Comparing community and single species distribution modelling methods for predicting distributions of mosquito vectors with different traits and invasion status

Background

Vector-borne diseases (VBD) cause substantial morbidity in Europe, and the risk of VBD outbreak and spread to naive populations is expected to increase [1]. Much of this risk is driven by the spread of invasive mosquitoes across Europe, and the increasing threat of naive mosquitoes spreading endemic diseases more efficiently. Study of mosquito ecology has informed the development of better predictive models for mosquito distributions through correlative relationships between environmental variables and mosquito occurrence. These Species Distribution Models (SDM) are now used to predict both present and future distributions of known mosquito vectors [2].

SDM's have been useful tools in informing policy decisions as well as identifying areas that are at risk of VBD transmission, but they operate under several caveats and key assumptions. Current SDMs fail to appropriately account for the complexity of disease transmission in mosquito populations, and the multiplicity of hosts and vectors that vector borne pathogens are spread by. Many diseases are spread by multiple mosquito species and therefore the importance of SDMs in informing public health policy are diminished when not considering this. Communities of mosquitoes which are prevalent vectors that overlap in space and time have the potential to intensify transmission by increasing the abundance of vectors in an area, lengthen the transmission period to longer portions of the year through differences in seasonality between mosquito species, and increase the persistence of transmissions by increasing the number of pathways for mosquitoes to carry arboviruses into coming season through differences in over-wintering and hibernation [1,3].

Determining how mosquitoes interact with each other will lead to better insights on mosquito community compositions that present the most risk for transmitting MBD, as mosquito are limited in distribution by biotic interactions as well as environmental or climatic ones [4]. Modeling the distribution of potential vector species as communities is therefore important for public health officials and intervention planning, and such modeling has prompted initiatives to further monitor mosquito distribution and VBD occurrence across Europe. Though more biologically realistic models are needed to accurately inform stakeholders of the risk that vectors and VBD may pose [5].

Novel statistical techniques that account for community responses to environmental and climatic drivers are now widely available, and have good utility in predicting species associations at various scales [6,7]. Additionally, the ability for these modeling techniques to include broad species traits that may better explain responses to environmental and climatic drivers. Because of this, fine scale responses of mosquitoes that are often missed in SDMs using coarse predictor data could be accounted for by species-specific morphological, physiological and behavioral traits that may be promising proxies of finer scale environmental relationships in trait based community models [8]. Additionally, the development of new joint species distribution models (JSDM) allows incorporation of all the above, and the ability to generate better predictions with sparse data, particularly for rarer species, by leveraging similarities in traits, biotic associations and phylogenetic relationships between groups. An approach applying JSDM to vector data is a novel way to make use of potentially sparse for disease vectors that are often under represented in global datasets, while additionally generating predictions bound to more realistic assumptions than traditional SDM approaches [9].

Objectives

Develop and apply novel statistical joint species distribution models for versus single species distribution models for-mosquito communities across Europe, combining distribution and seasonality data from VectorNet (a European data-sahring network, funded by ECDC and EFSA) with novel hydrological, host, landscape and trait data in order to:

- Determine optimal ways of accounting for sparse, spatially biased mosquito recording data when developing JSDMs and SSDMs.
- Analyse whether species responses to different suites of environmental variables are linked to broad ecological traits such as breeding site, host preference and invasion status.
- Determine whether species with particular traits are better predicted by JSDMs than SSDMs, and determining any trait-environment

Methods

This research question will utilise the current known vector distributions gathered by the VectorNet task force sanctioned by the ECDC and EFSA. From this data 22 individual species have suitable distributional data encompassing a range of invasion status and ecological traits including habitat and host preferences, and assemblages of vectors incriminated for several VBD (West Nile Virus, Rift Valler Fever, etc).

Single species distribution models will be drawn using several known robust methods, of which many are available in accessible ensemble modelling software packages (boosted regression trees, generalised additive models, generalised linear models, etc). These will then be compared to community driven models where species associations and biotic factors can be used to inform predictions of these distributions.

Broad scale traits of habitat preference, host preference and invasion status will be assimilated for all 22 mosquito species modelled. Of these traits, data is already gathered for the majority of traits, and records of habitat preferences and invasion statuses are known for all 22 species. I will also review all recent mosquito distributional models and score the types of predictor values used and how these related to the 22 species in the current dataset, to validate the types of predictors tot ake forward in modelling both JSDM and SDMs in this project. Similarly I will account for the spatial bias in recording effort for the VectorNet data set to generate unbiased results. Additionally, I will explore the use of novel environmental, land use and resource based predictors for use in the distribution models. Particularly I would like to explore the appropriateness of including resource based predictors that are essential for transition of mosquito life stages (breeding habitat, host availability, resting microclimates, etc) using hydrological data and land usage estimates.

Specifically we will analyse the predictive ability of JSDM vs stacked SDMs for mosquito communities across Europe, and determine whether JSDM are accurate enough to provide inform management of mosquito communities in policy. The predictive ability of the JSDMs will be compared using cross validation techniques, and independent validation data sets. Additionally I will employ measures of conditional-cross validation accuracy, to assess to what degree that including community data can better predict other species via expected species associations generated in JSDMs.

This will determine how accuracy of mosquito distributions, particularly for rare or under sampled species can be improved using community distribution models. Results from this project will also inform the rest of my PhD thesis by allowing me to draw conclusions about trait-environment relationships and species distributions; for example we may expect to see less reliance on natural hydrological resources in container breeding mosquitoes, or see an increased reliance on wooded habitats for those species that breed in tree-holes or natural pools. These associations will inform future questions on what traits are important for distributions of mosquitoes and their disease potential. Additionally, this project will also generate relevant and modern SDM ensemble models for many mosquito species across Europe in accordance with the latest data and inform public health officials on current expected distributions of potential vector species across Europe.

References

1. Barzon, L. Ongoing and emerging arbovirus threats in Europe. *J. Clin. Virol.* **107**, 38–47 (2018).
2. Kraemer, M. U. *et al.* The global distribution of the arbovirus vectors Aedes aegypti and Ae. albopictus. *eLife* **4**, e08347 (2015).
3. Medlock, J. M. *et al.* A Review of the Invasive Mosquitoes in Europe: Ecology, Public Health Risks, and Control Options. *Vector-Borne Zoonotic Dis.* **12**, 435–447 (2012).
4. Golding, N., Nunn, M. A. & Purse, B. V. Identifying biotic interactions which drive the spatial distribution of a mosquito community. *Parasit. Vectors* **8**, 367 (2015).
5. Purse, B. V. & Golding, N. Tracking the distribution and impacts of diseases with biological records and distribution modelling. *Biol. J. Linn. Soc.* **115**, 664–677 (2015).
6. Zurell, D., Pollock, L. J. & Thuiller, W. Do joint species distribution models reliably detect interspecific interactions from co-occurrence data in homogenous environments? *Ecography* **41**, 1812–1819 (2018).
7. König, C. *et al.* Scale dependency of joint species distribution models challenges interpretation of biotic interactions. *J. Biogeogr.* **n/a**,.
8. Chandrasegaran, K., Lahondère, C., Escobar, L. E. & Vinauger, C. Linking Mosquito Ecology, Traits, Behavior, and Disease Transmission. *Trends Parasitol.* **36**, 393–403 (2020).
9. Ovaskainen, O. *et al.* How to make more out of community data? A conceptual framework and its implementation as models and software. *Ecol. Lett.* **20**, 561–576 (2017).

Identifying the future risk of vector-borne disease though niche overlaps of hosts, vectors and viruses.

Background

West Nile Virus (WNV) is a disease increasing in frequency and intensity that causes a febrile fever, potential neuroin-flammatory diseases and death in a significant number of humans bitten by mosquitoes carrying the virus [1]. This disease poses an increasing risk to European nations with case rates reaching an all time high in 2018, and further expectation of this virus to become more prevalent and costly in the future [2,3]. WNV is carried by mosquitoes, and mosquito distribution is constrained by a number of biological factors. For example, the availability of blood meal hosts and therefore feeding habitat is linked closely with land use [4], and aquatic habitats for breeding are essential for mosquitoes and therefore regulated heavily by hydrology and topology of environments [5].

However, successful transmission of mosquito borne diseases (MBD) relies on both interactions between hosts, vectors and viruses, and the spatio-temporal overlap of all members in a transmission cycle. Specifically the transmission cycle of WNV relies on the presence of a competent vector mosquito (such as those from the *Culex pipiens* species complex) infected with WNV disease that has successfully incubated in the mosquito so that the virus may move to the salivary glands. Once a mosquito bites a host, the virus will then incubate in the blood until sufficient viremia are produced and reinfection of mosquitoes may occur through blood feeding. However, the competency of vectors can depend on species or subspecies (*Culex modestus* is a more competent vector of WNV than *Cx. pipiens*), and similarly the ability of WNV to replicate to levels in hosts so reinfection of mosquitoes can occur can vary by host species too (chickens are known carriers of WNV but can't reinfect mosquitoes) [6,7].

This means that if in two regions a different species of vector or hosts were present, the risk of WNV spillover in to human populations is different between those areas because of differences in virus-host, virus-vector and vector-host interactions and community composition. Additionally, WNV has an ecological niche that can influence infection rates. Temperatures between 25-30 degrees are known to reduce the extrinsic incubation period of WNV (meaning infection and replication can occur faster). Recent studies have identified the associated environmental characteristics of WNV disease incidence in humans, and several laboratory studies also confirm the impact of climatic variables such as low temperatures on WNV development leading to lower transmission potentials [8,9]. Though it must be noted that the inherent evolution of WNV in tandem with mosquito vectors means many of these factors will co-vary with one another, but differences in spread of WNV in North America despite similar vector species in colder climates of European regions suggest that some differences remain.

Despite the importance of community composition and interactions between host, vector and virus, little research currently focuses on these interacting components of disease transmission as a holistic group when generating risk maps for WNV disease outbreak, instead often examining the components as a pairs or individually. Some mechanistic frameworks attempt to account for these interactions, but often hosts are represented in a simplistic manner and spatio-temporal overlap of transmission components and human populations are poorly integrated [7,10]. As such, little work has examined how the spatiotemporal distribution of WNV, vectors, hosts and human populations may overlap, and what this means for potential disease risk in those areas. Combining distributional data information in a spatially explicit model it would be possible to generate detailed maps of WNV transmission potential across Europe. This data would better inform risk based policy initiatives for public health stake holders compared to current existing models, and furthermore would highlight areas where disease transmission may pose a risk in future due to spread of invasive mosquitoes, or range shifts in host species.

Objectives

Determine whether hotspots of spill over of WNV into human populations can be predicted from the distributions and overlap of hosts and vectors at "at risk" human populations quantified from SDMs and environmental niche overlap methods. Determine whether these quantified spatial interactions between hosts, vectors and people can inform mechanistic model frameworks for WNV transmission that currently have very simplified representations of host and vector distributions and seasonality across landscapes.

- Develop and validate SDMs and niche overlap models for all relevant hosts and vectors of WNV.
- Overlay with validated SDMs for hosts, vectors, and WNV with distributions of human population and activities to quantify population at risk.

- Analyse whether the distributions and overlap of hosts and vectors at "at risk" human populations are associated with and can be used to predict patterns in WNV outbreaks.
- Analyse the risk that species range shifts, climatic variables and invasive mosquito species may drive WNV disease rates in the future, and how this may impact policy and intervention.

Methods

Generate SDM (or stacked-SDMs/JSDMs depending on previous chapter output) for distributions of suspected and known WNV vector species (*Culex pipiens, Culex modestus, Culiseta annulata , Aedes albopictus*), known hosts of WNV (passerines, corvids) vector species that have a proven to hold sufficient viremia levels in the blood to reinfect mosquito species and amplify transmission cycles [11]. The distribution of WNV vector species will be generated from the latest VectorNet data (ECDC and EFSA), using typical distributional predictors for mosquito species climate, land use, and environmental data, informed by the previous chapter. Similar distributions of host species will be generated from freely available distributional data for individual species or groups of relevant hosts (Birdlife). Generating the environmental suitability model for WNV will follow similar methods as previous authors, identifying climatic envelopes at regional levels to generate characteristics associated with WNV infection [8,9].

Overlap metrics can be generated through a combination of overlapping projected distributions across geographical space and ordination techniques [12]. Such methods account for differences in spatial resolution and sampling area to achieve more realistic estimates of true niche overlap between taxa, and these methods are accessible through distributed packages in R [13]. Spatial mixed models will be used to determine whether these niche overlap distributions accurately predict current known distributions WNV disease, and their accuracy determined as such with traditional metrics for spatial model accuracy (AUC, etc). Further investigation and validation of overlap models can be determined by comparing mechanistic models expected hotspots for WNV disease transmission with our ecological niche overlap models, and discuss the ability for these niche overlap methods to be incorporated into hybrid mechanistic/correlative modelling methods [14].

References

1. Murray, K. *et al.* Risk factors for encephalitis and death from West Nile virus infection. *Epidemiol. Infect.* **134**, 1325–1332 (2006).
2. Bakonyi, T. & Haussig, J. M. West Nile virus keeps on moving up in Europe. *Eurosurveillance* **25**, (2020).
3. Barzon, L. Ongoing and emerging arbovirus threats in Europe. *J. Clin. Virol.* **107**, 38–47 (2018).
4. Talbot, B., Caron-Lévesque, M., Ardis, M., Kryuchkov, R. & Kulkarni, M. A. Linking Bird and Mosquito Data to Assess Spatiotemporal West Nile Virus Risk in Humans. *EcoHealth* **16**, 70–81 (2019).
5. Bond, J. G., Arredondo-Jiménez, J. I., Rodríguez, M. H., Quiroz-Martínez, H. & Williams, T. Oviposition habitat selection for a predator refuge and food source in a mosquito. *Ecol. Entomol.* **30**, 255–263 (2005).
6. Kilpatrick, A. M., LaDeau, S. L. & Marra, P. P. Ecology of West Nile Virus Transmission and its Impact on Birds in the Western Hemisphere. *The Auk* **124**, 1121–1136 (2007).
7. Vogels, C. B. F., Hartemink, N. & Koenraadt, C. J. M. Modelling West Nile virus transmission risk in Europe: effect of temperature and mosquito biotypes on the basic reproduction number. *Sci. Rep.* **7**, 5022 (2017).
8. Wimberly, M. C., Hildreth, M. B., Boyte, S. P., Lindquist, E. & Kightlinger, L. Ecological Niche of the 2003 West Nile Virus Epidemic in the Northern Great Plains of the United States. *PLOS ONE* **3**, e3744 (2008).
9. García-Carrasco, J.-M., Muñoz, A.-R., Olivero, J., Segura, M. & Real, R. Predicting the spatio-temporal spread of West Nile virus in Europe. *PLoS Negl. Trop. Dis.* **15**, e0009022 (2021).
10. Bessell, P. R. *et al.* Assessing the potential for Bluetongue virus 8 to spread and vaccination strategies in Scotland. *Sci. Rep.* **6**, 38940 (2016).
11. Jourdain, E. *et al.* Surveillance for West Nile Virus in Wild Birds from Northern Europe. *Vector-Borne Zoonotic Dis.* **11**, 77–79 (2010).
12. Broennimann, O. *et al.* Measuring ecological niche overlap from occurrence and spatial environmental data. *Glob. Ecol. Biogeogr.* **21**, 481–497 (2012).
13. Cola, V. D. *et al.* ecospat: an R package to support spatial analyses and modeling of species niches and distributions. *Ecography* **40**, 774–787 (2017).
14. Cheng, Y. *et al.* Evaluating the risk for Usutu virus circulation in Europe: comparison of environmental niche models and epidemiological models. *Int. J. Health Geogr.* **17**, 35 (2018).

Exploring the current availability of mosquito traits and the feasibility of including these in novel analyses based on ecological modelling frameworks.

Introduction

Mosquitoes are primary vectors for many debilitating arboviruses, with far reaching societal, economic and health impacts on much of the worlds population [1]. There is significant evidence that a vectors behaviour and physiology can determine VBD transmission and modulate the risk of disease in a given area. These behaviours, physiological and morphological components of an organism are known as functional traits, and study of functional traits has the potential to uncover ecological reasoning behind responses to climate and environmental change [2].

Trait based approaches in mosquitoes would be particularly useful, as traits can be used for proxies of finer scale environmental resource usage, meaning coarse environmental data such as land use or feeding resource abundance can give more accurate results from correlative species distribution modelling techniques [3]. Similarly, including traits in correlative models gives them greater transferability on average between eco-regions and species, making them more generalised and useful for species prediction of range shifts or distributions [2,3]. Understanding these trait-environment relationships, and their impact on life history traits in mosquitoes is important because they can describe the influence of environment on mosquito population dynamics. For instance, *Culex pipiens* development time increases nearly three-fold under increasing temperatures, increases in temperature are also associated with smaller sized adult mosquitoes and reduced fecundity which directly impacts vector competence for many disease [4,5]. The importance of integrating mosquito trait-environment relationships into current modelling frameworks to understand ecological responses cannot be overstated, yet has only recently been recognised [6,7].

Correlative methods have been used to investigate how tick morphology can relate to potential vectorability of diseases, suggesting that morphological and life history traits can be used to identify current vector species with high accuracy, and potential unrecognised vectors [8]. Similar studies using mosquitoes in South America have highlighted the utility of traits to inform potentially unrecognised vectors of arboviruses such as Zika, and can inform next steps for validation and incrimination of vectors identified through these methods which is particularly useful for new and emerging disease outbreaks [9]. However, studies currently trying to integrate trait data for mosquitoes in their analyses are severely limited by the availability of such traits, especially compared to the trait data available for other broad groups of species, and even other insect vector species [10–12]. This highlights the need for mosquito traits to better understand mosquito ecology, particularly as this group represents the most medically important insect taxa [7].

While many authors have highlighted the importance of including traits in future modelling techniques, none have examined the current state of trait data availability for mosquito species. The current literature understandably focuses on medically important mosquito species, which means a large breadth of trait data is likely to be under-represented in species with no current link to disease. Additionally, for well studied species trait data is often stored in primary or grey literature and in a format that is difficult for collaborators or future researchers to work with. Alleviating these problems will remain a huge challenge for the future, but one that is achievable considering the progress of similar projects for other groups of organisms. A first step in moving towards a trait based frameworks for mosquito ecology involves a systematic review of the current state of trait data for mosquitoes, and an exploratory analysis of how this trait data can be incorporated into future projects that considering the caveats of trait-environmental relationships and their impacts on VBD transmission and mosquito ecology.

Objectives

Much of the suggested work by Cator *et al.* (2020) and Chandrasegaran *et al.* (2020) highlights the importance of considering mosquito ecology at a larger scale, both from a bottom up and top down approach, to generate more realistic models to determine distributions, population dynamics and ecological relationships between both disease and mosquitoes. However, these approaches are data intense and rely on the production of a trait database spanning many physiological, behavioural and life history traits of mosquitoes, including those that are currently under sampled. Before work starts on gathering this into easier, more accessible formats for researchers, is first necessary to review the available data that could be used in analysis, and to identify areas where key data may be missing for other analyses in order to direct future research efforts.

- Assess the state, availability and accessibility of trait data in the primary and grey literature.
- Identify gaps in trait data for species that may be of epidemiological and ecological importance, and suggest methods to gather and assimilate this into usable reproducible sources, to direct future research efforts in generating an informative data source of sufficient scope for analytical needs.
- Identify the current state of trait data concerning trait variation between and within populations or lineages of mosquitoes, and identify the need for future research encompassing trait data across lineages and geographic regions, so that adequate nuance of trait-environment relationships are captured in future data sources.
- Determine how previous studies could incorporate currently available trait data into modelling techniques and frameworks.

References

1. WHO. *Global vector control response 2017-2030.* (2017).
2. Vesk, P. A., Morris, W. K., Neal, W. C., Mokany, K. & Pollock, L. J. Transferability of trait-based species distribution models. *Ecography* **44**, 134–147 (2021).
3. Syphard, A. D. & Franklin, J. Species traits affect the performance of species distribution models for plants in southern California. *J. Veg. Sci.* **21**, 177–189 (2010).
4. Ciota, A., Matacchiero, A., Kilpatrick, M. & Kramer, L. The Effect of Temperature on Life History Traits of Culex Mosquitoes. *J. Med. Entomol.* **51**, 55–62 (2014).
5. Reinhold, J. M., Lazzari, C. R. & Lahondère, C. Effects of the Environmental Temperature on Aedes aegypti and Aedes albopictus Mosquitoes: A Review. *Insects* **9**, 158 (2018).
6. Cator, L. J. *et al.* The Role of Vector Trait Variation in Vector-Borne Disease Dynamics. *Front. Ecol. Evol.* **8**, (2020).
7. Chandrasegaran, K., Lahondère, C., Escobar, L. E. & Vinauger, C. Linking Mosquito Ecology, Traits, Behavior, and Disease Transmission. *Trends Parasitol.* **36**, 393–403 (2020).
8. Yang, L. H. & Han, B. A. Data-driven predictions and novel hypotheses about zoonotic tick vectors from the genus Ixodes. *BMC Ecol.* **18**, 7 (2018).
9. Evans, M. V., Dallas, T. A., Han, B. A., Murdock, C. C. & Drake, J. M. Data-driven identification of potential Zika virus vectors. *eLife* **6**, e22053 (2017).
10. Homburg, K., Homburg, N., Schäfer, F., Schuldt, A. & Assmann, T. Carabids.org – a dynamic online database of ground beetle species traits (Coleoptera, Carabidae). *Insect Conserv. Divers.* **7**, 195–205 (2014).
11. Parr, C. L. *et al.* GlobalAnts: a new database on the geography of ant traits (Hymenoptera: Formicidae). *Insect Conserv. Divers.* **10**, 5–20 (2017).
12. Fraser, L. H. TRY—A plant trait database of databases. *Glob. Change Biol.* **26**, 189–190 (2020).