Data–model fusion to better understand emerging pathogens and improve infectious disease forecasting

Shannon L. LaDeau,[1,5] Gregory E. Glass,[2] N. Thompson Hobbs,[3] Andrew Latimer,[4] and Richard S. Ostfeld[1]

[1]*Cary Institute of Ecosystem Studies, Millbrook, New York 12545 USA*
[2]*Johns Hopkins School of Public Health, Department of Molecular Microbiology and Immunology, Baltimore, Maryland 21205 USA*
[3]*Colorado State University, Natural Resource Ecology Laboratory and Graduate Degree Program in Ecology,
Fort Collins, Colorado 80524 USA*
[4]*University of California, Department of Plant Sciences, Davis, California 95616 USA*

Abstract. Ecologists worldwide are challenged to contribute solutions to urgent and pressing environmental problems by forecasting how populations, communities, and ecosystems will respond to global change. Rising to this challenge requires organizing ecological information derived from diverse sources and formally assimilating data with models of ecological processes. The study of infectious disease has depended on strategies for integrating patterns of observed disease incidence with mechanistic process models since John Snow first mapped cholera cases around a London water pump in 1854. Still, zoonotic and vector-borne diseases increasingly affect human populations, and methods used to successfully characterize directly transmitted diseases are often insufficient. We use four case studies to demonstrate that advances in disease forecasting require better understanding of zoonotic host and vector populations, as well of the dynamics that facilitate pathogen amplification and disease spillover into humans. In each case study, this goal is complicated by limited data, spatiotemporal variability in pathogen transmission and impact, and often, insufficient biological understanding. We present a conceptual framework for data–model fusion in infectious disease research that addresses these fundamental challenges using a hierarchical state-space structure to (1) integrate multiple data sources and spatial scales to inform latent parameters, (2) partition uncertainty in process and observation models, and (3) explicitly build upon existing ecological and epidemiological understanding. Given the constraints inherent in the study of infectious disease and the urgent need for progress, fusion of data and expertise via this type of conceptual framework should prove an indispensable tool.

Key words: Bayesian; climate change; disease ecology; emergent; forecast; hierarchical; likelihood; pathogen; state-space models; transmission; vector; zoonotic.

Introduction

The need to understand and predict the timing and intensity of outbreaks of infectious disease has existed since humans first clustered around common resources. In the 19th century, John Snow (1855) systematically traced cholera mortalities to a shared London water pump in what is now often recognized as the first published epidemiological study (Evans 1976). At the time, Snow and his contemporaries were largely unaware of the ecological dynamics that led naturally occurring aquatic bacteria (*Vibrio cholerae*) to cause epidemic disease. Snow used scientific inquiry based on observations and an implicit spatiotemporal clustering model to demonstrate a connection between cholera cases and water supplies. Today we know that *V.*

cholerae occurs naturally in copepod hosts across many aquatic environments (Colwell 1996) and that the ecology of these copepod communities is an essential component of managing and forecasting cholera epidemics in humans (Lipp et al. 2002, Colwell et al. 2003).

From its inception, the formal study of infectious disease has depended on strategies for integrating theoretical process models with patterns of observed disease incidence. The availability of data and advances in technology in just the past few decades have greatly enhanced our ability to evaluate the ecological dynamics of infectious disease (e.g., Glass et al. 2000, Miller et al. 2006, Ostfeld et al. 2006) and have fostered tremendous progress in dynamic modeling of disease processes (Ferguson et al. 2001, Bjornstad et al. 2002, Grenfell et al. 2002, Keeling et al. 2003, Ferrari et al. 2008, Smith et al. 2008, He et al. 2010). However, even with the extraordinary advances in data availability, computational power, and algorithms for stochastic modeling, fusing data with mechanistic understanding in ways that
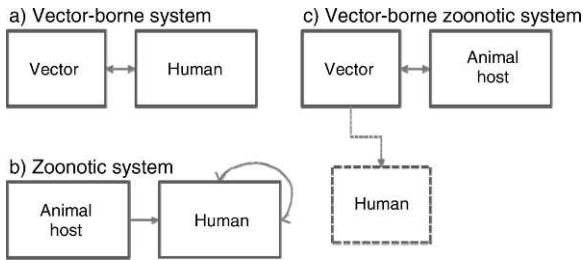
FIG. 1. Human disease systems involving (a) vector-borne (e.g., malaria), (b) zoonotic, and (c) vector-borne zoonotic pathogens. Lines denote pathogen transmission. Double-ended arrows denote bidirectional transmission. In panel (b), humans are infected through direct contact with wild or domestic animals and do not generally reinfect these reservoir hosts but may be infectious to other humans (i.e., severe acute respiratory syndrome [SARS]). In the system depicted in panel (c), humans are not important to pathogen persistence or transmission but may be infected if exposed to the vector (i.e., Lyme, West Nile virus [WNV]).

allow for forecasts in advance of epidemics remains a major challenge (Ostfeld et al. 2005, Kitron et al. 2006, Glass 2007), and is especially problematic in the case of emergent infectious diseases.

Emerging infectious diseases (EIDs) are caused by pathogens that have been newly identified (e.g., human immunodeficiency virus [HIV], severe acute respiratory syndrome [SARS], Lyme disease, H5N1 influenza), have undergone range expansions into naïve populations (e.g., West Nile virus, Dengue fever), or have evolved into more virulent or drug resistant strains (e.g., drug-resistant malaria and tuberculosis) (Woolhouse and Dye 2001). Several recent studies have demonstrated that reported EIDs have increased in recent decades and that a majority of EIDs are zoonotic (vertebrate reservoir) and vector-borne (Taylor et al. 2001, Wilcox and Gubler 2005, Jones et al. 2008). Following declines in the mid-20th century, mortalities in the United States attributed to infectious diseases began to rise in the 1980s, an increase that has been blamed on the spread of HIV and related infections (reviewed by Greger 2007). It is likely that exponential growth of the human population and human resource consumption (Daszak et al. 2000), globalization of travel and trade (Karesh et al. 2005), and changes in climate (Pascual and Bouma 2009) have already and will continue to facilitate this rise in EIDs.

Here we discuss approaches for integrating the information in data with the biological understanding in process models, paying particular attention to how data–model assimilation can advance understanding of infectious disease. We use a case study approach to examine data–model assimilation in vector-borne and zoonotic diseases affecting humans and present a conceptual modeling framework to support ecological disease research and generate forecasts of pathogen amplification and persistence in the environment, as well as human disease risk.

OVERVIEW OF DISEASE MODELING AND MODEL DATA FUSION

There is an inherent paradox in modeling epidemics. Models that can guide control efforts are most useful early in the course of an outbreak, when the data are most sparse (Matthews and Woolhouse 2005). If the models are effective in supporting wise choices of interventions that reverse the trajectory of infection, then data *remain* sparse. Thus, assimilation of data with models of disease outbreaks must be able to generate insight from limited data.

Pathogen transmission is the critical process behind disease outbreaks. Pathogen transmission involves the interaction of at least two organisms (Fig. 1) and is inherently a spatiotemporal process (Antolin 2008). It is also a process that cannot be directly observed. Unobservable processes are difficult to model because estimation of latent parameters relies on our ability to accurately relate them to quantities that can be measured. Available data are often discrete, while the underlying processes are continuous in nature. Numerous approaches have been employed to estimate transmission rates and predict disease intensity from observed disease incidence (i.e., successful transmission events for a fixed spatial and temporal interval). A majority of advances in understanding infectious dynamics have been made in endemic diseases with direct human-to-human transmission because data and understanding of human population dynamics are generally available (Anderson and May 1992, Bolker and Grenfell 1995, Bjornstad et al. 2002, Ferrari et al. 2008, He et al. 2010). Systems that include zoonotic reservoir hosts and vector transmission are more complex than directly transmitted human diseases and require some understanding of the ecological dynamics controlling host and vector populations, as well as the factors that influence disease incidence (Yates et al. 2002, Ostfeld et al. 2006, Begon et al. 2009). Sick animals are not easily studied in their natural environment and the historical time series that have supported many advances in infectious disease research (e.g., Anderson and May 1992) do not exist for wildlife populations or for newly emergent diseases.

Researchers have generally focused on either model-based (predominantly mathematical simulation) or data-based (with statistical hypothesis testing) approaches to develop understanding and generate predictions of ecological dynamics (Hobbs and Hilborn 2006), including disease incidence. Dynamic compartmental models, a common model-based approach in epidemiology, employ a linked system of differential equations to represent susceptible, exposed, infected, and recovered states (i.e., SEIR) in order to characterize transmission and disease progression in host populations (Anderson and May 1992). Data are not generally integrated formally with these mathematical process equations, instead critical parameter estimates are derived from other analyses and plugged into the model. Researchers

use model simulations to predict incidence "data" that can be compared to observed data.

A primary goal of compartmental modeling in epidemiology is estimation of the basic reproductive number, R_0, the number of new infections created by a single infected individual in a wholly susceptible population. Theoretically, an epidemic can only proceed if R_0 is greater than 1 (Anderson and May 1985). The parameter estimates that characterize disease transmission can be adjusted to explore how changes in these rates alter the probability that an epidemic will occur as well as its probable duration. Dynamic compartmental models can be valuable tools for assessing the relative importance of critical rates and have been crucial for evaluating vaccination strategies (Anderson and May1985) and controlling emerging human epidemics such as SARS, which we review later. The mathematical structure reflects several underlying biological assumptions, including the nature of transmission and the constancy of rate parameters in time and space. When these assumptions are not correct then the resulting R_0 values can be off by orders of magnitude (e.g., Wonham et al. 2006). Limitations associated with mathematical modeling of infectious transmission have been reviewed elsewhere (e.g., Wearing et al. 2005). These models can be very powerful *if* the biological processes are well defined *and* demographic and environmental stochasticity are either well-characterized or unimportant for meeting the goal of the research (Hohle et al. 2005, Britton and Lindenstrand 2009, He et al. 2010). However, capturing the stochasticity is especially important when the research goal is to forecast the timing and intensity of disease outbreaks. Recent "plug-and-play" methods use simulations from flexible classes of compartmental models to generate understanding of complex interactions and stochasticity in dynamic systems (Ionides et al. 2006, He et al. 2010), but currently are applied to understanding direct (human) transmission processes with relatively rich data. Dynamic compartmental modeling may be particularly limited when data are sparse and biological understanding poor, especially when forecasting infectious dynamics of vector-borne and zoonotic disease systems is the goal.

Data-based or phenomenological modeling of disease incidence has strengths and weaknesses complementary to dynamic mathematical models. Advances in remote sensing have greatly increased the amount and scale of data available to characterize environmental variables that can influence the timing and location of disease outbreaks (Kitron et al. 2006). Observed disease incidence (or host or pathogen abundance) can be readily compared to a suite of variables representing habitat type (Xiao et al. 2007, Brown et al. 2008), climate (Brownstein et al. 2003, Xiao et al. 2007), and dispersal pathways (Kilpatrick et al. 2006a). These data-based methods can certainly capture the heterogeneity in population abundance and infection dynamics, but may

not have enough biological integrity to extend beyond the unique time and place in which data were collected. Phenomenological models tend to rely on the assumption that the abiotic features measured actually determine disease incidence. When the disease is in the process of expanding in range, the abiotic conditions that match its current range might represent only a small subset of those that permit its existence. Thus, the potential for these types of approaches to provide accurate or useful forecasts of disease incidence relies on how well they capture the underlying biological processes that determine transmission rates, which can be difficult to evaluate. Still, phenomenological modeling has undoubtedly provided important insight into the distribution and spread of infectious disease and can be instrumental in defining the broad realm of possible interactions and hypotheses (Kilpatrick et al. 2006a, Xiao et al. 2007, Peterson and Williams 2008).

Forecasting disease outbreaks must be driven by biological understanding of the processes that determine pathogen transmission. When that understanding is limited and mechanistic data are sparse, such as is often the case with EIDs, efficient data–model integration to support inference and forecasting is crucial. Methods for assimilating data with models of infectious disease are dominated by likelihood approaches. Likelihood based approaches quantitatively evaluate data support for specific hypotheses and are important data–model assimilation tools in ecology (Hobbs and Hilborn 2006). Bayesian methods are increasingly applied; especially when ecological complexity is high and forecasting is a primary goal (Clark 2005). Hierarchical modeling (using either Bayesian or maximum likelihood methods) is gaining prominence across ecological disciplines (Clark and Gelfand 2006, Gelman and Hill 2007, Cressie et al. 2009).

A hierarchical (sometimes referred to as *multilevel*) model structure defines how information is shared across sampling units and processes. Although it is not yet widely used in disease ecology, this approach is particularly attractive because it allows researchers to accommodate important differences (e.g., among groups, individuals, or regions) while still allowing for shared characteristics (Clark et al. 2005, Cressie et al. 2009, Qian et al. 2010). Additionally, when the hierarchical structure is built around a conditional modeling framework (i.e., Bayesian or likelihood based) then multiple datasets and processes can be combined in a common analysis (LaDeau and Clark 2006, Clark et al. 2007). State-space models, which may also be hierarchically structured, are becoming standard for portraying population dynamics (Tavecchia et al. 2009, Wang 2009), but they have not been widely applied in modeling infectious diseases (Baadsgaard et al. 2004, He et al. 2010). The state-space approach to modeling populations can be described as the nesting of two models; an observation equation that relates the observed data to the unobserved but true state of the

population, and a process equation that represents understanding of the processes governing the state. The state-space formulation allows for partitioning of uncertainty that arises from our inability to perfectly observe the process (i.e., observation error) and uncertainty that results from the failure of our model to perfectly represent the process (process variance or model misspecification). This is especially important in time series applications where process error propagates over time but observation error does not (Calder et al. 2003). If the model is further structured to allow for individual heterogeneity in disease risk (for example) as well as population risk, then it is also hierarchical. Both state-space and hierarchical approaches to data–model assimilation are potentially important avenues for future advances in understanding the infectious dynamics of zoonotic and vector-born infectious disease.

In the subsequent sections we present case studies of disease systems that build in ecological complexity from SARS, a directly transmitted human disease of zoonotic origin (Fig. 1b) to Dengue Fever, a vector-transmitted, anthroponotic, human disease (Fig. 1a) to Lyme and West Nile virus (Fig. 1c) where the pathogens are maintained in the environment by transmission among zoonotic hosts and arthropod vectors, independent of humans. Human infections with Lyme disease or West Nile virus result when unusual or extreme events that may occur seasonally or sporadically cause critical thresholds to be surpassed (e.g., contact rates, transmission timing, or vector abundance). Our goal is to identify similar characteristics among these diverse pathogen systems that either have facilitated or limited successful data–model integration for inference or forecasting. Finally, we use these insights to construct a conceptual hierarchical modeling framework for organizing research to advance the understanding of zoonotic and vector-borne pathogens and improve forecasting of human disease risk.

CASE STUDIES

Severe acute respiratory syndrome (SARS)

Severe acute respiratory syndrome (SARS) is caused by a zoonotic coronavirus (family Coronaviridae) that emerged in human populations during autumn of 2002 in Guangdong province of China. Before the epidemic was contained in summer of 2003, it had spread to 27 countries, causing almost 9000 infections and 800 deaths (Anderson et al. 2004). SARS offers an unusually valuable example of coordinated response to an emerging disease by the public health, medical, and research communities worldwide. A coordinated effort to understand the disease system led to rapid quarantines and likely prevented a pandemic of far greater magnitude than was realized.

Assimilation of data from the epidemic with models portraying the process of transmission played an important role in determining whether ongoing control efforts were effective and in evaluating alternatives for intervention by public health agencies (e.g., Lipsitch et al. 2003, Lloyd-Smith et al. 2003, Riley et al. 2003). Since the SARS epidemic, there has been extensive research using databases assembled from around the world to enhance preparedness for future outbreaks of SARS and similar diseases (Colizza et al. 2007, Kwok et al. 2007, Kramer-Schadt et al. 2009). The contribution of modeling in controlling SARS has been reviewed elsewhere (Anderson et al. 2004, Gumel et al. 2004, Bauch et al. 2005), and we will not duplicate those efforts here. Rather, we focus on how models of disease processes were fused with data to estimate key epidemiological states and parameters during the epidemic and thereafter. We pay particular attention to representing heterogeneity in transmission and identify ways that model–data assimilation might be improved in the future.

Dynamic compartmental models using linked systems of differential equations to characterize disease transmission in human populations (i.e., SEIR models) offered a vital epidemiological framework for assembling observations on the progression of the SARS epidemic. Model parameters were estimated from observations of events during the course of the disease, including onset of symptoms, hospitalization, and death or recovery. Assimilation of these data with a modeling framework allowed estimation of epidemiological quantities of interest, in particular, the basic reproductive number, R_0 (Table 1), and the reproductive number R, the number of infections created by an infected individual in a population containing a mixture of infected and susceptible individuals. The reproductive number is critical to assessing the success of control efforts because it must be <1 to assure that the epidemic is subsiding. Several data–model assimilation techniques were used to estimate R_0 and R including classical (Lipsitch et al. 2003), hierarchical Bayes (McBryde et al. 2006, Lekone 2008), and maximum likelihood (Riley et al. 2003, Wang and Ruan 2004) methods. Despite differences in approaches, similar estimates of key parameters were obtained (Table 1). A critically important result obtained from the parameterized models was the discovery, relatively early in the epidemic, that control efforts would be effective in reversing the exponential increase of new cases (Lipsitch et al. 2003, Riley et al. 2003).

There were several lessons from modeling SARS that can be broadly applied to other problems in assimilating data with models of infectious disease. As with any stochastic process, separate realizations of the process, in this case the trajectory of epidemics, differed among locations (Wallinga and Teunis 2004). Some of these differences can be explained by high levels of heterogeneity in transmission, the phenomenon known as "super spreading," which emerged as a hallmark of the SARS epidemic (Galvani and May 2005, Lloyd-Smith et al. 2005). Although the average number of infections created by an infected individual was certainly less than

TABLE 1. Estimates of the basic reproductive number (R_0) for SARS based on process models fused with case history data.

Study	Location	Modeling approach	R_0	95% confidence envelope
Riley et al. (2003)	Hong Kong	SEIR + hospitalized model for multiple nodes in metapopulation. Maximum likelihood estimation.	2.7†	2.2, 3.7
McBryde et al. (2006)	Shanxi Province, China	Bayesian estimates of gamma distributed transition times. Incubation time treated as random effect. SEIR model structure with hospitalized state.	4.8‡	2.2, 8.8
Lipsitch et al. (2003)	Hong Kong	Bayesian analysis of rate of exponential growth in infected population and serial interval using prior information from Singapore epidemic.	2.2	1.5, 7.7§
Lekone (2008)	Hong Kong	Bayesian analysis of variation of Riley et al. (2003) model.	3.8	sd = 0.09
Donnelly et al. (2003)	Hong Kong	Analysis of time-delay distributions using maximum likelihood methods.	2.7	2.3, 3.7
Wallinga and Teunis (2004)	Hong Kong	Maximum likelihood analysis of case pairing.	3.6	3.1, 4.2
Wallinga and Teunis (2004	Singapore	Maximum likelihood analysis of case pairing.	3.1	2.3, 4.0
Wallinga and Teunis (2004)	Vietnam	Maximum likelihood analysis of case pairing.	2.4	1.8, 3.1
Wallinga and Teunis (2004)	Canada	Maximum likelihood analysis of case pairing.	2.7	1.8, 3.6

Note: SEIR stands for a susceptible, exposed, infected, recovered compartmental transmission model.
† Does not include super-spreading events.
‡ Estimate of R early in epidemic.
§ The 90% credible interval

10 (Table 1), a few, rare individuals infected more than 100 (Riley et al. 2003, Galvani and May 2005, Lloyd-Smith et al. 2005, Small et al. 2006). It appears that hospital environments that failed to isolate infected individuals early in the epidemic caused this heterogeneity: rates of infection in un-isolated hospital environments were an order of magnitude higher than in the non-hospital environment (Kuk and Tan 2009). Thus, estimates of infectivity of individuals showed that the three examples of "super spreading" actually did not have exceptionally high infectiousness; instead, super-spreading events seemed to result from characteristics of the environments where these individuals expressed symptoms (Lloyd-Smith et al. 2005, Kuk and Tan 2009).

Heterogeneity in transmission or contact rates must be included as sources of stochasticity in models of disease processes. Failure to do so causes false precision in parameter estimates and in forecasts. Several alternatives for dealing with these heterogeneities in model–data assimilation have been used in modeling the SARS epidemic. Heterogeneity in transmission may be more accurately captured by increasing the number of states in the model, thereby allowing for differences in transmission parameters for each state. For example Riley et al. (2003) included meta-population structure in their model of the SARS epidemic in Hong Kong. Splitting the infected class into hospitalized and un-hospitalized states represents another illustration of this approach (Lipsitch et al. 2003, Wang and Ruan 2004). A second approach was to employ more realistic models of host contact to better represent heterogeneities resulting from non-uniform mixing of infected and susceptible

individuals (e.g., Meyers et al. 2005, Small et al. 2006, Zhong et al. 2009). Finally, parameters may be defined as temporally varying as the epidemic proceeds, reflecting the heterogeneity in transmission created by public health interventions (Kuk and Tan 2009). All of these alternative approaches require expanding the number of parameters to be estimated, which leads to problems with parameter estimation when data are limited (as they inevitably are early in an epidemic). More importantly, all of these strategies for dealing with heterogeneity require sufficient understanding of transmission to represent the heterogeneity in a coherent way within the structure of the model. Rarely will such understanding exist during disease emergence.

An approach that was not widely used in modeling SARS (but see Lloyd-Smith et al. 2005, McBryde et al. 2006) is to acknowledge that heterogeneity among individuals exists as a result of a complex interaction of genetics, contact behavior, environment, and demographics. Rather than trying to explicitly account for these heterogeneities in model structure, we can simply treat them as random effects, admitting that we do not understand their sources (Clark and Bjornstad 2004, Clark et al. 2005). In this way we incorporate a distribution that portrays heterogeneities that we are sure exist, but which we may not fully understand. Such hierarchical approaches may provide more reliable estimates of uncertainty in future efforts to assimilate models with data on epidemics.

State-space models were not applied in modeling the SARS epidemic and observation error was not treated formally in any of the data assimilation approaches.

Efforts were made to determine if model results were sensitive to such errors, without formally estimating their magnitude (Lipsitch et al. 2003). It is well known that there are uncertainties in even the best case-reporting data: there are cases that go undiagnosed and those that are misdiagnosed. These errors are most likely to occur early in epidemics, when data are sparse and parameter estimates are most sensitive to errors. This argues that future data–model assimilation could benefit from state-space approaches that formally estimate uncertainty arising from observation error and process variance.

The SARS epidemic illustrates both the challenges and the benefits of formally assimilating data with models of emerging infectious disease. Because disease transmission was between humans and because it emerged in nations with relatively well-developed public health programs, data on the progression of SARS were far more extensive than could be expected for many diseases. In other words, the data were about as good as they get for outbreaks of a new disease. Despite the quality of the data, the models that generated inference to understand the epidemic were relatively simple, with few parameters and many simplifying assumptions. This simplicity did not prevent them from being useful, but it does caution excessive optimism for data–model assimilation efforts that involve sparse data and greater inherent uncertainty.

Dengue fever

Dengue fever is caused by a viral infection (family Flaviviridae) that is transmitted between mosquito vectors and humans. Although the ancestral transmission cycle of dengue viruses likely included forest primates, human populations are generally believed to be the primary reservoirs for current dengue outbreaks (Gubler 1998). The predominant vector transmitting dengue to humans is the mosquito *Aedes aegypti*, a species that lives in close proximity to humans and often breeds in man-made containers. The acute illness manifests as headache, fever, exhaustion, rash, and muscle and joint pains. There are four serotypes of dengue viruses and infection with heterologous serotypes in fairly close temporal proximity may produce more severe disease (dengue shock syndrome and dengue hemorrhagic fever). Although dengue has been endemic in Asia since at least since the 18th century, expansion of the mosquito vectors and global movement of human populations have greatly changed the distribution and intensity of dengue fever, which is now considered an emergent or reemerging disease across Africa, Australia, and the Americas (Gubler 1998). It is estimated that approximately10 million cases of uncomplicated dengue occur annually, with 500 000 cases of dengue hemorrhagic fever or shock syndrome primarily from tropical and subtropical regions.

As with most vector-borne pathogens, the ability to successfully predict dengue incidence is generally be-lieved to rely on how well we can predict vector abundance. However it is important to note that because humans are the primary dengue host, human behavior may be as important as vector abundance for accurate forecasts of pathogen amplification and disease risk. Models have been useful for understanding dengue transmission and for evaluating control strategies. Dynamic compartmental models for vector-borne diseases affecting humans are generally similar to the SEIR-type mathematical models described above, but with the added complexity of a vector population (MacDonald 1957, Anderson and May 1992). For instance, Burattini et al. (2008) linked a series of nine differential equations to characterize dengue infection in human hosts and mosquitoes. The authors used this simulation model to evaluate control options and showed that even a dramatic crash in the size of adult mosquito populations may not be sufficient to stop ongoing dengue epidemics (Burattini et al. 2008).

Much of the dengue research has focused on understanding how disease incidence is related to environmental variables that influence the abundance of mosquito populations (Peterson et al. 2005, Kearney et al. 2009, Wu et al. 2009). For example, climatic conditions are important predictors of dengue amplification due to the tight relationship between mosquito life cycles and temperature and precipitation. Mosquito development from egg to biting adult, survival rate at each life stage, and feeding behavior are all strongly dependent on temperature (Watts et al. 1987, Rueda et al. 1990, Focks et al. 1993). Furthermore, the extrinsic incubation period (EIP) of viral replication within the mosquito depends nonlinearly on temperature and is maximal at temperatures above 34°C, while the virus fails to be transmitted at temperatures below 30°C (Watts et al. 1987) in the laboratory. Due to the intimate connection between mosquito population dynamics, viral amplification, and climatic variables, much effort has been invested in developing methods to integrate climate data with observations of mosquitoes and dengue incidence. However, while climate data are often available and prolific at broad spatial scales, mosquito data are less abundant and are rarely characterized at the scale of climate data. Local and transient weather patterns and interactions with land cover are likely to control real-time mosquito abundances and can be difficult to ascertain from the available climate data. A formal framework for accommodating mismatches in scale in order to coherently integrate process models, historical disease incidence data, experimental results, and diverse environmental data would greatly advance understanding of the ecological characteristics that control dengue outbreaks, as well as many other vector-borne diseases.

Phenomenological models can often establish clear associations between a suite of plausible climate measurements and mosquito abundance (e.g., Reisen et al. 2008). However, even when statistically significant

associations are established, the resulting models do not always lead to accurate forecasts of vector population dynamics (Ailes 1998) or pathogen amplification. Many models use a common model structure (e.g., Ailes 1998):

$$\mathbf{X}(t) = B_1 Y_1(t - n_1) + B_2 Y_2(t - n_2) + \ldots + B_i Y_i(t - n_i)$$

where $\mathbf{X}(t)$ is vector abundance at time t and the Y's are environmental conditions at various but fixed time lags $(n_1, n_2, \ldots n_i)$. Many of the studies that have used climate variables to predict mosquito abundance choose a specific time lag, such as precipitation two weeks prior to the abundance estimate. Often these choices are made based on biological understanding (i.e., average mosquito development duration is two weeks) but fail to incorporate the natural variation behind published means. A focus on building parsimonious models and assuming a constant effect (B_i) at a fixed time interval in the past may lead to useful inference but can also result in error accumulation and propagation forward in predictions.

Visualization and modeling approaches that facilitate the exploration of biologically relevant hypothesis space are critical tools for understanding complex systems like infectious disease (e.g., Plowright et al. 2008). Graphical methods that include cross correlation mapping to visualize support for a range of plausible relationships between vector populations and environmental variation are effective tools for evaluating appropriate environmental space. Such approaches demonstrate that there can be extensive time intervals, on the order of weeks, when the relationship between the leading environmental conditions (i.e., temperature and precipitation) and the subsequent vector population abundances are approximately constant (Curriero et al. 2005, Shone et al. 2006).

Phenomenological models have also been used to link dengue incidence to climate variables. As was the case with the relationship between mosquito abundance and climate, statistically significant relationships do not always indicate sufficient model structure for accurate forecasts of disease incidence. For example, Nakhapakorn and Tripathi (2005) found that incidence of dengue was negatively related to temperature, which only makes biological sense at the upper temperature limits of *Aedes aegypti* survival, an uncommon condition during their study. Another study found that lagged precipitation was a significant predictor of early-wet-season dengue incidence but was not associated with a significant change in *Aedes aegypti* abundance (Foo et al. 1985), the theoretical mechanism for why precipitation would increase dengue transmission. Even within the same study, spatial location seems to play an important role in when and how climate variables are associated with dengue transmission (e.g., Promprou et al. 2005, Johansson et al. 2009).

A hierarchical approach taken by Johansson and colleagues (2009) to evaluate dengue transmission in Puerto Rico examined the association between climate and dengue transmission using a hierarchically structured model to explicitly evaluate both the short-term association between dengue incidence and monthly weather variation and the "global" influence of regional climate on local short-term associations. The authors also included an adaptive cubic spline component to control for the inherent (and potentially confounding) seasonality in weather variables. Johansson et al. (2009) demonstrated that spatial heterogeneity in the shorter-term relationships between weather and dengue transmission might best be understood within the context of spatial patterns in long-term climate characteristics. For example, the cumulative effect of temperature on dengue incidence was greatest in the cooler mountainous areas, while precipitation was most important in the dry southwestern coastal region. Neither temperature nor precipitation was an important predictor of dengue in the region where they were already normally high. The two-stage analysis employed by Johansson et al. (2009) to estimate average global effects and determinants of local variation is an approach that could be applied to a number of other environmentally mediated diseases. This also highlights how simple amalgamation of many studies across space could generate misleading or uninterpretable results.

Although dengue viruses have likely caused illness for much of human history, they have expanded in distribution and had the greatest impact on human populations in just the past few decades. Changing climate and increased globalization seem poised to support further intensification of dengue outbreaks (Kearney et al. 2009, Wu et al. 2009). Unlike the SARS example above, dengue control strategies must go beyond quarantining infected humans to direct manipulation of mosquito populations. This section demonstrates that a focus on model parsimony may help identify and define important relationships but ecological disease systems are ultimately complex and ignoring stochasticity in natural processes or estimation uncertainty is detrimental to forecasting efforts.

Lyme disease

Lyme disease is a zoonotic disease caused by spirochetal bacteria (*Borrelia burgdorferi*) and transmitted by several species of *Ixodes* ticks. Lyme disease was first recognized and formally described in the late 1970s following a cluster of juvenile arthritis cases in Lyme, Connecticut, USA (Barbour and Fish 1993). In the ensuing 30 years, Lyme disease has spread geographically from a coastal New England focus throughout much of the northeastern and mid-Atlantic regions, and adjacent southeastern Canada. The pathogen has also undergone dispersal from historical ranges in the upper Midwest of the United States and in Europe and human incidence rates (number of cases per capita per year) have increased dramatically throughout its range.

Human Lyme disease risk is generally correlated with the population density of infected ticks (specifically the

nymphal stage), although some studies indicate that the infection prevalence in the tick population (proportion of ticks infected) better predicts human incidence (Connally et al. 2006, Walk et al. 2009). *Borrelia* can persist in several wildlife (zoonotic) hosts and amplify whenever conditions allow. Data and understanding of tick biology and the ecological interactions between ticks and their zoonotic hosts are critical to the study of human Lyme disease risk. Humans are not the predominant host in the Lyme disease system (see Fig. 1c). Humans cannot infect each other or infect ticks. In eastern and central North America, the sole vector of Lyme disease is the blacklegged tick, *I. scapularis*. *Ixodes scapularis* ticks feed on vertebrate hosts three times, once each as a larva, nymph, and adult. Larval and nymphal ticks are extreme host generalists that will feed readily from dozens of potential zoonotic host species (Keirans et al. 1996, LoGiudice et al. 2003). Ticks that feed from an infected host might acquire an infection, which can persist and be transmitted during the nymphal (if infected as larvae) and adult blood meals. The life cycle of the blacklegged tick typically lasts for two years and includes three blood meals that are often taken from three different host species.

Dynamic compartmental models have been used to evaluate infectious dynamics in this system as well (Ogden et al. 2007), with similar benefits and limitations as described in the previous sections. Ogden et al. (2007) prioritized the importance of tick population dynamics and included a model component to simulate blacklegged tick population change by incorporating rates of survival, development, and host-finding of all life stages, as well as reproduction by adults. The component model for tick population dynamics (Ogden et al. 2005) included 48 different parameters for tick vital rates and two for host availability (white-footed mice and white-tailed deer), and was able to characterize locally observed population fluctuations of larval, nymphal, and adult ticks reasonably well. However, the resultant inference derived from even this complex, ecologically thoughtful model relied heavily on key assumptions. For example, tick population dynamics were most sensitive to mortality rates of immatures (larvae and nymphs), which were assumed to be time-invariant constants. Because of this assumption, warmer conditions that accelerate development automatically resulted in higher densities simply because ticks spent less time subject to constant daily probability of mortality (Ogden et al. 2005). These models thus predict northward expansion of blacklegged ticks into Canada as a consequence of anthropogenic climate change. However, they are unable to accommodate nonlinear or spatially heterogeneous relationships between climatic conditions and tick demography. For instance, a warmer winter climate could plausibly promote tick survival and population growth in northern latitudes while simultaneously reducing survival and population growth in southern latitudes. As seen in both the previous case study examples, heterogeneity in transmission rates in both space and time is a fundamental determinant of pathogen amplification and transmission that is often overlooked in even the most mechanistically complex modeling approaches.

Phenomenological models have been constructed to examine the spatial and temporal distribution of blacklegged ticks based on abiotic variables that can be remotely sensed. In general, these approaches use presence/absence type data for blacklegged ticks at one location to construct current distribution maps in order to delineate areas with ticks from those without ticks (Estrada-Pena 2002, Brownstein et al. 2003, 2005). Many of these models can capture current tick distribution with high sensitivity and specificity and are often used to predict the spread of Lyme disease risk into all abiotically suitable tick habitats that occur outside the current range. For a pathogen that is in the process of expanding in range (as is *Borrelia*), the abiotic conditions that match its current range might represent only a small subset of those that permit its existence. When this is true, any given snapshot of these conditions will be highly conservative in predicting the future range. Phenomenological models often fail to meaningfully address biological mechanisms that might affect tick populations and Lyme disease risk. The abiotic variables that enter models are selected not from a priori expectations but from the set of available data that can be produced from remote sensors. Even when high sensitivity and specificity can be achieved, the predictor variables are often biologically uninterpretable and not necessarily sufficient for accurate forecasting. A complex, fourth-order polynomial with minimum winter temperature was found to be the best predictor of habitat suitability for blacklegged ticks in Brownstein et al. (2003). Such modeled relationships can be highly informative for generating and defining the hypotheses regarding the complex ecological interactions and mechanisms driving spatiotemporal intensity of Lyme disease, but they are not always sufficient for evaluating those hypotheses or for generating accurate forecasts.

The ecological complexity in the Lyme disease system may best be approached through a comprehensive investigation of multiple biologically relevant model pathways explaining how, where, and when pathogen amplification occurs in zoonotic hosts and spills-over into human populations. There now exists a large library of related experiments and field observation studies that together generate strong inference (e.g., Plowright et al. 2008) regarding the complex ecological interactions that lead to spatiotemporal variability in Lyme disease (e.g., Schmidt and Ostfeld 2001, Schauber et al. 2005, Ostfeld et al. 2006). Synthesis of data sources from a broad group of researchers has shown that the timing and composition of host community dynamics are critical components of Lyme disease amplification and are best understood within the hosts' broader trophic interactions (Fig. 2). The diverse expertise and field data
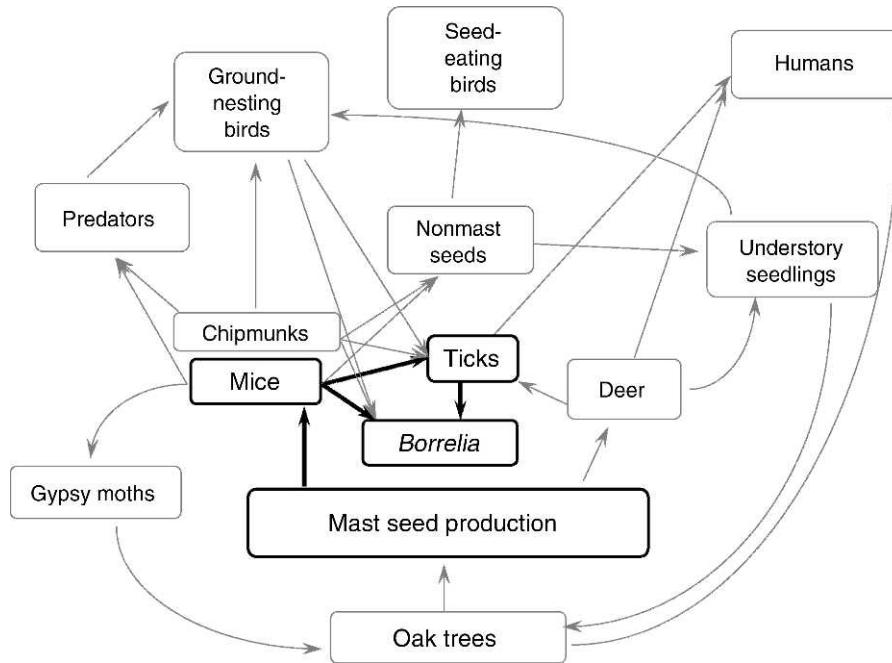
FIG. 2. The ecological interactions that define interannual variability in human infection with *Borrelia burgdorferi* in the northeastern United States (bold lines and boxes) are best understood when considered within the broader community (gray lines) of trophic interactions.

required to define and understand the relationships in Fig. 2 are clearly only possible through collaborative efforts and careful data–model assimilation. White-footed mice are the most competent reservoir for *Borrelia burgdorferi* and also the host on which the tick vector is most likely to survive while attempting a blood meal (Keesing et al., 2009). Consequently, high abundance of white-footed mice in midsummer when larval ticks are most active results in ample opportunities for blood meals that are likely to both infect larval ticks and promote survival to the nymphal stage. The population density of nymphs and infection prevalence with *B. burgdorferi* are correlated with population density of white-footed mice in the previous summer (Ostfeld et al. 2001, 2006). Summer population density of white-footed mice, in turn, is determined largely by acorn production (genus *Quercus*) the prior autumn (Elkinton et al. 1996, Ostfeld et al. 1996, Jones et al. 1998). Mouse populations with access to abundant seed resources have high overwinter survival rates, begin breeding earlier in the spring, and reach higher densities at the time ticks are seeking hosts (Wolff 1996, Ostfeld et al. 2006). Because of the strong trophic links between acorns and mice and between mice and both ticks and *Borrelia*, acorn production provides a valuable leading indicator of incidence of Lyme disease in humans two summers later (Schauber et al. 2005). Because acorn production by populations of oaks can be synchronized over areas of hundreds to thousands of square kilometers (Liebhold et al. 2004), the predictive power of acorn production at

any one site might be high even hundreds of kilometers away (Schauber et al. 2005).

There are a number of tick-borne diseases affecting humans globally and *Ixodes scapularis* carries at least four known zoonotic pathogens (Swanson and Norris 2007). Models that can characterize population fluctuations and infection rates for *Ixodes* ticks can advance understanding of multiple disease systems. Having detailed understanding of why and when tick populations increase and how this relates to human exposures is fundamental to disease management and forecasting. The Lyme disease system presents a clear example of how ecological understanding can be developed through an approach including both empirical field work and mathematical process models and how this broader understanding can inform local and regional public health protection. This system also illustrates that data are still needed to better understand host and vector abundances and demographic processes, at both fine and broad spatial scales. And finally, a formal framework to assimilate these diverse data with biological process models is required to elucidate the spatial determinants of current risk and to forecast future risk in a changing environment. Models that are supported by the abiotic and relevant biotic (e.g., seed production, host availability) data could be used in combination with land-use change models and regional climate change scenarios to predict specific changes in the distribution of ticks and the pathogens they transmit.

West Nile virus

West Nile virus (WNV) is an emergent zoonotic disease caused by a viral infection (family Flaviviridae) that was first detected in the western hemisphere in 1999, where it caused clusters of disease and mortality in human and bird populations in and around New York City (Lanciotti et al. 1999). West Nile virus spread rapidly across North America and since 2002, there have been an average 4089 ± 2738 (mean ± SD) human cases reported and 156 ± 89 deaths per year (data *available online*).[6] Persistence of WNV in the environment requires a continuous bird–mosquito–bird cycle that includes amplification of the virus in both the avian hosts and mosquito vectors. Several species of birds may be competent hosts (i.e., can be infected and reinfect a vector; Komar et al. 2003, Kilpatrick et al. 2007) and a subset of those are susceptible to WNV disease (Komar et al. 2003, LaDeau et al. 2007). Likewise, there are several mosquito species that may play important roles in amplifying and transmitting the virus between birds and to humans (Andreadis et al. 2004, Kilpatrick et al. 2005, Turell et al. 2005). Humans and other large mammals are not necessary for WNV persistence or amplification; they may be infected with WNV if bitten by an infectious mosquito but do not produce sufficient viremia to reinfect a mosquito. As few as one in roughly 100–200 undiagnosed WNV infections is ever reported (Tonjes 2008), and although human incidence is a readily available data source, disease risk cannot be effectively understood by evaluation of recorded human incidence alone.

The example of WNV in North America exemplifies the challenges associated with forecasting annual and spatial outbreaks of a newly emergent disease. Prior to 1999, West Nile virus had only been recorded in the eastern hemisphere where it caused sporadic and short-lived epidemics in humans and horses since the 1930s (Smithburn et al. 1940, Hubalek and Halouzka 1999). While the intensity of disease outbreaks in both humans and birds is variable across years and regions (LaDeau et al. 2008), at some locations WNV infections have now recurred annually for over a decade (see footnote 6). Understanding how the North American landscape and climate have facilitated the rapid spread and persistent amplification of WNV is essential to understand the pathogen's impact on avian communities, develop control strategies, and forecast human risk.

West Nile virus is now endemic across much of North America (Kilpatrick et al. 2007). That it persists from year to year in many locations confirms its continuous presence in either host or vector species throughout the year. Although there are numerous hypotheses regarding what extrinsic (e.g., weather, habitat) and intrinsic (e.g., host–vector population dynamics) factors might

drive spatiotemporal heterogeneity in WNV persistence, amplification and disease outbreaks (e.g., Kilpatrick et al. 2006b, c, Day and Shaman 2008, Platonov et al. 2008, LaDeau et al. 2010), the ability to capture the processes that determine annual and spatial intensity of WNV amplification and forecast disease outbreaks remains elusive. There is a growing library of research detailing the individual components of the WNV disease system: varying competence (ability to become infected and transmit infection) among mosquito vector species (Kilpatrick et al. 2005, Turell et al. 2005, Reisen et al. 2006b) and avian host species (Komar et al. 2003), strain-specific relationships between temperature and viral replication rates within mosquitoes (Reisen et al. 2006a, Kilpatrick et al. 2008), vector feeding preference (Andreadis et al. 2004, Kilpatrick et al. 2006b, c), host immunity (Fang and Reisen 2006), spatial heterogeneity in host seroprevalence (Komar et al. 2005, Bradley et al. 2008), and spatiotemporal records of human incidence (see footnote 6). The next crucial step is to develop a coherent framework to integrate these distinct components into a common model of the WNV system.

Correlative and phenomenological methods have been effective at generating quick inference from observations in order to guide public health efforts and warn of impending human risk. For example, mosquito density (Tachiiri et al. 2006) and early-season vector infection rates (Brownstein et al. 2004) both seem to indicate increased human risk of WNV infection in late summer. Evidence of WNV infections in avian communities in early and midsummer are also potential predictors of local spillover of WNV into human populations (Eidson et al. 2001, Guptill et al. 2003, Nielsen and Reisen 2007). As in the earlier case studies, the assumption that vector density is positively associated with human disease risk is widely accepted. Thus, when the early-warning signals mentioned above indicate that WNV amplification is occurring, mosquito control programs are most often enacted to reduce local mosquito abundances (e.g., Carney et al. 2008, Lothrop et al. 2008). Experimentation to test the efficacy of these early-warning signals or the importance of the biological processes they represent could potentially leave humans at risk and is not ethically feasible. As is the case with most infectious disease research, more creative methods are needed to synthesize the available data with alternative process models to evaluate mechanistic assumptions that characterize disease transmission and define how human risk is minimized.

As with the other case studies presented, dynamic compartmental models are a powerful tool for organizing our understanding of the WNV system, but have yet to fully integrate model structure with the available data. Wonham and colleagues (2006) reviewed several published compartmental models of WNV transmission, all of which used a slightly different structure of differential equations to represent transitions between susceptible and infected hosts and susceptible and

infected vectors. The authors found that each model was highly sensitive to the assumptions regarding the relationship between transmission rate and abundances of hosts and vectors (i.e., density dependent, frequency dependent, constant) and that allowing animal hosts to transition from infected to immune (versus to susceptible or dead) dramatically changed the predicted R_0 for the disease (Wonham et al. 2006). Unfortunately, data on host recovery and immunity is rare (but see Komar et al. 2003, Fang and Reisen 2006, Nemeth et al. 2008) and potential heterogeneities among recovery rates for species and regions need to be obtained before reliable predictions of R_0 can be made. Jiang et al. (2009) also used the WNV system to demonstrate that parameter estimates (e.g., of R_0) from standard dynamic compartmental models are also sensitive to starting conditions (i.e., the initial numbers of infected birds or mosquitoes).

Similar to the SARS case, the WNV example highlights the need for formal data–model assimilation in real-time to manage an epidemic as it occurs. Identifying realistic starting conditions or "true" transmission structure may be like hitting a moving target, as abundances and infection rates vary in time and space, and these limitations emphasize the importance of a data–model integration approach that is focused on exploration of alternative pathways over techniques that estimate support for yes/no tests. This case study again underlines that even complex models are only as good as the assumptions. When process error is present or when a parameter estimate from the literature fails to reflect empirical findings at the spatial scale of interest then simulation outcomes will not be realistic.

SYNTHESIS

Estimating latent parameters for processes that cannot be observed is a persistent challenge in the study of infectious disease. There is a common need across the case study examples for a formal framework to coherently integrate all relevant data with current and growing understanding of the system ecology, often in real time. Each case presented demonstrates continued need for data collected at relevant scales to better understand the importance of spatiotemporal heterogeneity and the mechanisms that define it. From the discussion of SARS, it is clear that public health risk can be effectively managed with minimal understanding of zoonotic reservoirs *if* the transmission pathway important to human epidemics is maintained by direct human-to-human contact (Fig. 1b). However, lack of ecological understanding of the zoonotic reservoirs and the variables that allowed for the initial spillover transmission to humans may mean that we have little ability to forecast when and where SARS or similar pathogens may jump to humans in the future. When zoonotic reservoirs and vector species are important for pathogen persistence and amplification in the environment (Fig. 1b and c), it becomes essential to coherently integrate

data on vector and host demography as well as infectious dynamics.

In general, we support the call by Plowright et al. (2008) for further development of strong inference approaches in infectious disease research. More specifically, we propose a formal framework to integrate all relevant information with a model structure that allows for latent processes, acknowledges incomplete biological understanding and has the flexibility to update process understanding as research advances. Next we describe a conceptual framework for data–model fusion in infectious disease research that addresses three primary challenges that we have identified in our discussion here, (1) integrate multiple data sources and spatial scales to inform latent processes, (2) partition uncertainty in process and observation models, and (3) explicitly build upon existing ecological and epidemiological understanding.

Conceptual framework for data–model fusion

We propose that progress in understanding and predicting zoonotic and vector-borne diseases can be made by using hierarchical or conditional statistical modeling (see Fig. 3) to link together three kinds of models: (1) data models for observations, (2) process models of disease dynamics that include critical latent parameters such as transmission rates and R_0, and (3) spatial models for mapping and predicting human disease prevalence. Each of these components is already well established. They have not, however, all been brought together in the context of zoonotic disease for data integration or epidemic forecasting. The conceptual framework presented in Fig. 3 presents opportunities for pooling information from multiple sources while weighting these in a consistent way based on their information content. Importantly, because the component model parts are assembled together conditionally, adding or removing parts and comparing alternative models can be straightforward (Clark 2005, Cressie et al. 2009). We have constructed a framework that combines the advantages of statistical models (i.e., explicit incorporation of data and quantification of uncertainty) with the key features of more traditional epidemiological models (i.e., incorporating mechanistic insights and exploring system dynamics like thresholds for emergence and persistence). Furthermore, the framework characterized by Fig. 3 explicitly relates ecological disease dynamics in zoonotic hosts and/or vectors to human disease prevalence in a stochastic spatial and temporal context.

The framework depicted by Fig. 3 may seem dauntingly complex. However, zoonotic and vector-borne infectious diseases are complex systems and failure to acknowledge the inherent heterogeneity and uncertainty is a major roadblock to advances in ecological forecasting. We are attempting to lay out a general structure that can accommodate the key modeling goals and lessons from the case studies, while
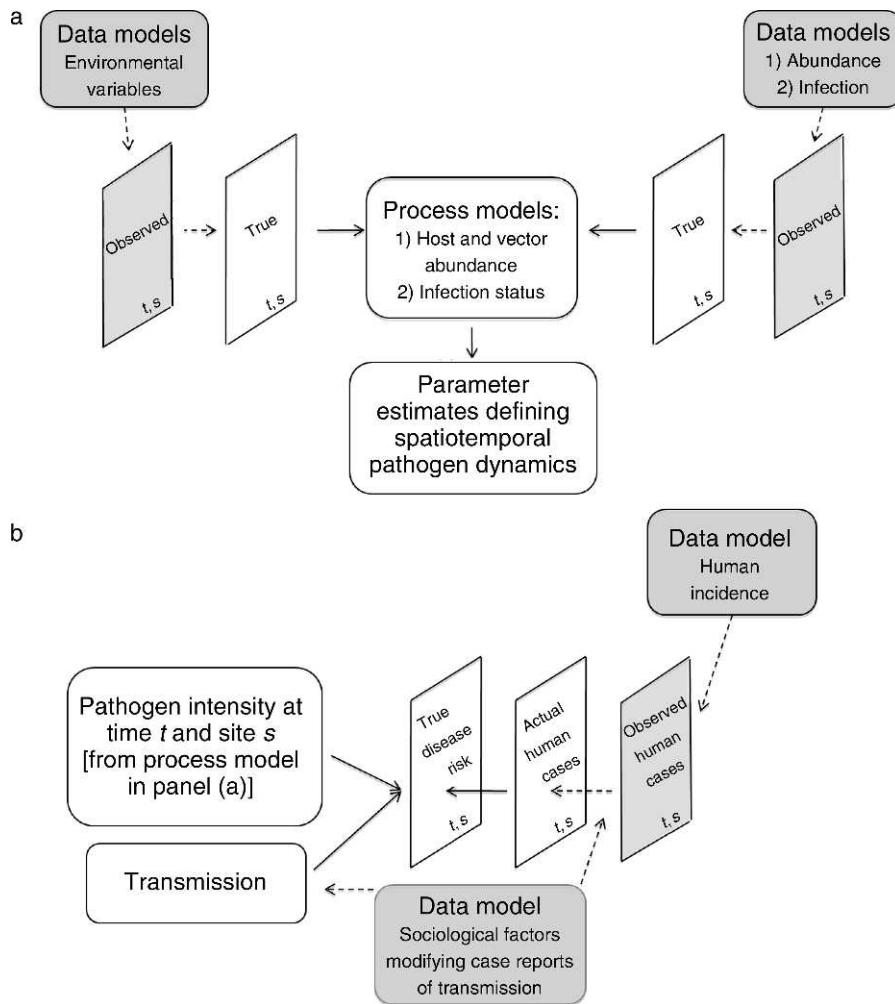
FIG. 3. A conceptual framework for inference and forecasting vector-borne zoonotic disease. (a) Modeling zoonotic/ecological components requires data models (shaded boxes, dashed lines) to relate observations to "true" latent variables and process models, shown in solid lines and unfilled boxes. Relevant data sources include environmental data layers (e.g., temperature, precipitation, habitat measures), and observations of host and vector abundances (counts), as well as observed infection rates. (b) An additional component model can be used to link processes (and uncertainties) in panel (a), to refine inference and forecasting of human disease risk. Data sources include observed human incidence and the data layers that describe bias in human case reporting (e.g., socioeconomic) or alter transmission and infection dynamics (e.g., recreational activities, immunosuppression).

providing a context for classifying and comparing models that implement different components of this idealized general model. While the overall conceptual framework is complex, it is built by combining tools that already exist in epidemiology, ecology, and statistics. In just the past decade great strides have been made that address each of the disease modeling goals discussed above. There have been methodological advances in spatiotemporal modeling (Neubert and Caswell 2000, Banerjee et al. 2004), disease mapping (Biggeri et al. 2006, Jin et al. 2007), dynamic spread models (Hooten and Wikle 2008), integrating data into dynamic models (Clark and Bjornstad 2004, Hobbs and Hilborn 2006, He et al. 2010), and in developing flexible "data models" to accommodate error structures and bias in observations (Congdon 2003, Ogle and Barber 2008, Cressie

et al. 2009). Model choice and validation for complex, hierarchical models remains a topic of current statistical research and we will not address it in detail here. However, in addition to recommending the "strong inference" approaches (e.g., Plowright et al. 2008), we stress that model structure, assumptions, and data support should be evaluated carefully at each stage. Specific model selection and choice methodologies will depend on the model structure and the research goals (Clark et al. 2007, Craigmile et al. 2009, He et al. 2010).

Data models.—A large part of the challenge in modeling zoonotic diseases is the need for diverse and extensive data, including host and vector abundances, demographic and infection rates, human case report records, climate, and land use data. Each data type can contribute toward understanding and predicting the

latent parameters critical to disease dynamics, but each is also subject to potentially large measurement error. One of the major advantages to using a conditional hierarchical modeling approach is the ability to flexibly accommodate such data types and their different sources of measurement error by explicitly characterizing the relationship between each type of data and the latent parameters they inform (dashed arrows in Fig. 3). Examples of such data models include likelihood-based mark–recapture models, which have been extended to incorporate among-individual and spatial and temporal variation (Royle and Link 2002, Royle 2009), and error rate models for infection assays (Joseph et al. 1995). Incidence data on human disease are also subject to error (e.g., He et al. 2010). Lyme disease reports, for example, may be low in areas with recent disease emergence and rise as physician awareness increases (Young 1998, Chen et al. 2006). The framework in Fig. 3 allows for such biases to be systematically dealt with; observed disease can be related to true incidence through a data model with a spatially and temporally varying latent parameter to characterize the probability that a case is reported given that it actually occurred. Furthermore, these probability parameters could be regressed on factors identified as important in previous studies, such as time since first disease report in the county, distance from a hospital, and human population density, potentially allowing us to also learn about the sources of reporting error and bias.

Process models.—The process functions may resemble the same dynamic compartmental models described in the case studies above. However, a key difference is that the process models in Fig. 3 are both formally fused with data and structured to allow for parameter stochasticity. For example, in a vector-borne disease system we could use a susceptible-infected (SI) model to characterize pathogen movement through susceptible uninfected and infected vectors, and relate the infection process to surveys of vector abundance and infection. The SI model would include a stochastic and potentially spatially and temporally varying transmission parameter, and be implemented as a state-space model in which populations of susceptible and infected vectors are related through data error models to the raw survey data. A more detailed model could then add population structure and could also include an analogous model for primary reservoir hosts. When these process models capture key dynamics of the disease's epidemiology, they will allow us to explore thresholds for disease persistence in a population and test how close to such thresholds real populations tend to be. In Lyme disease, for example, rate of infection of the tick vector appears to be correlated well with local tick abundance (Chen et al. 2006). This makes sense mechanistically: in a standard SI model, higher tick abundance means more bites per host, (assuming constant host density), which increases the proportion of infected hosts because each host is "sampling" multiple ticks and has a greater chance of

encountering an infected one. Higher host infection rates will feed back to increase the rates of infection in ticks, because there will be a higher probability that each tick's blood meal comes from an infected host (Caraco et al. 2002). So the proportion of infected ticks and tick abundance are dynamically related both directly and through feedbacks, and their nonlinear covariation will inform the process model for infection dynamics to provide inference about the critical latent variables of transmission and amplification rates. As with current dynamic compartment models, the precision and accuracy of the parameters will be directly related to the biological assumptions that structure the model and to the data used to construct the model. However, the structure of this framework will allow us to evaluate data and process structure explicitly through the partitioned process and observation error terms.

The hierarchical approach, especially if implemented in a Bayesian framework, enables these models to accommodate spatial and temporal rate variation directly. For example, a standard epidemiological model can be modified to relate key parameters to environmental factors, and to allow the relationships between environment (e.g., minimum winter temperature) and parameters to vary spatially (Johansson et al. 2009). As with many data models, spatially varying parameter models have been extensively studied in the statistics field, but applying them to zoonotic and vector-borne disease and in combination with other hierarchical levels is novel and could make these models more useful for managing and forecasting disease.

Spatial models of human disease incidence.—In spatial models of disease occurrence, the disease reports arise as a stochastic process from a latent intensity surface, with spatial structure in the disease process, the measurement process, or both. In the conceptual framework we are proposing, the relative risk of human disease occurrence is represented by a latent intensity surface that varies as a function of a few key parameters from the process models characterizing infection in vector and host populations. This disease intensity surface can be then related to hypothesized causal factors such as density and infection rates of vectors, and with environmental factors, including climate and land use, as well as population and behavior data, including contact networks (Farnsworth et al. 2006). For spatially aggregated or areal unit data (e.g., county-level disease counts) standard spatial modeling tools include conditional autoregressive (CAR) models that allow correlation among areas defined as "neighboring" (Banerjee et al. 2004).

By testing alternative model structures informed by existing information, we should be able to use the proposed framework to identify key links between population processes (the disease organism, the vector, the animal hosts), extrinsic environmental factors (land use, climate), and human infections. The spatial flexibility of the models, to the extent sufficient data

are available to support this, will be useful for exploring variation in these relationships in space and time, and for distinguishing relationships that are relatively constant from others that change with environmental gradients or population characteristics. Using models in this exploratory way brings us back around to where epidemiology began with John Snow and the water pump, now with better tools but faced with the harder challenge of finding the key sources and drivers of infection for complex emerging diseases.

CONCLUSIONS

Our review flags areas where improved data-gathering would substantially improve our ability to model disease dynamics, make inferences about latent parameters, and predict human disease incidence patterns. These data needs are important regardless of the modeling framework. To improve the epidemiological/ecological core process models in our conceptual framework, it will be important to gather ecological data on host and vector population structure and behavior in order to better quantify basic empirical parameters like population growth rates, as well as choosing the correct functional forms for latent parameters (e.g., Wonham et al. 2006). The case studies discussed here demonstrate that more comprehensive spatial and temporal information on vector and host abundances and infection rates would greatly improve the capacity for forecasting ecological dynamics that influence pathogen amplification and spillover to humans.

A persistent issue with developing timely inference for management and forecasts of EID outbreaks is that the data needed to fully parameterize process models (e.g., incidence, transmission rates, mortality rates) are rare until the disease reaches epidemic proportions. Better understanding of the population dynamics of common host and vector species for pathogens already present in North America (i.e., birds, mice, mosquitoes, ticks) is vital to managing these diseases but may also be important for limiting EIDs still to come. We echo many other researchers in our call for support of long term monitoring and interdisciplinary collaboration (e.g., Crowl et al. 2008). Public health departments, veterinarians, ecologists and epidemiologists all gather distinctive information that should be integrated (with input from statisticians, mathematicians, and computer scientists) in a model framework like the one proposed here, to produce better mechanistic insights and, eventually, better forecasts. Given strong constraints on research on EID (no human experiments, many parameters and little data), and the urgent need to make progress, this kind of fusion of data and expertise via hierarchical models should prove an indispensable tool.

ACKNOWLEDGMENTS

LITERATURE CITED

Ailes, M. C. 1998. Failure to predict abundance of saltmarsh mosquitoes *Aedes sollicitans* and *A-taeniorhynchus* (Diptera: Culicidae) by using variables of tide and weather. Journal of Medical Entomology 35:200–204.

Anderson, R. M., C. Fraser, A. C. Ghani, C. A. Donnelly, S. Riley, N. M. Ferguson, G. M. Leung, T. H. Lam, and A. J. Hedley. 2004. Epidemiology, transmission dynamics and control of SARS: the 2002–2003 epidemic. Philosophical Transactions of the Royal Society Series B 359:1091–1105.

Anderson, R. M., and R. M. May. 1985. Vaccination and herd-immunity to infectious-diseases. Nature 318:323–329.

Anderson, R. M., and R. M. May. 1992. Infectious diseases of humans: dynamics and control. Oxford University Press, Oxford, UK.

Andreadis, T. G., J. F. Anderson, C. R. Vossbrinck, and A. J. Main. 2004. Epidemiology of West Nile virus in Connecticut: a five-year analysis of mosquito data 1999–2003. Vector-Borne and Zoonotic Diseases 4:360–378.

Antolin, M. F. 2008. Unpacking beta: within-host dynamics and the evolutionary ecology of pathogen transmission. Annual Review of Ecology Evolution and Systematics 39:415–437.

Baadsgaard, N. P., S. Hojsgaard, Y. T. Grohn, and Y. H. Schukken. 2004. Forecasting clinical disease in pigs: comparing a naive and a Bayesian approach. Preventive Veterinary Medicine 64:85–100.

Banerjee, S., B. P. Carlin, and A. E. Gelfand. 2004. Hierarchical modeling and analysis for spatial data. CRC Press, Boca Raton, Florida, USA.

Barbour, A. G., and D. Fish. 1993. The biological and social phenomenon of Lyme disease. Science 260:1610–1616.

Bauch, C. T., J. O. Lloyd-Smith, M. P. Coffee, and A. P. Galvani. 2005. Dynamically modeling SARS and other newly emerging respiratory illnesses—past, present, and future. Epidemiology 16:791–801.

Begon, M., S. Telfer, M. J. Smith, S. Burthe, S. Paterson, and X. Lambin. 2009. Seasonal host dynamics drive the timing of recurrent epidemics in a wildlife population. Proceedings of the Royal Society B 276:1603–1610.

Biggeri, A., E. Dreassi, D. Catelan, L. Rinaldi, C. Lagazio, and G. Cringoli. 2006. Disease mapping in veterinary epidemiology: a Bayesian geostatistical approach. Statistical Methods in Medical Research 15:337–352.

Bjornstad, O. N., B. F. Finkenstadt, and B. T. Grenfell. 2002. Dynamics of measles epidemics: estimating scaling of transmission rates using a time series SIR model. Ecological Monographs 72:169–184.

Bolker, B., and B. Grenfell. 1995. Space, persistence and dynamics of measles epidemics. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences 348:309–320.

Bradley, C. A., S. E. J. Gibbs, and S. Altizer. 2008. Urban land use predicts West Nile virus exposure in songbirds. Ecological Applications 18:1083–1092.

Britton, T., and D. Lindenstrand. 2009. Epidemic modelling: aspects where stochasticity matters. Mathematical Biosciences 222:109–116.

Brown, H., M. Duik-Wasser, T. Andreadis, and D. Fish. 2008. Remotely-sensed vegetation indices identify mosquito clusters of West Nile virus vectors in an urban landscape in the northeastern United States. Vector-Borne and Zoonotic Diseases 8:197–206.

Brownstein, J. S., T. R. Holford, and D. Fish. 2003. A climate-based model predicts the spatial distribution of the Lyme disease vector *Ixodes scapularis* in the United States. Environmental Health Perspectives 111:1152–1157.

Brownstein, J. S., T. R. Holford, and D. Fish. 2004. Enhancing West Nile virus surveillance, United States. Emerging Infectious Diseases 10:1129–1133.

Brownstein, J. S., D. K. Skelly, T. R. Holford, and D. Fish. 2005. Forest fragmentation predicts local scale heterogeneity of Lyme disease risk. Oecologia 146:469–475.

Burattini, M. N., M. Chen, A. Chow, F. A. B. Coutinho, K. T. Goh, L. F. Lopez, S. Ma, and E. Massad. 2008. Modelling the control strategies against dengue in Singapore. Epidemiology and Infection 136:309–319.

Calder, C., M. Lavine, P. Muller, and J. S. Clark. 2003. Incorporating multiple sources of stochasticity into dynamic population models. Ecology 84:1395–1402.

Caraco, T., S. Glavanakov, G. Chen, J. E. Flaherty, T. K. Ohsumi, and B. K. Szymanski. 2002. Stage-structured infection transmission and a spatial epidemic: a model for Lyme disease. American Naturalist 160:348–359.

Carney, R. M., S. Husted, C. Jean, C. Glaser, and V. Kramer. 2008. Efficacy of aerial spraying of mosquito adulticide in reducing incidence of West Nile virus, California, 2005. Emerging Infectious Diseases 14:747–754.

Chen, H. Y., H. H. Stratton, T. B. Caraco, and D. J. White. 2006. Spatiotemporal Bayesian analysis of Lyme disease in New York State, 1990–2000. Journal of Medical Entomology 43:777–784.

Clark, J. S. 2005. Why environmental scientists are becoming Bayesians. Ecology Letters 8:2–14.

Clark, J. S., and O. N. Bjornstad. 2004. Population time series: process variability, observation errors, missing values, lags, and hidden states. Ecology 85:3140–3150.

Clark, J. S., G. A. Ferraz, N. Oguge, H. Hays, and J. DiCostanzo. 2005. Hierarchical Bayes for structured, variable populations: from recapture data to life-history prediction. Ecology 86:2232–2244.

Clark, J., and A. Gelfand. 2006. Hierarchical modelling for the environmental sciences—statistical methods and applications. Oxford University Press, New York, New York, USA.

Clark, J. S., M. Wolosin, M. Dietze, I. Ibanez, S. LaDeau, M. Welsh, and B. Kloeppel. 2007. Tree growth inference and prediction from diameter censuses and ring widths. Ecological Applications 17:1942–1953.

Colizza, V., A. Barrat, M. Barthelemy, and A. Vespignani. 2007. Predictability and epidemic pathways in global outbreaks of infectious diseases: SARS case study. BMC Medicine 5:34.

Colwell, R. R. 1996. Global climate and infectious disease: the cholera paradigm. Science 274:2025–2031.

Colwell, R. R., A. Huq, M. S. Islam, K. M. A. Aziz, M. Yunus, N. H. Khan, A. Mahmud, R. B. Sack, G. B. Nair, J. Chakraborty, D. A. Sack, and E. Russek-Cohen. 2003. Reduction of cholera in Bangladeshi villages by simple filtration. Proceedings of the National Academy of Sciences USA 100:1051–1055.

Congdon, P. 2003. Applied Bayesian modelling. John Wiley and Sons, Chichester, UK.

Connally, N. P., H. S. Ginsberg, and T. N. Mather. 2006. Assessing peridomestic entomological factors as predictors for Lyme disease. Journal of Vector Ecology 31:364–370.

Craigmile, P. F., C. A. Calder, H. Li, R. Paul, and N. Cressie. 2009. Hierarchical model building, fitting, and checking: a behind-the-scenes look at a Bayesian analysis of arsenic exposure pathways. Bayesian Analysis 4:1–36.

Cressie, N., C. A. Calder, J. S. Clark, J. M. V. Hoef, and C. K. Wikle. 2009. Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling. Ecological Applications 19:553–570.

Crowl, T. A., T. O. Crist, R. R. Parmenter, G. Belovsky, and A. E. Lugo. 2008. The spread of invasive species and infectious disease as drivers of ecosystem change. Frontiers in Ecology and the Environment 6:238–246.

Curriero, F. C., S. M. Shone, and G. E. Glass. 2005. Cross correlation maps: a tool for visualizing and modeling time lagged associations. Vector-Borne and Zoonotic Diseases 5:267–275.

Daszak, P., A. A. Cunningham, and A. D. Hyatt. 2000. Wildlife ecology—emerging infectious diseases of wildlife—threats to biodiversity and human health. Science 287:443–449.

Day, J. F., and J. Shaman. 2008. Using hydrologic conditions to forecast the risk of focal and epidemic arboviral transmission in peninsular Florida. Journal of Medical Entomology 45:458–465.

Donnelly, C. A., et al. 2003. Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong. Lancet 361:1761–1766.

Eidson, M., N. Komar, F. Sorhage, R. Nelson, T. Talbot, F. Mostashari, R. McLean, and S. West. 2001. Nile Virus avian mortality. Crow deaths as a sentinel surveillance system for West Nile virus in the Northeastern United States, 1999. Emerging Infectious Diseases 7:615–620.

Elkinton, J. S., W. M. Healy, J. P. Buonaccorsi, G. H. Boettner, A. M. Hazzard, H. R. Smith, and A. M. Liebhold. 1996. Interactions among gypsy moths, white-footed mice, and acorns. Ecology 77:2332–2342.

Estrada-Pena, A. 2002. Increasing habitat suitability in the United States for the tick that transmits Lyme disease: a remote sensing approach. Environmental Health Perspectives 110:635–640.

Evans, A. S. 1976. Causation and disease—Henle-Koch postulates revisited. Yale Journal of Biology and Medicine 49:175–195.

Fang, Y., and W. K. Reisen. 2006. Previous infection with West Nile or St. Louis encephalitis viruses provides cross protection during reinfection in house finches. American Journal of Tropical Medicine and Hygiene 75:480–485.

Farnsworth, M. L., J. A. Hoeting, N. T. Hobbs, and M. W. Miller. 2006. Linking chronic wasting disease to mule deer movement scales: a hierarchical Bayesian approach. Ecological Applications 16:1026–1036.

Ferguson, N. M., C. A. Donnelly, and R. M. Anderson. 2001. Transmission intensity and impact of control policies on the foot and mouth epidemic in Great Britain. Nature 414: 329–329.

Ferrari, M. J., R. F. Grais, N. Bharti, A. J. K. Conlan, O. N. Bjornstad, L. J. Wolfson, P. J. Guerin, A. Djibo, and B. T. Grenfell. 2008. The dynamics of measles in sub-Saharan Africa. Nature 451:679–684.

Focks, D. A., D. G. Haile, E. Daniels, and G. A. Mount. 1993. Dynamic life table model for Aedes-aegypti—analysis of the literature and model development. Journal of Medical Entomology 30:1003–1017.

Foo, L., T. Lim, H. Lee, and R. Fang. 1985. Rainfall, abundance of Aedes and dengue infection in Selangor, Malaysia. Southeast Asian Journal of Tropical Medicine and Public Health 16:560–568.

Galvani, A. P., and R. M. May. 2005. Epidemiology—dimensions of superspreading. Nature 438:293–295.

Gelman, A., and J. Hill. 2007. Data analysis using regression and multilevel/hierarchical models. Cambridge University Press, New York, New York, USA.

Glass, G. E. 2007. Rainy with a chance of plague: forecasting disease outbreaks from satellites. Future Virology 2:225–229.

Glass, G. E., J. E. Cheek, J. A. Patz, T. M. Shields, T. J. Doyle, D. A. Thoroughman, D. K. Hunt, R. E. Enscore, K. L. Gage, C. Irland, C. J. Peters, and R. Bryan. 2000. Using remotely sensed data to identify areas at risk for hantavirus pulmonary syndrome. Emerging Infectious Diseases 6:238–247.

Greger, M. 2007. The human/animal interface: emergence and resurgence of zoonotic infectious diseases. Critical Reviews in Microbiology 33:243–299.

Grenfell, B. T., O. N. Bjornstad, and B. F. Finkenstadt. 2002. Dynamics of measles epidemics: scaling noise, determinism,

and predictability with the TSIR model. Ecological Monographs 72:185–202.

Gubler, D. J. 1998. Dengue and dengue haemorrhagic fever. Clinical Microbiology Reviews 11:480–496.

Gumel, A. B., S. G. Ruan, T. Day, J. Watmough, F. Brauer, P. van den Driessche, D. Gabrielson, C. Bowman, M. E. Alexander, S. Ardal, J. H. Wu, and B. M. Sahai. 2004. Modelling strategies for controlling SARS outbreaks. Proceedings of the Royal Society B 271:2223–2232.

Guptill, S. C., K. G. Julian, G. L. Campbell, S. D. Price, and A. A. Marfin. 2003. Early-season avian deaths from West Nile virus as warnings of human infection. Emerging Infectious Diseases 9:483–484.

He, D. H., E. L. Ionides, and A. A. King. 2010. Plug-and-play inference for disease dynamics: measles in large and small populations as a case study. Journal of the Royal Society Interface 7:271–283.

Hobbs, N. T., and R. Hilborn. 2006. Alternatives to statistical hypothesis testing in ecology: a guide to self teaching. Ecological Applications 16:5–19.

Hohle, M., E. Jorgensen, and P. D. O'Neill. 2005. Inference in disease transmission experiments by using stochastic epidemic models. Journal of the Royal Statistical Society Series C-Applied Statistics 54:349–366.

Hooten, M. B., and C. K. Wikle. 2008. A hierarchical Bayesian non-linear spatio-temporal model for the spread of invasive species with application to the Eurasian Collared-Dove. Environmental and Ecological Statistics 15:59–70.

Hubalek, Z., and J. Halouzka. 1999. West Nile fever—a reemerging mosquito-borne viral disease in Europe. Emerging Infectious Diseases 5:643–650.

Ionides, E. L., C. Breto, and A. A. King. 2006. Inference for nonlinear dynamical systems. Proceedings of the National Academy of Sciences USA 103:18438–18443.

Jiang, J. F., Z. P. Qiu, J. H. Wu, and H. P. Zhu. 2009. Threshold conditions for West Nile virus outbreaks. Bulletin of Mathematical Biology 71:627–647.

Jin, X., S. Banerjee, and B. P. Carlin. 2007. Order-free coregionalized lattice models with application to multiple disease mapping. Journal of the Royal Statistical Society Series B 60:817–883.

Johansson, M. A., F. Dominici, and G. E. Glass. 2009. Local and global effects of climate on dengue transmission in Puerto Rico. PLoS Neglected Tropical Diseases 3(2)::e382.

Jones, C. G., R. S. Ostfeld, M. P. Richard, E. M. Schauber, and J. O. Wolff. 1998. Chain reactions linking acorns to gypsy moth outbreaks and Lyme disease risk. Science 279:1023–1026.

Jones, K. E., N. G. Patel, M. A. Levy, A. Storeygard, D. Balk, J. L. Gittleman, and P. Daszak. 2008. Global trends in emerging infectious diseases. Nature 451:990–U994.

Joseph, L., T. W. Gyorkos, and L. Coupal. 1995. Bayesian-estimation of disease prevalence and the parameters of diagnostic-tests in the absence of a gold standard. American Journal of Epidemiology 141:263–272.

Karesh, W. B., R. A. Cook, E. L. Bennett, and J. Newcomb. 2005. Wildlife trade and global disease emergence. Emerging Infectious Diseases 11:1000–1002.

Kearney, M., W. P. Porter, C. Williams, S. Ritchie, and A. A. Hoffmann. 2009. Integrating biophysical models and evolutionary theory to predict climatic impacts on species' ranges: the dengue mosquito *Aedes aegypti* in Australia. Functional Ecology 23:528–538.

Keeling, M. J., M. E. J. Woolhouse, R. M. May, G. Davies, and B. T. Grenfell. 2003. Modelling vaccination strategies against foot-and-mouth disease. Nature 421:136–142.

Keesing, F., J. Brunner, S. Duerr, M. Killilea, K. LoGiudice, K. Schmidt, H. Vuong, and R. S. Ostfeld. 2009. Hosts as ecological traps for the vector of Lyme disease. Proceedings of the Royal Society B 276:3911–3919.

Keirans, J. E., H. J. Hutcheson, L. A. Durden, and J. S. H. Klompen. 1996. *Ixodes scapularis*: redescription of all active stages, distribution, hosts, geographical variation, and medical and veterinary importance. Journal of Medical Entomology 33:297–318.

Kilpatrick, A. M., A. A. Chmura, D. W. Gibbons, R. C. Fleischer, P. P. Marra, and P. Daszak. 2006a. Predicting the global spread of H5N1 avian influenza. Proceedings of the National Academy of Sciences USA 103:19368–19373.

Kilpatrick, A. M., P. Daszak, M. J. Jones, P. P. Marra, and L. D. Kramer. 2006b. Host heterogeneity dominates West Nile virus transmission. Proceedings of the Royal Society B 273:2327–2333.

Kilpatrick, A. M., L. D. Kramer, S. R. Campbell, E. O. Alleyne, A. P. Dobson, and P. Daszak. 2005. West Nile virus risk assessment and the bridge vector paradigm. Emerging Infectious Diseases 11:425–429.

Kilpatrick, A. M., L. D. Kramer, M. J. Jones, P. P. Marra, and P. Daszak. 2006c. West Nile virus epidemics in North America are driven by shifts in mosquito feeding behavior. PLoS Biology 4:606–610.

Kilpatrick, A. M., S. L. LaDeau, and P. P. Marra. 2007. Ecology of West Nile virus transmission and its impact on birds in the western hemisphere. Auk 124:1121–1136.

Kilpatrick, A. M., M. A. Meola, R. M. Moudy, and L. D. Kramer. 2008. Temperature, viral genetics, and the transmission of West Nile virus by *Culex pipiens* mosquitoes. PLoS Pathogens 4(6):e1000092.

Kitron, U., J. A. Clennon, M. C. Cecere, R. E. Gurtler, C. H. King, and G. Vazquez-Prokopec. 2006. Upscale or downscale: applications of fine scale remotely sensed data to Chagas disease in Argentina and schistosomiasis in Kenya. Geospatial Health 1:49–58.

Komar, N., S. Langevin, S. Hinten, N. Nemeth, E. Edwards, D. Hettler, B. Davis, R. Bowen, and M. Bunning. 2003. Experimental infection of North American birds with the New York 1999 strain of West Nile virus. Emerging Infectious Diseases 9:311–322.

Komar, N., N. A. Panella, S. A. Langevin, A. C. Brault, M. Amador, E. Edwards, and J. C. Owen. 2005. Avian hosts for West Nile virus in St. Tammany Parish, Louisiana, 2002. American Journal of Tropical Medicine and Hygiene 73:1031–1037.

Kramer-Schadt, S., N. Fernandez, D. Eisinger, V. Grimm, and H. H. Thulke. 2009. Individual variations in infectiousness explain long-term disease persistence in wildlife populations. Oikos 118:199–208.

Kuk, A. Y. C., and C. C. Tan. 2009. Estimating the time-varying rate of transmission of SARS in Singapore and Hong Kong under two environments. Journal of the American Statistical Association 104:88–96.

Kwok, K. O., G. M. Leung, W. Y. Lam, and S. Riley. 2007. Using models to identify routes of nosocomial infection: a large hospital outbreak of SARS in Hong Kong. Proceedings of the Royal Society B 274:611–618.

LaDeau, S. L., and J. S. Clark. 2006. Elevated CO_2 and tree fecundity: role of tree size, interannual variability, and population heterogeneity. Global Change Biology 12:822–833.

LaDeau, S. L., A. M. Kilpatrick, and P. P. Marra. 2007. West Nile virus emergence and large-scale declines of North American bird populations. Nature 447:710–U713.

LaDeau, S. L., P. P. Marra, A. M. Kilpatrick, and C. A. Calder. 2008. West Nile virus revisited: consequences for North American ecology. BioScience 58:937–946.

Lanciotti, R. S., et al. 1999. Origin of the West Nile virus responsible for an outbreak of encephalitis in the northeastern United States. Science 286:2333–2337.

Lekone, P. E. 2008. Bayesian analysis of Severe Acute Respiratory Syndrome: the 2003 Hong Kong epidemic. Biometrical Journal 50:597–607.

Liebhold, A., V. Sork, M. Peltonen, W. Koenig, O. N. Bjornstad, R. Westfall, J. Elkinton, and J. M. H. Knops. 2004. Within-population spatial synchrony in mast seeding of North American oaks. Oikos 104:156–164.

Lipp, E. K., A. Huq, and R. R. Colwell. 2002. Effects of global climate on infectious disease: the cholera model. Clinical Microbiology Reviews 15:757–770.

Lipsitch, M., T. Cohen, B. Cooper, J. M. Robins, S. Ma, L. James, G. Gopalakrishna, S. K. Chew, C. C. Tan, M. H. Samore, D. Fisman, and M. Murray. 2003. Transmission dynamics and control of severe acute respiratory syndrome. Science 300:1966–1970.

Lloyd-Smith, J. O., A. P. Galvani, and W. M. Getz. 2003. Curtailing transmission of severe acute respiratory syndrome within a community and its hospital. Proceedings of the Royal Society Series B 270:1979–1989.

Lloyd-Smith, J. O., S. J. Schreiber, P. E. Kopp, and W. M. Getz. 2005. Superspreading and the effect of individual variation on disease emergence. Nature 438:355–359.

LoGiudice, K., R. S. Ostfeld, K. A. Schmidt, and F. Keesing. 2003. The ecology of infectious disease: effects of host diversity and community composition on Lyme disease risk. Proceedings of the National Academy of Sciences USA 100:567–571.

Lothrop, H. D., B. B. Lothrop, D. E. Gomsi, and W. K. Reisen. 2008. Intensive early season adulticide applications decrease arbovirus transmission throughout the Coachella Valley, Riverside County, California. Vector-Borne and Zoonotic Diseases 8:475–489.

MacDonald, G. 1957. The epidemiology and control of malaria. Oxford University Press, London, UK.

Matthews, L., and M. Woolhouse. 2005. New approaches to quantifying the spread of infection. Nature Reviews Microbiology 3:529–536.

McBryde, E. S., G. Gibson, A. N. Pettitt, Y. Zhang, B. Zhao, and D. L. S. McElwain. 2006. Bayesian modelling of an epidemic of severe acute respiratory syndrome. Bulletin of Mathematical Biology 68:889–917.

Meyers, L., B. Pourbohloul, M. Newman, D. Skowronski, and R. Brunham. 2005. Network theory and SARS: predicting outbreak diversity. Journal of Theoretical Biology 232:71–81.

Miller, M. W., N. T. Hobbs, and S. J. Tavener. 2006. Dynamics of prion disease transmission in mule deer. Ecological Applications 16:2208–2214.

Nakhapakorn, K., and N. Tripathi. 2005. An information value based analysis of physical and climatic factors affecting dengue fever and dengue haemorrhagic fever incidence. International Journal of Health Geographics 4:13–13.

Nemeth, N. M., G. E. Kratz, R. Bates, J. A. Scherpelz, R. A. Bowen, and N. Komar. 2008. Naturally induced humoral immunity to West Nile virus infection in raptors. Ecohealth 5:298–304.

Neubert, M. G., and H. Caswell. 2000. Demography and dispersal: calculation and sensitivity analysis of invasion speed for structured populations. Ecology 81:1613–1628.

Nielsen, C. F., and W. K. Reisen. 2007. West Nile virus-infected dead corvids increase the risk of infection in Culex mosquitoes in domestic landscapes. Journal of Medical Entomology 44:1067–1073.

Ogden, N. H., M. Bigras-Poulin, C. J. O'Callaghan I. K. Barker, K. Kurtenbach, L. R. Lindsay, and D. F. Charron. 2007. Vector seasonality, host infection dynamics and fitness of pathogens transmitted by the tick Ixodes scapularis. Parasitology 134:209–227.

Ogden, N. H., M. Bigras-Poulin, C. J. O'Callaghan I. K. Barker, L. R. Lindsay, A. Maarouf, K. E. Smoyer-Tomic, D. Waltner-Toews, and D. Charron. 2005. A dynamic population model to investigate effects of climate on geographic range and seasonality of the tick Ixodes scapularis. International Journal for Parasitology 35:375–389.

Ogle, K., and J. J. Barber. 2008. Bayesian data–model integration in plant physiological and ecosystem ecology. Progress in Botany 69:281–311.

Ostfeld, R., E. Schauber, C. Canham, F. Keesing, C. Jones, and J. Wolff. 2001. Effects of acorn production and mouse abundance on abundance and Borrelia burgdorferi-infection prevalence of nymphal Ixodes scapularis ticks. Vector-Borne and Zoonotic Diseases 1:55–63.

Ostfeld, R. S., C. D. Canham, K. Oggenfuss, R. J. Winchcombe, and F. Keesing. 2006. Climate, deer, rodents, and acorns as determinants of variation in Lyme-disease risk. PLoS Biology 4:1058–1068.

Ostfeld, R. S., G. E. Glass, and F. Keesing. 2005. Spatial epidemiology: an emerging (or re-emerging) discipline. Trends in Ecology and Evolution 20:328–336.

Ostfeld, R. S., C. G. Jones, and J. O. Wolff. 1996. Of mice and mast. BioScience 46:323–330.

Pascual, M., and M. J. Bouma. 2009. Do rising temperatures matter? Ecology 90:906–912.

Peterson, A. T., C. Martinez-Campos, Y. Nakazawa, and E. Martinez-Meyer. 2005. Time-specific ecological niche modeling predicts spatial dynamics of vector insects and human dengue cases. Transactions of the Royal Society of Tropical Medicine and Hygiene 99:647–655.

Peterson, A. T., and R. A. J. Williams. 2008. Risk mapping of highly pathogenic avian influenza distribution and spread. Ecology and Society 13:15.

Platonov, A. E., M. V. Fedorova, L. S. Karan, T. A. Shopenskaya, O. V. Platonova, and V. I. Zhuravlev. 2008. Epidemiology of West Nile infection in Volgograd, Russia, in relation to climate change and mosquito (Diptera: Culicidae) bionomics. Parasitology Research 103:S45–S53.

Plowright, R. K., S. H. Sokolow, M. E. Gorman, P. Daszak, and J. E. Foley. 2008. Causal inference in disease ecology: investigating ecological drivers of disease emergence. Frontiers in Ecology and the Environment 6:420–429.

Promprou, S., M. Jaroensutasinee, and K. Jaroensutasinee. 2005. Climatic factors affecting dengue haemorrhagic fever incidence in southern Thailand. Dengue Bulletin 29:41–48.

Qian, S. S., T. F. Cuffney, I. Alameddine, G. McMahon, and K. H. Reckhow. 2010. On the application of multilevel modeling in environmental and ecological studies. Ecology 91:355–361.

Reisen, W. K., D. Cayan, M. Tyree, C. A. Barker, B. Eldridge, and M. Dettinger. 2008. Impact of climate variation on mosquito abundance in California. Journal of Vector Ecology 33:89–98.

Reisen, W. K., Y. Fang, and V. M. Martinez. 2006a. Effects of temperature on the transmission of West Nile virus by Culex tarsalis (Diptera: Culicidae). Journal of Medical Entomology 43:309–317.

Reisen, W. K., Y. Fang, and V. M. Martinez. 2006b. Vector competence of Culiseta incidens and Culex thriambus for West Nile virus. Journal of the American Mosquito Control Association 22:662–665.

Riley, S., et al. 2003. Transmission dynamics of the etiological agent of SARS in Hong Kong: impact of public health interventions. Science 300:1961–1966.

Royle, J. A. 2009. Analysis of capture–recapture models with individual covariates using data augmentation. Biometrics 65:267–274.

Royle, J. A., and W. A. Link. 2002. Random effects and shrinkage estimation in capture–recapture models. Journal of Applied Statistics 29:329–351.

Rueda, L. M., K. J. Patel, R. C. Axtell, and R. E. Stinner. 1990. Temperature-dependent development and survival rates of Culex quinquefasciatus and Aedes aegypti. Journal of Medical Entomology 27:892–898.

Schauber, E. M., R. S. Ostfeld, and A. S. Evans. 2005. What is the best predictor of annual Lyme disease incidence: weather, mice, or acorns? Ecological Applications 15:575–586.

Schmidt, K. A., and R. S. Ostfeld. 2001. Biodiversity and the dilution effect in disease ecology. Ecology 82:609–619.

Shone, S. M., F. C. Curriero, C. R. Lesser, and G. E. Glass. 2006. Characterizing population dynamics of *Aedes sollicitans* using meteorological data. Journal of Medical Entomology 43:393–402.

Small, M., C. K. Tse, and D. M. Walker. 2006. Super-spreaders and the rate of transmission of the SARS virus. Physica D-Nonlinear Phenomena 215:146–158.

Smith, T., N. Maire, A. Ross, M. Penny, N. Chitnis, A. Schapira, A. Studer, B. Genton, C. Lengeler, F. Tediosi, D. De Savigny, and M. Tanner. 2008. Towards a comprehensive simulation model of malaria epidemiology and control. Parasitology 135:1507–1516.

Smithburn, K, and T. Hughes, A. W. Burke, and J. H. Paul. 1940. A neurotropic virus isolated from the blood of a native of Uganda. American Journal of Tropical Medicine and Hygiene 20:471–492.

Snow, J. 1855. On the mode of communication of cholera. John Churchill, London, UK.

Swanson, K. I., and D. E. Norris. 2007. Co-circulating microorganisms in questing *Ixodes scapularis* nymphs in Maryland. Journal of Vector Ecology 32:243–251.

Tachiiri, K., B. Klinkenberg, S. Mak, and J. Kazmi. 2006. Predicting outbreaks: a spatial risk assessment of West Nile virus in British Columbia. International Journal of Health Geographics 5(21)::1–21.

Tavecchia, G., P. Besbeas, T. Coulson, B. Morgan, and T. Clutton-Brock. 2009. Estimating population size and hidden demographic parameters with state-space modeling. American Naturalist 173:722–733.

Taylor, L. H., S. M. Latham, and M. E. J. Woolhouse. 2001. Risk factors for human disease emergence. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences 356:983–989.

Tonjes, D. J. 2008. Estimates of worst case baseline West Nile virus disease effects in a suburban New York county. Journal of Vector Ecology 33:293–304.

Turell, M. J., D. J. Dohm, M. R. Sardelis, M. L. O. Guinn, T. G. Andreadis, and J. A. Blow. 2005. An update on the potential of North American mosquitoes to transmit West Nile virus. Journal of Medical Entomology 42:57–62.

Walk, S. T., G. Xu, J. W. Stull, and S. M. Rich. 2009. Correlation between tick density and pathogen endemicity, New Hampshire. Emerging Infectious Diseases 15:585–587.

Wallinga, J., and P. Teunis. 2004. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. American Journal of Epidemiology 160:509–516.

Wang, G. M. 2009. Signal extraction from long-term ecological data using Bayesian and non-Bayesian state-space models. Ecological Informatics 4:69–75.

Wang, W. D., and S. G. Ruan. 2004. Simulating the SARS outbreak in Beijing with limited data. Journal of Theoretical Biology 227:369–379.

Watts, D. M., D. S. Burke, B. A. Harrison, R. E. Whitmire, and A. Nisalak. 1987. Effect of temperature on the vector efficiency of *Aedes aegypti* for dengue-2 virus. American Journal of Tropical Medicine and Hygiene 36:143–152.

Wearing, H. J., P. Rohani, and M. J. Keeling. 2005. Appropriate models for the management of infectious diseases. PLoS Medicine 2:621–627.

Wilcox, B. A., and D. J. Gubler. 2005. Disease ecology and the global emergence of zoonotic pathogens. Environmental Health and Preventive Medicine 10:263–272.

Williams, E. S., M. W. Miller, T. J. Kreeger, R. H. Kahn, and E. T. Thorne. 2002. Chronic wasting disease of deer and elk: a review with recommendations for management. Journal of Wildlife Management 66:551–563.

Williams, E. S., and S. Young. 1980. Chronic wasting disease of captive mule deer spongiform encephalopathy. Journal of Wildlife Diseases 16:89–98.

Wolff, J. O. 1996. Population fluctuations of mast-eating rodents are correlated with production of acorns. Journal of Mammalogy 77:850–856.

Wonham, M. J., M. A. Lewis, J. Renclawowicz, and P. Van den Driessche. 2006. Transmission assumptions generate conflicting predictions in host–vector disease models: a case study in West Nile virus. Ecology Letters 9:706–725.

Woolhouse, M. E. J., and C. Dye. 2001. Population biology of emerging and re-emerging pathogens—preface. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences 356:981–982.

Wu, P. C., J. G. Lay, H. R. Guo, C. Y. Lin, S. C. Lung, and H. J. Su. 2009. Higher temperature and urbanization affect the spatial patterns of dengue fever transmission in subtropical Taiwan. Science of the Total Environment 407:2224–2233.

Xiao, X. M., M. Gilbert, J. Slingenbergh, F. Lei, and S. Boles. 2007. Remote sensing, ecological variables, and wild bird migration related to outbreaks of highly pathogenic H5N1 avian influenza. Journal of Wildlife Diseases 43:S40–S46.

Yates, T. L., et al. 2002. The ecology and evolutionary history of an emergent disease: Hantavirus pulmonary syndrome. BioScience 52:989–998.

Young, J. D. 1998. Underreporting of Lyme disease. New England Journal of Medicine 338:1629–1629.

Zhong, S. B., Q. Y. Huang, and D. J. Song. 2009. Simulation of the spread of infectious diseases in a geographical environment. Science in China Series D-Earth Sciences 52:550–561.