Zoonotic infectious diseases are inextricably linked to their environment. In the case of vector-borne pathogens, environmental determinants control the distribution and abundance of vertebrate reservoirs, vectors, and pathogens (Kitron 1998; Pavlovsky 1966). The dynamics of their transmission is a function of several abiotic and biotic processes that affect the ecosystem as a whole and control the survival of the arthropod vector (Hay et al. 2000). The emergence of these diseases can be attributed to a response of an ecosystem to pressure resulting from environmental change (Reeves et al. 1994; Rogers and Randolph 2000). Thus, the spread of a number of vector-borne diseases can be correlated with natural and human-induced changes on the Earth system. Elucidating the relationship between environment and vector is essential for measuring human risk and target- ing effective surveillance and control measures.

A number of landscape features are closely associated with zoonotic diseases. In particular, climate, land cover, and landscape patterns are important epidemiologic determinants (Frank et al. 1998; Lindgren et al. 2000; Randolph 1993). The idea of using landscape ecology in the context of epidemiology was first brought about by Pavlovsky (1966). He introduced the concept of natural focality, defined by the idea that microscale disease foci are determined by the entire ecosystem (Galuzo 1975). With the recent availability of new technologies such as remote sensing and geographic information systems along with advances in spatial and temporal statistics, the theories of landscape ecology can be used analytically (O’Neill et al. 1999). Landscape features can be mapped and used as predictors of the pathogen, vector, and host reservoir presence and abundance (Daniel and Kolar 1990; Dister et al. 1997; Kitron and Kazmierczak 1997). Because an accurate understanding of the spatial distribution of both the pathogen and the vector is integral to disease prevention strategies, spatially explicit models founded on basic ecologic principles are invaluable tools in epidemiology and public health.

The deer tick, *Ixodes scapularis*, presents an ideal example of a disease vector that depends profoundly on climate and landscape patterns (Frank et al. 1998; Maupin et al. 1991; Mount et al. 1997; Ostfeld et al. 1996). *I. scapularis* is the primary vector of *Borrelia burgdorferi*, the agent of Lyme disease, in North America (Dennis et al. 1998; Keirans et al. 1996). Lyme disease is currently the most prevalent vector-borne disease in the United States, with more than 100,000 cases reported by the U.S. Centers for Disease Control and Prevention since its discovery in 1982 (Orlowski et al. 2000). In addition, *I. scapularis* is a known vector of other tick-borne diseases, including human babesiosis (Spigelman et al. 1985) and human granulocytic ehrlichiosis (Des Vignes and Fish 1997; Schwartz et al. 1997). Climatic variation has an essential function in determining *I. scapularis* population maintenance and distribution. Abiotic factors, including temperature and humidity, are likely to regulate off-host tick survival (Bertrand and Wilson 1996; Needham and Teel 1991). Water stress and temperature are major causes of mortality in nonfeeding ticks because the seasonal patterns of these variables control both developmental success and rates for all stages (Needham and Teel 1991). Because 98% of the *I. scapularis* life cycle occurs off of the host, climate should play a major role in the distribution of tick populations across the United States (Fish 1993). However, the complex relationship between the tick vector and the environment hinders a detailed understanding of the ecologic constraints on the distribution of *I. scapularis*.

Moreover, there is still no consensus on the precise geographic distribution of Lyme disease in the United States because of increased human case surveillance, overdiagnosis, underreporting, and human travel. In addition, the underlying ecologic data supporting vector distribution are limited and incomplete because of uneven sampling and a lack of standardized field techniques (Dennis et al. 1998; Fish and Howard 1999). Without an accurate depiction of where Lyme disease exists, and where it is emerging, the targeting of efficient prevention and control strategies would be inaccurate. A more precise understanding of the spatial distribution of *I. scapularis* would enable the appropriate targeting of prevention efforts at populations at risk at the appropriate times of year. Although a distribution map (Dennis et al. 1998) exists for the United States, both the criteria for classification and the category definitions are problematic. The county-resolution data consist of a compilation of data from questionnaires, published literature, and the collection records of the U.S. National Tick Collection, Institute of Arthropodology and Parasitology, Georgia Southern University. The map is arbitrarily classified with the following categories: established (six or more *I. scapularis* presence and distribution); widespread (five to six *I. scapularis* presence and distribution); and limited (one to four *I. scapularis* presence and distribution) areas.

An understanding of the spatial distribution of the black-legged tick, *Ixodes scapularis*, is a fundamental component in assessing human risk for Lyme disease in much of the United States. Although a county-level vector distribution map exists for the United States, its accuracy is limited by arbitrary categories of its reported presence. It is unknown whether reported positive areas can support established populations and whether negative areas are suitable for established populations. The steadily increasing range of *I. scapularis* in the United States suggests that all suitable habitats are not currently occupied. Therefore, we developed a spatially predictive logistic model for *I. scapularis* in the 48 conterminous states to improve the previous vector distribution map. We used ground-observed environmental data to predict the probability of established *I. scapularis* populations. The autologistic analysis showed that maximum, minimum, and mean temperatures as well as vapor pressure significantly contribute to population maintenance with an accuracy of 95% (*p < 0.0001*). A cutoff probability for habitat suitability was assessed by sensitivity analysis and was used to reclassify the previous distribution map. The spatially modeled relationship between *I. scapularis* presence and large-scale environmental data provides a robust suitability model that reveals essential environmental determinants of habitat suitability, predicts emerging areas of Lyme disease risk, and generates a new map of Lyme disease risk that is a robust suitability model that reveals essential environmental determinants of habitat suitability, predicts emerging areas of Lyme disease risk, and generates a new map of Lyme disease risk that is more precise and detailed than previous maps.
reported in any stage (considered a critical mass) or more than one life stage identified (considered a reproductive population), reported (fewer than six I. scapularis identified and one life stage identified), and absent (no collections of I. scapularis). Additionally, it is not known whether areas of reported occurrence could support continuous population maintenance, or whether absent areas are unsuitable for establishment or are simply the result of inadequate sampling. Despite these limitations, the map provided an essential source of information that guided the Advisory Committee on Immunization Practices recommendations for vaccinating the public against Lyme disease (Fish and Howard 1999).

Because of the important role of an accurate vector distribution map for Lyme disease prevention, we developed a spatially predictive logistic model for I. scapularis in the United States to improve the current reported vector distribution map, and highlight areas of potential emerging disease risk. This model builds upon other vector distribution studies that have used environmental data to enhance base surveillance data (Cumming 2000; Estrada-Peña 2002) by accounting for the effects of spatial autocorrelation through a Bayesian approach and providing a statistical means for creating an easily interpretable classified risk map. We used this spatially explicit habitat suitability model to dissect the relative importance of seasonal temperature and humidity in determining the biologic constraints of I. scapularis distribution. This study is also unique because we validated the environmental model by strategic field sampling. Our model of the relationship between tick habitat suitability and large-scale environmental data can be used to predict the current and future I. scapularis habitat suitability and to provide the basis for a detailed ecologic risk map for Lyme disease as well as other diseases transmitted by I. scapularis.

Materials and Methods

We used ground-observed environmental data to predict the probability of established I. scapularis populations. We obtained the data from a 0.5° × 0.5° global data set of 30-year average monthly climatic surfaces, based on daily measurements for the period of 1961–1990, available through the Climatic Research Unit at the University of East Anglia (New et al. 1999). Climate surfaces were derived from interpolation of station data as a function of latitude, longitude, and elevation using thin-plate splines. This data set has been used previously in the study of tick-borne encephalitis foci in Europe (Randolph and Rogers 2000). Variables selected for analysis include minimum, maximum, and mean monthly temperature and monthly vapor pressure. The climate data were imported using ERDAS IMAGINE (ERDAS 2001) and processed using ESRI ArcGIS (ESRI 2001). Data for each 0.5° pixel corresponding to the conterminous United States were summarized by calculating the cell statistics for each variable, including mean, maximum, minimum, and standard deviation (SD).

We selected the U.S. distribution map of I. scapularis as the dependent variable in the model (Dennis et al. 1998). This distribution map was converted into the 0.5° grid, assigning each cell the category that took up the greatest amount of area. Only areas classified as absent and established were considered in the analysis because the true status of I. scapularis in the reported category is unknown.

Next, we derived a logistic model for the relationship between environment and known established I. scapularis populations. The 16 independent variables, including the mean, maximum, minimum, and SD of each of the four monthly environmental factors, were first individually examined to test the linearity assumption for a continuous regressor variable. Log odds ratio plots were created for each variable grouped into deciles to determine if higher-order polynomial terms or variable transformations were necessary. The decision was formally verified by a goodness-of-fit test, comparing the likelihood ratio of each potential model for a given factor to that when treating the variable as a nominal categorical variable divided into deciles (Holford 2002). The best-fitting model was selected using a chi-square goodness-of-fit test (Holford 2002). The log odds ratio plots also have biologic significance because they display the relationship between the environmental variables and established populations of I. scapularis. Variables that required greater than a fourth-order polynomial were removed because of uncertain biologic relevance. A binary logistic regression model was fit with SAS software using the remaining variables and associated interaction terms (SAS 2001). The analysis was carried out using a stepwise selection with Pentry = 0.15 and P removal = 0.2 (Hosmer and Lemeshow 1989).

Because adjacent areas tend to have similar climatic conditions and the probability of occurrence in one location is not independent of occurrence in neighboring locations, the number of degrees of freedom (DF) reduces and the chance of a type I error increases (Legendre 1993). By dealing with spatial autocorrelation, we can show that the I. scapularis range is nonrandom and dependent on climatic variables (Augustin et al. 1996). This consideration will also give a better indication of the relative importance of environmental factors (Augustin et al. 1996). Spatial autocorrelation in the probability of establishment derived from the initial regression was assessed, therefore, by Moran’s I using Crimestat (Levine 2000). We also evaluated the extent of the correlation by a semivariogram with a linear-to-sill model using GS+ (Gamma Design Software 1998).

We accounted for spatial autocorrelation by applying an autologistic approach (Augustin et al. 1996). A moving window was used to calculate the average probability of occupation among the set of neighbors defined by the limit of correlation, weighted by the inverse of the Euclidean distance. This average probability is called the autologistic term and is added as an additional covariate in the logistic model (Augustin et al. 1996; Osborne et al. 2001). The autologistic term acts as a smoothing filter, removing isolated pixels and consolidating habitat patches. Because vector population status in the reported locations is unknown, we incorporated a modified Gibbs sampler to estimate the distribution in these unknown areas (Augustin et al. 1996, 1998). This Monte Carlo–type method involves iterating the procedure of fitting the autologistic model, deriving the probability surface for all locations, and then recalculating the autologistic term until stability. We implemented this procedure with a program written with Microsoft Visual C++.

The receiver operating characteristics (ROC) plot was used to independently assess accuracy (Cumming 2000; Osborne et al. 2001). This method graphs sensitivity versus 1–specificity over all possible cutoff probabilities. The area under the curve for ROC (AUC) is a measure of overall fit, where 0.5 indicates a chance performance (Fielding and Bell 1997). We generated the plot for the autologistic model using Simstat (Provalis Research 2000). A probability cutoff point for habitat suitability assessed by sensitivity analysis determined whether a given cell could support an established vector population.

Strategic field sampling at locations of varying probability confirmed the validity of the environmental model for established I. scapularis populations. Sampling was confined to a portion of Northeast United States, including the states of Pennsylvania, Maryland, New Jersey, and Delaware (Figure 1). This area was ideal for sampling because the established population probability is highly variable among cells. Validation did not include areas previously defined as absent or reported because it is uncertain whether I. scapularis has expanded into these areas yet. We determined the population status at each location by measuring nymphal abundance with field sampling.

To estimate total tick abundance, we designed a multistage sampling scheme. The first stage involved stratified random sampling, where 20 cells were selected for sampling with equal allocation to probability groups divided by quintiles. The second sampling stage involved sampling I. scapularis within each cell selected in the first stage. To standardize the sampling effort, a state park or state forest was selected at each sample grid cell. As a result, similar deciduous forest habitat would be sampled at each site. We estimated nymphal
abundance by dragging a 1-m² flannel cloth, fixed to a wooden handle, through vegetation (Daniels et al. 2000; Milne 1943). For each park, we randomly selected 10 transects of 100 m² for dragging, inspecting the cloth at 20 m intervals. We performed sampling daily between 900 and 1800 hr throughout June 2002, the period of greatest host-seeking activity (Daniels et al. 2000). A one-tailed Fisher exact test was used to test for a positive association between positive tick collections and suitable habitat defined by the sensitivity analysis. Both the validity and the predictive value of the model were also assessed.

Results
The converted I. scapularis distribution 0.5° grid map gave a total of 3,628 cells, with 11% established area (n = 416), 77% absent area (n = 2,785), and 12% remaining reported (n = 427), analogous to the distribution of the classified county data. Univariate analysis of the 16 dependent environmental variables showed a strong positive association with tick presence with a fourth-order polynomial regression (\(R^2 = 0.97\); Figure 2), whereas the other variables were shown to have more complex relationships. Goodness-of-fit testing removed seven of the 16 environmental descriptors.

The initial logistic model using only established and absent locations uncovered a significant relationship between tick presence and the nine remaining variables and associated interaction terms (\(\chi^2 = 314\); DF = 8; \(p < 0.00001\)). The stepwise analysis left eight significant variables in the model, representing the four environmental correlates used (Table 1). Maximum monthly temperature and SD of vapor pressure had the greatest influence on population maintenance (Table 1). The model also predicted the suitability status of the additional reported locations.

The probability of established tick populations derived from the initial regression exhibited significant spatial autocorrelation by Moran's I (\(p < 0.00001\)). The limit of the autocorrelation, assessed by a semivariogram with a linear-to-sill model, was 2,297 km. The autologistic model was then applied with a moving window radius of this limit. The regression coefficients converged after five iterations to produce the final probability surface (Figure 3). The final logistic model was significant (\(p < 0.0001\)) with all eight variables remaining in the model (Table 1). The inclusion of spatial autocorrelation in the model reveals minimum monthly temperature to play a more significant role in sustaining tick populations than was described by the non-spatial model.

The ROC plot for the autologistic model significantly outperformed the chance model with an accuracy of 0.9508 (\(p < 0.00005\); Figure 4). Sensitivity analysis produced a probability threshold of 21%, because it gave a maximum sensitivity of 88% and a specificity of 89%. This cutoff was used to reclassify the existing distribution map. Of the reported locations (n = 427), 66% were defined as established, and 11% of the absent areas (n = 2,327) were defined as suitable. All other reported and absent areas were considered unsuitable. Areas previously defined as established maintained the same classification. We therefore propose a new distribution map for I. scapularis in the United States (Figure 5) with the categories established, suitable for colonization but not yet introduced, and unsuitable for colonization.

For the model verification, 20 parks were sampled for nymphal abundance across four states (Figure 6): 5 parks were sampled in Maryland, 11 in Pennsylvania, 3 in New Jersey, and 1 in Delaware. A total of 413 I. scapularis nymphs were collected at all the study sites, with an average abundance of 0.018 nymph/L. The sampling sites were then grouped according to the probability threshold. The sites above the threshold (n = 16) had an average tick abundance of 0.026 (SD = 0.019), with only one site with no collections (Figure 6). There were no ticks found in sites below the threshold (n = 4). The one-tailed Fisher exact test revealed a significant association between the presence of nymphs and suitable habitat (\(p < 0.01\)). The model produced a sensitivity of 100% (15 of 15), a specificity of 80% (4 of 5), and overall accuracy of 95% (19 of 20). The false positive and false negative rates were 6.25% (1 of 16) and 0% (0 of 4), respectively.
Discussion
The spatially modeled relationship between I. scapularis presence and large-scale environmental data served both to reveal the essential environmental determinants of I. scapularis population establishment and to generate the spatial pattern of habitat suitability across the United States. We used the seasonality of temperature and humidity to explore the role of climate upon I. scapularis distribution because of their relationship with host-seeking behavior and off-host mortality. The accuracy of the model (AUC = 95%) indicates that variations in these climatic conditions play a major role in determining the range of I. scapularis.

Both univariate and multivariate analyses quantified the relationship between previously proposed environmental constraints and established I. scapularis populations. The univariate analyses computed the relationship between single variables and established populations. Minimum temperature was the only variable to have a noncomplex positive relationship (Figure 2). The importance of minimum temperature is reflected in the idea that it represents the environmental conditions at the lower limit of tick survival. The log odds plot revealed the possibility of a threshold winter condition, below which development is unsuccessful. An average monthly minimum temperature below –7°C in the winter is predicted to prevent an area from maintaining established populations.

The multivariate logistic model revealed the relative contribution of environmental variables in explaining suitability for I. scapularis established populations. Our analysis suggests the importance of climatic extremes and variation in vapor pressure as major indicators of habitat suitability. However, we cannot be certain that other variables that were excluded by the variable selection algorithm do not also play a role because of the collinearity of these climatic factors. The primary use of this model is prediction, which is affected very little by collinearity (Neter et al. 1996). Removing the effect of spatial autocorrelation gives a better reflection of the relative importance of each factor. When accounting for space, minimum temperature was the only variable to increase in importance, once again indicating a significant biologic role. Maximum temperature and vapor pressure also played a significant role in determining the range of I. scapularis despite the complex relationships displayed by the univariate analyses. Maximum temperature was the most influential variable in the model, indicating an important role in sustaining tick populations. Higher temperatures augment both the developmental and hatching rates while hindering overall survival and oviposition success (Needham and Teel 1991).

Because of this defined relationship between environment and tick population maintenance, climate change may be involved in controlling the future distribution of the Lyme disease vector (Estrada-Peña 2002; Shope 1991). Variation in climatic conditions may affect the range of I. scapularis by altering host-seeking activity and vector population density. This model could therefore forecast the future distribution of this vector by analyzing the trends in both natural and anthropogenic environmental changes.

The derived relationship also assessed environmental suitability for I. scapularis populations for all locations across the conterminous United States. This suitability map builds upon data from the previously published vector distribution map (Dennis et al. 1998). Although their map comprehensively covers all counties of the United States, there are clear disadvantages. First, the categories used to define established populations are determined according to an arbitrary threshold number of collected ticks. There is no biologic significance attributable to the presence or absence of six collected ticks. Second, the suitability of areas for established populations cannot be determined from the map. A subset of the absent areas may still represent colonizable areas, and reported locations may represent either adventitious specimens in unsuitable areas or the initiation of a reproducing population. Third, the reported map highlights the problem of the nonreportability of negative data, preventing a distinction between absent areas and unsampled areas. Using climatic variables that create the appropriate conditions for I. scapularis population maintenance can therefore provide more accurate predictions of the current and potential future distribution of I. scapularis.

Our improved vector distribution map evaluates whether a particular location can support a continuous population of I. scapularis. Therefore, unsuitable areas may have introductions but not allow for completion of...
the tick’s life cycle (Lindsay et al. 1998). Areas that are suitable but not currently sustained are those that will experience the greatest increase in _I. scapularis_ population density, because introductions should result in reproducing populations. The presence of these areas validates the idea that _I. scapularis_ continues to expand its range (Dennis et al. 1998; Keirans et al. 1996). According to the model, notable increases in vector distribution are expected in Virginia, North Carolina, Georgia, Minnesota, Iowa, and Michigan. Interestingly, the probability surface displays low suitability on the West Coast, where _Ixodes pacificus_, the vector of Lyme disease in the western United States, is situated. Thus, the model provides evidence that the environmental constraints on the ranges of the closely related species are distinctly different.

Because locations that were originally designated as established were maintained in the new map, certain areas with low habitat suitability were still designated as established areas. For instance, in Missouri, isolated pixels are classified as established even though the entire state is predicted to have an extremely low probability of established _I. scapularis_ populations. It cannot be determined whether samples in these areas most likely represent samples of temporarily introduced ticks or founding populations, revealing the problems with original classification criteria. However, this information is still retained in the new classification map, which makes this model a conservative revision.

The field sampling design validated the environmental _I. scapularis_ model with empirical data. The ability of the model to predict habitat suitability was reflected in the significant positive association between nymphal presence and above-threshold suitability (p < 0.01). Interestingly, the model predicted all sites where ticks had been collected as having suitability above threshold, yielding a sensitivity of 100%. This result highlights the idea that climate determines the appropriate conditions for the presence of reproducing populations. However, other factors, such as host density and species composition, may be more important factors controlling the tick population size than is climate.

Our goal in this work was to identify the environmental determinants regulating _I. scapularis_ populations in space, which in turn dictates Lyme disease risk. However, this model should not be considered the only layer necessary to identify areas of risk. Not all areas that can support populations of _I. scapularis_ can also maintain an enzootic cycle of _Borrelia burgdorferi_. Therefore, a second layer that illustrates infection prevalence in the ticks would be an essential component of a complete Lyme disease risk map. Because of the inaccuracies in human case report data, estimates of tick infection prevalence might also be predicted through the use of environmental data. In addition, other information, such as canine seroprevalence for _B. burgdorferi_ and host species composition data, could be used to construct an infection prevalence layer (Daniels et al. 1993; Fish and Howard 1999).

This habitat suitability model for _I. scapularis_ provides an essential first step toward a more precise geographic distribution of Lyme disease and a stronger evidence base for determining human risk in specific endemic regions. Our methods can also provide a template for mapping other vector-borne zoonotic diseases. Given the limitations of a human reporting system, a more complete assessment of risk can be provided by developing a spatial model of environmental suitability. The output of such a model can enable improved predictions of emerging risk, as well as aid in implementation of efficient control strategies and target disease prevention efforts toward high-risk populations.

**REFERENCES**


ERDAS. 2001. IMAGINE, Version 8.5. Atlanta, GA: Leica Geosystems, LLC.


ERDAS. 2001. IMAGINE, Version 8.5. Atlanta, GA: Leica Geosystems, LLC.


Fox D, Howard C. 1999. Methods used for creating a national


Pavlovska EN. 1996. Natural Nodality of Transmissible Diseases, with Special Reference to the Landscape Epidemiology of Zoanthropopause, Urbana, IL: University of Illinois Press.


