Modelling spatial concordance between Rocky Mountain spotted fever disease incidence and habitat probability of its vector *Dermacentor variabilis* (American dog tick)

Samuel F. Atkinson[1], Sahotra Sarkar[2], Aldo Aviña[3], Jim A. Schuermann[4], Phillip Williamson[5]

[1]*Institute of Applied Science, Department of Biological Sciences, University of North Texas, 1155 Union Circle #310559, Denton, TX 76203, USA;* [2]*Integrative Biology and Philosophy, University of Texas, 1 University Station, C3500, Austin, TX 78712, USA;* [3]*Department of Environmental and Occupational Health, School of Public Health, University of North Texas Health Science Center, 3500 Camp Bowie Boulevard, Fort Worth, TX 76107, USA;* [4]*Infectious Disease Control Unit, Texas Department of State Health Services, PO Box 149347, Austin, TX 78714, USA;* [5]*Forensic and Investigative Genetics, Graduate School of Biomedical Sciences, University of North Texas Health Science Center, 3500 Camp Bowie Boulevard, Fort Worth, TX 76107, USA*

Abstract. The spatial distribution of *Dermacentor variabilis*, the most commonly identified vector of the bacterium *Rickettsia rickettsii* which causes Rocky Mountain spotted fever (RMSF) in humans, and the spatial distribution of RMSF, have not been previously studied in the south central United States of America, particularly in Texas. From an epidemiological perspective, one would tend to hypothesise that there would be a high degree of spatial concordance between the habitat suitability for the tick and the incidence of the disease. Both maximum-entropy modelling of the tick's habitat suitability and spatially adaptive filters modelling of the human incidence of RMSF disease provide reliable portrayals of the spatial distributions of these phenomenons. Even though rates of human cases of RMSF in Texas and rates of *Dermacentor* ticks infected with *Rickettsia* bacteria are both relatively low in Texas, the best data currently available allows a preliminary indication that the assumption of high levels of spatial concordance would not be correct in Texas (Kappa coefficient of agreement = 0.17). It will take substantially more data to provide conclusive findings, and to understand the results reported here, but this study provides an approach to begin understanding the discrepancy.

Keywords: Rocky Mountain spotted fever, *Dermacentor variabilis*, *Rickettsia rickettsii*, spatial distributions, concordance modelling and mapping, USA.

Introduction

RMSF is a potentially fatal human illness, and has been reportable since the 1920s. It is a life-threatening disease caused by *Rickettsia rickettsii*, an intracellular bacterium that is spread to human beings by infected ticks. RMSF is the most commonly reported tick-borne rickettsial disease in the United States of America (USA) (Dantas-Torres, 2007), is among the most virulent infections identified in human beings (Dumler and Walker, 2005) and it is the most common fatal tick-borne disease in the USA (Dalton et al., 1995). The United States Centers for Disease Control and Prevention (CDC) reports that the incidence of

RMSF has increased during the last decade, from less than 2 cases per million persons in 2000 (approximately 500 cases per year at the beginning of the decade) to over 8 cases per million persons in 2008, the highest rate ever reported, with a total of approximately 1,500 cases reported in 2010 (CDC, 2011). Fortunately, the proportion of cases that result in death has declined from a high of nearly 28% in the mid-1940s to less than 0.5% in 2008. The distribution of RMSF spans the continental USA, with 46 states reporting cases to the CDC since the 1960s through national surveillance (Dalton et al., 1995). These national surveillance data have shown that the disease is not equally distributed among all reporting states, with incidence highest in the southeast and south-central states. Between the years 2000 and 2010 there were 266 reported cases of RMSF in Texas (Fig. 1). The American dog tick, *Dermacentor variabilis*, is the most commonly identified tick species (but not the only species) responsible for transmitting *R. rickettsii* to humans (Parola et al., 2009). *Dermacentor* spp.

Corresponding author:
Samuel F. Atkinson
Institute of Applied Science, Department of Biological Sciences
University of North Texas, 1155 Union Circle #310559
Denton, TX 76203, USA
Tel. +1 940 565 2694; Fax +1 940 565 4297
E-mail: atkinson@unt.edu
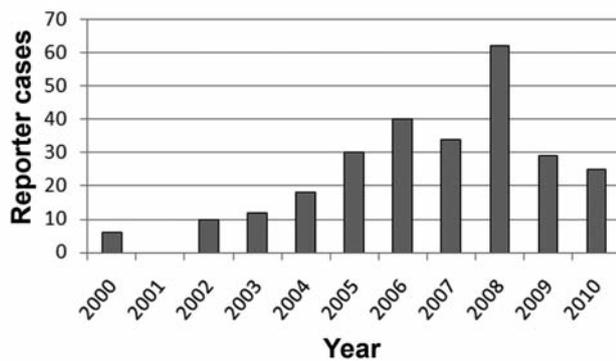
Fig. 1. Trend of RMSF in Texas, 2000-2010.

have a three-stage life cycle (Sonenshine, 1993; Dharmarajan et al., 2010), involving larvae and nymphs which predominantly parasitise small mammals (e.g. mice) and adults which prefer larger mammals (e.g. raccoons). After mating on the third host, females engorge and drop off to deposit eggs in the environment (Sonenshine, 1991). Reproduction in *Dermacentor* is predominantly sexual and while parthenogenesis has been reported, the survival of parthenogenetic individuals is usually low (Oliver, 1971). Because of its importance as a vector of numerous pathogens such as *Rickettsia rickettsii,* (Jongejan and Uilenberg, 2004)*, Dermacentor* has been studied in terms of its spatial distribution (Stein et al., 2008), life cycle (Troughton and Levin, 2007), host preferences (Kollars et al., 2000) and demography (Monello and Gompper, 2007). However, studies examining the spatial concordance between its habitat suitability and the incidence of RMSF remain limited. A better understanding of this relationship should illuminate the decisions concerning prevention and treatment of the disease.

We have recently reported (Williamson et al., 2010) the results of a study where a total of 903 ticks, representing 11 tick species, which had been submitted to the University of North Texas Health Science Center's (UNTHSC) Tick-Borne Disease Research Laboratory for analysis of both tick identification and rates of *Rickettsia* infection. These ticks had been attached to, and removed from humans, and were subjected to molecular analysis for detection of *Rickettsia* spp. bacteria. A total of 6.5% of the 355 *Dermacentor* ticks in that study were infected with *Rickettsia* bacteria.

Because neither tick presence nor RMSF incidence are evenly spread across Texas, we became interested in understanding the spatial concordance between the two types of data. The study reported here developed and compared two maps of Texas related to RMSF.

The first map represents the suitability of environmental conditions for *Dermacentor* ticks (habitat suitability mapping), while the second map represents the incidence rate of RMSF across the state (disease incidence mapping). From an epidemiological perspective, one would tend to hypothesise that there would be a high degree of spatial concordance between the habitat suitability and the disease incidence maps. Even though rates of human cases of RMSF in Texas and rates of infected ticks in Texas are both relatively low, the best data currently available indicates that the assumption of high levels of spatial concordance would not be correct in Texas. Because the results presented here must be considered as preliminary, it will take substantially more data to provide conclusive findings.

Materials and methods

This research involved mapping the probability of occurrence that the disease vector exists in the environment (i.e. tick habitat suitability mapping), mapping the incidence of RMSF in the human population (disease incidence mapping) and examining the spatial concordance between the suitability map and incidence map.

Habitat probability mapping

Because *Dermacentor* ticks are the most commonly associated vector of the bacterium *Rickettsia*, the presence of *Dermacentor* increases the risk contracting RMSF (Dalton et al., 1995). However, public health efforts to determine the spatial distribution of tick populations in Texas (and in many other parts of the world) are limited due to insufficient data collection (McDade and Newhouse, 1986). Fortunately, ticks have certain environmental requirements for survival (Monello and Gompper, 2007; Stein et al., 2008), so the potential distribution of tick populations can be estimated with species distribution models that allow mapping habitat suitability (that is the probability of occurrence) based on presence-only occurrence data and the spatial distribution of environmental variables.

Tick habitat suitability maps can be thought of as models of the probability of occurrence of ticks (dependent variable) based on a series of environmental covariates (independent variables). Following a well-established risk assessment protocol in the epidemiological literature (González et al., 2010; Sarkar et al., 2010), these maps can then be used as a surro-

gate, or indicator, of the risk of encountering infected ticks. Previous literature (Monello and Gompper, 2007) suggests that the environmental covariates most indicative of *Dermacentor* ticks are: maximum, minimum and mean temperature; relative humidity; vapor pressure; soil types; and vegetation density.

The study reported here utilized most of the environmental covariants mentioned (vapor pressure was not available, and soil type was investigated, but it had no influence on resultant map so it was removed from the model). The environmental covariants, in conjunction with tick occurrence data, were modelled using a maximum-entropy algorithm incorporated in the MaxEnt software package. MaxEnt works well with limited, presence-only datasets, and quickly processes data to create probability distributions based on environmental parameters and species occurrence (Phillips et al., 2006; Phillips and Dudík, 2008)

Among the several alternative protocols available for species distribution mapping (Franklin and Miller, 2009), MaxEnt was chosen for this study because a recent comprehensive comparative assessment found its predictive performance to be consistently among the highest performing methods (Elith et al., 2006). MaxEnt estimates a distribution across geographical space (Phillips et al., 2006; Phillips and Dudík, 2008) or, equivalently, probability densities in covariate space (Elith et al., 2006).

MaxEnt requires as input the geographical coordinates of the locations where the species of interest has been observed (i.e. the presence sites). Next, it examines the data for each environmental covariate at each of those locations (this is termed "conditional density of the covariates at the presence sites"). Then, for a series of randomly selected locations in the study area called background sites (which may or may not contain some of the presence sites), the data for each environmental covariate (called "marginal density of covariates across the study area") are examined. The model computes a metric called the area under the curve (AUC) value, which measures the quality of a ranking of sites' suitability to support the species

under consideration (Fielding and Bell, 1997). The AUC is the probability that a randomly chosen presence site will be ranked above a randomly chosen absence site. A random ranking has on average an AUC of 0.5, and a perfect ranking achieves the best possible AUC of 1.0. The most comprehensive reviews to date (Elith et al., 2006; Pawar et al., 2007) accepts a value of 0.7 as indicting the model is predictively adequate.

The tick occurrence data were generated between 2004 and 2010 from submissions to the Texas Department of State Health Services (TDSHS). Testing of TDSHS samples by UTNHSC is an ongoing endeavour and the study reported here not only includes the 903 ticks collected between 2004 and 2008 in our original study (Williamson et al., 2010), but also includes 555 additional tick submissions collected between 2008 and 2010. The occurrence data were aggregated by TDSHS to the zip-code level. Molecular screening for tick-borne pathogens, including *R. rickettsii*, was performed at UNTHSC using the method of Williamson et al. (2010). These data were spatially tagged with the zip code of where the tick was encountered. For modelling purposes, it is unrealistic to assume that for zip codes with multiple tick occurrences, each human-tick encounter occurred at the same location (e.g. the spatial centroid of the zip code area). Because MaxEnt requires point data (a specific longitude/latitude or X/Y point coordinate based on various projection systems) the tick occurrence data were re-assigned an X/Y location by randomly distributing individual tick occurrences within the corresponding zip code. This allowed tick occurrences to be more realistically distributed across the entire zip code, rather than just at its centroid. The occurrence data were randomly distributed across the zip codes 30 times, with each resultant distribution processed by MaxEnt. Those results were averaged to produce our habitat suitability map, and ultimately sampled on a 500 x 500 m grid to match the disease incidence mapping resolution to allow an investigation in the spatial concordance between the two data sets.

Table 1. Data, sources and reasoning for inclusion in modelling.

Data	Source	Reasoning
Elevation	National Elevation Dataset	Determine if elevation is a factor of tick presence in Texas
Temperatures	WorldClim.org	Ticks sensitive to temperature extremes
Relative humidity	PRISM Climate Group, calculated from dew point and ambient temperature.	Tick presence/activity is dependent on water vapor
Land use/land cover	National Land Cover Dataset	Forest cover relates to tick suitability
Soil texture	Soil Survey Geographical Database	Soil texture influences moisture content

The environmental parameters entered into our MaxEnt model are shown in Table 1 (National Elevation Dataset, 2010; WorldClim, 2010). These parameters were chosen based on the studies discussed earlier, which identified several key ecological variables associated with the presence of *Dermacentor* ticks. By mapping the distribution of *Dermacentor* ticks at geographical point scale, something which has not been done across an area the size of Texas before, we have developed a surrogate of RMSF exposure risk based on locations and attendant environmental data associated with tick submissions.

Disease incidence mapping

The second aspect of our study was to develop a map expressing RMSF incidence in Texas. RMSF is a nationally notifiable disease, reported to the U.S. CDC because it is considered a disease for which regular, frequent and timely information regarding individual cases is considered necessary for its prevention and control. When a physician diagnosis a case of RMSF-based on CDC's criteria (clinical, laboratory, healthcare record or death certificate) to determine whether a case should be reported, those data are reported to the appropriate state health department. In Texas, physicians report the disease to the TDSHS, and they are reported at the zip code level. The data we used represent all reported cases of RMSF in Texas between 2000 and 2010. These data were provided by TDSHS, with all human identifiable information removed, so that incidence rates could be calculated. When the data are normalised by population size in each zip code to report incidence rates of RMSF, the data may be susceptible to instability due to the small population sizes in some zip codes – the so called "small numbers problem" (MacEachren et al., 1998; Jerrett et al., 2010). For example, a zip code with a population of 75 people and one reported case of RMSF results in a rate of 1,333 cases per 100,000 people – an extraordinarily high incidence rate. Once these rates are calculated, and a simple map of an incidence rate by zip code is produced, an unrealistic view of the disease may appear. In order to reduce the influence of the small numbers problem, we smoothed the disease count data using kernel density estimation methods. These methods compute rates for overlapping circular areas – spatial filters – centered on a regular grid of points. Maps based on this approach are known as kernel density estimation maps. Using kernel density estimation allows the normalization of incidence rates that would otherwise be over- or under-estimated, cre-

ating a more accurate representation of disease incidence (Rushton et al., 2004; Tiwari and Rushton, 2005). More detailed discussions of spatial filters are available (Brillinger, 1994; Kafadar, 1994), and several illustrations of the method can be found (Lai et al., 2004; Yang et al., 2006). Here, our spatial filters (kernel) were centered on a 10-mile grid superimposed on the state, and grew in size by capturing surrounding zip codes when the centroid of a zip code with a reported case of RMSF fell within the kernel, until a minimum population size of 10,000 of the captured zip codes was reached. This approach results in kernels of varying size based on the underlying population density, thus the approach is referred to as spatially adaptive filters. Larger sized filters occur in sparsely populated areas and smaller filters occur in densely populated areas. Larger filters provide more smoothing, but also result in loss of geographical detail whereas smaller filters provide greater geographical detail. Because the adaptive filter method is based on varying filter sizes that adapt to population density, resultant maps provide high geographical resolution in those areas where such detail is expected while maintaining rate stability in rural areas with sparse populations (Tiwari, 2011).

The number of RMSF cases in all zip codes captured by the spatial filter, and the total population in those zip codes was used to generate a more representative incidence rate across Texas. Because each grid point in the study area now had a calculated incidence rate, incidence rates could be interpolated across Texas using the inverse distance weighted method (Tiwari and Rushton, 2005). Ultimately, the map was resampled into 500 x 500 m grid cells for spatial concordance analysis.

Spatial concordance mapping

The spatial agreement between our habitat probability and disease incidence maps was measured using an overlay mapping technique where categorical data from two input maps generate a third categorical data map for which each output category represents every combination of the two input map categories. The technique was originally developed to detect and quantify land use change that occurs in an area over time (Jensen and Lulla, 1987; Serra et al., 2003). The technique compares the assigned categories from each input map at a given location (i.e. a pixel), and uses a decision matrix such as the one shown in Table 2 to assign an output class for that same location in the output map. Because the technique provides a measure of concordance and

Table 2. Spatial concordance matrix, consisting of **h** habitat and **i** incidence classes for **N** pixels.

Spatial concordance	Habitat probability map class					Row total
	1 (lowest 20%)	2 (2nd quintile)	3 (3rd quintile)	4 (4th quintile)	5 (highest 20%)	
Incidence rate map class						
1 (lowest 20%)	$1_{h1,i1}$	$2_{h2,i1}$	$3_{h3,i1}$	$4_{h4,i1}$	$5_{h5,i1}$	*row 1*
2 (2nd quintile)	$6_{h1,i2}$	$7_{h2,i2}$	$8_{h3,i2}$	$9_{h4,i2}$	$10_{h5,i2}$	*row 2*
3 (3rd quintile)	$11_{h1,i3}$	$12_{h2,i3}$	$13_{h3,i3}$	$14_{h4,i3}$	$15_{h5,i3}$	*row 3*
4 (4th quintile)	$16_{h1,i4}$	$17_{h2,i4}$	$18_{h3,i4}$	$19_{h4,i4}$	$20_{h5,i4}$	*row 4*
5 (highest 20%)	$21_{h1,i5}$	$22_{h2,i5}$	$23_{h3,i5}$	$24_{h4,i5}$	$25_{h5,i5}$	*row 5*
Column total	*col 1*	*col 2*	*col 3*	*col 4*	*col 5*	*N*

also produces a derived map, it offers both visual and quantitative expressions of map agreement.

Here, the categories in the habitat suitability map, represented by columns in the decision matrix were: 1-pixels with the lowest 20% of tick habitat probability, through 5-pixels with the highest 20% of habitat probability, with classes 2, 3 and 4 representing the three middle classes. A similar categorization for the RMSF incidence rate was utilized, where rows in the decision matrix represent: 1-pixels with the lowest 20% incidence rates, through 5 (the highest 20% of incidence rates). These two input categorized maps were overlaid and the resultant output map contained 25 possible classes: $1_{1,1}$-pixels with the lowest 20% incidence rate and the lowest 20% habitat probability; class $2_{1,2}$-pixels with the lowest 20% incidence rate and the second lowest 20% habitat probability; and so on through class $25_{5,5}$-pixels with the highest incidence and the highest habitat probability.

Overall concordance is simply the sum of the count of pixels along the major diagonal of the decision matrix, divided by the total number of pixels analysed. If two maps are in perfect spatial concordance, all pixels would fall on the major diagonal of the decision matrix (for this case, $1_{1,1}$; $7_{2,2}$; $13_{3,3}$; $19_{4,4}$; $25_{5,5}$), resulting in a concordance of 1.0. Perfect spatial concordance is rare, and every non-concordant pixel falls in one of the off-diagonal cells in the decision matrix. The column and row totals around the margin of the decision matrix (referred to as marginals) can be used to compute the concordance for each class. Finally, the Kappa coefficient of agreement, K, provides a measure of the amount of agreement between two maps, and accounts for chance or random agreement between the

two maps (Congalton et al., 1983):

$$K = \frac{N\Sigma x_d - \Sigma(x_r \times x_c)}{N^2 - \Sigma(x_r \times x_c)} \qquad \text{Eq. 1}$$

where N is the total number of pixels analysed, x_d the count on the diagonal for row i and column h, x_r the total count on row i, and x_c = total count on column h.

The conditional error in agreement between the likelihood of a RMSF incidence class correctly falling into the appropriate habitat probability class (e.g. likelihood that RMSF incidence class 1 will fall into *Dermacentor* habitat probability class 1) can be calculated:

$$E_L = 1 - \frac{x_d}{x_c} \qquad \text{Eq. 2}$$

where E_L is the conditional error for *Dermacentor* habitat probability class i falling into RMSF incidence class i.

Likewise, the conditional error in agreement between the likelihood of a *Dermacentor* habitat probability class correctly falling into the appropriate RMSF incidence class (e.g. likelihood that *Dermacentor* habitat probability class 1 will fall into RMSF incidence class 1) can be calculated:

$$E_H = 1 - \frac{x_d}{x_r} \qquad \text{Eq. 3}$$

where E_H is the conditional error for RMSF incidence class i falling into *Dermacentor* habitat probability class i.
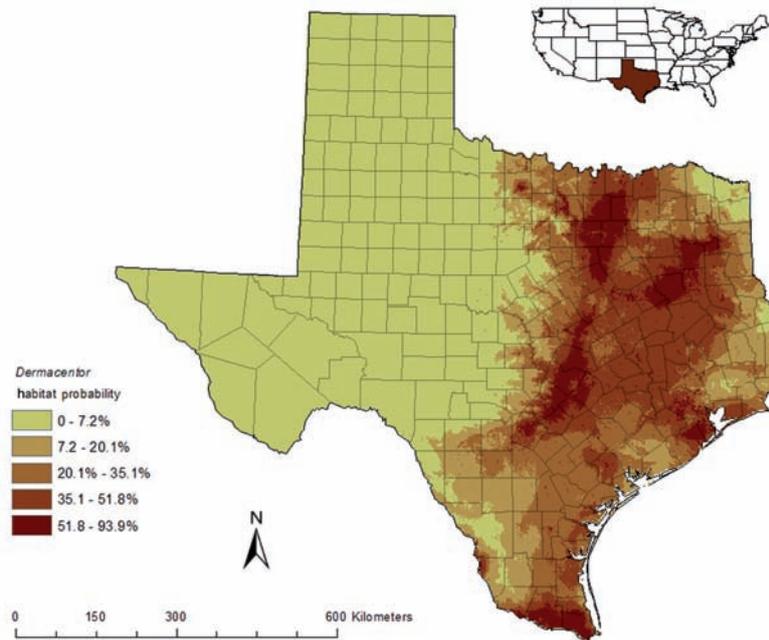
Fig. 2. Estimated distribution of *Dermacentor variabilis* habitat, based on ticks collected between 2004 and 2010 and MaxEnt modelling of habitat probability.

Results

Texas has a land area of approximately 678,000 km² (~262,000 mi²) and a population exceeding 25 million. This overall density (approximately 37 people per km²) is not evenly divided across the state – nearly a third of the state's population lives in just 10 cities. Additionally, the state is divided into a fairly distinct eastern and western climatic regime: the eastern half tends to have more potential precipitation than potential evapotranspiration, while the western half tends to be the opposite. These two characteristics appear to have a substantial influence on the maps generated for this research.

Habitat probability map

Fig. 2 shows the MaxEnt output of the estimated distribution of *Dermacentor* tick habitat, based on environmental characteristics modelled from the collection of ticks between 2004 and 2010. The areas of highest habitat probability occur in eastern and northeastern Texas. The AUC number of the map in Fig. 2 is 0.843, and therefore the habitat probability map for *Dermacentor* ticks can be considered predictively adequate based upon Elith, Graham (Elith et al., 2006) and Pawar, Koo (Pawar et al., 2007) criteria. Table 3 indicates that approximately 95% of Texas land area is predicted to have less than a 50%

probability of having suitable habitat for *Demacentor* ticks. Only 0.3% of Texas has a 75% or greater probability of having suitable habitat, yet Texas incidence rate is almost exactly the nation's median rate for all states.

Table 3. MaxEnt modelling results for *Dermacentor* tick habitat in Texas.

Probability for suitable *Dermacentor* habitat	Percent of Texas' land	Cumulative percent
0% to 5%	50.34%	50.34%
> 5% to 10%	6.41%	56.75%
>10% to 15%	4.28%	61.02%
>15% to 20%	4.25%	65.28%
>20% to 25%	4.83%	70.11%
>25% to 30%	5.15%	75.26%
>30% to 35%	5.19%	80.45%
>35% to 40%	5.39%	85.84%
>40% to 45%	5.16%	90.99%
>45% to 50%	3.13%	94.12%
>50% to 55%	2.10%	96.23%
>55% to 60%	1.62%	97.84%
>60% to 65%	0.95%	98.79%
>65% to 70%	0.43%	99.22%
>70% to 75%	0.23%	99.45%
>75% to 80%	0.27%	99.72%
>80% to 85%	0.18%	99.90%
>85% to 90%	0.08%	99.98%
>90% to 95%	0.02%	100.00%
>95% to 100%	0.00%	100.00%
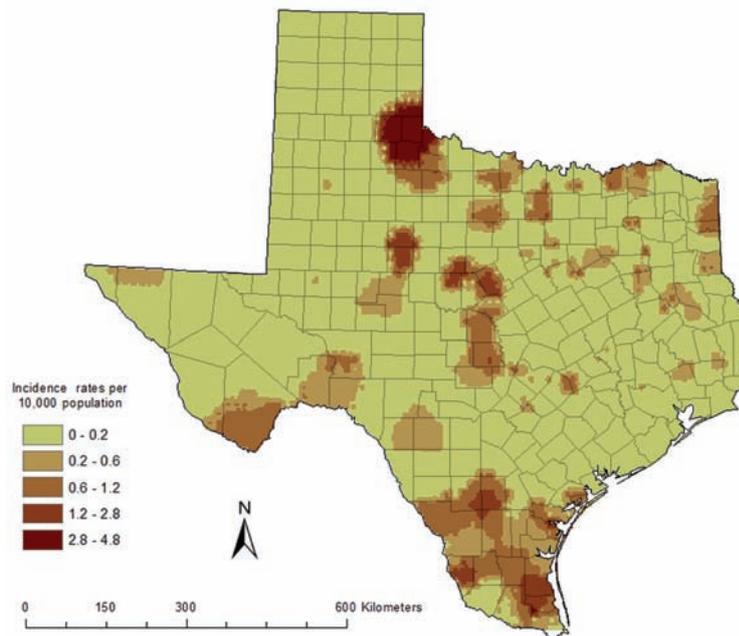
Fig. 3. Rates of RMSF per 10,000 population, based on cumulative cases reported from 2000-2010.

RMSF incidence map

Over the past decade the incidence of RMSF across Texas ranges from zero to 4.77 cases per 10,000 people between 2000 and 2010. Fig. 3 presents the map of RMSF incidence rates in Texas. The highest burdens occur primarily in the western half of Texas, away from both the most densely populated cities and areas of high tick habitat probability. The highest incidence rate during the 2000s occurred in a rural area in the Texas panhandle, about mid-way between Amarillo and Wichita Falls.

There are a few other areas in Texas with relatively high incidence rates, for example on the northern border with Oklahoma and the southern border with Mexico, but all of these areas are removed from the most densely populated and large cities in Texas. Approximately 96% of the states area experienced a rate of less than 1 case per 10,000 since 2000 (Table 4). Only 0.6% of the state's area experienced a rate higher than 4 cases per 10,000.

Concordance map

Fig. 4 shows the result of overlay mapping between *Dermacentor* tick habitat probability mapping and RMSF rates, while Table 5 summarises the overall and class-specific concordance between these two maps. Fig. 4 shows (in black) only areas where there was

Table 4. Spatially adaptive filters model of RMSFincidence in Texas, 2000 to 2010.

2000 to 2010 incidence rate per 10,000 population	Percent of Texas' land	Cumulative percent
0 to 0.5	87.17%	87.17%
>0.5 to 1.0	8.90%	96.07%
>1 to 1.5	1.64%	97.71%
>1.5 to 2.0	0.90%	98.61%
>2.0 to 2.5	0.24%	98.85%
>2.5 to 3.0	0.13%	98.98%
>3.0 to 3.5	0.16%	99.15%
>3.5 to 4.0	0.26%	99.41%
>4.0 to 4.5	0.40%	99.81%
>4.5 to 5.0	0.19%	100.00%

concordance between the top 20% of the state with the highest tick habitat suitability and the top 20% of the state with the highest RMSF incidence. While 20% of the states area is approximately 135,600 km² (~52,300 mi²), only 20,499 km² (~7,900 mi²), or 3% of the state has both high incidence of RMSF and high habitat probability for *Dermacentor* ticks.

The overall concordance of these two maps is only 33.6%, and nearly half of that agreement (14.7%) falls in those areas with both the lowest habitat probability and the lowest incidence rates. The Kappa statistic, which removes chance agreement, indicates that there is only 17% agreement between the two maps
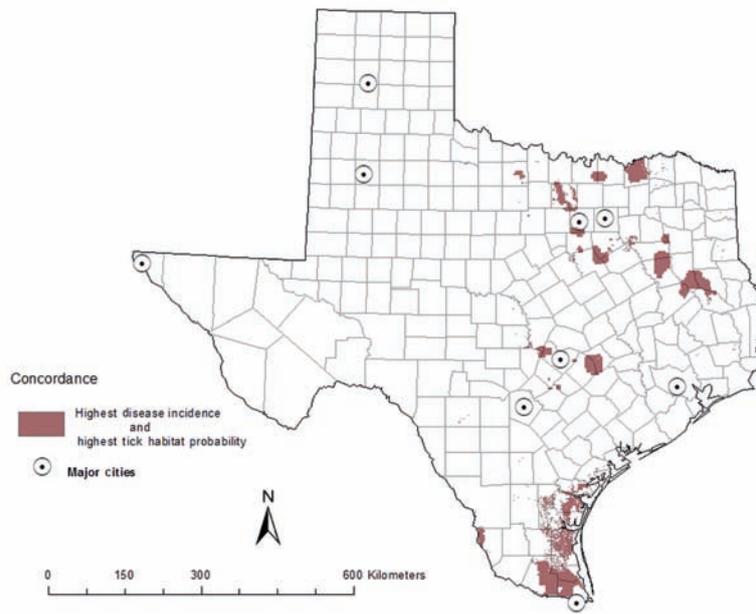
Fig. 4. Areas of spatial concordance (approximately 3.0% of Texas) where the highest 20% of *Dermacentor* habitat probability and the highest 20% of RMSF incidence rates overlap (K = 0.17).

(K = 0.17). Kappa values less than 0.4 is considered to represent poor agreement (Landis and Koch, 1977).

The conditional error in agreement for a RMSF incidence class falling into the appropriate *Dermacentor* tick habitat probability class ranges from a low of 26.9% error (disease incidence class 1 with habitat probability class 1) to a high of 84.7% error (disease incidence class 5 and habitat probability class 5). The conditional error in agreement for a *Dermacentor* tick habitat probability class falling into the appropriate RMSF incidence class ranges from 20.4% error (habitat probability class 1 and disease incidence classes 1)

to a high of 85.0% error (habitat probability class 5 and disease incidence class 5).

Discussion

Currently we are unsure of why some areas exhibit higher RMSF incidence rates than others, nor why there is so little concordance between high *Dermacentor* tick habitat probability and the incidence of the disease. Some can be due to populations' accessibility to hike and bike trails, greenbelts and campgrounds (i.e. areas where people are exposed to ticks). Ideally, an accurate estimation of the level of access and

Table 5. Spatial concordance matrix of RMSF incidence by *Dermacentor* tick habitat probability across Texas. Percentages represent what proportion of Texas' 678,000 km² falls within each pair of map classes.

	Spatial concordance	Habitat probability map class					Row total
		1 (lowest 20%)	2 (2ⁿᵈ quintile)	3 (3ʳᵈ quintile)	4 (4ᵗʰ quintile)	5 (highest 20%)	
Incidence rate map class	1 (lowest 20%)	14.7%	5.3%	0.0%	0.0%	0.0%	20.0%
	2 (2ⁿᵈ quintile)	0.8%	5.9%	3.8%	5.0%	4.7%	20.2%
	3 (3ʳᵈ quintile)	1.1%	1.6%	5.1%	5.1%	6.9%	19.8%
	4 (4ᵗʰ quintile)	1.4%	3.3%	5.0%	4.9%	5.5%	20.1%
	5 (highest 20%)	0.4%	4.9%	6.4%	5.1%	3.0%	19.8%
	Column total	18.4%	21.0%	20.3%	20.1%	20.1%	100.0%

activities that people have within areas of high tick probability could be included in our concordance mapping. Unfortunately, these types of data are rarely collected, and if they are, they are typically only collected for places such as state or national parks, not necessarily the types of areas with highly suitable tick habitat.

Another reason may be that areas that would normally not be suitable for *Dermacentor* ticks according to MaxEnt modelling may actually be suitable due to agricultural and irrigation processes that change the landscape in a way that creates suitable moisture and vegetation regimes for the ticks. These agricultural processes were not incorporated in the tick habitat model. Additionally, RMSF cases may be transported from (or to) Mexico, affecting the disease incidence estimates on the USA-Mexico border region. However, more localized research must be done to accurately assess the origins of RMSF in specific regions.

Additionally, in Texas, the level at which data are reported for individuals infected with RMSF is at the zip code level of the residence of the patient. Even if physicians were instructed to not report the zip code of the patient's home, but rather the location of where they were bitten by the tick, the ability of the patient to accurately know when and where the tick bite occurred may be problematic. Those that have ever been bitten by a tick know that rarely do you know the moment of the bite, but rather much later when they examine their body when changing clothes or bathing.

Our study, an initial investigation based on a relatively small data set, has provided a preliminary model which revealed that areas which might be considered to have an increased risk of encountering *Dermacentor* ticks due to more suitable habitat are located in the eastern portions of Texas, while the pattern of RMSF incidence is more prevalent in central Texas. Reasons for this difference need to be further investigated. As of now, we speculate that it may be due to insufficient tick sampling, people traveling from low tick habitat probability areas to areas with high tick habitat probability where they become exposed to the disease, leading to a potential error in reporting a patient's residence location *versus* where the actual tick encounter occurred, and/or the possibility that *R. rickettsii* is being spread by another species of tick. Further, some anecdotal information suggests that south Texas has a relatively high deer population, and incorporating deer density data into MaxEnt modelling may provide additional insights. The areas that we have shown to have a high degree of concordance for high disease incidence rates and high tick habitat

probability (see Fig. 4) should be targeted for more focused research.

Adding more accurate tick collection data and more precise case location data including the possibility of patient travel information into areas of high tick probability will help in understanding why our model suggests a spatial disagreement between *Dermacentor* distribution and RMSF incidence in Texas. Nevertheless, these results indicate that even with preliminary data, by utilizing advanced mapping techniques many more observations can be made about RMSF in Texas than would have been previously possible. By implementing a spatially adaptive filters analysis to human disease incidence, and utilising MaxEnt habitat probability modelling, it has been possible to provide some initial, unexpected observations about RMSF in Texas.

Acknowledgements

References

Brillinger DR, 1994. Examples of scientific problems and data analyses in demography, neurophysiology, and seismology. J Comput Graph Stat 3, 1-22.

CDC, 2011. Rocky Mountain spotted fever, statistics and epidemiology. Atlanta: Centers for Disease Control and Prevention.

Congalton RG, Oderwald RG, Mead RA, 1983. Assessing Landsat classification accuracy using discrete multivariate analysis statistical techniques. Photogramm Eng Remote Sens 49, 1671-1678.

Dalton MJ, Clarke MJ, Holman RC, Krebs JW, Fishbein DB, Olson JG, Childs JE, 1995. National surveillance for Rocky Mountain spotted fever, 1981-1992: epidemiologic summary and evaluation of risk factors for fatal outcome. Am J Trop Med Hyg 52, 405-413.

Dantas-Torres F, 2007. Rocky Mountain spotted fever. Lancet Infect Dis 7, 724-732.

Dharmarajan G, Beasley JC, Rhodes OE Jr, 2010. Spatial and temporal factors affecting parasite genotypes encountered by hosts: empirical data from American dog ticks (*Dermacentor variabilis*) parasitising raccoons (*Procyon lotor*). Int J Parasitol 40, 787-795.

Dumler JS, Walker DH, 2005. Rocky Mountain spotted fever-changing ecology and persisting virulence. N Engl J Med 353, 551-553.

Elith J, Graham CH, Anderson RP, Dudik M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick JR, Lehmann A, Li

J, Lohmann LG, Loiselle BA, Manion G, Moritz C, Nakamura M, Nakazawa Y, Overton JM, Peterson TA, Phillips SJ, Richardson K, Scachetti-Pereira R, Schapire RE, Soberon J, Williams S, Wisz MS, Zimmermann NE, 2006. Novel methods improve prediction of species distributions from occurrence data. Ecography 29, 129-151.

Fielding AH, Bell JF, 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environ Conserv 24, 38-49.

Franklin J, Miller JA, 2009. Mapping species distributions: spatial inference and prediction. Cambridge University Press Cambridge.

González C, Wang O, Strutz SE, González-Salazar C, Sánchez-Cordero V, Sarkar S, 2010. Climate change and risk of leishmaniasis in North America: predictions from ecological niche models of vector and reservoir species. PLoS Negl Trop Dis 4, e585.

Jensen JR, Lulla DK, 1987. Introductory digital image processing: a remote sensing perspective.

Jerrett M, Gale S, Kontgis C, 2010. Spatial modeling in environmental and public health research. Int J Environ Res Public Health 7, 1302-1329.

Jongejan F, Uilenberg G, 2004. The global importance of ticks. Parasitology 129, S3-S14.

Kafadar K, 1994. Choosing among two-dimensional smoothers in practice. Comput Stat Data Anal 18, 419-439.

Kollars TM Jr, Oliver JH Jr, Masters EJ, Kollars PG, Durden LA, 2000. Host utilization and seasonal occurrence of *Dermacentor*. Exp Appl Acarol 24, 631-43.

Lai PC, Wong CM, Hedley AJ, Lo SV, Leung PY, Kong J, Leung GM, 2004. Understanding the spatial clustering of severe acute respiratory syndrome (SARS) in Hong Kong. Environ Health Perspect 112, 1550-1556.

Landis JR, Koch GG, 1977. The measurement of observer agreement for categorical data. Biometrics 33, 159-174.

MacEachren AM, Brewer CA, Pickle LW, 1998. Visualizing geo-referenced data: representing reliability of health statistics. Environ Plan A 30, 1547-1562.

McDade JE, Newhouse VF, 1986. Natural history of *Rickettsia rickettsii*. Ann Rev Microbiol 40, 287-309.

Monello RJ, Gompper ME, 2007. Biotic and abiotic predictors of tick (*Dermacentor variabilis*) abundance and engorgement on free-ranging raccoons (*Procyon lotor*). Parasitology 134, 2053-2062.

National Elevation Dataset, 2010. The national map seamless server. U.S. Geological Survey.

National Land Cover Dataset, 2010. Multi-resolution land characteristics consortium. U.S. Geological Survey.

Oliver JH Jr, 1971. Parthenogenesis in mites and ticks (*Arachnida: Acari*). Am Zool 11, 283-299.

Parola P, Labruna MB, Raoult D, 2009. Tick-borne rickettsioses in America: unanswered questions and emerging diseases. Curr Infect Dis Rep 11, 40-50.

Pawar S, Koo MS, Kelley C, Ahmed MF, Chaudhuri S, Sarkar S, 2007. Conservation assessment and prioritization of areas in northeast India: priorities for amphibians and reptiles. Bio Conserv 136, 346-361.

Phillips SJ, Anderson RP, Schapire RE, 2006. Maximum entropy modelling of species geographical distributions. Ecol Modell 190, 231-259.

Phillips SJ, Dudík M, 2008. Modelling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography 31, 161-175.

PRISM Climate Group, 2010. Precipitation 2004-2008. Oregon State University.

Rushton G, Peleg I, Banerjee A, Smith G, West M, 2004. Analysing geographical patterns of disease incidence: rates of late-stage colorectal cancer in Iowa. J Med Syst 28, 223-236.

Sarkar S, Strutz SE, Frank DM, Rivaldi CL, Sissel B, Sánchez-Cordero V, 2010. Chagas disease risk in Texas. PLoS Negl Trop Dis 4, e836.

Serra P, Pons X, Saurí D, 2003. Post-classification change detection with data from different sensors: some accuracy considerations. Int J Remote Sens 24, 3311-3340.

Sonenshine DE, 1991. Biology of ticks, Vol. 1. Oxford University Press, pp. 447.

Sonenshine DE, 1993. Biology of ticks, Vol. 2. Oxford University Press, pp. 465.

Stein KJ, Waterman M, Waldon JL, 2008. The effects of vegetation density and habitat disturbance on the spatial distribution of ixodid ticks (*Acari: Ixodidae*). Geospat Health 2, 241-252.

Tiwari C, 2011. Web-based disease mapping and analysis program. Webdmap.

Tiwari C, Rushton G, 2005. Using spatially adaptive filters to map late stage colorectal cancer incidence in Iowa. Developments in Spatial Data Handling, 665-676.

Troughton DR, Levin ML, 2007. Life cycles of seven ixodid tick species (*Acari: Ixodidae*) under standardised laboratory conditions. J Med Entomol 44, 732-740.

U.S. Department of Agriculture, 2010. Soil Survey Geographical (SSURGO). Database.

Williamson PC, Billingsley PM, Teltow GJ, Seals JP, Turnbough MA, Atkinson SF, 2010. *Borrelia*, *Ehrlichia*, and *Rickettsia* spp. in ticks removed from persons, Texas, USA. Emerg Infect Dis 16, 441-6.

WorldClim, 2010. Global Climate Data.

Yang DH, Goerge R, Mullner R, 2006. Comparing GIS-based methods of measuring spatial accessibility to health services. J Med Syst 30, 23-32.