Consensus and conflict among ecological forecasts of Zika virus outbreaks in the United States

Colin J. Carlson[1,*], Eric Dougherty[1], Mike Boots[2], Wayne Getz[1,3], and Sadie J. Ryan[3,4,5]

[1]Department of Environmental Science, Policy and Management, University of California Berkeley, Berkeley, CA 94720-3112, USA
[2]Department of Integrative Biology, University of California Berkeley, Berkeley, CA 94720-3112, USA
[3]Schools of Mathematical Sciences (Getz) and Life Sciences (Ryan) University of KwaZulu, Natal, South Africa
[4]Department of Geography, University of Florida, Gainesville, FL 32601
[5]Emerging Pathogens Institute, University of Florida, Gainesville, FL, 32610
[*]cjcarlson@berkeley.edu

ABSTRACT

Ecologists are increasingly involved in the pandemic prediction process. In the course of the 2015-2016 outbreak of Zika virus, several ecological models were developed to forecast the potential global distribution of the disease. Conflicting results produced by alternative methods are unresolved, hindering the development of appropriate public health forecasts. We compare ecological niche models and experimentally-driven mechanistic forecasts for Zika transmission in the continental United States, a region of high model conflict. Based on plausible epidemiological parameters, we use stochastic county-level simulations to demonstrate the epidemiological consequences of conflict among ecological models. We also propose a basic consensus method that could resolve conflicting models of potential outbreak geography and seasonality, but we note that it fails to produce substantially more realistic outbreak trajectories. Our results illustrate the unacceptable margin of uncertainty contained within these predictions, which could misrepresent the risk faced by millions of people in the United States within a single year. In the short term, ecologists face the task of developing better *post hoc* consensus that accurately forecasts spatial patterns of Zika virus outbreaks. Ultimately, methods are needed that bridge the gap between ecological and epidemiological approaches to predicting transmission and realistically capture both outbreak size and geography.

Introduction

In the urgent setting of pandemic response, ecologists have begun to play an increasingly important role[1]. The processes that underly transmission of vector-borne or zoonotic diseases are fundamentally ecological, as well as social. In particular, key parameters governing the vector-borne transmission cycle, including transmission rates, vector lifespan, extrinsic incubation period, and other aspects of vector-borne disease epidemiology, are often controlled by temperature, precipitation, and other bioclimatic variables[2]. Ecological niche modeling is a typically phenomenological method that uses occurrence data to make inferences about geographic boundaries of potential transmission as a function of these hidden relationships[3]. Within niche modeling approaches, there are conflicting views regarding which algorithms are appropriate to use in the context of particular applications[4–6], and consensus methods have hardly advanced beyond basic model averaging[7]. As an increasingly popular alternative, mechanistic ecological models have been developed that extrapolate geographic projections from experimental results[8], but these can be data-intensive and highly sensitive to parameterization[8–10]. The two approaches—phenomenological and top-down, or mechanistic and bottom-up—should be roughly congruent when correctly implemented with sufficient data and predictors[11]. Yet discrepancies between the two approaches in practice highlight a tension in ecology, between deductive approaches that infer ecology from observed broad-scale patterns, and inductive approaches that scale ecological experiments to predict real patterns. In the context of pandemic response, the trade-off has acute stakes: early access to ecological predictions can help pandemic efforts, but inaccurate information based on limited data could drive mis-allocation of public health resources.[12]. There is, therefore, clearly a need to develop better consensus methods, but even before that, a clear need to understand the epidemiological implications of the differences between these models.

The Zika virus (henceforth Zika) pandemic that began in Brazil in December 2015 highlights the unusual and sensitive challenges of pandemic response. A number of characteristics make Zika unique from a public health standpoint. The appearance of microcephaly, and more broadly Zika congenital syndrome, is perhaps the most significant impact of the outbreak. At least 11,000 confirmed cases of Zika have affected pregnant women, leading to roughly 10,000 cases of birth defects, including microcephaly, in Brazil alone[13]. As of April 6, 2017, a total of 207,557 confirmed cases of autochthonous

transmission (out of 762,036 including suspected cases) have been recorded in the Americas[14]. Moreover, Zika is exceptional among vector-borne diseases in that it has developed a sexual pathway of transmission in humans (comparable examples, such as canine leishmaniasis, are incredibly rare[15]). The rapid spread of Zika virus from Brazil throughout the Americas has posed a particular problem for ecologists involved in pandemic response, as several different ecological niche models (ENMs)[12,16,17] and a handful of mechanistic forecasts[10] have been developed to project the potential full spread of the pathogen. So far, autochthonous transmission has been recorded throughout most of Central America and the Caribbean, with cases as far north as the southern tips of Texas and Florida. The geographic region with the most evident conflict between these predictions in the Americas is in the United States (see Figure 1).

The impacts of conflict among different risk assessments cannot be understated. For the United States, the lack of a consensus among different models renders the literature less credible to policymakers, as predictions span a range from 13 counties at risk[12] to the entire United States for at least part of the year[10]. At the time of writing, the majority of public health agencies in the United States appear to be preparing for the eventuality of Zika, based either on no prior geographic information, or basic data on the range of *Aedes* mosquitoes[18,19]. Millions of dollars have already been invested in state- and city-level Zika preparation, even in areas without recently-recorded *Aedes* presence; and pesticide spraying for vector control has already had unanticipated consequences, including killing millions of honeybees[20]. Domestic efforts to prepare for Zika are not unreasonable in the absence of a consensus prediction about Zika's likely final range; the continued importation of new cases into every state in the U.S. likely amplifies the perceived threat of local outbreaks, especially given the pathway of sexual transmission (which could conceivably start stuttering chains[21] outside regions of vector-borne transmission). However, an informed response to Zika in the United States requires both a greater consensus about at-risk areas, and a more precise understanding of current uncertainty among different methods of ecological forecasting.

Here, we compare four ecological predictions for the extent and duration of possible Zika virus transmission in the United States, and show that there is no one straightforward way of resolving uncertainty among them. By overlaying epidemiological simulations on ecological forecasts at the county scale, we show that these tools are compatible but, in concert, generate simulated outbreaks on a scale that is far from plausible. In the process, our exercise illustrates that relying on any one ecological model adds a hidden layer of uncertainty to epidemiological forecasts, highlighting the need to develop better consensus methods—and to develop ecological and epidemiological tools in a more integrated approach that better approximates observed outbreak patterns.

Results

Ecological forecasts for Zika suitability span the range of thirteen counties to all 3108 counties in the continental United States (Table 1), and within that range is a dramatic spread of outbreak scenarios. For ENM-based projections, the margin of error between mean trajectories spans an order of magnitude, with a total difference of 168 million cases between Carlson and Samy (Figure 1). Areas predicted by other methods to be at the greatest risk from Zika virus are roughly agreed upon between the models, with southern California and the Gulf Coast represented most significantly as outbreak hotspots among the three models. Agreement among all three models is limited, but is most significantly clustered in these areas, especially in the southern tip of Florida and Los Angeles County.

A roughly comparable range of predictions—but significantly greater overall case burden—is contained within the entire span of implementations of Mordecai's Bayesian model (Figure 2). Whereas ENM approaches indicate a somewhat restricted geographic range for possible outbreaks, the Mordecai model suggests that even in a conservative scenario (using minimum temperatures, and 97.5% posterior probability), the majority of *Aedes aegypti*'s range is at least seasonally suitable for Zika transmission. A far greater range of variation is contained within the minimum-temperature-based model scenario, which encompasses roughly half of the land area of the continental U.S. In contrast, the three scenarios based on maximum temperatures are geographically indistinguishable, though worsening projections do extend the seasonality of transmission and thereby produce somewhat longer-tailed epidemics (Figure 2a).

Between all nine ecological scenarios considered (three niche models and six mechanistic scenarios), an overwhelming spread of possible epidemics could be predicted for the United States (Figure 3). The accompanying spatial pattern of case burden also varies between interpretations; while the spatial patterns are roughly identical for Carlson, Messina, and Samy, the temporal dimension introduced by mapping the Mordecai model onto monthly temperature grids dramatically affects how cases are ultimately distributed—and produces a reduction in epidemic size in some scenarios (Figure 4). In fact, the most conservative Mordecai scenario (97.5% confidence with minimum temperature) falls between Carlson and Messina in terms of case burden, despite predicting more than four times more counties with transmission suitability than the Messina model. Across all models, forecasts predict that the majority of the case burden will still be seen along the Gulf Coast and in southern California. The value added from interfacing ecology and epidemiology is especially evident here; for example, while the original Carlson *et al.* study noted the most significant suitable area was in southern Florida and failed to comment on the potential importance of southern California, the most significant epidemic predicted by most models is in Los Angeles county.

The exception is the most conservative Mordecai scenario (Figure 2b), the only model of the nine in which Los Angeles is designated unsuitable for transmission—a fairly important discrepancy, given that the county is the most populous in the United States, and correspondingly contributes substantially to epidemic size in every other scenario (Figure 4a-c).

In an effort to illustrate a method of resolving these conflicting predictions, we present a final "consensus model" that incorporates all four modeling studies. Consensus methods are limited for ecological niche models[7], so we adopt one possible approach, a majority rule at the county scale across Carlson, Messina, and Samy (i.e., in Figure 1b, any county value at or above 2 is "suitable", and any below is "unsuitable"). Building on this "majority rule model," for counties that are marked suitable by the ecological niche models, we superimpose the monthly transmission values from Mordecai's most conservative scenario, which most closely matches the geographic extent predicted by the ecological niche models (Figure 2a versus Figure 4a-c). This filtered "seasonal majority rule" algorithm incorporates the temporal dimension of transmission that is added by our implementation of Mordecai's model, while maintaining consensus among the niche models, and unsurprisingly produces a comparatively conservative outbreak prediction (Figure 5). The inclusion of the temporal component from the mechanistic model reduces case burden by almost two-thirds (Table 2), and excludes a handful of counties in the process (which were suitable in the ENM approach but not suitable for a single month in the mechanistic model). Most notably, Los Angeles county (which is suitable for no months of the year in the Mordecai model) is excluded despite being suitable in all three ENMs, which contributes substantially to the overall reduction of projected case totals in the seasonal majority rule approach.

Discussion

The models we present here fail to explain why Zika case totals have been so low in the United States, given the wide suitable area suggested by the majority of models. One explanation is that the ongoing spread of the pandemic could simply be progressing slowly, in which case outbreaks in the United States over the next decade could be on the same scale as Brazil or Colombia. (This could especially prove true if the high force of infection from traveler cases—which have so far been an order of magnitude more common in the U.S.—drives more significant outbreaks than we model here by initializing simulations with a single case.) More realistically, a number of factors likely prevent the United States from experiencing an outbreak on the scale we simulate: better access to health care, preemptive vector control as part of Zika preparedness efforts, and significant fine-scale heterogeneity limiting mosquito populations in well populated areas (a factor that some models can accommodate[22], but niche models at the global scale do not). The last of these is most easily addressed through ecological tools, and finer-scale validation of downscaled ecological models is an important next step for ongoing forecasting. At the county scale, more detailed GIS data are needed to identify probable areas of suitable vector density; identifying those areas can reduce the population at risk (used to parameterize models) from an entire county's population down to just those living in high-risk (or non-zero risk) areas.

A number of further developments are needed to bridge important eco-epidemiological gaps and improve Zika forecasting both in the United States and worldwide. The Mordecai *et al.* model illustrates that transmission is unlikely to be a year-round property of most areas, especially in temperate zones; and our exercise shows that reducing the months of possible transmission does significantly reduce total outbreak size. Time-specific ecological niche models have been used with great success to predict the dynamics of dengue[23], another *Aedes*-borne disease, and have been applied as a proxy for Zika risk[24]. However, these models will need to be developed for Zika as more data on enzootic cycles outside the Americas become available; and time-specific ecological niche models will pose an additional challenge for consensus building with mechanistic time-sensitive models like Mordecai *et al.*'s. Travel-based models have shown promise for other diseases[25] and have been applied in a limited capacity with dengue models to predict Zika risk[24]. These types of models can be applied with Zika-specific niche models for more detailed forecasts of traveler-driven outbreaks at the edges of suitability. But a more detailed epidemiological link is needed between traveler force of infection and the scale of subsequent local outbreaks; so far, that causation has only been investigated in reverse.[26] Finally, the role of sexual transmission requires deeper investigation. Early work suggested sexual transmission might be a substantial factor explaining the explosive South American outbreak[27,28], but recent work has suggested sexually-transmitted outbreaks are unlikely[29], even if sexual transmission increases the severity of vector-borne outbreaks[30]; others still argue these risk are "understated."[31,32] Some work at the county level has already begun predicting Zika risk based on other sexually transmitted diseases[33], but for this to be useful to policymakers, a basic and accurate model of importance of sexual transmission is still needed[34].

Ecological niche modeling (ENM) is a comparatively new statistical method in ecology, and it has only recently been applied to emerging infectious diseases. In fewer than two decades, the statistical power of ENMs has grown exponentially, especially as increasingly complex methods for machine learning have been applied to the problem. The dozen or so methods currently employed offer a wide palette of options for potential modelers to choose from, and compounded with the wide range of potential environmental and social covariates, seemingly limitless combinations of possible models can be produced from a single dataset, each of which is statistically rigorous enough to be published. High-profile targets, such as emerging infectious diseases with ecologically-driven occurrences, such as vector-borne and other zoonotic diseases, frequently inspire conflicting

models, but in mainstream species distribution modeling research, the impacts of those conflicts are often treated in as an academic problem. For infectious disease mapping, such conflict has conspicuous stakes that produce downstream uncertainty for stakeholders, clinicians, and policymakers.

The conflict among different models exposes an underlying tension between two different intentions of disease mapping. One paradigm focuses on accuracy (especially specificity), and follows a similar paradigm to mainstream invasion biology research in that it attempts to most accurately guess the final boundaries of incipient range expansion. Overprediction and underprediction are weighted as equal problems in this approach; the task of appropriate allocation of clinical resources is equally impeded by both margins of error. Another, alternative paradigm assumes that a Type II error is of far greater significance than a Type I error, from a preparedness standpoint; and reacts especially to the stakes of under-prediction by targeting predictions at any area that could, theoretically, sustain outbreaks. In this study, Carlson and Messina are more representative of the first paradigm, while Samy and Mordecai match the latter, especially in the way that we incorporate them into our analyses. In reality, all disease distribution models fall somewhere on a continuum between the two paradigms, and modelers following best practices are likely to produce primarily objective results. But to the degree that no forecasting effort is fully unsupervised, and basic decisions (like including or excluding current outbreak data) introduce opportunities for subjectivity, conflict between these two approaches is likely to be an ongoing disciplinary problem beyond Zika. At a minimum, understanding the epidemiological implications of ecological uncertainty is a key step towards improving ecologists' performance in pandemic preparedness.

At the present time, the most common practice to address the ecology-epidemiology interface in the niche modeling literature is the use of population-at-risk (PAR) methods. Basic area-under-the-model population estimates are perhaps the simplest and most readily comparable possible epidemiological metric; only Messina *et al.* present a global PAR (2.17 billion people) based on their Zika virus niche model. Bogoch *et al.* revised that figure in a more regional assessment for Africa, Asia, and the Pacific that included traveler populations and a seasonal component to transmission but to do so, substituted existing dengue models in place of actual Zika models.[24] If implemented more frequently, population-at-risk methods could be a simple *post hoc* way of comparing different ecological forecasts. However, these methods might accidentally introduce more alarm than they communicate risk (just as using susceptible populations, without any associated model of transmission, is a fairly uninformative proxy for an epidemic projection in mainstream epidemiology).

The exercise carried out here illuminates one of the primary weaknesses of ecological niche modeling methods; namely, though ENMs have great value for defining the plausible outer bounds of transmission, they are largely unable to clarify the distribution of risk within these suitable areas. Our methods here are meant to illustrate the danger of extrapolating the epidemiological impacts of an infectious disease based on the course categorization of spatial units into "suitable" and "unsuitable." Without consideration of the differential suitability within the boundaries set by and ENM (using one of the many possible methods currently available), there is no way to accurately calculate the magnitude of the epidemic supposedly being described. Rather, we must treat all potentially suitable areas as having exactly the same risk of an outbreak. Subsequent propagation of an outbreak based on reasonable epidemiological parameters results in completely unreasonable estimates of the size and character of the epidemic. The range of projected epidemics varies among the ENM approaches by more than an order of magnitude, but even the smallest outbreak prediction is still five orders of magnitude higher than the case totals observed during the last outbreak season (223 real cases versus roughly 12 million simulated cases).

The incompatibility of these two types of models represents a major deficiency in our ability to forecast disease spread. A logical next step in integrating eco-epidemiology should therefore be to directly interface epidemiological predictions with niche models to more accurately characterize spatiotemporal patterns of risk. This, however, may require a vast methodological advancement. In the meantime, there are a number of potential avenues for exploration that may aid in our ability to effectively apply these existing methods to emerging infectious diseases. Consensus models, like the simple majority-rule model presented here, may be the first step towards decomposing suitability into something more epidemiologically-relevant. Development of alternative consensus models should aim to further clarify the level of suitability beyond the simple binomial categorization offered by ENM methods alone. The inclusion of a temporal component (i.e., the use of the conservative Mordecai projections of suitability for mosquitoes) enables some decomposition of the ENM results. Furthermore, the frameworks underlying consensus models should be adaptable as additional case data is made available. The outputs of ENMs represent static instantiations of dynamic processes, whether they describe species ranges of the transmission niches of emerging infectious diseases. The ability of these models to contribute to our understanding of pathogens entering novel regions or hosts will hinge upon their flexibility in incorporating near-real-time data[35].

Methods

Ecological Models

Three ecological niche models (ENMs) have been published that explicitly map the possible distribution of Zika virus, using a different combination of occurrence data, environmental predictors, and statistical approaches[12,16,17]. They offer three

scenarios, of varying severity, especially as measured within the United States (Table 1). Other models have also been widely used in epidemiological work as a proxy for the distribution of Zika, such as an ecological niche model of *Aedes aegypti* and *Ae. albopictus*[36] (fairly commonly used, e.g.,[33,37]; or see[38], which presents its own *Aedes* ENM that becomes a risk map of Zika transmission), or dengue-specific niche models (recently used by Bogoch *et al.* in two separate publications[24,39]). Most ecological niche models indicate the range of Zika virus should be more restricted than that of its vectors, and published evidence suggests there may be significant differences between the known and potential distributions of dengue and Zika[12], so we exclude these proxy methods from our study and focus instead on modeling studies that explicitly use Zika occurrence data.

Carlson et al.

Carlson *et al.*[12] developed an ensemble niche model constructed using the R package BIOMOD2. The resulting model uses seven of ten possible methods (general linear models, general additive models, classification tree analysis, flexible discriminant analysis, multiple adaptive regression splines, random forests, and boosted regression trees), notably omitting maximum entropy (MaxEnt). Their primary model uses only occurrence data from outside the Americas, but here we adapt their secondary model which incorporates data from Messina *et al.* (below) to show the lack of the sensitivity of the method to that additional data, especially in the United States. The only environmental predictors used are the BIOCLIM dataset[40] and a vegetation index (NDVI). In the Carlson *et al.* model, suitable range for Zika virus is predicted to be limited to the southern tip of Florida and small patches of Los Angeles and the San Francisco Bay area. Only a total of 13 counties have any suitable area in this model; at the county scale, this model has the greatest concordance with observed outbreak patterns of autochthonous transmission in the United States during 2016.

Messina et al.

Messina *et al.*[16] use an ensemble boosted regression trees approach with occurrence points primarily from South American outbreak data. The model incorporates prior information about *Aedes* distributions. For example, pseudoabsences are preferentially generated in areas of lower *Aedes* suitability. Their model uses six environmental predictors: two direct climate variables, two indices of dengue transmission based on temperature (one for *Ae. albopictus* and one for *Ae. aegypti*), a vegetation index (EVI), and a binary land cover classifier (urban or rural). Their model predicts that suitable range for Zika virus encompasses a substantial portion of the Gulf Coast, including the entirety of Florida and as far west as eastern Texas. Their study is also the first to estimate population-at-risk, placing the global figure at 2.17 billion people.

Samy et al.

Samy *et al.*[17] use MaxEnt to build two main sub-models. The first is a conventional ENM approach using environmental predictors (precipitation, temperature, EVI, soil water stress, "aridity," and elevation). The second, a more unconventional approach in the niche modeling literature, separates out socioeconomic predictors (among them population density, night light from satellite imagery, and a function of expected travel time called "accessibility"). According to Samy *et al.*, areas suitable by both criteria have "all factors in place for autochthonous transmission." But in areas where one factor is in place, they suggest, outbreaks could still occur, even if social risk factors exist but the environment is unsuitable. Here, we use that model ("Model 4"), the widest extent predicted by their study, which also projects high environmental suitability in the Gulf region, very similar to that of Messina *et al.*. This model also produced isolated suitable patches based on social factors, which predominantly occur at urban centers.

Mordecai et al.

Mordecai *et al.*[10] produced a Bayesian model of transmission of *Aedes*-borne viruses (dengue, chikungunya, and Zika) in the Americas that we adapt as a mechanistic geographic forecast for subsequent analyses. In their main model, an R_0 modeling framework is constructed based on models for vector borne diseases, building upon the Kermack-McKendrick R_0 model for malaria[41]. In this model, the majority of parameters describing the life cycle of mosquitoes and parasite development within the mosquitoes are sensitive to temperature. Mordecai *et al.* used data derived from the literature to parameterize the shape of the temperature response for each temperature sensitive parameter. These are based on laboratory observations of *Aedes aegypti* and *Aedes albopictus*, and infections with dengue, chikungunya, and Zika at constant temperatures through the range of possible values. Because these are bioenergetic functions, curve fitting exercises to derive appropriate models of the non-linear relationships underlie the parameterization of the overall transmission model. A non-linear overall relationship between transmission (R_0) and temperature is fitted in a Bayesian inference framework, and from it two endpoints of a "suitable range" can be extrapolated within which $R_0 > 0$. Those ranges can be adjusted for different levels of posterior probability, and can be used as a suitability threshold that can be projected onto gridded temperature data. In the Mordecai *et al.* publication, the most conservative probability level ($> 97.5\%$) was then mapped onto long-term mean monthly average temperatures in the Americas, derived from Worldclim data[40], to estimate the number of months transmission was possible for *Ae. aegypti* and *Ae. albopictus*[10]. Additional maps were also constructed of the number of months of possible transmission for $R_0 > 0$ at posterior probabilities of 50% and 2.5%, and are found in the supplemental material. Here, we use all three probability levels from the

Ae. aegypti model, to project the terms of the number of months of predicted transmission potential by mapping the model onto WorldClim temperature gridded data for long-term monthly minimum and maximum temperatures (six possible combinations).

Consensus Mapping Methods

In a preliminary effort to present a consensus forecast based on current ecological understanding, we use two alternative methods to develop county-scale predictions from the models included in our analysis. The first ("majority rule") excludes the Mordecai model, and simply applies a majority rule to the Carlson, Messina, and Samy county shapefiles (i.e., any county with agreement between a majority of the niche models for either presence or absence). In the second model ("seasonal majority rule"), we take the counties predicted by the majority rule method and restrict their suitability to the months predicted in the strictest Mordecai model (97.5% confidence) for minimum temperatures. That process excludes 13 of the counties deemed suitable according to the simple majority rule, but which are predicted to be unsuitable year-round in the Mordecai model.

Epidemiological Model

To simulate potential Zika outbreaks in the United States, we adopt the modeling framework used by Gao *et al.*, which incorporates both sexual and vector-borne transmission[42]. We selected Gao *et al.*'s framework because, while fairly simple, it includes a number of important features of the epidemiology of Zika, including the high rate of asymptomatic cases, and lingering (primarily sexual) transmission by post-symptom "convalescent" cases. Because the transmission term is normalized by dividing by total population size, the model itself is scale-free. Thus, the values associated with each compartment could be represented as proportions rather than the number of individuals. The model divides the human population into six compartments with levels: susceptible (S), exposed (E), symptomatically (I_1) or asymptomatically (A) infected, convalescent (I_2), and recovered (R), where he h and v refer to the human host and mosquito vector populations, respectively:

$$\frac{dS_h}{dt} = -ab\frac{I_v}{N_h}S_h - \beta\frac{\kappa E_h + I_{h1} + \tau I_h 2}{N_h}S_h \tag{1}$$

$$\frac{dE_h}{dt} = \theta(ab\frac{I_v}{N_h}S_h + \beta\frac{\kappa E_h + I_{h1} + \tau I_h 2}{N_h}S_h) - \nu_h E_h \tag{2}$$

$$\frac{dI_{h1}}{dt} = \nu_h E_h - \gamma_{h1} I_{h1} \tag{3}$$

$$\frac{dI_{h2}}{dt} = \gamma_{h1} I_{h1} - \gamma_{h2} I_{h2} \tag{4}$$

$$\frac{dA_h}{dt} = (1-\theta)(ab\frac{I_v}{N_h}S_h + \beta\frac{\kappa E_h + I_{h1} + \tau I_h 2}{N_h}S_h) - \gamma_h A_h \tag{5}$$

$$\frac{dR_h}{dt} = \gamma_{h2} I_{h2} + \gamma_h A_h \tag{6}$$

$$N_h = S_h + E_h + I_{h1} + I_{h2} + A_h + R_h \tag{7}$$

Vectors are governed by a complimentary set of equations but only divided into susceptible, exposed, and infected classes:

$$\frac{dS_v}{dt} = \mu_v N_v - ac\frac{\eta E_h + I_{h1}}{N_h}S_v - \mu_v S_v \tag{8}$$

$$\frac{dE_v}{dt} = ac\frac{\eta E_h + I_{h1}}{N_h}S_v - (\nu_v + \mu_v)E_v \tag{9}$$

$$\frac{dI_v}{dt} = nu_v E_v - mu_v I_v \tag{10}$$

$$N_v = S_v + E_v + I_v \tag{11}$$

Rather than use the fitted parameters from any given country's outbreak, parameters for the above models were randomly generated from a set of uniform prior distributions specified by Gao *et al.* as reasonable priors based on the literature (Table 2). Evidently, these models are significantly discrepant with outbreaks in the continental U.S. so far, with fewer than 300 cases of local transmission recorded in 2016 (and in fact, our simulations are far more severe in terms of final case burden than estimates for Brazil or Colombia). However, the purpose of applying this epidemiological model across the spatial extent predicted by each niche model is both to illustrate the uncertainty that goes unstated in presenting such ENMs and to intimate the necessity of developing and parameterizing these models in concert.

County-Level Simulations

Every spatial model of Zika risk was summarized at the U.S. county scale, such that if a single pixel within a county polygon was projected to be suitable under a given model, the county was marked suitable for outbreaks. For the Mordecai models, the maximum value (months suitable per year) of all pixels within a county was assigned as the value. For example, if a single 25 km^2 cell in a particular county was suitable for a single month, simulations were run for one month with mosquitoes present and the remaining 11 with a mosquito population of zero. While this approach has the potential to overestimate populations vulnerable to mosquito-borne transmission, it adds a number of key strengths. Aggregating information at the county scale absorbs some of the relative spatial uncertainty of predictions at the pixel scale, and may account for source-sink dynamics for vector-borne outbreaks driven by heterogeneity in vector density and competence. Moreover, the county level is the finest standard scale at which public health infrastructure is likely to decide whether interventions like vector control are necessary. Finally, sexual transmission can spread from cells with suitable vectors to vector-free areas, and as a function of both sexual transmission and underlying mobility, outbreaks are therefore unlikely to be contained to a given pixel. Other Zika modeling work in the United States validates that the county scale is an appropriate resolution for considering the risk of sexual transmission[33,35].

Population data for each county was taken from projections to the year 2016 based on the 2010 United States Census, and were set as the total susceptible human population at the start of a year. The mosquito population was set at five times the baseline human population, as per Gao *et al.*[42]. Outbreaks were simulated stochastically at the county level using the Gao *et al.* model, initiated with a single infected person per county. We randomly selected a value for each of the parameters in the Gao *et al.* model for which a range was provided, using a uniform distribution (Table 3). For each modeling model, 100 simulations were run in each county designated as suitable. For the three ENMs, county models were run for a "model year" (twelve months of thirty days each), and had no interactive effect on each other. For the Mordecai models, the full vector- and sexually-transmitted epidemic models were run for the number of months (thirty days each) that were predicted suitable. After that period, the total vector population N_v was set to 0, effectively ending vector-borne transmission, but models continued so that sexual transmission was ongoing up to 360 days. All simulations were run in R 3.3.2, and all scripts and county simulation data are available as supplementary files.

Data Availability Statement

All data generated or analysed during this study are included in this published article (and its Supplementary Information files).

References

1. Morse, S. S. *et al.* Prediction and prevention of the next pandemic zoonosis. *The Lancet* **380**, 1956–1965 (2012).
2. Jetten, T. H. & Focks, D. A. Potential changes in the distribution of dengue transmission under climate warming. *The American journal of tropical medicine and hygiene* **57**, 285–297 (1997).
3. Escobar, L. E. & Craft, M. E. Advances and limitations of disease biogeography using ecological niche modeling. *Frontiers in Microbiology* **7** (2016).
4. Hernandez, P. A., Graham, C. H., Master, L. L. & Albert, D. L. The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography* **29**, 773–785 (2006).
5. Segurado, P. & Araujo, M. B. An evaluation of methods for modelling species distributions. *Journal of Biogeography* **31**, 1555–1568 (2004).

6. Qiao, H., Soberón, J. & Peterson, A. T. No silver bullets in correlative ecological niche modelling: insights from testing among many potential algorithms for niche estimation. *Methods in Ecology and Evolution* **6**, 1126–1136 (2015).

7. Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R. K. & Thuiller, W. Evaluation of consensus methods in predictive species distribution modelling. *Diversity and distributions* **15**, 59–69 (2009).

8. Johnson, L. R. *et al.* Understanding uncertainty in temperature effects on vector-borne disease: a bayesian approach. *Ecology* **96**, 203–213 (2015).

9. Kearney, M. & Porter, W. Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. *Ecology letters* **12**, 334–350 (2009).

10. Mordecai, E. *et al.* Temperature determines zika, dengue and chikungunya transmission potential in the americas. *bioRxiv* 063735 (2016).

11. Kearney, M. R., Wintle, B. A. & Porter, W. P. Correlative and mechanistic models of species distribution provide congruent forecasts under climate change. *Conservation Letters* **3**, 203–213 (2010).

12. Carlson, C., Dougherty, E. & Getz, W. An ecological assessment of the pandemic threat of zika virus. *PLoS Neglected Tropical Diseases* **10**, e0004968 (2016).

13. PAHO/WHO. Zika epidemiological report: Brazil. march 2017. Tech. Rep., Pan-American Health Organization, Washington, D.C. (2017).

14. PAHO/WHO. Zika suspected and confirmed cases reported by countries and territories in the americas cumulative cases, 2015-2017. updated as of 6 april 2017. Tech. Rep., Pan-American Health Organization, Washington, D.C. (2017).

15. Naucke, T. J. & Lorentz, S. First report of venereal and vertical transmission of canine leishmaniosis from naturally infected dogs in germany. *Parasites & vectors* **5**, 67 (2012).

16. Messina, J. P. *et al.* Mapping global environmental suitability for zika virus. *Elife* **5**, e15272 (2016).

17. Samy, A. M., Thomas, S. M., El Wahed, A. A., Cohoon, K. P. & Peterson, A. T. Mapping the global geographic potential of zika virus spread. *Mem Inst Oswaldo Cruz* .

18. Lindstrom, N. Zika cases top 200 in pennsylvania. *Tribune Live* URL https://shar.es/1QKP8O.

19. Chelsia Rose Marcius, D. R. Zika virus prompts mta plans for mosquito-killing larvicide in subway's stagnant pools of water, cuomo announces. *New York Daily News* URL http://nydn.us/2axoHxO.

20. Blinder, A. Aimed at zika mosquitoes, spray kills millions of honeybees. *New York Times* URL https://nyti.ms/2olguVh.

21. Blumberg, S. & Lloyd-Smith, J. O. Inference of r_0 and transmission heterogeneity from the size distribution of stuttering chains. *PLoS Comput Biol* **9**, e1002993 (2013).

22. Perkins, T. A., Scott, T. W., Le Menach, A. & Smith, D. L. Heterogeneity, mixing, and the spatial scales of mosquito-borne pathogen transmission. *PLoS Comput Biol* **9**, e1003327 (2013).

23. Peterson, A. T., Martínez-Campos, C., Nakazawa, Y. & Martínez-Meyer, E. 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 (2005).

24. Bogoch, I. I. *et al.* Potential for zika virus introduction and transmission in resource-limited countries in africa and the asia-pacific region: a modelling study. *The Lancet Infectious Diseases* **16**, 1237–1245 (2016).

25. Balcan, D. *et al.* Seasonal transmission potential and activity peaks of the new influenza a (h1n1): a monte carlo likelihood analysis based on human mobility. *BMC medicine* **7**, 45 (2009).

26. Ogden, N. H. *et al.* Risk of travel-related cases of zika virus infection is predicted by transmission intensity in outbreak-affected countries. *Parasites & Vectors* **10**, 41 (2017).

27. Coelho, F. C. *et al.* Higher incidence of zika in adult women than adult men in rio de janeiro suggests a significant contribution of sexual transmission from men to women. *International Journal of Infectious Diseases* **51**, 128–132 (2016).

28. Towers, S. *et al.* Estimate of the reproduction number of the 2015 zika virus outbreak in barranquilla, colombia, and estimation of the relative role of sexual transmission. *Epidemics* **17**, 50–55 (2016).

29. Yakob, L., Kucharski, A., Hue, S. & Edmunds, W. J. Low risk of a sexually-transmitted zika virus outbreak. *The Lancet infectious diseases* **16**, 1100–1102 (2016).

30. Baca-Carrasco, D. & Velasco-Hernández, J. X. Sex, mosquitoes and epidemics: An evaluation of zika disease dynamics. *Bulletin of Mathematical Biology* **78**, 2228–2242 (2016).

31. Allard, A., Althouse, B. M., Hébert-Dufresne, L. & Scarpino, S. V. The risk of sustained sexual transmission of zika is underestimated. *bioRxiv* 090324 (2016).

32. Folkers, K., Caplan, A. & Igel, L. Zika, sexual transmission and prudent public health policy. *Public Health* **148**, 66–68 (2017).

33. Shacham, E., Nelson, E. J., Hoft, D. F., Schootman, M. & Garza, A. Potential high-risk areas for zika virus transmission in the contiguous united states. *American journal of public health* e1–e8 (2017).

34. Regan, D. G. & Wilson, D. P. Modelling sexually transmitted infections: less is usually more for informing public health policy. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **102**, 207–208 (2008).

35. Castro, L. A. *et al.* Assessing real-time zika risk in the united states. *BMC Infectious Diseases* **17**, 284 (2017).

36. Kraemer, M. U. *et al.* The global distribution of the arbovirus vectors aedes aegypti and ae. albopictus. *Elife* **4**, e08347 (2015).

37. Lessler, J. *et al.* Assessing the global threat from zika virus. *Science* **353**, aaf8160 (2016).

38. Santos, J. & Meneses, B. M. An integrated approach for the assessment of the aedes aegypti and aedes albopictus global spatial distribution, and determination of the zones susceptible to the development of zika virus. *Acta tropica* **168**, 80–90 (2017).

39. Bogoch, I. I. *et al.* Anticipating the international spread of zika virus from brazil. *Lancet (London, England)* **387**, 335–336 (2016).

40. Hijmans, R., Cameron, S., Parra, J., Jones, P. & Jarvis, A. The worldclim interpolated global terrestrial climate surfaces. version 1.3 (2004).

41. Mordecai, E. A. *et al.* Optimal temperature for malaria transmission is dramatically lower than previously predicted. *Ecology letters* **16**, 22–30 (2013).

42. Gao, D. *et al.* Prevention and control of zika as a mosquito-borne and sexually transmitted disease: a mathematical modeling analysis. *Scientific reports* **6** (2016).

Acknowledgements

We thank Lewis Bartlett, Eva Harris, and others for helpful comments and guidance. This work was supported by the Rutgers University Center for Discrete Mathematics & Theoretical Computational Science (DIMACS) Mathematics of Planet Earth 2013+ Workshop on Zika & DIMACS MPE 2013+ Workshop on Appropriate Complexity Modeling of the Impacts of Global Change on Ecosystems, and the Centers for Disease Control and Prevention Epidemic Predictions Initiative (CDC EPI), a Center funded by NSF (EF-0553768). SJR was supported by NSF DEB EEID 1518681 and NSF DEB RAPID 1641145.

Author contributions statement

CJC conceived of the study, and contributed the ecological niche models. SJR contributed the mechanistic models. ERD designed code for epidemiological simulations. CJC and ERD contributed to figures. All authors contributed to the writing and editing of the manuscript.

Competing interests statement

The authors declare no competing interests.

Table 1. A comparison of the different ecological forecasts. Four different methods, each performing well based on sufficient data and predictors, produce highly contrasting results. Out of a total of 3108 counties in the continental U.S., only four have experienced outbreaks (Brownsville, TX; Miami-Dade, FL; Palm Beach, FL; and Broward County, FL), with 223 cases of local transmission (6 in TX, 217 in FL). The Carlson model comes closest to predicting the geography of those outbreaks most accurately; but all epidemiological models "overpredict" based on the current extent of outbreaks. (Mordecai results are split for the highest bound with minimum temperatures, and the lowest bound for maximum temperatures, to give the full range of predictions. NA = Not Applicable; NR = Not Reported)

	Carlson	Messina	Samy	Mordecai (97.5% min)	Mordecai (2.5% max)
n_{points}	242	323	NR	NA	NA
$n_{predictors}$	15	6	15	NA	NA
AUC	0.970	0.829	NR	NA	NA
Counties Predicted	13	465	1616	1937	3108
County Population at Risk	19,653,445	95,359,408	270,249,781	218,444,263	320,957,062
Mean Outbreak Size	12,871,005	63,622,367	181,290,371	37,598,099	198,910,979
Median Outbreak Size	14,552,250	64,038,273	181,732,629	37,312,233	197,731,918

Table 2. Majority rule based consensus models, meant to resolve uncertainty between the forecasts and provide a middle scenario. The main majority rule model combines the Carlson, Messina, and Samy forecasts; the seasonal majority rule model assigns monthly suitability values to that forecast, based on the minimum temperature 97.5% Mordecai model.

	Majority Rule	Seasonal Majority Rule
Counties Predicted	383	370
County Population at Risk	93,195,970	87,632,865
Mean Outbreak Size	60,259,904	24,267,441
Median Outbreak Size	60,940,111	24,407,885

Table 3. Epidemiological parameters for the SEIIAR & SEI models presented in the main text; values taken directly from from Table 1 of Gao *et al.*'s (2017) study.

Parameter	Description	Range
a	Mosquito biting rate ($mosquito^{-1}day^{-1}$)	$(0.3, 1)$
b	Mosquito to human transmission rate ($bite^{-1}$)	$(0.1, 0.75)$
c	Human to mosquito transmission rate ($bite^{-1}$)	$(0.3, 0.75)$
β	Human to human (sexual) transmission rate	$(0.001, 0.1)$
$1/\gamma_{h1}$	Infectious period (acute)	$(3, 7)$
$1/\gamma_{h2}$	Infectious period (convalescent)	$(14, 30)$
$1/\gamma_h$	Infectious period (asymptomatic)	$(5, 10)$
η	Exposed human to mosquito transmission proportion	$(5, 10)$
θ	Proportion symptomatic infections	$(0.1, 0.27)$
κ	Exposed human to human transmission proportion	$(0, 1)$
$1/\mu$	Mosquito lifespan (day)	$(4, 35)$
$1/\nu_h$	Intrinsic incubation period (day)	$(2, 7)$
$1/\nu_v$	Extrinsic incubation period (day)	$(8, 12)$
τ	Convalescent human to human transmission proportion	$(0, 1)$

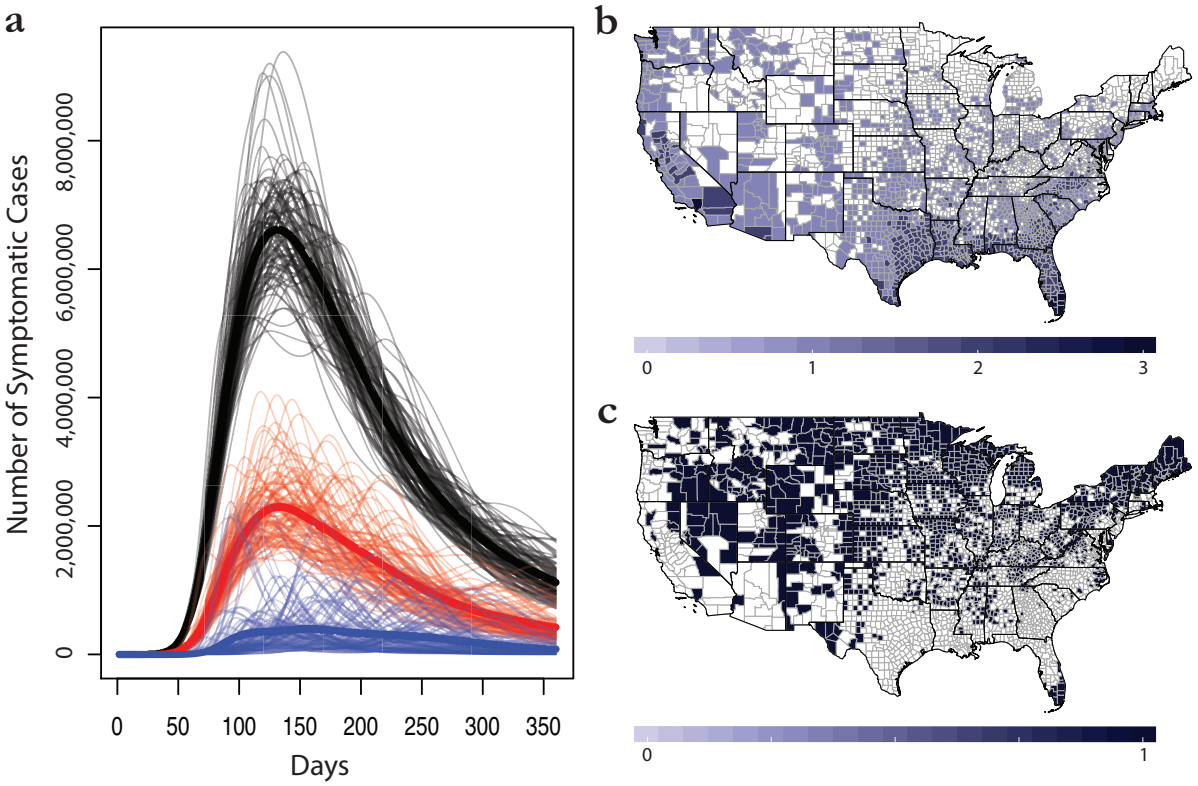

Figure 1. The margin of error in ecological niche models for Zika virus. (a) Average epidemiological forecasts associated with county data for Carlson (*blue*), Messina (*red*), and Samy (*black*), against a backdrop of overlapping individual simulations for each (grey). (b) The individual predictions of each model are given as presence or absence values; a maximum score of 3 indicates all models agree on presence, while a score of 0 indicates all models agree on absence. (c) Has consensus been achieved? At the county scale, dark blue indicates consensus among niche models; white indicates controversy.

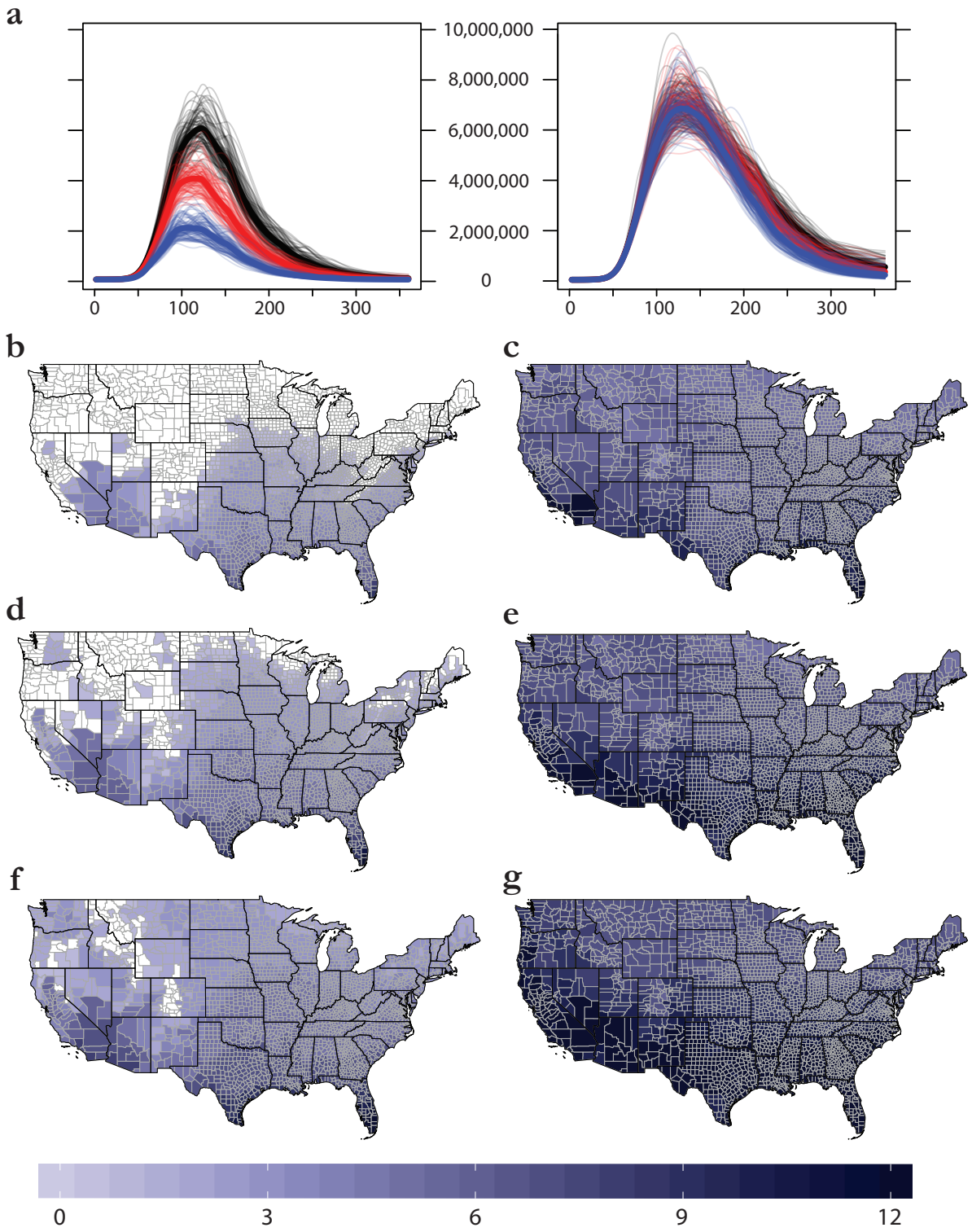

Figure 2. The margin of error within a single Bayesian mechanistic model for Zika virus, applied to minimum (left) and maximum (right) monthly temperatures. (a) 100 outbreak simulations for 97.5% (blue), 50% (red), and 2.5% (black) confidence intervals. (b-f) The number of months each county is predicted to be suitable for Zika virus transmission ($R_0 > 0$) for 97.5% (b,c), 50% (d,e), and 2.5% (f,g) scenarios.

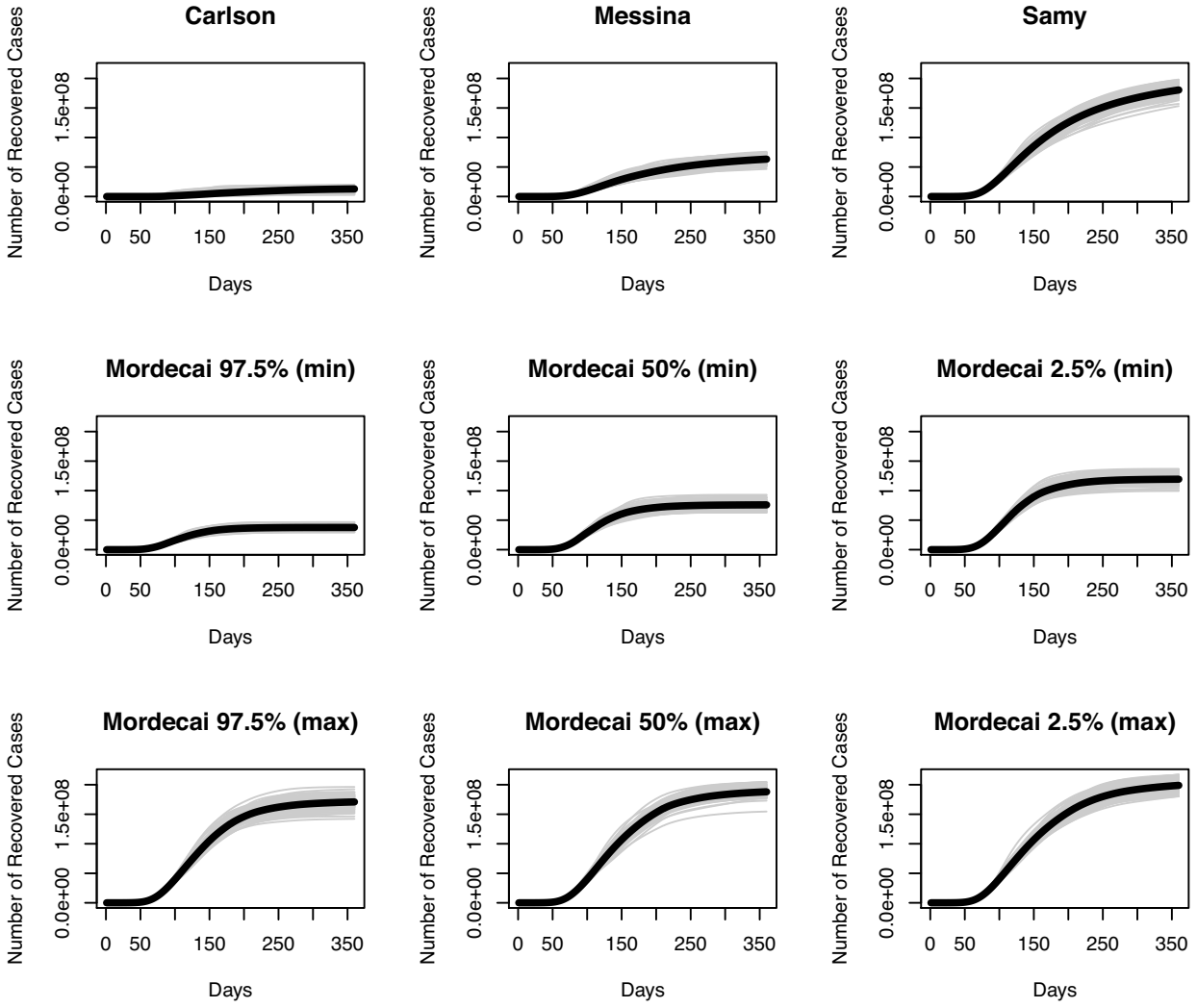

Figure 3. Nine possible trajectories for outbreaks in the United States: three based on ecological niche models, and six based on Bayesian mechanistic forecasts. (*y-axis on log scale*)

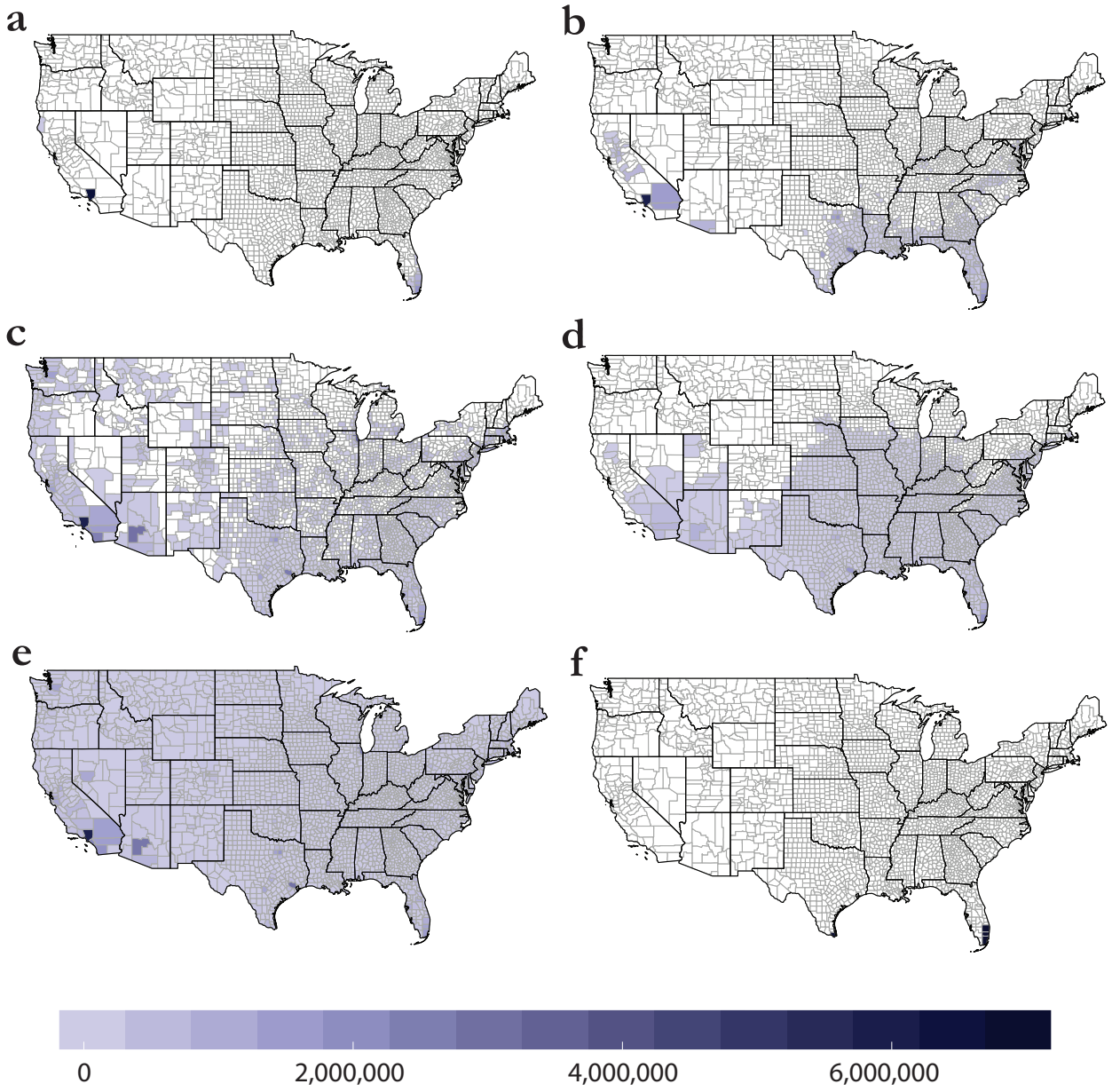

Figure 4. Case totals by county for (a) Carlson, (b) Messina, (c) Samy, (d), Mordecai 97.5% confidence (minimum temperatures), and (e) Mordecai 2.5% confidence (max temperatures), compared against (f) counties with reported autochthonous transmission in 2016 (three in Florida, one in Texas).

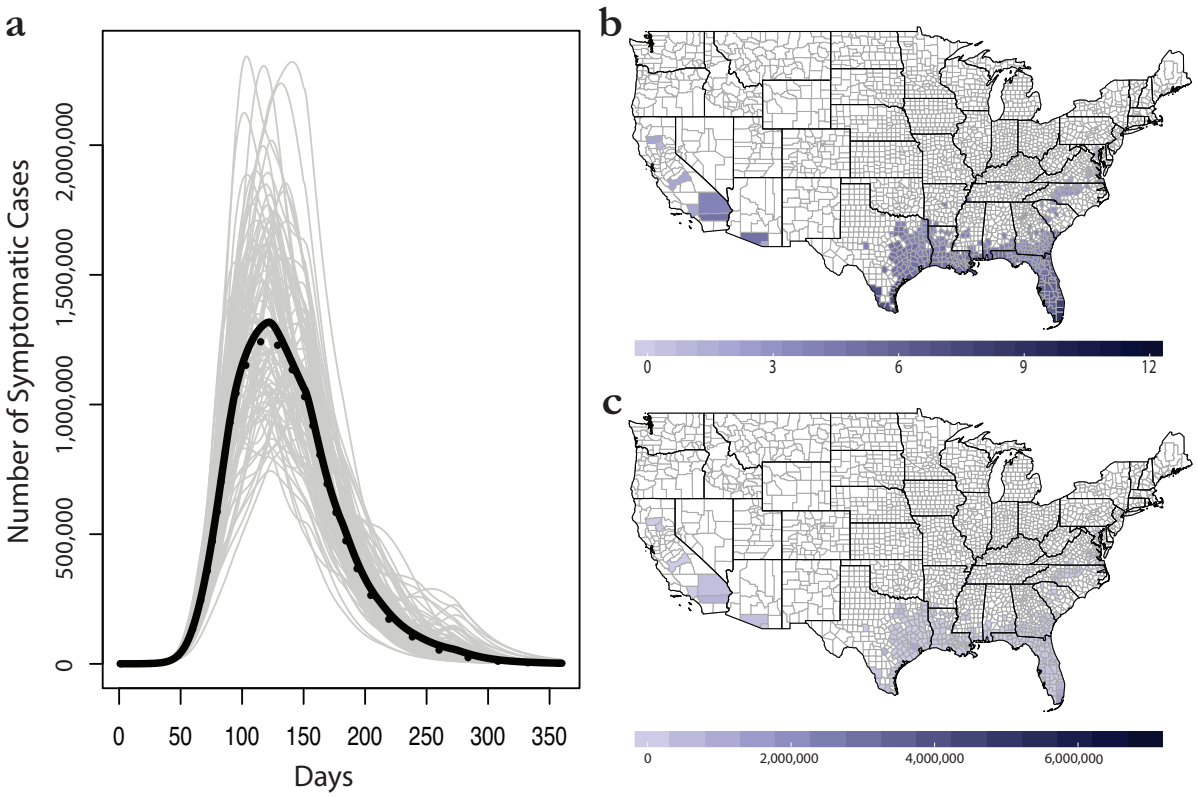

Figure 5. The seasonal majority rule method for consensus building across ecological forecasts. (a) Mean (black) and median (dashed) trajectories for 100 epidemic simulations. (b) The majority rule map: shading represents the number of months each county is marked suitable for outbreaks. (c) Final average case totals in the seasonal majority rule method.