</i>	2004	County	$X_{s,12}(t)$	
	<i>Aedes trivittatus</i>	2004	County	$X_{s,13}(t)$	
	<i>Anopheles punctipennis</i>	2004	County	$X_{s,14}(t)$	
	<i>Anopheles quadrimaculatus</i>	2004	County	$X_{s,15}(t)$	
<i>Culex quinquefasciatus</i>	2004	County	$X_{s,16}(t)$		

For further discussion, including the source of each factor, see [16]

counties $s \in \{1, \dots, S\}$ and years $t \in \{1, \dots, T\}$. Methodologies for modeling spatial and temporal dependence have received much recent attention in the statistics literature [20–27]. Among many choices, Bayesian hierarchical models have been prominent due to their flexibility. In a Bayesian hierarchical model, spatial and temporal dependence is modelled in a hierarchy

via a series of random effect terms with prescribed structures; see [20, 21] for a modern review of spatio-temporal models. Typically, when modeling the spatial or spatial-temporal dependent count data via parametric models, a Poisson distribution is preferred [21–24]. The following hierarchical regression model is used here:





$$Y_s(t) | n_s(t), p_s(t) \sim \text{Poisson}\{n_s(t)p_s(t)\}, \tag{1}$$

$$\log\{p_s(t)\} = \beta_0 + \sum_{k=1}^{16} \beta_k X_{s,k}(t) + \xi_s(t), \tag{2}$$

where $\log(\cdot)$ denotes natural logarithm, $\mathbf{X}_s(t) = (X_{s,1}(t), \dots, X_{s,16}(t))'$ is a vector of covariate information for county s at time t ('denotes matrix transpose), $\boldsymbol{\beta} = (\beta_0, \dots, \beta_p)'$ is a vector of regression coefficients, $p_s(t)$ denotes the heartworm prevalence of county s at time t , the symbol \sim means has the distributional type, and $|$ indicates given quantities. Equation (1) indicates that $Y_s(t)$ has a Poisson distribution with mean $n_s(t)p_s(t)$. In addition, it is assumed that the positive test counts (i.e. $Y_s(t)$) are conditionally independent of each other given the number of tests, factor information, and random effects. This does not imply that $Y_s(t)$ is independent across varying space s or time t .

To relate prevalence to the factors and build spatial and temporal dependence, the model in (2) is proposed, as is common in Poisson regressions [21–24]. There are many ways to induce spatial and temporal dependence from the random effects $\{\xi_s(t)\}$. One natural and popular choice is the conditional autoregressive (CAR) structure

$$\begin{aligned} \xi_1 &= \phi_1; \\ \xi_t | \xi_{t-1}, \varphi &= \varphi \xi_{t-1} + \phi_t, \text{ for } t = 2, \dots, T; \end{aligned} \tag{3}$$

$$\phi_t \sim \text{CAR}(\tau^2; \rho), \text{ for } t = 1, \dots, T, \tag{4}$$

where $\boldsymbol{\xi}_t = (\xi_1(t), \dots, \xi_S(t))'$ and $\boldsymbol{\phi}_t = (\phi_1(t), \dots, \phi_S(t))'$ are random vectors. Equation (4) indicates that the spatial effects (i.e. $\boldsymbol{\phi}_t$, for $t = 1, \dots, T$) are independent and identically distributed random vectors that follow a conditional autoregressive (CAR) model [25], which is a popular choice for modeling spatial dependence [26].

More specifically, let $\boldsymbol{\phi} = (\phi_1, \dots, \phi_S)'$ denote a random vector which follows a CAR model. There are several varieties of CAR models. Typically, the CAR model is specified via a series of univariate conditional distributions. Spatial dependence is induced through a neighboring system involving geographically adjacent counties. Our version of the CAR model, taken from [26], uses

$$\begin{aligned} \phi_k | \boldsymbol{\phi}_{-k}, \tau^2, \rho, \mathbf{W} &\sim N\left(\rho \frac{\sum_{i=1}^S w_{k,i} \phi_i}{\sum_{i=1}^S w_{k,i}}, \frac{\tau^2}{\sum_{i=1}^S w_{k,i}}\right), \\ \text{for } k &= 1, \dots, S \end{aligned} \tag{5}$$

Here, $\boldsymbol{\phi}_{-k} = (\phi_1, \dots, \phi_{k-1}, \phi_{k+1}, \dots, \phi_S)'$ is a vector that contains county effects for all counties except the k th one and $N(\mu, \sigma^2)$ denotes a normally distributed quantity with mean μ and variance σ^2 . In addition, $\mathbf{W} = \{w_{k,i}\}$ is an $S \times S$ dimensional matrix that describes the neighborhood structure of all counties. Specifically, the entries of \mathbf{W} are either zero or unity; $w_{k,i} = 1$ if and only if the i th and k th counties share some common border.

The parameter τ^2 in (5) is a scaling variance parameter. In fact, it can be seen that the conditional variance of ϕ_k given its neighbor’s random effects is inversely proportional to the number of counties bordering county k . Hence, counties with more neighboring counties tend to have a smaller variance, which is reasonable since data from the bordering counties helps predict the prevalence in the said county.

In Eq. (5), $\rho \in [0, 1]$ is an autocorrelation parameter that governs correlation between bordering counties. Notice that the conditional expectation of ϕ_k is the weighted arithmetic average of the neighboring random effects, multiplied by ρ . When $\rho = 0$, the conditional expectations of ϕ_s are zero and all random effects are independent of each other; antipodally, ρ close to unity indicates strong spatial dependence between bordering counties.

In (3), time-dependence is modeled through a temporal autoregressive model of order one (AR(1)), which is a time series staple [28]. Here, it describes prevalence for a fixed county across different years. The parameter φ is the temporal correlation between consecutive years and lies within $(-1, 1)$. This ensures a causal and stationarity solution to the time series model [28], which is needed in estimation.

From (5), it is possible to explicitly identify the joint distribution of ϕ , which is multivariate normal:

$$\phi \sim N(\mathbf{0}, \mathbf{\Gamma}), \quad \mathbf{\Gamma} = \tau^2(\mathbf{D} - \rho\mathbf{W})^{-1},$$

where \mathbf{W} is the previously mentioned neighborhood matrix and $\mathbf{D} = \{d_{i,i}\}$ is an $S \times S$ diagonal matrix whose i th diagonal element is the number of neighboring counties for county i .

Bayesian techniques are used to estimate the model parameters, which are β, φ, ρ , and τ^2 . Thus, to complete the Bayesian hierarchical model, the following prior distributions for these parameters are introduced:

$$\begin{aligned} \beta_k &\sim N(0, 1000), \text{ for } k = 0, \dots, 16; \\ \varphi &\sim \text{Uniform}(-1, 1); \\ \rho &\sim \text{Uniform}(0, 1); \\ \tau^{-2} &\sim \text{Gamma}(0.5, 0.05). \end{aligned}$$

Prior distributions for φ and ρ are taken as uninformative in that all admissible possibilities are equally likely. Priors for the regression coefficients $\beta_0, \dots, \beta_{16}$ are taken as diffuse so that inferences for these parameters are based primarily on the data. The prior for τ^{-2} is chosen as a conjugate prior (the posterior and prior distributions are from the same distributional family) for ease of computation. The random effects and model parameters are estimated based on posterior samples from a Markov chain Monte Carlo (MCMC) simulation. The MCMC simulation for our model uses a combination of Gibbs

and Metropolis-Hastings steps. In the implementation of the algorithm, the test data for non-reporting counties was viewed as being latent and was subsequently sampled along with the model parameters. To run our MCMC simulation and assess significance of model parameters, a program was developed and implemented in **R** and **C++**.

Results

Model assessment

The spatio-temporal Poisson regression model in (1) has 16 explanatory factors, all of which may not have predictive power. To assess this issue, a full model with all 16 factors was first fitted. Credible intervals, Bayesian analogs to confidence intervals in frequentist statistics, were then created for the parameters of interest. Table 1 summarizes our full model findings, showing 16 regression coefficients estimates (posterior median) and their 95 % highest posterior density (HPD) intervals; for further details about credible and HPD intervals, see [29, 30].

Table 2 implies that not all factors are significant, e.g. 95 % HPD intervals of annual precipitation, elevation, percentage surface water coverage, and all mosquito presence factors except *A. albopictus* contain zeroes. To develop a parsimonious model with only significant factors, all explanatory variables whose 95 % HPD intervals contain zeroes were removed and the model was refitted. This leaves a “reduced model” with the six explanatory factors: annual temperature, annual relative humidity, percentage forest coverage, population density, median household income and *A. albopictus* absence/presence. Parameter estimates (posterior median) and 95 % HPD intervals for the regression parameters for the reduced model are shown in Table 3. The estimates (posterior median) of the other model parameters are $\varphi = 0.914$, $\rho = 0.998$, and $\tau^2 = 0.802$.

Most of the significant factors have an intuitive interpretation. For example, the positive regression coefficient for the temperature and relative humidity factors implies that heartworm is more prevalent in warmer and humid locations. On the other hand, as is seen by negative regression coefficients, heartworm prevalence decreases with increasing population densities and median household incomes. Given the presence of relative humidity in the model, it is not overly surprising that precipitation drops out of the model fit. Finally, the negative regression coefficient on *A. albopictus* presence is not a contradiction: in the presence of all other factors (which include space and time prevalence histories), presence of this mosquito is associated with lessened heartworm prevalence. It is worthwhile to note that in a separate analysis (results not shown) a model with only

Table 2 Parameter estimates from the full model

Factor	Estimate	95 % HPD interval
Annual temperature	0.052	[0.038, 0.066]
Annual precipitation	0.008	[-0.031, 0.047]
Annual relative humidity	0.007	[0.003, 0.013]
Elevation	0.013	[-0.013, 0.039]
Percentage forest coverage	2.482	[1.664, 3.317]
Percentage surface water coverage	0.036	[-0.215, 0.277]
Population density	-5.086×10^{-5}	$[-6.744 \times 10^{-5}, -3.367 \times 10^{-5}]$
Median household income	-0.018	[-0.021, -0.016]
<i>Aedes aegypti</i>	-0.095	[-0.255, 0.059]
<i>Aedes albopictus</i>	-0.158	[-0.237, -0.071]
<i>Aedes canadensis</i>	0.185	[-0.039, 0.402]
<i>Aedes sierrensis</i>	-0.112	[-0.414, 0.204]
<i>Aedes trivittatus</i>	0.169	[-0.094, 0.414]
<i>Anopheles punctipennis</i>	-0.065	[-0.321, 0.182]
<i>Anopheles quadrimaculatus</i>	-0.076	[-0.246, 0.109]
<i>Culex quinquefasciatus</i>	0.099	[-0.099, 0.295]

A. albopictus was fitted, and the accompanying regression coefficient was non-negative.

To assess the overall performance of our reduced model, Fig. 3 graphically portrays our fitted model by plotting the average (over all 5 years) of model-estimated prevalence for each county after smoothing (standard Kriging with default parameters were used here). The model-estimated prevalence in Fig. 3 compares well to the head-banging smoothed baseline in Fig. 2. In fact, the correlation between the Figs 2 and 3 graphics is 0.727 (only counties reporting at least one test during the 5 year study period were used in this calculation). Clarifying further, our correlation between the two observation sets $\{A_s\}_{s=1}^S$ and $\{B_s\}_{s=1}^S$ is

Table 3 Parameter estimates from the reduced model

Parameter	Median	95 % HPD interval
Annual temperature	0.042	[0.027, 0.062]
Annual relative humidity	0.007	[0.002, 0.012]
Percentage forest coverage	2.599	[1.82, 3.473]
Population density	-5.177×10^{-5}	$[-7.074 \times 10^{-5}, -3.550 \times 10^{-5}]$
Median household income	-0.018	[-0.021, -0.016]
<i>Aedes albopictus</i>	-0.165	[-0.246, -0.081]

$$\text{Corr}(\{A_s\}, \{B_s\}) = \frac{\sum_{s=1}^S n_s (A_s - \bar{A})(B_s - \bar{B})}{\sqrt{\sum_{s=1}^S n_s (A_s - \bar{A})^2 \sum_{s=1}^S n_s (B_s - \bar{B})^2}}, \tag{6}$$

where

$$\bar{A} = \frac{\sum_{s=1}^S n_s A_s}{\sum_{s=1}^S n_s}, \quad \bar{B} = \frac{\sum_{s=1}^S n_s B_s}{\sum_{s=1}^S n_s}$$

are the sample-size weighted averages of $\{A_s\}_{s=1}^S$ and $\{B_s\}_{s=1}^S$, and n_s is the number of tests conducted in county s . Since the correlation here is between smoothed and model-estimated prevalence (these are non sample size dependent quantities), the weights were taken as $n_s \equiv 1$ (and not the county-by-county sample sizes). The 0.727 correlation achieved indicates that the regression model has explained most of the data structure.

The fitted model has a number of uses. In the next section, it is used to construct annual heartworm prevalence forecasts. The model could also be used to estimate how climate change could alter heartworm disease risk.

Forecasting

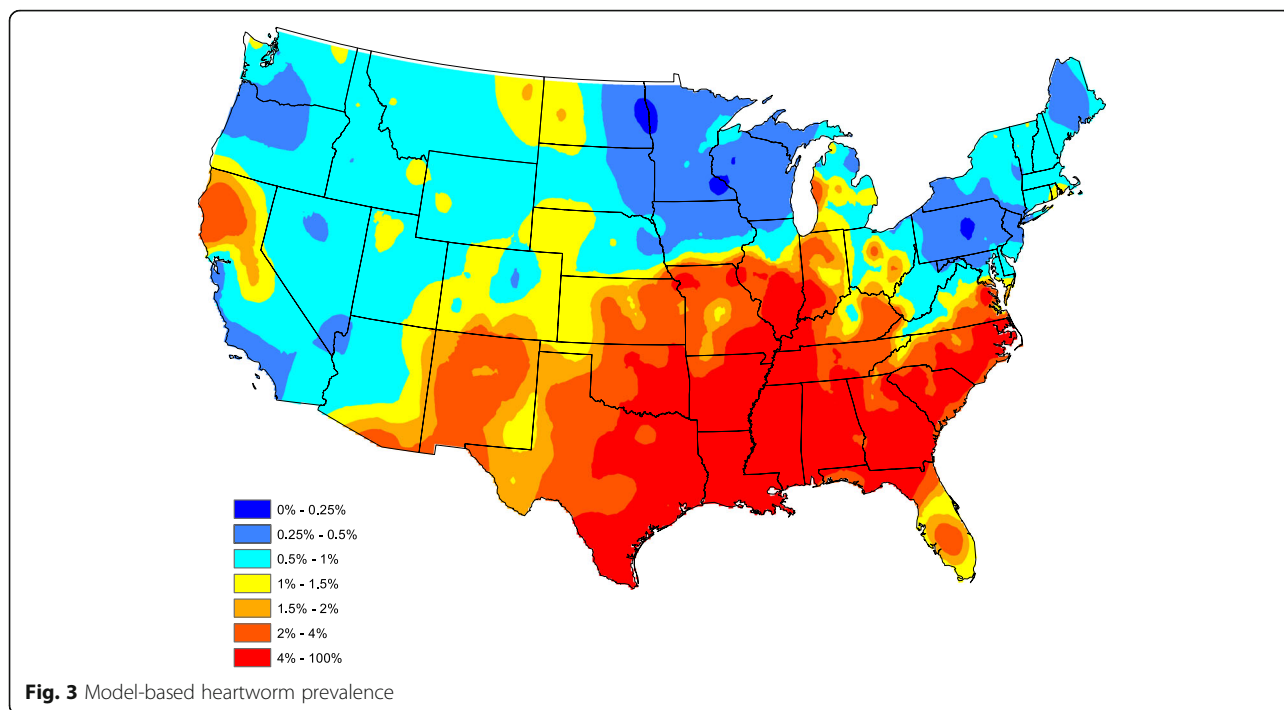
This section shows how to use our model to forecast next year’s regional heartworm prevalence. For this, all six significant explanatory factors and the spatial-temporal effects will need to be forecasted for the forthcoming year. To see how our forecast performs, the 2015 test and factor data was removed from the analysis, and the proposed six-factor model was refitted using data from 2011 to 2014 only. Our forecasts simply “plug in” 2015 forecasted factors for *A. albopictus*, annual temperature, annual relative humidity, percent forest coverage, population density, median household income, and a randomly generated random effect component into our model; for further details see [29, 30].

Two of the six factors are relatively stable over time: county forestation and the presence of *A. albopictus*. For these two factors, the most recent observations are used as 2015’s forecasted values.

To forecast annual temperature, historical temperature records were collected from 1895 to 2014 for each county and modeled as an autoregressive model of order one. The AR(1) model for an annual temperature series $\{F_t\}$ (previously denoted by $\{X_{s,1}(t)\}$ in Section 3 for county s) obeys the difference equation

$$F_t = \delta + \gamma F_{t-1} + \omega_t,$$

where $\{\omega_t\}$ is zero mean white noise; for further time series forecasting information, see [28]. The AR(1) model can be fitted to the temperature observations



using practically any statistical software package. Let $\hat{\delta}$ and $\hat{\gamma}$ denote estimates of δ and γ , respectively. A prediction of the annual temperature at year $t + 1$ from temperatures from year 1 to year t is

$$\hat{F}_{t+1} = \hat{\delta} + \hat{\gamma}F_t.$$

In our forecast, \hat{F}_{t+1} is used as next year’s forecasted temperature factor. Figures 4 and 5 compare forecasted and observed annual temperatures for 2015. The correlation between these two figures in (6) is $r = 0.996$, which is very good ($n_s \equiv 1$ here).

A simple linear regression was used to forecast next year’s relative humidity (previously denoted by $\{X_{s,3}(t)\}$) and median household income (previously denoted by $\{X_{s,8}(t)\}$) in each county. Historical relative humidities from 2006 to 2014 and median household incomes from 1997 to 2014 were used to fit a regression model of form

$$I_t = \alpha + \kappa t + \eta_t$$

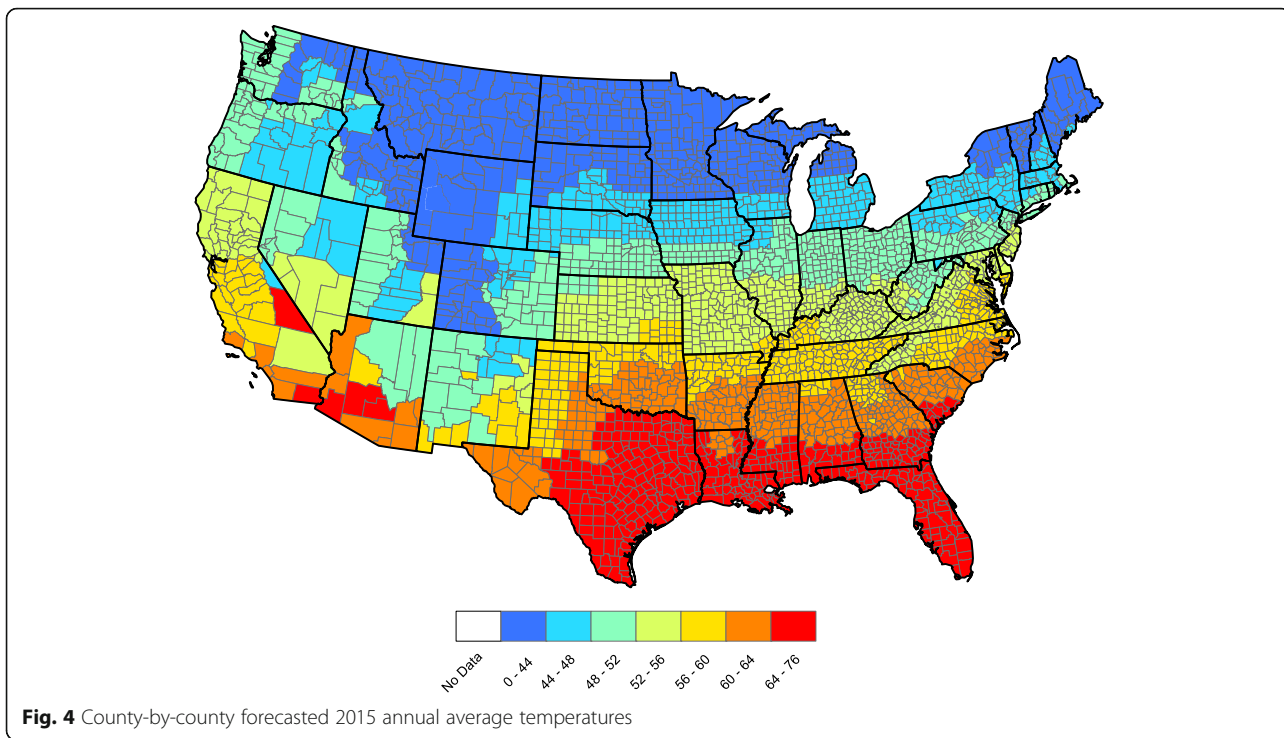
for each county. Here, $\{I_t\}$ denotes the relative humidity (previously denoted by $\{X_{s,3}(t)\}$) or median household income (previously denoted by $\{X_{s,8}(t)\}$), $\{\eta_t\}$ is zero-mean random noise. Least squares estimators of α and κ , denoted by $\hat{\alpha}$ and $\hat{\kappa}$, respectively, were computed from the data at each county. The forecasted value for year $t + 1$ is simply

$$\hat{I}_{t+1} = \hat{\alpha} + \hat{\kappa}(t + 1).$$

Forecasting the county population density for next year requires the county areas and their recent population counts. The US Census provides good county population estimates for 2010, but not in years since 2010. Estimated state populations were obtained for each state between 1969 and 2014. A simple linear regression was fitted to these data for each state and 2015 state populations were forecasted. This forecasted state population was then partitioned to the counties within the state at a proportion that agrees with 2010 Census proportions.

To forecast the next year’s spatial and temporal random effects, formula (3) is applied. Since the ϕ_t s are independent and identically distributed over various years, based on the values of τ^2 and ρ (available from the posterior samples), for the current time t , ϕ_{t+1} can be generated from the multivariate normal distribution $N(\mathbf{0}, \tau^2(\mathbf{D} - \rho\mathbf{W})^{-1})$. Then ξ_{t+1} is calculated by $\xi_{t+1} = \varphi\xi_t + \phi_{t+1}$. This process is repeated for each value of (ρ, τ^2, φ) , which are available from the posterior sample, thus yielding predictive posterior samples of the next year’s random effect ξ_{t+1} . See [29, 30] for additional detail on obtaining predictive posterior samples.

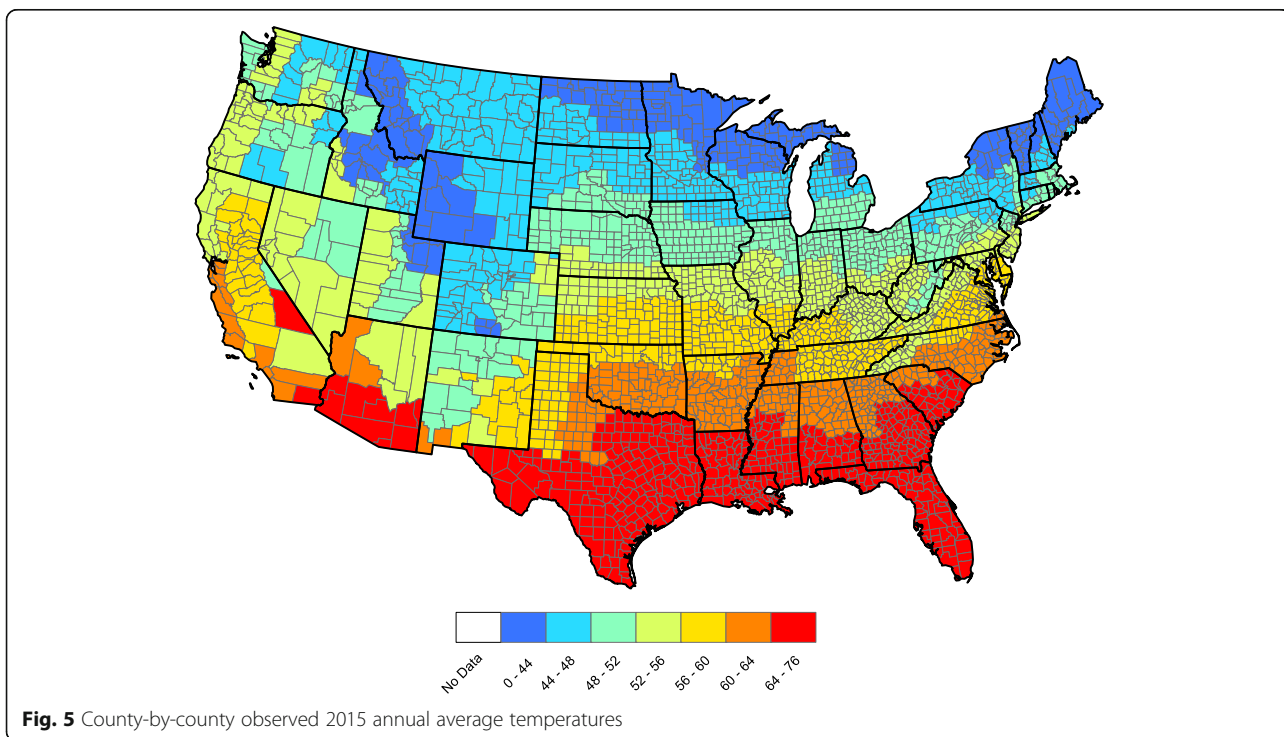
Figures 6 and 7 compare observed and forecasted heartworm prevalence during 2015. One can discern where heartworm is forecasted to be higher/lower than normal by comparing Figs. 2 and 7. The correlation



between the Figs. 6 and 7 prevalence, as measured in (6), is 0.940. Hence, the model is accurately forecasting in locations that report more tests. Finally, Fig. 8 presents our heartworm forecast for 2016. When 2016 concludes, we will be able to compare this forecast to 2016 test results.

Discussion

Generally, the management of emerging infectious diseases is approached reactively, with efforts focused on managing outbreaks after onset. The ability to reliably forecast transmission risk, particularly for diseases influenced by dynamic factors such as climate, could shift



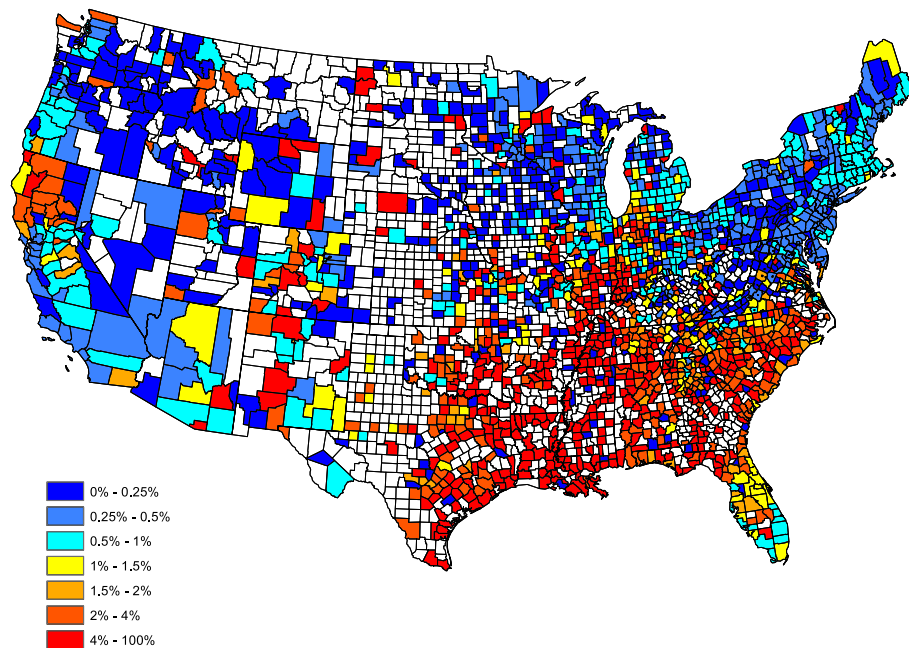


Fig. 6 Observed heartworm prevalence for 2015

our paradigm from reaction to prevention. This is particularly true for vector-borne diseases, as specific environmental needs for vector survival are well documented [31]. One approach to infectious disease modeling is to use these factors to predict transmission and model the data in both space and time. This has been used

successfully to estimate the incidence of malaria during eradication campaigns in Namibia and cutaneous leishmaniasis in high-risk areas of Columbia [32, 33].

Although preventable, heartworm disease is a relatively common and serious vector-borne disease of domestic dogs. Annual disease incidence, as reported by IDEXX

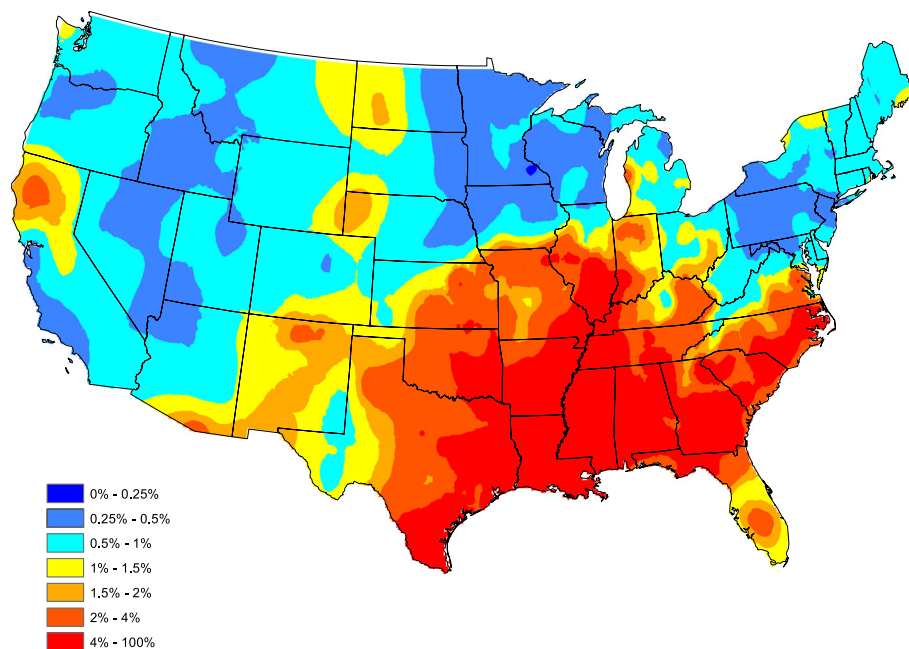
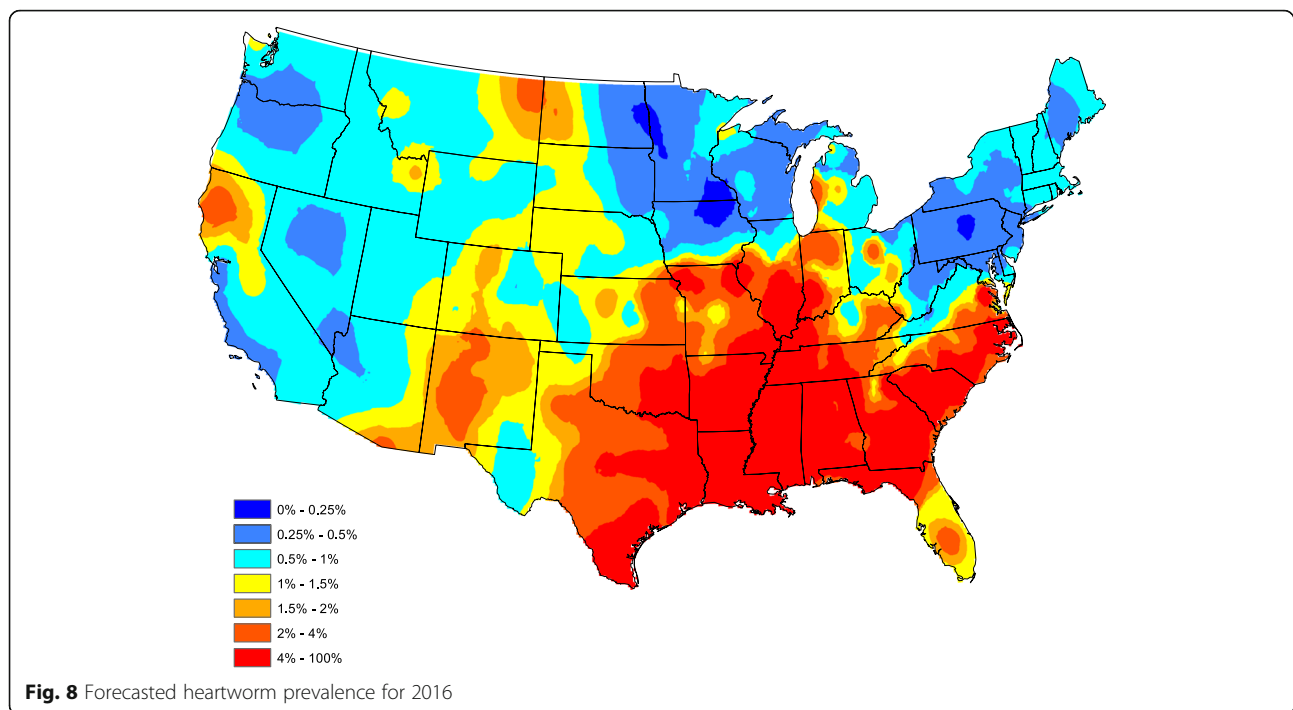


Fig. 7 Forecasted heartworm prevalence for 2015



and Antech, averages greater than 100,000 new cases annually. Annual data likely represent the true annual incidence of heartworm infection in domestic dogs: when diagnosed with heartworm, most dogs are either treated, or in some cases euthanized, due to poor outcome or financial constraints [34]. While fulminant infection with *D. immitis* may be due to lack of owner compliance in use of preventatives, it also may be due to misunderstanding the disease risk; mosquito vectors are known to be dynamic in their range and survival under changing climatic conditions.

To enhance veterinary client education and illuminate the benefits of preventatives, factors associated with *D. immitis* transmission [16, 35] were identified and used to develop a spatial-temporal conditional autoregressive forecast model of heartworm prevalence. A comparison of observed versus forecasted heartworm prevalence was made in 2015 and was quite accurate. This may be attributed, in part, to the fact that many of the factors influencing heartworm prevalence do not change significantly from year to year (e.g. forest coverage, population density, household income). While temperature and humidity change annually and are important disease risk factors [36], these factors are still reasonably predictable; however, environmental or climate catastrophes (e.g. regional climate shifts, flooding, hurricanes) could impact heartworm incidence. Finally, mosquito populations can fluctuate greatly from year to year as they depend on numerous local land-use and environmental factors. Some competent vectors of *D. immitis* are still expanding their range in the US [37].

Several of the mosquito presence/absence factors were not included in our final model; this may be because our currently available data are only presence/absence, whereas mosquito abundance, a purportedly more powerful factor, varies annually at a local level. More accurate mosquito counts would likely yield more accurate forecasts. In addition, human activities such as treatment abatement programs may impact mosquito abundances. Since the introduction of West Nile virus into the US, localities have developed or expanded mosquito control programs, including reducing breeding habitats and application of pesticides. With increased concern over the Chikungunya and Zika viruses, it is possible that increased mosquito control may be initiated in the coming year(s). If such programs are initiated, mosquito abundance counts should take these programs into account.

Conclusion

In conclusion, our 2016 heartworm disease forecast (Fig. 8) has some noteworthy implications for veterinary practitioners, including an increased prevalence in northern California, eastern Montana, and central New Mexico. A relatively small increase in risk is predicted in some areas where heartworm is likely under appreciated, such as parts of the Dakotas and Nebraska. Importantly, our data indicate that all regions of the lower 48 United States have some risk for heartworm infection. Our maps and forecasts provide veterinarians with evidence-based recommendations for use of preventive in non-

endemic regions of the US and support the recommendation of year-round use of preventive in high risk areas. Ultimately, we believe that these methods can be used to forecast multiple vector-borne diseases with veterinary and human health impacts, including Lyme disease, ehrlichiosis and anaplasmosis. Currently, the Companion Animal Parasite Council (CAPC, www.capcvet.org) provides monthly updates of heartworm prevalence on a county by county scale. Through a combination of real-time updates and forecasting efforts, we hope to see fewer cases of heartworm disease in dogs and cats in the future.

Abbreviations

AR(1): First order temporal autoregression; CAPC: Companion Animal Parasite Council; CAR: Conditional autoregressive model; HPD: Highest posterior density; MCMC: Markov chain Monte Carlo; ML: Macrocytic lactones

Acknowledgements

The IDEXX and Antech corporations are thanked for their data contributions. The Companion Animal Parasite Council financially supported this work. Two referees are thanked for comments that improved this paper.

Funding

YL and RBL were partially supported by Grant DMS 1407480 from the National Science Foundation. CM was partially supported by Grant R01 AI121351 from the National Institutes of Health.

Availability of data and material

The data used in this study are available from www.capcvet.org.

Authors' contributions

All authors made substantial contributions to the manuscript. YL, CSM, and RBL performed the statistical analysis and contributed to the write-up. DDB, SN, and MJY wrote parts of the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable (retrospective study using data that are publically available).

Author details

¹College of Veterinary Medicine, Cornell University, Ithaca, NY, USA. ²Department of Mathematical Sciences, Clemson University, Clemson, SC, USA. ³Department of Molecular Biomedical Sciences, North Carolina State University, Raleigh, NC, USA. ⁴Southeastern Cooperative Wildlife Disease Study, Department of Population Health, College of Veterinary Medicine and the Warnell School of Forestry and Natural Resources, The University of Georgia, Athens, GA, USA.

Received: 26 July 2016 Accepted: 19 September 2016

Published online: 10 October 2016

References

- Morchon R, Carreton E, González-Miguel J, Mellado-Hernández I. Heartworm disease (*Dirofilaria immitis*) and their vectors in Europe - new distribution trends. *Front Physiol.* 2012;3:196.
- Simon F, Morchon R, Gonzalez-Miguel J, Marcos-Atxutegi C, Siles-Lucas M. What is new about animal and human *dirofilariosis*? *Trends Parasitol.* 2009;25:404–9.
- Bowman DD, Atkins CE. Heartworm biology, treatment, and control. *Vet Clin North Am Small Anim Pract.* 2009;39:1127–58.
- Bowman DD, Georgi JR. *Georgis' Parasitology for Veterinarians.* 9th ed. Saint Louis: Saunders/Elsevier; 2009.
- Colby KN, Levy JK, Dunn KF, Michaud RI. Diagnostic, treatment, and prevention protocols for canine heartworm infection in animal sheltering agencies. *Vet Parasitol.* 2011;176:333–41.
- Maxwell E, Ryan K, Reynolds C, Pariaut R. Outcome of a heartworm treatment protocol in dogs presenting to Louisiana State University from 2008 to 2011: 50 cases. *Vet Parasitol.* 2014;206:71–7.
- Wolstenholme AJ, Evans CC, Jimenez PD, Moorhead AR. The emergence of macrocyclic lactone resistance in the canine heartworm, *Dirofilaria immitis*. *Parasitol.* 2015;142:1249–59.
- Venco L, Marchesotti F, Manzocchi S. Feline heartworm disease: A 'Rubik's-cube-like' diagnostic and therapeutic challenge. *J Vet Cardiol.* 2015;17 Suppl 1:S190–201.
- Lee AC, Montgomery SP, Theis JH, Blagburn BL, Eberhard ML. Public health issues concerning the widespread distribution of canine heartworm disease. *Trends Parasitol.* 2010;26:168–73.
- Orihel TC, Eberhard ML. Zoonotic filariasis. *Clin Microbiol Rev.* 1998;11:366–81.
- Little SE, Beall MJ, Bowman DD, Chandrashekar R, Stamaris J. Canine infection with *Dirofilaria immitis*, *Borrelia burgdorferi*, *Anaplasma* spp., and *Ehrlichia* spp. in the United States, 2010–2012. *Parasit Vectors.* 2014;7:257.
- Levy JK, Lappin MR, Glaser AL, Birkenheuer AJ, Anderson TC, Edinboro CH. Prevalence of infectious diseases in cats and dogs rescued following Hurricane Katrina. *J Am Vet Med Assoc.* 2011;238:311–7.
- Yabsley MJ, Dresden-Osborne C, Pirkle EA, Kirven JM, Noblet GP. Filarial worm infections in shelter dogs and cats from northwestern South Carolina, USA. *Comp Parasitol.* 2004;71:154–7.
- Gates MC, Nolan TJ. Factors influencing heartworm, flea, and tick preventative use in patients presenting to a veterinary teaching hospital. *Prev Vet Med.* 2010;93:193–200.
- Bourguinat C, Lee AC, Lizundia R, Blagburn BL, Liotta JL, Kraus MS, et al. Macrocyclic lactone resistance in *Dirofilaria immitis*: Failure of heartworm preventives and investigation of genetic markers for resistance. *Vet Parasitol.* 2015;210:167–78.
- Wang D, Bowman DD, Brown HE, Harrington LE, Kaufman PE, McKay T, et al. Factors influencing US Canine heartworm (*Dirofilaria immitis*) prevalence. *Parasit Vectors.* 2014;7:264–82.
- Antech Diagnostics, Incorporated: <http://www.antechdiagnostics.com/Main/Home.aspx>
- IDEXX Laboratories, Incorporated: http://www.idexx.com/view/xhtml/en_us/corporate/home.jsf
- Stich RW, Blagburn BL, Bowman DD, Carpenter C, Cortinas MR, Ewing SA, et al. Quantitative factors proposed to influence the prevalence of canine tick-borne agents in the United States. *Parasit Vectors.* 2014;7:417–24.
- Cressie N, Wikle CK. *Statistics for Spatio-temporal data.* New York: John Wiley & Sons; 2011.
- López-Quilez A, Muñoz F. Review of spatio-temporal models for disease mapping. In: Final report for the EUROHEIS 2 project. 2009. <http://www.uv.es/~famarmu/doc/Euroheis2-report.pdf>.
- Martínez-Beneito MA, López-Quilez A, Botella-Rocamora P. An autoregressive approach to spatio-temporal disease mapping. *Stat Med.* 2008;27:2874–89.
- Xia H, Carlin BP. Spatio-temporal models with errors in covariates: mapping Ohio lung cancer mortality. *Stat Med.* 1998;17:2025–43.
- Nobre AA, Schmidt AM, Lopes HF. Spatio-temporal models for mapping the incidence of malaria in Pará. *Environmetrics.* 2005;16:291–304.
- Besag J. Spatial interaction and the statistical analysis of lattice systems. *J R Stat Soc Series B.* 1974;36:192–236.
- Banerjee S, Carlin BP, Gelfand AE. *Hierarchical modeling and analysis for spatial data.* 2nd ed. Boca Raton: Chapman and Hall/CRC; 2014.
- Brockwell PJ, Davis RA. *Introduction to time series and forecasting.* 2nd ed. New York: Springer; 2002.
- Harrison J, West M. *Bayesian forecasting & dynamic models.* New York: Springer; 1999.
- Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. *Bayesian data analysis.* 3rd ed. Boca Raton: Chapman & Hall/CRC; 2014.
- Nene V, Wortman JR, Lawson D, Haas B, Kodira C, Tu ZJ, et al. Genome sequence of *Aedes aegypti*, a major arbovirus vector. *Science.* 2007;316(5832):1718–23.
- Parham PE, Waldock J, Christophides GK, Hemming D, Agosto F, Evans KJ, et al. Climate, environmental and socio-economic change: weighing up the

- balance in vector-borne disease transmission. *Philos Trans R Soc Lond B Biol Sci.* 2015;370(1665). doi:10.1098/rstb.2013.0551.
32. Alegana VA, Wright JA, Nahzat SM, Butt W, Sediqi AW, Habib N, et al. Modelling the incidence of *Plasmodium vivax* and *Plasmodium falciparum* malaria in Afghanistan 2006–2009. *PLoS One.* 2014;9(7), e102304.
 33. Perez-Florez M, Ocampo CB, Valderrama-Ardila C, Alexander N. Spatial modeling of cutaneous *leishmaniasis* in the Andean region of Colombia. *Mem Inst Oswaldo Cruz.* 2016. doi:10.1590/0074-2760160074.
 34. Polak KC, Smith-Blackmore M. Animal shelters: managing heartworms in resource-scarce environments. *Vet Parasitol.* 2014;206(1–2):78–82.
 35. Brown HE, Harrington LC, Kaufman PE, McKay T, Bowman DD, Nelson CT, et al. Key factors influencing canine heartworm, *Dirofilaria immitis*, in the United States. *Parasit Vectors.* 2012;5:245.
 36. Beugnet F, Chalvet-Monfray K, Loukos H. FleaTickRisk: a meteorological model developed to monitor and predict the activity and density of three tick species and the cat flea in Europe. *Geospat Health.* 2009;4(1):97–113.
 37. Ogden NH, Milka R, Caminade C, Gachon P. Recent and projected future climatic suitability of North America for the Asian tiger mosquito *Aedes albopictus*. *Parasit Vectors.* 2014;7:532.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

