Ecological Informatics 52 (2019) 150–158

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

Predicting larval tick burden on white-footed mice with an artificial neural network

Stacy Mowry[a,*], Felicia Keesing[b], Ilya R. Fischhoff[a], Richard S. Ostfeld[a]

[a] Cary Institute of Ecosystem Studies, 2801 Sharon Turnpike, Millbrook, NY 12564, USA
[b] Bard College, PO Box 5000, Annandale-on-Hudson, New York 12504, USA

ARTICLE INFO

Keywords:
Aggregation
Blacklegged ticks
Ixodes scapularis
White-footed mouse
Peromyscus leucopus
Larval burden
Artificial neural network (ANN)

ABSTRACT

White-footed mice are important hosts for immature blacklegged ticks (*Ixodes scapularis*) and the most competent reservoir hosts for several tick-borne pathogens, including the agent of Lyme disease, in eastern North America. The distribution of larval ticks on individual mice tends to be highly heterogeneous, potentially resulting in few individual hosts causing the majority of host-to-tick transmission events. In this study, we created an artificial neural network (ANN) model using a 20 year data set from Millbrook, NY, to understand which attributes of mice or the environment predict high larval burden. Furthermore, we performed a sensitivity analysis to explore the importance of, and interactions between, the most influential attributes. Our analysis indicated that highest larval burden is predicted in warmer and drier than average years when host abundance is low, and that climatic conditions and host density are far more important in predicting larval burden than traits of individual mice, a finding that could have human health implications within the context of a warming climate. Practically, our results suggest that instead of basing tick-control treatments on particular attributes of hosts, treatments should be targeted based on climate factors. Additionally, our results highlight the importance of including variable interactions in models aiming to predict vector (tick) aggregation, and, most broadly, demonstrate the utility of ANNs in understanding aggregation of ticks and other vectors.

1. Introduction

In eastern and central North America, blacklegged ticks (*Ixodes scapularis)* are key vectors in the spread of zoonotic pathogens including the agents of Lyme disease, human babesiosis, and human granulocytic anaplasmosis. These pathogens are transmitted to ticks during blood meals on infected vertebrate hosts. Of the many vertebrate hosts on which these ticks feed, the white-footed mouse (*Peromyscus leucopus*) plays a particularly important role in the spread of tick-borne diseases. Not only are larval burdens on mice 2–3× higher than on other rodent hosts (Schmidt et al., 1999), likely due to higher survival (Keesing et al., 2009), but also, the white-footed mouse is the most competent reservoir for *Borrelia burgdorferi s.l.* (Lane et al., 1991; LoGiudice et al., 2003), the bacterial agent responsible for Lyme disease, as well as for *Babesia microti* and *Anaplasma phagocytophilum*, the agents of babesiosis and anaplasmosis, respectively (Hersh et al., 2012; Levi et al., 2016).

Within a host population, vector burdens are often heterogeneous such that a minority of hosts feed the majority of vectors, while many hosts feed very few (Woolhouse et al., 1997). In fact, it is widely accepted that parasite distributions within host populations are generally most adequately described by the negative binomial distribution (Anderson and May, 1978; Shaw et al., 1998), which is characterized by a sharp peak and long right tail. Larval aggregation is particularly important because it allows for a scenario in which a single infected host can infect a large group of naïve larvae, making a small proportion of hosts potentially responsible for a vast percentage of transmission events. Furthermore, it is well established that the potential rate of spread of an infection in a host population, R_0, increases with the degree of vector aggregation (Woolhouse et al., 1997). Based on the potential impact of these few, key hosts, there has been widespread interest in determining which factors are associated with high tick burden on select hosts.

Within the past four decades, researchers have identified various attributes of hosts and the environment associated with high tick burden. For example, some studies (Ostfeld et al., 1996; Schmidt et al., 1999; Devevey and Brisson, 2012) have found that males tend to have higher mean tick burdens than females and that heavier, older hosts tend to have higher burdens than lighter, younger hosts (Brunner and Ostfeld, 2008). Additionally, it has been well established that larval *Ixodes scapularis* ticks have strong seasonal life histories (Spielman

* Corresponding author.
E-mail address: mowrys@caryinstitute.org (S. Mowry).

https://doi.org/10.1016/j.ecoinf.2019.04.002
Received 4 February 2019; Received in revised form 11 April 2019; Accepted 13 April 2019
Available online 20 April 2019

et al., 1985) characterized by a small, early-summer peak in host-seeking activity and a much larger peak in the late summer (Spielman et al., 1985; Ostfeld et al., 1996). Furthermore, presumably due to the strong seasonal life history, multiple studies (Brunner and Ostfeld, 2008; Kiffner et al., 2010) have found seasonality (time of year) to be the single most important factor when predicting tick burden.

Although there has been significant progress in explaining heterogeneity in parasite aggregation among vertebrate hosts, many studies have produced contradictory results, and the relative importance of environmental attributes versus individual-host attributes on tick burden remains controversial. For example, some studies have concluded that individual-host traits play little to no role in predicting tick burden (Calabrese et al., 2011; Lutermann et al., 2011), instead finding more support for the importance of tick density and spatial aggregation. Conversely, others have found that individual-host attributes are instrumental in predicting tick burden (Dallas et al., 2012; Devevey and Brisson, 2012). It seems possible that tick burdens are a complex function of random spatial aggregation of host-seeking ticks, individual-host characteristics, host abundance or dispersion, and climate factors (Brunner and Ostfeld, 2008; Kiffner et al., 2010). The lack of consistency between studies may derive from the complexity of the interactions between the various attributes (Brunner and Ostfeld, 2008; Sackett, 2018).

Artificial neural networks (ANNs) have long been recognized as powerful tools to explore complex, non-linear problems and have proved their utility in a range of ecological areas from freshwater systems (Brosse et al., 1999) to species richness (Monteil et al., 2005). Importantly, the models offer a way to uncover patterns in data without explicitly having to define the relationships between the model's variables (Jorgensen, 1999; Lek et al., 1996) and, unlike commonly utilized statistical models, ANNs do not have any predefined restrictive assumptions regarding variable distributions or underlying relationships between dependent and independent variables (Chang, 2005).

Here we present a simple ANN that predicts larval burden on white-footed mice based upon individual-specific, population level, and climate attributes. We followed our ANN models with sensitivity analyses to gain more insight into the importance of, and interactions between, each of the attributes. We explore the general usefulness of ANNs and follow-up sensitivity analyses for exploring complex, interacting determinants of larval burden on key reservoir hosts.

2. Materials and methods

2.1. Field data

We used 20 years of data from a small-mammal trapping program at the Cary Institute of Ecosystem Studies campus in Dutchess County, New York. This data set includes 16,864 observations of newly trapped white-footed mice. As ANNs cannot process data records with missing components, we excluded any incomplete observations (i.e., observations where sex or age was not recorded) resulting in 16,258 complete observations. The data were collected on six permanent trapping grids, each consisting of 242 Sherman traps arranged in pairs along an 11×11 grid covering ~2.25 ha (Brunner and Ostfeld, 2008). Each grid was trapped for 2 consecutive nights between 6 and 16 times per year between April and October from 1995 to 2015. In 1995 and 1997 mouse density on three of the grids was manipulated. In 1995 both mice and chipmunks were removed during June and July and in 1997 only mice were removed from mid June through July (Brunner and Ostfeld, 2008). For each capture, an animal's weight, sex age, reproductive status, pregnancy and lactation status (for females) and larval *I. scapularis* count on the head and ears were recorded. To estimate mouse density, we used Minimum Number Alive (MNA), the minimum number of animals known to be alive during a sampling period. For small mammals trapped on a similar schedule, MNA has been shown to be highly correlated with the commonly used Jolly-Seber model: a method to estimate population size of marked animals in an open population (Hilborn et al., 1976). The data from the mark-recapture study provided values for the variable we are interested in predicting: the burden of larval ticks on individual white-footed mice. These data also provided values for some of the candidate variables that might explain larval tick burdens, including population density of mice, host age, sex, reproductive condition, and season.

2.2. Climate data

In addition to individual and population attributes of mice and the season of their capture, we also explored climatic conditions as possible determinants of tick burdens. Daily temperature data were obtained from the Cary Institute archive and Cumulative Degree Days (CDDs) were calculated as the cumulative degrees above 0 °C (based on the average daily temperature) from the first of the year up to and including the date of the trapping session. We used the Palmer Hydrological Drought Index (PHDI) as our measure of drought; this is a long term drought measure, based on soil moisture, reservoir, and ground water levels. New York Hudson Valley region drought data were obtained from the National Oceanic and Atmospheric Administration climate database (NOAA, 2018).

2.3. Data processing for NN

To normalize variable effect within our ANN, all quantitative variables were scaled between −0.5 and 0.5 using a linear max-min transformation. Additionally, all qualitative variable values were converted to separate, binary variables. To avoid conflation between day of the year and CDD, CDD was converted to ΔCDD, defined as the CDD at the date of the capture above or below the 20 year average for that date. Similarly, to avoid conflation between age and weight, weight was converted to Δweight, defined as the grams below or above the average weight within each age class.

2.4. Model formation

Our ANN was fit using trapping data and the R package neuralnet (Fritsch et al., 2018) in R version 3.3.1. (R CoreTeam, 2016). Neural networks are non-linear functions, broken into a series of hidden layers, which take input data and transform them to desired outputs. Each node in a layer of a neural network transforms a linear combination of a node-specific bias term and the weighted input data based on a defined, and not necessarily linear, activation function. The output from each node then becomes a new input for nodes in the next layer. The final layer of the network calculates the output by, once again, transforming a linear combination of its node-specific bias term and the weighted output data from the last hidden layer, based on another activation function. A loss function, defined as the difference between observed values and the networks output, is calculated and iteratively used to modify the weights and biases in the network through a process known as backpropogation. A detailed description of the mathematical theory of neural networks can be found in Bishop (1996). In our network, we used resilient backpropogation with weight back tracking, a sum of squares error function, and sigmoid and linear activation functions for the hidden and output layer, respectively. There is general consensus that most problems can be sufficiently modeled with a 3 layer network (Bishop, 1995); therefore, in order to facilitate interpretability, all of our models contained only a single hidden layer.

In order to determine optimal network architecture and variable selection, we compared Akaike Information Criterion (AIC) values (as calculated in Panchal et al., 2010) between candidate networks with different input variable and hidden node combinations. Specifically, we began with three versions of the full model containing all potential variables (adult, sub-adult, sex, weight, reproductive condition, pregnancy status, lactation status, CDD, day of year, MNA and PHDI) with

11, 10 or 9 nodes in our hidden layer. Although, generally, models with more nodes yield a better fit, to avoid over fitting, the number of nodes in the hidden layer should not exceed the number of input variables in the neural network (Heaton, 2005). Therefore, we fit all models with *n*, *n-1* or *n-2* nodes, where *n* is the number of input variables included in the model. After fitting each model with all potential variables, we then implemented a step-down approach, dropping single variables sequentially to construct additional candidate models and, ultimately, retained the model with the lowest AIC value. Because neural networks are nonlinear, their loss functions have many local minima: parameterizations where the function is the lowest in its immediate neighborhood, but not the lowest in the function's entire range. This means the network may converge to a local minimum instead of the global minimum, the parameterization where the function is the lowest in its entire range. To limit influence of non-optimal models, defined as models where the network converged to a local minimum, each network was fit 10 times with random initial values, and the average AIC value between the 10 models was considered in our model comparison.

Once model architecture was established, we constructed 20 networks with the determined architecture and retained the network with the lowest error rate to decrease the likelihood of convergence to a local minimum. Finally, we determined optimal model threshold, the value that the improvement in the error function must exceed for the model to continue to iterate, based on R^2 between actual and predicted larval burden. We stopped model iterations at the value at which additional epochs of training yielded minimal improvement in model performance (Appendix, Table 2). In order to maximize training data available, we utilized the resubstitution method, which uses the same data to fit the model and to estimate model error. We acknowledge that this methodology may overestimate model performance as defined by accuracy of predictions (Twomey and Smith, 1996).

It is widely accepted that warmer years advance larval phenology (Levi et al., 2015). Therefore, in order to validate our model, we broke our original data set into 5 groups of ΔCDD: very high (ΔCDD between −286 and −166), high(ΔCDD between −166 and −40), moderate (ΔCDD between −40 and 82), low (ΔCDD between 82 and 203), and very low (ΔCDD between 203 and −327), and examined the timing of the beginning of the large larval peak for each group, with the expectation that when ΔCDD is higher, the larval peak will begin at an earlier day of the year, reflecting the advanced phenology caused by warmer years.

2.5. Model analysis

Our model analysis aimed to answer the following three questions. First, under which conditions should we expect high, defined as at least 10 larvae, and highest, defined as the model's 100 highest predicted burdens, larval burden on *P. leucopus*? Second, which attributes are most influential in predicting larval burden? And, third, how do key attributes interact to influence larval burden? To address all questions, we examined full model output using a simulated data set that contains all combinations of discretized input variables, within ranges measured in our trapping data, as our input variables. To construct the simulated data set, we discretized each input variable, except day of the year, within its range (i.e., between maximum and minimum values recorded in our trapping data for continuous variables, and between 0 and 1 for binary variables). A priori, we hypothesized that the effect of host abundance and host weight would be linear, so we discretized host abundance and weight into 5 quantiles evenly distributed from the minimum value recorded in our data to the maximum value recorded in our data in order to examine model behavior over the full range of possible input values (0%, 25%, 50%, 75% and 100%). Also a priori, we hypothesized that the effect of climate variables could be more complex so we discretized ΔCDD and PHDI into 11 quantiles again evenly distributed from the minimum value recorded in our data to the maximum value recorded in our data (0%, 10%, 20%, …, 100%), allowing us to detect possible intricate patterns based on climate variables modeled by our ANN. To simulate conditions at the small larval peak, we held day of year constant at day 156 (June 5). To simulate conditions at the large larval peak, we varied day of the year from day 244 to day 260 (August 30–September 17), to account for the effect of CDD on the timing of larval phenology, and considered the average larval burden over the 17 day span as our output. Our simulated data set had 54,450 unique variable combinations. Within our trapping data, only about 8% and 6% of sub-adult and adult females were lactating or pregnant, respectively, during the main larval peak. Based on this finding, we only considered pregnant and lactating females during the small peak, when, according to our data, about 50% and 43% of sub-adult or adult females were pregnant or lactating, respectively.

To address our first question, under which conditions should we expect high and highest larval burden on *P. leucopus,* we first evaluated the 100 data points that yielded the highest model output, predicted larval burden, from our simulated data set. To more generally determine conditions that predict high larval burden, we defined high as at least 10 larvae on a single host and inputted our simulated data for each host subset (i.e., juvenile males, adult females) into our model (Eq. (1)). To analyze model behavior, we constructed multi-faceted heat maps using the R package ggplot2 (Wickham et al., 2018) with our model output, either average larval burden at the large larval peak or larval burden at the earlier peak, as our dependent variables.

To address our second question, which host traits and extrinsic conditions are most influential in predicting tick burden, we calculated each variable's overall effect to determine which variables most impacted our model. Although techniques such as Garson's and Lek's algorithms have been developed to interpret variable importance within ANNs, these methodologies often reach different conclusions, and do not adequately interpret interactions (Olden and Jackson, 2002). Therefore, to interpret our variable importance, we performed a comprehensive sensitivity analysis. For each variable, except day of year, we constructed *k* single variable models based on our original multivariable ANN, where average peak larval burden was a function of the focal variable and *k* is the total number of discrete combinations of non-focal variables (note: variables were discretized for our sensitivity analysis exactly as they were discretized for model analysis). We created *k* single variable models, as opposed to creating one single variable model with all non-focal variables held constant at the middle of their range, to account for variable interactions. We then calculated the sensitivity for each of the *k* models as the difference between the function maximum and the function minimum when varying the focal variable from −0.5 to 0.5 for continuous variables or 0 or 1 for binary variables. We took the average sensitivity over the *k* models to obtain overall effect for the focal variable.

Finally, to interpret variable interactions, and investigate our third question, we examined interaction multi-faceted heat maps for each variable, except day of year. Each cell in these heat maps represents one of the *k* combinations of non-focal variables. The difference between function maximum and function minimum, for that set of conditions, served as the dependent variable.

3. Results

3.1. Descriptive statistics

3.1.1. Larval burden

During the 20 years, 16,864 white-footed mice were sampled for demographic information and checked for larval *I. scapularis.* The mean number of larvae per individual was 5.97 (range 0–242, mode = 0, median = 2) and larval burden more closely followed a negative binomial distribution than a normal distribution. The vast majority of individuals had very few larvae (64% of individuals had a larval burden of < 4) and very few individuals had many larvae (< 2% had a larval burden > 50). The mean number of larvae on males was 6.57 (range

0–217, mode = 0, median = 2), while the mean number of larvae on females was 5.01 (range 0–242, mode = 0, median = 2). Based on age class, the mean number of larvae on juveniles was 3.45 (range 0–172, mode = 0, median = 1), the mean number on sub-adults was 4.97 (range 10–169, mode = 0, median = 1), and the mean number on adults was 7.63 (range 0–242, mode = 0, median = 3).

3.1.2. Averages of relative explanatory variables

The average weight of adults was 21.08 g, the average weight of sub-adults was 16.37 g, and the average weight of juveniles was 12.16 g; Δweight is in relation to these values. Average CDD at day 244 (August 30), the approximate start of the main seasonal peak of larval tick abundance was 2952.31° (average CDD values for each day of the year given in Appendix Fig. 1); ΔCDD is in relation to these values.

3.2. Full model

Our final model, described in Eq. (1), includes all variables except whether male or female reproductive features were visible, contains 10 nodes in the hidden layer, and was fit with a threshold of 1.

$$y = 10^{b_o + \sum_{n=1}^{10} w_n \frac{1}{1 + e^{-b_n + \sum_{i=1}^{10} w_{in} x_i}}} - 1 \quad (1)$$

b_o, b_n, w_n, and w_{in} represent the bias term in the output layer, the bias terms in the hidden layer, the connection weights between the hidden nodes and output, and the connection weights between the input variables and hidden nodes, respectively. Full model comparison results and AIC values are given in Table 1 and full model weights and bias terms are given in Appendix Table 1. Overall, the ANN produced an R^2 value of 0.52 signifying that our model accounted for 52% of the variation seen in the data. Furthermore, the model predicted about 24% of the larval burdens correctly, and predicted an additional 57% within 5 larvae. Full distribution of model output error terms is given in Table 2. An analysis of residuals showed that the model is not very accurate when predicting larval burdens of 0, or larval burdens greater than about 50 (Appendix Fig. 1). Inspection of the day of the year that the large larval peak begins (defined as the day on which the vertex of the concave up parabola, initiating the peak, falls) within each ΔCDD group (very high, high, moderate, low and very low) showed that our ANN successfully detected that warmer years see an earlier peak than cooler years (Appendix Fig. 2), offering validation for our model.

Table 1
Support for models of larval burden on P.leucopus. Summary of AIC and ΔAIC values for each of the candidate models considered in model comparison. Full model included all of the following variables: sex, whether or not the individual is a sub-adult as opposed to juvenile (sub), whether or not the individual is an adult as opposed to sub-adult, weight (wt), whether the vagina was visible (vag), whether testes were visible, pregnancy status (pg), lactation status (lac), cumulative degree days above or below 20 year average from 1995–2015 (CDD), day of year, minimum number of hosts alive (MNA), and Palmer Hydrological Drought Index (PHDI).

Model	AIC	Δ AIC
full - (vag + testes)	−45,702.800	0
full - lac	−45,687.700	15.114
full - vag	−45,653.200	49.543
full - (vag + lac)	−45,643	59.817
full -sub	−45,641.600	61.151
full - testes	−45,637.200	65.582
full - (sub + sex)	−45,631	71.782
full - (lac + testes)	−45,610.500	92.259
Full	−45,594.200	108.626
full - adult	−45,591.800	110.952
full - pg	−45,591.800	111.028
full - sex	−45,590.800	111.978
full - (vag + sex)	−45,573.600	129.173
full - (vag + sub)	−45,545.100	157.725
full - (vag + testes + sex)	−45,543.900	158.864
full - (lac + sub)	−45,541.700	161.097
full - (testes + sex)	−45,528.800	173.967
full - (vag + testes + sub)	−45,525.800	177.018
full – wt	−45,506.300	196.480
full - (sub + testes)	−45,480	222.777
full - (lac + sex)	−45,465	237.823
full - (vag + testes + lac)	−45,162.600	540.213
full - PHDI	−45,020.900	681.855
full - MNA	−44,933.200	769.621
full - CDD	−44,693.800	1008.950

Table 2
Distribution of residuals between actual larval burden from raw data and predicted larval burden from model output. Values indicate the percentages of predicted values that fall into each residual category. (i.e., 23.67% of our predicted values from our model output were correct, 57.19% were off by 1 – 5 larvae, etc.)

Residual	Percentage
0	23.67%
1–5	57.19%
6–10	9.48%
10–25	7.87%
26–50	2.12%
> 50	0.84%

3.3. Model analysis

3.3.1. Question 1: under which conditions does our model predict high larval burdens?

During the large (main) larval peak, 82% of hosts that our ANN predicted to have the highest larval burden were sub-adults or adults. 67% were males and 72% were of at least average weight. One hundred percent of hosts that our ANN predicted to have the highest larval burden were predicted when host abundance was 1 and 83% and 79% were predicted during moderately warm (ΔCDD 143°–203 °C above average), and dry (PHDI between −3.83 and −2.41) years (Fig. 1). The model predicted at most 86 larvae on a single host.

The model predicted similar general patterns across all subgroups (R^2 values between model output at the large larval peak for adult males and all other subgroups are given in Appendix Table 3); therefore, we analyzed output associated with a single subgroup, adult males, to understand when the model predicts high larval burden, defined as > 10 larvae per host.

When MNA is very low, our ANN predicts > 10 larvae per individual, regardless of host weight CDD or PHDI, with highest larval burden predicted in a relatively warm and dry climate. When MNA is moderate, generally, our ANN predicts more than10 larvae per host when ΔCDD > 0 with little influence of host weight or PHDI. As MNA further increases, however, our ANN only predicts > 10 larvae per host when the climate is both warm and relatively dry, and as host body size declines, a drier climate is required (Fig. 2).

3.3.2. Question 2: which attributes are most influential in predicting larval burden?

According to our sensitivity analysis, ΔCDD was the most influential variable in our model, influencing the predicted larval burden by an average of about 18 larvae, over all discrete combinations of other variables. MNA and PHDI were also important, affecting larval burden predictions by about 12 and 8 larvae on average, respectively. In general, individual-specific attributes were less important than other attributes. Of the individual-specific attributes, weight had the largest effect on model output, influencing predicted larval burden by about 4 larvae on average. Sex and sub-adult status each affected larval burden

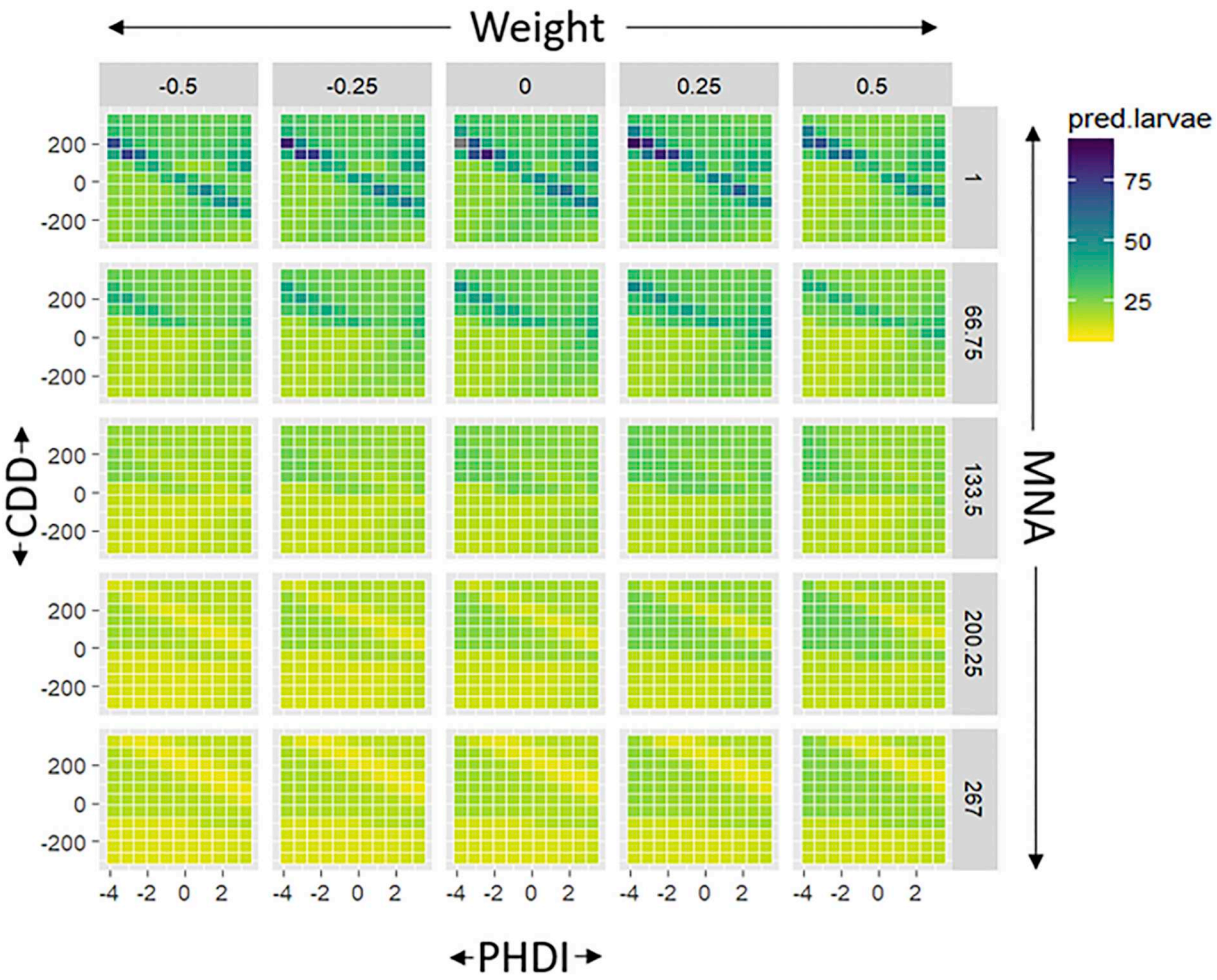

Fig. 1. Heat map of full model output of predicted larval burden on adult, male white-footed mice during the large larval peak (averaged over days 244–260). Large x-axis displays Δweight from the average weight within the subgroup at 0, 25th, 50th, 75th and 100th percentile. Small x-axis displays PHDI at 0, 10th, 20th, …, 100th percentile ranging from −4 to 4 (note: full PHDI scale ranges from −10 to 10). Large y-axis displays MNA of white-footed mice at 0, 25th, 50th, 75th and 100th percentile, ranging from 1 individual to 267. Small y-axis displays ΔCDD from the 20 year average (from 1995 to 2015) at 0, 10th, 20th, …, 100th percentile ranging from 333° below average to 333° above average.

by about 2 larvae (with males predicted to have more larvae than females and sub-adults more than juveniles) on average. Whether a female was pregnant or a host was an adult (as opposed to a sub-adult), had very little influence on predicted larval burden: each increasing predictions by on average < 1 larva. Finally, lactation decreased predicted burden by about 2 larvae on average (Table 3).

3.3.3. Question 3: how do attributes interact to influence larval burden?

According to our model, there is substantial variation in the influence of ΔCDD, PHDI and MNA on predicted larvae based on the state of other variables. ΔCDD, for example, has a very large influence on larval burden in years with a drier climate (PHDI < −0.3) and low host abundance (MNA = 1), but a much smaller influence in wetter years or years when host abundance is high (Fig. 3a). Similarly, PHDI has a much larger effect on larval burden in years with moderate to high temperatures (CDD 82°–265 °C above average), especially when host abundance is low (MNA = 1) (Fig. 3b). Finally, host abundance has the largest influence on predicted larval burden in warmer than average years (CDD > 20 °C above average), particularly those with a drier climate (PHDI < −2) (Fig. 3c). The impact of weight on the model was relatively constant across all conditions, and there were comparatively minimal interactions between sex, age, pregnancy or lactation and any other variables, reflecting their small overall influence on larval burden.

4. Discussion

Aggregation of vectors or parasites on hosts is known to increase the potential rate of spread of an infection in a host population (Woolhouse et al., 1997). Within the realm of tick-borne diseases, larval aggregation on white-footed mice, particularly competent reservoirs for *Borrelia burgdorfori s.l.* (Lane et al., 1991; LoGiudice et al., 2003), increases the spread of the spirochete bacterium. Therefore, understanding which attributes predict high larval burden on white-footed mice is important because it could allow for more targeted control efforts, ultimately reducing the spread of Lyme disease to human populations.

The purpose of our study was to create and analyze a simple ANN that predicts larval burden on white-footed mice based on individual-specific, population, and climate attributes in order to uncover complex, possibly overlooked patterns governing larval burden and understand the importance of, and interactions between, these attributes. Larval burdens within our data were highly aggregated: while most individuals had very low larval burdens, a few fed many larvae. Overall, our model, which included day of year, host age class, host sex, host weight, host pregnancy status, host lactation status, host abundance, and measures of annual temperature and drought, accounted for 52% of the variation in larval burden. Although our model predicted nearly 81% of host larval burden within 5 larvae, our model predictions were not very accurate when actual larval burden was 0 or > 50. Poor

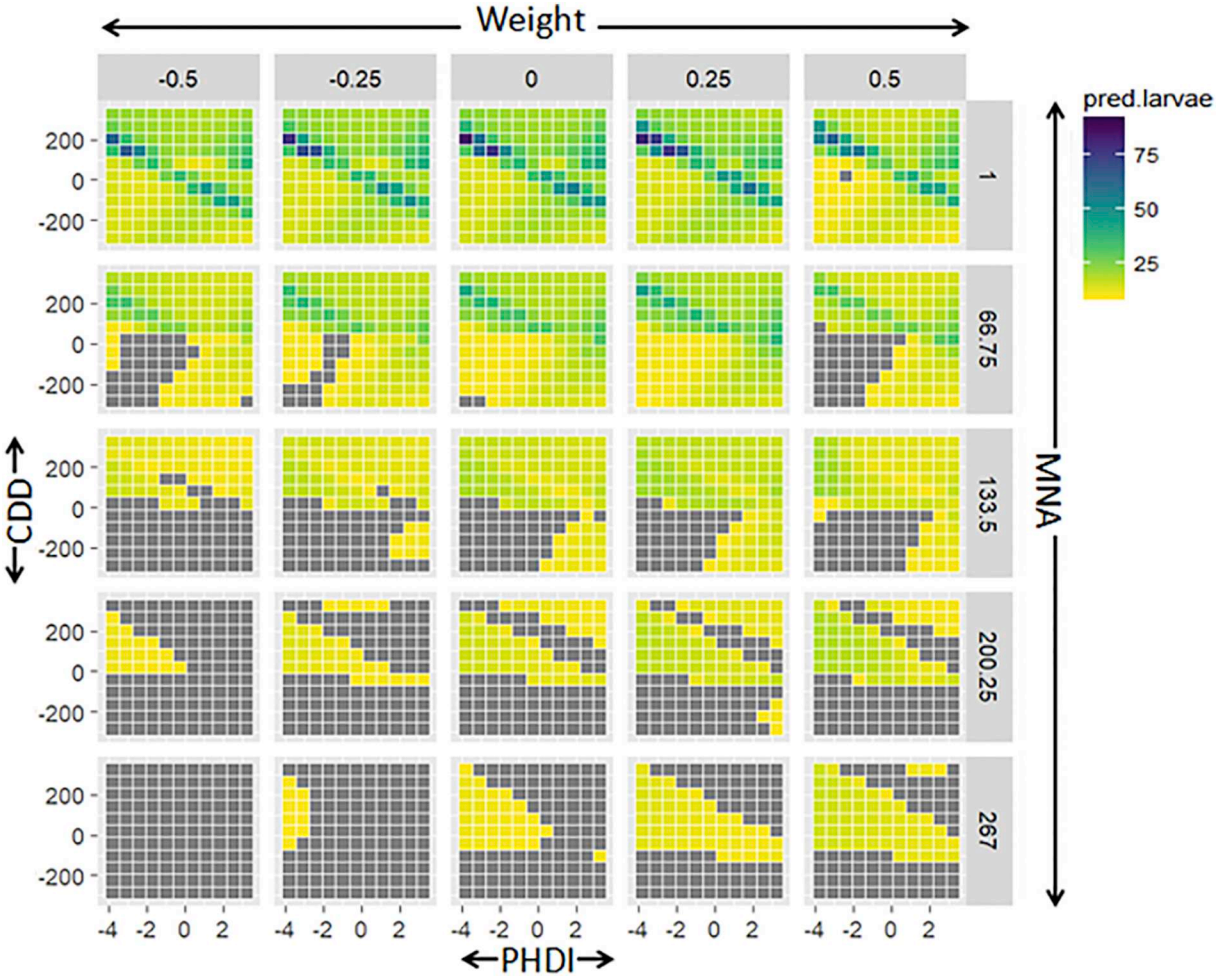

Fig. 2. Heat map of high larval burden (> 10 larvae/host) on adult, male white-footed mice during the large larval peak (averaged over days 244–260). Large x-axis displays Δweight from the average weight within subgroup at the 0, 25th, 50th, 75th and 100th percentiles. Small x-axis displays PHDI at the 0, 10th, 20th, …, 100th percentiles ranging from −4 to 4 (note: full PHDI scale ranges from −10 to 10). Large y-axis displays MNA of white-footed mice at the 0, 25th, 50th, 75th and 100th percentiles, ranging from 1 individual to 267. Small y-axis displays ΔCDD from the 20 year average (from 1995 to 2015) at the 0, 10th, 20th, …, 100th percentiles ranging from 333° below average to 333° above average. Only conditions where model predicts high larval burden (> 10 larvae) are displayed.

Table 3
Results from the sensitivity analysis calculated as, Effect = the average difference between function maximum and the function minimum when varying each variable from −0.5 to 0.5 for continuous variables or 0 or 1 for binary variables over *k* models representing discrete combinations of other variables (CDD = cumulative degree days, MNA = minimum number of hosts alive, PHDI = Palmer Hydrological Drought index).

Variable	Effect
CDD	17.62
MNA	12.41
PHDI	8.43
ΔWeight	3.54
Sex	2.11
Sub-adult	1.77
Lactation[a]	1.77
Pregnancy[a]	0.63
Adult	0.54

[a] Indicates that variable was only considered during the early small larval peak.

accuracy, especially among extreme values indicates that our model was either missing important attributes affecting larval aggregation, such as an individual's ability to groom or its physiological status, or, perhaps offers partial support for the conclusion that tick burden is largely caused by simple bad luck (Calabrese et al., 2011), described as inhabiting a home range with many aggregated larvae.

Our model predicted highest larval burden on non-juvenile males of at least average weight, when host abundance is very low, during moderately warm and dry years. Furthermore, our ANN detected similar patterns when predicting > 10 larvae on a single host, among all subgroups. When modeling larval burden at the main peak, our ANN nearly always predicts > 10 larvae per host, regardless of the host's age, sex or weight, or the climate conditions if host density is low. As host density increases, however, our ANN predicts > 10 larvae on a single host only in warmer than average years, regardless of host weight. When host density is very high, the model predicts > 10 larvae per host only when the climate is warm and dry, with a drier climate required for the model to predict > 10 larvae on smaller hosts.

Intraspecific dilution provides a reasonable mechanism for this finding in that the higher the host density, the smaller the probability that any larva will feed on a particular individual. This explanation suggests that a warmer, drier climate always predicts high larval burden, but the effect is masked when average larval burden is diluted due to a large host population. Similarly, PHDI consistently had a larger

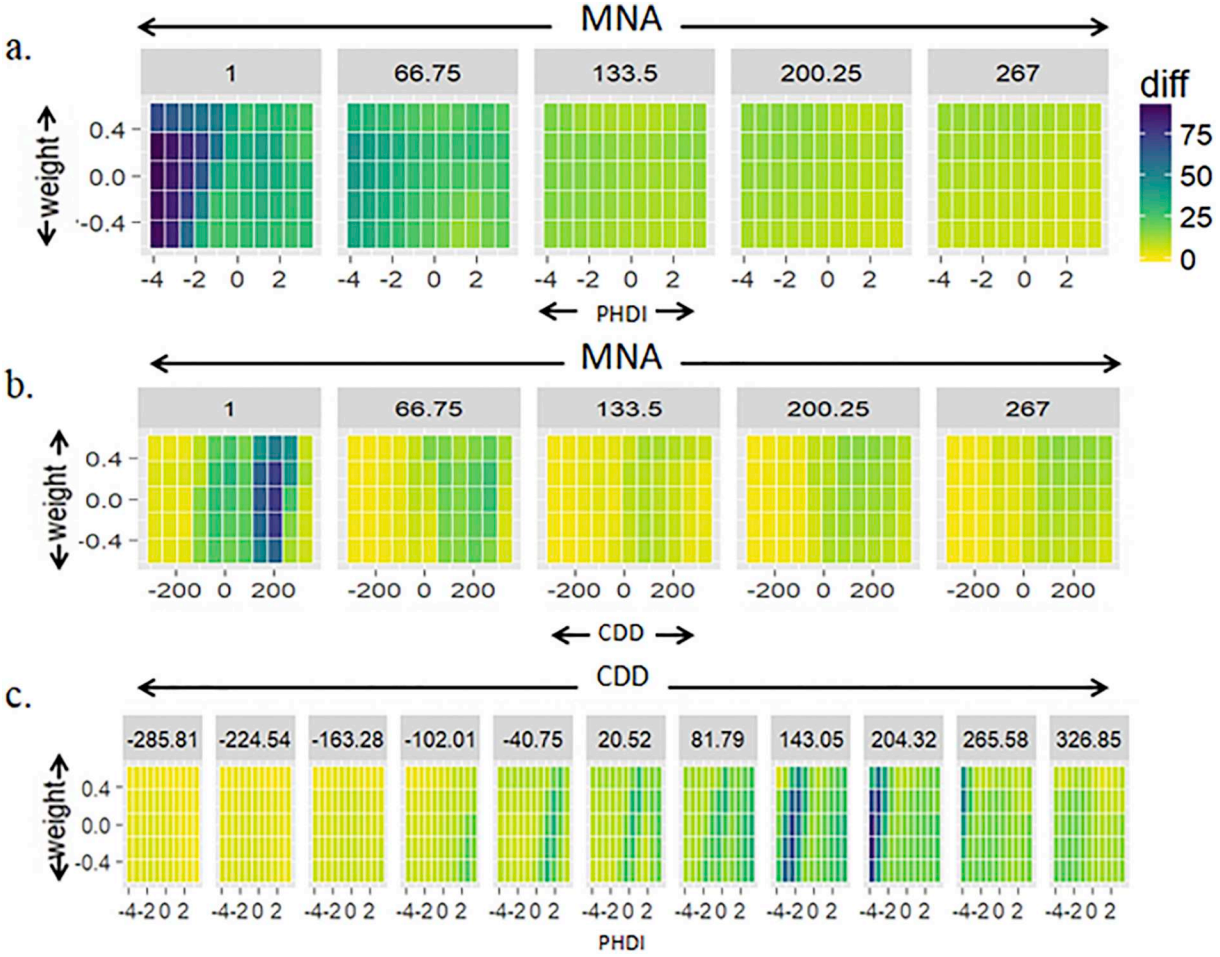

Fig. 3. Heat maps of each continuous variable's effect, defined as the difference between the function maximum and the function minimum when varying each variable from −0.5 to 0.5, as a function of other continuous variables. a. Effect of ΔCDD as a function of MNA, PHDI and Δweight. The large x-axis displays MNA of white-footed mice at the 0, 25th, 50th, 75th and 100th percentiles, ranging from 1 individual to 267. Small x-axis displays PHDI at the 0, 10th, 20th, …, 100th percentiles ranging from −4 to 4 (note: full PHDI scale ranges from −10 to 10). Y-axis displays Δweight from the average weight within the subgroup at the 0, 25th, 50th, 75th and 100th percentiles. b. Effect of PHDI as a function of MNA, ΔCDD, and Δweight. The large x-axis displays MNA of white-footed mice at the 0, 25th, 50th, 75th and 100th percentiles, ranging from 1 individual to 267. Small x-axis displays ΔCDD from the 20 year average (from 1995 to 2015) at the 0, 10th, 20th, …, 100th percentiles ranging from 333° below average to 333° above average. Y-axis displays Δweight from the average weight within the subgroup at the 0, 25th, 50th, 75th and 100th percentiles. c. Effect of MNA as a function of ΔCDD, PHDI and Δweight. The large x-axis displays ΔCDD from the 20 year average (from 1995 to 2015) at the 0, 10th, 20th, …, 100th percentiles ranging from 333° below average to 333° above average. The small x-axis displays PHDI at the 0, 10th, 20th, …, 100th percentiles ranging from −4 to 4 (note: full PHDI scale ranges from −10 to 10). Y-axis displays Δweight from the average weight within the subgroup at the 0, 25th, 50th, 75th and 100th percentiles.

effect on larval burden when CDD was above average, suggesting the existence of a temperature threshold, below which larval burden is so low that the effect of drought becomes masked. According to our sensitivity analysis, climate attributes, on average, affect larval burden by 8–18 larvae, host density affects larval burden by, on average, 12 larvae, while individual-attributes such as sex, age and mass affect larval burden, on average, by fewer than 4 larvae.

Indeed, these findings lend further support to the conclusion previously reached that individual-specific traits do not account for enough variation in burdens to warrant special focus (Brunner and Ostfeld, 2008). Instead, our results suggest that to increase the likelihood of treating individual white-footed mice with high larval burdens, tick control treatments, such as biological or chemical acaricides, should be applied based on climatic attributes, specifically in warmer than average and slightly dry years. Controlling mouse abundance, rather than tick burdens on mice, could potentially also reduce opportunities by ticks to feed on this host species. However, our results suggest that reduced mouse abundance would increase the per-mouse larval burden, potentially offsetting the efficacy of mouse control. In addition, strong density-dependent reproduction by white-footed mice could render mouse-control short-lived (Wang et al., 2008). Larval blacklegged ticks feed on many other mammal and bird hosts; therefore, our modeling approach could be applied to tick burdens on those hosts as well. However, none of these other hosts is as permissive to blacklegged tick feeding or as efficient a reservoir for *B. burgdorfori s.l.* infection as are mice (Keesing et al., 2009).

In recent years, the earth's global mean temperature has consistently been above historic averages (Reidmiller et al., 2018; Stocker et al., 2018); given projected climate change, warmer than average years, conducive to high larval burden on white-footed mice according to our model, will be more and more common in much of the Lyme endemic zone in the coming decades. In general, the consequences of global climate change on infectious disease, and more specifically, tick-borne disease, are not well understood (Ostfeld and Brunner, 2015). This is in part due to the need for long term studies and additional empirical evidence (Eisen, 2008) and in part due to the overwhelming number of components and interactions involved in the biology of vectors, hosts and pathogens (Ogden and Lindsay, 2016; Ostfeld and Brunner, 2015).

Warm conditions can enhance tick host-seeking, accelerate development, increase or decrease survival, and potentially affect natural enemies of ticks (Ogden et al., 2004; Dautel et al., 2008; Tagliapietra et al., 2011; Burtis et al., 2016). Because drier conditions are not expected to directly increase tick abundance or host-seeking behavior, we suspect that the higher tick burdens associated with drier conditions arise from changes in mouse behavior, although further studies are warranted. Warmer, wetter conditions associated with anthropogenic climate change are expected to generally increase tick abundance locally and permit range expansion, increasing tick-borne disease risk (Bennet et al., 2006; Materna et al., 2008; Ostfeld and Brunner, 2015). Our results suggest that warmer, drier conditions could increase tick burdens on mice potentially resulting in increased tick survival (Keesing et al., 2009) and increased rates of pathogen acquisition (Hersh et al., 2012).

Finally, our study demonstrates the utility of ANNs in modeling tick aggregation and parasite aggregation more generally. Our methodology presents novel approaches towards interpreting ANNs, which have often been described as “black boxes” due to the difficulty of interpreting relationships between input variables and model output and the consequential difficulty of gaining insight from the model into the underlying mechanisms governing its behavior (Anderson, 1995). The variable interactions uncovered by our ANN could be included in future models that require specific interaction terms to be defined a priori. The current lack of consensus regarding the utility of various attributes in predicting larval burden may result from overlooking these interactions. Overall, continued application of ANNs towards the question of parasite burden, paired with methods to interpret model output, will allow for detection of patterns not discernible in traditional statistical models, and further our understanding of the many, complex mechanisms that ultimately affect pathogen transmission.

Ethics statement

Declarations of interest: none. The CIES Institutional Animal Care and Use Committee approved the protocols for live-capturing small mammals.

Acknowledgements

We are grateful to Kelly Oggenfuss and the many field assistants who collected these data over > 20 years. This work was supported by several grants from the Long Term Research in Environmental Biology (LTREB) Program at the National Science Foundation and the Steven and Alexandra Cohen Foundation. This is a contribution to the program of Cary Institute of Ecosystem Studies.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecoinf.2019.04.002.

References

Anderson, J.A., 1995. An Introduction to Neural Networks. MIT, Cambridge, MA, pp. 650.
Anderson, R.M., May, R.M., 1978. Regulation and stability of host-parasite population interactions: I. Regulatory processes. J. Anim. Ecol. 47 (1), 219.
Bennet, L., Halling, A., Berglund, J., 2006. Increased incidence of Lyme borreliosis in southern Sweden following mild winters and during warm, humid summers. Eur. J. Clin. Microbiol. Infect. Dis. 25 (7), 426–432.
Bishop, C., 1995. Neural Networks for Pattern Recognition. Oxford University Press, New York, NY, pp. 132.
Bishop, C. (1996). Neural Computing Research Group. Neural Networks: Pattern Recognition Perspective. New York: Oxford University Press. Retrieved from https://www.microsoft.com/en-us/research/wpcontent/uploads/1996/01/neural_networks_pattern_recognition.pdf
Brosse, S., Guegan, J., Tourenq, J., Lek, S., 1999. The use of artificial neural networks to assess fish abundance and spatial occupancy in the littoral zone of a mesotrophic lake. Ecol. Model. 120 (2–3), 299–311.
Brunner, J., Ostfeld, R., 2008. Multiple causes of variable tick burdens on small-mammal hosts. Ecology 89 (8), 2259–2272.
Burtis, J., Sullivan, P., Levi, T., Oggenfuss, K., Fahey, T., Ostfeld, R., 2016. The impact of temperature and precipitation on blacklegged tick activity and Lyme disease incidence in endemic and emerging regions. Parasit. Vectors 9 (1), 606.
Calabrese, J., Brunner, J., Ostfeld, R., 2011. Partitioning the aggregation of parasites on hosts into intrinsic and extrinsic components via an extended Poisson-Gamma mixture model. PLoS One 6 (12), e29215–e29231.
Chang, L., 2005. Analysis of freeway accident frequencies: Negative binomial regression versus artificial neural network. Saf. Sci. 43 (8), 541–557.
Dallas, T., Foré, S., Kim, H., 2012. Modeling the influence of *Peromyscus leucopus* body mass, sex, and habitat on immature *Dermacentor variabilis* burden. J. Vector Ecol. 37 (2), 338–341.
Dautel, H., Dippel, C., Kämmer, D., Werkhausen, A., Kahl, O., 2008. Winter activity of *Ixodes ricinus* in a Berlin forest. Int. J. Med. Microbiol. 298 (1), 50–54.
Devevey, G., Brisson, D., 2012. The effect of spatial heterogenity on the aggregation of ticks on white-footed mice. Parasitology 139 (07), 915–925.
Eisen, L., 2008. Climate change and tick-borne diseases: a research field in need of long-termempirical field studies. Int. J. Med. Microbiol. 298 (1), 12–18.
Fritsch, S., Guenther, F., Suling, M. & Mueller, S. (2018). [online] Cran.r-project.org Available at: https://cran.r-project.org/web/packages/neuralnet/neuralnet.pdf. Accessed November 5, 2018.
Heaton, J., 2005. Introduction to Neural Networks With Java. Heaton Research, Chesterfield, MO, pp. 158.
Hersh, M.H., Tibbets, M., Strauss, M., Ostfeld, R., Keesing, F., 2012. Reservoir competence of wildlife host species for *Babesia microti*. Emerg. Infect. Dis. 18 (2), 1951–1957.
Hilborn, R., Redfield, J.A., Krebs, C.J., 1976. On the reliability of enumeration for mark and recapture census of voles. Can. J. Zool. 43 (3), 1019–1024.
Jorgensen, S.E., 1999. State-of-the-art of ecological modelling with emphasis on development of structural dynamic models. Ecol. Model. 120 (3), 75–96.
Keesing, F., Brunner, J., Duerr, S., Killilea, M., LoGiudice, K., Schmidt, K., et al., 2009. Hosts as ecological traps for the vector of Lyme disease. Proc. R. Soc. B Biol. Sci. 276 (1675), 3911–3919.
Kiffner, C., Vor, T., Hagedorn, P., Niedrig, M., Rühe, F., 2010. Factors affecting patterns of tick parasitism on forest rodents in tick-borne encephalitis risk areas, Germany. Parasitol. Res. 108 (2), 323–335.
Lane, R.S., Piesman, J., Burgdorfer, W., 1991. Lyme borreliosis: relation of its causative agent to its vectors and hosts in North America and Europe. Annu. Rev. Entomol. 36 (1), 587–609.
Lek, S., Delacoste, M., Baran, P., Dimopoulos, I., Lauga, J., Aulagnier, S., 1996. Application of neural networks to modelling nonlinear relationships in ecology. Ecol. Model. 90 (2), 39–52.
Levi, T., Keesing, F., Oggenfuss, K., Ostfeld, R., 2015. Accelerated phenology of blacklegged ticks under climate warming. Philos. Trans. R. Soc.B Biol. Sci. 370 (1665), 2013–2016.
Levi, T., Keesing, F., Holt, R., Barfield, M., Ostfeld, R., 2016. Quantifying dilution and amplification in a community of hosts for tick-borne pathogens. Ecol. Appl. 26 (2), 484–498.
LoGiudice, K., Ostfeld, R., Schmidt, K., Keesing, F., 2003. The ecology of infectious disease: effects of host diversity and community composition on Lyme disease risk. Proc. Natl. Acad. Sci. 100 (2), 567–571.
Lutermann, H., Medger, K., Horak, I., 2011. Abiotic and biotic determinants of tick burdens in the eastern rock sengi (*Elephantulus myurus*). Med. Vet. Entomol. 26 (3), 255–262.
Materna, J., Daniel, M., Metelka, L., Harčarik, J., 2008. The vertical distribution, density and the development of the tick *Ixodes ricinus* in mountain areas influenced by climate changes (The Krkonoše Mts., Czech Republic). Int. J. Med. Microbiol. 298 (1), 25–37.
Monteil, C., Deconchat, M., Balent, G., 2005. Simple neural network reveals unexpected patterns of bird species richness in forest fragments. Landsc. Ecol. 20 (5), 513–527.
NOAA. (2018). [online] Available at: https://www7.ncdc.noaa.gov/CDO/CDODivisionalSelect.jsp. Accessed Nov 20, 2018.
Ogden, N., Lindsay, L., 2016. Effects of climate and climate change on vectors and vector-borne diseases: ticks are different. Trends Parasitol. 32 (8), 646–656.
Ogden, N., Lindsay, L., Beauchamp, G., Charron, D., Maarouf, A., O'Callaghan, C., et al., 2004. Investigation of relationships between temperature and developmental rates of tick *Ixodes scapularis* (*Acari: Ixodidae*) in the laboratory and field. J. Med. Entomol. 41 (4), 622–633.
Olden, J., Jackson, D., 2002. Illuminating the “black box”: a randomization approach for understanding variable contributions in artificial neural networks. Ecol. Model. 154 (1–2), 135–150.
Ostfeld, R., Brunner, J., 2015. Climate change and *Ixodes* tick-borne diseases of humans. Philos. Trans. B 370 (1665).
Ostfeld, R., Miller, M., Hazler, K., 1996. Causes and consequences of tick (*Ixodes scapularis*) burdens on white-footed mice (*Peromyscus leucopus*). J. Mammal. 77 (1), 266–273.
Panchal, G., Ganatra, A., Kosta, Y., Panchal, D., 2010. Searching most efficient neural network architecture using Akaike's Information Criterion (AIC). Int. J. Comput. Appl. 1 (5), 54–57.
R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistica Computing, Vienna, Austria. http://www.R-project.org/
Reidmiller, D.R., Avery, C.W., Easterlink, D.R., Kunkel, K.E., Lewis, K.L.M., Maycock, T.K., Stewart, B.C., 2018. Impacts, Risks, and Adaptation in the United States: Fourth National Climate Assessment, Volume. U.S. Global Change Research Program, Washington, DC, USA.
Sackett, C., 2018. Does the host matter? Variable influence of host traits on parasitism

rates. International Journal for Parasitology 48 (1), 27–39.
Schmidt, K., Ostfeld, R., Schauber, E., 1999. Infestation of *Peromyscus leucopus* and Tamiasstriatus by *Ixodes scapularis* (*Acari: Ixodidae*) in relation to the abundance of hosts and parasites. J. Med. Entomol. 36 (1), 749–757.
Shaw, D., Grenfell, B., Dobson, A., 1998. Patterns of macroparasite aggregation in wildlife host populations. Parasitology 117 (6), 597–610.
Spielman, A., Wilson, M., Levine, J., Piesman, J., 1985. Ecology of *Ixodes Dammini* borne human babesiosis and lyme disease. Annu. Rev. Entomol. 30 (1), 439–460.
Stocker, T.F., Qin, D., Plattner, G.K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgely, P.M., 2018. IPCC, 2013: Climate Change 2013: The Physical Science Basis, Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
Tagliapietra, V., Rosà, R., Arnoldi, D., Cagnacci, F., Capelli, G., Montarsi, F., et al., 2011. Saturation deficit and deer density affect questing activity and local abundance of *Ixodes ricinus* (*Acari, Ixodidae*) in Italy. Vet. Parasitol. 183 (1–2), 114–124.
Twomey, J., Smith, A., 1996. Verification and validation. In: Artificial Neural Networks for Civil Engineers: Fundamentals and Applications. American Society of Civil Engineers, New York, NY, pp. 12.
Wang, G., Wolff, J., Vessey, S., Slade, N., Witham, J., Merritt, J., et al., 2008. Comparative population dynamics of *Peromyscus leucopus* in North America: influences of climate, food, and density dependence. Popul. Ecol. 51 (1), 133–142.
Wickham, H., Chang, W., Henry, L., Penderson, T., Takahashi, K. & Wilke, C. (2018). Create Elegant Data Visualisations Using the Grammar of Graphics. [online] Available at: https://ggplot2.tidyverse.org. Accessed December 4, 2018.
Woolhouse, M., Dye, C., Etard, J., Smith, T., Charlwood, J., Garnett, G., Hagan, P., Hii, J., Ndhlovu, P., Quinnell, R., Watts, C., Chandiwana, S., Anderson, R., 1997. Heterogeneities in the transmission of infectious agents: Implications for the design of control programs. Proc. Natl. Acad. Sci. 94 (1), 338–342.