

# Perspectives on Applying Artificial Intelligence to Model Outbreak Risk: Contextualizing The Impact of Climate Change on Vector-borne Disease

Albert A Gayle<sup>1</sup>

1. Department of Public Health and Clinical Medicine, Section of Sustainable Health, Umeå University, SE-90187 Umeå, Sweden

## Abstract

As recent history has shown, changing climate not only threatens to increase the spread of known disease, but also the emergence of new and dangerous phenotypes. This occurred most recently with West Nile virus: a virus previously known for mild febrile illness rapidly emerged to become a major cause of mortality and long-term disability throughout the world. As we move forward, into increasingly uncertain times, public health research must begin to incorporate a broader understanding of the determinants of disease emergence – what, how, why, and when. The increasing mainstream availability of high-quality open data and high-powered analytical methods presents promising new opportunities. Up to now, quantitative models of disease outbreak risk have been largely based on just a few key drivers, namely climate and large-scale climatic effects. Such limited assessments, however, often overlook key interacting processes and downstream determinants more likely to drive local manifestation of disease. Such pivotal determinants may include local host abundance, human behavioral variability, and population susceptibility dynamics. The results of such analyses can therefore be misleading in cases where necessary downstream requirements are not fulfilled. It is therefore important to develop models that include climate and higher-level climatic effects alongside the downstream non-climatic factors that ultimately determine individual disease manifestation. Today, few models attempt to comprehensively address such dynamics: up until very recently, the technology simply hasn't been available. Herein, we present an updated overview of current perspectives on the varying drivers and levels of interactions that drive disease spread. We review the predominant analytical paradigms, discuss their strengths and weaknesses, and highlight promising new analytical solutions. Our focus is on the prediction of arboviruses, particularly West Nile virus, as these diseases represent the pinnacle of epidemiological complexity – solution to which would serve as an effective “gatekeeper”. We present the current state-of-the-art with respect to known drivers of arbovirus outbreak risk and severity, differentially highlighting the impact of climate and non-climatic drivers. The reality of multiple classes of drivers interacting at different geospatial and temporal scales requires advanced new methodologies. We therefore close out by presenting and discussing some promising new applications of AI. Given the reality of accelerating disease risks due to climate change, public health and other related fields must begin the process of updating their research programs to incorporate these much needed, new capabilities.

**Keywords:** climate change; vector-borne disease; artificial intelligence; explainable AI; geospatial modeling; infectious disease; arbovirus

## Introduction

Climate change and variability has now been implicated in the range expansion of a number of vector-borne infectious diseases across the globe<sup>1</sup>. Copious empirical studies have been conducted to understand which variables contribute most to emergence and spread of disease<sup>2</sup>. Such studies generally aim to establish statistical relationships between long-term climate trends, contemporaneous meteorological conditions, and disease. Some studies will include the presence and deterministically relevant activities of known disease vectors. This particular theme has been covered quite exhaustively: strong statistical and causal links have been demonstrated between vectors (mosquitoes and other arthropods) and climatic conditions, such as temperature and rainfall. This has, in turn, been linked to the incidence of vector-borne diseases, such as dengue and West Nile virus. And such information has indeed proven valuable for prediction of outbreak severity and proactive intervention<sup>3,4</sup>. However, there are few who would argue that the current degree of geospatial and temporal precision of prediction does not need to be improved. Up to now, progress beyond the status quo has been stymied by methodological limitations – traditional research methodologies were simply not designed to handle the volume of data needed to deliver precision required<sup>5</sup>. Recent methodological advancements along with increased open data availability promise to change this<sup>6</sup>. Herein, we present an updated overview of current perspectives on the varying drivers and levels of interactions that drive disease spread. We review the predominant analytical paradigms, discuss their strengths and weaknesses, and highlight promising new analytical solutions. Our focus is on the prediction of arboviruses, particularly West Nile virus, as these diseases represent the pinnacle of epidemiological complexity – solution to which would serve as an effective “gatekeeper”. We present the current state-of-the-art with respect to known drivers of arbovirus outbreak risk and severity, differentially highlighting the impact of climate and non-climatic drivers. The reality of multiple classes of drivers interacting at different geospatial and temporal scales requires advanced new methodologies.

## Modeling of climate and infectious diseases

The past 20 years have been characterized by a notable trend in the spread of infectious disease into previously naïve populations; and this spread has been projected to continue in line with climate change forecasts<sup>7</sup>. The enhancement and spread of infectious disease is driven by climate in various ways<sup>8</sup>. One notable example is the ongoing change to the habitats and behavior of both hosts and vectors<sup>9–11</sup>. This has been most notable in the case of mosquito borne disease, particularly those capable of being carried long-distance by regular host migration<sup>12–14</sup>. Climate change also has potential impacts on human mobility, which has an obvious association with disease spread<sup>15–17</sup>.

Much research attention has been focused on more precisely understanding these dynamics. There are two broad categories of quantitative models applied for this purpose<sup>18</sup>: statistical and process-oriented. Statistical models focus mainly on the analysis of time-series and patterns associated with disease outbreak data<sup>19</sup>. However, statistical models lack any explicit structure with respect to mechanistic drivers. As a result, while generally regarded as having high predictive power, they tend to suffer from problematic generalizability<sup>5</sup>. Such models are often used for development of early warning systems and for intervention (resource) planning<sup>20</sup>. An extension of the statistical modeling paradigm is referred to as statistical, or machine, learning. Such models take advantage of advances in computational power to allow for a more data-driven approach that obviates the need for *a priori* model specification and feature selection<sup>21</sup> (Simonsen L., et al. 2016). Decision trees are an increasingly popular implementation that provide several convenient advantages (*Table 1*). These models have met with great success in ecological modeling, where precise assumptions with respect to temporal associations and causality are of a lesser priority than pure predictive power and feature discrimination<sup>22</sup>. Classification trees are also becoming increasingly popular for geospatial disease

risk modeling<sup>23</sup>, which share many similarities to ecological models in terms of methodology and predictors<sup>24</sup>.

<b>Decion tree methods</b>	<b>Description</b>
Decision Trees	A graphical representation of possible solutions to a decision based on certain conditions or thresholds being met. Each decision path is guided by the degree to which the evidence provided meets or exceeds successive threshold conditions.
Bagging (or Bootstrapping)	An ensemble meta-algorithm that combines predictions from multiple decision trees algorithmically, using a majority voting heuristic. Models can be based on subsets drawn randomly from a larger dataset. This is employed in cross-validation.
Random Forest	Similar to bagging but allows for feature randomization. Each model to represent a differet set of features. Multiple decsion tree are generated with a randomized feature set and then the results algorithmically combined to generate a prediction.
Boosting	Extends either the previous approaches by allowing each new model to learn from previous iterations. Error minimization heuristics guide each iteration: models with high validity are "boosted" and sequentially augmented with new features
Gradient Boosting	Boosting based algorithm that employs predictive heuristics (gradient descent) to guide each successive model iteration.
Extreme Gradient Boosting	Employs various internal optimizations to improve execution times by up to a factor of 100 and increase the stability of estimations. This allows for larger data sets to be applied.

Table 1: Such models are simple to run, deliver accurate predictions, and provide for easy interpretation<sup>22</sup>. They are being increasingly applied in many fields, including infectious disease modeling<sup>21</sup>. A major advantage of these methods is the extreme tolerance for data sets that are large and of disparate data types. Categorical and continuous variables can be modeled side-by-side with no need for pre-processing. Input variables are automatically stratified or "binned" and so dummy variables do not need to be encoded and outliers do not need to be removed. Missing values are similarly handled. Objective functions can include either classification or regression—or any other model for which a linear formulation exists, e.g the Cox proportional hazards model. This result is a statistically robust modeling process that allows for more data to be used with less overall effort. However, these models do not provide parameter estimates out of the box. As a result, there is limited value for applications that require both direction and magnitude of effect. These models are also prone to over-fitting and often require careful curation to ensure out-of-sample validity. "eXtreme Gradient Boosting" (XGB) is a recently introduced method that has shown exceptional promise with respect to the limitations cited above. Extensions to this method allow for robust effect estimation as well<sup>25</sup>. XGB is now regarded by many as having made all prior variants obsolete.

By comparison, process-oriented models set out to more precisely quantify and characterize the relationships between observed and suspected causal determinants of outbreak risk as well as other associated factors<sup>26</sup>. Such mechanistic models attempt to extend theory and practical knowledge to the problem of understanding, predicting, and controlling the factors ultimately influencing the manifestation and spread of disease<sup>27</sup>. Questions that might be addressed within the context of such models include—a) association between vector abundance and disease manifestation<sup>28</sup>, b) climate-

and environment-associated biotic factors that influence “a”<sup>29</sup>, c) population vulnerability factors that influence “b”<sup>30,31</sup>, d) human behavioral and mobility patterns that influence “c”<sup>16,32</sup>, and e) the effects of various control/clinical interventions on all of the above<sup>33</sup>. However, it is often very difficult to parametrize such theory-driven models – i.e., precisely align them to the reality observed by means of empirical data (Figure 1).

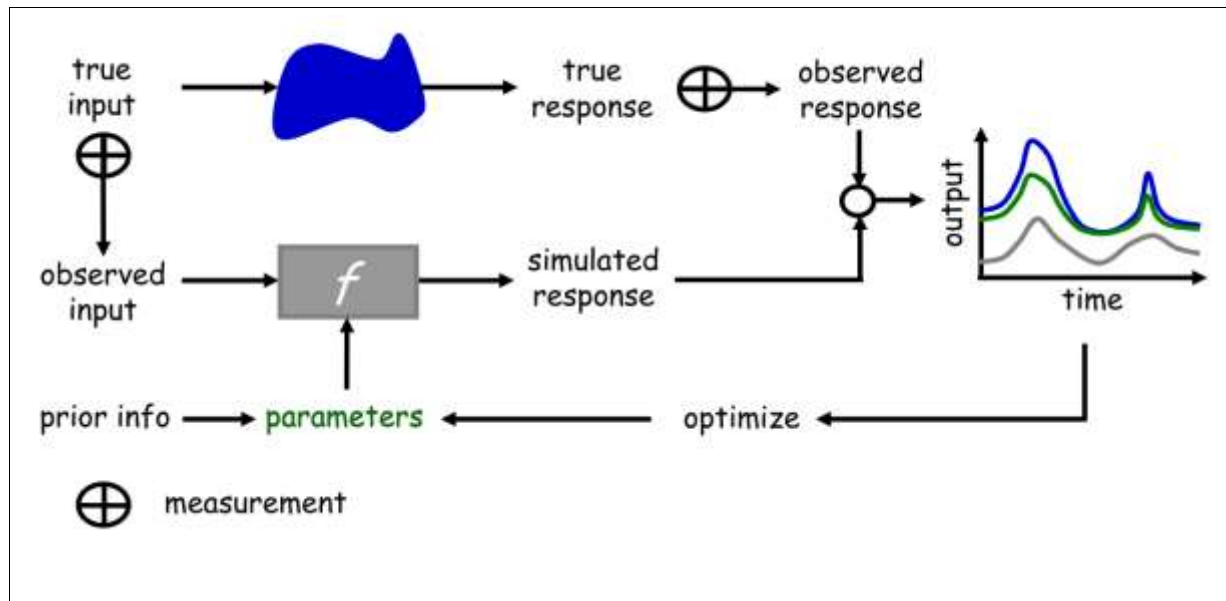


Figure 1. Parameterization can be seen as an iterative process that begins with functional assumptions and attempts to improve them by collecting observations. However, this often requires data that is simply unavailable due to resource constraints or scale. Effective parameterization requires data and time. Researchers are therefore often forced to rely on limited data or even pure supposition. Parameter values may often also differ according to the spatiotemporal scope being considered. Such models must therefore often be recalibrated when applied to contexts beyond their original scope. This problem tends to increase with model complexity. Parameterization under such circumstances therefore often substantial additional efforts in terms of data collection and computation. This is, of course, preferable to having too little detail in the model, as the mechanisms driving the response of interest may not be sufficiently represented. Luo et al (2011)<sup>34</sup> offers a solution in the form of data assimilation (DA), also referred to as online learning. DA uses streamed data to 1) continuously update model parameters based on observed data, and 2) quantify and integrate errors and patterns in unexplained variability. DA can be thought of as a machine-learning based form of continuous parameter optimization. DA be deployed to incorporate a broader spectrum of data, in real-time, for predictions that are more reliable across contexts and scale.

Beyond that, climate change scenarios based on long-term variability as well as anthropogenic causes have been used to estimate the long-term expansion or contraction of geospatial disease risks, particularly those associated with arthropod vectors (i.e., arboviruses). Such studies have found positive associations between determinants of arbovirus outbreak risk and climate change, in terms of a) changes to vector habitat suitability<sup>35</sup>, b) socioeconomic development<sup>36</sup>, and c) anthropogenic changes to the environment<sup>37</sup>. Human mobility and increased trade has been shown to be associated with increased outbreak risk<sup>38</sup>, and has in fact been linked to the introduction of both disease<sup>32</sup> and disease-competent vectors in Europe<sup>39</sup>. Taken together, this clearly indicates the need to develop models that are comprehensive both in terms of mechanistic assumptions and included predictor data.

Despite the history of successful research efforts, progress has been stymied by the frequent observation that, for many climate-related variables, associations are time-dependent and appear to change depending on time-lag and seasonality<sup>18</sup>. Moreover, effect estimates are known to vary as a

function of the geospatial scope and scale assumed for analysis<sup>40,41</sup>. In addition, the impact of underreporting and detection is an ongoing concern. Barker (2019) attributes this to stochastic noise due to under reporting of mild symptoms and the otherwise limited number of reported cases. This creates imprecision in the predictive object used to train the model and can lead to difficulties in both calibration and interpretation<sup>42</sup>. Beyond that, efforts to understand the role of underlying drivers of climatic variability have up to now been limited. Such drivers have been shown to contribute to the spread of infectious disease in both the American and Old World contexts. ENSO (“el Niño Southern Oscillation”) is well established as a driver of dengue and malaria<sup>43,44</sup>. Likewise, NAO (“North Atlantic Oscillation”) has been associated with a host of infectious diseases in Europe<sup>45</sup>. Predictions of the emergence and severity of disease that are spatiotemporally precise are therefore difficult to achieve in practice. This difficulty increases in direct proportion to the complexity of the dynamics associated with disease transmission. Transmission dynamics are relatively simple in cases where transmission is driven primarily by host-to-host contact – e.g., influenza – but unfortunately this is not always the case.

### **Climate and West Nile Virus as an example**

West Nile Virus (WNV) has attracted much attention as of late for this very reason<sup>46</sup>. In the United States alone, it has caused over 50,830 reported cases of human disease since its introduction in 1999. About half of these (24,657) have been of the deadly neuroinvasive variant, which boasts a death rate of about 9%<sup>47</sup>. This includes 5674 cases and 286 deaths during a widespread outbreak in 2012. The virus has long been present in Europe but outbreaks have been sporadic and relatively mild. This suddenly changed in 1996 where a deadly outbreak of the neuroinvasive variant was responsible for about 400 cases in Romania alone<sup>48</sup>. The lineage 2 strain then emerged in 2008 and has since spread rapidly across the continent, with about 10,647 cases confirmed to date<sup>49</sup>. This spread has been attributed to changes to conditions in the northern latitudes of both Europe and America that have increased the capacity for endemic mosquito-vectors to transmit the disease<sup>50</sup>. The invasion and establishment of non-native, but disease-competent mosquito species has likewise also been associated with improved habitat suitability due to weather that is warmer and wetter<sup>51–53</sup>. However, this alone fails to explain the noted increase in syndromic severity as a proportion of infected cases.

Within the European context, multiple species of mosquito vectors have been associated with the WNV transmission cycle. In addition to importation via migrant hosts, local amplification requires suitable local reservoir as well. However, the extent to which specific resident species impact WNV outbreak risk remains largely a mystery in the European context. Contrasted to the United States, where several reservoir species have been firmly established, the situation in Europe is still a matter of much uncertainty. A broad range of possible amplification hosts have been suggested: in addition to birds, amphibians and reptiles have also been found to contribute to local amplification. Mammals, which include humans, horses, and even domestic pets generally do not develop levels of viremia sufficient to transmit the virus to biting mosquitoes. While transmission of WNV generally requires a mosquito vector, the literature has cited dozens of discrete interactions that contribute to WNV outbreak risk. For example, host-to-host transmission has been observed among both humans and birds.

		Transmission (From)				
		Migratory Hosts	Local Reservoir Hosts	Host-specific vectors	Opportunistic vectors	Humans
Infection (to)	Migratory Hosts	1. contact, 2. food (oral), 3. water (fecal)	1. contact, 2. food (oral), 3. water (fecal)	1. Blood-meal	1. Blood-meal	x
	Local Reservoir Hosts	1. contact, 2. food (oral), 3. water (fecal)	1. contact, 2. food (oral), 3. water (fecal)	1. Blood-meal	1. Blood-meal	x
	Host-specific vectors	1. Blood-meal	1. Blood-meal	1. Vertical transmission	x	x
	Opportunistic vectors	1. Blood-meal	1. Blood-meal	x	1. Vertical transmission	x
	Humans	x	x	x	1. Blood-meal	1. Blood donation, 2. organ transplant, 3. transplacental, 4. breast feeding, 5. sexual contact

Figure 2. Simplified transmission matrix for West Nile virus involves 25 cells, each representing routes that contribute to the WNV transmission cycle for a given interaction<sup>54,55</sup>.

This degree of determinant complexity is difficult to handle by means of traditional process oriented or statistical models. Multicollinearity tends to be a function of model complexity, and often leads to difficulty in model construction or interpretation<sup>56</sup>. Geospatial or scale-dependent effect variability is another notable limitation<sup>57</sup>: it is an unavoidable statistical fact that estimations will vary according to the scope and size of the sample. Spatiotemporal precision is further hampered by the lack of quality data at sufficient levels of granularity and completeness<sup>20</sup>. One consequence is that features with greater levels of data availability, such as climate, are most convenient for disease modeling and tend to be highlighted accordingly. Meanwhile, features with high degrees of local variability – and therefore more difficult to accurately sample – tend to be omitted or overgeneralized. This includes mechanistically critical factors associated with amplification hosts, vectors, and human susceptibility<sup>58</sup>. And process-oriented models suffer from the same limitations: absent sufficient volume and variety of observational data, such models cannot be effectively parameterized.

### Climate in the context of other drivers of outbreak risk

Beyond climate, several key components are needed to understand the dynamics that determine the emergence and spread of infectious disease. These include non-climate drivers which facilitate 1) introduction and spread, 2) potential evolution, and 3) the susceptibility of available host populations. Such drivers facilitate transmission through an interactive pathway of underlying mechanisms and processes. Each of these factors enjoy differing degrees of causal proximity: some causal effects occur proximal to and in direct connection to the event, others occur further upstream. All drivers of outbreak risk can be understood to work synergistically: the effect of one is often mediated through processes linked to other drivers. This often makes it difficult to isolate and quantify independent effects. Cross-scale effects, such as the interaction between climate and the behavior of local vectors and hosts, are another area where understanding is limited<sup>41</sup>. As previously described, the most causally direct drivers exhibit high geospatial variability and localization, and available data is limited<sup>59</sup>. New methods are therefore needed that can make better use of available data, and better capture and quantify these crucial local effects, both independently and in synergy with or downstream from other drivers<sup>60</sup>. Promising new applications of machine learning and AI may represent possible solutions, but are presently only in the early stages of development.

(`Synthetic data`: AI-derived, based on what the underlying, determinant data `should` look like, given the prediction: free of statistical error or biases, therefore better fit to statistical realities)

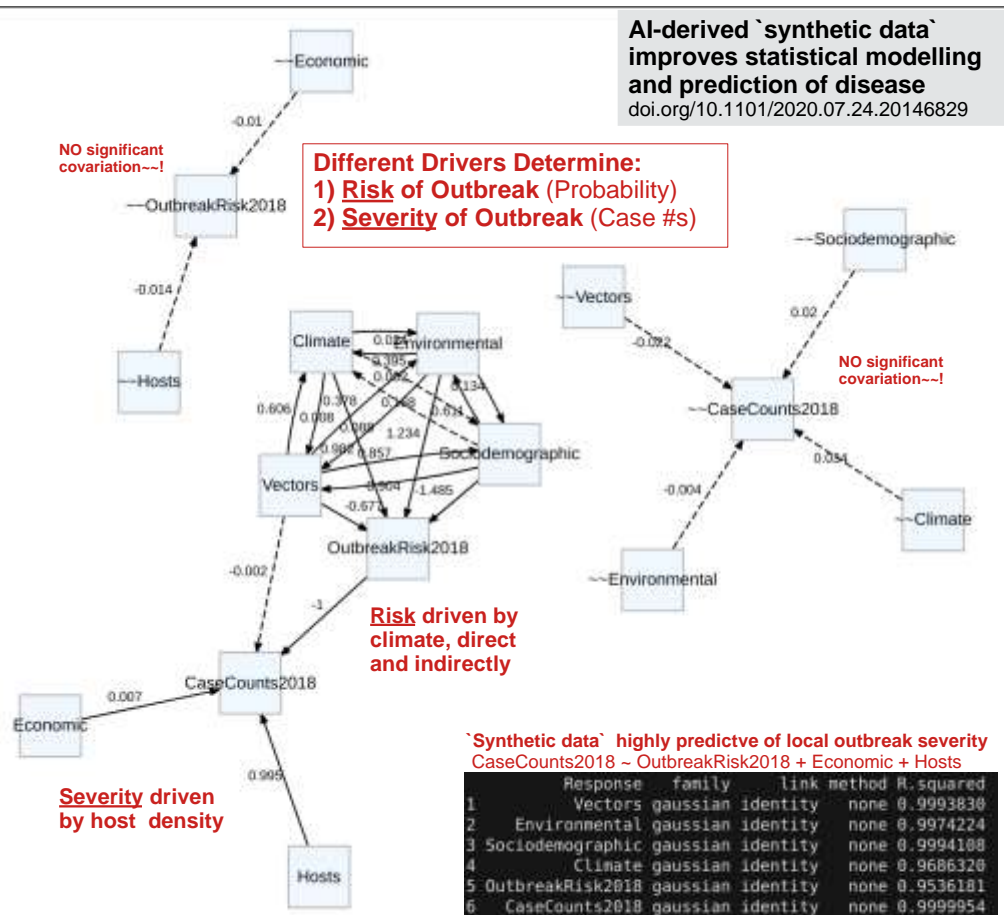


Figure 3. AI-powered confirmatory path analysis suggests potential causal associations

One distal effect associated with climate pertains to viral evolution. Changes to viral fitness have been cited as responsible for increases in the spread and neurotropic severity of WNV in Europe since 2008<sup>61</sup>. This mutated “lineage 2” has since supplanted the previous variant and has been responsible for increasing disease burden since<sup>46</sup>. Historically, genetic drift and evolution have been recognized as important factors for epidemics and spread, and will likely continue to play an important role in the future. The potential relationship between climate change and the viral evolution of arboviruses has not garnered much research attention. However, given the well-established dependence of vectorial capacity on temperature, viruses could very well likely experience accelerated transmission and replication cycles, therefore facilitating more rapid evolutionary processes. This would therefore imply that changing climate not only increases the risk of emergence and spread, but also the rate at which new, more virulent phenotypes potentially emerge.

### Where do we go from here?

The degree to which such challenges limit performance of a given model is in direct proportion to the complexity and variability of the system being modeled. And no other infectious disease can claim the degree of transmission chain complexity and variability as demonstrated by WNV. WNV has consequently been declared “notoriously hard to predict” and likened to an “impetuous child”<sup>62</sup>. And given the reality of sudden and large surges in outbreak severity and range, such as that observed in 2018, new solutions are clearly required. That said, WNV is most definitely a “worst case” modeling target – current technology is more than sufficient for the most common diseases (i.e., influenza). However, as demonstrated by the COVID-19 crisis, new pathogenic threats – with unknown transmission dynamics – can emerge at any time. It is therefore imperative that predictive modeling capabilities can cope with worst case dynamics – at a minimum. Looking forward, old modeling

paradigms can no longer be relied upon, absent substantial efforts to allow for predictive precision at local scales by expanding the analytical scope of dynamics addressed.

Recently developments in data availability and computational approaches have been suggested to hold the answer<sup>63</sup>. Supported by the increasing availability of high-resolution data from open sources<sup>64</sup>, models that incorporate an interdisciplinary assortment of drivers are becoming a reality<sup>65,66</sup>. Looking forward, such efforts are expected to evolve to incorporate state-of-the-art predictive algorithms with novel methods for deriving effect estimates and exploring underlying associations. For example, a novel method has recently been introduced that generates statistically optimal local model explanations<sup>25</sup>. Local effect estimation allows for global model structure to be assessed at heretofore unimaginable level of analytical precision and granularity<sup>67</sup>. This will allow for more precise tracking and prediction of outbreaks as well as greater ability to predict disease burden, especially within the context of the threat of climate change and the likely effect it may have with respect to the emergence of new pathogenic threats.

## Acknowledgments

A.A.G. received funding from the European Center for Disease Control and Prevention (contract no ECDC.9504). We would like to thank the European Center for Disease Control for kindly providing funding, case and vector data, and support. And special thanks to the team at Exploratory.io who provided technical guidance and support.

## Author contributions

The authors worked collaboratively and contributed equally to the article.

## Competing interests

The authors declare no competing interests.

## Citations

- Esser, H. J. *et al.* Risk factors associated with sustained circulation of six zoonotic arboviruses: a systematic review for selection of surveillance sites in non-endemic areas. *Parasit. Vectors* **12**, 265 (2019).
- Paz, S. Effects of climate change on vector-borne diseases: an updated focus on West Nile virus in humans. *Emerg. Top. Life Sci.* **3**, 143–152 (2019).
- Rocklöv, J. & Dubrow, R. Climate change: an enduring challenge for vector-borne disease prevention and control. *Nat. Immunol.* **21**, 479–483 (2020).
- Hussain-Alkhateeb, L. *et al.* Early warning and response system (EWARS) for dengue outbreaks: Recent advancements towards widespread applications in critical settings. *PLOS ONE* **13**, e0196811 (2018).
- Barker, C. M. Models and Surveillance Systems to Detect and Predict West Nile Virus Outbreaks. *J. Med. Entomol.* **56**, 1508–1515 (2019).
- Potter, K. M., Koch, F. H., Oswalt, C. M. & Iannone, B. V. Data, data everywhere: detecting spatial patterns in fine-scale ecological information collected across a continent. *Landsc. Ecol.* **31**, 67–84 (2016).



7. Semenza, J. C. & Suk, J. E. Vector-borne diseases and climate change: a European perspective. *FEMS Microbiol. Lett.* **365**, fnx244 (2018).
8. Paz, S. & Semenza, J. C. Environmental Drivers of West Nile Fever Epidemiology in Europe and Western Asia—A Review. *Int. J. Environ. Res. Public Health* **10**, 3543–3562 (2013).
9. Robson, D. & Barriocanal, C. Ecological conditions in wintering and passage areas as determinants of timing of spring migration in trans-Saharan migratory birds. *J. Anim. Ecol.* **80**, 320–331 (2011).
10. Koch, L. K. *et al.* Modeling the habitat suitability for the arbovirus vector *Aedes albopictus* (Diptera: Culicidae) in Germany. *Parasitol. Res.* **115**, 957–964 (2016).
11. Haest, B., Hüppop, O., Pol, M. van de & Bairlein, F. Autumn bird migration phenology: A potpourri of wind, precipitation and temperature effects. *Glob. Change Biol.* **25**, 4064–4080 (2019).
12. Medlock, J. M. & Leach, S. A. Effect of climate change on vector-borne disease risk in the UK. *Lancet Infect. Dis.* **15**, 721–730 (2015).
13. Van Den Bossche, D. *et al.* Chikungunya virus and West Nile virus infections imported into Belgium, 2007–2012. *Epidemiol. Infect.* **143**, 2227–2236 (2015).
14. Gabriel, M. *et al.* Increase in West Nile virus infections imported to Germany in 2012. *J. Clin. Virol.* **58**, 587–589 (2013).
15. Struchiner, C. J., Rocklöv, J., Wilder-Smith, A. & Massad, E. Increasing Dengue Incidence in Singapore over the Past 40 Years: Population Growth, Climate and Mobility. *PLOS ONE* **10**, e0136286 (2015).
16. Arthur, R. F., Gurley, E. S., Salje, H., Bloomfield, L. S. P. & Jones, J. H. Contact structure, mobility, environmental impact and behaviour: the importance of social forces to infectious disease dynamics and disease ecology. *Philos. Trans. R. Soc. B Biol. Sci.* **372**, 20160454 (2017).
17. Ramadona, A. L., Tozan, Y., Lazuardi, L. & Rocklöv, J. A combination of incidence data and mobility proxies from social media predicts the intra-urban spread of dengue in Yogyakarta, Indonesia. *PLoS Negl. Trop. Dis.* **13**, e0007298 (2019).
18. Davis, J. K. *et al.* Improving the prediction of arbovirus outbreaks: A comparison of climate-driven models for West Nile virus in an endemic region of the United States. *Acta Trop.* **185**, 242–250 (2018).
19. Trawinski, P. R. & Mackay, D. S. Meteorologically conditioned time-series predictions of West Nile virus vector mosquitoes. *Vector-Borne Zoonotic Dis.* **8**, 505–522 (2008).
20. Morin, C. W. *et al.* Unexplored Opportunities: Use of Climate- and Weather-Driven Early Warning Systems to Reduce the Burden of Infectious Diseases. *Curr. Environ. Health Rep.* **5**, 430–438 (2018).
21. Simonsen, L., Gog, J. R., Olson, D. & Viboud, C. Infectious Disease Surveillance in the Big Data Era: Towards Faster and Locally Relevant Systems. *J. Infect. Dis.* **214**, S380–S385 (2016).

22. Elith, J., Leathwick, J. R. & Hastie, T. A working guide to boosted regression trees. *J. Anim. Ecol.* **77**, 802–813 (2008).
23. Hay, S. I. *et al.* Global mapping of infectious disease. *Philos. Trans. R. Soc. B Biol. Sci.* **368**, 20120250 (2013).
24. Hess, A., Davis, J. K. & Wimberly, M. C. Identifying environmental risk factors and mapping the risk of human West Nile virus in South Dakota. *AGU Fall Meet. Abstr.* **13**, (2017).
25. Lundberg, S. M. *et al.* From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2**, 56–67 (2020).
26. Semenza, J. C. *et al.* Determinants and Drivers of Infectious Disease Threat Events in Europe. *Emerg. Infect. Dis.* **22**, 581–589 (2016).
27. Semenza, J. C. Prototype Early Warning Systems for Vector-Borne Diseases in Europe. *Int. J. Environ. Res. Public Health* **12**, 6333–6351 (2015).
28. Liu-Helmerson, J., Stenlund, H., Wilder-Smith, A. & Rocklöv, J. Vectorial Capacity of *Aedes aegypti*: Effects of Temperature and Implications for Global Dengue Epidemic Potential. *PLOS ONE* **9**, e89783 (2014).
29. Liljepold, K., Rocklöv, J., Liu-Helmerson, J., Sewe, M. & Semenza, J. C. More arboviral disease outbreaks in continental Europe due to the warming climate? *J. Travel Med.* **26**, (2019).
30. Murray, K. O. *et al.* Survival Analysis, Long-Term Outcomes, and Percentage of Recovery up to 8 Years Post-Infection among the Houston West Nile Virus Cohort. *PLoS ONE* **9**, (2014).
31. Harrigan, R. J. *et al.* Economic conditions predict prevalence of West Nile virus. *PLoS One* **5**, e15437 (2010).
32. Massad, E. *et al.* Estimating the probability of dengue virus introduction and secondary autochthonous cases in Europe. *Sci. Rep.* **8**, 4629 (2018).
33. Roiz, D. *et al.* Integrated Aedes management for the control of Aedes-borne diseases. *PLoS Negl. Trop. Dis.* **12**, e0006845 (2018).
34. Luo, Y. *et al.* Ecological forecasting and data assimilation in a data-rich era. *Ecol. Appl.* **21**, 1429–1442 (2011).
35. Morin, C. W. & Comrie, A. C. Modeled response of the West Nile virus vector *Culex quinquefasciatus* to changing climate using the dynamic mosquito simulation model. *Int. J. Biometeorol.* **54**, 517–529 (2010).
36. Åström, C. *et al.* Potential Distribution of Dengue Fever Under Scenarios of Climate Change and Economic Development. *EcoHealth* **9**, 448–454 (2012).
37. Hales, S., de Wet, N., Maindonald, J. & Woodward, A. Potential effect of population and climate changes on global distribution of dengue fever: an empirical model. *The Lancet* **360**, 830–834 (2002).
38. Rocklöv, J. *et al.* Using Big Data to Monitor the Introduction and Spread of Chikungunya, Europe, 2017. *Emerg. Infect. Dis.* **25**, 1041–1049 (2019).

39. Rezza, G. *et al.* Infection with chikungunya virus in Italy: an outbreak in a temperate region. *The Lancet* **370**, 1840–1846 (2007).
40. Morin, C. W. & Comrie, A. C. Regional and seasonal response of a West Nile virus vector to climate change. *Proc. Natl. Acad. Sci.* **110**, 15620–15625 (2013).
41. Keyel, A. C. *et al.* Seasonal temperatures and hydrological conditions improve the prediction of West Nile virus infection rates in *Culex* mosquitoes and human case counts in New York and Connecticut. *PLOS ONE* **14**, e0217854 (2019).
42. Laperriere, V., Brugger, K. & Rubel, F. Simulation of the seasonal cycles of bird, equine and human West Nile virus cases. *Prev. Vet. Med.* **98**, 99–110 (2011).
43. Poveda, G., Jaramillo, A., Gil, M. M., Quiceno, N. & Mantilla, R. I. Seasonally in ENSO-related precipitation, river discharges, soil moisture, and vegetation index in Colombia. *Water Resour. Res.* **37**, 2169–2178 (2001).
44. Hales S, Weinstein P, Soares Y & Woodward A. El Niño and the dynamics of vectorborne disease transmission. *Environ. Health Perspect.* **107**, 99–102 (1999).
45. Hubálek, Z. North Atlantic weather oscillation and human infectious diseases in the Czech Republic, 1951–2003. *Eur. J. Epidemiol.* **20**, 263–270 (2005).
46. Browne, C. & Medlock, J. M. West Nile fever in Europe in 2018: an emerging problem or just an anomaly? *Vet. Rec.* **185**, 365–368 (2019).
47. Final Cumulative Maps and Data | West Nile Virus | CDC. <https://www.cdc.gov/westnile/statsmaps/cumMapsData.html> (2019).
48. Factsheet about West Nile virus infection. *European Centre for Disease Prevention and Control* <https://www.ecdc.europa.eu/en/west-nile-fever/facts/factsheet-about-west-nile-fever>.
49. Historical data by year - West Nile fever seasonal surveillance. *European Centre for Disease Prevention and Control* <https://www.ecdc.europa.eu/en/west-nile-fever/surveillance-and-disease-data/historical>.
50. Hoover, K. C. & Barker, C. M. West Nile virus, climate change, and circumpolar vulnerability. *WIREs Clim. Change* **7**, 283–300 (2016).
51. Bisanzio, D. *et al.* Spatio-temporal patterns of distribution of West Nile virus vectors in eastern Piedmont Region, Italy. *Parasit. Vectors* **4**, 230 (2011).
52. Kilpatrick, A. M. Globalization, Land Use, and the Invasion of West Nile Virus. *Science* **334**, 323–327 (2011).
53. Cunze, S., Kochmann, J., Koch, L. K. & Klimpel, S. *Aedes albopictus* and its environmental limits in Europe. *PLoS One* **11**, e0162116 (2016).
54. Komar, N. West Nile virus: epidemiology and ecology in North America. *Adv. Virus Res.* **61**, 185–234 (2003).
55. Komar, N. *et al.* Experimental infection of North American birds with the New York 1999 strain of West Nile virus. *Emerg. Infect. Dis.* **9**, 311 (2003).
56. Trawinski, P. R. & Mackay, D. S. Identification of environmental covariates of West Nile virus vector

- mosquito population abundance. *Vector-Borne Zoonotic Dis.* **10**, 515–526 (2010).
57. Jelinski, D. E. & Wu, J. The modifiable areal unit problem and implications for landscape ecology. *Landscape Ecol.* **11**, 129–140 (1996).
58. Parham, P. E. *et al.* Climate, environmental and socio-economic change: weighing up the balance in vector-borne disease transmission. *Philos. Trans. R. Soc. B Biol. Sci.* **370**, 20130551 (2015).
59. Battle, K. E., Pigott, D. M. & Hay, S. I. Re: Response to Hay (2013) Global mapping of infectious disease. (2013).
60. Kraemer, M. U. G. *et al.* Progress and Challenges in Infectious Disease Cartography. *Trends Parasitol.* **32**, 19–29 (2016).
61. Bakonyi, T. *et al.* Explosive spread of a neuroinvasive lineage 2 West Nile virus in Central Europe, 2008/2009. *Vet. Microbiol.* **165**, 61–70 (2013).
62. Petersen, L. R. & Fischer, M. Unpredictable and Difficult to Control — The Adolescence of West Nile Virus. *N. Engl. J. Med.* **367**, 1281–1284 (2012).
63. Hay, S. I., George, D. B., Moyes, C. L. & Brownstein, J. S. Big Data Opportunities for Global Infectious Disease Surveillance. *PLOS Med.* **10**, e1001413 (2013).
64. Gorelick, N. *et al.* Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **202**, 18–27 (2017).
65. Karki, S., Brown, W. M., Uelmen, J., Ruiz, M. O. & Smith, R. L. The drivers of West Nile virus human illness: fine scale dynamic effects of weather, mosquito infection, social, and biological conditions. *bioRxiv* 2019.12.16.877951 (2019) doi:10.1101/2019.12.16.877951.
66. J, U. *et al.* Assessing Ultra-Fine-Scale Factors to Improve Human West Nile Virus Disease Models in the Chicago Area. (2020) doi:10.21203/rs.3.rs-19812/v1.
67. Attaway, D. F. *et al.* Mosquito habitat and dengue risk potential in Kenya: alternative methods to traditional risk mapping techniques. (2014) doi:10.4081/gh.2014.10.