

**Assessing spatio-temporal risks of vector-borne diseases:  
an interdisciplinary view integrating  
ecological and epidemiological models**

Dissertation

zur Erlangung des akademischen Grades einer

Doktorin der Naturwissenschaften (Dr. rer. nat.)

an der Fakultät für Biologie, Chemie und Geowissenschaften

der

Universität Bayreuth

vorgelegt von

Yanchao Cheng

geb. in Heilongjiang, China

Bayreuth, 2020



This doctoral thesis was prepared at the department of Biogeography at the University of Bayreuth from November 2015 until December 2020 and was supervised by Prof. Dr. Carl Beierkuhnlein.

This is a full reprint of the dissertation submitted to obtain the academic degree of Doctor of Natural Sciences (Dr. rer. nat.) and approved by the Faculty of Biology, Chemistry and Geosciences of the University of Bayreuth.

Date of submission: December 2<sup>nd</sup> 2020

Date of defence: March 22<sup>nd</sup> 2021

Acting dean: Prof. Dr. Matthias Breuning

Doctoral committee:

Prof. Dr. Carl Beierkuhnlein (reviewer)

Prof. Dr. Kurt Chudej (reviewer)

Prof. Dr. Cyrus Samimi (chairman)

Prof. Dr. Heike Feldhaar



# Contents

<b>Summary</b>	<b>1</b>
<b>Zusammenfassung</b>	<b>3</b>
<b>Introduction</b>	<b>5</b>
<b>1 Motivation</b>	<b>5</b>
<b>2 Structure of this thesis</b>	<b>6</b>
<b>3 General introduction</b>	<b>7</b>
3.1 Infectious diseases	7
3.2 Vector-borne diseases	8
3.3 Vector-borne diseases in face of global change	11
3.4 Modelling the potential outbreak risks of vector-borne diseases	12
<b>4 Ecological niche models</b>	<b>13</b>
4.1 General workflow	13
4.2 Occurrence/presence locations	13
4.3 Explanatory variables	14
4.4 How do ecological niche models use spatial information?	14
4.5 More than a black box: How do ecological niche models work?	19
4.6 Products from ecological niche models – spatial risk maps	22
<b>5 Epidemiological models</b>	<b>23</b>
5.1 General workflow	23
5.2 Basic reproduction number: $R_0$	23
5.3 Survival function	24

5.4 Next generation matrix	25
5.5 Incidence-based methods	28
5.6 Products from epidemiological models – spatio-temporal risk maps	28
6 Comparison of ecological niche models and epidemiological models	30
7 Synopsis of the following manuscripts	32
8 Challenges for future studies	34
8.1 Positional errors in ecological niche models	34
8.2 The complexity of epidemiological models	35
8.3 Assessment of epidemiological models' performance	36
8.4 The quality of vector and host parameters	37
9 List of articles and declaration of own contribution	38
10 References of introduction	40
<b>Manuscript 1</b>	<b>47</b>
<b>Manuscript 2</b>	<b>76</b>
<b>Manuscript 3</b>	<b>106</b>
<b>Appendix 1</b>	<b>139</b>
<b>List of abbreviations</b>	<b>148</b>
<b>Other academic activities</b>	<b>149</b>
<b>Acknowledgements</b>	<b>151</b>
<b>(Eidesstattliche) Versicherungen und Erklärungen</b>	<b>152</b>

## Summary

Vector-borne diseases are infectious diseases that are transmitted among vertebrate hosts by (typically arthropod) vectors. Among the whole world's population, 80% is at risk of one or more vector-borne diseases, leading to an annual death toll of 700 000. These striking numbers are calling for urgent actions to prevent vector-borne diseases from emerging further. However, to apply preventions, we need to know *where* a risk exists; and if possible, *when* the prevention should take place.

The key to those two primary questions are risk maps, which are typically generated with ecological niche models or epidemiological models. Ecological niche models require occurrence records of the transmissions and the respective environmental variables (mostly long-term-averaged) to build a correlative model. This correlative model can be projected to a different spatial extent, or into future climate scenarios, etc., showing the spatial outbreak risk. Epidemiological models, on the other hand, look into the transmission process and thus require a good understanding of the transmission cycle of the investigated vector-borne disease. Epidemiological models can work with time-series data, and produce spatio-temporal risk maps based on the basic reproduction number  $R_0$ . In practice, both ecological niche models and epidemiological models have their respective strengths and drawbacks. In this thesis, I contribute to the improvement of both approaches by analyzing some of their drawbacks and making suggestions for new standards.

For ecological niche models, the correlative models are highly dependent on the quality of occurrence records. In this thesis, I investigate how positional error, i.e. substituting the geographical centroid of the respective administrative spatial unit for unknown occurrence records, affects model performance in the context of varying grain size of environmental data. I quantify the decrease of model performance caused by the use of geographical centroids and varying grain size, respectively. As a consequence, I suggest that special cautions should be given when geographical centroids are applied as substitutes; when possible, central tendency values should be preferred.

For epidemiological models, I review the common ways to generate risk maps and illustrate them with an example. I demonstrate that using different temporal aggregation methods affects the comparability and the quantity information of the resulting maps; and that via different visualization methods, two fundamentally different maps can appear very similar, and *vice versa*. Consequently, I highlight the importance of using appropriate temporal aggregations and visualizations and give suggestions for best practice. I recommend to show both intensity and duration of the risk, using small time-steps to show spatio-temporal dynamics when possible.

Pushing towards new standards for best practice in vector-borne disease risk mapping, I directly compare ecological niche models and epidemiological models, using Usutu virus as an example. The results from the parallel-model approach shows that relying on a single model for assessing vector-borne disease risk may lead to incomplete conclusions. For future research, it is crucial to realize this and aim to apply different modelling approaches for risk-assessment of under-studied emerging pathogens like Usutu virus.



# Zusammenfassung

Durch Vektoren übertragene Krankheiten sind Infektionskrankheiten, die von Wirbeltierwirten durch Vektoren (typischerweise Arthropoden) übertragen werden. Bei 80% der gesamten Weltbevölkerung besteht das Risiko einer oder mehrerer durch Vektoren übertragener Krankheiten, was zu einer jährlichen Zahl von 700 000 Todesopfern führt. Diese bemerkenswerten Zahlen erfordern dringende Maßnahmen, um das weitere Auftreten von Vektor-übertragenen Krankheiten zu verhindern. Um Präventionsmaßnahmen einleiten zu können, müssen wir jedoch wissen, wo ein Risiko besteht, und wenn möglich, wann die Prävention erforderlich ist.

Der Schlüssel zu diesen beiden Hauptfragen sind Risikokarten, die typischerweise mit ökologischen Nischenmodellen oder epidemiologischen Modellen erstellt werden. Ökologische Nischenmodelle erfordern Vorkommensdaten der Übertragungen sowie relevante Umweltvariablen (meist langfristig gemittelt), um ein korrelatives Modell zu erstellen. Dieses korrelative Modell kann in einen anderen räumlichen oder zeitlichen (mittels Klimaszenarien) Kontext projiziert werden, um das räumliche Ausbruchsrisko aufzuzeigen. Epidemiologische Modelle untersuchen dagegen den Übertragungsprozess und erfordern daher ein gutes Verständnis des Übertragungszyklus der untersuchten Vektor-übertragenen Krankheit. Epidemiologische Modelle können mit Zeitreihendaten arbeiten und räumlich-zeitliche Risikokarten basierend auf der Basisreproduktionszahl  $R_0$  erstellen. In der Praxis haben sowohl ökologische Nischenmodelle als auch epidemiologische Modelle ihre jeweiligen Vor- und Nachteile. In dieser Arbeit trage ich zur Verbesserung beider Ansätze bei, indem ich einige ihrer Schwächen analysiere und Vorschläge für neue Standards mache.

Bei ökologischen Nischenmodellen hängen die korrelativen Modelle stark von der Qualität der Vorkommensdaten ab. In dieser Arbeit untersuche ich, wie räumliche Fehler durch das Ersetzen unbekannter Vorkommensdaten durch den geografischen Schwerpunkt der jeweiligen administrativen Raumeinheit die Modellleistung im Kontext mit variierender Korngröße von Umgebungsdaten beeinflussen. Ich quantifiziere die Abnahme der Modellleistung, die durch die Verwendung von geografischen Schwerpunkten bzw. durch unterschiedliche Korngrößen verursacht wird. Infolgedessen schlage ich vor, besondere Vorsichtsmaßnahmen zu treffen, wenn geografische Schwerpunkte als Ersatz verwendet werden. Wenn möglich, sollten Lageparameter (räumlicher Mittelwert, Median) bevorzugt werden.

Für epidemiologische Modelle überprüfe ich die gängigen Methoden zur Erstellung von Risikokarten und illustriere sie anhand eines konkreten Beispiels. Ich zeige, dass die Verwendung verschiedener zeitlicher Aggregationsmethoden die Vergleichbarkeit und den quantitativen Informationsgehalt der resultierenden Karten beeinflusst. Ich zeige weiterhin, dass verschiedene Visualisierungsmethoden zwei grundlegend unterschiedliche Karten sehr ähnlich erscheinen lassen können und umgekehrt. Infolgedessen unterstreiche ich die Bedeutung der Verwendung geeigneter zeitlicher Aggregations- und Visualisierungsmethoden und identifiziere bewährte Verfahren. Ich empfehle, sowohl die Intensität als

auch die Dauer des Risikos zu zeigen und wenn möglich kleine Zeitschritte zu verwenden, um die räumlich-zeitliche Dynamik zu zeigen.

Auf dem Weg zu neuen Standards für bewährte Verfahren bei der Risikokartierung Vektor-übertragener Krankheiten vergleiche ich direkt ökologische Nischenmodelle und epidemiologische Modelle am Beispiel des Usutu-Virus. Die Ergebnisse des parallelen Modellierungsansatzes zeigen, dass die Verwendung eines einzigen Modells zur Bewertung des Risikos von Vektor-übertragener Krankheiten zu unvollständigen Schlussfolgerungen führen kann. Für die zukünftige Forschung ist es entscheidend, dies zu realisieren und verschiedene Modellierungsansätze für die Risikobewertung von kaum untersuchten neu auftretenden Krankheitserregern wie dem Usutu-Virus anzuwenden.

# Introduction

## 1 Motivation

Vector-borne diseases are infectious diseases that are typically transmitted by arthropods via blood meals on different hosts (Jamison et al., 2015). Humans can be both, source and sink, but also other organisms can serve as sources. In face of global change, especially climate change and intercontinental transportation, vector-borne diseases have been emerging increasingly during the last few decades – for example the recent Zika epidemic in Latin America (Wikan & Smith, 2016; Zhang et al., 2017) or the emergence of West Nile virus in Europe (Rizzoli et al., 2015; Ziegler et al., 2015). Preferably, preventive measures should be taken *before* an epidemic – or even pandemic – is forming, as seen during the currently ongoing COVID-19 pandemic. To enable this, maps showing potential outbreak risk of the investigated vector-borne disease are a crucial tool.

Risk maps for vector-borne diseases are usually, if not always, based on modelling approaches. There are mainly two groups of modelling approaches commonly applied for generating risk maps, namely ecological/environmental niche models and process-based epidemiological models (Tjaden et al., 2018). Evidently, these two modelling approaches are established in different scientific communities (ecology vs. epidemiology) (Escobar, 2020; Escobar & Craft, 2016). Nevertheless, these approaches are not absolutely different in terms of the mechanisms that are either mimicked or assessed in their spatial results. In the past, theoretical ecologists like Alfred Lotka and Robert May greatly influenced the development of modern epidemiological models, proving the value of interdisciplinary thinking. Despite this, these two kinds of model are almost always applied separately and their outcome is rarely compared. This lack of interdisciplinarity is calling for attention. Generally, the generation and visualization of risk maps is dependent on algorithms, data, spatial scale and other factors. Different approaches will necessarily yield deviating products. One product may partially contradict another one, as it is unlikely that they will match. Intuitively, the difference between those two modelling approaches needs to be investigated and, if possible, quantified.

In this thesis, I approach common methodological challenges in both ecological niche models and epidemiological models that have remained largely ignored until now. I aim to deliver the basis for the development of useful future standards. Finally, I actively contribute to improving interdisciplinarity by performing one of the very first direct comparisons between the kinds of models typically used for the creation of risk maps.

## 2 Structure of this thesis

In chapter 3 of this thesis, I first introduce the main challenges regarding spatio-temporal risk assessment of epidemics or even pandemics of infectious diseases. To explain the challenges in a gradual manner, I briefly talk about a series of commonly used and important terms in epidemiology. This section gives an overall impression of how the different groups of infectious diseases are defined and classified, and sheds light on the basic elements included in and general patterns of vector-borne disease transmission cycle. After the introduction of epidemiological terms, I refer to factors that affect vector or host population, distribution, or transmission between vectors and hosts, etc., which further affect the potential outbreak risk of vector-borne diseases.

In terms of increasing outbreak risk of investigated vector-borne diseases, some of the essential questions that need to be answered are 1) *where* a risk for an outbreak exists and 2) *when* the risk is high. This leads to the main topic of this thesis – how do we answer these *where* and *when* questions?

In the following chapters, I introduce two modelling approaches that are often used to answer these questions: ecological niche modelling (chapter 4) and epidemiological modelling (chapter 5). I explain the theoretical basis as well as the applications of these two fundamentally different modelling approaches, using either a conceptual or a simplified model as an example. In these chapters, I aim to explain the two modelling approaches in a down-to-earth manner, with the intention of letting this thesis serve as an easy guide for new modelers as well.

Thereafter, I compare ecological niche models and epidemiological models, and illustrate the necessity and importance of practicing integrated modelling approach (chapter 6). This comparison of those two modelling approaches is the first step towards an interdisciplinary view. In chapter 7, I link the current understanding of those two modelling approaches and future challenges with my three manuscripts.

Challenges for future studies aiming to improve model performance of both modelling disciplines and achieving an interdisciplinary view, are included in Chapter 8.

### 3 General introduction

Throughout human history, several epidemics and even pandemics of infectious diseases took place related with large numbers of deaths (Johnson & Mueller, 2002; Taubenberger & Morens, 2006; WHO, last accessed 26, Nov, 2020c). Even in the current era, infectious diseases are causing millions of deaths on a yearly basis (WHO, last accessed 26, Nov, 2020e). In 2020, the COVID-19 pandemic resulted in more than one million fatalities globally until the end of September (ECDC, 2020). The harsh fact is that at the time of writing this thesis, this number is still growing, since the COVID-19 pandemic has not been globally contained yet (last accessed on 30<sup>th</sup> of September) (CSSE, 2020; ECDC, 2020).

To prevent disease outbreaks from emerging, a series of prevention measures can be undertaken, such as improving general precautions (e.g. suggest people not to visit epidemic zones), vaccinations, vector control (for vector-borne diseases such as dengue, malaria, etc.), etc. (CDC, last accessed 26, Nov, 2020d; WHO, last accessed 26, Nov, 2020f). The next step is to control and retain an epidemic, which is way more challenging and in most cases also less efficient, as seen during the COVID-19 pandemics, e.g. (Onder et al., 2020; Remuzzi & Remuzzi, 2020; Wu et al., 2020). It is not difficult to see that, whenever possible, preventions should be first in place before mitigation. However, all prevention measures are dependent on the availability of information: When and where caution needs to be taken? In order to answer these questions, modelling approaches are often applied. However, prior to the modelling approaches, it is essential to know the basic background knowledge necessary to understand the underlying mechanisms of disease transmission and the models used to answer these questions. Consequently, Chapter 3 introduces this information as following.

#### 3.1 Infectious diseases

The term “infectious disease” is defined as a disease due to a pathogenic agent (Porta, 2014), which can be transmitted directly or indirectly from one individual to another (Krickeberg et al., 2012; WHO, last accessed 26, Nov, 2020b). The group of pathogenic agents includes viruses, bacteria, fungi, parasites, and prions (Porta, 2014). For instance, tuberculosis is an infectious disease caused by *Mycobacterium tuberculosis* bacteria (WHO, last accessed 26, Nov, 2020d); flu (influenza) is caused by viruses (CDC, last accessed 26, Nov, 2020e); Malaria is caused by parasites from the genus of *Plasmodium* (CDC, last accessed 26, Nov, 2020b); the ringworm disease is caused by a fungus (CDC, last accessed 26, Nov, 2020c); the mad cow (bovine spongiform encephalopathy) disease is caused by a prion (CDC, last accessed 26, Nov, 2020a).

The direct transmission of pathogens is possible for some diseases where contacts between an infected individual to an uninfected one are resulting in transmission. This includes physical touching, kissing, biting, sexual intercourse, etc. (Ahrens & Pigeot, 2007). The direct projection of droplet spray onto the mucous membranes (of the eyes, nose, or mouth) also belongs to the category of direct transmission (Ahrens & Pigeot, 2007; Porta, 2014).

Indirect transmission does not include any forms of direct physical contact or close interaction; it is the transmission via a “third party”. The third party can be stationary: e.g. contaminated inanimate material or objects (e.g. toys, surgical instruments); water, food, milk; biological products such as blood, serum, tissues, etc.; while it can also be non-stationary: e.g. microbial aerosols (airborne), or vectors (vector-borne) (Ahrens & Pigeot, 2007).

A noticeable confusion in terms of infectious disease is about cancers. For instance, in a rare case, four patients got cancer after they received organ transplants from the same donor, who had early stage cancer (Matser et al., 2018). In this case, it is not so difficult to understand that this cancer does not constitute an infectious disease, since there was no pathogen involved. However, there are also some kinds of cancer that are clearly associated with certain pathogens. For instance, cervical, anal, and penile cancers are associated with human papillomavirus (HPV) infection, liver cancer is associated with Hepatitis B virus or Hepatitis C virus infections (Mayo Clinic, last accessed 26, Nov, 2020). These viruses can be transmitted from one individual to another. There are even vaccines against HPV infection, with a clear aim of reducing the risk of cervical cancer (WHO, last accessed 26, Nov, 2020a). Nevertheless, cancers are not classified as infectious diseases. Another controversial example is Alzheimer’s disease (Itzhaki et al., 2020). There were four Alzheimer cases reported to be associated with the injection of a growth hormone that was contaminated with a prion through preparations (in the past the growth hormone was derived from human cadavers) (Abbott, 2016). However, Alzheimer is typically not considered as an infectious disease, either.

In the field of epidemiology, there are some terms that are closely related to *infectious disease*, such as *communicable disease* or *contagious disease*. It should be noticed that an infectious disease is communicable, but not necessarily contagious (Ahrens & Pigeot, 2007), e.g. vector-borne diseases belong to the class of infectious disease, but they are not contagious. The definition of infectious disease from (WHO, last accessed 26, Nov, 2020b) is in accordance to communicable disease (Ahrens & Pigeot, 2007): a disease whose causal agent can be transmitted from successive hosts to healthy subjects, from one individual to another (WHO, last accessed 26, Nov, 2020b). In this thesis, I apply the definition of infectious disease from the WHO:

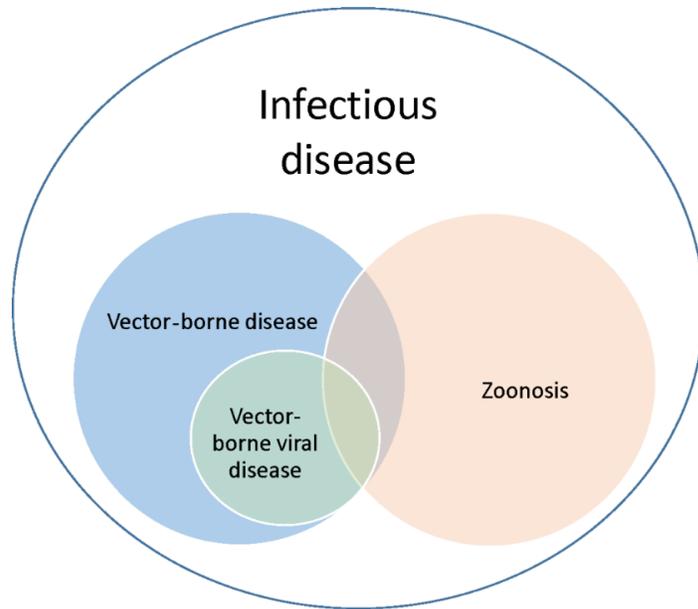
“Infectious diseases can be transmitted from one to another, directly or indirectly.”

### **3.2 Vector-borne diseases**

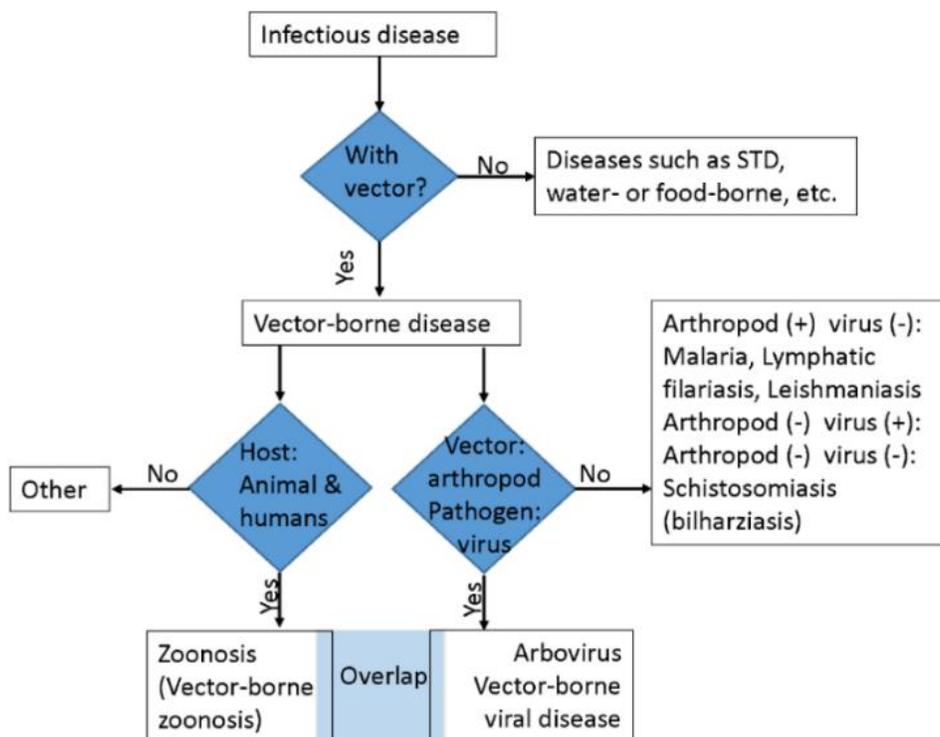
Vector-Borne Diseases (VBDs) account for more than 17% of all known infectious diseases (WHO, last accessed 26, Nov, 2020e). Every year, VBDs cause more than 700 000 human deaths globally, especially in tropical and sub-tropical regions (WHO, last accessed 26, Nov, 2020e). As its name suggests, VBDs are diseases of which the causative agents are transmitted between vertebrate hosts by vectors (Braks et al., 2011).

The vectors are living organisms that typically feed on blood. These vectors take blood meals from mammals including humans (Porta, 2014). The vectors include not only insects such as mosquitoes, sandflies, triatomine bugs, tsetse flies and black flies, but also other types of small animals such as ticks and water snails (WHO, last accessed 26, Nov, 2020e). While taking blood meals from infected individuals, the vectors can ingest the respective pathogen. The pathogen will be replicated within the vectors for a while (this time period is called extrinsic incubation period, EIP). Once this infected vector bites another individual, the pathogen will be further transmitted.

The individuals being fed on by the vectors are referred to as the hosts. The hosts can be humans, horses, bats, birds, etc. Some VBDs only affect a certain host species, while other VBDs affect a huge range of host species. When an infectious disease affects both humans and other vertebrate hosts, it is also called a zoonosis; as for a VBD, it enters the category of vector-borne and zoonotic disease (Porta, 2014) (Figure 3.1). There are some VBDs that only infected other vertebrate hosts in the past, while unexpectedly started infecting humans as well (e.g. Usutu virus and West Nile virus (Ashraf et al., 2015; Mackenzie et al., 2004; Weissenböck et al., 2002; Williams et al., 1964)). There are also VBDs that only caused asymptomatic infections among humans in the past, but evolved into life threatening diseases, e.g. Zika (Ferguson et al., 2016). This means that the categorization of a given VBD can change over time. As the pathogen evolve, the host range can increase, the symptoms may differ, and the direction of pathogen evolution is unpredictable.



a) A conceptual vein figure on some commonly used terms on infectious disease.



b) A flowchart showing the classification of infectious disease.

Figure 3.1 Commonly used terms in epidemiology and the classifications

### 3.3 Vector-borne diseases in face of global change

Global change affects emergence and re-emergence of infectious diseases including VBDs', directly or indirectly (Mackey et al., 2014; Rizzoli et al., 2019). Considering the recent range expansions of some avian pathogens, climate change, global transportation such as human movements and commerce may have played an important role (Fuller et al., 2012; Smith & Guégan, 2010). As with rapid human societal development and globalization, increasing changes in demographics, populations, and the environment also add to the challenges for VBDs control (Mackey et al., 2014). For instance, climate change affects environmental suitability for vectors such as mosquitoes (Fischer et al., 2013), it also affects migrating routes of migratory birds (Fuller et al., 2012; Knudsen et al., 2011).

With the fast growth of global transportation, insect vectors and pathogens are spread to places they could not reach in the past (Cabral et al., 2019; Pettersson et al., 2016). With the aid of global air travel and seaborne trade, they can overcome geographic barriers and travel great distances in short periods of time (Tatem et al., 2006). For instance, Asian tiger mosquito (*Aedes albopictus*) is emerging in Europe as an invasive species after being introduced to the continent through the trade of used tires (Medlock et al., 2012; Thomas et al., 2014). The pathogens carried by the respective vectors can easily cross natural barriers such as oceans and high mountains.

Invasive species also have the potential of facilitating the emerging of VBDs (Crowl et al., 2008). Invasive species are "species of animals, plants, fungi, or microorganisms translocated into environments outside their natural range", where they compete and interact with native species (Chinchio et al., 2020). When pathogens are introduced into new geographic locations or host species, this is also called "pathogen pollution" (Daszak, 2000). Due to global change, especially increasing global transportation, the frequency and magnitude of the invasions by invasive species and infectious diseases has increased (Strickland et al., 2015). Invasive species affect ecological dynamics at multiple spatial scales via interactions with native species (Crowl et al., 2008; Gonzalez-Moreno et al., 2015; Mollot et al., 2017). This is not limited to the well-known examples of highly invasive mosquito vectors (e.g., *Ae. albopictus*, *Ae. japonicus*). For VBDs, invasive species also affect host-parasite interactions (Westby et al., 2019).

Migratory birds pose another threat on introduction of novel pathogens. Seasonal bird migration affects interactions between hosts and pathogens. Migratory birds work as airplanes for pathogens, especially when they stop en route along rivers or at ponds and mix with local birds e.g. (Wu & Perrings, 2017). Migratory birds can transport vectors and pathogens across vast distances (Cohen et al., 2015), and possibly transmit pathogens to local bird species or mosquitoes. In some cases, birds play a critical role in the dispersal of certain vector-borne viruses (Brown et al., 2007), e.g. the transmission of West Nile virus in the US (Moon et al., 2019). Ticks can establish in any migratory bird stopover and breeding sites (Heffernan et al., 2014). Some virus transmission is directly associated with the magnitude and direction of daily bird movement in a local area (Brown et al., 2007).

The potential importance of climate change during the non-breeding season lies in constraining the response of migratory species to temperature changes at both the trailing and leading edges of their

breeding distributions (Rushing et al., 2020). Climate change may affect VBDs from several aspects. For instance, heat waves and sea level change can affect VBDs directly (Watson et al., 2005). Climate change may affect the incidence of VBDs through its effect on the geographic distribution and population density of vectors and hosts, as well as the prevalence of infection by pathogens and the pathogen load in individual hosts and vectors (Mills et al., 2010). Many bird species have extended their geographic range northward and adjusted their migratory routes due to climate change (Patterson & Guerin, 2013). Especially for VBDs, climate change may significantly impact the distribution of the arthropods vectors such as mosquitoes (Watson et al., 2005). Furthermore, climate change will likely affect vector, pathogen, and reservoir ecology, which could contribute to changes in the range limits and the intensity of disease transmission (Patterson & Guerin, 2013), and redistribute some VBDs to areas not previously affected (Watson et al., 2005).

### **3.4 Modelling the potential outbreak risks of vector-borne diseases**

In history, there have been many catastrophic outbreaks of infectious diseases such as the Spanish flu, Malaria, etc. (Johnson & Mueller, 2002; Taubenberger & Morens, 2006; WHO, last accessed 26, Nov, 2020c). Many of these diseases were vector-borne, and millions of people died from them over the centuries. However, this kind of disaster was not understood, nor predictable. Through the development of epidemiological theories, computer science, and new technologies such as remote sensing during the last few decades, it has become possible to simulate the spatial and temporal outbreak risk of some infectious diseases. Similar to a weather forecast, this risk mapping can provide information on when and where the outbreak risk is high, e.g. (Zinszer et al., 2012). Note that the term *risk* here means the potential of an outbreak in space or time, not necessarily a probability or likelihood; it can be quantified in certain cases, but not always.

To answer the question of when and where the outbreak risks of infectious diseases calls for proactive management and precaution, modelling approaches are needed to assess upcoming developments. The complexity of the transmission processes of VBDs requires the application of modelling approaches at different spatial and temporal scales (Tjaden et al., 2018), such as ecological niche models and process-based epidemiological models.

## 4 Ecological niche models

Ecological/environmental niche modelling (ENMs) approaches are widely used in the field of conservation ecology, disturbance ecology, biogeography, etc. (Evans et al., 2010; Mainali et al., 2015; Peterson, 2014; Regos et al., 2018; Villero et al., 2017). They are also often applied to project potential spatial distribution of investigated species and diseases, especially when the spatial aspect of risk is emphasized (Tjaden et al., 2018). ENMs are correlative models built with occurrence locations of investigated species (disease transmission can be treated as a species) and respective explanatory ecological/environmental variables. Those correlative models can be projected to future scenarios or to hind-cast past distributions of the investigated species. Not surprisingly, ENMs are also often referred as spatial distribution models, though these two terms are not quite synonymous to each other (Peterson & Soberón, 2012). Peterson and Soberón (2012) argued that any use beyond the actual distribution of investigated species would reach the realm of niche modelling, e.g. projecting future distribution of a species, or potential distribution under climate change. Consequently, in this thesis, I use the term “ecological niche modelling”.

### 4.1 General workflow

ENMs typically require geographical *occurrence locations* of the investigated species. These locations are then related to several *explanatory variables* (spatial raster data describing the environmental and/or socio-economic conditions in the study area), forming a correlative model of the species' ecological niche. This model can be projected onto regions where the environmental suitability of the species is unknown, resulting in a map showing probability of presence or environmental suitability.

### 4.2 Occurrence/presence locations

Occurrence locations of the investigated species are spatial (geographical) locations where this species were recorded. They are also referred as *presence locations*, especially when mentioned together with the counterpart state: absence. According to how these occurrence locations are collected, they can be grouped into three categories: presence-only, presence-absence, and occupancy detection (Gelfand & Shirota, 2019; Wang & Stone, 2019). For presence-only sampling, the occurrence locations are arbitrarily recorded, e.g. citizen science based occurrence locations are from presence-only sampling method. When the sampling was done through a systematic manner, e.g. dividing the study area into several sampling units, and noting down where the target species was found and where not, then it is presence-absence sampling (Gelfand & Shirota, 2019). Of course, the recorded “presence” are generally more trustworthy than the recorded “absence”, as the latter consist of “true absence” and “false absence”. There are also “false presence”, but this is not as common. False presence can happen when the species is misidentified (Miller et al., 2011), e.g. misidentify dog foot prints as fox's during snow-track surveys, or misidentify badgers as raccoons, etc. The difference between presence-only sampling and presence-absence sampling can sometimes be ambiguous, while the most distinct difference is that by presence-absence sampling,

the locations are often planned before the sampling takes place (Gelfand & Shirota, 2019). For instance, a dataset with presence-only records can be yielded by discarding the absence records from a presence-absence dataset. By discarding the absence records, the type of data has changed, but the sampling method not.

If the sampling procedure is not only following a spatially systematic design, but also including several rounds of revisiting, this sampling method is generally called occupancy detection (MacKenzie et al., 2017; Rota et al., 2009; Wang & Stone, 2019). This kind of sampling is rather rare, as it is more time-consuming and needs much more effort compared to the former two sampling methods, especially revisiting the sampling plots. Occupancy detections are often applied to estimate relative species abundance, population dynamics, individual animal territory, etc. (Hines et al., 2010; MacKenzie et al., 2003; Royle et al., 2005).

Among these three sampling methods, occupancy detection sampling requires the most effort, followed by presence-absence sampling, and presence-only sampling requires the least effort. It is not difficult to see that the occupancy detection method achieves the richest information content, and presence-only sampling the poorest. Due to the huge effort required by occupancy detection and presence-absence sampling, most of cases, only occurrence records collected by presence-only sampling are available. However, a typical problem with presence-only sampling is that the occurrence records are clustered (Kramer-Schadt et al., 2013; Varela et al., 2014). The sampling effort can be dramatically different across the investigated area (Fourcade et al., 2014; Stolar & Nielsen, 2015), and very often this sampling effort information is not available (Collins et al., 2017; Jones et al., 2010; Park & Davis, 2017).

### **4.3 Explanatory variables**

Explanatory variables are the variables which are associated with and supposed to contribute to the captured distributional pattern of the target species. These variables form a multi-dimensional environmental space, and the ENMs construct the niche of the target species by the respective occurrence locations (Peterson & Soberón, 2012). consequently, for explanatory variables, it needs to be explained or at least explainable that why they are chosen. These variables are supposed to define the niche/distribution of the target species. For instance, while modeling the potential distribution of mosquitoes, temperature relevant variables would be considered, as temperature affects mosquito survival.

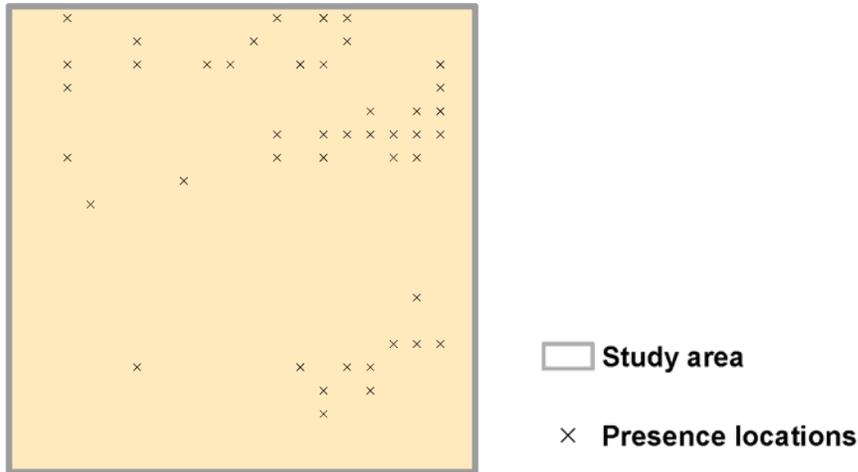
While selecting the explanatory variables, however, it needs to be borne in mind that correlation does not imply causation, i.e. a factor being correlated with the observed pattern does not mean this factor really causes this pattern. If a certain variable improves the ENM performance, but this variable does not have an explainable effect on the investigated species, following Occam's razor, this variable should not be included. To select the variable candidates, some primary questions should be asked: (1) Can this variable define the niche of the investigated species (binary question, either yes or no)? (2) How does it contribute to defining the niche (should be explainable)?

Another potentially critical point is the use of proxy variable in ENMs. For instance, some species only occur above or below a certain elevation, suggesting the inclusion of a digital elevation model as an explanatory variable in an ENM for such species. However, rather than elevation itself, the true limiting factor is often temperature, UV exposure, or water availability, etc. (Reyes et al., 2020; Watermann et al., 2020). As elevation is highly correlated with these factors, it can serve as a proxy for them. For a relative small study area, using a proxy instead of the direct factor may not change the ENM output, as the correlation between these two factors is likely to be consistent across a small area. However, when the correlation is inconsistent across the study area, using the proxy will lead to wrong conclusions.

#### **4.4 How do ecological niche models use spatial information?**

When we look at the input and output of ENMs, we may have the impression that ENMs work with spatial information directly. This is not the case. The occurrence locations are recorded as geographic coordinates. These coordinates are then related to the explanatory variables, and ENMs work with those explanatory variables values. The coordinates are only used to extract the respective variable values, navigating the ENMs where to perform the value extractions.

In the following, I show the primary idea of a correlative ENM with a series of simplified conceptual figures. There are many different types of ENMs. To understand what tasks ENMs handle, and how they process data, I apply this simple example to show the generalizations among ENMs.

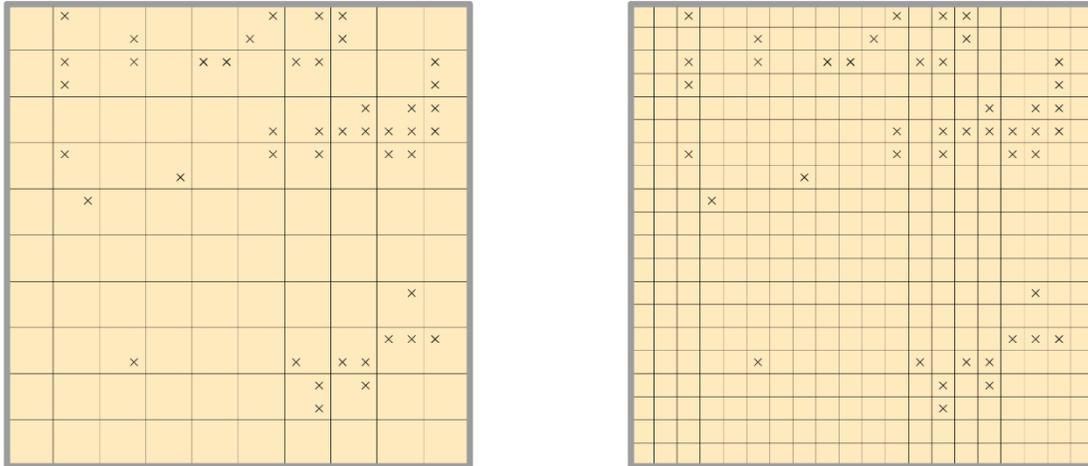


**Figure 4.1.** An example of presence-only occurrence records (presence locations) in a rectangular study area. The crosses stand for presence locations, and the yellow color for locations where presence-absence state is unknown.

For a given study area (which can be of any shape and any size), there are some known presence-only occurrence locations of species A, while there are still places for which the presence-absence state is unknown (Figure 4.1). Assume that species A's niche is defined by three known environmental variables. Now the task is to predict where inside the study area the environment is suitable for species A, and where it is not.

To handle the question of where the environment is suitable and where it is not, we need a spatial reference system. A simple solution is to divide this whole area into equal sections and number them. These grid cells can be large, as in Figure 4.2 (a), sometimes with more than one occurrence records in a single unit; they can also be small, as in Figure 4.2 (b).

In terms of numbering them, for Figure 4.2 (a), we can easily do that from left to right using simple numbers, as there are only 100 pieces in total. However, we still need to connect the occurrence locations and the grid cells. To achieve this, it makes sense to locate the occurrence records using the same coordinate system. The commonly used longitude-latitude geographic information system serves this purpose very well. The occurrence records collected from the field can be noted down using longitude-latitude coordinates, or any other kind of geographic information system coordinates (for instance, Universal Transverse Mercator coordinate system), as long as these locations can be located in the respective system.



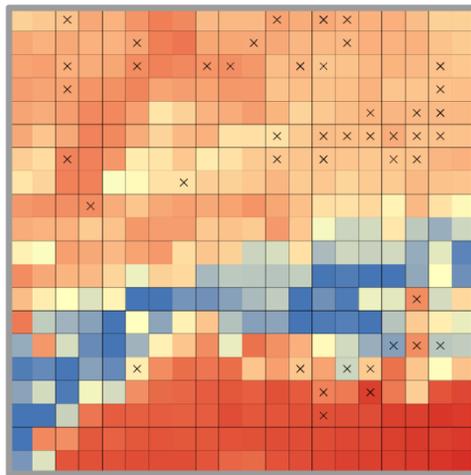
**Study area**
× **Presence locations**

**(a)** Divide the study area into 100 pieces

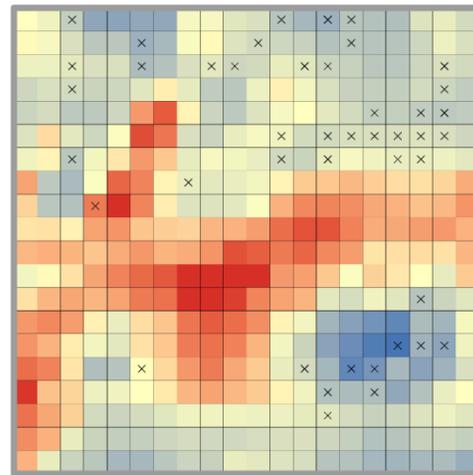
**(b)** Divide the study area into 400 pieces

**Figure 4.2.** Examples on different grid cells. (a) big grid cell and (b) small grid cell. The crosses stand for presence locations, and the yellow color for locations where presence-absence state is unknown.

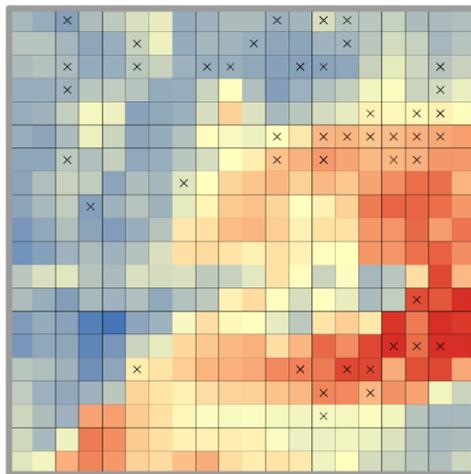
Then next question is: Where to store the explanatory variable values? Ideally, these values should be related to the occurrence records easily. To achieve this, these explanatory values are normally stored in rasterized grid cells (Figure 4.3). By using the same geographical coordinate system, these grid cells and the occurrence records are easily matched. Consequently, the occurrence records and the explanatory variable values from the grid cells are matched as well.



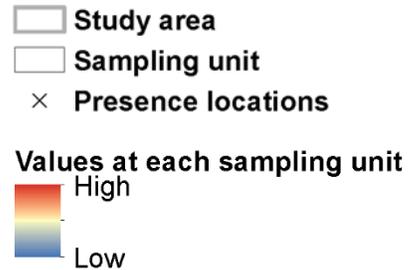
(a) Explanatory variable 1



(b) Explanatory variable 2



(c) Explanatory variable 3



**Figure 4.3.** A simplified conceptual figure of typical input data for ecological niche modelling, with presence-only occurrence records and three explanatory variables. From the spatial pattern of occurrence records and respective explanatory variables values, we can already guess intuitively where the environment is generally suitable and where it is not. When using ENMs, we can form a formula describing the relationship of occurrence and respective explanatory variables' values, and project this formula onto places where the presence-absence status is unknown.

#### 4.5 More than a black box: How do ecological niche models work?

To take a closer look at how ENMs work and why sampling bias may affect ENM output, we can start with the probability distribution of getting a presence record. Assume that 1) when we visit a big, far away forest, we have 0.6 chance seeing at least one roe deer, qualifying as a presence record; 2) when we visit a small, nearby forest, we have 0.2 chance getting a presence record of roe deer. The probability of getting a presence record for either scenario follows a binominal distribution. If we visit each forest 100 times, the expected presence records are 60 and 20, respectively. However, if we visit the big far away forest 10 times, the small near-by forest 100 times, the expected presence records are 6 and 20, respectively.

To generalize the possible scenarios: Assume that we divide the big forest into  $n$  sections, in a unit of time, the probability of seeing a roe deer in each section is  $p$ , and in  $x$  sections we have observed the roe deer. The expectation of seeing roe deer in a unit of time is  $\lambda$ . The observations (the total number is  $Y$ ) of roe deer follow a binominal distribution and can be described with a simple equation:

$$P_{(Y=x)} = C_n^x p^x (1-p)^{(n-x)}, \quad x=1,2, \dots n. \quad \text{eqn. 2-1)}$$

$$\lambda = np \quad \text{eqn. 2-2)}$$

Replace  $p$  with  $\frac{\lambda}{n}$ , eqn. 2-1) can be written as:

$$\begin{aligned} C_n^x p^x (1-p)^{(n-x)} &= \frac{(n-x)!}{x!} * \left(\frac{\lambda}{n}\right)^x * \left(1 - \frac{\lambda}{n}\right)^{(n-x)} \\ &= \frac{\lambda^x}{x!} * \frac{(n-x)!}{n^x} * \left(1 - \frac{\lambda}{n}\right)^{(n-x)} \end{aligned} \quad \text{eqn. 2-3)}$$

With increasing  $n$ ,  $p$  is decreasing; when  $n$  is very large, parts of eqn. 2-3) can be simplified:

$$\lim_{n \rightarrow \infty} \frac{(n-x)!}{n^x} = 1 \quad \text{eqn. 2-4)}$$

$$\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^{(n-x)} = \lim_{n \rightarrow \infty} \left\{ \left(1 + \frac{1}{-\frac{n}{\lambda}}\right)^{-\frac{n}{\lambda}} \right\} = e^{-\lambda} \quad \text{eqn. 2-5)}$$

Then we get the Poisson distribution:

$$\lim_{n \rightarrow \infty} C_n^x p^x (1-p)^{(n-x)} = \frac{\lambda^x}{x!} e^{-\lambda} \quad \text{eqn. 2-6)}$$

Now the only parameters needed are  $\lambda$  and  $x$ .

For a random number of occurrence locations/small sections ( $n$  denotes all sections within the study area,  $x$  denotes total number of observations) in the study area ( $S$ ), intensity (averaged expectation)  $\bar{\lambda} = np_n$ . Note that the intensity itself is also a probability, it has two dimensions. In this example, we ignore the temporal dimension by only looking at one unit of time. For a more detailed two-dimensional version see (Hefley & Hooten, 2016). Then we get:

$$[x|\lambda_{(n)}] = \frac{\bar{\lambda}^x}{x!} e^{-\bar{\lambda}} \quad \text{eqn. 2-7)}$$

$$\bar{\lambda} = \int_S \lambda_{(n)} d_n \quad \text{eqn. 2-8)}$$

For observed occurrence records locations  $Y$ , each location is denoted with  $y_i, i = 1, \dots, x$ .

$$[Y|\lambda_{(n)}, x] = \prod_{i=1}^x \frac{\lambda_{(y_i)}}{\bar{\lambda}} = \prod_{i=1}^x \frac{\lambda_{(y_i)}}{\int_S \lambda_{(n)} d_n} \quad \text{eqn. 2-9)}$$

Combine the total number of occurrence locations ( $x$ ) and the observed occurrence locations ( $y_i$ ):

$$[Y|\lambda_{(n)}] = \frac{\bar{\lambda}^x}{x!} e^{-\bar{\lambda}} * \prod_{i=1}^x \frac{\lambda_{(y_i)}}{\int_S \lambda_{(n)} d_n} \quad \text{eqn. 2-10)}$$

Simplify the calculation (to replace the  $\prod$  with  $\sum$ ) via log transformation:

$$l(\lambda; Y) = \sum_{i=1}^x \log \lambda_{(y_i)} - \int_S \lambda_{(n)} d_n - \log(x!) \quad \text{eqn. 2-11)}$$

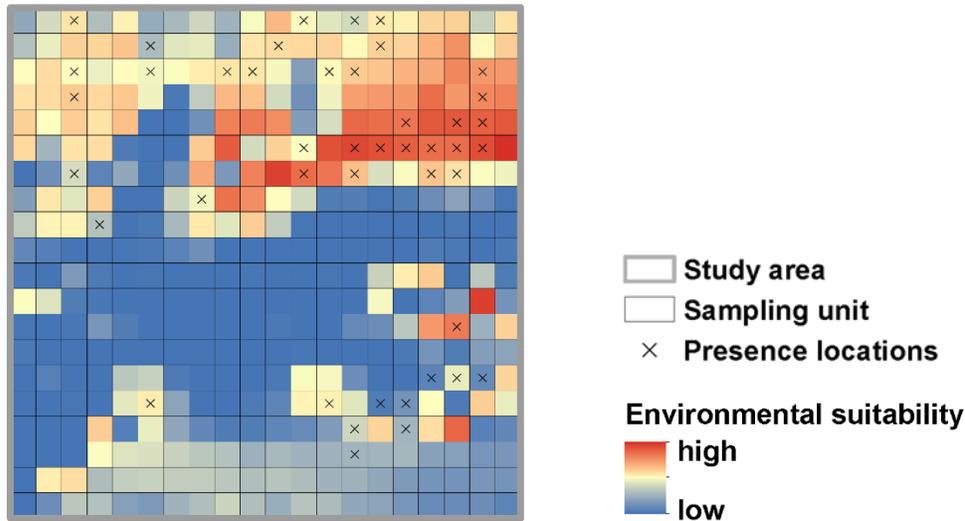
If a linear combination of covariates is applied, the intensity function can be written as (Dorazio, 2014; Hefley et al., 2017; Renner et al., 2015):

$$\log(\lambda_{(y_i)}) = \beta_0 + \mathbf{v}(y_i)' \boldsymbol{\beta} \quad \text{eqn. 2-12 )}$$

Or as suggested by Warton et al. (Warton & Shepherd, 2010):

$$\log(\lambda_{(y_i)}) = \beta_0 + \sum_{j=1}^k v_{ij} \beta_j \quad \text{eqn. 2-13 )}$$

Here,  $\beta_0$  is the intercept,  $\mathbf{v}(y_i)$  is a  $k \times 1$  vector (a total number of  $k$  explanatory variables) that contains covariates at any location  $y_i$ ; and  $\boldsymbol{\beta}$  (also a  $k \times 1$  vector) is the respective regression coefficients ( $\boldsymbol{\beta} = \beta_1, \dots, \beta_k$ ) (Hefley et al., 2017; Warton & Shepherd, 2010).  $v_{ij}$  denote variables values,  $Y$  is a subset of  $n$ , consists of a total number of  $x$  occurrence records. Thereafter, pseudo-absence locations can be introduced, and a variety of methods can be applied to estimate those regression coefficients  $\boldsymbol{\beta}$  (Pearce & Boyce, 2006). A correlative model can be achieved and projected onto regions where the environmental suitability is unknown (Figure 4.4). If  $\mathbf{v}(y_i)$  is with errors, the estimation of  $\boldsymbol{\beta}$  will be consequently affected. It can be thus predicted, using occurrence locations with positional errors may lead to a decrease in model performance of ENMs (Manuscript 1).



**Figure 4.4.** Environmental suitability of species A as the output from Ecological niche models.

#### **4.6 Products from ecological niche models – spatial risk maps**

For VBDs, ENMs can be adopted to model the spatial distribution of the respective vectors or hosts, especially across large spatial extent, e.g. (Brownstein et al., 2003; Johnson et al., 2019). It has been increasingly recommended to apply ENMs in epidemiology, since ENMs are helpful in understanding the distribution and biogeography of diseases (Escobar, 2020; Johnson et al., 2019), and the risk maps generated from ENMs can achieve fine spatial resolution without the loss of information (Peterson, 2006). More importantly, future projections of the effects of climate change on the investigated VBD's distribution can be achieved via ENMs as well (Carvalho et al., 2017; Fischer et al., 2013; Tjaden et al., 2018).

## 5 Epidemiological models

Process-based epidemiological models are also referred to as mathematical models, process-based models, mechanistic models and compartment models (Delamater et al., 2019; Heffernan et al., 2005). Depending on which aspect is emphasized, they are referred to accordingly. For instance, while being compared with ENMs, the term of “epidemiological models” will be preferred. For simplicity, in this thesis I stick to the term of “epidemiological models” (from here on EMs). EMs typically look into processes of infectious disease transmission. For VBDs, different health states and health state changes of both vectors and hosts are described with a series of parameters and variables, such as EIP, transmission rate, etc. Those EMs are typically applied with fine temporal resolution meteorological data, e.g. daily mean temperatures (Manuscript 2). Consequently, those models are good at capturing small temporal scale extreme weather event.

### 5.1 General workflow

As mentioned in earlier sections, factors such as climate change and globalization also bring in more challenges in disease control. Consequently, EMs are increasingly developed and widely applied nowadays. EMs include all the models that look into transmission mechanics of infectious diseases, while sometimes ENMs are also counted. In general, however, EMs refer to the models that are built with mathematical equations, which describe different health states of vectors and hosts. Depending on which property of EMs is emphasized, they are also referred to as compartmental models, mathematical models, compartmental epidemic models, mechanic models, process-based models or mechanistic models.

### 5.2 Basic reproduction number: $R_0$

The basic reproduction number  $R_0$  is an important threshold quantity parameter, which has even been called the “most important quantity in disease epidemiology” (Diekmann et al., 2010). Originally,  $R_0$  was defined as the average number of secondary cases caused by an infected individual during its lifetime in a completely susceptible population (often referred as a naïve population) (Diekmann et al., 1990; Heffernan et al., 2005; Rubel et al., 2008). When  $R_0 > 1$ , an outbreak of the investigated VBD can take place, otherwise, the transmission will die out on its own (Delamater et al., 2019; Diekmann et al., 1990; Hethcote, 2000; Liu et al., 2018).

$R_0$  has been treated as the spreading speed of certain disease (Ridenhour et al., 2014), and been consequently referred as “basic reproduction rate” (Dietz, 1993), as in this case  $R_0$  denotes the expected number of cases per generation. For an infectious disease,  $R_0$  determines the herd immunity threshold that is required to prevent it from further spreading (Dietz, 1993; Guerra et al., 2017), as seen earlier in 2020 for dealing with the COVID-19 pandemic, though the general message

was then misinterpreted by media (Anderson et al., 2020; Medley, 2020). The concept of the “basic reproduction” can be dated back to demography, where the “net reproduction rate” plays an important role (Dietz, 1993). However, depending on the method applied while constructing the EM, the interpretation of  $R_0$  also differs.

$R_0$  can be measured through disease surveillance systems (Marques et al., 1994) or estimated via EMs (Delamater et al., 2019; Ridenhour et al., 2018; Siettos & Russo, 2013). For VBDs, the originally defined  $R_{0 \text{ classic}}$  is difficult to assess directly from surveillance systems due to the significant workload required to conduct both monitoring and surveillance and to determine the complexity of transmission patterns of VBDs (Delamater et al., 2019). Instead, estimates of  $R_0$  are often based on EMs that calculate measures which share the threshold property (if  $R_0 > 1$ , the disease can spread). These measures are generally also called  $R_0$ , although they may differ considerably from the original definition in several aspects (see below) (Dietz, 1993; Heffernan et al., 2005; Yang, 2014). To construct EMs, there are mainly three methods available. They are all based on some kind of mathematical equations. Depending on which method is applied to calculate  $R_0$ , the interpretations differ accordingly, as explained in the following sections.

### 5.3 Survival function

The *survival function* method describes, through mathematical functions, whether or not a new infection can survive. A survival function calculates the *probability* of secondary infections by considering relevant parameters such as infectivity and survival probability (Heffernan et al., 2005). This method sticks to the original definition of  $R_0$ , i.e.  $R_0$  estimated with this method denotes the average number of secondary infections resulting from a single introduction in the infected individual’s lifetime. And of course, a transmission can be sustained if  $R_0 > 1$ , otherwise not, which is in line with the original definition as well.

For a VBD, the average number of contacts between the vectors and the hosts is  $C$ , the average infection rate through each contact is  $I$ , the infectious time period of the infected individual is  $t$ , then

$$R_0 = \int_0^t C \times I dt \quad \text{eqn. 3-1 )}$$

In eqn. 3-1), both  $C$  and  $I$  can be constant values or variables; when  $C$  and  $I$  are constant values, this equation can be written as:

$$R_0 = CIt \quad \text{eqn. 3-2 )}$$

The components can be temperature or time-dependent, e.g. the infection rate can vary through time, the contact rate between vectors and hosts can be associated with temperature, etc. As we can see, the  $R_0$  estimated here is indeed the “expected secondary cases”. For a simple case like this, this method is very straightforward. However, the complexity of this equation can increase very quickly, especially when the transmission cycle is complicated, e.g. (Holy et al., 2011; Johansson et al., 2012; Mordecai et al., 2017; Ng et al., 2017; Parham & Michael, 2010; Perkin et al., 2019). For example, the  $R_0$  equation from (Ng et al., 2017) is no longer as simple:

$$R_0 = R_0^{HM} R_0^{MH} = \varphi \alpha^2 \beta_{HM} \beta_{MH} L V \gamma \quad \text{eqn. 3-3 )}$$

$\varphi$ : Mosquito density per human

$\alpha$  : Daily biting rate

$\beta$  : Transmissibility, HM from human to mosquitoes and MH *vice versa*

$L$ : Average adult mosquito lifespan in days

$V$ : Duration of the human infectious period

$\gamma$  : Proportion of mosquitoes surviving the Extrinsic incubation period

$H$ : human

$M$ : mosquito

Especially when these parameters vary through time or are dependent on temperature, the  $R_0$  equation can be very complex. In this equation, only cross-transmission (from human to mosquito or the other way around) is considered. However, for many diseases, the transmission cycle also includes vertical transmission (from mother to offspring), adding another layer of complexity. It is not impossible but difficult to build an EM, based upon the survival function method, that includes both vertical and cross-transmissions. Consequently, other methods for constructing EMs have been developed that are easier to handle in practice.

## 5.4 Next generation matrix

The *next generation matrix* (NGM) method is anything but a straightforward approach. NGM-based EMs use a series of ordinary differential equations (ODEs) to describe different health states of vectors/hosts, or any kind of sub-populations (e.g. age group). In the NGM, only “states-at-infection” are contributing to the calculation of  $R_0$ . The basic reproduction number  $R_0$  is defined as the dominant eigenvalue of this NGM (Diekmann et al., 1990).

The  $R_{0\text{NGM}}$  calculated with the NGM method, not surprisingly, has a different interpretation from the ones calculated with the survival function method. This  $R_0$  is introduced as “the typical number of secondary cases”, i.e. each “generation” is  $R_0$  times as big as the preceding one (Diekmann et al., 1990). Note that here the “generation” is not a natural generation of vectors or hosts, but for the “infections;” the new infections are classified as a “species” in this case. This  $R_{0\text{NGM}}$  is a geometric mean through “generations.” Since introduced,  $R_{0\text{NGM}}$  has soon become common practice concerning VBDs (Ridenhour et al., 2014).

For a VBD,  $v$  denotes vectors and  $h$  hosts,  $n$  the transmission between them, e.g.  $n_{vh}$  denotes the transmission from vectors to hosts.

The next generation matrix can be written as:

$$\begin{vmatrix} n_{hh} & n_{hv} \\ n_{vh} & n_{vv} \end{vmatrix} \quad \text{Matrix 1)}$$

This 2x2 matrix has two eigenvalues:

$$R_{01} = \frac{1}{2} [(n_{hh} + n_{vv}) - \sqrt{(n_{hh} + n_{vv})^2 - 4(n_{hh}n_{vv} - n_{hv}n_{vh})}] \quad \text{eqn. 3-4)}$$

$$R_{02} = \frac{1}{2} [(n_{hh} + n_{vv}) + \sqrt{(n_{hh} + n_{vv})^2 - 4(n_{hh}n_{vv} - n_{hv}n_{vh})}] \quad \text{eqn. 3-5)}$$

$R_0$  is defined as the largest eigenvalue, thus  $R_0 = R_{02}$ .

$$R_0 = \frac{1}{2} [(n_{hh} + n_{vv}) + \sqrt{(n_{hh} + n_{vv})^2 - 4(n_{hh}n_{vv} - n_{hv}n_{vh})}] \quad \text{eqn. 3-6)}$$

From this equation, it is not difficult to see that, in the absence of vertical transmission (i.e.  $n_{hh} = 0$  and  $n_{vv} = 0$ ), the matrix is:

$$\begin{vmatrix} 0 & n_{hv} \\ n_{vh} & 0 \end{vmatrix} \quad \text{Matrix 2)}$$

The equation of  $R_0$  then is even simpler:

$$R_0 = \sqrt{n_{hv}n_{vh}} \quad \text{eqn. 3-7 )}$$

The equation 3-7) is the geometric mean of expected secondary cases from vector to host and *vice versa*. When an NGM-based model is as simple as this, it is not so different from the survival-function-based model. In this example, the elements such as  $n_{hv}$  or  $n_{vh}$  on their own are expressed with survival functions. Indeed, the eqn. 3-7) is the square root of the eqn. 3). However, this kind of connection between EMs built with different methods is not ensured for all situations. Here is just an example, and it is only true when the transmission cycle is simple and clear like this case.

This  $R_{0\text{NGM}}$  shares the property of threshold quantity with the  $R_{0\text{classic}}$  (as mentioned above, when  $R_0 > 1$ , the disease can spread, otherwise dies out), which can be proved by its relationship with the Malthusian parameter (or “natural rate of increase”, often denotes as  $r$ , see (Dietz, 1993)). If we treat the new infections of a VBD as a species, the rate of “epidemiological birth” is the natural increase rate ( $r$ ) of the infections, from the perspective of “population”. The  $r$  can well depict the growth of new infections, but is very often not possible to achieve (more details see (Diekmann et al., 2010)), whereas the  $R_{0\text{NGM}}$  can be calculated through the NGM. The  $R_{0\text{NGM}}$  is applied indicating the “population dynamic” of the new infections, as the *implicit* relationship between  $R_{0\text{NGM}}$  value and the  $r$  stands (Diekmann et al., 2010):

$r > 0$ , if and only if  $R_0 > 1$ ;

$r = 0$ , if and only if  $R_0 = 1$ ;

$r < 0$ , if and only if  $R_0 < 1$ .

$R_{0\text{NGM}}$  only indicates if the introduction of a new VBD can survive (the sign of  $r$ , positive or negative), and a high  $R_{0\text{NGM}}$  value does not automatically equal fast exponential increase of incidence (Diekmann et al., 2010). In addition, it has been suggested that the transmission is still possible for some EMs even if  $R_{0\text{NGM}} < 1$  (near this threshold) (Ma & Li, 2009; van den Driessche & Watmough, 2002). In consequence, this must be considered when interpreting EM outputs or transforming the results into spatial maps.

Breaking down the transmission between vectors and hosts into finer processes will result in relatively complex EMs. The calculation of eigenvalue is no longer as simple, nor straight forward. Those complex NGM are often solved by a Jacobian transformation (Diekmann et al., 1998; Diekmann et al., 2010; van den Driessche & Watmough, 2002).

To build the Jacobian matrix, health states (e.g. susceptible, exposed, infected-infectious, recovered) of vectors and hosts, described with ODEs, are grouped into two parts: transmission and

transition (Diekmann et al., 2010). The *transmission* part includes new infections from all directions; and *transition* part includes health state changes. A detailed explanation on this can be found in (Diekmann et al., 2010). There, recipes for constructing NGM-based models are also provided. The Jacobian transformation method efficiently simplifies the calculation of the eigenvalues (see the small-domain NGM in Diekmann 2010), especially when multiple health states, age groups, or multiple vectors or hosts are included.

For instance, following the recipe suggested by (Diekmann et al., 2010), I constructed a model on West Nile virus transmission, including both local birds and migratory birds (See Appendix 1 and Chapter 8). The method suggested by Diekmann, Heesterbeek et al. 2010 is relatively easy to handle compared to the classical method suggested by van den Driessche (van den Driessche & Watmough, 2002), and it also provides reasonable biological interpretations.

## 5.5 Incidence-based methods

In addition to those two commonly used ways explained above, it is also possible to approach  $R_0$  from incidence-based methods, e.g. by estimating time varying effective reproduction number from past incidence (Peña García & Christofferson, 2019; Perkin et al., 2019). However, this method does not look into the processes of disease transmission. Instead, it estimates the number of cases in the previous generation by recorded incidence rate, within a certain area or administration region (Jentes et al., 2016). This method assumes a certain distribution of the time interval, i.e. the time between an induced secondary case and a primary case (Jentes et al., 2016; Peña García & Christofferson, 2019; Perkin et al., 2019). Based on this assumed time interval,  $R_0$  can be estimated for previous generations. As it highly relies on incidence rate data and does not look into transmission processes of VBDs, this method has not been applied as often as those two common methods explained above.

## 5.6 Products from epidemiological models – spatio-temporal risk maps

To date, it is common to see single values or time-series curves of  $R_0$  for a certain place being reported from the output of EMs (Manuscript 2). This kind of product from EMs shows temporal risk dynamics for one or several places, providing information on *when* potential outbreak of the investigated disease is high, e.g. (Mordecai et al., 2017; Ng et al., 2017; Rubel et al., 2008). However, to monitor the potential outbreak risk of a VBD, this information on its own is not enough. It is essential to also know *where* the risk is higher; more ideally, when and where (Manuscript 2).

Spatial risk maps can also be produced by EMs, resulting in spatial-temporal risk maps (Manuscript 2). There are some methods available for achieving this: (1) Calculating the  $R_0$  for some scattered places (e.g. the location of weather stations in several cities) and then mapping this value into a

certain political region (Hartemink, Purse et al. 2009). (2) Calculating the correlation between  $R_0$  and a highly correlated variable, then applying this correlation onto the variable layer (gridded raster file), e.g. (Wu, Duvvuri et al. 2013). (3) Calculating the  $R_0$  for each gridded cell of a raster file, which is often a temperature observation layer, resulting in spatial risk maps directly, e.g. (Holy, Schmidt et al. 2011, Cadar, Luehken et al. 2017).

## 6 Comparison of ecological niche models and epidemiological models

Generating risk maps for potential outbreak risk of VBDs is crucial for public health (Escobar & Craft, 2016). Those risk maps for VBDs can be achieved by using both ENMs and EMs (Escobar & Craft, 2016; Johnson et al., 2019; Tjaden et al., 2018). Of course, those two modelling approaches each have their own pros and cons (Tjaden et al., 2018). Relying on a single model for assessing potential outbreak risk of VBDs may lead to incomplete conclusions (Manuscript 3). As (Escobar & Craft, 2016) point out, ecology and epidemiology are highly related disciplines, and both modelling approaches can contribute to improvements regarding disease risk mapping (Escobar & Craft, 2016; Tjaden et al., 2018). Consequently, it should be considered to draw an interdisciplinary view and make use of both models' advantages.

As explained earlier (in chapter 4), ENMs use occurrence data and environmental data to make a correlative model of the potential suitable environmental conditions (Warren & Seifert, 2011). They have many potential applications regarding the geography (for vectors, hosts, pathogens, etc.) and ecology of disease transmission (Peterson, 2006, 2014), and are increasingly used to gain understanding of biogeography of diseases (Escobar & Craft, 2016). Compared with EMs, ENMs have the advantage that they can easily include a wild variety of explanatory variables, such as land use types and precipitation variables. In particular, they can include not only the factors that has a positive effect on the investigated species' potential distribution, but also the ones with negative effect (Although theoretically, it is also possible to include factors with negative effect in EMs, it has not been commonly done yet). In addition, ENMs do not require an in-depth understanding of each variable included in the modelling process. ENMs are typically run with mid- or long-term climate data. In general, ENMs can well reflect spatial distribution of the investigated species, especially that they can achieve risk maps at fine spatial resolutions (Peterson, 2006, 2014). The drawback of using climatic data is that extreme weather events such as drought or heat waves cannot be captured. However, those factors are very important for VBDs.

EMs, on the other hand, are process-based. They include different health stages of and the interaction between vectors and hosts (Diekmann et al., 2010; Heffernan et al., 2005; van den Driessche & Watmough, 2002; van den Driessche & Yakubu, 2018). EMs typically use meteorological data with fine temporal resolutions, such as daily or monthly mean temperature data (Manuscript 2). Consequently, they have the advantage of capturing short term extreme weather events or temperature fluctuations. However, it should still be kept in mind that EMs usually use daily mean temperature as input data, i.e. diurnal variations are *not* captured. Nevertheless, compared with ENMs, EMs are better at depicting temporal aspect of potential outbreak risk of VBDs (Tjaden et al., 2018). The drawback of EMs is that the spatial transferability of those models is very limited (Manuscript 2). In addition, EMs are highly dependent on the understanding the disease transmission cycle, and respective parameters for both vectors and hosts (Manuscript 2 & 3). Especially for some new emerging VBDs, those parameters are seldom available, e.g. Usutu virus (Manuscript 3).

Clearly, an integrated/interdisciplinary model could benefit from both models' advantages. However, due to the magnitude of the workload required for running both models, an integrated model has not been done yet. A possible way to do this is to apply a hierarchical approach, where the spatial distribution of the potential transmission or the respective vectors could first be estimated by an ENM, then zoomed into a finer scale at the high risk areas via EMs. However, this approach requires both occurrence locations of the respective VBDs, and also a good understanding of disease transmission cycles, i.e. parameters and variables, etc. For well-studied VBDs, this approach can work as a live early warning forecast with spatio-temporal information of the potential outbreak risk. The results from ENMs, such as the spatial distribution of vectors or hosts, can also work as input data for epidemiological modelling approaches as well. The integrated EMs can offer more fine-grained insights on the progression of outbreaks, with the potential for short-term forecasts based on weather models. In addition, ENMs can estimate not only where climate is suitable, but also where it is unsuitable. The unsuitable regions can be thus excluded, or classified as low-risk regions. In this case, the results from ENMs can be applied to classify the study area into different risk levels. However, these possibilities have not been tested yet. Consequently, Manuscript 3 compares those two modelling approaches as a first step towards constructing an interdisciplinary early warning system.

## 7 Synopsis of the following manuscripts

In the following, I will give a short overview of the manuscripts included in this thesis.

**Manuscript 1** deals with a long existing but often ignored problem in ecological niche modelling (ENM) – the usage of geographical centroid locations. ENMs typically use occurrence locations to extract the respective values of explanatory variables, and thus construct a correlative model of the investigated species' potential environmental suitability. When precise occurrence locations are not available, the geographical centroid locations of respective Administrative Spatial Units (ASUs) are commonly applied as a substitute. However, the values of explanatory variables at those substitutes may differ considerably from that at occurrence locations.

Manuscript 1 investigates how the use of ASU centroids in ENMs affects model performance, and which role the size of ASUs plays in the context of varying grain size of explanatory variables. In order to answer those questions, a two-factorial study design was applied with artificial ASUs of three different sizes and environmental data of four commonly used grain sizes, repeated over 3 study regions. To best control other factors that may affect ENM performance, a virtual species was generated. In this case, the exact explanatory variables are known, the occurrence locations are precise and drawn from even sampling. More importantly, the model performance can be assessed with non-threshold measures such as Spearman's rank correlation, since the "true suitability map" of the virtual species is available. The *value frequency mismatch* was applied as a novel indicator of potential projected niche mismatch. This method calculates, for each explanatory variable, the value frequency mismatch between ASU centroids and occurrence locations. Compared with conventional statistical tests (such as ANOVA or Kruskal-Wallis-test) that show whether a statistical difference between centroid-based and true location-based environmental data exists or not, this method focuses on quantifying the differences between the two.

Manuscript 1 shows that the usage of ASU centroids leads to value frequency mismatch of explanatory variables between the true locations and the centroids. This mismatch increases with increasing ASU size; as large ASUs tend to have large spatial heterogeneity. Compared with the ASU size, grain size matters not as much. In general, ENMs built with ASU centroids over-estimated the species' distributional range. Manuscript 1 highlights that in order to gain optimal model performance, substitutes such as central tendency values should be considered when possible. Special cautions should be given when ASU centroids are applied as substitutes.

**Manuscript 2** gives an insight into the preparation of spatio-temporal risk maps from EMs. EMs typically use the basic reproduction number ( $R_0$ ), a threshold quantity, to indicate potential outbreak risk of the investigated infectious disease. The output from EMs often is a series of risk maps of fine (e.g. daily or monthly) temporal resolutions. To provide an overall view of the risk, these EM outputs are often compiled into a single, summarizing map. However, depending on the specific temporal aggregation and visualization applied, the yielded spatial risk map may differ considerably.

The commonly used temporal aggregation methods are either averaging  $R_0$  over a certain time period (intensity) or counting the cumulative number of days with  $R_0 > 1$  (duration). Despite the fact that risk maps have been increasingly derived from EMs in last two decades, until now there were no standardized methods or guidelines available for this. Consequently, Manuscript 2 points out the importance of using appropriate temporal aggregations and visualizations, analyzes the advantages and disadvantages of each approach, and gives suggestions for best practice. Similarly, the importance of choosing an appropriate method for visualizing the data in a map, in practice, is often neglected. Via different visualizations, two fundamentally different maps can appear very similar, and *vice versa*. Though continuous representations of spatial risk (e.g. average  $R_0$ , cumulative number of days with  $R_0 > 1$ ) are most commonly used in maps, this is not always the best interpretation of EMs.

To improve the interdisciplinary communication regarding spatio-temporal risk maps generation from EMs, Manuscript 2 put forward general remarks as following: 1) different temporal aggregation methods lead to different interpretations; 2) similar-looking spatial patterns do not necessarily bear the same meaning; 3) visualization methods considerably affect how results are perceived, and thus should be applied with caution. Manuscript 2 suggests that both intensity and duration of the VBD outbreak risk should be included when possible, and preferably with small time-step intensity maps to show spatio-temporal dynamics. Categorized maps can be an option when comparisons cross EMs or different diseases are needed.

In **Manuscript 3**, two fundamentally different models are applied to map the potential outbreak risk of Usutu Virus (USUV) in Europe. Manuscript 3 aims at two goals: to estimate the potential risk for USUV transmission in Europe, and to compare ENM and EM approaches using USUV as a case study. This manuscript investigates how much, to our best knowledge, the outputs from those two disciplines differ. To do so, the EM on USUV was run with time-series of daily mean temperature data across Europe. The ENM was run with long-term bio-climatic variables, derived from the same source. Risk maps for USUV circulation were generated from both modelling approaches.

Manuscript 3 highlights the necessity of integrating different approaches to map the potential outbreak risks of VBDs. This is especially true for newly emerging pathogens like USUV, where the knowledge about important processes is still limited. A hierarchical or an integrated modelling approach could benefit from both models' advantages. For instance, the potential spatial distribution of the outbreak risk could first be estimated by an ENM as a general overview. Thereafter, for high risk areas, the temporal risk patterns can be modeled with EMs. ENMs can also be nested in an EM, e.g. to estimate the abundance of vectors and hosts, or even to exclude environmental unsuitable regions.

## 8 Challenges for future studies

As seen in chapter 4 and 5, both modelling approaches are rooted in mathematics. However, those mathematical processes and algorithms are only one part of the whole modelling processes. It is not difficult to see that the performance of either (or any) model is not decided by the respective algorithms alone. Instead, it depends on a series of factors, such as the quality of input data, the match of the input data type and the model algorithm, the calibration of the models, etc. To improve model performance, effort should be invested in those aspects.

### 8.1 Positional errors in ecological niche models

When precise occurrence locations are not available, geographical centroids of administrative spatial units with known occurrences are often applied as substitutes in ENMs solution (Collins et al., 2017; Jones et al., 2010; Park & Davis, 2017). This centroid method minimizes the largest possible spatial distance between the substitute location and the unknown true location. However, it does not necessarily minimize the mismatch of explanatory variables' values at the centroid locations and the true occurrence locations. This kind of positional errors brought in by the usage of centroids leads to a decrease in model performance, and has the potential for introducing extreme values of the respective explanatory variables (manuscript 1). In contrary to this centroid method, it has been suggested that using central tendency values (e.g., mean or median) across the respective administrative spatial unit, is a better solution (Collins et al., 2017; Jones et al., 2010; Park & Davis, 2017). Obviously, this central tendency method minimizes the largest potential difference between the substitute explanatory values and the values at the true locations (manuscript 1).

Tackling this positional error problem from another angle, Hefley et al. have suggested a change-of-support (COS) method (Hefley et al., 2017). Different from those two methods which attempt to find the "least-worst" substitutes, this COS method aims to find the "likely-best". As suggested by some researchers, the ecological niche modelling approaches are unified as the methods to estimate parameters of an inhomogeneous Poisson point process distribution (Hefley et al., 2017; Renner et al., 2015; Warton & Shepherd, 2010). In general, those inhomogeneous Poisson point process distribution are using continuous support; while Hefley et al. have applied a discrete distribution by using areal spatial support instead (the COS method). Basically, this COS method does not look at each spatial location (a single pair of coordinates), instead, it counts the number of occurrence records in each spatial unit.

In (Hefley et al., 2017), they have put forward this COS method more as a theoretical approach than a real tool to correct positional errors (only two variables and with linear effects). More effort should be put into investigating the performance of this COS method, and integrate it in future ecological niche modelling approaches. In Manuscript 1, I have investigated the effect of using

centroid data as substitutes of occurrence locations on the performance ENMs. As a follow-up study, this COS method can be tested with a virtual species, in a similar manner as Manuscript 1. Those three alternative methods can be thus compared in a controlled setting: 1) using centroid locations, 2) using central tendency values, and 3) using COS method instead. In the case of using the COS method is the best among those three methods, more effort should be given to integrate this method into ENMs.

## **8.2 The complexity of epidemiological models**

Modelling with VBDs, there is an increasing number of EMs, with different levels of complexity. For West Nile viral disease, several EMs have been constructed, with different levels of complexity (Bergsman et al., 2016; Bhowmick et al., 2020; Laperriere et al., 2011). For instance, the model constructed by (Laperriere et al., 2011) has eight compartments, including one host species and one vector species; while the model by (Bhowmick et al., 2020) has fifteen compartments, including two host species, one of which was even classified into two types. Bergsman's model also included both migratory birds and local birds as hosts, but it did not include the health state of "exposed" (Bergsman et al., 2016). Consequently, Bergsman's model only has nine compartments. Before even more complicated EMs being constructed, a critical question needs to be answered: does the model performance increase with model complexity? If the increase of model complexity does not increase the performance of the respective models, according to Occam's razor, we should avoid increasing the model complexity.

In order to answer this question for EMs for West Nile viral disease, the following approach is planned for the future: The three models mentioned above and the model attached as appendix (the example of complex model in chapter 5) will be updated with same variable and parameter settings where possible, so that the effect of increasing complexity can be isolated. There are three pairs of comparisons (Figure 8-1), and for each pair, the direction of complexity increase is clear. The complexity always only increases at one level – either through an increase of health states (i.e. from Figure 8-1 a) to b)), or through the introduction of a new host type (i.e. from Figure 8-1 b) to c) and d) to b)).

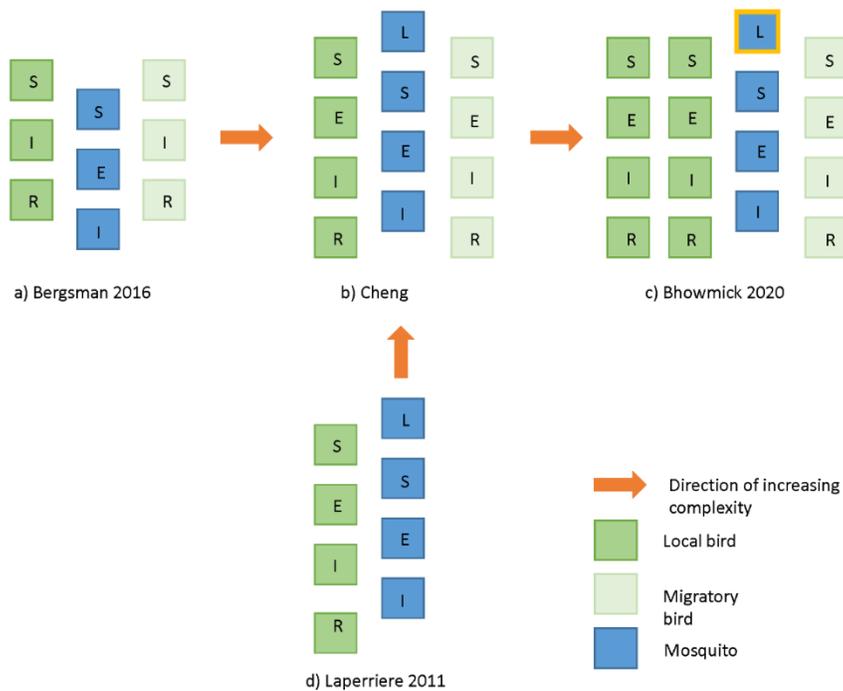


Figure 8-1. Four epidemiological models of West Nile virus transmission. The orange arrow indicates the direction of increasing complexity. Compartment abbreviations are: L = larvae state (of mosquitoes), S = susceptible, E = exposed, I = infectious, R = recovered. Compared with model a), model b) and c) includes the additional health stage of “Exposed” for both local and migratory birds, and adds the stage of “Larvae” to mosquitoes; while model c) additionally differentiates local birds into “clinical” and “subclinical” groups. Model b) differs from model d) by including migratory birds. The original model from Bhowmick in c) does not include the state “Larvae” for vectors. However, to make the complexity only increase from b) → c), the state “Larvae” needs to be included.

To assess how much the models’ output differ, i.e. how much they overlap, a possible solution is to use the quantile method suggested in Manuscript 2. If the overlap is very high (e.g. more than 90%), apparently, there is no need to increase the model complexity; if the overlap is very low (i.e. less than 50%), there must be something going wrong. Once we can assess model performance, it is also possible to discuss about the trade-off between model performance and complexity of models.

### 8.3 Assessment of epidemiological models’ performance

To investigate how the increase of model complexity affects model performance, the first challenge is to quantify the model performance. Although EMs are increasingly applied to estimate potential at-risk areas of VBDs (Escobar & Craft, 2016; Johnson et al., 2019; Peterson, 2014; Tjaden et al., 2018), the performance of those models are seldom assessed, if ever.

For VBDs with good quality of occurrence records, it is possible to make use of those occurrence records, e.g. to draw a certain amount of true occurrence locations of the respective VBDs (with locations and date info), and locations of potential-absence (where active surveillance is taking place). Once the true presence and highly likely absence (it is still possible to be false negative) are available, the model performance can be assessed.

#### **8.4 The quality of vector and host parameters**

For process-based EMs, the model performance is highly depended on the quality of input data, i.e. the vector and host parameters. Parameters such as EIP, biting rate of vectors and vector-to-host ratio are very important for EMs. However, for diseases caused by under-studied pathogens such as Usutu virus, Dengue and Chikungunya, those parameters are either not available or only available at a small spatial extent within the whole spatial distribution area. For instance, the parameters for the vectors of West Nile virus are usually mostly taken from (Reisen et al., 2006), more than a decade old. For some VBDs, such as Usutu virus, the vector and host ranges are still increasing, requiring additional laboratory work.

## 9 List of articles and declaration of own contribution

### Manuscript 1

**Authors:** Yanchao Cheng, Nils Benjamin Tjaden, Anja Jaeschke, Stephanie Margarete Thomas and Carl Beierkuhnlein

**Title:** Using centroids of spatial units in ecological niche modelling: effects on model performance in context with grain size of environmental data

**Journal and status:** *Global Ecology and Biogeography*, (2021) 30 (3): 611-621

**Authors' contribution:** **YC**, NT, ST, AJ, and CB conceptualized the study. **YC** and NT processed and analyzed the data and prepared the figures. **YC** wrote the first draft and prepared the appendices. NT wrote parts of the results section. CB, ST and AJ supervised the analyses. All authors discussed the methodology and results, and reviewed and edited the manuscript. **YC** is the corresponding author.

### Manuscript 2

**Authors:** Yanchao Cheng, Nils Benjamin Tjaden, Anja Jaeschke, Stephanie Margarete Thomas and Carl Beierkuhnlein

**Title:** Deriving risk maps from epidemiological models of vector borne diseases: state-of-the-art and suggestions for best practice

**Journal and status:** *Epidemics*, (2020) 33:100411

**Authors' contribution:** **YC**, NT conceptualized the review. **YC** did the literature searches, wrote the first draft, ran the epidemiological models, drafted the table and prepared the visualizations. NT had the initial idea for the review, co-wrote the results section, prepared parts of Figure 2 and edited the table and the manuscript. ST, AJ and CB reviewed and provided critical feedback on the draft. All authors discussed the concept, preliminary results and figures at various stages. All authors discussed and revised the manuscript. All authors read and approved the final version of the manuscript. **YC** is the corresponding author.

### Manuscript 3

**Authors:** Yanchao Cheng, Nils Benjamin Tjaden, Anja Jaeschke, Renke Lühken, Ute Ziegler, Stephanie Margarete Thomas and Carl Beierkuhnlein

**Title:** Evaluating the risk for Usutu virus circulation in Europe: comparison of environmental niche models and epidemiological models

**Journal and status:** *International Journal of Health Geographics*, (2018) 17:35

**Authors' contribution:** **YC**, NT, ST, AJ, RL, and CB developed the concept of the study. **YC**, UZ and RL compiled the occurrence records. NT and **YC** processed the climate data and adapted the SEIR model. **YC** and NT ran the Maxent models. CB, ST and AJ supervised the modelling process. **YC** prepared the figures. All authors discussed the preliminary results and figures at various stages of the modelling process. **YC** and NT wrote the original draft of the manuscript. All authors discussed and revised the manuscript. All authors read and approved the final version of the manuscript. **YC** is the corresponding author.

This manuscript is also part of Dr. Nils Tjaden's doctoral thesis.

## 10 References of introduction

- Abbott, A., 2016. The red-hot debate about transmissible Alzheimer's. *Nature*, 531, 294–297.
- Ahrens, W., & Pigeot, I. (Eds.). (2007). *Handbook of Epidemiology* (2nd edition ed.). Berlin, Heidelberg, New York: Springer.
- Anderson, R. M., Vegvari, C., Truscott, J., & Collyer, B. S., 2020. Challenges in creating herd immunity to SARS-CoV-2 infection by mass vaccination. *The Lancet*, 396, 1614-1616.
- Ashraf, U., Ye, J., Ruan, X. D., Wan, S. F., Zhu, B. B., & Cao, S. B., 2015. Usutu virus: an emerging flavivirus in Europe. *Viruses*, 7, 219-238.
- Bergsman, L. D., Hyman, J. H., & Manore, C. A., 2016. A mathematical model for the spread of West Nile virus in migratory and resident birds. *Mathematical Biosciences and Engineering*, 13, 401-424.
- Bhowmick, S., Gethmann, J., Conraths, F. J., Sokolov, I. M., & Lentz, H. H. K., 2020. Locally temperature - driven mathematical model of West Nile virus spread in Germany. *Journal of Theoretical Biology*, 488, 110117.
- Braks, M., van der Giessen, J., Kretzschmar, M., van Pelt, W., Scholte, E. J., Reusken, C., ... Sprong, H., 2011. Towards an integrated approach in surveillance of vector-borne diseases in Europe. *Parasites & Vectors*, 4, 192.
- Brown, C. R., Brown, M. B., Moore, A. T., & Komar, N., 2007. Bird movement predicts Buggy Creek virus infection in insect vectors. *Vector-Borne and Zoonotic Diseases*, 7, 304-314.
- Brownstein, J. S., Holford, T. R., & Fish, D., 2003. A climate-based model predicts the spatial distribution of the Lyme disease vector *Ixodes scapularis* in the United States. *Environmental Health Perspectives*, 111, 1152-1157.
- Cabral, B. P., Fonseca, M. D. D., & Mota, F. B., 2019. Long term prevention and vector control of arboviral diseases: What does the future hold? *International Journal of Infectious Diseases*, 89, 169-174.
- Carvalho, B. M., Rangel, E. F., & Vale, M. M., 2017. Evaluation of the impacts of climate change on disease vectors through ecological niche modelling. *Bulletin of Entomological Research*, 107, 419-430.
- CDC, last accessed 26, Nov, 2020a. Bovine Spongiform Encephalopathy (BSE), or Mad Cow Disease. <https://www.cdc.gov/prions/bse/index.html>.
- CDC, last accessed 26, Nov, 2020b. Malaria <https://www.cdc.gov/malaria/about/faqs.html>.
- CDC, last accessed 26, Nov, 2020c. Ringworm. <https://www.cdc.gov/fungal/diseases/ringworm/index.html>.
- CDC, last accessed 26, Nov, 2020d. Travelers' Health. <https://wwwnc.cdc.gov/travel/yellowbook/2020/list/maps>.
- CDC, last accessed 26, Nov, 2020e. Understanding Influenza Viruses. <https://www.cdc.gov/flu/about/viruses/index.htm>.
- Chinchio, E., Crotta, M., Romeo, C., Drewe, J. A., Guitian, J., & Ferrari, N., 2020. Invasive alien species and disease risk: An open challenge in public and animal health. *Plos Pathogens*, 16, 7.
- Cohen, E. B., Aucklan, L. D., Marra, P. P., & Hamer, S. A., 2015. Avian migrants facilitate invasions of neotropical ticks and tick-borne pathogens into the United States. *Applied and Environmental Microbiology*, 81, 8366-8378.
- Collins, S. D., Abbott, J. C., & McIntyre, N. E., 2017. Quantifying the degree of bias from using county-scale data in species distribution modeling: Can increasing sample size or using county-averaged environmental data reduce distributional overprediction? *Ecology and Evolution*, 7, 6012-6022.

- Crowl, T. A., Crist, T. O., Parmenter, R. R., Belovsky, G., & Lugo, A. E., 2008. The spread of invasive species and infectious disease as drivers of ecosystem change. *Frontiers in Ecology and the Environment*, 6, 238-246.
- CSSE, 2020. COVID-19 dashboard by the center for systems science and engineering at Johns Hopkins university.
- Daszak, P., 2000. Emerging infectious diseases of wildlife - Threats to biodiversity and human health (vol 287, pg 443, 2000). *Science*, 287, 1756-1756.
- Delamater, P. L., Street, E. J., Leslie, T. F., Yang, Y. T., & Jacobsen, K. H., 2019. Complexity of the basic reproduction number ( $R_0$ ). *Emerging Infectious Diseases*, 25, 1-4.
- Diekmann, O., Gyllenberg, M., Metz, J. A. J., & Thieme, H. R., 1998. On the formulation and analysis of general deterministic structured population models - I. Linear theory. *Journal of Mathematical Biology*, 36, 349-388.
- Diekmann, O., Heesterbeek, J. A. P., & Metz, J. A. J., 1990. On the definition and the computation of the basic reproduction ratio  $R_0$  in models for infectious-diseases in heterogeneous populations. *Journal of Mathematical Biology*, 28, 365-382.
- Diekmann, O., Heesterbeek, J. A. P., & Roberts, M. G., 2010. The construction of next-generation matrices for compartmental epidemic models. *Journal of the Royal Society, Interface*, 7, 873-885.
- Dietz, K., 1993. The estimation of the basic reproduction number for infectious diseases. *Statistical Methods in Medical Research*, 2, 23-41.
- Dorazio, R. M., 2014. Accounting for imperfect detection and survey bias in statistical analysis of presence-only data. *Global Ecology and Biogeography*, 23, 1472-1484.
- ECDC, 2020. COVID-19 situation update worldwide, as of 29 September 2020. <https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases>.
- Escobar, L. E., 2020. Ecological Niche Modeling: An Introduction for Veterinarians and Epidemiologists. *Frontiers in Veterinary Science*, 7.
- Escobar, L. E., & Craft, M. E., 2016. Advances and Limitations of disease biogeography using ecological niche modeling. *Frontiers in Microbiology*, 7.
- Evans, J. M., Fletcher, R. J., & Alavalapati, J., 2010. Using species distribution models to identify suitable areas for biofuel feedstock production. *Global Change Biology Bioenergy*, 2, 63-78.
- Ferguson, N. M., Cucunuba, Z. M., Dorigatti, I., Nedjati-Gilani, G. L., Donnelly, C. A., Basanez, M. G., ... Lessler, J., 2016. Countering the Zika epidemic in Latin America. *Science*, 353, 353-354.
- Fischer, D., Thomas, S. M., Suk, J. E., Sudre, B., Hess, A., Tjaden, N. B., ... Semenza, J. C., 2013. Climate change effects on Chikungunya transmission in Europe: geospatial analysis of vector's climatic suitability and virus' temperature requirements. *International Journal of Health Geographics*, 12, 51.
- Fourcade, Y., Engler, J. O., Rödder, D., & Secondi, J., 2014. Mapping species distributions with MAXENT Using a geographically biased sample of presence data: A performance assessment of methods for correcting sampling bias. *PLoS One*, 9, e97122.
- Fuller, T., Bensch, S., Müller, I., Novembre, J., Pérez-Tris, J., Ricklefs, R. E., ... Waldenström, J., 2012. The ecology of emerging infectious diseases in migratory birds: an assessment of the role of climate change and priorities for future Research. *Ecohealth*, 9, 80-88.
- Gelfand, A. E., & Shirota, S., 2019. Preferential sampling for presence/absence data and for fusion of presence/absence data with presence-only data. *Ecological Monographs*, 89, e01372.
- Gonzalez-Moreno, P., Diez, J. M., Richardson, D. M., & Vila, M., 2015. Beyond climate: disturbance niche shifts in invasive species. *Global Ecology and Biogeography*, 24, 360-370.

- Guerra, F. M., Bolotin, S., Lim, G., Heffernan, J., Deeks, S. L., Li, Y., ..., Crowcroft, N. S., 2017. The basic reproduction number ( $R_0$ ) of measles: a systematic review. *Lancet Infectious Diseases*, 17, e420-e428.
- Heffernan, J. M., Lou, Y. J., & Wu, J. H., 2014. Range expansion of *Ixodes scapularis* ticks and of *borrelia burgdorferi* by migratory birds. *Discrete and Continuous Dynamical Systems-Series B*, 19, 3147-3167.
- Heffernan, J. M., Smith, R. J., & Wahl, L. M., 2005. Perspectives on the basic reproductive ratio. *Journal of the Royal Society, Interface*, 2, 281-293.
- Hefley, T. J., Brost, B. M., & Hooten, M. B., 2017. Bias correction of bounded location errors in presence-only data. *Methods in Ecology and Evolution*, 8, 1566-1573.
- Hefley, T. J., & Hooten, M. B., 2016. Hierarchical species distribution models. *Current Landscape Ecology Reports*, 1, 87-97.
- Hethcote, H. W., 2000. The mathematics of infectious diseases. *SIAM Review*, 42, 599-653.
- Hines, J. E., Nichols, J. D., Royle, J. A., MacKenzie, D. I., Gopalaswamy, A. M., Kumar, N. S., ..., Karanth, K. U., 2010. Tigers on trails: occupancy modeling for cluster sampling. *Ecological Applications*, 20, 1456-1466.
- Holy, M., Schmidt, G., & Schröder, W., 2011. Potential malaria outbreak in Germany due to climate warming: risk modelling based on temperature measurements and regional climate models. *Environmental Science and Pollution Research*, 18, 428-435.
- Itzhaki, R. F., Golde, T. E., Heneka, M. T., & Readhead, B., 2020. Do infections have a role in the pathogenesis of Alzheimer disease? *Nature Reviews Neurology*, 16, 193-197.
- Jamison, A., Tuttle, E., Jensen, R., Bierly, G., & Gonsler, R., 2015. Spatial ecology, landscapes, and the geography of vector-borne disease: A multi-disciplinary review. *Applied Geography*, 63, 418-426.
- Jentes, E. S., Lash, R. R., Johansson, M. A., Sharp, T. M., Henry, R., Brady, O. J., ... Brunette, G. W., 2016. Evidence-based risk assessment and communication: a new global dengue-risk map for travellers and clinicians. *Journal of Travel Medicine*, 23, taw062.
- Johansson, M. A., Arana-Vizcarrondo, N., Biggerstaff, B. J., Gallagher, N., Marano, N., & Staples, J. E., 2012. Assessing the risk of international spread of yellow fever virus: a mathematical analysis of an urban outbreak in Asuncion, 2008. *American Journal of Tropical Medicine and Hygiene*, 86, 349-358.
- Johnson, E. E., Escobar, L. E., & Zambrana-Torrel, C., 2019. An ecological framework for modeling the geography of disease transmission. *Trends in Ecology & Evolution*, 34, 655-668.
- Johnson, N., & Mueller, J., 2002. Updating the accounts: global mortality of the 1918-1920 "Spanish" influenza pandemic. *Bulletin of the History of Medicine*, 76, 105-115.
- Jones, S. G., Ashby, A. J., Momin, S. R., & Naidoo, A., 2010. Spatial implications associated with using euclidean distance measurements and geographic centroid imputation in health care research. *Health Services Research*, 45, 316-327.
- Knudsen, E., Linden, A., Both, C., Jonzen, N., Pulido, F., Saino, N., ... Stenseth, N. C., 2011. Challenging claims in the study of migratory birds and climate change. *Biological Reviews*, 86, 928-946.
- Kramer-Schadt, S., Niedballa, J., Pilgrim, J. D., Schroder, B., Lindenborn, J., Reinfelder, V., ... Wilting, A., 2013. The importance of correcting for sampling bias in Maxent species distribution models. *Diversity and Distributions*, 19, 1366-1379.
- Krickeberg, K., Phạm, V. n. T. n., & Phạm, T. M. H. h., 2012. *Epidemiology: key to prevention*. New York: Springer.
- Laperriere, V., Brugger, K., & Rubel, F., 2011. Simulation of the seasonal cycles of bird, equine and human West Nile virus cases. *Preventive Veterinary Medicine*, 98, 99-110.

- Liu, Q. H., Ajelli, M., Aleta, A., Merler, S., Moreno, Y., & Vespignani, A., 2018. Measurability of the epidemic reproduction number in data-driven contact networks. *Proceedings of the National Academy of Sciences of the United States of America*, *115*, 12680-12685.
- Ma, Z., & Li, J. (Eds.). (2009). *Dynamical modeling and analysis of epidemics*. Singapore: World Scientific Publishing.
- MacKenzie, D. I., Nichols, J. D., Hines, J. E., Knutson, M. G., & Franklin, A. B., 2003. Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. *Ecology*, *84*, 2200-2207.
- MacKenzie, D. I., Nichols, J. D., Royle, J. A., Pollock, K. H., Bailey, L. L., & Hines, J. E., 2017. *Occupancy estimation and modeling: Inferring patterns and dynamics of species occurrence*: Academic Press.
- Mackenzie, J. S., Gubler, D. J., & Petersen, L. R., 2004. Emerging flaviviruses: the spread and resurgence of Japanese encephalitis, West Nile and dengue viruses. *Nature Medicine*, *10*, S98-S109.
- Mackey, T. K., Liang, B. A., Cuomo, R., Hafen, R., Brouwer, K. C., & Lee, D. E., 2014. Emerging and reemerging neglected tropical diseases: A review of key characteristics, risk factors, and the policy and innovation environment. *Clinical Microbiology Reviews*, *27*, 949-979.
- Mainali, K. P., Warren, D. L., Dhileepan, K., McConnachie, A., Strathie, L., Hassan, G., ... Parmesan, C., 2015. Projecting future expansion of invasive species: comparing and improving methodologies for species distribution modeling. *Global Change Biology*, *21*, 4464-4480.
- Marques, C. A., Forattini, O. P., & Massad, E., 1994. The basic reproduction number for dengue fever in Sao-Paulo state, Brazil - 1990-1991 epidemic. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *88*, 58-59.
- Matser, Y. A. H., Terpstra, M. L., Nadalin, S., Nossent, G. D., de Boer, J., van Bommel, B. C., ... Bemelman, F. J., 2018. Transmission of breast cancer by a single multiorgan donor to 4 transplant recipients. *American Journal of Transplantation*, *18*, 1810-1814.
- Mayo Clinic, last accessed 26, Nov, 2020. Liver cancer. <https://www.mayoclinic.org/diseases-conditions/liver-cancer/symptoms-causes/syc-20353659>.
- Medley, G. F., 2020. Herd immunity confusion. *The Lancet*, *396*, 1634-1635.
- Medlock, J. M., Hansford, K. M., Schaffner, F., Versteirt, V., Hendrickx, G., Zeller, H., ... Van Bortel, W., 2012. A review of the invasive mosquitoes in Europe: ecology, public health risks, and control options. *Vector borne and zoonotic diseases (Larchmont, N.Y.)*, *12*, 435-447.
- Miller, D. A., Nichols, J. D., McClintock, B. T., Grant, E. H. C., Bailey, L. L., & Weir, L. A., 2011. Improving occupancy estimation when two types of observational error occur: non-detection and species misidentification. *Ecology*, *92*, 1422-1428.
- Mills, J. N., Gage, K. L., & Khan, A. S., 2010. Potential influence of climate change on vector-borne and zoonotic diseases: A review and proposed research plan. *Environmental Health Perspectives*, *118*, 1507-1514.
- Mollot, G., Pantel, J. H., & Romanuk, T. N., 2017. The Effects of Invasive Species on the Decline in Species Richness: A Global Meta-Analysis. In D. A. Bohan, A. J. Dumbrell, & F. Massol (Eds.), *Networks of Invasion: A Synthesis of Concepts* (Vol. 56, pp. 61-84).
- Moon, S. A., Cohnstaedt, L. W., McVey, D. S., & Scoglio, C. M., 2019. A spatio-temporal individual-based network framework for West Nile virus in the USA: Spreading pattern of West Nile virus. *Plos Computational Biology*, *15*, e1006875.
- Mordecai, E. A., Cohen, J. M., Evans, M. V., Gudapati, P., Johnson, L. R., Lippi, C. A., ... Weikel, D. P., 2017. Detecting the impact of temperature on transmission of Zika, dengue, and chikungunya using mechanistic models. *Plos Neglected Tropical Diseases*, *11*, e0005568.
- Ng, V., Fazil, A., Gachon, P., Deuymes, G., Radojević, M., Mascarenhas, M., ... Ogden, N. H., 2017. Assessment of the probability of autochthonous transmission of chikungunya virus in

- Canada under recent and projected climate change. *Environmental Health Perspectives*, 125, 067001.
- Onder, G., Rezza, G., & Brusaferro, S., 2020. Case-Fatality Rate and Characteristics of Patients Dying in Relation to COVID-19 in Italy. *Jama-Journal of the American Medical Association*, 323, 1775-1776.
- Parham, P. E., & Michael, E., 2010. Modeling the effects of weather and climate change on malaria transmission. *Environmental Health Perspectives*, 118, 620-626.
- Park, D. S., & Davis, C. C., 2017. Implications and alternatives of assigning climate data to geographical centroids. *Journal of Biogeography*, 44, 2188-2198.
- Patterson, C. D., & Guerin, M. T., 2013. The effects of climate change on avian migratory patterns and the dispersal of commercial poultry diseases in Canada - Part II. *Worlds Poultry Science Journal*, 69, 163-181.
- Pearce, J. L., & Boyce, M. S., 2006. Modelling distribution and abundance with presence-only data. *Journal of Applied Ecology*, 43, 405-412.
- Peña García, V. H., & Christofferson, R. C., 2019. Correlation of the basic reproduction number ( $R_0$ ) and eco-environmental variables in Colombian municipalities with chikungunya outbreaks during 2014-2016. *Plos Neglected Tropical Diseases*, 13, e0007878.
- Perkin, T. A., Rodriguez-Barraquer, I., Manore, C., Siraj, A. S., Espana, G., Barker, C. M., ... Reiner, R. C., 2019. Heterogeneous local dynamics revealed by classification analysis of spatially disaggregated time series data. *Epidemics*, 29, 100357.
- Peterson, A. T., 2006. Ecologic niche modeling and spatial patterns of disease transmission. *Emerging Infectious Diseases*, 12, 1822-1826.
- Peterson, A. T., 2014. *Mapping disease transmission risk: enriching models using biogeography and ecology*. Baltimore: Johns Hopkins University Press.
- Peterson, A. T., & Soberón, J., 2012. Species distribution modeling and ecological niche modeling: Getting the concepts right. *Natureza & Conservacao*, 10, 102-107.
- Pettersson, J. H. O., Eldholm, V., Seligman, S. J., Lundkvist, A., Falconar, A. K., Gaunt, M. W., ... de Lamballerie, X., 2016. How did Zika virus emerge in the Pacific islands and Latin America? *Mbio*, 7, e01239-01216.
- Porta, M. (Ed.) (2014). *A dictionary of epidemiology* (6 ed.). Oxford: Oxford University Press.
- Regos, A., Imbeau, L., Desrochers, M., Leduc, A., Robert, M., Jobin, B., ... Drapeau, P., 2018. Hindcasting the impacts of land-use changes on bird communities with species distribution models of bird atlas data. *Ecological Applications*, 28, 1867-1883.
- Reisen, W. K., Fang, Y., & Martinez, V. M., 2006. Effects of temperature on the transmission of West Nile virus by *Culex tarsalis* (Diptera : Culicidae). *Journal of Medical Entomology*, 43, 309-317.
- Remuzzi, A., & Remuzzi, G., 2020. COVID-19 and Italy: what next? *Lancet*, 395, 1225-1228.
- Renner, I. W., Elith, J., Baddeley, A., Fithian, W., Hastie, T., Phillips, S. J., ... Warton, D. I., 2015. Point process models for presence-only analysis. *Methods in Ecology and Evolution*, 6, 366-379.
- Reyes, T. H., Esparza, E., Crestani, G., Limonchi, F., Cruz, R., Salinas, N., ... Cosio, E., 2020. Physiological responses of maca (*Lepidium meyenii* Walp.) plants to UV radiation in its high-altitude mountain ecosystem. *Scientific Reports*, 10, 2654.
- Ridenhour, B., Kowalik, J. M., & Shay, D. K., 2014. Unraveling  $R_0$ : Considerations for public health applications. *American Journal of Public Health*, 104, E32-E41.
- Ridenhour, B., Kowalik, J. M., & Shay, D. K., 2018. The basic reproductive number ( $R_0$ ): considerations for its application in public health. *American Journal of Public Health*, 108, S455-S465.

- Rizzoli, A., Jimenez-Clavero, M. A., Barzon, L., Cordioli, P., Figuerola, J., Koraka, P., ... Tenorio, A., 2015. The challenge of West Nile virus in Europe: knowledge gaps and research priorities. *Eurosurveillance*, 20, 28-42.
- Rizzoli, A., Tagliapietra, V., Cagnacci, F., Marini, G., Arnoldi, D., Rosso, F., ..., Rosa, R., 2019. Parasites and wildlife in a changing world: The vector-host- pathogen interaction as a learning case. *International Journal for Parasitology-Parasites and Wildlife*, 9, 394-401.
- Rota, C. T., Fletcher, R. J., Dorazio, R. M., & Betts, M. G., 2009. Occupancy estimation and the closure assumption. *Journal of Applied Ecology*, 46, 1173-1181.
- Royle, J. A., Nichols, J. D., & Kery, M., 2005. Modelling occurrence and abundance of species when detection is imperfect. *Oikos*, 110, 353-359.
- Rubel, F., Brugger, K., Hantel, M., Chvala-Mannsberger, S., Bakonyi, T., Weissenböck, H., ..., Nowotny, N., 2008. Explaining Usutu virus dynamics in Austria: Model development and calibration. *Preventive Veterinary Medicine*, 85, 166-186.
- Rushing, C. S., Royle, J. A., Ziolkowski, D. J., & Pardieck, K. L., 2020. Migratory behavior and winter geography drive differential range shifts of eastern birds in response to recent climate change. *Proceedings of the National Academy of Sciences of the United States of America*, 117, 12897-12903.
- Siettos, C. I., & Russo, L., 2013. Mathematical modeling of infectious disease dynamics. *Virulence*, 4, 295-306.
- Smith, K. F., & Guégan, J. F., 2010. Changing geographic distributions of human pathogens. In D. J. Futuyma, H. B. Shafer, & D. Simberloff (Eds.), *Annual Review of Ecology, Evolution, and Systematics*, Vol 41 (Vol. 41, pp. 231-250).
- Stolar, J., & Nielsen, S. E., 2015. Accounting for spatially biased sampling effort in presence-only species distribution modelling. *Diversity and Distributions*, 21, 595-608.
- Strickland, C., Dangelmayr, G., Shipman, P. D., Kumar, S., & Stohlgren, T. J., 2015. Network spread of invasive species and infectious diseases. *Ecological Modelling*, 309, 1-9.
- Tatem, A. J., Hay, S. I., & Rogers, D. J., 2006. Global traffic and disease vector dispersal. *Proceedings of the National Academy of Sciences of the United States of America*, 103, 6242-6247.
- Taubenberger, J. K., & Morens, D. M., 2006. 1918 influenza: the mother of all pandemics. *Emerging Infectious Diseases*, 12, 15-22.
- Thomas, S. M., Tjaden, N. B., van den Bos, S., & Beierkuhnlein, C., 2014. Implementing cargo movement into climate based risk assessment of vector-borne diseases. *International Journal of Environmental Research and Public Health*, 11, 3360-3374.
- Tjaden, N. B., Caminade, C., Beierkuhnlein, C., & Thomas, S. M., 2018. Mosquito-borne diseases: Advances in modelling climate-change impacts. *Trends in Parasitology*, 34, 227-245.
- van den Driessche, P., & Watmough, J., 2002. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Mathematical Biosciences*, 180, 29-48.
- van den Driessche, P., & Yakubu, A. A., 2018. Demographic population cycles and  $R_0$  in discrete-time epidemic models. *Journal of Biological Dynamics*, 12, 961-982.
- Varela, S., Anderson, R. P., Garcia-Valdes, R., & Fernandez-Gonzalez, F., 2014. Environmental filters reduce the effects of sampling bias and improve predictions of ecological niche models. *Ecography*, 37, 1084-1091.
- Villero, D., Pla, M., Camps, D., Ruiz-Olmo, J., & Brotons, L., 2017. Integrating species distribution modelling into decision-making to inform conservation actions. *Biodiversity and Conservation*, 26, 251-271.
- Wang, Y., & Stone, L., 2019. Understanding the connections between species distribution models for presence-background data. *Theoretical Ecology*, 12, 73-88.

- Warren, D. L., & Seifert, S. N., 2011. Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. *Ecological Applications*, 21, 335-342.
- Warton, D. I., & Shepherd, L. C., 2010. Poisson point process models solve the "pseudo-absence problem" for presence-only data in ecology. *Annals of Applied Statistics*, 4, 1383-1402.
- Watermann, L. Y., Hock, M., Blake, C., & Erfmeier, A., 2020. Plant invasion into high elevations implies adaptation to high UV-B environments: a multi-species experiment. *Biological Invasions*, 22, 1203-1218.
- Watson, R. T., Patz, J., Gubler, D. J., Parson, E. A., & Vincent, J. H., 2005. Environmental health implications of global climate change. *Journal of Environmental Monitoring*, 7, 834-843.
- Weissenböck, H., Kolodziejek, J., Url, A., Lussy, H., Rebel-Bauder, B., & Nowotny, N., 2002. Emergence of Usutu virus, an African mosquito-borne flavivirus of the Japanese encephalitis virus group, central Europe. *Emerging Infectious Diseases*, 8, 652-656.
- Westby, K. M., Sweetman, B. M., Van Horn, T. R., Biro, E. G., & Medley, K. A., 2019. Invasive species reduces parasite prevalence and neutralizes negative environmental effects on parasitism in a native mosquito. *Journal of Animal Ecology*, 88, 1215-1225.
- WHO, last accessed 26, Nov, 2020a. Guidelines for the prevention and control of cervical cancer. <https://www.who.int/reproductivehealth/topics/cancers/hpv-vaccination/en/>.
- WHO, last accessed 26, Nov, 2020b. Infectious disease.
- WHO, last accessed 26, Nov, 2020c. Plague. [http://www.who.int/vaccine\\_research/diseases/zoonotic/en/index4.html](http://www.who.int/vaccine_research/diseases/zoonotic/en/index4.html).
- WHO, last accessed 26, Nov, 2020d. Tuberculosis. <https://www.who.int/news-room/fact-sheets/detail/tuberculosis>.
- WHO, last accessed 26, Nov, 2020e. Vector-borne diseases. <https://www.who.int/news-room/fact-sheets/detail/vector-borne-diseases>.
- WHO, last accessed 26, Nov, 2020f. Vector control. <https://www.who.int/vector-control/en/>.
- Wikan, N., & Smith, D. R., 2016. Zika virus: History of a newly emerging arbovirus. *Lancet Infectious Diseases*, 16, E119-E126.
- Williams, M. C., Knight, E. M., Haddow, A. J., & Simpson, D. I. H., 1964. Isolation of West Nile virus from man and of Usutu virus from the bird-biting mosquito *Mansonia aurites* (Theobald) in Entebbe area of Uganda. *Annals of Tropical Medicine and Parasitology*, 58, 367-374.
- Wu, T., & Perrings, C., 2017. Conservation, development and the management of infectious disease: avian influenza in China, 2004-2012. *Philosophical Transactions of the Royal Society B-Biological Sciences*, 372, 20160126.
- Wu, Y. C., Chen, C. S., & Chan, Y. J., 2020. The outbreak of COVID-19: An overview. *Journal of the Chinese Medical Association*, 83, 217-220.
- Yang, H. M., 2014. The basic reproduction number obtained from Jacobian and next generation matrices - A case study of dengue transmission modelling. *Biosystems*, 126, 52-75.
- Zhang, Q., Sun, K. Y., Chinazzi, M., Piontti, A. P. Y., Dean, N. E., Rojas, D. P., ... Vespignani, A., 2017. Spread of Zika virus in the Americas. *Proceedings of the National Academy of Sciences of the United States of America*, 114, E4334-E4343.
- Ziegler, U., Jöst, H., Müller, K., Fischer, D., Rinder, M., Tietze, D. T., ... Groschup, M. H., 2015. Epidemic spread of Usutu virus in southwest Germany in 2011 to 2013 and monitoring of wild birds for Usutu and West Nile viruses. *Vector-Borne and Zoonotic Diseases*, 15, 481-488.
- Zinszer, K., Verma, A. D., Charland, K., Brewer, T. F., Brownstein, J. S., Sun, Z. Y., ..., Buckeridge, D. L., 2012. A scoping review of malaria forecasting: past work and future directions. *Bmj Open*, 2.

# Manuscript 1

## Centroid data in Ecological Niche Modelling: Effects on Model Performance in Context with Grain Size

In press in *Global Ecology and Biogeography*, (2020), DOI: 10.1111/geb.13240

Yanchao Cheng<sup>1</sup>, Nils Benjamin Tjaden<sup>1</sup>, Anja Jaeschke<sup>1</sup>, Stephanie Margarete Thomas<sup>1,2</sup>, Carl Beierkuhnlein<sup>1,2</sup>

1 Department of Biogeography, University of Bayreuth, Universitätsstr. 30, 95447, Bayreuth, Germany

2 BayCEER, Bayreuth Center for Ecology and Environmental Research, Bayreuth, Germany

**Keywords:** Administrative Spatial Unit, centroid, county, ecological niche model, grain size, Maxent, spatial heterogeneity, species distribution model, virtual species

## **Abstract**

**Aim** Ecological Niche Models (ENM) typically require point locations of species' occurrence as input data. Where exact locations are not available, geographical centroids of respective Administrative Spatial Units (ASU) are often used as a substitute. We investigated how the use of ASU centroids in ENMs affects model performance, which role the size of ASUs plays, and which effects different grain sizes of explanatory variables have.

## **Location**

Europe.

## **Major taxa studied**

Virtual species.

## **Methods**

We set up a two-factorial study design with artificial ASUs of three different sizes and environmental data of four commonly used grain sizes, repeated over 3 study regions. To control other factors that may affect ENM performance, we created a virtual species with a known response to environmental variables, precise and even sampling and a known spatial distribution. We ran a series of Maxent models for the virtual species based on centroids and precise occurrence locations under varying ASU- and grain sizes.

## **Results**

The use of ASU centroids introduces a value frequency mismatch of the explanatory variables between centroids and true occurrence locations, and it has a negative effect on ENM performance. Value frequency mismatch, negative effect on ENM-performance and over-prediction of the species' range all increase with ASU size. The effect of grain size of environmental data, on the other hand, was small in comparison.

## **Main conclusions**

ENMs built upon ASU centroids can suffer considerably from the introduced error. For ASUs that are sufficiently small or show low spatial heterogeneity of explanatory variables, ASU centroids can still be a viable and convenient surrogate for precise occurrence locations. When possible, however, central tendency values (median, mean) that represent the whole ASU rather than just a single point location need to be considered.

## 1 Introduction

Ecological niche models (ENMs), based on niche theory, are widely used in many fields such as invasion and conservation ecology, biogeography, as well as epidemiology (Elith & Leathwick, 2009; Escobar & Craft, 2016; Liu et al., 2018; Peterson, 2014). They are often employed to estimate the spatial distribution of certain species (Elith et al., 2006; Elith & Leathwick, 2009; Elith et al., 2011) or diseases (Tjaden et al., 2018), and thus also known as “species distribution models”. ENMs typically use geographical occurrence locations of the target species as input data. These locations are then related to a series of explanatory variables (spatial raster data describing the environmental and/or socio-economic conditions in the study area), forming a correlative model of the species’ environmental niche. This model can be projected onto regions where the presence-absence state of the species is unknown, resulting in a map showing probability of presence or environmental suitability.

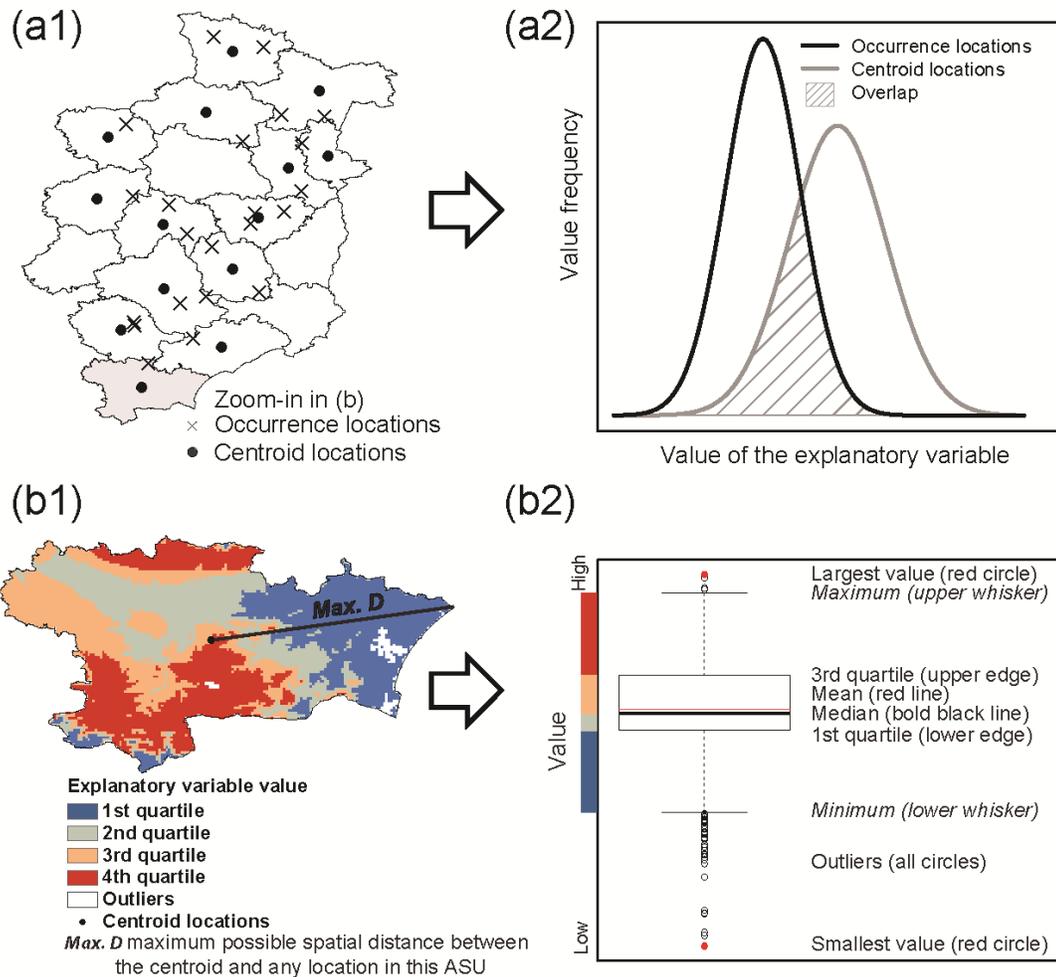
By default, ENMs assume that the whole study area was sampled consistently with precisely recorded geographic locations of occurrence, and that the selected explanatory variables can represent the species well (Yackulic et al., 2013). In practice, however, sampling bias is inevitable, especially when the input data has to be assembled from different sources that are based on different sampling methods (Liu et al., 2018; Lobo & Tognelli, 2011; Qiao et al., 2017; Stolar & Nielsen, 2015). While the sampling bias caused by *uneven* sampling can be reduced by rarifying or filtering the occurrence records (Castellanos et al., 2019; Gábor et al., 2020; Kramer-Schadt et al., 2013), *imprecisely* recorded occurrence locations are very difficult (although not entirely impossible) to correct (Hefley et al., 2017). In certain cases, e.g. when using citizen-science databases or local monitoring systems, the occurrence locations of the species may be of a coarse precision or only available at municipal or county level (i.e. related to geographical surfaces of differing sizes). For epidemiological data, such missing spatial precision ensures information privacy.

Internet databases like the Global Biodiversity Information Facility (GBIF; [www.gbif.org](http://www.gbif.org)), gather and compile species occurrence data from different sources (scientific, governmental, citizen-science) across a tremendous geographical extent and across national boundaries. However, the precision of the occurrence locations in this kind of databases is not always satisfying (Liu et al., 2018), and the record precision differs considerably depending on how the occurrence records were collected and processed (in certain cases, the precision might even be unknown) (Collins et al., 2017). In other databases, only very coarse administrative level information is available, i.e. instead of geographical locations, the occurrence records are assigned to counties or postal regions. For instance, the European Centre for Disease Prevention and Control (ECDC) and the European Food Safety Authority (EFSA) maintain a joint collection of occurrence records of epidemiologically relevant mosquito, tick and sand fly species in their VectorNet database. This highly relevant database covers the entire European Union and adjoining countries, but only maps showing local administrative units are publically available (<https://www.ecdc.europa.eu/en/disease->

vectors/surveillance-and-disease-data). Similarly, occurrences of species are often reported as inventories of protected areas that can differ considerably in size. For the sake of simplicity, hereafter we will refer to all kinds of administrative areas as “administrative spatial unit” (ASU).

When occurrence records are available at the level of entire ASUs only, the geographical centroids of the ASU are often used in ENMs as a substitute for precise point locations (Collins et al., 2017; Park & Davis, 2017). As mentioned above, the ENMs allocate explanatory variables’ values at the respective geographical occurrence locations and form a correlative model of the species’ environmental niche. The use of centroid locations introduces geographical distance between the true (but unknown) occurrence locations and the geographical centroids representing them. This induces a mismatch in the values of explanatory variables (Figure 1 (a)): it is very unlikely (although not impossible) for the true geographical location of the observed record to exactly match the environmental conditions at the centroid location (Figure 1 (b)). This means that between each pair of true location and geographical centroid, there is likely a mismatch in values of explanatory variables. It can consequently be expected that substituting geographical centroids for true occurrence locations also leads to a change in the over-all frequency of values of explanatory variables. The correlative model, built with the shifted values, will further lead to a biased prediction for the species’ distribution, and would probably lead to overprediction.

While finding a substitute for a geographically unknown occurrence location, drawing the geographical centroid of the ASU minimizes the largest possible spatial distance between the substitute location and the unknown true location (Figure 1 (b1)). However, this does not necessarily minimize the difference in environmental conditions (i.e., values of explanatory variables) at the two locations. In fact, it is entirely possible that among all possible locations within an ASU, the environmental conditions at the ASU centroid are the worst possible substitute for the conditions at the true location – especially in areas where spatial heterogeneity is high. Approaching the substitute from another angle (Figure 1 (b2)), using a central tendency value (median, mean) of each explanatory variable across the entire ASU has been presented as a better option (Park & Davis, 2017). Instead of minimizing the largest possible spatial distance, central tendency values minimize the largest value mismatch directly. The boxplot in Figure 1 (b2) illustrates this on the basis of the median: When using the median as the substitute, very likely (97% in this example) the largest potential value mismatch is half of the range between the two bars. Possibly (50%) the value of the occurrence location falls in the box. In this case, the largest potential value mismatch is even smaller. It is obvious that using central tendency values reduces the possibility of introducing extreme values. Not surprisingly, it has been shown that the central tendency values outperform the variable values at the centroids (Collins et al., 2017; Park & Davis, 2017). Despite this, however, ASU centroids are still widely being used (e.g., (Evans et al., 2010; Fois et al., 2018; Gao & Cao, 2019; Johnson et al., 2017; Quiner & Nakazawa, 2017). It is thus worth to further investigate if and under what circumstances the much simpler approach of using ASU centroids can lead to satisfying results.



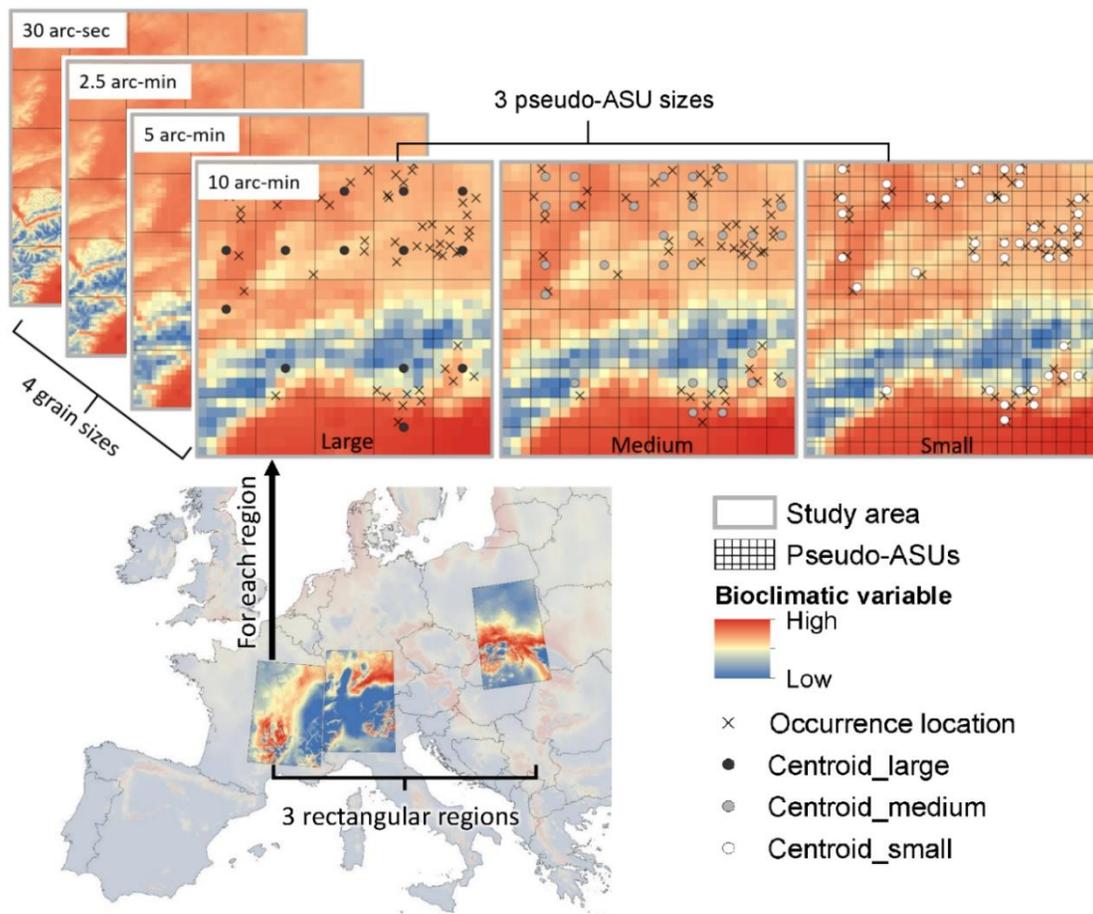
**Figure 1.** (a) Value frequency mismatch of an explanatory variable resulting from the use of ASU (administrational spatial unit) centroids. (a1) is a group of ASU (here: counties) with true occurrence locations and respective centroid locations. Note that for each ASU only one centroid location will be kept, as there exists only one. (a2) is the value frequency curve mismatch between these two groups of locations, concerning an explanatory variable. (a1) and (a2) illustrate our hypotheses on how geographical distance between occurrence locations and ASU centroids leads to value frequency mismatch. (b) Zooming in to each ASU, for a single pair of occurrence location and centroid location. (b1) shows that using ASU centroids minimizes the largest potential spatial distance (thick black line) between the centroid location (black dot) and any possible unknown occurrence location in the given ASU. However, this does not mean that the difference in values between those two points is minimized. (b2) shows the variation of values of an explanatory variable across all the grid cells within the ASU. More than 97% of values fall into the range between the whiskers, and 50% of the values fall into the rectangular box.

Here, we investigate how the application of geographical centroids affects the ENM results, a factor that has not received much attention so far. Of course, ENMs are also affected by a series of other factors. This includes the selection of explanatory variables, the specific modelling algorithm, model

settings and the spatial resolution or pixel size of explanatory variables, from here on referred to as “grain size” (Connor et al., 2018; Fourcade et al., 2018; García-Callejas & Araújo, 2016; Moudrý & Simova, 2012; Nezer et al., 2017; Record et al., 2018; Warren & Seifert, 2011; Yates et al., 2018). To best control these factors, we generated a virtual species with true, known occurrence locations and a known response to a fixed set of variables. Similar to the use of ASU centroids, the choice of grain size of explanatory variables may also cause a mismatch of explanatory variable value. For a given location, the explanatory variable value may differ at different grain sizes, though the difference is in general small due to spatial autocorrelation. Hence, we included a series of commonly used grain sizes of explanatory variables. Focusing on the bias resulting from the use of ASU centroids, we hypothesize that: 1) Using ASU centroids as substitutes for true occurrence locations leads to a value frequency mismatch of explanatory variables between the true locations and the centroids. An increased spatial heterogeneity within the ASU elevates that mismatch, assuming that larger ASUs tend to have higher spatial heterogeneity. 2) When using ASU centroids, the ENM’s performance decreases with increasing ASU size. 3) The size of the ASUs affects the model performance more than the grain size of explanatory variables. 4) The use of ASU centroids leads to an over-estimation of the modelled species’ distributional range.

## 2 Material and methods

To investigate how much ASU size and grain size affect ENM performance, a two factorial design with three replicates was applied (Figure 2). For this, a virtual species was generated based on three explanatory variables across Europe (see below for details). According to the presence-absence map of this virtual species, three squared regions (sized  $5^\circ \times 5^\circ$ ) were selected within which the virtual species occupancy was about 50%. For each squared region, pseudo-ASUs of three different sizes were constructed by dividing it evenly into 25 (large), 100 (medium), and 400 (small) squares (Figure 2). The size range of these pseudo-ASUs corresponds to those of low-level administrative units across Europe. There, the Nomenclature of Territorial Units for Statistics (NUTS, <https://ec.europa.eu/eurostat/web/nuts/background>) consist of three levels (NUTS 1–3). NUTS 3 is for small regions with a population size threshold of 150 000 – 800 000. The average area of the NUTS 3 units is about 7000 km<sup>2</sup>. The large pseudo-ASU size applied in this study is about 8000 km<sup>2</sup>, which can be treated as an equivalent of the average NUTS3 administrative units across Europe. The medium pseudo-ASU size is about 2000 km<sup>2</sup>, and the small pseudo-ASU size is about 500 km<sup>2</sup>.



**Figure 2.** The two-factorial study design with three replicates. Three rectangular regions (with varying grain size and pseudo-ASU size) were used to assess the general trend of bias resulting from the use of ASU centroids together with varying grain size. For each region, 200 random locations were drawn to keep the sampling effort even, and only presence locations among these locations were kept (black cross). The whole setup was repeated with the three bioclimatic variables, which the virtual species was generated with (see methods).

The hypotheses were first tested with these artificial pseudo-ASUs from the three rectangular regions (with three pseudo-ASU sizes) (Figure 2). Afterwards, data from real countries (Germany and France) with irregular ASU size and shape was used to confirm the previous results in a real-life environment (Appendix S1, Fig. S1.1). The varying pseudo-ASU sizes for the regions were applied to detect the general trend of centroid-arisen bias. France and Germany were chosen as test cases as they are of very different NUTS 3 ASU sizes. For France, the average area of NUTS 3 ASUs is about 6000 km<sup>2</sup>. For Germany, it is 1200 km<sup>2</sup>. As the NUTS 3 ASU size of Germany is much smaller than that of France, we expected the ENM models based on Germany's NUTS 3 centroids to outperform the ones based on French NUTS 3 centroids.

For the rectangular regions as well as France and Germany, a series of commonly used grain sizes was taken into consideration (four grain sizes 0.5, 2.5, 5, 10 arc-min, roughly equivalent to 0.5, 10, 40 and 200 km<sup>2</sup> respectively), because the grain size (raster resolution) of explanatory variables in ENMs also affects model performance (Connor et al., 2018; Guisan et al., 2007; Lauzeral et al., 2013; Manzoor et al., 2018). As both factors affect models on similar spatial scales, it is necessary to view the effects of ASU size and grain size in context with each other.

## **2.1 Virtual species**

A virtual species was generated using the “virtualspecies” package version 1.4.2 (Leroy et al., 2016) in R 3.4.2 (R Core Team, 2015), with a spatial resolution of 0.5 arc-min (details on explanatory variables see section “Explanatory variables”). Virtual species generation can be understood as defining a niche of a virtual species by limiting the determining environmental variables, setting the response to the variables, and setting prevalence or tolerance levels. By applying a virtual species, 1) the exact explanatory variables are known, 2) the occurrence locations are precise, 3) the whole study area is evenly sampled without sampling bias, 4) the true spatial distribution probability and presence-absence maps are available. These advantages make virtual species an ideal tool for testing our hypotheses.

To generate a virtual species, a certain amount of explanatory variables is needed, as well as parameters such as the response of the virtual species to each variable. Depending on the parameters and the presence-absence conversion method applied, different species distribution patterns can be achieved. The spatial distribution of the virtual species, both logistic distribution map and presence-absence map, can be exported as raster files for further use. A dataset of presence-absence or presence-only locations can be generated in order to simulate real-world sampling of occurrence records in the field. In this study, 200 random locations were drawn for each rectangular region to simulate sampling locations of an unbiased field campaign. By allocating the same amount of sampling locations, the simulated sampling effort for each region is the same, thus the model performance is comparable across the regions. Locations where the species was recorded as “absent” were discarded and the remaining presence locations (ca. 60 per region) were used to build the ENMs. For more details about the virtual species in this study see Appendix S2.

## **2.2 Explanatory variables**

To keep the virtual species simple, only bioclimatic variables were taken into account. For this, the standard set of 19 bioclimatic variables was acquired from [www.worldclim.org](http://www.worldclim.org) (Fick & Hijmans, 2017), with grain sizes of 0.5, 2.5, 5 and 10 arc-min. Three bioclimatic variables were selected according to the following criteria: 1) the set must include both hydrological and thermal factors, which are essential to most life forms; 2) the variables should not be closely related to each other (De Marco & Nóbrega, 2018) (i.e.  $|\text{Pearson's } r| > 0.7$  (see Appendix S3, Table S3.1), calculated with the European extent (Appendix S2, Fig. S2.2) of the virtual species,). As a consequence, three

bioclimatic variables, namely Annual Mean Temperature (Bio1), Annual Precipitation (Bio12), and Precipitation Seasonality (Bio15), were chosen (details see Appendix S2, Fig. S2.3 and S2.4).

### **2.3 Value frequency mismatch in explanatory variables due to the use of ASU centroids**

To visualize the value frequency mismatch, the three explanatory variables' values were extracted at the centroid locations of the 3 different pseudo-ASU sizes separately per region. This was repeated for the 4 different grain sizes (Figure 2). The value frequencies of the explanatory variables were then described through a kernel density curve (which can be understood as response curve of the virtual species to the variable) using the R package "caTools" version 1.17.1.2 (Jarek Tuszynski, 2014). Relative overlap of the curve for the centroid locations with the corresponding curve for the original occurrence records was calculated using "caTools". Mismatch between those curves was then calculated as  $1 - \text{overlap}$ , so that a mismatch of 0 means identical curves and 1 means no overlap at all.

The spatial heterogeneity of each explanatory variable was assessed by calculating its standard deviation within the respective pseudo-ASU and for the respective grain size, using the "raster" package version 2.6.7 (Hijmans, 2019) in R. The spatial heterogeneity was expected to increase with the size of pseudo-ASUs. For each explanatory variable, a linear regression was applied to describe the correlation between frequency curve mismatch and spatial heterogeneity.

### **2.4 Ecological niche model**

Maxent, an ENM algorithm widely used and known for its good performance with small occurrence location datasets (Baldwin, 2009; Elith et al., 2006; Hernandez et al., 2006), was chosen in this study. To make the models comparable, the settings were kept the same for all runs, for the three rectangular patches, Germany and France. Default model settings were applied (with 10000 background locations), with 10 replicates. Instead of commonly used methods such as the true skill statistic (TSS) or the area under the curve of receiver operating characteristics (AUC), model performance was assessed using Spearman's rank correlation coefficient (Spearman's rho). Spearman's rho is obviously a better choice than AUC, and compared to TSS, Spearman's rho has the advantage of being threshold-independent. As the true spatial distribution probability map for the virtual species is available, Spearman's rho between the environmental suitability predicted by the model and the true probability of presence can easily be achieved (via R package "pspearman" (Savicky, 2014)). In this case, Spearman's rho can range from 0 (no correlation) to 1 (perfect linear positive correlation, the compared models are identical).

To calculate Spearman's rho, the larger grain-sized (2.5, 5 and 10 arc-min) outputs from Maxent models were resampled to 0.5 arc-min resolution using the "nearest neighbor" method (R package "raster" (Hijmans, 2019)). Essentially, this means cutting the large raster cells into smaller ones, while keeping the original values without interpolation or loss of information. The model results were then compared with the true (distribution) probability map of the virtual species (the virtual species was generated with 0.5 arc-min resolution, more details see above and Appendix S2).

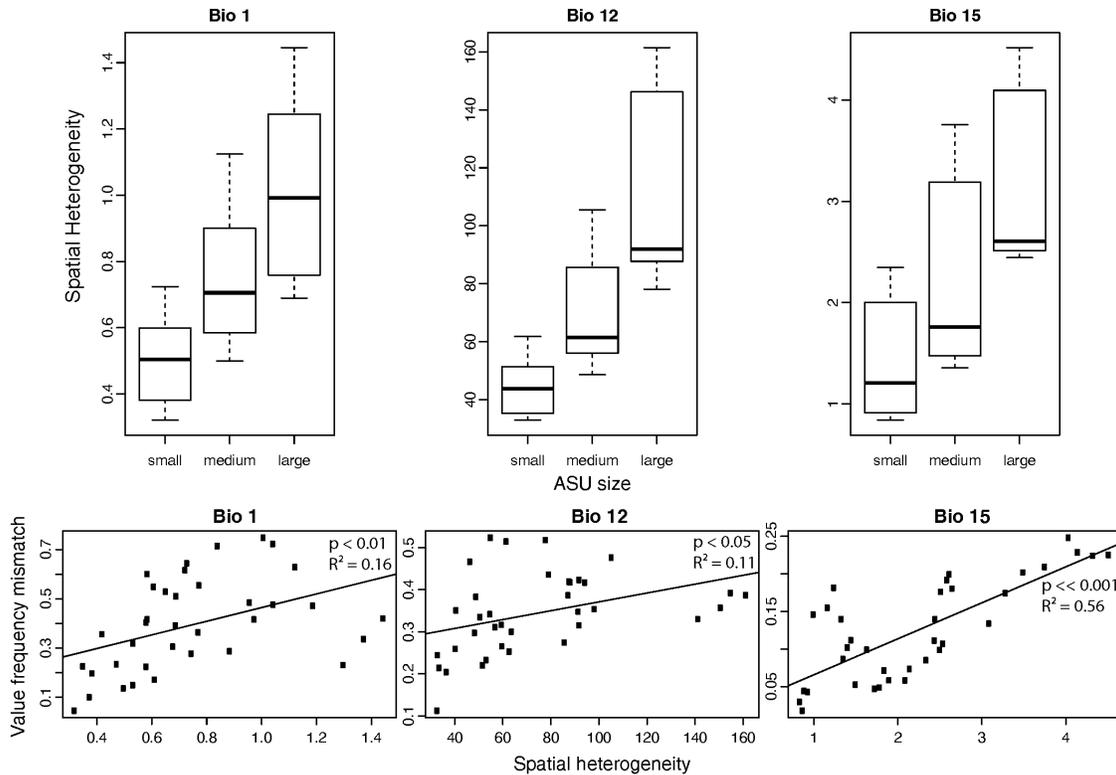
## 2.5 Calculation of over-prediction ratio

The model results were transformed into binary presence-absence maps according to the thresholds: Maximum training sensitivity plus specificity logistic threshold (MaxSSS) (Liu et al., 2005; Liu et al., 2016), Equal training sensitivity and specificity logistic threshold (eqSS) (Liu et al., 2005; Nenzen & Araujo, 2011) and 10 percentile training presence logistic threshold (10 percentile) (Pearson et al., 2004). Over-prediction was then calculated as the ratio of raster cells classified as “presence” in the model output vs. the original virtual species (i.e.:  $(Presence_{modeled} - Presence_{original}) / Presence_{original}$ ). Here, a value of 0 suggests that the distributional range predicted by the model has the same size as the one defined in the virtual species. Values larger or smaller than 0 mean that the predicted range is larger (over-prediction) or smaller (under-prediction) than that of the original species, respectively.

## 3 Results

### 3.1 The larger the pseudo-ASU size, the larger the value frequency mismatch of the explanatory variables

Using ASU centroids resulted in a mismatch of value frequency curves of the explanatory variables. This mismatch increases with the spatial heterogeneity within the respective ASU. A statistically significant positive relationship between variable mismatch and spatial heterogeneity was revealed through linear regression analysis for all three bioclimatic variables (Figure 3, top; Bio1:  $p < 0.01$ ,  $R^2 = 0.16$ ; Bio12:  $p < 0.05$ ,  $R^2 = 0.11$ ; Bio15:  $p < 0.001$ ,  $R^2 = 0.56$ ). The over-all spatial heterogeneity of an ASU increases with its size (Figure 3, bottom).

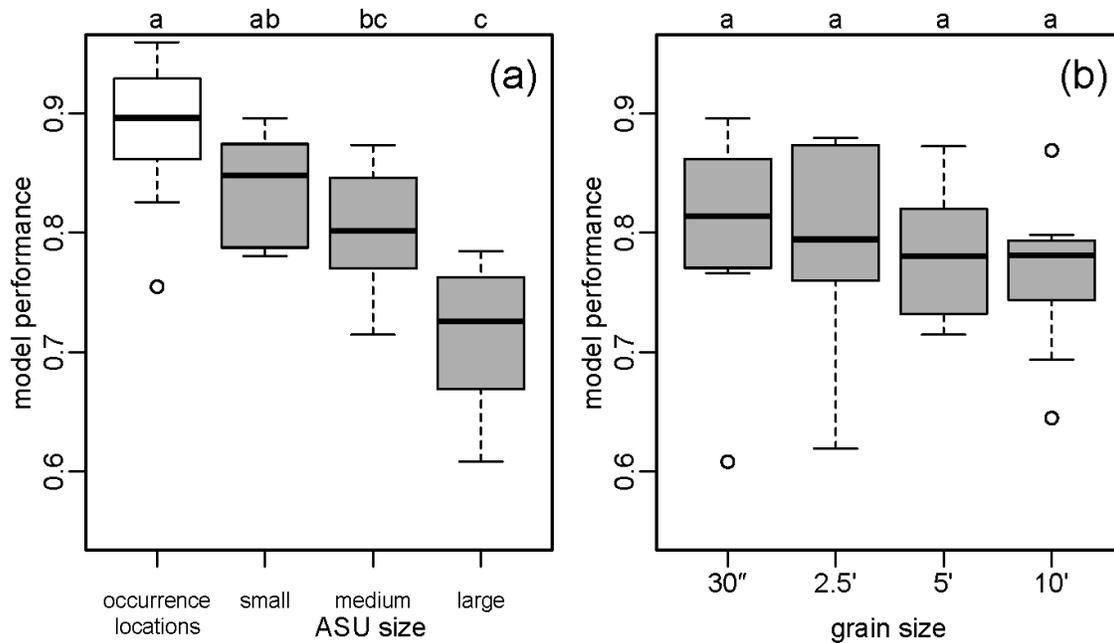


**Figure 3.** Top: The spatial heterogeneity of each explanatory variable (Bio 1, Bio 12, Bio 15) increases with increasing centroid region size. Bottom: Value frequency curve mismatch for explanatory variables (Bio 1 Annual Mean Temperature, Bio 12 Annual Precipitation, Bio 15 Precipitation Seasonality) at occurrence locations versus centroid locations increases with elevated spatial heterogeneity. Each black dot represents one pair of comparison on the recorded variable values between centroid locations and real locations. For each variable, the spatial heterogeneity is measured by the standard deviation of that variable within the individual centroid regions.

### 3.2 For ENM-performance, ASU size matters more than the grain size

Ecological niche models built with the original occurrence locations showed strong correlations of predicted environmental suitability with the true distribution of the virtual species, suggesting good model performance (Figure 4 (a)). For these ENMs, the model performance decreased with increasing grain size (see Appendix S4, Fig. S4.5). For those ENMs built with centroids, a Kruskal-Wallis Rank Sum Test with multiple comparison post-hoc test revealed statistically significant ( $p < 0.05$ ) effects of ASU size on model performance. There is a clear trend of model performance decreasing with increasing ASU size (Figure 4 (a)). No clear model performance pattern was observed for the different grain sizes (Figure 4 (b)). A direct comparison using two-way ANOVA (see Appendix S4, Table S4.2; only include the ENMs built with centroids, i.e. the three gray boxes in Figure 4 (a)) reveals that while ASU size can explain more than half (52%) of the variability in model performance ( $f(2) = 16.414$ ,  $p << 0.001$ ); grain size appears to have almost no effect ( $f(3) = 0.876$ ,

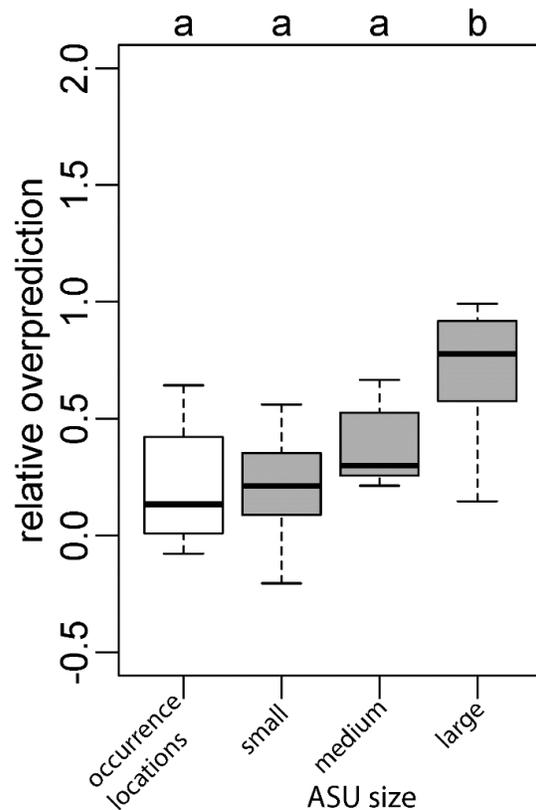
$p = 0.467$ ); the interaction between grain size and ASU size shows no significance ( $f(6) = 0.561$ ,  $p = 0.757$ ), either.



**Figure 4.** ASU size affects model performance more than grain size. (a): Grey: performance of the ENM in dependence of ASU size. White, for reference: ENM performance when using true occurrence locations. Lower case letters above the boxes indicate differences between groups as indicated by a Kruskal-Wallis Rank Sum Test with multiple comparison post-hoc test. (b): Performance of the ENM in dependence of grain size (i.e. spatial resolution of environmental data). Grey boxes in (a) and (b) refer to the same set of models. Model performance was assessed through the correlation coefficient (Spearman’s rho) between the environmental suitability predicted by the ecological niche model and the true species distribution defined for the virtual species. Model performance ranges from 0 to 1. The larger the value, the better the model performance. Lower case letters above the boxes indicate that an ANOVA followed by a Tukey Honest Significant Differences post-hoc test revealed no statistically significant differences between any of the groups.

### 3.3 Over-prediction of species spatial distribution due to the use of centroid data

Almost all ENM runs in this study, including those performed with true occurrence locations, over-predicted the virtual species’ occurrence. Based on the MaxSSS threshold, over-prediction tends to be stronger with increasing ASU size (Figure 5). However, this increase is statistically significant only for the large ASUs ( $p < 0.01$  based on ANOVA followed by a Tukey Honest Significant Differences post-hoc test). This is consistent with the results obtained from the eqSS and 10 percentile thresholds (Appendix S4, Fig. S4.6).



**Figure 5.** Relative over-prediction ratio (ENM model results vs. “true” virtual species occurrence) for models built on point locations as well as centroids of differently sized regions. Lower case letters above the boxes indicate differences between groups as indicated by an ANOVA followed by a Tukey Honest Significant Differences post-hoc test.

### 3.4 Real-world application example using French and German ASU centroids

The ENM built with centroids of NUTS 3 administrative units in Germany outperforms the model for France, with Spearman’s rho value of 0.818 and 0.790, respectively. For the ENM built with true occurrence locations, Spearman’s rho for Germany and France is 0.894 and 0.924, respectively (Table 1). When occurrence locations are available, the model performance decreases with increasing grain size. However, when only centroid locations are available, fine grain size was not always the best. For France, the ENM with 2.5 arc-min had the best model performance (Table 1). This is in accordance with the pattern shown in Figure 4: ASU size matters more than grain size.

**Table 1.** Model performance of the real-world examples using French and German ASUs. Performance of the models was assessed by calculating Spearman’s rho for the correlation of the predicted probability of presence with the known true probability of presence of the virtual species. Bold: best-performing models for centroid- and true location-based models for France and Germany, respectively.

country	Grain size	Model performance (centroids)	Model performance (true occurrence locations)
France	10 arc-min	0.725	0.852
France	5 arc-min	0.798	0.927
France	2.5 arc-min	<b>0.858</b>	0.942
France	30 arc-sec	0.779	<b>0.974</b>
Germany	10 arc-min	0.807	0.837
Germany	5 arc-min	0.812	0.897
Germany	2.5 arc-min	0.826	0.914
Germany	30 arc-sec	<b>0.828</b>	<b>0.927</b>

#### 4 Discussion

In this study, we looked into the mechanism of how the use of centroids affects ENM performance. Though there have been studies focusing solely on centroid size (Collins et al., 2017; Park & Davis, 2017) or grain size (Connor et al., 2018; Lauzeral et al., 2013; Manzoor et al., 2018) respectively, we investigated and compared how much ASU size affects ENM performance not only on its own but also in context with grain size. Our results confirm that, in general, larger ASUs have higher spatial heterogeneity, and higher spatial heterogeneity is associated with higher value frequency mismatch of explanatory variables between the true locations and centroids. When using ASU centroids, larger ASUs also lead to larger decrease of ENM’s performance. Compared with ASU size, grain size does not affect ENM’s performance as much. The use of ASU centroids leads to overprediction of modelled species’ distribution range, and the overprediction ratio shows the tendency of increasing with ASU size.

Using centroids of ASUs as a substitute for true occurrence records in ENMs introduces errors. Spatial distance between the centroids and the true, unknown occurrence locations leads to a mismatch of explanatory variables’ values at these locations. These mismatched values at the centroid locations lead to a mismatch of explanatory variables’ value frequency curves, which further results in a mismatch between the projected niche and the true niche. Our results show that the absolute size of ASUs affects the value frequency mismatch between true locations and centroids. How strong this effect is, ultimately depends on the explanatory variables’ spatial heterogeneity (here: standard deviation) within the ASUs. Park & Davis (2017) found that spatial heterogeneity in climatic variables was mainly governed by the heterogeneity in topography in the US. Their findings that ASU (county) size only had minimal effects on spatial heterogeneity does not contradict our results, due to their different, non-nested study design. The absolute ASU size or grain size alone cannot determine how much the explanatory variables’ values mismatch with the

values at the true occurrence locations, but in general, larger ASU size leads to larger value mismatch (Figure 3).

Similarly, coarser grain size leads to a deterioration in model performance (Appendix S4, Fig. S4.5). This is in line with previous findings (Connor et al., 2018; Guisan et al., 2007; Manzoor et al., 2018). However, although the grain size does affect model performance, its effect was found to be small compared that of the ASU size (Figure 4 and Appendix S4, Table S4.2). When centroids are drawn from ASUs with large extent, the ENM's performance can hardly be improved by using a fine grain size (Figure 4). However, when using centroids drawn from a small extent or using the true occurrence locations, a fine grain size is preferable (Table 1). Note that the "small" pseudo-ASU size used in this study is  $0.25^\circ \times 0.25^\circ$  (ca. 400 km<sup>2</sup>). This is not much larger than the coarsest grain size (ca. 170 km<sup>2</sup>) in this study, which corresponds to the resolution of some commonly used environmental data sets (e.g., E-OBS (Cornes et al., 2018)). For ASUs of this size, the value mismatch between the centroids and true locations is very small, and the effect of ASU size on the value mismatch cannot be distinguished from that introduced by a large grain size.

The use of a virtual species in this study means that the environmental suitability and the presence-absence status of the species across the study region are known, and the species' occurrence records are precise. This makes the comparison between models based on true locations and models based on ASU centroids feasible and reliable. The difference between models based on observed point locations from field data versus centroid-based locations has previously been quantified (Collins et al., 2017). However, using observed locations, additional effects resulting from sampling bias or uncertainty cannot be excluded. Generating a virtual species and drawing precise geographical locations ensures that the observed differences between model results are due to the use of centroid locations itself and that they can be quantified. As the true environmental suitability of the virtual species is known, it can be used as a benchmark for the performance of the models by directly calculating Spearman's rho. This obviates the use of threshold-based performance measures such as TSS or the commonly used but controversial AUC (Allouche et al., 2006; Tjaden et al., 2018).

The value frequency mismatch was for the first time applied as an indicator of potential projected niche mismatch. This method calculates, for each explanatory variable, the value frequency mismatch between ASU centroids and occurrence locations. Compared with conventional statistical tests (such as ANOVA or Kruskal-Wallis-test) that show whether a statistical difference between centroid-based and true location-based environmental data exists or not, this method focuses on quantifying the differences between the two. As ENMs typically process variables' values in a continuous way, (for a given variable value at a single location, the ENMs calculate a probability of presence rather than a binary presence-absence value), the true/false information alone is clearly not sufficient to assess the effect of using ASU centroids. While test statistics fail to measure how much two groups of data differ, the continuous method makes it possible to capture the general trend of the value mismatch and compare it across groups. For instance, our results suggest that a larger ASU size leads to larger value frequency mismatch. If a binary-result-only method was applied

in this study, this general trend could not be captured, nor evaluated. It should be noted that the value frequency mismatch ranges from 0 to 100 %, i.e. though the value frequency mismatch increases with spatial heterogeneity within the ASU, this increase is not always linear. Here a linear model was applied to show the general trend, but it should not be interpreted as an indication for how much the mismatch will be when larger spatial heterogeneity occurs.

As the virtual species was generated with the grain size of 0.5 arc-min, it is to be expected that the output from the combination of this grain size and original locations has the best performance. While generating the virtual species, we assumed that the grain size employed is the smallest unit that the virtual species can survive, as an individual (similar to e.g. a deer, a bird) or a population (e.g. ants, bees). However, modelling with real species, it should be questioned which grain size should be utilized. It has been suggested to use a grain size smaller than 1 km when possible, especially for habitat specialists (Manzoor et al., 2018). Nevertheless, the question of choosing optimal grain size needs to be further investigated.

It is not surprising that centroids of ASUs are chosen as a substitute when no precise occurrence locations are available. After all, the changes in the modelling workflow required by this approach are minimal compared with the calculation of central tendency measures across ASUs. Using the geographical centroid of an ASU minimizes the largest possible spatial distance between the centroid and any (unknown) occurrence location within the ASU (Figure 1 (b1)). Considering spatial autocorrelation, the value mismatch between the unknown true location and the chosen substitute (centroid) has a chance of being limited as well. However, central tendency values (e.g. mean or median) of the variable value within the ASU are a better alternative for minimizing the value mismatch, as they limit the potential value distance directly rather than indirectly through spatial distance (Figure 1 (b2)). Although the value mismatch cannot be eliminated completely, using central tendency values lowers the probability of introducing extreme values or outliers. In accordance with previous studies (Park & Davis, 2017), we thus suggest to use central tendency values as substitutes wherever possible. However, when the ASU size is very small (e.g.  $0.25^\circ \times 0.25^\circ$ ) or the environment within the ASU is very homogenous, geographical centroids can work as good substitutes. In this case, it is worth comparing the variables' values from the centroid locations with those from available occurrence locations. If the centroids lead to value outliers, central tendency values should be used instead or the outliers discarded. Of course, this assumes that only a fraction of the available records consists of un-precisely recorded location data.

When using a limited number of ASU centroids *in addition* to precise locations, there are two critical aspects that need to be considered. Firstly, these coarse centroids are typically assigned the same weight as the precise location by the ENMs, which may result in a distorted model of the spatial distribution of the species. This could be ameliorated by down-weighting centroid locations, provided that the chosen modelling algorithm allows for that. Secondly, for each ASU, only one centroid record will be kept by the ENMs. As a consequence, regions with only one record are treated the same as regions with several records, and the abundance information (if available at all) is neglected. It has been shown that a mixture of precise and centroid-based data would be

more robust against the issues demonstrated in this work (Collins et al., 2017), but to what degree and how the ratio of the two data types affect the robustness needs to be clarified in future studies.

It should be noted that the insights gained here only apply to models built with continuous variables. Categorical predictors like land use classes typically show sharp edges on the map, so that even small differences in the spatial location of occurrence records can lead to dramatically different values being assigned to them. Thus, it seems reasonable to assume that models built with categorical variables would be affected more strongly by ASU size, but further investigations are needed to confirm this. Similarly, all our analyses were conducted at an intermediate (sub-continental) spatial scale, as this is the scale where ASU centroids are most likely to be used. Further research is needed to verify whether the conclusions drawn from this can be transferred to coarser or finer scales.

## 5 Conclusions

Whether ASU centroids can be a viable surrogate for precise occurrence locations depends on the ASUs' sizes and how heterogeneous they are in terms of environmental explanatory variables. For instance, in the northern German flatlands, where ASUs are small and the environment comparably homogenous, the use of centroid locations is much less of a problem than in the alpine regions of France, where ASUs are large and environmental gradients steep. If possible, central tendency values should be considered as a more robust alternative. As our results suggest that effects of using ASU centroids outweigh effects of grain size, it is important for modellers to recognize this source of error. In order to enable researchers to assess whether the use of centroid locations is appropriate for a specific project, new methods and guidelines need to be developed.

**Acknowledgement:** Yanchao Cheng was funded by China Scholarship Council, No. 201506040059. Stephanie Thomas was funded by the Bavarian State Ministry of the Environment and Consumer Protection and the Bavarian State Ministry of Health and Care through the BayVirMos project (TKP 01KPB-73560).

**Author contributions:** YC, NT, ST, AJ, and CB conceptualized the study. YC and NT processed and analyzed the data and prepared the figures. YC wrote the first draft and prepared the appendices. NT wrote parts of the results section. CB, ST and AJ supervised the analyses. All authors discussed the methodology and results, and reviewed and edited the manuscript.

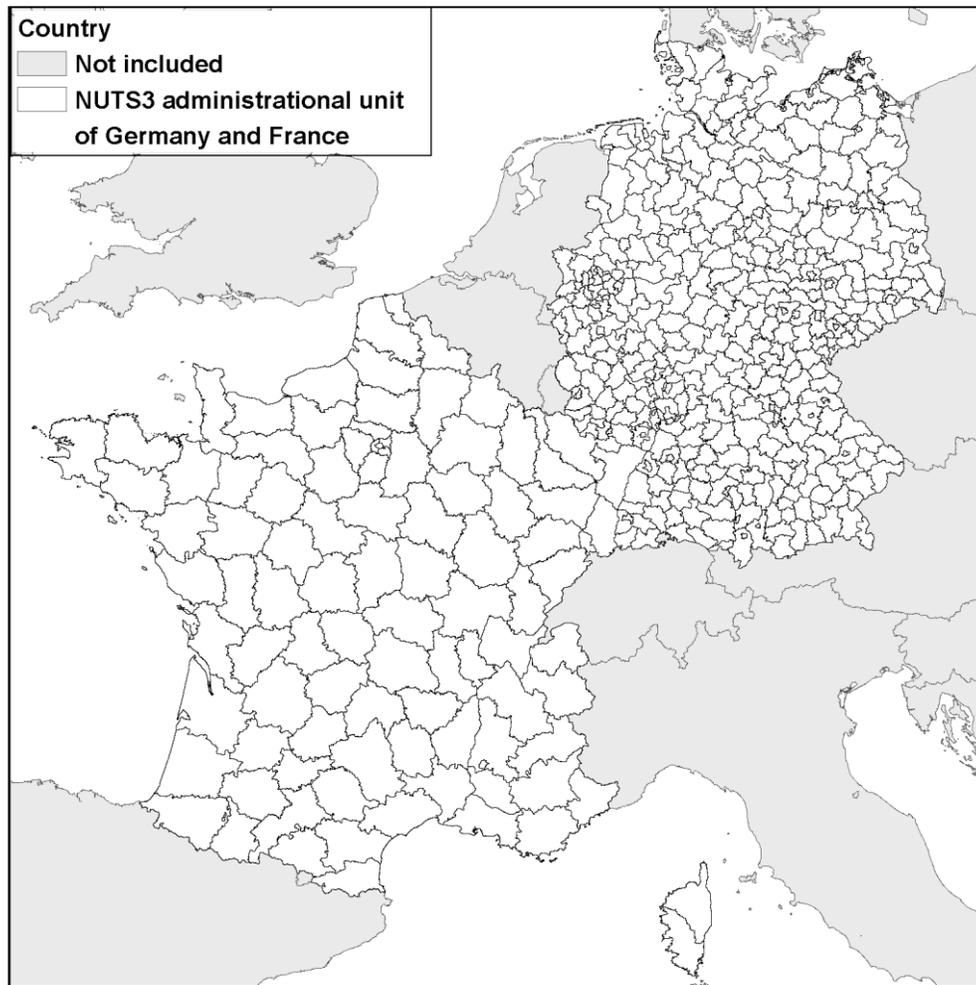
## References

- Allouche, O., Tsoar, A., & Kadmon, R., 2006. Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, *43*, 1223-1232.
- Baldwin, R. A., 2009. Use of maximum entropy modeling in wildlife research. *Entropy*, *11*, 854-866.
- Castellanos, A. A., Huntley, J. W., Voelker, G., & Lawing, A. M., 2019. Environmental filtering improves ecological niche models across multiple scales. *Methods in Ecology and Evolution*, *10*, 481-492.
- Collins, S. D., Abbott, J. C., & McIntyre, N. E., 2017. Quantifying the degree of bias from using county-scale data in species distribution modeling: Can increasing sample size or using county-averaged environmental data reduce distributional overprediction? *Ecology and Evolution*, *7*, 6012-6022.
- Connor, T., Hull, V., Viña, A., Shortridge, A., Tang, Y., Zhang, J. D., ... Liu, J. G., 2018. Effects of grain size and niche breadth on species distribution modeling. *Ecography*, *41*, 1270-1282.
- Cornes, R. C., van der Schrier, G., van den Besselaar, E. J. M., & Jones, P. D., 2018. An Ensemble Version of the E-OBS Temperature and Precipitation Data Sets. *Journal of Geophysical Research: Atmospheres*, *123*, 9391-9409.
- De Marco, P. J., & Nóbrega, C. C., 2018. Evaluating collinearity effects on species distribution models: An approach based on virtual species simulation. *Plos One*, *13*, e0202403.
- Elith, J., Graham, C. H., Anderson, R. P., Dudík, M., Ferrier, S., Guisan, A., ... Zimmermann, N. E., 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, *29*, 129-151.
- Elith, J., & Leathwick, J. R., 2009. Species distribution models: Ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, *40*, 677-697.
- Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., & Yates, C. J., 2011. A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, *17*, 43-57.
- Escobar, L. E., & Craft, M. E., 2016. Advances and limitations of disease biogeography using ecological niche modeling. *Frontiers in Microbiology*, *7*, 1174.
- Evans, J. M., Fletcher, R. J., & Alavalapati, J., 2010. Using species distribution models to identify suitable areas for biofuel feedstock production. *Global Change Biology Bioenergy*, *2*, 63-78.
- Fick, S. E., & Hijmans, R. J., 2017. WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*, 4302-4315.
- Fois, M., Cuenca-Lombraña, A., Fenu, G., Cogoni, D., & Bacchetta, G., 2018. Does a correlation exist between environmental suitability models and plant population parameters? An experimental approach to measure the influence of disturbances and environmental changes. *Ecological Indicators*, *86*, 1-8.
- Fourcade, Y., Besnard, A. G., & Secondi, J., 2018. Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. *Global Ecology and Biogeography*, *27*, 245-256.
- Gábor, L., Moudrý, V., Barták, V., & Lecours, V., 2020. How do species and data characteristics affect species distribution models and when to use environmental filtering? *International Journal of Geographical Information Science*, *34*, 1567-1584.
- Gao, X., & Cao, Z., 2019. Meteorological conditions, elevation and land cover as predictors for the distribution analysis of visceral leishmaniasis in Sinkiang province, Mainland China. *Science of the Total Environment*, *646*, 1111-1116.
- García-Callejas, D., & Araújo, M. B., 2016. The effects of model and data complexity on predictions from species distributions models. *Ecological Modelling*, *326*, 4-12.
- Guisan, A., Graham, C. H., Elith, J., Huettmann, F., Dudík, M., Ferrier, S., ... Zimmermann, N. E., 2007. Sensitivity of predictive species distribution models to change in grain size. *Diversity and Distributions*, *13*, 332-340.
- Hefley, T. J., Brost, B. M., & Hooten, M. B., 2017. Bias correction of bounded location errors in presence-only data. *Methods in Ecology and Evolution*, *8*, 1566-1573.
- Hernandez, P. A., Graham, C. H., Master, L. L., & Albert, D. L., 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography*, *29*, 773-785.

- Hijmans, R. J. (2019). *raster: Geographic Data Analysis and Modeling. R package version 3.0-7*. <https://CRAN.R-project.org/package=raster>.
- Jarek Tuszynski. (2014). *caTools: Tools: moving window statistics, GIF, Base64, ROC AUC, etc. R package version 1.17.1.2*. <https://CRAN.R-project.org/package=caTools>.
- Johnson, T. L., Haque, U., Monaghan, A. J., Eisen, L., Hahn, M. B., Hayden, M. H., ... Eisen, R. J., 2017. Modeling the environmental suitability for *Aedes (Stegomyia) aegypti* and *Aedes (Stegomyia) albopictus* (Diptera: Culicidae) in the contiguous United States. *Journal of Medical Entomology*, *54*, 1605-1614.
- Kramer-Schadt, S., Niedballa, J., Pilgrim, J. D., Schröder, B., Lindenborn, J., Reinfelder, V., ... Wilting, A., 2013. The importance of correcting for sampling bias in MaxEnt species distribution models. *Diversity and Distributions*, *19*, 1366-1379.
- Lauzeral, C., Grenouillet, G., & Brosse, S., 2013. Spatial range shape drives the grain size effects in species distribution models. *Ecography*, *36*, 778-787.
- Leroy, B., Meynard, C. N., Bellard, C., & Courchamp, F., 2016. Virtualspecies, an R package to generate virtual species distributions. *Ecography*, *39*, 599-607.
- Liu, C. R., Berry, P. M., Dawson, T. P., & Pearson, R. G., 2005. Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*, *28*, 385-393.
- Liu, C. R., Newell, G., & White, M., 2016. On the selection of thresholds for predicting species occurrence with presence-only data. *Ecology and Evolution*, *6*, 337-348.
- Liu, C. R., White, M., & Newell, G., 2018. Detecting outliers in species distribution data. *Journal of Biogeography*, *45*, 164-176.
- Lobo, J. M., & Tognelli, M. F., 2011. Exploring the effects of quantity and location of pseudo-absences and sampling biases on the performance of distribution models with limited point occurrence data. *Journal for Nature Conservation*, *19*, 1-7.
- Manzoor, S. A., Griffiths, G., & Lukac, M., 2018. Species distribution model transferability and model grain size - finer may not always be better. *Scientific Reports*, *8*.
- Moudrý, V., & Simova, P., 2012. Influence of positional accuracy, sample size and scale on modelling species distributions: a review. *International Journal of Geographical Information Science*, *26*, 2083-2095.
- Nenzen, H. K., & Araujo, M. B., 2011. Choice of threshold alters projections of species range shifts under climate change. *Ecological Modelling*, *222*, 3346-3354.
- Nezer, O., Bar-David, S., Gueta, T., & Carmel, Y., 2017. High-resolution species-distribution model based on systematic sampling and indirect observations. *Biodiversity and Conservation*, *26*, 421-437.
- Park, D. S., & Davis, C. C., 2017. Implications and alternatives of assigning climate data to geographical centroids. *Journal of Biogeography*, *44*, 2188-2198.
- Pearson, R. G., Dawson, T. P., & Liu, C., 2004. Modelling species distributions in Britain: a hierarchical integration of climate and land-cover data. *Ecography*, *27*, 285-298.
- Peterson, A. T., 2014. *Mapping disease transmission risk: Enriching models using biogeography and ecology*. Baltimore: Johns Hopkins University Press.
- Qiao, H. J., Escobar, L. E., Saupe, E. E., Ji, L. Q., & Soberón, J., 2017. A cautionary note on the use of hypervolume kernel density estimators in ecological niche modelling. *Global Ecology and Biogeography*, *26*, 1066-1070.
- Quiner, C. A., & Nakazawa, Y., 2017. Ecological niche modeling to determine potential niche of Vaccinia virus: A case only study. *International Journal of Health Geographics*, *16*, 28.
- R Core Team. (2015). *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria*. Available from: <http://www.R-project.org/>.
- Record, S., Strecker, A., Tuanmu, M. N., Beaudrot, L., Zarnetske, P., Belmaker, J., ..., Gerstner, B., 2018. Does scale matter? A systematic review of incorporating biological realism when predicting changes in species distributions. *Plos One*, *13*, e0194650.
- Savicky, P. (2014). *pspearman: Spearman's rank correlation test. R package version 0.3-0*. <https://CRAN.R-project.org/package=pspearman>.
- Stolar, J., & Nielsen, S. E., 2015. Accounting for spatially biased sampling effort in presence-only species distribution modelling. *Diversity and Distributions*, *21*, 595-608.
- Tjaden, N. B., Caminade, C., Beierkuhnlein, C., & Thomas, S. M., 2018. Mosquito-borne diseases: Advances in modelling climate-change impacts. *Trends in Parasitology*, *34*, 227-245.

- Warren, D. L., & Seifert, S. N., 2011. Ecological niche modeling in Maxent: The importance of model complexity and the performance of model selection criteria. *Ecological Applications*, 21, 335-342.
- Yackulic, C. B., Chandler, R., Zipkin, E. F., Royle, J. A., Nichols, J. D., Grant, E. H. C...., Veran, S., 2013. Presence-only modelling using MAXENT: When can we trust the inferences? *Methods in Ecology and Evolution*, 4, 236-243.
- Yates, K. L., Bouchet, P. J., Caley, M. J., Mengersen, K., Randin, C. F., Parnell, S., ... Sequeira, A. M. M., 2018. Outstanding challenges in the transferability of ecological models. *Trends in Ecology & Evolution*, 33, 790-802.

## Appendix S1. Administrative Spatial Unit of Germany and France



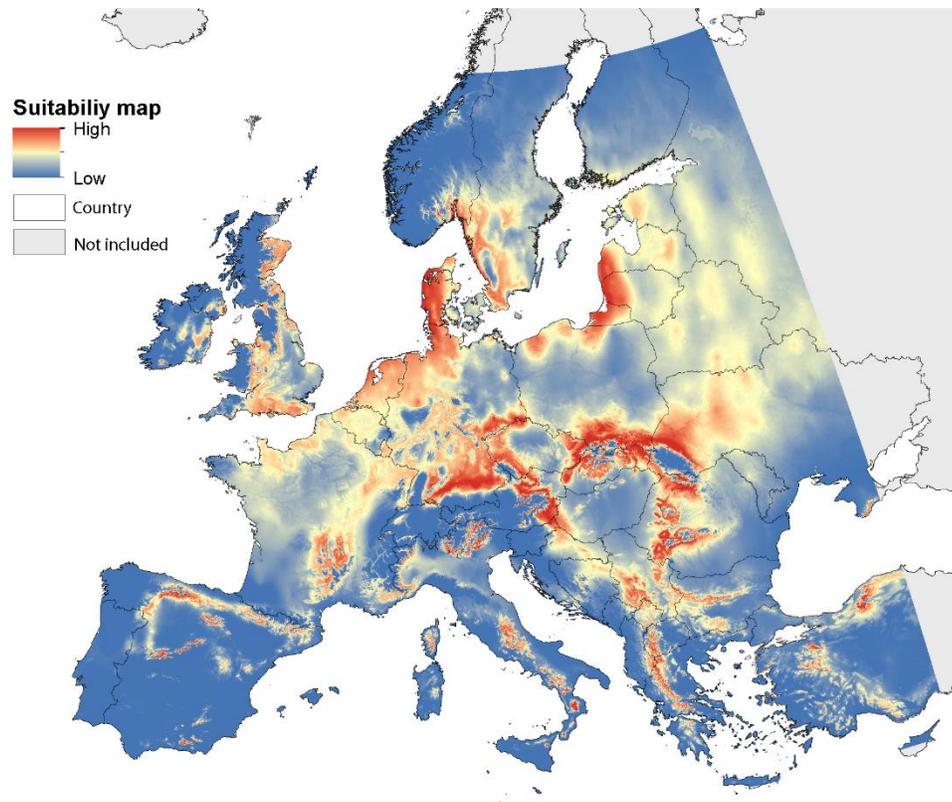
**Figure S1.1.** The NUTS 3 administrative units of Germany and France, that served as test extent. The same virtual species as in the main part of the manuscript was applied for these areas. For both Germany and France, 400 presence-absence locations were drawn, and only presence locations were selected. For France, the average area of NUTS 3 counties is about 6000 km<sup>2</sup>. For Germany, it is 1200 km<sup>2</sup>.

## Appendix S2: Generation of the Virtual Species

For the generation of the virtual species, we used the R package: 'virtualspecies' version 1.4-2 for R version 3.4.2, following the tutorial available at <http://borisleroy.com/files/virtualspecies-tutorial.html>.

**Step 1.** We acquired 19 bioclimatic variables from <https://worldclim.org/data/worldclim21.html>, of the 30 s grain size (see the website "bio 30s", Fick & Hijmans, 2017 ). The names and explanations of these 19 variables can be found in **Appendix 3** or at <https://worldclim.org/data/bioclim.html>.

We downloaded these 19 bioclimatic variables, and cropped them to the European extent. For that, we applied a rectangular mask with the extent of -11 to 35°E, 35 to 65°N, excluding the African continent. The cropped data was converted from GeoTiff to \*.asc raster files.

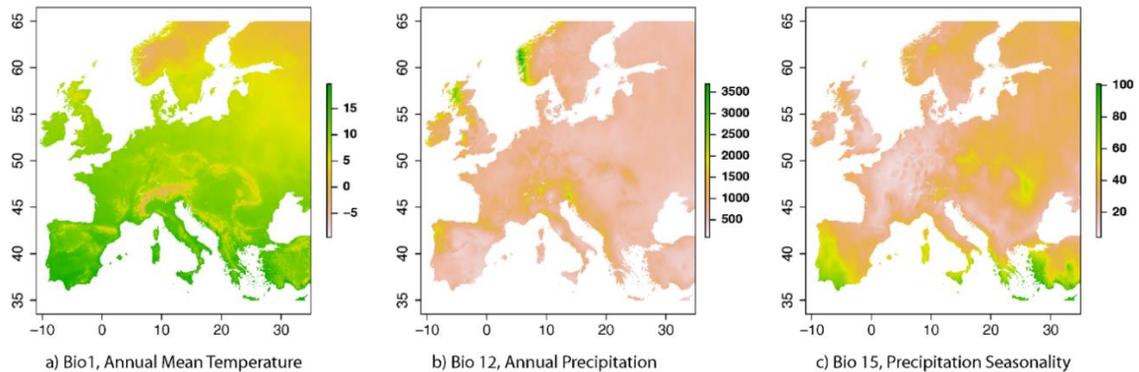


**Figure S2.2.** The extent that the virtual species was generated with. The value shows environmental suitability for the virtual species. Low suitability means values close to 0, and high suitability to 1.

**Step 2.** Calculate spatial correlation among these 19 bioclimatic variables, see **Appendix 3** for the Pearson's correlation coefficient table. Based on that ( $| \text{Pearson's } r | < 0.7$ ), we selected three bioclimatic variables as the basis of the virtual species: **Bio1 (Annual Mean Temperature)**, **Bio12 (Annual Precipitation)**, **Bio15 (Precipitation Seasonality)**.

Of course, for most of species distribution models, more explanatory variables are required. For this virtual species we only chose three variables to keep data process simple.

Now we plot these three variables (1) to see the value range and general pattern of the value distribution, and (2) to make sure the unit of the bioclimatic variables layers are the same as the virtual species generation process (the equations). Note that for Annual precipitation (Figure S2.2 B), the very high value (in green) in the north probably occurs in winter, different from that of the southern part.



**Figure S2.3** Overview of the three explanatory variables used to construct the virtual species.  
a) Bio1 = annual mean temperature, b) Bio12 = annual precipitation, c) Bio15 = precipitation seasonality.

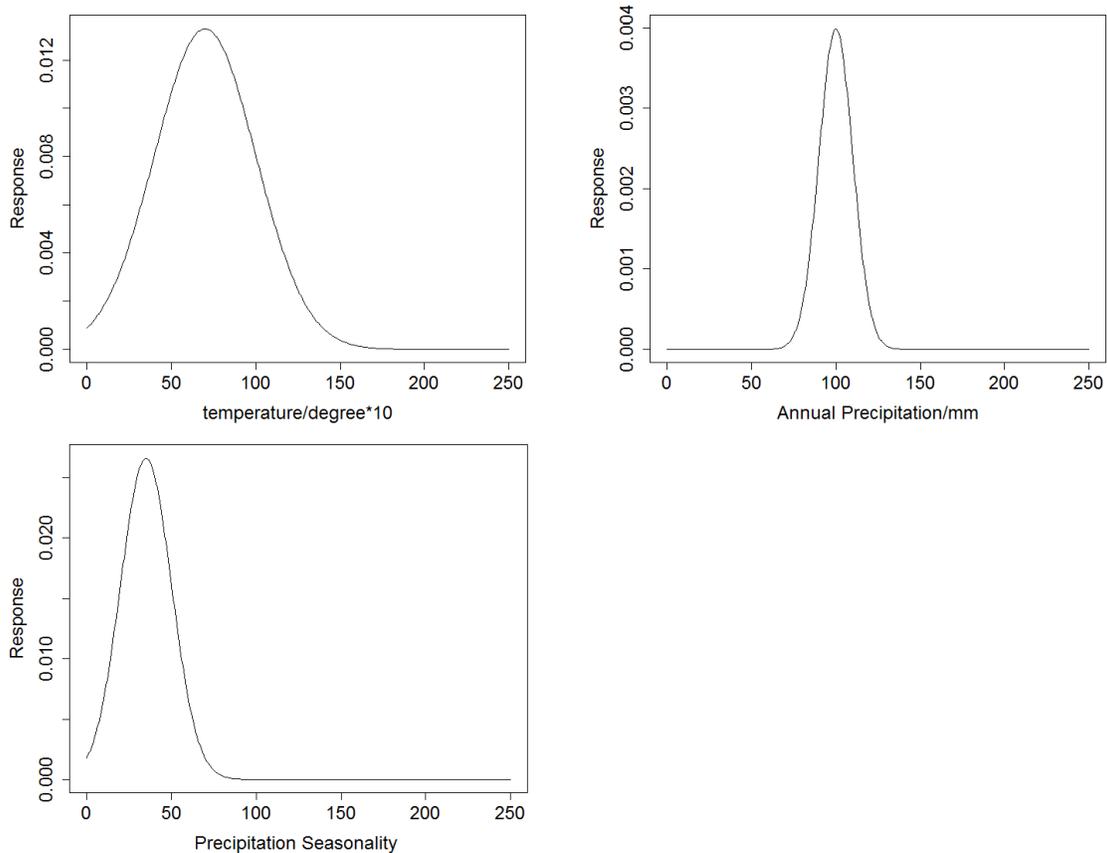
**Step 3.** Set the response curve for each variable using the `formatFunctions()` command and generate the virtual species using `generateSpFromFun()`:

```
my.parameters <- formatFunctions(bio1 = c(fun = 'dnorm', mean = 8, sd = 3),
                                bio12 = c(fun = 'dnorm', mean = 800, sd = 150),
                                bio15 = c(fun = 'dnorm', mean = 40, sd = 30))

species1 <- generateSpFromFun(raster.stack = my.variables,
                              parameters = my.parameters,
                              formula = "bio1 * bio12 * bio15")
```

Here I just multiply these three explanatory variables together as the formula. Actually, it is not so different from adding them together.

Our aim of generating a virtual species is to test our hypotheses, to run ENMs with a *known* species. For running ENMs, we do not want a virtual species which occupies every corner across our study area, nor a species only with very sparse distribution. Those response curves were thus defined and adjusted.



**Figure S2.4.** Response (curves) to the three bio-climate variables defined for the virtual species. For Annual Mean Temperature, the X-axis was multiplied by 10. It does not affect the virtual species, only for presentation here.

**Step 4.** Set alpha and beta values needed for converting the continuous environmental suitability map into a binary presence-absence map. Following the recommendations in the tutorial to achieve a reasonable distribution map, we set them to:  $\beta = 0.5$ ,  $\alpha = -0.09$

beta controls the inflexion point, beta and alpha together define the shape of the transformation curve. As shown by Figure S2.2, the environmental suitability is actually with logistic values, ranging from 0 to 1. To get a binary presence-absence map, some kind of transformation method is needed. The parameters alpha and beta here are for this purpose: defining a transformation curve instead of a single value as threshold. Theoretically, beta can be of any value between 0 and 1, in our case we set it at 0.5, so the inflexion point is in the center of the logistic value range.

See also the explanation from the tutorial of Virtual species: *“a lower beta will increase the probability of finding suitable conditions for the species (wider distribution range). A higher beta will decrease the probability of finding suitable conditions (smaller distribution range)”*.

alpha drives the ‘slope’ of the curve: e.g.  $\alpha = -1$ , it is a linear transformation curve;  $\alpha = -0.00001$ , the transformation curve is almost like a threshold curve. We set  $\alpha = -0.09$ , it is also somehow a medium level of slope.

For more information: <http://borisleroy.com/files/virtualspecies-tutorial.html#introduction-why-should-we-avoid-a-threshold-conversion>, see also (Meynard & Kaplan, 2013).

**Step 5.** Draw occurrence locations. This was done by a simple function available from Virtual Species package, using the command `sampleOccurrences, type="presence-absence"`. For each rectangular region, about 60 presence locations were achieved.

Three rectangular regions at size of  $5^{\circ} \times 5^{\circ}$  were chosen. Up to date, there are not really golden rules available regarding the ideal size of study area. To conduct the two factorial design, we need at least three replicates. In our case, to get three rectangular regions fitting our criterion across Europe, there were not many choices. We set the size of the rectangular regions at  $5^{\circ} \times 5^{\circ}$ , as seen in Figure 2. It can also be  $6^{\circ} \times 6^{\circ}$  or larger, but the main problem here is can they all fit in across the extent of Europe.

In addition, by setting the size at  $5^{\circ} \times 5^{\circ}$ , the number of ASUs we got was 25 (large), 100 (medium), 400 (small). If we use  $6^{\circ} \times 6^{\circ}$ , it would be 36 (large), 144 (medium) and 576 (small). Apparently, the former one is more intuitive and easier to grasp.

**Step 6.** Export the continuous suitability map (Figure S2.2), the presence-absence map, and true occurrence records to be used as the basis for the analyses.

## References

- Fick, S.E. and R.J. Hijmans, 2017. WorldClim 2: new 1km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* 37 (12): 4302-4315.
- Meynard, C. N., & Kaplan, D. M. (2013). Using virtual species to study species distributions and model performance. *Journal of Biogeography*, 40, 1-8.

### Appendix S3. Pearson's Correlation Coefficient

**Table 3.1.** Pearson's correlation coefficient (calculated for the rectangular region defined in Appendix 2)

	Bio 1	Bio 2	Bio 3	Bio 4	Bio 5	Bio 6	Bio 7	Bio 8	Bio 9	Bio 10	Bio 11	Bio 12	Bio 13	Bio 14	Bio 15	Bio 16	Bio 17	Bio 18	Bio 19
<b>Bio 1</b>	1.00																		
Bio 2	0.60	1.00																	
Bio 3	0.74	0.58	1.00																
Bio 4	-0.45	0.04	-0.77	1.00															
Bio 5	0.86	0.79	0.46	0.04	1.00														
Bio 6	0.87	0.28	0.83	-0.81	0.51	1.00													
Bio 7	-0.22	0.37	-0.52	0.94	0.30	-0.67	1.00												
Bio 8	-0.01	0.01	-0.42	0.55	0.22	-0.29	0.51	1.00											
Bio 9	0.82	0.55	0.80	-0.57	0.65	0.82	-0.35	-0.44	1.00										
Bio 10	0.90	0.70	0.46	-0.01	0.99	0.58	0.21	0.23	0.66	1.00									
Bio 11	0.94	0.45	0.87	-0.73	0.65	0.98	-0.53	-0.24	0.86	0.70	1.00								
<b>Bio 12</b>	-0.16	-0.39	0.15	-0.49	-0.46	0.18	-0.59	-0.47	0.03	-0.42	0.07	1.00							
Bio 13	-0.08	-0.29	0.15	-0.40	-0.32	0.18	-0.47	-0.41	0.09	-0.29	0.09	0.93	1.00						
Bio 14	-0.33	-0.50	-0.01	-0.37	-0.60	0.00	-0.52	-0.29	-0.24	-0.56	-0.12	0.81	0.60	1.00					
<b>Bio 15</b>	0.30	0.35	0.15	0.09	0.42	0.14	0.20	-0.02	0.34	0.39	0.21	-0.16	0.17	-0.63	1.00				
Bio 16	-0.09	-0.31	0.15	-0.43	-0.35	0.19	-0.51	-0.42	0.09	-0.32	0.09	0.95	0.99	0.63	0.13	1.00			
Bio 17	-0.29	-0.48	0.03	-0.41	-0.58	0.05	-0.55	-0.34	-0.18	-0.54	-0.07	0.85	0.64	0.99	-0.61	0.67	1.00		
Bio 18	-0.66	-0.61	-0.45	0.08	-0.73	-0.45	-0.14	0.11	-0.68	-0.72	-0.55	0.60	0.49	0.76	-0.42	0.51	0.73	1.00	
Bio 19	0.17	-0.15	0.42	-0.63	-0.15	0.46	-0.63	-0.63	0.42	-0.11	0.38	0.86	0.88	0.53	0.12	0.89	0.58	0.15	1.00

Numbers in yellow: Pearson's correlation coefficient higher than 0.70.

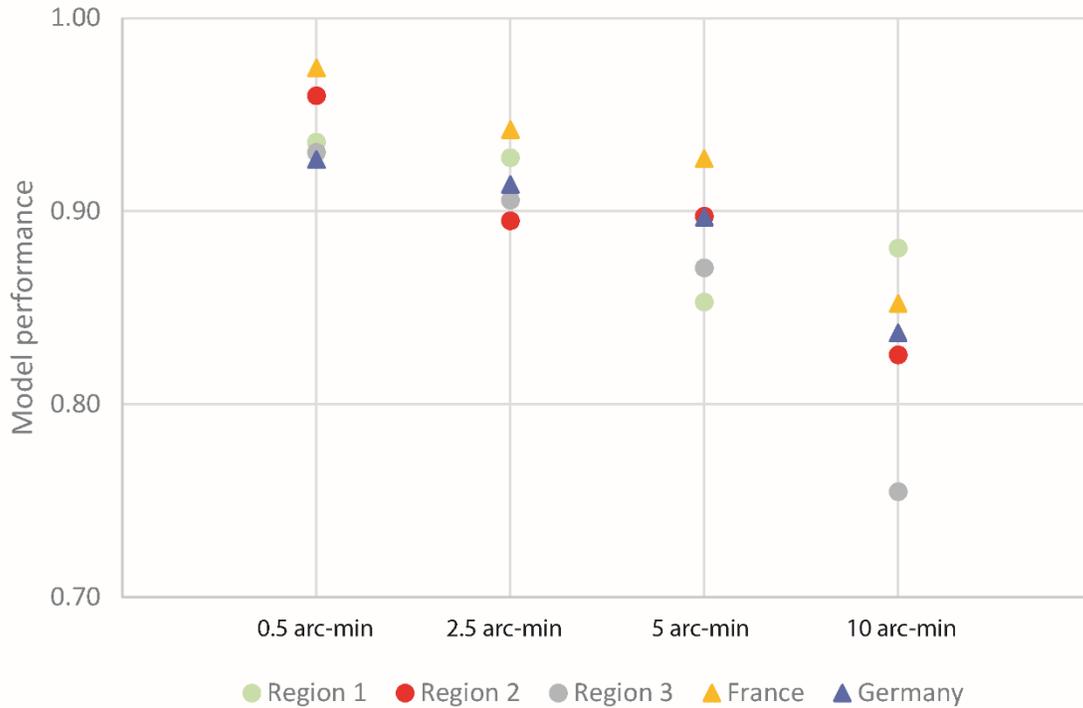
Variables in orange: Chosen variables for generating the virtual species.

**Full names of the bioclimate variables**

- Bio 1 = Annual Mean Temperature
- Bio 2 = Mean Diurnal Range (Mean of monthly (max temp - min temp))
- Bio 3 = Isothermality (Bio 2/Bio 7) (×100)
- Bio 4 = Temperature Seasonality (standard deviation ×100)
- Bio 5 = Max Temperature of Warmest Month
- Bio 6 = Min Temperature of Coldest Month
- Bio 7 = Temperature Annual Range (Bio 5-Bio 6)
- Bio 8 = Mean Temperature of Wettest Quarter
- Bio 9 = Mean Temperature of Driest Quarter
- Bio 10 = Mean Temperature of Warmest Quarter
- Bio 11 = Mean Temperature of Coldest Quarter
- Bio 12 = Annual Precipitation
- Bio 13 = Precipitation of Wettest Month
- Bio 14 = Precipitation of Driest Month
- Bio 15 = Precipitation Seasonality (Coefficient of Variation)
- Bio 16 = Precipitation of Wettest Quarter
- Bio 17 = Precipitation of Driest Quarter
- Bio 18 = Precipitation of Warmest Quarter
- Bio 19 = Precipitation of Coldest Quarter

## Appendix S4 Calculations

### Model performance at different grain sizes



**Figure S4.5.** Model performance at different grain sizes, with true occurrence locations. Region 1, 2 and 3 are the rectangular regions as shown in Figure 2. Resolution does affect model performance. However, this effect is very small compared to the effects resulting from varying pseudo-ASU sizes. The virtual species was generated at 0.5 arc-min. With true occurrence locations and this 0.5 arc-min grain size, the model performance is the best.

## Two-way ANOVA

Table 4.2 Two-way ANOVA results

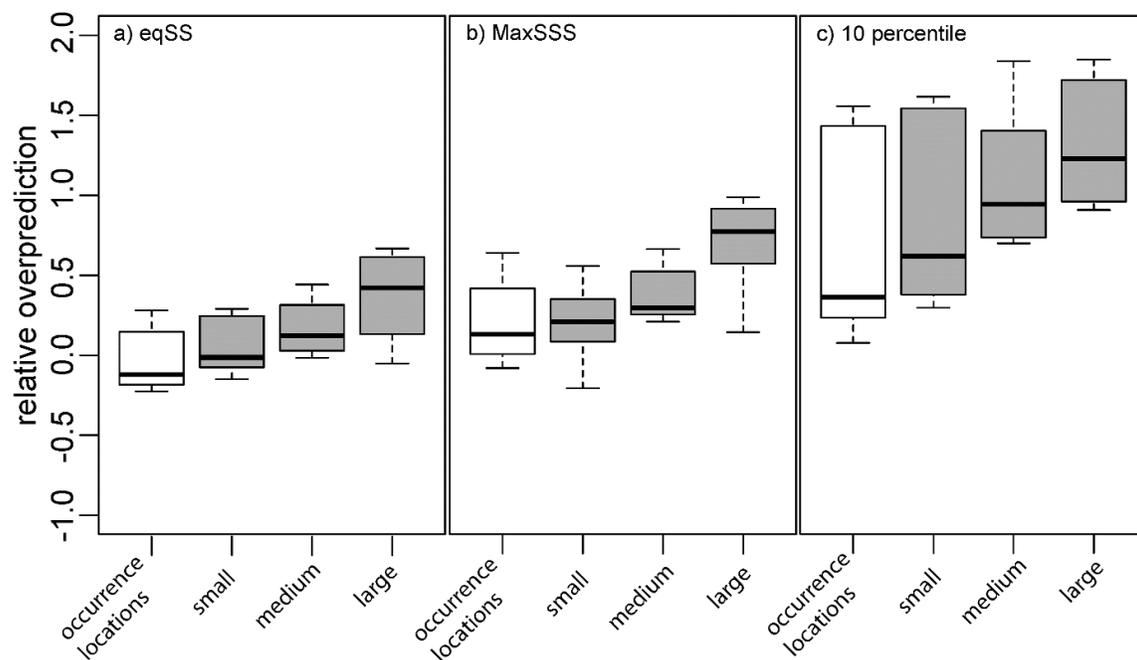
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Grain size	3	0.00778	0.00259	0.876	0.467
ASU	2	0.09716	0.04858	16.414	3.22e-05 ***
Grain size:ASU	6	0.00997	0.00166	0.561	0.757
Residuals	24	0.07103	0.00296		

---  
 Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Shapiro-Wilk normality test

data: aov\_residuals  
 W = 0.96511, p-value = 0.3074

**Overprediction ratio, calculated with different thresholds.**



**Figure S4.6.** Relative over-prediction ratio (ENM model results vs.true virtual species occurrence) for models built on point locations as well as centroids of differently sized regions, calculated according to thresholds of a) Equal training sensitivity and specificity logistic threshold (eqSS), b) Maximum training sensitivity plus specificity logistic threshold (MaxSSS) and c) 10 percentile training presence logistic threshold (10 percentile).

## Manuscript 2

Deriving risk maps from epidemiological models of vector borne diseases: state-of-the-art and suggestions for best practice

Published in *Epidemics*, (2020) 33:100411

Yanchao Cheng<sup>1</sup>, Nils Benjamin Tjaden<sup>1</sup>, Anja Jaeschke<sup>1</sup>, Stephanie Margarete Thomas<sup>1,2</sup>, Carl Beierkuhnlein<sup>1,2</sup>

1 Department of Biogeography, University of Bayreuth, Universitätsstr. 30, 95447, Bayreuth, Germany

2 BayCEER, Bayreuth Center for Ecology and Environmental Research, Bayreuth, Germany

**Keywords:** basic reproduction number; basic reproductive ratio; mechanistic model;  $R_0$ ; risk map

## Abstract

Epidemiological models (EMs) are widely used to predict the temporal outbreak risk of vector-borne diseases. EMs typically use the basic reproduction number ( $R_0$ ), a threshold quantity, to indicate risk. To provide an overall view of the risk, these model outputs can be transformed into spatial risk maps, using various aggregation methods (e.g. average  $R_0$  over time, cumulative number of days with  $R_0 > 1$ ). However, there is no standardized methodology available for this. Depending on the specific aggregation methods used, the yielded spatial risk map may have considerably different interpretations. Additionally, the method used to visualize the aggregated data also affects the perceived spatial patterns. In this review, we compare commonly used aggregation and visualization methods and discuss the respective interpretation of risk maps.

Research publications using epidemiological modelling methods were drawn from Web of Science. Only publications containing maps of  $R_0$  transformed from EMs were considered for the analysis. An example EM was applied to illustrate how aggregation and visualization methods affect the final presentations of risk maps.

Risk maps can be generated to show duration, intensity and spatio-temporal dynamics of potential outbreak risk of VBDs. We show that 1) different temporal aggregation methods lead to different interpretations; 2) similar spatial patterns do not necessarily bear the same meaning; 3) visualization methods considerably affect how results are perceived, and thus should be applied with caution. We recommend mapping both intensity and duration of the VBD outbreak risk, using small time-steps to show spatio-temporal dynamics when possible.

## Introduction

### *Vector-borne diseases and risk mapping approaches*

Vector-borne diseases (VBDs), causing more than 700,000 deaths each year (WHO, 2017), have become a huge concern for public health. The vectors include different groups of arthropods such as mosquitoes, ticks, or sandflies, while the hosts are typically vertebrate animals including humans. Many vector-borne pathogens have been limited to specific areas, e.g. malaria epidemics in Africa (Taffese et al., 2018). With global change, however, vector-borne pathogens can increasingly invade naïve populations of vectors and hosts in formerly disease-free areas, causing disease outbreaks and leading to new challenges in disease surveillance and control (Caminade et al., 2019; Gage et al., 2008; Myers et al., 2000; Tatem et al., 2006). International air travel is a major contributor to the global spread of pathogens. International cargo transportation facilitates the transport of vectors to regions that were previously not reachable due to dispersal limitations (Thomas et al., 2014), which now poses threats through the introduction of related pathogens. Climate change impacts the transmission of pathogens, e.g. affecting the length of the transmission season, or enabling diseases to emerge in areas where they could not survive before (ECDC, 2019; WHO, 2017). Consequently, the chances of VBDs emerging in novel places are increasing. To better monitor and control VBDs, it is essential to know where and when a VBD can potentially occur; and if possible, how severe the situation is.

Regarding these questions, a variety of modelling approaches have been introduced and applied. Generally, current modelling studies can be categorized by their use of either correlative or process-based models, either of which has pros and cons (Tjaden et al., 2018). The spatial aspects of risk are typically modelled using correlative ecological niche models (also known as species distribution models) based on species (vectors, virus, infected vectors or hosts) occurrence records and respective explanatory variables describing their environment (Escobar & Craft, 2016; Peterson, 2014). The explanatory variables can include a broad range of factors such as bioclimatic variables, land use, and any other variables that may affect spatial distribution patterns. The outputs of the ecological niche models are maps of the potential spatial distribution of the target species. As the explanatory variables are typically based on long-term (e.g. yearly or decadal) average data, ecological niche models are mainly applied for gaining estimates of long-term (decadal) trends in risk for large (continental) areas. Ecological niche models are well-accepted for their good performance on spatial distribution mapping. However, correlative models typically do not resolve the short-term dynamics of real-world outbreaks. This is where the process-based epidemiological models discussed in this paper promise an advantage.

Epidemiological models (EMs, also referred to as mechanistic or mathematical models, compartmental epidemic models or compartmental models (Siettos & Russo, 2013)) aim to identify risks at fine temporal resolution at a specific location. EMs can assess the course of an outbreak by simulating the processes that drive the transmission cycle of the disease. EMs do not require occurrence records of the investigated disease or pathogen when constructing the model. Instead,

they are dependent on detailed knowledge of the respective processes upon which they are based. One major environmental variable that often drives these processes is ambient temperature (Tjaden et al., 2018).

The temporal resolution of an EM depends on the available data for temperature or other key driving forces (Calistri et al., 2016). It can be annual (Wu et al., 2013), monthly (Hartemink et al., 2009), biweekly (Hartley et al., 2012), or even finer. Spatially, these models are often limited to a specific place or region where processes are well understood (Du et al., 2017; Ferguson et al., 2016). Nevertheless, EMs are increasingly used to estimate spatial aspects of risk as well (Tjaden et al., 2018). In this study, we review how values of  $R_0$  obtained from EMs can be transformed into meaningful maps. We demonstrate that the interpretation of such maps differs depending on the underlying model algorithm and the chosen method of aggregation. The visualization methods need to be chosen carefully as well.

### ***The basic reproduction number ( $R_0$ ): a threshold parameter***

The basic reproduction number ( $R_0$ ) is a characteristic value for epidemiological assessments (Hethcote, 2000; Ma & Li, 2009; Smith et al., 2014). It was originally defined as the expected number of secondary cases caused by a single infected individual (host) during its lifetime within a susceptible population (Diekmann et al., 1990; Dietz, 1993; Liu et al., 2018). The investigated disease can spread when  $R_0 > 1$ . Otherwise, it cannot (Diekmann et al., 1990; Hethcote, 2000; Liu et al., 2018; Smith et al., 2014). For an infectious disease,  $R_0$  can be used to determine the herd immunity threshold that is required to prevent it from further spreading (Dietz, 1993; Guerra et al., 2017; Massad et al., 2001).  $R_0$  has also been used as an indicator for the spreading speed of diseases (Ridenhour et al., 2014).

$R_0$  can be measured directly through disease surveillance systems (Marques et al., 1994) or estimated via EMs (Delamater et al., 2019; Ridenhour et al., 2018; Siettos & Russo, 2013). For VBDs, however,  $R_0$  is difficult to assess directly from surveillance systems due to the significant workload required to conduct both monitoring and surveillance and to determine the complexity of transmission patterns of VBDs (Delamater et al., 2019). Instead, estimates of  $R_0$  are often based on EMs. These estimates are generally also called  $R_0$ , although it must be noted that they may differ considerably from the original definition in several aspects (Delamater et al., 2019; Li et al., 2011). While they all share the threshold at  $R_0=1$ , absolute values of  $R_0$  are not comparable between different types of EM (Dietz, 1993; Heffernan et al., 2005; Li et al., 2011; Yang, 2014). For instance, values of  $R_0$  calculated for a VBD with the popular next generation matrix (NGM) approach represent the mean number of infections in both vectors and hosts per generation (Martcheva, 2015). This is different from the original definition of  $R_0$  that refers to the number of individuals (hosts) that get infected by a single infected host throughout its life time. Only the survival function approach reliably calculates values of  $R_0$  that are consistent with this definition (Li et al., 2011).

### ***Generating spatial risk maps from epidemiological models: temporal aggregation and visualization***

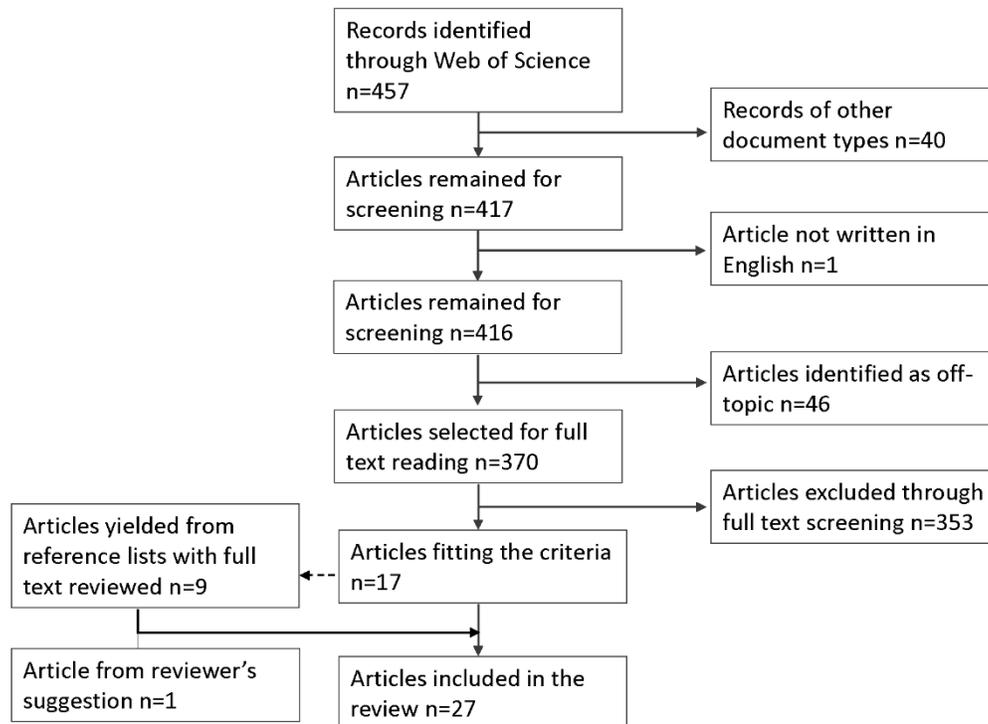
For the calculation of  $R_0$  for VBDs, EMs typically use time series of temperature data for specific locations, which are acquired at daily resolution from weather stations (Cadar et al., 2017; Cheng et al., 2018; Hartemink et al., 2009; Holy et al., 2011; Mordecai et al., 2017; Racloz et al., 2008). This enables capturing of weather extremes and assessment of temporal risk fluctuations. To assess the spatial aspects of risk, time series of point-wise weather station data can be replaced with time series of spatial temperature data representing the weather conditions across the whole study area (typically stored as geographical raster data). The series of spatial  $R_0$  maps derived from this provides spatio-temporal predictions of transmission risk. Since this can result in an unpractically large number of figures (e.g. 365 maps for one year of daily-resolution data), some form of temporal post-modelling aggregation of the series of  $R_0$  maps is required. Careful attention must be given during this process, as the method for temporal aggregation of raw EM outputs into a single map affects the interpretation and usefulness of the final figure.

In addition to temporal data aggregation methods, visualization methods affect the final presentation of risk maps as well. After all, as every map necessarily is a distorted model of the real world (Monmonier, 1991), decisions on *how* to distort reality for the benefit of the viewer should be made consciously and coherently. The first thing to consider here is whether the raw (temporally aggregated) data should be displayed as an unclassed map, where each data value corresponds to a unique point on a continuous color gradient. The alternative is a classed map, where the range of raw data values is broken up into a limited number of groups or classes, each of them represented by a unique color. How many classes to use and how to define them are necessary follow-up questions in this case (Slocum et al., 2009 ). Finally, an appropriate color scheme needs to be found (Brewer, 1994).

In this study, we investigate how the outputs of  $R_0$ -based EMs have been transformed into spatial risk maps in the literature. Commonly used methods are identified and illustrated using an established NGM-based EM as an example. We evaluate the strengths and weaknesses of the different methods and make recommendations for both useful methods for temporal aggregation and informative visualization schemes. With these analyses, we raise awareness and contribute to the development of a more standardized methodology for future usage.

## Methods

A systematic literature search was performed via Web of Science (Clarivate Analytics, 2019) topic searching, using core collections only (last accessed on January 17<sup>th</sup>, 2020). Two sets of intentionally broad key word combinations were applied: 1) “vector AND borne AND basic AND reproduct\*” and 2) “basic AND reproduct\* AND climate AND change AND (risk OR map OR spatial OR distribut\* OR transmit\*)”, resulting in 457 publications (Figure 1).



**Figure 1.** Flowchart of the article selection process concerning spatio-temporal risk maps produced with epidemiological models. First, only articles published in peer-reviewed scientific journals were considered. Articles that were not written in English and those where screening of abstract and/or title indicated that they were outside the scope of this review were removed. The remaining articles were screened visually for the existence of  $R_0$ -based maps in the full-text downloads, resulting in 17 articles fitting the inclusion criteria. The same procedure was then applied to the literature referencing/referenced in these articles, revealing another nine relevant literature records.

The publications were filtered via following criteria: they should be 1) research articles (as reviews and book chapters may cause double counting), 2) written in English and 3) include spatial risk maps (showing  $R_0$ ) that were generated from process-based EMs (Figure 1).

Among these 457 publications yielded from the primary search, 416 were of article type and written in English. Another 46 articles were identified as off-topic based on title or abstract. The remaining 370 articles were downloaded for manual full-text screening for the presence of maps representing  $R_0$ . Manual screening was chosen as our primary selection method for the simple reason that maps

are difficult to find through search strings but easily recognizable for the human eye. This procedure yielded seventeen articles.

To broaden the search even more, the literature referenced by these 17 research articles as well as the literature citing them were screened and tracked manually, resulting in nine more articles that fit the inclusion criteria. In consequence, a total number of 27 research articles was analyzed in this review.

The publications were analyzed according to the following details: 1) the temporal aggregation method applied to summarize raw  $R_0$  values over time; 2) the visualization method used to display the aggregated model output data; 3) the method used for estimating  $R_0$  (i.e. NGM or probabilistic); 4) the data source (e.g. time series of daily temperature data or satellite images) that was used by the EM; and 5) the specific VBD investigated and the respective study area.

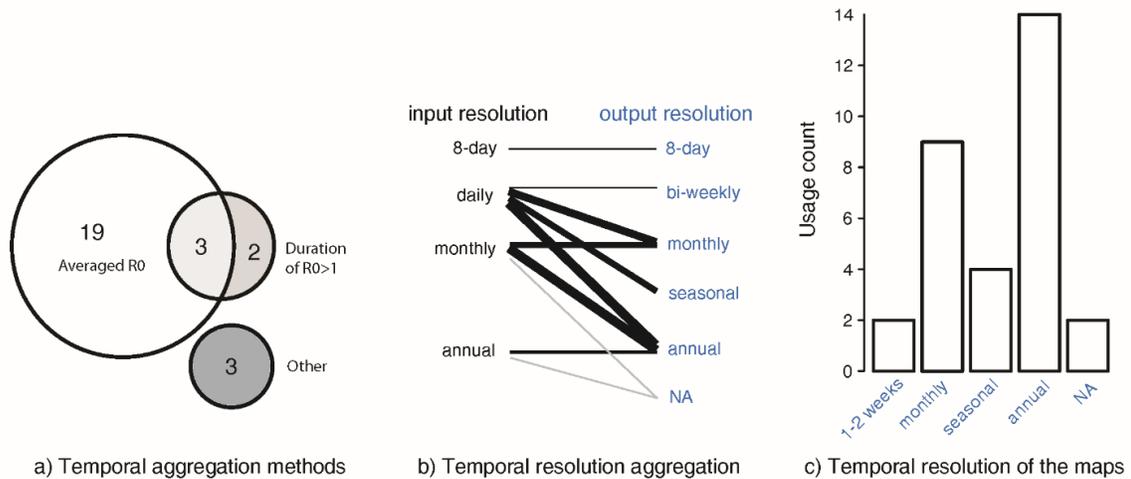
To illustrate the differences among the recorded transformation methods, these methods were demonstrated with an NGM-based EM developed by Rubel et al. (Rubel et al., 2008) for Usutu virus and previously used for a cross-discipline model comparison (Cheng et al., 2018). The model was run using rastered 0.25 degree resolution daily temperature data for 2018 from the E-OBS data set (Haylock et al., 2008). With the same original EM output, a series of spatial risk maps were generated using different spatial transformation methods. More details about the NGM-based EM applied are provided in Supplement S1.

## Results

### *Summary of recorded methods for temporal aggregation and visualization*

Overall, the NGM method was applied in 19 studies, whereas the remaining 8 studies followed a probabilistic approach (see Supplement S2). From the 27 research papers analyzed in this review, two types of temporal aggregation methods for the creation of maps from  $R_0$  values were recorded (Figure 2 a): the averaged intensity (mean values of  $R_0$  over time) on the one hand and the duration of the at-risk state (number of days with  $R_0 > 1$ ) on the other hand. In all but three studies, some kind of temporal aggregation was applied. Three publications used both averaged intensity and the duration of the potential risk in parallel (Brugger & Rubel, 2013; Cheng et al., 2018; Ng et al., 2017). The most frequently used method was the averaged intensity (Figure 2 a). Some form of averaging was performed in 22 studies, using annual, seasonal (transmission season), monthly, biweekly, or roughly weekly (8 days) averages of  $R_0$  (Figure 2 b). All of the aggregated NGM-based EMs applied this method. In some cases, average values of  $R_0$  were calculated indirectly, e.g. by first calculating the relationship between  $R_0$  and temperature, then calculating the annual average  $R_0$  based on long-term average temperature data (Cordovez et al., 2014; Kakmeni et al., 2018; Wu et al., 2013). Only five studies included maps that show the duration of the at-risk state ( $R_0 > 1$ ), three of which also show average  $R_0$  in parallel. Among these, two studies displayed the total number of days with  $R_0 > 1$  throughout the study period (Brugger & Rubel, 2013; Cheng et al., 2018), while the remaining three (Holy et al., 2011; Mordecai et al., 2017; Ng et al., 2017) mapped the number of (consecutive)

months with  $R_0 > 1$ . One study (Racloz et al., 2008) calculated  $R_0$  based on monthly mean temperatures with no further aggregation. Two further studies only calculated  $R_0$  for a single point in time, so that no aggregation was necessary for the final map (Hartemink et al., 2011; Moraga et al., 2015).



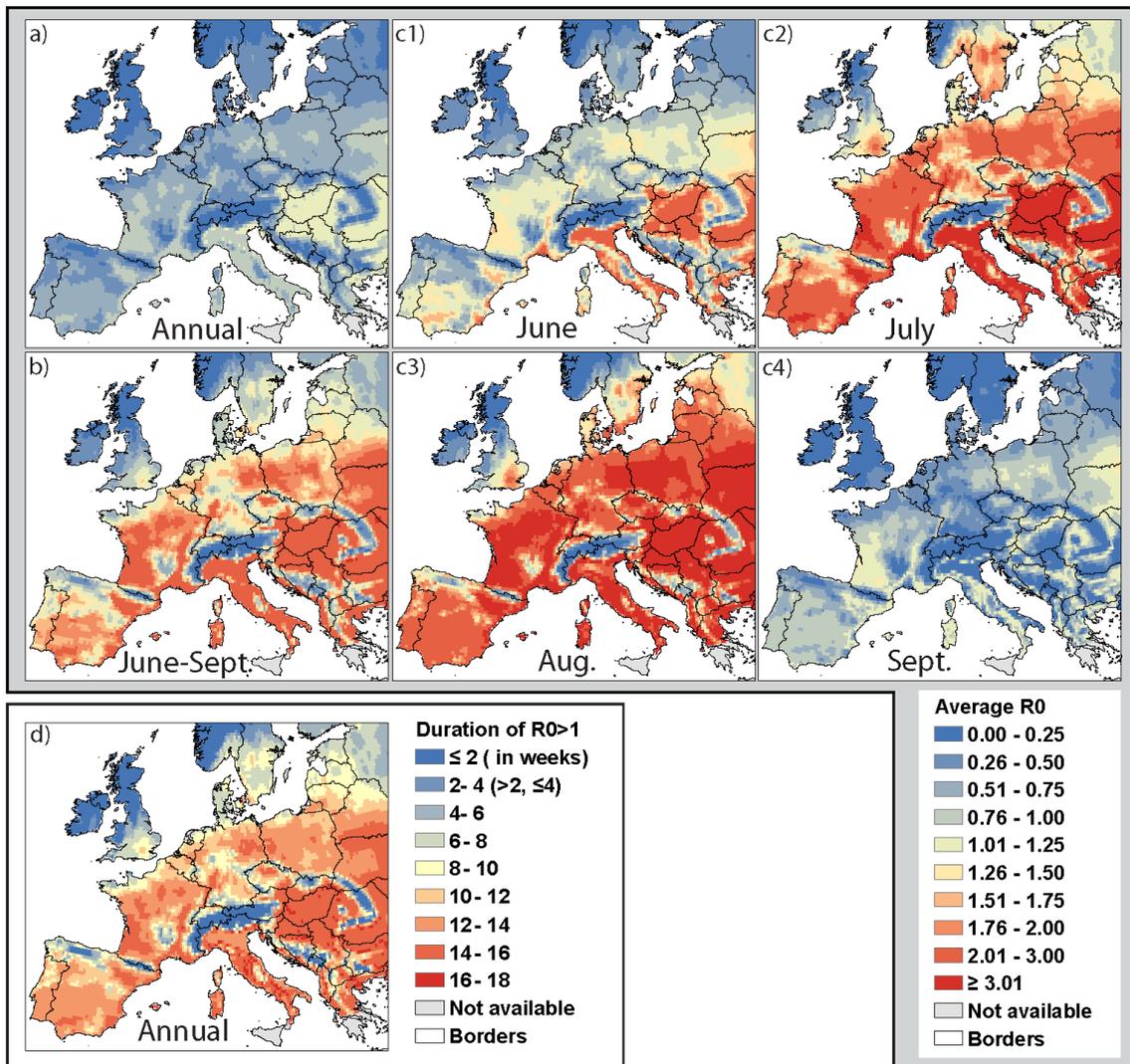
**Figure 2.** Overview of the studies using epidemiological models to produce spatial risk maps for vector-borne diseases. a) Studies grouped by the temporal aggregation methods applied. Out of 27 total, 19 studies displayed temporal averages of  $R_0$ , 2 displayed the duration of the at-risk state ( $R_0 > 1$ ) and 3 more showed both. 3 studies did not apply any temporal aggregation method. b) Relationship between temporal resolution in input data and final maps. Thickness of the connecting lines is equivalent to the number of studies aggregating in this way. c) Temporal resolution of the final maps.

Among the recorded studies, 13 were based on daily resolution weather data, 10 on monthly data, 3 on annual mean temperature and 1 on an 8-day resolution satellite data (Figure 2 b). The resulting maps typically showed annual ( $n=14$ ) or monthly ( $n=9$ ) representations of  $R_0$  (Figure 2 c). Seasonal ( $n=4$ ) and sub-monthly ( $n=2$ ) maps were rarer, and in two cases the time frame represented by the maps remained unclear. EMs based on daily resolution data were transformed into final maps of bi-weekly, monthly, seasonal or annual resolution. EMs based on monthly resolution data were transformed into maps of either monthly or annual resolution. In two cases, coarse resolution temperature data was first interpolated to gain daily estimates of  $R_0$ , which were then aggregated again so that the maps met the coarse temporal resolution of the temperature data (Calistri et al., 2016; Zhang et al., 2017).

Regarding visualization methods, 9 studies presented unclassified maps with continuous color ramps. The remaining 18 studies presented classed maps, 7 of which used a relatively large amount ( $\geq 10$ ) of classes. Averaged intensity maps are typically visualized using diverging color ramps with cold colors for  $R_0 < 1$  and warm colors for  $R_0 > 1$ .

**Exemplary demonstration of common aggregation and visualization methods**

The aggregation methods affect the resulting values and respective interpretations (e.g. intensity or duration) across the risk map; the visualization methods, on the other hand, affect the perceived map patterns. To illustrate how these two factors affect the final presentation of risk maps, a series of figures was derived from the same EM for Usutu virus (Figure 3 and 4).



**Figure 3.** Temporal aggregation methods: Different spatial risk maps derived from the same NGM-based epidemiological model for Usutu virus in Europe, based on gridded daily temperature observations for 2018. Demonstrating the two common aggregation methods, a total number of 365 daily raw  $R_0$  maps were transformed into: a) annual average  $R_0$ , b) average  $R_0$  over the transmission season, June – September, c) monthly average  $R_0$  during the transmission season, d) duration of the at-risk-state ( $R_0 > 1$ ) in weeks.

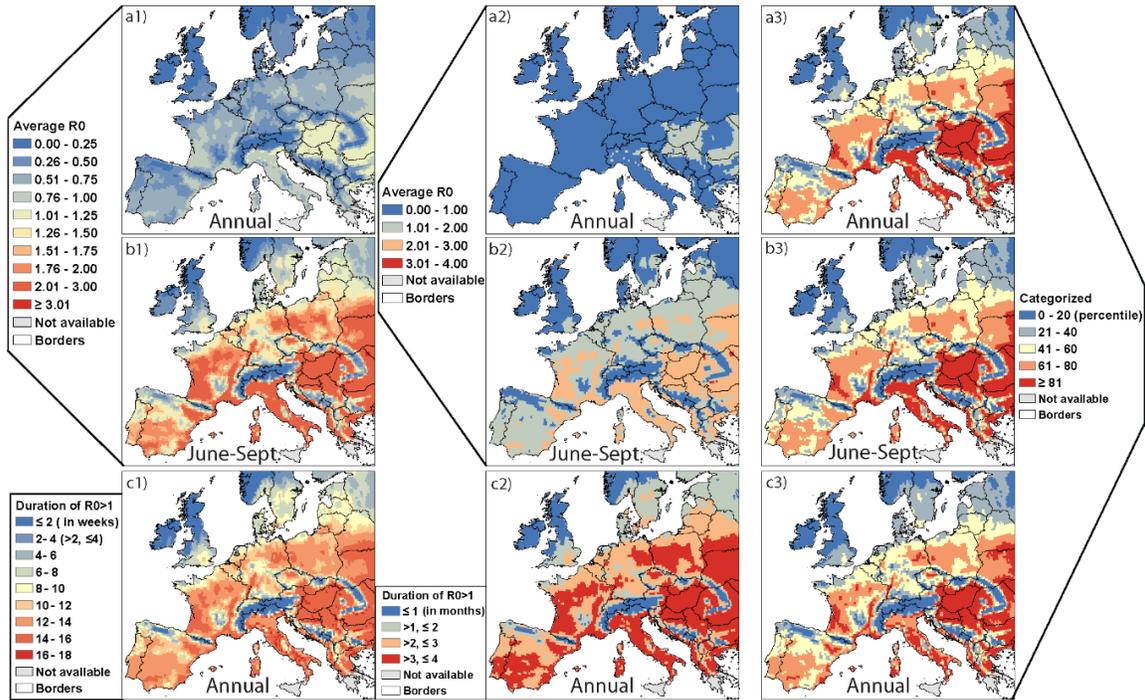
### ***Temporal aggregation***

The averaged  $R_0$  maps (Figure 3 a, b, c) show the intensity of potential risk averaged over a certain time period. The annual average  $R_0$  map (Figure 3 a)) portrays the spatial risk pattern, without any temporal information on the potential start of the transmission season or when during the year the risk is particularly high. The seasonal average  $R_0$  map (Figure 3 b)) mainly shows the spatial risk pattern as well, although with a larger  $R_0$  value range compared with the annual map. As for temporal risk information, the seasonal average  $R_0$  map has a defined transmission season. The series of monthly average  $R_0$  maps (Figure 3 c1) - c4)) captures the spatio-temporal risk dynamic, and shows a sudden decline of risk at the end of the transmission season. The duration map (Figure 3 d)) shows the duration of the at-risk state by means of the total number of days with  $R_0 > 1$ . In this example, most parts of Europe show more than 10 weeks of  $R_0 > 1$ .

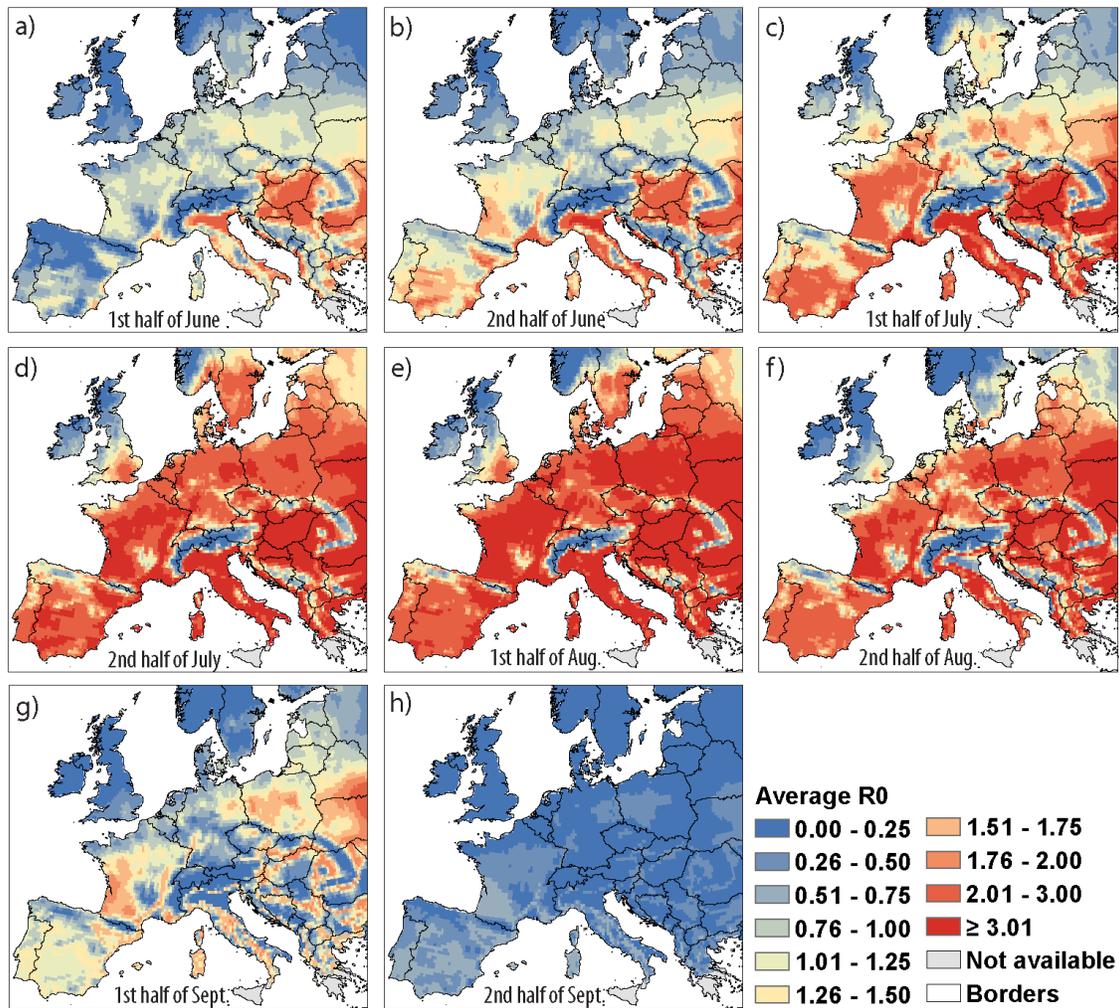
### ***Visualization***

We varied visualization settings of Figure 3 a), b) and d), as shown in Figure 4. With certain visualization setting, some maps can show very similar spatial patterns, e.g. Figure 4 b1) and c1). However, just by changing the value interval, the spatial patterns can certainly differ, e.g. Figure 4 b2) and c2). On the other hand, maps with different spatial patterns under certain visualization settings, e.g. Figure 4 a2), b2) and c2), can also show similar spatial patterns, as seen in Figure 4 a3), b3) and c3).

In addition to the monthly average  $R_0$  maps over the transmission season (Figure 3 c)), more detailed spatio-temporal risk maps were carried out by using biweekly time-steps (Figure 5). Using finer temporal resolution, these maps capture not only the spatial risk patterns, but also identify the peak of risk. The overall risk across Europe gradually increases from the beginning of June until the first half of August, which is when it begins to drop. Until the second half of September (Figure 5 h)) the average  $R_0$  across Europe is below 1.



**Figure 4.** Effects of different visualization methods on the final presentation of the map. All maps were derived from the same NGM-based epidemiological model for Usutu virus in Europe, based on gridded daily temperature observations for 2018. Three rows show a) annual average  $R_0$ , b) average  $R_0$  over the transmission season, June-Sept., and c) duration of conditions with  $R_0 > 1$ . Three columns show classification methods: 1) equal intervals method with fine value intervals and a large number of classes, 2) equal intervals method with coarse value intervals and a low number of classes and 3) percentile-based classification.



**Figure 5.** Biweekly average  $R_0$  over the transmission season, June – September. All maps were derived from the same NGM-based epidemiological model for Usutu virus in Europe, based on gridded daily temperature observations for 2018.

## Discussion

The ultimate purpose of generating risk maps is to show where and when the potential risk of a VBD outbreak is high. To achieve this, it is essential to apply appropriate and practical temporal aggregation methods when producing these maps with EMs. In this review, we provide a detailed comparison of the commonly used aggregation methods, and demonstrate their use with an EM for Usutu virus. In addition, we highlight the importance of reasonable visualizations.

### *Methods for temporal aggregation of $R_0$*

When transforming the raw EM outputs into spatial risk maps, the purpose of these risk maps needs to be considered. In order to monitor the general spatial risk patterns of a VBD, the relative risk

might be more important than the absolute value of  $R_0$ . In this case, long-term (annual or even longer) or mid-term (transmission season) average  $R_0$  risk maps and the total number of days with  $R_0 > 1$  can work well. However, to take preventative measures to stop VBD emergence, it is essential to know both the spatial and temporal patterns of transmission risk. For this purpose, a series of short-term average  $R_0$  maps is preferable. For all these applications, it should be borne in mind that  $R_0$  is calculated under the assumption of a completely susceptible population, i.e. for the early stages of a disease event. Furthermore, in the EMs discussed here,  $R_0$  is calculated for each cell of the spatial raster individually; i.e. there is no simulated spread from one cell to another. Consequently, these maps can only be interpreted as a potential for an outbreak to occur. In order to predict the actual course and magnitude of an outbreak, much more sophisticated models would be needed.

In the majority of the articles reviewed here, the authors aggregate  $R_0$  in time by calculating average values over various periods. When applying this method, the primary task is to find a reasonable time period to investigate. It is important to note here that while  $R_0$  itself is a threshold quantity, *average*  $R_0$  is not. While high average  $R_0$  (qualify  $R_0 > 1$ ) indicates high risk, low average  $R_0$  does not necessarily mean that transmission cannot occur. Even if the mean  $R_0$  over a certain period is below 1, there could be several consecutive days or weeks with  $R_0 > 1$  which may lead to an outbreak. This is especially problematic for long-term or mid-term average  $R_0$  maps. Hence, maps based on longer-term average values of  $R_0$  are not useful for determining the presence or absence of risk (let alone absolute risk) for ongoing transmission at a given place in the study area. They can, however, be useful to estimate the relative spatial differences in risk across the study area (e.g. Figure 3a) and b)). The short-term (monthly) average  $R_0$  maps provide temporal information together with the spatial risk maps, capturing the spatio-temporal risk dynamics (e.g. Figure 3 b), c)). With even smaller time steps, averaged  $R_0$  values may still be able to reflect the absolute values of  $R_0$  reasonably well. We use a two-week time step as an approximation of the extrinsic incubation period (EIP) of Usutu virus in Figure 5. This kind of map can be useful to show the beginning and end of a high risk season. In any case, it needs to be kept in mind that absolute values of  $R_0$  calculated by different methods are not comparable (Li et al., 2011), which also applies to average values derived from them. For  $R_{0\text{ NGM}}$  calculated by EMs using the popular NGM method, there is the additional complication that the daily  $R_{0\text{ NGM}}$  is already the geometric mean per generation (Dietz, 1993). When averaging the daily  $R_{0\text{ NGM}}$  using the arithmetic mean, it has the potential to show a distorted picture of the real-world situation. Additionally,  $R_{0\text{ NGM}}$  indicates whether or not the intrinsic growth rate  $r$  of the investigated VBD is positive, but the relationship between  $R_0$  and  $r$  is not explicit and differs depending on the respective NGM (Diekmann et al., 2010; Dietz, 1993). Therefore, the average  $R_{0\text{ NGM}}$  maps should be interpreted with special caution, and those from different NGMs are not comparable.

The less frequently applied accumulation-based methods do not treat the threshold quantity  $R_0$  as a continuous value. Instead, they show the spatial risk pattern as the duration of at-risk states (such as number of days with  $R_0 > 1$ ). This method leads to results that are comparable across different

model types (including EMs built with different NGM) as long as they share the threshold at  $R_0 = 1$ . The drawback of this method is that it does not take into account the intensity of risk: a day with  $R_0 = 1.1$  and a day with  $R_0 = 111$  would both be equally counted as one day of  $R_0 > 1$ , though higher  $R_0$  values generally indicate higher theoretical risk (Diekmann et al., 2010). In certain cases, a period with  $R_0$  being continuously above 1 could last for months across the whole study area, e.g. Figure 3 d). However, for disease control measures, it is also important to know where the potentially most-severely affected areas will be. Since this method ignores the quantity of  $R_0$ , valuable information is lost. Consequently, solely relying on this method would not be sufficient in many scenarios.

The average  $R_0$  maps and the maps showing the total number of days with  $R_0 > 1$  are not comparable in terms of risk, as one shows risk intensity while the other shows duration. In our test case, these two types of maps show similar patterns (Figure 3 b) and d)), but this will not always be the case (see Figure 4 b2) and c2)). These two types of maps should not be interpreted the same way. For instance, the high correlation between the transmission season average  $R_0$  map and the total number of days with  $R_0 > 1$  map indicates that where average  $R_0$  is higher, the duration of risk is also longer, or vice versa. This does not necessarily mean one can replace or resemble the other. As the EMs are typically driven by temperature variables, it is not surprising that the average  $R_0$  is correlated to risk duration, i.e. a long duration of high temperature would probably result in both a high average  $R_0$  and a longer transmission season. Allowing for other parameters (such as human/vector population density) to vary in space as well may break this perceived pattern.

### ***Visualization methods***

In the vast majority of the reviewed studies, the authors opted for a classed map rather than one showing continuous values. This is in line with common practice in geography, where classed maps are often preferred based on the argument that the reader would not be able to precisely locate a pixel's color value on a continuous color bar in the legend anyway (Slocum et al., 2009 ). A classed map thus provides a simplification that helps the reader to estimate which numerical values an area of a certain color represents. Given that several of the articles reviewed here have classed maps with 10 or more classes, it seems worth pointing out that this effect is lost if the value intervals are too small. This leads to a large number of classes that cannot be distinguished visually any more (Figure 4 a1), b1), c1)), making these maps are visually equivalent to an unclassed map. However, they do not share their main advantage of each color value in the map representing only one unique data value. If a classed map is being produced, the number of classes and the color scheme of the legend should be chosen in a way that enables the reader to easily distinguish the different classes.

Modern map making software supports a variety of methods that can be used to classify different types of data for display in a map (Slocum et al., 2009 ). The most straight-forward option for aggregated  $R_0$  data is the classic "equal intervals" approach, where the range of values is broken up into value intervals of equal size. It is easy to calculate and the legend can be understood intuitively. However, the chosen size of the intervals (or: number of classes) strongly affects how the presented data is being perceived by the reader. For instance, Figure 4 b1) and c1) show two different

temporal aggregates of the same model (transmission season average and duration of the conditions with  $R_0 > 1$ , respectively). They are being presented using equal intervals with a large number of classes and are perceived as showing very similar spatial patterns. Figure 4 b2) and c2) show the same data again, using the same equal intervals methods but with less classes. Due to how the values are distributed in the two data sets, the perceived patterns are now dramatically different. In contrast to this Figure 4 a1) and b1) show an average of  $R_0$  over the whole year and the transmission season respectively. Using the exact same value intervals for both maps, the different patterns are clearly visible. The same is true for the equivalent maps using a lower number of classes (Figure 4 a2) and b2)). Another challenge is to define reasonable limits for the area to be investigated with this method, as the geographical extent affects the value range and value distribution of the pixels within the extent, which will in turn affect the classification of the pixels. The equal intervals approach can thus be useful to compare average  $R_0$  for different time periods within the same model (e.g. annual vs. transmission season or July vs. August). Its use for comparisons across temporal transformation methods or between models from different studies is severely limited, though.

In order to enable comparisons across different models, quantile-based classification should be considered as a useful alternative to the equal intervals method. Here, the data values are first ranked and then grouped so that an equal number of data points falls into each class (Slocum et al., 2009 ). In our example, the annual and transmission season average of  $R_0$  as well as the duration of the conditions with  $R_0 > 1$  were classified into five quantiles, expressed as percentiles in the legend (Figure 4 a3), b3) and c3)). Each of them shows an estimate of relative risk across the study region. Through that, comparisons across different studies using different methods are being made possible. Similarly, this method could enable a cross-disease overview, i.e. though the absolute values of  $R_0$  for different VBD are not comparable, it is still helpful to know which VBD is more likely to occur in a certain area. In our test case, the two classed risk maps (Figure 4 a3) and b3)) show virtually identical patterns, which can be explained by the general correlation between the annual mean temperature and the mean temperature over the warmest quarter (transmission season). Even the duration based map (Figure 4 a3), b3) and c3)) looks very similar. However, this does not mean that one map can replace the other. Those two maps still cover different time-periods and have different interpretations. For average  $R_0$  over transmission season, a transmission season is defined. It should also be noted that high relative risk does not necessarily mean high definite risk. For instance, were the map of average  $R_0$  over the second half of September (Figure 5 h) transformed into a quantile-based risk map, all values within the highest quantile would indeed still be below 1. For the sake of clarity, the data values corresponding to the class limits should thus be noted in the legend, along with or even replacing the percentile labels.

When the quantiles method or any other method that does not lead to regular value intervals (such as “natural breaks” or “optimal”, (Slocum et al., 2009 ) is used, it is important to report how the classification was achieved. Not only would the reproducibility of the study be limited – under certain circumstances, the wary reader may even suspect a map to be manipulated (Monmonier,

1991). Only when their thresholds are reasonable, the risk maps are interpretable. Unfortunately, several of the studies reviewed here are incomplete in this regard.

### ***Temporal dynamics***

The temporal resolution of risk maps is typically not a free choice. Instead, it is very often dictated by the input data. For instance, with daily resolution temperature data as an input, it is easy to produce maps that show the duration of conditions with  $R_0 > 1$  as well as and risk maps with small time steps. There are plenty of options to aggregate fine temporal resolution to coarse ones. However, this does not work in the opposite direction. When the input data is annual average temperature, it is not possible to produce duration maps, nor a seasonal averaged intensity of risk. Interpolating monthly data into daily values may work if the goal is a map of long-term average  $R_0$  (Calistri et al., 2016; Zhang et al., 2017), but it is certainly not a useful method for estimating the number of days with  $R_0 > 1$ .

### ***Summary and suggestions***

In order to prepare risk maps for potential outbreaks from the raw output of EMs, two main steps are required. First, one or more appropriate methods for temporal aggregation of  $R_0$  values need to be chosen. For a summarizing spatial overview, both maps of duration and average long-term intensity of risk can be useful. Strictly speaking, however, only the duration-based approach leads to results that are comparable across different types of EMs. This method also has a rather straight-forward interpretation, making it the best choice for predicting areas of potential disease emergence in non-endemic regions so that preventive measure can be taken. A series of maps based on short-term average  $R_0$ , on the other hand, can be useful to determine the time and place of peak transmission potential within endemic areas.

The second important, but often neglected aspect is how the maps are visualized. Unclassed maps avoid any loss in information, but are often difficult to interpret due to the lack of abstraction. By using different classification methods with the same data, the resulting maps can show very different patterns. Among the methods available, the equal intervals approach is the obvious straight-forward solution. However, quantile-based classification approach results in more meaningful maps and better comparability across different studies. When classification methods other than equal intervals are used, this needs to be documented properly, if only to avoid the suspicion of purposeful manipulation.

Ultimately, there is always a trade-off to be made between minimizing information loss, facilitating understanding and minimizing the amount of maps. If in doubt, we recommend to combine 1) an overview map of the duration of conditions with  $R_0 > 1$ , visualized using quantile-based classification with a 2) series of maps showing the change of short-term (monthly, bi-weekly) average  $R_0$  throughout the transmission season, visualized using the same set of equal intervals for the entire set. Preferably, this would be accompanied by a digital supplement containing the (aggregated) data stored in a commonly used geodata format such as GeoTIFF or NetCDF.

## Acknowledgement

We would like to thank Taylor Blair for the proof reading. We acknowledge the E-OBS dataset from the EU-FP6 project ENSEMBLES (<http://ensembles-eu.metoffice.com>) and the data providers in the ECA&D project (<http://www.ecad.eu>).

## Author contributions

YC, NT conceptualized the review. YC did the literature searches, wrote the first draft, ran the epidemiological models, drafted the table and prepared the visualizations. NT had the initial idea for the review, co-wrote the results section, prepared parts of Figure 2 and edited the table and the manuscript. ST, AJ and CB reviewed and provided critical feedback on the draft. All authors discussed the concept, preliminary results and figures at various stages. All authors discussed and revised the manuscript. All authors read and approved the final version of the manuscript.

**Funding Statement:** Yanchao Cheng was funded by China Scholarship Council, No. 201506040059. Stephanie Thomas was funded by the Bavarian State Ministry of the Environment and Consumer Protection and the Bavarian State Ministry of Health and Care through the BayVirMos project (TKP 01KPB-73560). This publication was funded by the German Research Foundation (DFG) and the University of Bayreuth in the funding programme Open Access Publishing.

## References

- Brewer, C. A., 1994. Color Use Guidelines for Mapping and Visualization. In A. M. MacEachren & D. R. Fraser Taylor (Eds.), *Visualization in Modern Cartography* (pp. 123-147). Pergamon.
- Brugger, K., Rubel, F., 2013. Bluetongue disease risk assessment based on observed and projected *Culicoides obsoletus* spp. vector densities. *PLoS ONE*, 8, e60330.
- Cadar, D., Lühken, R., van der Jeugd, H., Garigliany, M., Ziegler, U., Keller, M., Lahoreau, J., Lachmann, L., Becker, N., Kik, M., Munnink, B. B. O., Bosch, S., Tannich, E., Linden, A., Schmidt, V., Koopmans, M. P., Rijks, J., Desmecht, D., Groschup, M. H., Reusken, C., Schmidt-Chanasit, J., 2017. Widespread activity of multiple lineages of Usutu virus, western Europe, 2016. *Eurosurveillance*, 22, 11-17.
- Calistri, P., Savini, L., Candeloro, L., Di Sabatino, D., Cito, F., Bruno, R., Danzetta, M. L., 2016. A transitional model for the evaluation of west Nile virus transmission in Italy. *Transboundary and Emerging Diseases*, 63, 485-496.
- Caminade, C., McIntyre, K. M., Jones, A. E., 2019. Impact of recent and future climate change on vector-borne diseases. *Annals of the New York Academy of Sciences*, 1436, 157-173.
- Cheng, Y. C., Tjaden, N. B., Jaeschke, A., Lühken, R., Ziegler, U., Thomas, S. M., Beierkuhnlein, C., 2018. Evaluating the risk for Usutu virus circulation in Europe: comparison of environmental niche models and epidemiological models. *International Journal of Health Geographics*, 17.
- Clarivate Analytics, 2019. Web of Science. Retrieved from <https://www.webofknowledge.com/>
- Cordovez, J. M., Rendon, L. M., Gonzalez, C., Guhl, F., 2014. Using the basic reproduction number to assess the effects of climate change in the risk of Chagas disease transmission in Colombia. *Acta Tropica*, 129, 74-82.
- Delamater, P. L., Street, E. J., Leslie, T. F., Yang, Y. T., Jacobsen, K. H., 2019. Complexity of the basic reproduction number ( $R_0$ ). *Emerging Infectious Diseases*, 25, 1-4.

- Diekmann, O., Heesterbeek, J. A. P., Metz, J. A. J., 1990. On the definition and the computation of the basic reproduction ratio  $R_0$  in models for infectious-diseases in heterogeneous populations. *Journal of Mathematical Biology*, 28, 365-382.
- Diekmann, O., Heesterbeek, J. A. P., Roberts, M. G., 2010. The construction of next-generation matrices for compartmental epidemic models. *Journal of the Royal Society, Interface*, 7, 873-885.
- Dietz, K., 1993. The estimation of the basic reproduction number for infectious diseases. *Statistical Methods in Medical Research*, 2, 23-41.
- Du, Z. C., Zhang, W. J., Zhang, D. M., Yu, S. C., Hao, Y. T., 2017. Estimating the basic reproduction rate of HFMD using the time series SIR model in Guangdong, China. *PLoS ONE*, 12, e0179623.
- ECDC, 2019. Vector-borne diseases. Retrieved from <https://ecdc.europa.eu/en/climate-change/climate-change-europe/vector-borne-diseases>
- Escobar, L. E., Craft, M. E., 2016. Advances and limitations of disease biogeography using ecological niche modeling. *Frontiers in Microbiology*, 7, 1174.
- Ferguson, N. M., Cucunuba, Z. M., Dorigatti, I., Nedjati-Gilani, G. L., Donnelly, C. A., Basanez, M. G., Nouvellet, P., Lessler, J., 2016. Countering the Zika epidemic in Latin America. *Science*, 353, 353-354.
- Gage, K. L., Burkot, T. R., Eisen, R. J., Hayes, E. B., 2008. Climate and Vectorborne Diseases. *American Journal of Preventive Medicine*, 35, 436-450.
- Guerra, F. M., Bolotin, S., Lim, G., Heffernan, J., Deeks, S. L., Li, Y., Crowcroft, N. S., 2017. The basic reproduction number ( $R_0$ ) of measles: a systematic review. *Lancet Infectious Diseases*, 17, e420-e428.
- Hartemink, N. A., Purse, B. V., Meiswinkel, R., Brown, H. E., de Koeijer, A., Elbers, A. R. W., Boender, G. J., Rogers, D. J., Heesterbeek, J. A. P., 2009. Mapping the basic reproduction number ( $R_0$ ) for vector-borne diseases: A case study on bluetongue virus. *Epidemics*, 1, 153-161.
- Hartemink, N. A., Vanwambeke, S. O., Heesterbeek, H., Rogers, D., Morley, D., Pesson, B., Davies, C., Mahamdallie, S., Ready, P., 2011. Integrated mapping of establishment risk for emerging vector-borne infections: A case study of canine leishmaniasis in southwest France. *PLoS ONE*, 6, e20817.
- Hartley, D. M., Barker, C. M., Le Menach, A., Niu, T. C., Gaff, H. D., Reisen, W. K., 2012. Effects of temperature on emergence and seasonality of west Nile virus in California. *American Journal of Tropical Medicine and Hygiene*, 86, 884-894.
- Haylock, M. R., Hofstra, N., Tank, A. M. G. K., Klok, E. J., Jones, P. D., New, M., 2008. A European daily high-resolution gridded data set of surface temperature and precipitation for 1950-2006. *Journal of Geophysical Research-Atmospheres*, 113.
- Heffernan, J. M., Smith, R. J., Wahl, L. M., 2005. Perspectives on the basic reproductive ratio. *Journal of the Royal Society, Interface*, 2, 281-293.
- Hethcote, H. W., 2000. The mathematics of infectious diseases. *SIAM Review*, 42, 599-653.
- Holy, M., Schmidt, G., Schroder, W., 2011. Potential malaria outbreak in Germany due to climate warming: Risk modelling based on temperature measurements and regional climate models. *Environmental Science and Pollution Research International*, 18, 428-435.
- Kakmeni, F. M. M., Guimapi, R. Y. A., Ndjomatchoua, F. T., Pedro, S. A., Mutunga, J., Tonnang, H. E. Z., 2018. Spatial panorama of malaria prevalence in Africa under climate change and interventions scenarios. *International Journal of Health Geographics*, 17.
- Li, J., Blakeley, D., Smith, R. J., 2011. The Failure of  $R_0$ . *Computational and Mathematical Methods in Medicine*, 2011, Article ID 527610.
- Liu, Q. H., Ajelli, M., Aleta, A., Merler, S., Moreno, Y., Vespignani, A., 2018. Measurability of the epidemic reproduction number in data-driven contact networks. *Proceedings of the National Academy of Sciences of the United States of America*, 115, 12680-12685.
- Ma, Z., Li, J. (Eds.). (2009). *Dynamical modeling and analysis of epidemics*. Singapore: World Scientific Publishing.
- Marques, C. A., Forattini, O. P., Massad, E., 1994. The basic reproduction number for dengue fever in Sao Paulo state, Brazil - 1990-1991 epidemic. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 88, 58-59.
- Martcheva, M., 2015. *An Introduction to Mathematical Epidemiology*. New York, Heidelberg, Dordrecht, London: Springer.

- Massad, E., Coutinho, F. A. B., Burattini, M. N., Lopez, L. F., 2001. The risk of yellow fever in a dengue-infested area. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 95, 370-374.
- Monmonier, M., 1991. *How to lie with maps*. Chicago and London: The University of Chicago Press.
- Moraga, P., Cano, J., Baggaley, R. F., Gyapong, J. O., Njenga, S. M., Nikolay, B., Davies, E., Rebollo, M. P., Pullan, R. L., Bockarie, M. J., Hollingsworth, T. D., Gambhir, M., Brooker, S. J., 2015. Modelling the distribution and transmission intensity of lymphatic filariasis in sub-Saharan Africa prior to scaling up interventions: Integrated use of geostatistical and mathematical modelling. *Parasites & Vectors*, 8, 560.
- Mordecai, E. A., Cohen, J. M., Evans, M. V., Gudapati, P., Johnson, L. R., Lippi, C. A., Miazgowiec, K., Murdock, C. C., Rohr, J. R., Ryan, S. J., Savage, V., Shocket, M. S., Ibarra, A. S., Thomas, M. B., Weikel, D. P., 2017. Detecting the impact of temperature on transmission of Zika, dengue, and chikungunya using mechanistic models. *PLoS Neglected Tropical Diseases*, 11, e0005568.
- Myers, M. F., Rogers, D. J., Cox, J., Flahault, A., Hay, S. I., 2000. Forecasting disease risk for increased epidemic preparedness in public health. *Advances in Parasitology*, 47, 309-330.
- Ng, V., Fazil, A., Gachon, P., Deuymes, G., Radojević, M., Mascarenhas, M., Garasia, S., Johansson, M. A., Ogden, N. H., 2017. Assessment of the probability of autochthonous transmission of chikungunya virus in Canada under recent and projected climate change. *Environmental Health Perspectives*, 125, 067001.
- Peterson, A. T., 2014. *Mapping disease transmission risk: enriching models using biogeography and ecology*. Baltimore: Johns Hopkins University Press.
- Racloz, V., Venter, G., Griot, C., Stark, K. D. C., 2008. Estimating the temporal and spatial risk of bluetongue related to the incursion of infected vectors into Switzerland. *BMC Veterinary Research*, 4, 42.
- Ridenhour, B., Kowalik, J. M., Shay, D. K., 2014. Unraveling  $R_0$ : Considerations for public health applications. *American Journal of Public Health*, 104, E32-E41.
- Ridenhour, B., Kowalik, J. M., Shay, D. K., 2018. The basic reproductive number ( $R_0$ ): considerations for its application in public health. *American Journal of Public Health*, 108, S455-S465.
- Rubel, F., Brugger, K., Hantel, M., Chvala-Mannsberger, S., Bakonyi, T., Weissenböck, H., Nowotny, N., 2008. Explaining Usutu virus dynamics in Austria: Model development and calibration. *Preventive Veterinary Medicine*, 85, 166-186.
- Siettos, C. I., Russo, L., 2013. Mathematical modeling of infectious disease dynamics. *Virulence*, 4, 295-306.
- Slocum, T. A., McMaster, R. B., Kessler, F. C., Howard, H. H., 2009 Data classification. In T. A. Slocum, R. B. McMaster, F. C. Kessler, & H. H. Howard (Eds.), *Thematic cartography and geovisualization* (3<sup>rd</sup> ed., pp. 57-75): Pearson.
- Smith, D. L., Perkins, T. A., Reiner, R. C., Barker, C. M., Niu, T. C., Chaves, L. F., Ellis, A. M., George, D. B., Le Menach, A., Pulliam, J. R. C., Bisanzio, D., Buckee, C., Chiyaka, C., Cummings, D. A. T., Garcia, A. J., Gattton, M. L., Gething, P. W., Hartley, D. M., Johnston, G., Klein, E. Y., Michael, E., Lloyd, A. L., Pigott, D. M., Reisen, W. K., Ruktanonchai, N., Singh, B. K., Stoller, J., Tatem, A. J., Kitron, U., Godfray, H. C. J., Cohen, J. M., Hay, S. I., Scott, T. W., 2014. Recasting the theory of mosquito-borne pathogen transmission dynamics and control. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 108, 185-197.
- Taffese, H. S., Hemming-Schroeder, E., Koepfli, C., Tesfaye, G., Lee, M. C., Kazura, J., Yan, G. Y., Zhou, G. F., 2018. Malaria epidemiology and interventions in Ethiopia from 2001 to 2016. *Infectious Diseases of Poverty*, 7, 103.
- Tatem, A. J., Rogers, D. J., Hay, S. I., 2006. Global transport networks and infectious disease spread. *Advances in Parasitology*, 62, 293-343.
- Thomas, S. M., Tjaden, N. B., van den Bos, S., Beierkuhnlein, C., 2014. Implementing cargo movement into climate based risk assessment of vector-borne diseases. *International Journal of Environmental Research and Public Health*, 11, 3360-3374.
- Tjaden, N. B., Caminade, C., Beierkuhnlein, C., Thomas, S. M., 2018. Mosquito-borne diseases: Advances in modelling climate-change impacts. *Trends in Parasitology*, 34, 227-245.
- WHO, 2017. Vector-borne diseases. Retrieved from <https://www.who.int/news-room/fact-sheets/detail/vector-borne-diseases>

- Wu, X. T., Duvvuri, V. R., Lou, Y. J., Ogden, N. H., Pelcat, Y., Wu, J. H., 2013. Developing a temperature-driven map of the basic reproductive number of the emerging tick vector of Lyme disease *Ixodes scapularis* in Canada. *Journal of Theoretical Biology*, 319, 50-61.
- Yang, H. M., 2014. The basic reproduction number obtained from Jacobian and next generation matrices - A case study of dengue transmission modelling. *Biosystems*, 126, 52-75.
- Zhang, Q., Sun, K. Y., Chinazzi, M., Piontti, A. P. Y., Dean, N. E., Rojas, D. P., Merler, S., Mistry, D., Poletti, P., Rossi, L., Bray, M., Halloran, M. E., Longini, I. M., Vespignani, A., 2017. Spread of Zika virus in the Americas. *Proceedings of the National Academy of Sciences of the United States of America*, 114, E4334-E4343.

## Supplement S1

### Introduction

In this study, a Next Generation Matrix (NGM) based Epidemiological Model (EM) was applied to test the transformation method recorded. The model was originally built by (Rubel et al., 2008) and adapted by (Cheng et al., 2018) for an interdisciplinary model comparison approach.

This model concerns a vector-borne disease caused by a *flavivirus*, Usutu virus. The main vectors are *Culex* mosquitoes. The common hosts include birds, dogs, horses, bats, humans, but mostly black birds (*Turdus merula*). Consequently, in the model, only mosquitoes and black birds are taken into account.

### Model description

Each health state of mosquitoes / black birds can be described by Ordinary differential equations (ODEs).

Population growth of black birds:

$$\frac{dN_B}{dt} = r_B N_B = b_B N_B - m_B N_B$$

$N_B$  is the total number of black birds,  $r_B$  is the population growth rate,  $b_B$  is the birth rate and  $m_B$  is the mortality rate (B stands for black birds).

Follow logistic population growth (density dependent model):

$$\frac{dN_B}{dt} = r_B \left(1 - \frac{N_B}{K_B}\right) N_B$$

As

$$r_B = b_B - m_B$$

it can be written as:

$$\frac{dN_B}{dt} = \left(b_B - (b_B - m_B) \frac{N_B}{K_B}\right) N_B - m_B N_B$$

$K_B$  stands for environmental capacity. It can be understood as the maximum number of individuals that can be supported by the environment under ideal conditions.

The population of "larval" mosquitoes (includes all aquatic stages of *Culex* mosquitoes, only females taken into account) also follows logistic population growth:

$$\frac{dL_M}{dt} = (b_L N_M - m_L L_M) \left(1 - \frac{L_M}{K_M}\right) - b_M L_M$$

$L_M$  is the total number of larvae,  $b_L$  is the birth rate of larvae,  $m_L$  is the mortality rate of larvae,  $b_M$  is the "birth rate" of mosquitoes (transformation from larvae to adult mosquitoes). Note here: although also following logistic population growth, mosquito growth is divided to aquatic and terrestrial stages, thus the equation looks different from black birds'.

Total density of terrestrial stages of *Culex* mosquitoes ( $N_M$ ):

$$\frac{dN_M}{dt} = b_M L_M - m_M N_M$$

$m_M$  is the mortality rate of *Culex* mosquitoes.

Cross-infection between mosquitoes and black birds:

$$\lambda_B = \beta_B \frac{I_B}{K_B} = \kappa P_B \frac{I_B}{K_B}$$

$$\lambda_M = \beta_M \frac{I_M}{K_B} = \kappa P_M \frac{I_M}{K_B}$$

$\lambda_B$  denote the possible fraction of cross-transmission from birds to mosquitoes, and  $\lambda_M$  vice versa.  $\beta_B$  is the product of biting rate ( $\kappa$ ) and transmission possibility from birds to mosquitoes ( $P_B$ ), and  $\beta_M$  vice versa. Transmission possibility from mosquitoes to birds is  $P_M$ .

Then the different health states of birds can be described by following ODEs:

1. The susceptible black bird population ( $S_B$ )

$$\frac{dS_B}{dt} = \left( b_B - (b_B - m_B) \frac{N_B}{K_B} \right) N_B - m_B N_B - \delta_M \lambda_M S_B$$

2. The exposed black bird population ( $E_B$ )

$$\frac{dE_B}{dt} = \delta_M \lambda_M S_B - m_B E_B - \gamma_B E_B$$

$\delta_M$  : Percentage of non-hibernating mosquitoes

$\gamma_B$  : The exposed – infected/infectious rate of birds, and  $1/\gamma_B$  is the intrinsic-incubation period

3. The infected black bird population ( $I_B$ )

$$\frac{dI_B}{dt} = \gamma_B E_B - m_B I_B - \alpha_B I_B$$

$\alpha_B$ : The removal rate, removed from the previous health state, either get recovered (immunized) or dead

4. The black bird deaths ( $D_B$ ) due to USUV infection

$$\frac{dD_B}{dt} = \nu_B \alpha_B I_B$$

$\nu_B$ : the percentage of bird deaths due to USUV infection

5. The recovered black bird population ( $R_B$ )

$$\frac{dR_B}{dt} = (1 - \nu_B) \alpha_B I_B - m_B R_B$$

And

$$N_B = S_B + E_B + I_B + R_B$$

Note: In this model both horizontal and vertical virus transmission in birds are not taken into account, so the transmission is limited to through mosquitoes' blood meal.

Similarly, the different health states of mosquitoes are described as following:

6. The larval population of *Culex* mosquitoes:

$$\frac{dL_M}{dt} = (b_L N_M - m_L L_M) \left(1 - \frac{L_M}{K_M}\right) - b_M L_M$$

7. The susceptible mosquito population:

$$\frac{dS_M}{dt} = b_M L_M - m_M S_M - \delta_M \lambda_B S_M$$

8. The exposed mosquito population:

$$\frac{dE_M}{dt} = \delta_M \lambda_B S_M - m_M E_M - \gamma_M E_M$$

$\gamma_M$ : The exposed – infected/infectious rate of mosquitoes, and  $1/\gamma_M$  is the extrinsic-incubation period

9. The infected mosquito population:

$$\frac{dI_M}{dt} = \gamma_M E_M - m_M I_M$$

And

$$N_M = S_M + E_M + I_M$$

Note: Infectious mosquitoes remain in the infectious state and will not get recovered.

Following the **next generation matrix** method, the final  $R_0$  equation:

$$R_0 = \sqrt{\left[ \frac{\delta_M \gamma_M \beta_M}{(\gamma_M + m_M) m_M} \frac{S_B}{K_B} \right] \left[ \frac{\delta_M \gamma_B \beta_B}{(\gamma_B + m_B) (\alpha_B + m_B)} \frac{S_M}{K_B} \right]}$$

This  $R_0$  equation is indeed the **largest eigenvalue** of the respective next generation matrix.

### Parameters

See (Rubel et al., 2008).

### References

- Cheng, Y. C., Tjaden, N. B., Jaeschke, A., Luhken, R., Ziegler, U., Thomas, S. M., Beierkuhnlein, C., 2018. Evaluating the risk for Usutu virus circulation in Europe: comparison of environmental niche models and epidemiological models. *International Journal of Health Geographics*, 17.
- Rubel, F., Brugger, K., Hantel, M., Chvala-Mannsberger, S., Bakonyi, T., Weissenböck, H., Nowotny, N., 2008. Explaining Usutu virus dynamics in Austria: Model development and calibration. *Preventive Veterinary Medicine*, 85, 166-186.

**Supplement S2: Overview of the articles reviewed.**

Map characteristics, Visualization	Model Type	Data source (temperature), description	VBD, Country/ Region	References
<b>Intensity -- Average <math>R_0</math> over time</b>				
Spatial reference: raster, 4 km × 4 km cells Temporal reference: annual (multi-year average) Color scheme: 5 classes (interval: 1), diverging color gradient from green ( $R_0 < 1$ ) to red ( $R_0 > 1$ )	NGM	Monthly mean temperature from 30 meteorological stations for 1971–2000. $R_0$ values per station derived from mean annual number of degree-days > 0°C and interpolated spatially.	Lyme disease, Canada	Wu et al., 2013
Spatial reference: raster, unknown resolution Temporal reference: annual (multi-year average) Color scheme: continuous color ramp, diverging color gradient from white ( $R_0 = 1$ ) – yellow – red – black ( $R_0 = 7$ )	NGM	Annual mean temperature for current and future climatic conditions from <a href="https://worldclim.org">https://worldclim.org</a> .	Chagas disease, Colombia	Cordovez et al., 2014
Spatial reference: raster, 2.5' × 2.5' cells Temporal reference: annual (multi-year average) Color scheme: continuous color ramp, diverging color gradient from green ( $R_0 = 0$ ) to red ( $R_0 = 9$ )	NGM	Monthly average, minimum and maximum temperatures for current and future climatic conditions from <a href="https://worldclim.org">https://worldclim.org</a> and <a href="http://ccafs-climate.org">http://ccafs-climate.org</a> .	Malaria, Africa	Kakmeni et al., 2018
Spatial reference: raster, unknown resolution Temporal reference: annual (multi-year average) Color scheme: 8 classes (interval: 0.2) from yellow to dark red	Probabilistic	Current rainfall and temperature profiles for Tanzania and the predictions of HadCM3 for 2080 under A2a and B2a emission scenarios (data from WorldClim 2009).	Malaria, Tanzania	Parham & Michael, 2010
Spatial reference: raster, 4 km × 4 km cells Temporal reference: annual (multi-year average) Color scheme: 14 classes (for $R_0 > 1$ only, value interval: 1), diverging color gradient from dark blue to red	NGM	Daily temperature from ANUSPLIN observations 1971–2000 and projected climate from the CRCM4.2.3 regional climate model for 2011–2040 and 2041–2070.	Lyme disease, North America	Ogden et al., 2014

<p>Spatial reference: raster, 0.25° × 0.25° cells  Temporal reference: annual (multi-year average)  Color scheme: 12 classes, multi-color gradient blue-green-yellow-red</p>	NGM	<p>Temperature and rainfall data from E-OBS (1961–2008, <a href="http://www.ecad.eu">http://www.ecad.eu</a>) and two ensembles of climate model simulations from ENSEMBLES RT3 (future, <a href="http://ensemblesrt3.dmi.dk/">http://ensemblesrt3.dmi.dk/</a>).</p>	<p>Bluetongue disease, Europe</p>	<p>Guis et al., 2012</p>
<p>Spatial resolution: raster, 10'  Temporal reference: annual (multi-year average)  Color scheme: R<sub>0</sub> &lt; 1 in white, R<sub>0</sub> ≥ 1 continuous diverging multi-color gradient from dark green to yellow</p>	NGM	<p>Daily estimates of temperature and water vapour for 1961–1990 from the New LocClim software converted to 10-day average values.</p>	<p>Crimean–Congo haemorrhagic fever, western Palearctic</p>	<p>Estrada-Pena et al., 2013</p>
<p>Spatial resolution: raster, 1°×1°  Temporal reference: annual (multi-year average)  Color scheme: R<sub>0</sub> &lt; 1 NA, R<sub>0</sub> ≥ 1 green to dark red, continuous color scheme</p>	NGM	<p>Monthly temperature and rainfall. Observed (1950–2015): CRU v3.4 (0.5° × 0.5° resolution). Future: North American Multi-Model Ensemble Project (NMME, 1°×1° resolution).</p>	<p>Zika, Latin America and the Caribbean</p>	<p>Munoz et al., 2017</p>
<p>Spatial reference: raster, 1° × 1° cells  Temporal reference: seasonal average (spring, summer, autumn, winter)  Color scheme: continuous color ramp, diverging color gradient from purple (R<sub>0</sub> = 0) to red (R<sub>0</sub> = 0.9)</p>	NGM	<p>Daily rainfall, minimum and maximum temperature (2002 – 2004) from the South African Weather Service (SAWS).</p>	<p>Malaria, South Africa</p>	<p>Abiodun et al., 2018</p>
<p>Spatial reference: raster, 0.25° × 0.25° cells  Temporal reference: seasonal average (June – September)  Color scheme: 11 classes (interval: 0.25), diverging color gradient from white (R<sub>0</sub> = 0) – green – red (R<sub>0</sub> = 2.75)</p>	NGM	<p>Daily mean temperature from the E-OBS dataset (<a href="https://www.ecad.eu">https://www.ecad.eu</a>, interpolated from weather stations).</p>	<p>Usutu disease, Western Europe</p>	<p>Cadar et al., 2017</p>
<p>Spatial reference: a) 107 locations of sampling points on a 20 km × 20 km grid, b) polygons of postal code areas, c) 21 locations of sampling points  Temporal reference: a) 1 month, b) multi-year average, c) monthly averages  Color scheme: light (R<sub>0</sub> &lt; 1) and dark (R<sub>0</sub> &gt; 1) color for point locations; 5 classes (increasing interval sizes) with sequential single-hue color gradient for polygons</p>	NGM	<p>Monthly temperature data from 10 Dutch meteorological stations for point locations. For areal maps, time series of MODIS Land Surface Temperatures and vegetation indices were used for spatial estimates of vector species' abundance.</p>	<p>Bluetongue disease, Netherlands</p>	<p>Hartemink et al., 2009</p>

<p>Spatial resolution: raster, 2.5'</p> <p>Temporal reference: monthly (multi-year averages)</p> <p>Color scheme: continuous diverging multi-color gradient blue-yellow-red</p>	NGM	<p>Monthly and daily mean temperatures from PRISM Climate Group as 30-year (1981–2010) climatic normals.</p>	<p>Bluetongue disease, California, USA</p>	<p>Mayo et al., 2016</p>
<p>Spatial reference: raster, unknown resolution</p> <p>Temporal reference: January and July (multi-year average)</p> <p>Color scheme: continuous double gradient: <math>R_0 &lt; 1</math> in shades of gray, <math>R_0 \geq 1</math> yellow to orange</p>	Probabilistic	<p>Daily climate data for all cities from long-term climate models created by the Climate Research Unit of East Anglia University, United Kingdom.</p>	<p>Yellow fever, global</p>	<p>Johansson et al., 2012</p>
<p>Spatial resolution: point locations of mosquito traps (710 locations for the year 2012 and 183 for 2013) with kernel density estimates</p> <p>Temporal reference: monthly mean (averaged from April to November)</p> <p>Color scheme: binary, <math>R_0 &lt; 1</math> in black, <math>R_0 \geq 1</math> in red, kernel density in continuous shades of orange</p>	Probabilistic	<p>30-day interval average temperature data spatially interpolated from 10 meteorological stations of the National Observatory of Athens (NOA).</p>	<p>Malaria, Greece</p>	<p>Pergantas et al., 2017</p>
<p>Spatial resolution: raster, <math>0.25 \times 0.25^\circ</math> cells</p> <p>Temporal reference: monthly mean (multi-year average, June-September)</p> <p>Color scheme: 5 classes, diverging multi-color gradient from yellow (<math>R_0 &lt; 0.5</math>) to dark red</p>	Probabilistic	<p>Daily temperature from E-OBS 12.0 (<a href="http://www.ecad.eu">http://www.ecad.eu</a>), 2006 - 2015.</p>	<p>Zika, Europe</p>	<p>Rocklöv et al., 2016</p>
<p>Spatial resolution: raster, <math>0.25^\circ \times 0.25^\circ</math></p> <p>Temporal reference: monthly (March, June, September, December)</p> <p>Color scheme: continuous double gradient from blue (starts from the lowest value 0) to red (the highest value)</p>	NGM	<p>Monthly mean air temperature data at a spatial resolution of <math>0.5^\circ \times 0.5^\circ</math>, interpolated into daily values</p>	<p>Zika, The Americas</p>	<p>Zhang et al., 2017</p>
<p>Spatial resolution: raster, 5 km <math>\times</math> 5 km</p> <p>Temporal reference: monthly mean (from January to November)</p> <p>Color scheme: 5 classes, diverging multi-color gradient blue-red, <math>R_0 &lt; 1</math> in blue, <math>R_0 \geq 1</math> in orange-red.</p>	NGM	<p>Daily maximum and minimum temperatures from the NASA Terrestrial Observation and Prediction System</p>	<p>Rift Valley fever, California, USA</p>	<p>Barker et al., 2013</p>

Spatial reference: raster, 1 km × 1 km cells Temporal reference: bi-weekly Color scheme: 9 classes (interval: 0.25), sequential single-hue color gradient from ( $R_0 = 0-0.25$ ) to ( $R_0 = 1.5-1.75$ )	NGM	Daily mean air temperatures from the NASA Terrestrial Observation and Prediction System.	West Nile disease, USA (California)	Hartley et al., 2012
Spatial reference: raster, 1 km × 1 km cells Temporal reference: 8 day periods (May – September) Color scheme: 16 classes, diverging color gradient in grey-scale	NGM	Daily MODIS land surface temperature interpolated from 8-day values (2009-2011).	West Nile disease, Italy	Calistri et al., 2016

**Duration of conditions where  $R_0 > 1$**

Spatial reference: raster, 1 km × 1 km cells Temporal reference: No. of months of $R_0 \geq 1$ per year, 30-year average Color scheme: 6 classes, diverging multi-color gradient	Probabilistic	Mean monthly air temperature for 1961–1990 and 1991–2007 from meteorological stations, transformed to surface maps using regression kriging. Future projections of temperature derived from models (REMO, WettReg).	Malaria, Germany	Holy et al., 2011
Spatial reference: raster, unknown resolution Temporal reference: No. of consecutive months of $R_0 \geq 1$ per year Color scheme: 13 classes, sequential single-hue color gradient	Probabilistic	Daily temperature data from meteorological stations averaged at the country level by epidemic week. Maps produced using monthly mean temperature from <a href="https://worldclim.org">https://worldclim.org</a>	Zika, dengue, and chikungunya, The Americas	Mordecai et al., 2017

**Both average  $R_0$  over time and duration of  $R_0 > 1$**

Spatial reference: raster, 1 km × 1 km cells Temporal reference: a) mean monthly $R_0$ for June – September; b) No. of days $R_0 > 1$ for 2010 & 2011 Color scheme: a) 12 classes (log scale), diverging color gradient from green ( $R_0 = 0.1$ ) to red ( $R_0 = 10$ ); b) 6 classes (interval: 20), diverging color gradient from green (0 days) to red (120 days)	NGM	Daily gridded temperature and precipitation data from the ALADIN weather prediction model, 2010–2011.	Bluetongue disease, Austria	Brugger & Rubel, 2013
--	-----	---	-----------------------------	-----------------------

<p>Spatial reference: raster, 0.25° × 0.25° cells  Temporal reference: a) yearly mean no. of days <math>R_0 &gt; 1</math> for 2003–2016; b) seasonal mean <math>R_0</math> for June – September for individual years  Color scheme: a) 9 classes (interval: 30), diverging color gradient from blue (0–30 days) to red (241-270 days); b) 6 classes (interval: 0.5), diverging color gradient from blue (<math>R_0 = 0–0.5</math>) to red (<math>R_0 = 2.51–3</math>)</p>	NGM	Daily mean temperature data from the E-OBS dataset ( <a href="http://www.ecad.eu">http://www.ecad.eu</a> , interpolated from weather stations).	Usutu disease, Europe	Cheng et al., 2018
<p>Spatial reference: raster, 5' × 5' cells  Temporal reference: based on monthly multi-year average temperatures a) <math>R_0</math> of the warmest month; b) number of months with <math>R_0 &gt; 1</math>  Color scheme: a) 5 classes, sequential single-hue color gradient from yellow to red; b) 4 classes, sequential single-hue color gradient from white to red</p>	Probabilistic	Daily temperature spatially interpolated from weather stations.	Chikungunya, Canada	Ng et al., 2017
<b>No temporal aggregation/other</b>				
<p>Spatial reference: raster, unknown resolution  Temporal reference: Monthly estimates of <math>R_0</math>  scheme: 4 classes from “low” to “high”, 4 colors, no legend</p>	Probabilistic	$R_0$ calculated for monthly mean temperature data from 50 meteorological stations, then smoothed through kriging.	Bluetongue disease, Switzerland	Racloz et al., 2008
<p>Spatial reference: raster, 1 km × 1 km cells  Temporal reference: NA, July assumed  Color scheme: 11 classes (interval: mostly 0.25), diverging color gradient from blue (<math>R_0 = 0-0.1</math>) – yellow – red (<math>R_0 = 5-10</math>)</p>	NGM	Average July temperature from <a href="https://worldclim.org">https://worldclim.org</a> , daytime and nighttime land surface temperature from MODIS Terra satellite data.	Canine leishmaniasis, France	Hartemink et al., 2011
<p>Spatial reference: raster, 1 km × 1 km cells  Temporal reference: NA  Color scheme: 7 classes (irregular intervals), diverging color gradient from blue (<math>R_0 = 2-2.7</math>) – green – red (<math>R_0 = 14.7-30</math>)</p>	NGM	Estimates of mean, minimum and maximum temperature and precipitation derived from meteorological stations; averaged long-term estimates of land surface temperature.	Lymphatic filariasis, sub-Saharan Africa	Moraga et al., 2015

## References

- Abiodun, G. J., Witbooi, P., Okosun, K. O., 2018. Modelling the impact of climatic variables on malaria transmission. *Hacettepe Journal of Mathematics and Statistics*, 47, 219-235.
- Barker, C. M., Niu, T. C., Reisen, W. K., Hartley, D. M., 2013. Data-Driven Modeling to Assess Receptivity for Rift Valley Fever Virus. *PLoS Neglected Tropical Diseases*, 7.
- Brugger, K., Rubel, F., 2013. Bluetongue disease risk assessment based on observed and projected *Culicoides obsoletus* spp. vector densities. *PLoS ONE*, 8, e60330.
- Cadar, D., Lühken, R., van der Jeugd, H., Garigliany, M., Ziegler, U., Keller, M., Lahoreau, J., Lachmann, L., Becker, N., Kik, M., Munnink, B. B. O., Bosch, S., Tannich, E., Linden, A., Schmidt, V., Koopmans, M. P., Rijks, J., Desmecht, D., Groschup, M. H., Reusken, C., Schmidt-Chanasit, J., 2017. Widespread activity of multiple lineages of Usutu virus, western Europe, 2016. *Eurosurveillance*, 22, 11-17.
- Calistri, P., Savini, L., Candeloro, L., Di Sabatino, D., Cito, F., Bruno, R., Danzetta, M. L., 2016. A transitional model for the evaluation of west Nile virus transmission in Italy. *Transboundary and Emerging Diseases*, 63, 485-496.
- Cheng, Y. C., Tjaden, N. B., Jaeschke, A., Lühken, R., Ziegler, U., Thomas, S. M., Beierkuhnlein, C., 2018. Evaluating the risk for Usutu virus circulation in Europe: comparison of environmental niche models and epidemiological models. *International Journal of Health Geographics*, 17.
- Cordovez, J. M., Rendon, L. M., Gonzalez, C., Guhl, F., 2014. Using the basic reproduction number to assess the effects of climate change in the risk of Chagas disease transmission in Colombia. *Acta Tropica*, 129, 74-82.
- Estrada-Pena, A., Ruiz-Fons, F., Acevedo, P., Gortazar, C., de la Fuente, J., 2013. Factors driving the circulation and possible expansion of Crimean-Congo haemorrhagic fever virus in the western Palearctic. *Journal of Applied Microbiology*, 114, 278-286.
- Guis, H., Caminade, C., Calvete, C., Morse, A. P., Tran, A., Baylis, M., 2012. Modelling the effects of past and future climate on the risk of bluetongue emergence in Europe. *Journal of the Royal Society, Interface*, 9, 339-350.
- Hartemink, N. A., Purse, B. V., Meiswinkel, R., Brown, H. E., de Koeijer, A., Elbers, A. R. W., Boender, G. J., Rogers, D. J., Heesterbeek, J. A. P., 2009. Mapping the basic reproduction number ( $R_0$ ) for vector-borne diseases: A case study on bluetongue virus. *Epidemics*, 1, 153-161.
- Hartemink, N. A., Vanwambeke, S. O., Heesterbeek, H., Rogers, D., Morley, D., Pesson, B., Davies, C., Mahamdallie, S., Ready, P., 2011. Integrated mapping of establishment risk for emerging vector-borne infections: A case study of canine leishmaniasis in southwest France. *PLoS ONE*, 6, e20817.
- Hartley, D. M., Barker, C. M., Le Menach, A., Niu, T. C., Gaff, H. D., Reisen, W. K., 2012. Effects of temperature on emergence and seasonality of west Nile virus in California. *American Journal of Tropical Medicine and Hygiene*, 86, 884-894.
- Holy, M., Schmidt, G., Schroder, W., 2011. Potential malaria outbreak in Germany due to climate warming: Risk modelling based on temperature measurements and regional climate models. *Environmental Science and Pollution Research International*, 18, 428-435.
- Johansson, M. A., Arana-Vizcarrondo, N., Biggerstaff, B. J., Gallagher, N., Marano, N., Staples, J. E., 2012. Assessing the risk of international spread of yellow fever virus: a mathematical analysis of an urban outbreak in Asuncion, 2008. *American Journal of Tropical Medicine and Hygiene*, 86, 349-358.
- Kakmeni, F. M. M., Guimapi, R. Y. A., Ndjomatchoua, F. T., Pedro, S. A., Mutunga, J., Tonnang, H. E. Z., 2018. Spatial panorama of malaria prevalence in Africa under climate change and interventions scenarios. *International Journal of Health Geographics*, 17.
- Mayo, C., Shelley, C., MacLachlan, N. J., Gardner, I., Hartley, D., Barker, C., 2016. A deterministic model to quantify risk and guide mitigation strategies to reduce bluetongue virus transmission in California dairy cattle. *PLoS ONE*, 11, e0165806.
- Moraga, P., Cano, J., Baggaley, R. F., Gyapong, J. O., Njenga, S. M., Nikolay, B., Davies, E., Rebollo, M. P., Pullan, R. L., Bockarie, M. J., Hollingsworth, T. D., Gambhir, M., Brooker, S. J., 2015. Modelling the distribution and transmission intensity of lymphatic filariasis in sub-Saharan Africa prior to scaling up interventions: Integrated use of geostatistical and mathematical modelling. *Parasites & Vectors*, 8, 560.

- Mordecai, E. A., Cohen, J. M., Evans, M. V., Gudapati, P., Johnson, L. R., Lippi, C. A., Miazgowicz, K., Murdock, C. C., Rohr, J. R., Ryan, S. J., Savage, V., Shocket, M. S., Ibarra, A. S., Thomas, M. B., Weikel, D. P., 2017. Detecting the impact of temperature on transmission of Zika, dengue, and chikungunya using mechanistic models. *PLoS Neglected Tropical Diseases*, *11*, e0005568.
- Munoz, A. G., Thomson, M. C., Stewart-Ibarra, A. M., Vecchi, G. A., Chourio, X., Najera, P., Moran, Z., Yang, X. S., 2017. Could the Recent Zika Epidemic Have Been Predicted? *Frontiers in Microbiology*, *8*.
- Ng, V., Fazil, A., Gachon, P., Deuymes, G., Radojević, M., Mascarenhas, M., Garasia, S., Johansson, M. A., Ogden, N. H., 2017. Assessment of the probability of autochthonous transmission of chikungunya virus in Canada under recent and projected climate change. *Environmental Health Perspectives*, *125*, 067001.
- Ogden, N. H., Radojevic, M., Wu, X. T., Duvvuri, V. R., Leighton, P. A., Wu, J. H., 2014. Estimated effects of projected climate change on the basic reproductive number of the lyme disease vector *Ixodes scapularis*. *Environmental Health Perspectives*, *122*, 631-638.
- Parham, P. E., Michael, E., 2010. Modeling the effects of weather and climate change on malaria transmission. *Environmental Health Perspectives*, *118*, 620-626.
- Pergantas, P., Tsatsaris, A., Malesios, C., Kriparakou, G., Demiris, N., Tselentis, Y., 2017. A spatial predictive model for malaria resurgence in central Greece integrating entomological, environmental and social data. *PLoS ONE*, *12*, e0178836.
- Racloz, V., Venter, G., Griot, C., Stark, K. D. C., 2008. Estimating the temporal and spatial risk of bluetongue related to the incursion of infected vectors into Switzerland. *BMC Veterinary Research*, *4*, 42.
- Rocklöv, J., Quam, M. B., Sudre, B., German, M., Kraemer, M. U. G., Brady, O., Bogoch, I., Liu-Helmersson, J., Wilder-Smith, A., Semenza, J. C., Ong, M., Aaslav, K. K., Khan, K., 2016. Assessing seasonal risks for the introduction and mosquito-borne spread of zika virus in Europe. *Ebiomedicine*, *9*, 250-256.
- Wu, X. T., Duvvuri, V. R., Lou, Y. J., Ogden, N. H., Pelcat, Y., Wu, J. H., 2013. Developing a temperature-driven map of the basic reproductive number of the emerging tick vector of Lyme disease *Ixodes scapularis* in Canada. *Journal of Theoretical Biology*, *319*, 50-61.
- Zhang, Q., Sun, K. Y., Chinazzi, M., Piontti, A. P. Y., Dean, N. E., Rojas, D. P., Merler, S., Mistry, D., Poletti, P., Rossi, L., Bray, M., Halloran, M. E., Longini, I. M., Vespignani, A., 2017. Spread of Zika virus in the Americas. *Proceedings of the National Academy of Sciences of the United States of America*, *114*, E4334-E4343.

## Manuscript 3

### Evaluating the Risk for Usutu virus circulation in Europe: Comparison of Environmental Niche Models and Epidemiological Models

Published in *International Journal of Health Geographics*, (2018) 17:35

Yanchao Cheng<sup>1</sup>, Nils Benjamin Tjaden<sup>1</sup>, Anja Jaeschke<sup>1</sup>, Renke Lühken<sup>2</sup>, Ute Ziegler<sup>3</sup>, Stephanie Margarete Thomas<sup>1</sup>, Carl Beierkuhnlein<sup>1</sup>

<sup>1</sup>Department of Biogeography, University of Bayreuth, Universitätsstr. 30, 95447 Bayreuth, Germany

<sup>2</sup>Bernhard Nocht Institute for Tropical Medicine, World Health Organization Collaborating Centre for Arbovirus and Hemorrhagic Fever Reference and Research, Hamburg, Germany

<sup>3</sup>Friedrich-Loeffler-Institut, Institute of Novel and Emerging Infectious Diseases, Südufer 10, 17493 Greifswald – Insel Riems

**Keywords:** Usutu, Maxent, SEIR, vector-borne disease, risk map, Europe, basic reproduction number,  $R_0$ , ENM

## Abstract

**Background** Usutu virus (USUV) is a mosquito-borne flavivirus, reported in many countries of Africa and Europe, with an increasing spatial distribution and host range. Recent outbreaks leading to regional declines of European common blackbird (*Turdus merula*) populations and a rising number of human cases emphasize the need for increased awareness and spatial risk assessment.

**Methods** Modelling approaches in ecology and epidemiology differ substantially in their algorithms, potentially resulting in diverging model outputs. Therefore, we implemented a parallel approach incorporating two commonly applied modelling techniques: 1) Maxent, a correlation-based environmental niche model and 2) a mechanistic epidemiological susceptible-exposed-infected-removed (SEIR) model.

Across Europe, surveillance data of USUV-positive birds from 2003 - 2016 was acquired to train the environmental niche model and to serve as test cases for the SEIR model. The SEIR model is mainly driven by daily mean temperature and calculates the basic reproduction number  $R_0$ . The environmental niche model was run with long-term bio-climatic variables derived from the same source in order to estimate climatic suitability.

**Results** Large areas across Europe are currently suitable for USUV transmission. Both models show patterns of high risk for USUV in parts of France, in the Pannonian Basin as well as northern Italy. The environmental niche model depicts the current situation better, but with USUV still being in an invasive stage there is a chance for under-estimation of risk. Areas where transmission occurred are mostly predicted correctly by the SEIR model, but it mostly fails to resolve the temporal dynamics of USUV events. High  $R_0$  values predicted by the SEIR model in areas without evidence for real-life transmission suggest that it may tend towards over-estimation of risk.

**Conclusions** The results from our parallel-model approach highlight that relying on a single model for assessing vector-borne disease risk may lead to incomplete conclusions. Utilizing different modelling approaches is thus crucial for risk-assessment of under-studied emerging pathogens like USUV.

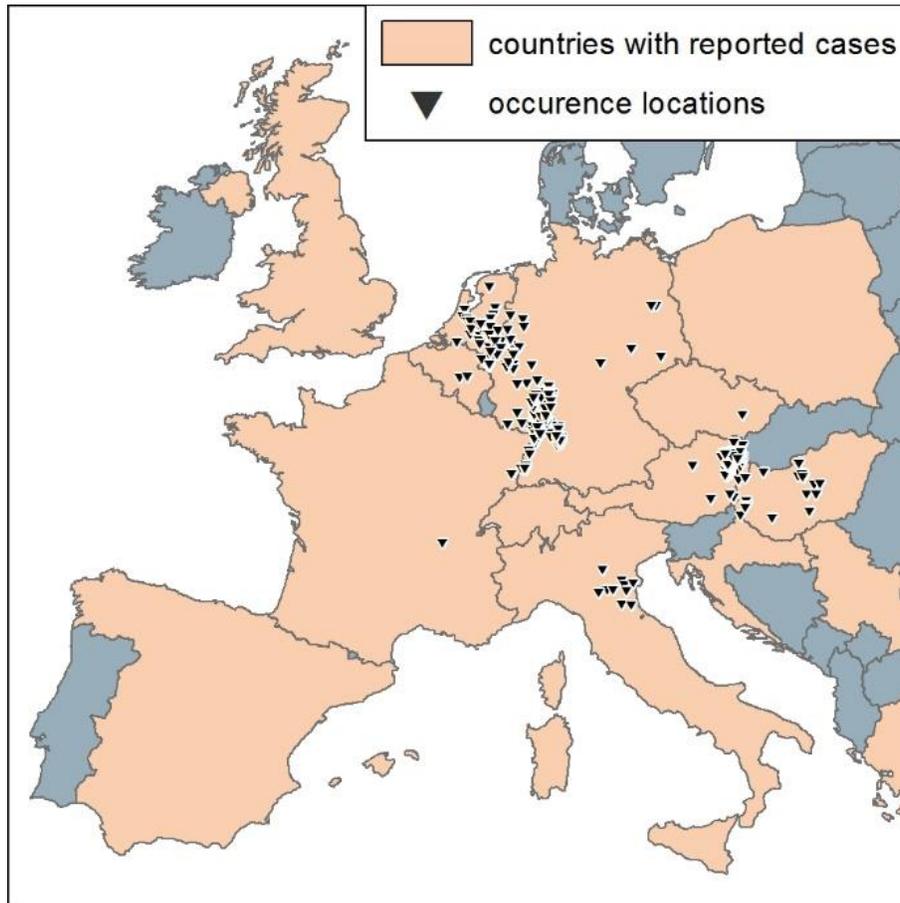
## Background

Vector-borne diseases (VBDs) are of growing importance. Due to global transport, long-distance travels, population growth, environmental and climatic changes, VBDs are emerging all over the world (Gage et al., 2008; Mangili & Gendreau, 2005; Tatem et al., 2006; Wu et al., 2017). In addition to human-mediated spread, mobile species such as migratory birds are promoting long-distance transport of pathogens (Engel et al., 2016). If the local conditions at the introduction sites (e.g. hosts, vectors, and climate) are suitable, the pathogen can establish and evolve quickly, resulting in rapid local spread (Kilpatrick & Randolph, 2012). Usutu virus (USUV) is an example where both processes resulted in the recent arrival and spread of a zoonotic mosquito-borne virus in Europe (Engel et al., 2016).

USUV is a flavivirus (Williams et al., 1964) belonging to the Japanese encephalitis virus serocomplex (Poidinger et al., 1996). As a member of the family Flaviviridae, USUV is a single-stranded RNA virus closely related to Murray Valley encephalitis virus, Japanese encephalitis virus, and West Nile virus (WNV) (Poidinger et al., 1996). It was first isolated in 1959 from *Culex neavei* mosquitoes in Swaziland and named after the Usutu river (Williams et al., 1964). Its most important vectors are mosquito species of the genus *Culex* (Nikolay, 2015). Since the first record, USUV has been reported for several African countries (e.g. Senegal, Central African Republic, Nigeria, Uganda) and detected in mosquitoes, birds, and humans (Nikolay et al., 2011). In Europe USUV has been detected in 15 countries, with increasing spatial distribution and host range (Ashraf et al., 2015; Barbic et al., 2013; Escribano-Romero et al., 2015; Nikolay, 2015; Rijks et al., 2016; Vittecoq et al., 2013) (Figure 1). The earliest evidence of USUV in Europe came from a dead common blackbird (*Turdus merula*) found in Italy in 1996, although this case was not identified as such until 2013 (Weissenböck et al., 2013). The first USUV epidemic in Europe was a series of dead common blackbirds reported from Austria in 2001 (Weissenböck et al., 2002). In the subsequent years, USUV was reported in further European countries. USUV or corresponding antibodies were detected in horses, bats, dogs (Barbic et al., 2013; Cadar et al., 2014; Durand et al., 2016), and at least 58 bird species, with common blackbirds as dominant avian host (Ashraf et al., 2015).

In 2009, the first human case of USUV infection in Europe was reported in Italy (Pecorari et al., 2009), followed by further human cases in Germany (Allering et al., 2012; Cadar, Maier, et al., 2017), Croatia (Vilibic-Cavlek et al., 2014), Austria (Bakonyi et al., 2017), and France (Simonin et al., 2018). Human cases are commonly characterized by mild symptoms including fever, rash, jaundice, headache, nuchal rigidity, hand tremor and hyperreflexia (Cavrini et al., 2009; Pecorari et al., 2009; Santini et al., 2015; Vilibic-Cavlek et al., 2014). However, at least in immunosuppressed patients USUV can cause a neuro-invasive infection (Pecorari et al., 2009), and it has recently been suspected to have caused idiopathic facial paralysis (Simonin et al., 2018). In addition to that, USUV infections were also detected from blood donors and healthy forestry workers in Germany and Italy (Allering et al., 2012; Cadar, Maier, et al., 2017; Percivalle et al., 2017), suggesting that asymptomatic infections can occur among humans. Recent data from Italy indicate that human USUV infections may not be a sporadic event and can even be more frequent than WNV infections in areas where

both viruses co-circulate (Calzolari et al., 2012; Grottole et al., 2017; Nikolay, 2015). Furthermore, due to cross reactions in antibody tests, the number of human USUV cases may be underestimated through confusion with other flaviviruses (Santini et al., 2015). As a consequence, the actual distribution of USUV and associated number of cases is likely to be larger than currently known (Lühken et al., 2017).



**Figure 1. USUV in Europe.** Orange areas: European countries where cases of USUV have been reported, regardless of species and method of confirmation. Triangles: Spatially explicit records of USUV occurrence 2003–2016 before spatial rarefication. These are locations where individual USUV-positive dead birds have been found, confirmed by reverse transcription polymerase chain reaction (RT-PCR).

The transmission cycle with birds as enzootic hosts creates a complex setting related to the risk for human health. First, migratory birds may transport the pathogen over large distances and can cause repeated re-introduction of the virus into a specific region that is not appropriate to maintain an outlasting population of the pathogen (Engel et al., 2016). Second, common blackbirds are the predominant host (Ashraf et al., 2015; Nikolay, 2015). This species is very common across Europe and has grown accustomed to urban habitats, exhibiting high population densities in human settlements (Møller et al., 2014). This means that vectors only need to cover short spatial distances

between infected birds and humans – and the widespread mosquito species *Cx. pipiens* is a known bridge vector between mammals, birds and humans (Börstler et al., 2016; Muñoz et al., 2011). In consequence, USUV is becoming an increasing threat for Europe as a mosquito-borne and zoonotic disease. Measures should be undertaken to improve or even create awareness towards zoonotic VBDs. For this purpose, spatial representations of risk are needed.

Models for vector borne viral diseases can be generated at various spatial and temporal scales (Nils Benjamin Tjaden et al., 2018). Maps of vector occurrence or disease transmission risk derived from them can be used to direct vector surveillance and control programs as well as to inform public health officials, medicine practitioners and the general public about potential risks. Current approaches can be divided into two basic groups: correlative models (e.g. environmental niche models) and process-based models (e.g. epidemiological models). Both types of models have their own strengths and weaknesses (Nils Benjamin Tjaden et al., 2018). Correlative environmental niche models, on the one hand, typically utilize species occurrence records and environmental predictor variables to estimate the current and future potential spatial distribution of a target species (Elith et al., 2006) or disease (Bhatt et al., 2013; Nsoesie et al., 2016; Peterson, 2014; Samy et al., 2016; Samy et al., 2014; Nils B. Tjaden et al., 2017). They do not require a-priori knowledge about the specific effects single variables have, and are typically used on coarser spatio-temporal scales (Nils Benjamin Tjaden et al., 2018). Process-based epidemiological models, on the other hand, aim to simulate the entire transmission process. Using knowledge gained from laboratory experiments or field observations, they require a deeper understanding of disease dynamics. As all models for VBD have their individual strengths and weaknesses, it is best practice not to rely on a single approach, but draw a conclusion from a consensus of multiple different models (Nils Benjamin Tjaden et al., 2018). Although both model categories are widely used when modeling VBDs (Nils Benjamin Tjaden et al., 2018), comparisons of different models' outputs are typically made within those categories (e.g. (Cianci et al., 2015)), and a comparison across categories is still missing.

To date only a limited number of USUV models for spatially confined areas exist. Based on an epidemiological model for WNV, Rubel et al. (Rubel et al., 2008) developed a mechanistic susceptible-exposed-infected-removed (SEIR) model for USUV in Vienna (Austria) (Brugger & Rubel, 2009; Reiczigel et al., 2010; Rubel et al., 2008), which was later successfully applied to Germany and neighboring countries (Cadar, Lühken, et al., 2017). This model is mainly driven by daily mean temperature, and to enable the comparison of modeled bird deaths and observed bird deaths, it was originally carried out with interpolated monthly mean temperature values so as to achieve the same temporal resolution as the available bird death data (Rubel et al., 2008). A different, environmental niche model-based approach was followed by Lühken et al. (Lühken et al., 2017), who adopted boosted regression trees to assess the spatio-temporal risk for USUV in Germany by estimating the risk in each grid cell.

Here we present, for the first time, USUV risk maps covering the entirety of the European mainland. Using two models in parallel, we utilize the mechanistic SEIR model by Rubel et al. (Rubel et al., 2008) as well as a newly developed environmental niche model based on the machine-learning

technique Maxent. Instead of using interpolated monthly mean temperature values for a single location, rasterized daily mean temperature was used to run the SEIR model. In order to increase comparability between the models, the same data source was also applied for the use of Maxent. Spatial risk maps were generated by both models. By using models from these two different groups, we are aiming at 1) estimating the potential risk for USUV transmission under current climate conditions in Europe and 2) investigating the differences between the outputs of two widely-used modelling approaches, which could be a first step towards interdisciplinary model comparison.

## **Methods**

### ***Study area and USUV occurrence records***

In this study, we focus on current European occurrence records of USUV in the years of 2003 - 2016, from the earliest to the latest USUV cases available. The investigation area is limited by the natural coastlines, as well as through the reported USUV locations in Eastern Europe (Figure 1).

To achieve a good data quality, only locations of USUV-positive birds confirmed by reverse transcription polymerase chain reaction (RT-PCR) were taken into account. This was done because 1) data from USUV-positive mammals or mosquitoes are collected quite unsystematic, i.e. data on USUV-positive birds are most consistent and comparable between the different European countries, and 2) other methods such as antibody analysis may not be able to distinguish USUV from other closely related flaviviruses such as WNV (Jöst et al., 2011). According to this rule, a total number of 376 USUV records was collected. USUV-positive data in Germany were collected by the German Mosquito Control Association (KABS), the Nature and Biodiversity Conservation Union (NABU), the local veterinary authorities and/or by the local state veterinary laboratories (Becker et al., 2012; Cadar, Lühken, et al., 2017; Ziegler et al., 2016; Ziegler et al., 2015). Records for other European countries were derived from the literature (Additional File 1): Geographical coordinates published in the literature were directly entered into the database. Precise site descriptions were digitized using Google Earth Pro, and high-quality occurrence maps were geo-referenced using ESRI ArcGIS 10.2.2.

### ***Climate Data***

Time series of daily mean temperature data, required by the SEIR model, were acquired from the E-OBS dataset version 15.0 (Haylock et al., 2008) on a regular latitude-longitude grid with a spatial resolution of 0.25° (about 20 km). E-OBS provides gridded daily temperature and precipitation data for Europe based on data from weather stations. To compare the results from the SEIR model and the environmental niche model properly, bio-climatic variables, which are required by the environmental niche model, were generated from the E-OBS dataset as well. Therefore, time series of daily minimum, maximum temperature and daily precipitation sums were acquired in addition to daily mean temperature.

Since the occurrence records for USUV cover the years of 2003 - 2016, these time series were trimmed accordingly. Considering that the spatial coverage of the E-OBS time series varies over time, grid cells with more than 10% missing data were excluded from our analyses. Monthly mean values were derived using the “raster” package (Hijmans) for R 3.2.1 (R Core Team) and 19 bioclimatic variables were calculated in SAGA-GIS version 2.1.4 (Conrad et al., 2015) for use with the environmental niche model.

### ***Environmental Niche Model: Maxent***

For the environmental niche model, we used Maxent 3.3.3k (Phillips et al., 2006). Maxent is a powerful machine-learning technique that is widely used (Nils Benjamin Tjaden et al., 2018) to model the potential distribution of species, especially when the occurrence data are sparse (Baldwin, 2009). Using occurrence records and environmental predictor variables as input data, Maxent generates maps of environmental suitability for transmission of USUV. Ranging between 0 for the lowest and 1 for the highest suitability, these maps can optionally be converted into presence/absence maps by applying a threshold value.

Maxent models are fitted assuming that all locations in the landscape are equally likely to be sampled. However, when the occurrence records are collected with different methods, sampling bias is inevitable. Compared to other methods, systematic sampling, also called spatial filtering of biased records (Kramer-Schadt et al., 2013), has a good performance regardless of species and bias type (Fourcade et al., 2014; Kramer-Schadt et al., 2013). It was applied by using the SDM tool box (Brown, 2014), an addon for ESRI ArcGIS that provides advanced tools and convenience functions for the Maxent workflow. To determine an appropriate spatial filtering resolution (the minimum distance between any two locations), the following rules were taken into consideration: 1) The spatial filtering process should decrease the bias distribution, but the remaining records should still represent the observed spatial patterns well. 2) There should be enough records left to run Maxent after spatial filtering. Consequently, the spatial filtering resolution was set to 20 km (about 0.25°), and 92 USUV records left after filtering in order to achieve optimum results and to avoid artefacts (Figure 1).

Selection of the environmental predictors for the model followed a two-step approach (Table 1). First, 8 out of the 19 bio-climatic variables that were deemed unsuitable for the task were excluded due to the following ecological reasons: BIO2 and 3 (“mean diurnal range” and “isothermality”) were excluded because while daily fluctuations in temperature are important for the mosquito life cycle and transmission dynamics, the monthly averages available here were considered unsuitable for capturing such short-term fluctuations. BIO12 (“annual precipitation”) was excluded because summer and winter precipitation play very different roles in this context and should be considered separately. All variables referring to the wettest/driest quarter or month of the year (BIO8, 9, 13, 14, 16, and 17) were excluded because seasonal precipitation patterns vary largely across Europe. As such, the wettest time of the year can be summer in some regions and winter in others, making this kind of variable unsuitable for larger scale analyses. The remaining eleven variables were

further reduced through the built-in Jackknife feature in Maxent with a ten-fold cross-validation run, following the recommendations of Elith et al. (Elith et al., 2011). In the end, a combination of five variables was chosen, consisting of annual mean temperature, minimum temperature of coldest month, mean temperature of coldest quarter, precipitation seasonality, and precipitation of warmest quarter. We used default settings for Maxent (10000 background locations, 500 iterations), but disabled the use of “threshold” and “hinge” features, that would have led to over-fitting due to an inappropriate amount of model complexity.

**Table 1.** Excluded and selected environmental predictor variables for the environmental niche model.

Abbreviation	Variables
Excluded - Monthly minima and maxima are not suitable to estimate daily fluctuations:	
BIO2	Mean Diurnal Range (Mean of monthly (max temp - min temp))
BIO3	Isothermality (BIO2/BIO7) × 100
Excluded - Summer and winter precipitation are important to distinguish for mosquitoes and disease transmission dynamics:	
BIO12	Annual Precipitation
Excluded - Wettest/driest time of the year can be in different seasons across Europe:	
BIO8	Mean Temperature of Wettest Quarter
BIO9	Mean Temperature of Driest Quarter
BIO13	Precipitation of Wettest Month
BIO14	Precipitation of Driest Month
BIO16	Precipitation of Wettest Quarter
BIO17	Precipitation of Driest Quarter
Excluded by jackknife:	
BIO4	Temperature Seasonality (standard deviation × 100)
BIO5	Max Temperature of Warmest Month
BIO7	Temperature Annual Range (BIO5-BIO6)
BIO10	Mean Temperature of Warmest Quarter
BIO19	Precipitation of Coldest Quarter
Model input:	
BIO1	Annual Mean Temperature
BIO6	Min Temperature of Coldest Month
BIO11	Mean Temperature of Coldest Quarter
BIO15	Precipitation Seasonality (Coefficient of Variation)
BIO18	Precipitation of Warmest Quarter

Maxent, like many other environmental niche model approaches, generates pseudo-absence (“background”) locations to make up for the lack of field records of true absence of the target species. Careful selection of the area from which these background locations are allowed to be drawn from is an important part of model creation, as it can affect model performance and results. According to Barve et al. (Barve et al., 2011), this should be done by requiring the background locations to be within the area the species could realistically disperse to. We followed a buffer-

based method (VanDerWal et al., 2009) by setting a series of buffer radii from 0.5° to 24° (see Additional File 2), given the grid cell size of 0.25°. It is suggested to take the radius when the model performance stops increasing (VanDerWal et al., 2009). In addition to the built-in AUC (area under the receiver operator characteristic curve), true skill statistic (TSS) was also calculated as an indicator of model performance (Additional File 2). A radius of 12° was chosen as suggested, with the final model reaching an AUC of 0.92 and a TSS score of 0.78, both suggesting good model performance. In this model, the minimum temperature of the coldest month had the strongest contribution to the model (58%), followed by precipitation of the warmest quarter (21%) and annual mean temperature (13%). The threshold for distinguishing predicted presence and absence was based on the receiver operator characteristic (ROC), choosing the point along the ROC curve that maximized the sum of sensitivity and specificity. We chose this criterion also known as “maxSSS” because it is objective (Liu et al., 2005), widely used, performs consistently well with presence-only data (Liu et al., 2015; Liu et al., 2013) and delivers threshold values that are relatively low (Liu et al., 2015), facilitating the high sensitivity desired in risk assessment studies.

### ***Epidemiological model: SEIR***

The SEIR model used in this study was developed by Rubel et al. (Rubel et al., 2008) for Vienna (Austria) and surrounding areas based on data from different parts of the world. The model simulates the seasonal life cycles and inter-species USUV infections of the main vector and host species, *Cx. pipiens* and *T. merula* respectively. Health states of birds and mosquitoes are classified into nine compartments (larvae state of mosquitoes, health states susceptible/latent infected/infectious of mosquitoes and birds as well as recovered and dead birds, see (Rubel et al., 2008)), and described by ordinary differential equations (see Additional File 3). The basic reproduction number  $R_0$  is then calculated as the dominant eigenvalue of the next-generation matrix as described in (Diekmann et al., 1990), resulting in (see Table 2 for model parameters and Additional File 3 for details):

$$R_0 = \sqrt{\left[ \frac{\delta_M \gamma_M \beta_M}{(\gamma_M + m_M) m_M} \frac{S_B}{K_B} \right] \left[ \frac{\delta_M \gamma_B \beta_B}{(\gamma_B + m_B) (\alpha_B + m_B)} \frac{S_M}{K_B} \right]}$$

The SEIR model is mainly driven by variables responding to temperature. Further drivers are latitude, calendar day, and parameters with constant values (Rubel et al., 2008).

The original SEIR R-code of the model was upgraded to work on a spatial grid rather than a single point location, and daytime length was calculated for each grid cell based on the geographical latitude of its center. Instead of interpolating daily data from monthly mean temperature, the model was run with true daily temperature data from the E-OBS dataset (Haylock et al., 2008). As an extensive literature review did not yield any new information, all other variables and parameters originally used by Rubel et al. were maintained in this study.

**Table 2.** Variables and parameters in the  $R_0$  equation, following (Rubel et al., 2008).

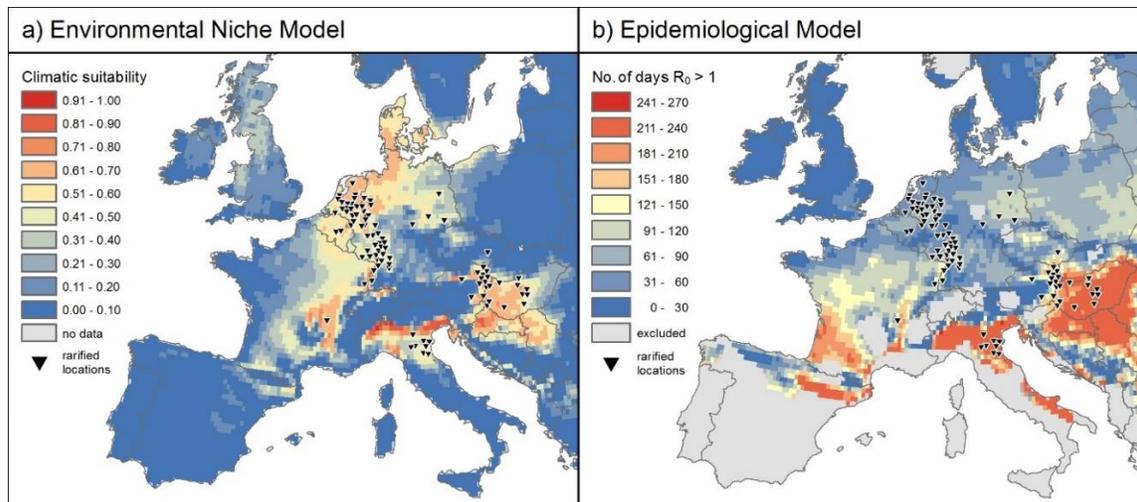
Parameter		Value
<b>Mosquitoes</b>		
mortality rate	$m_M$	$m_M(T) = 0.00025T^2 - 0.0094T + 0.10257$ $T$ : Daily Mean Temperature
biting rate	$\kappa$	$\kappa(T) = \frac{0.344}{1 + 1.231 \exp(-0.184(T - 20))}$
product of biting rate ( $\kappa$ ) and transmission possibility from mosquitoes to birds ( $P_M$ )	$\beta_M$	$\beta_M(T) = P_M \kappa(T)$ $P_M=1$
Percentage of non-hibernating mosquitoes	$\delta_M$	$\delta_M = 1 - \frac{1}{1 + 1775.7 \exp[1.559(D - 18.177)]}$ $D = 7.639 \arcsin \left[ \tan(\epsilon) \tan(\varphi) + \frac{0.0146}{\cos(\epsilon) \cos(\varphi)} \right] + 12$ $\epsilon = 0.409 \sin \left( \frac{2\pi(d - 80)}{365} \right)$ $D$ : Daytime length, $\epsilon$ : Declination, $\varphi$ : Geographic latitude
exposed – infected/infectious rate	$\gamma_M$	$\gamma_M(T) = 0.0093T - 0.1352, T \geq 15^\circ$ $\gamma_M(T) = 0, T < 15^\circ$
susceptible mosquito population	$S_M$	Dynamic value, see Additional File 3
<b>Birds</b>		
mortality rate	$m_B$	0.0012
removal rate: fraction of infected birds either recovering or dying	$\alpha_B$	0.182
exposed – infected/infectious rate	$\gamma_B$	0.667
product of biting rate ( $\kappa$ ) and transmission possibility from birds to mosquitoes ( $P_B$ )	$\beta_B$	$\beta_B(T) = P_B \kappa(T)$ $P_B=0.125$
susceptible black bird population	$S_B$	Dynamic value, see Additional File 3
environmental capacity	$K_B$	see Additional File 3

As the SEIR model for USUV was created for and calibrated within a temperate climate, water availability or precipitation were not considered a limiting factor by the developers. However, this assumption is not applicable for the entire study area, as the dry summers of Mediterranean climates can lead to a different, two peaked activity pattern of *Cx. pipiens* mosquitoes (Roiz et al., 2014). Consequently, the model was applied only to regions with a climate that is classified as cold or temperate with warm to hot summers but no dry season (Cfa, Cfb, Dfa and Dfb in the Köppen-Geiger system (Kottek et al., 2006; Peel et al., 2007)) (Figure 2b).

The basic reproduction number  $R_0$  (the number of secondary cases arising from a single infection in an otherwise uninfected population) of USUV calculated by the SEIR model is a threshold value: if  $R_0 > 1$ , an outbreak is possible after a single introduction of the pathogen; whereas if  $R_0 < 1$ , the introduced virus population will die out (Diekmann et al., 1990). The daily  $R_0$  value of each cell within the spatial raster was calculated within the time span of 2003-01-01 to 2016-12-31. From this, the average yearly number of days with  $R_0 > 1$  was calculated for each raster cell and the maxSSS threshold was calculated for direct comparison with the environmental niche model based on the same presence and background locations that were used in the Maxent model. In addition to that, the average daily  $R_0$  value of the main transmission season (June to September) was calculated for each year and raster cell.

## Results

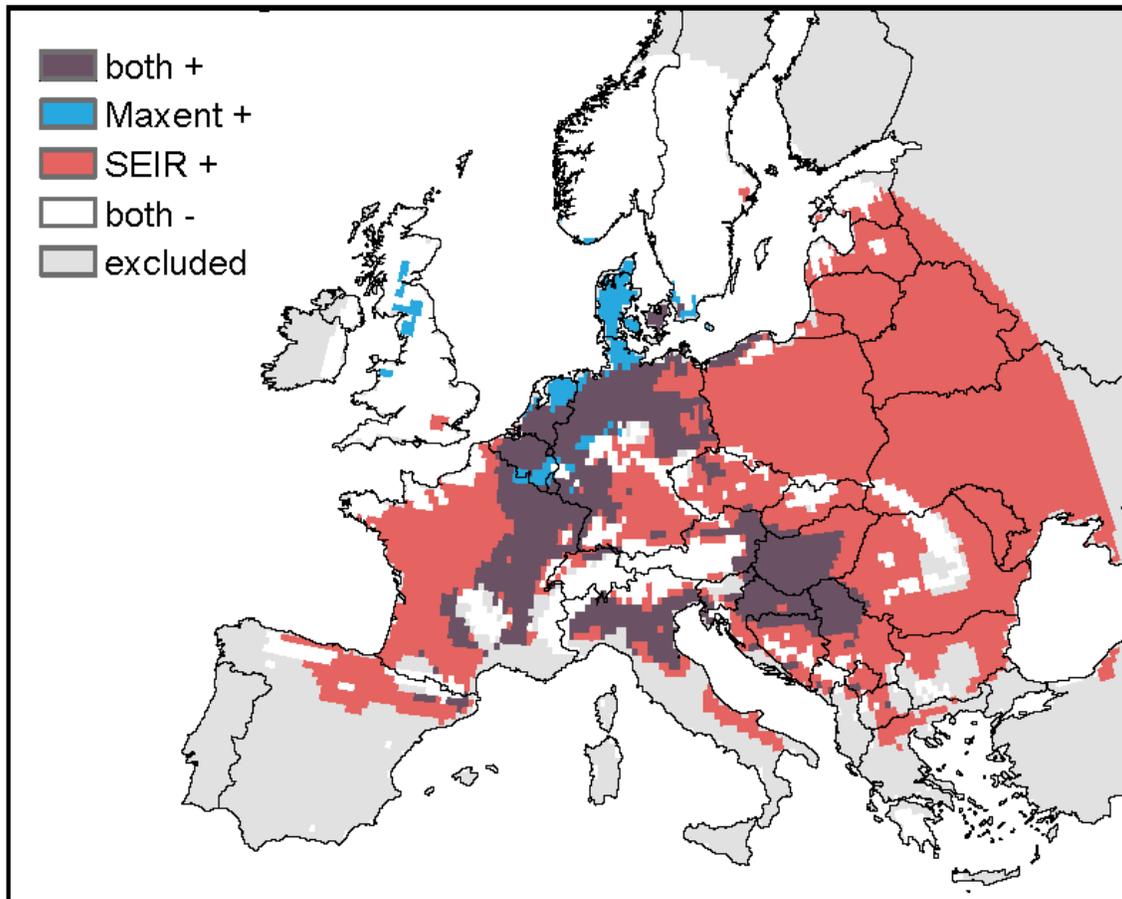
The potential geographic distribution of USUV predicted by both models on the continental European scale are shown in continuous form in Figure 2, and as a direct comparison based on the maxSSS thresholds (environmental niche model: 0.35 in Maxent's logistic output format, epidemiological model: 40 days of  $R_0 > 1$ ) in Figure 3. While there are differences between the two models in parts of the study area, 15% of the study area are projected to be suitable by both approaches. The northern Italian outbreak region in and around the Po Valley is identified as a highly suitable area for USUV by both models. The same is true for eastern Austria, the Pannonian Basin and adjoining areas, as well as a narrow strip along the Rhône river in France. Large parts of north-eastern France, the Benelux states and western and northern Germany are predicted to be at least somewhat suitable by both models. On the other hand, environmental niche model and SEIR agree on low risk being present in northern and mountainous regions (such as Sweden, Norway and the British Isles), where relatively low average and minimum temperatures keep the probability of transmission low.



**Figure 2. Potential geographic distribution of USUV in Europe.** a) Climatic suitability estimated by the environmental niche model, and b) the yearly mean absolute number of days of  $R_0 > 1$  simulated by the epidemiological SEIR model. Gray areas in b) denote regions with a dry season that were not included in the SEIR model. Both models use the same E-OBS climate data for 2003 - 2016. Locations of recorded cases for the environmental niche model were rarified (in comparison to Figure 1) to avoid spatial autocorrelation (see Methods).

In general, the environmental niche model accurately determines the occurrences of birds found positive with USUV. Compared to the SEIR, it suggests elevated climatic suitability for USUV to the north and west of the Jura Mountains as well as northwards along the Rhine and the North Sea coast until southern Denmark (Figure 2a). Following the maxSSS threshold, the environmental niche model predicts at total of 17% of the study area to be suitable for transmission (sensitivity: 0.946, specificity: 0.852). 2% of the entire area are considered suitable only by the environmental niche model and not by the SEIR, including most parts of Denmark and adjoining parts of northern Germany, northern Netherlands, southern Belgium and a few areas in northern Britain (Figure 3).

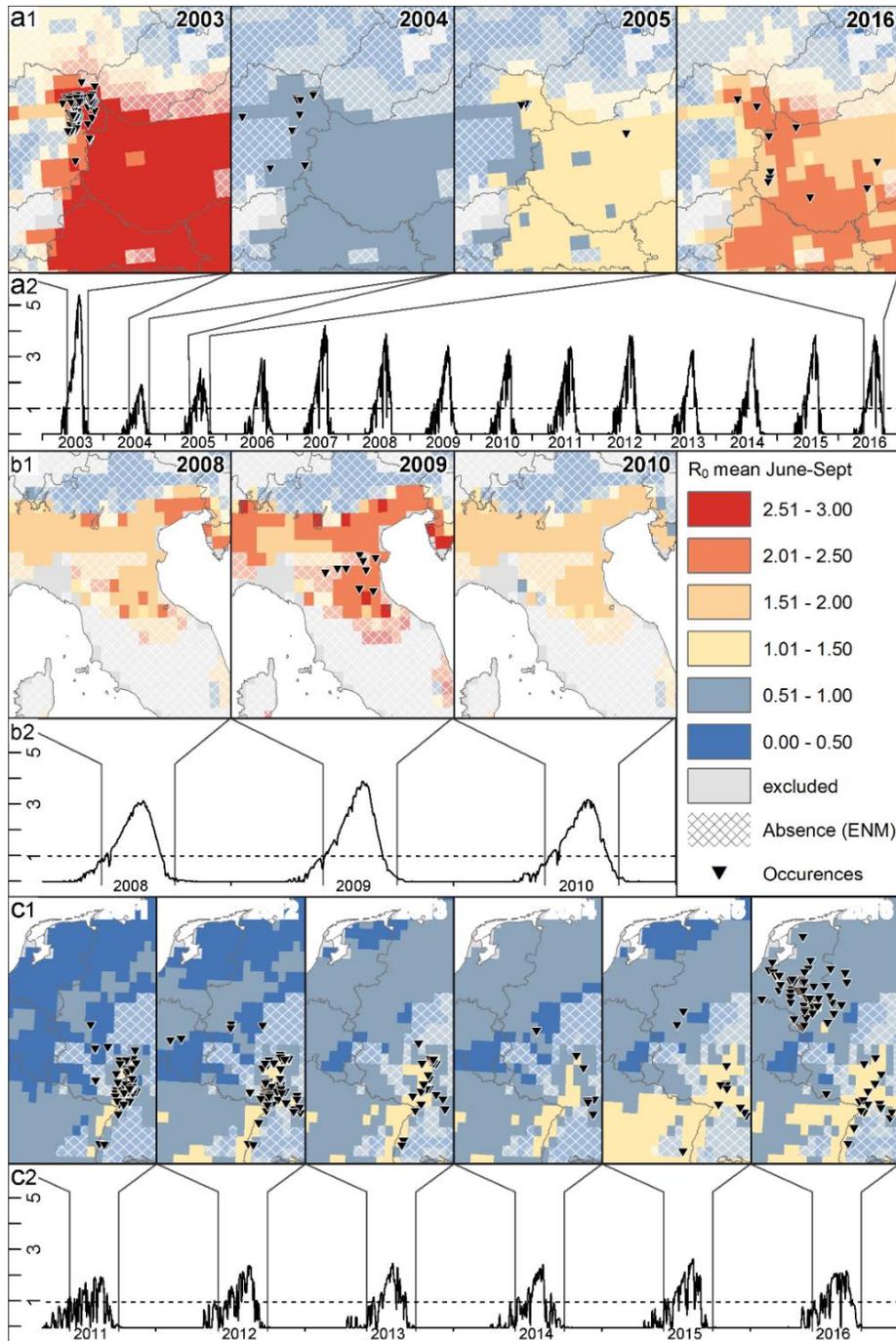
In contrast, the average yearly number of days with  $R_0 > 1$  derived from the SEIR suggests a high risk for USUV in southwestern France and southeastern Italy, but shows relatively low risk in the northern Germany-Netherlands-Belgium region (Figure 2b). North of the Pyrenees, the former French regions of Aquitaine and Midi-Pyrénées show a high transmission potential as well. Medium values mainly occur in Poland and northeastern Germany, along the Upper Rhine Valley and in central France. For the outbreak area in the Netherlands and northern Germany, the SEIR in this form suggests relatively low risk of transmission. However, following the maxSSS threshold, most of this region can still be classified as suitable for USUV transmission (Figure 3). A total of 67% of the whole study area lies above the threshold for this model, resulting in a sensitivity that is slightly higher (0.989) than that of the environmental niche model but a very low specificity (0.274).



**Figure 3. Areas of agreement and disagreement of both models.** Dark purple areas denote regions where both models predict suitable conditions for USUV-transmission based on the maxSSS threshold. In the blue and red areas, only the environmental niche model and SEIR predict suitable conditions, respectively. In white areas none of the models predicts suitable environmental conditions, while gray areas were excluded from further analyses because they are outside the climatic zones the SEIR model was developed for.

Zooming in towards the main areas of observed USUV transmission allows a closer inspection of the models. In the Austrian-Hungarian outbreak area, Maxent predicts climatic suitability values sufficient for USUV all observed occurrences (Figure 4 a1). The SEIR model predicts the highest  $R_0$  values for the largest USUV event in 2003 (Figure 4 a2) and considerably lower values for the following two years with less observed cases (Figure 4 a). Relatively high  $R_0$  values are observed again for the last USUV event in 2016. Interestingly, though, values for the USUV-free years of 2006 – 2015 are higher than those of 2004/5 (Figure 4a2).

In Italy, Maxent is able to predict the general outbreak area (Figure 4 b1). The SEIR model predicts elevated  $R_0$  values for the year of 2009 where USUV occurred, but similarly high values for the USUV-free years before and after (Figure 4 b2).



**Figure 4. Temporal patterns of the average  $R_0$  values for three selected regions of Europe.** a) Austria and the Pannonian Basin, b) northern Italy, and c) Germany and the Netherlands. 1) Spatial representation of both models for years with USUV events. Color coding in the maps shows the average daily  $R_0$  values throughout June to September for the given years. Gray areas denote climate types with dry seasons, thus the SEIR model was not applied there. Cross-hatching indicates areas where the environmental niche model suggests absence of USUV, based on climate data for the whole time period from 2003-2016. 2) Time series curves illustrate the daily  $R_0$  value, averaged over all occurrence records of the respective region for each given year.

In the largest outbreak area in western Germany and the Benelux states, Maxent closely resembles the observed pattern of USUV occurrence (Figure 4 c1). Compared to the other two regions, the SEIR model in these areas shows much lower average and absolute  $R_0$  values as well as higher temporal variability throughout the transmission season (Figure 4 c2). Average  $R_0$  values for the transmission season rise above 1 and match the occurrence records well in the Rhine Valley but stay below 1 in the northern parts of the area, i.e. the Netherlands and northwestern Germany.

## Discussion

In face of emerging VBDs and rapid spread into new regions with suitable climatic conditions, models that show the current geographic regions at risk are required to allow local health authorities to be prepared. However, modelling approaches can differ substantially in philosophy, structure, and algorithms. Pros and cons of different approaches are evident and, obviously, there is not one single approach to be preferred for every pathogen, area or timespan.

In this study, two fundamentally different models were applied to describe the current emergence of USUV in Europe. This disease exhibits a series of complex interactions between the virus, vectors and host species (Nikolay, 2015). Process-based models offer direct links between model outcome and underlying mechanisms, which makes interpretation of the observed spatial patterns relatively straightforward. However, exact knowledge on the parameters of USUV transmission is still scarce. With large numbers of USUV-positive birds reported from distinct geographical hot spots, the application of biogeographical distribution models may be a viable alternative. In order to identify coinciding and deviating model output, we ran the analyses based on the same climate data and following standard processes to detect regions at risk for the transmission of USUV.

The large-scale spatial patterns predicted by the two models (Figure 2 & 3) are quite similar close to the observed USUV events – with the notable exception of northern Germany and the Netherlands. Here, the environmental niche model favors higher latitudes as far north as Denmark, while the epidemiological model suggests good conditions for transmission in southwestern France and northeastern Spain (Figure 2b) and at least suitable conditions for most parts of Eastern Europe (Figure 3). Given the observed recent increase in temperatures across Europe and the projected further increase during the upcoming century [IPCC] (Cramer et al., 2001), it can be expected that both models under-estimate future potential for USUV transmission to some degree. If precipitation patterns change dramatically so as to affect mosquito populations, the SEIR model may not be a reliable option any more in some regions. Similarly, both models are not suitable to predict today's potential for USUV transmission in areas that are climatically very different from the study region.

### ***Environmental Niche Model***

As the environmental niche model is strongly driven by existing spatial records, it is not surprising that it reflects the current distribution of USUV records better. However, it has to be kept in mind that there is no consistent monitoring of USUV across Europe, leading to biases in the occurrence records. For instance, many USUV events were reported in Italy, Austria, Hungary, and Croatia (though no RT-PCR positive birds), but to date no USUV case was reported in their neighbor countries – Slovenia and Slovakia. Due to the same reason, only bird cases were included in our approach, as it is the least biased dataset in Europe, compared to USUV cases from wild mammals (e.g. bats and wild boars) or humans. Furthermore, we restricted our USUV dataset to USUV cases confirmed by RT-PCR counts, as other methods bear the possibility of false positives that would lead to overestimation of risk. Given the high activity of West Nile Virus in the area that could easily be mistaken for USUV in antibody tests, the gain from avoiding false positives should outweigh the loss from potentially excluding some true positives. Even though Maxent is relatively insensitive to sampling bias compared to other environmental niche models (Baldwin, 2009) and records were spatially rarified in this study, the modelling output would still be inevitably affected, e.g. in Italy, where occurrence records are comparably sparse.

In addition, USUV is still spreading in Europe and likely does not occupy its entire environmental niche yet, which may lead to under-estimation of risk through the environmental niche model in areas that may be climatically suitable, but have not been reached yet (compare e. g. (Elith et al., 2010)). The quality and accessibility of observed records of occurrence of vectors, hosts and especially pathogens is a major practical obstacle for the development of models of the environmental niche model family. Only a consistent and advanced monitoring system covering a selection of representative areas across Europe could give more accurate and reliable occurrence records to produce risk maps. Consequently, the environmental niche model performance can be improved as more occurrence data with high quality are available and the sampling bias is minimized. Ideally, such a monitoring system is centralized, open access and would not only focus on birds or mosquitoes but also include mammalian hosts such as rodents or bats to cover different types of potentially circulating pathogens. Especially the latter have been suspected to be underestimated but important hosts for other viral zoonotic diseases (Calisher et al., 2006). As USUV outbreaks typically cease with the arrival of winter, hibernating bats could enable overwintering of the virus. However, coordinated efforts are also needed for centralized and open access to the occurrence records resulting from these improved measures [35].

### ***Epidemiological Model***

As an absence of records does not necessarily indicate an absence of risk, it makes sense to use a mechanistic model to point out regions such as southwestern France, where transmission appears to be possible. The SEIR model captured the USUV events in the Pannonian Basin and Po Valley regions well, though the events in Germany and the Netherlands area were not represented correctly. Hence, it must be questioned whether the current knowledge on processes, mechanisms

and underlying parameters is sufficient to explain USUV transmission patterns and outbreaks. Although an extensive literature review was conducted with the aim of improving and updating the parameters for the SEIR model, no information supporting the integration of additional processes, drivers or variables was found. Therefore, all the parameters and variables used already in the 2008 study of Rubel et al. (Rubel et al., 2008) were kept unchanged, even though some of them are probably not suitable for the whole study area. For instance, population density as well as birth and mortality rates of common blackbirds are unlikely to be constant across the whole study area. An advanced, open-access monitoring system as discussed above could also be of great use for this.

Furthermore, although precipitation is known to affect mosquito life cycles and disease transmission dynamics (Kang et al., 2017; Morin et al., 2013), the applied SEIR model does not take this into account. The SEIR model for USUV was originally developed and calibrated for temperate climates. It is thus possible that certain ecological factors (e.g. precipitation), which are not limiting in the calibration area but could be limiting elsewhere, are not included in the model. In our study we restrained the extent for the SEIR model by excluding climate types with dry seasons in order to avoid making predictions for regions the model is not suitable for. Future models should aim to improve the population model components for vectors and hosts, leading to a more universally useful model. In addition, explicit parameters for USUV are not available yet and had to be substituted by data for the related WNV. For instance, no information about the extrinsic incubation period and its relation to ambient temperature is currently available. Data from a single experiment on a single strain of another virus (i.e. West-Nile virus) (Reisen et al., 2006) is far from optimal, as it has been shown that these experiments are subject to large uncertainty for various reasons (N. B. Tjaden et al., 2013). This is a common problem, though, since updated and realistic experiments are sorely needed for many VBDs (Nils Benjamin Tjaden et al., 2018). Future models could account for some of this uncertainty by incorporating stochastic variations instead of relying on fixed values, as it has already been done e.g. for Chikungunya (Ng et al., 2017).

Another point worth considering is that so far there is no standardized way of converting the daily values of  $R_0$  calculated by the SEIR model for each grid cell into interpretable maps. Obviously, some amount of temporal aggregation needs to be applied in order to gain low dimensional, printable maps. In practice, this ranges from  $R_0$  being displayed as averages for single months (e.g. (Rocklöv et al., 2016)) up to  $R_0$  values being averaged over 30-year periods (e.g. (Ogden et al., 2014)). Here, we chose to display average  $R_0$  values for single transmission seasons, which apparently failed to predict the 2016 USUV event in Northwest Europe (Figure 4 c)). However,  $R_0$  is a threshold value. Thus, while a value of  $R_0 > 1$  indicates high risk of disease spread, an average  $R_0 < 1$  for the same period does not necessarily mean no or even low risk, depending on how the length of that period was chosen and how often the threshold was exceeded. This is a serious drawback of SEIR model results to visualize the spatial-explicit risk of pathogen transmission. Hence, an alternative way of illustrating these models is concentrating on the duration of time where  $R_0 > 1$ . Here, we chose to count the (average) number of days per year where  $R_0 > 1$ , but this can also be done on other temporal scales (e.g. months (Mordecai et al., 2017)). In our case, this value apparently fails to

capture the outbreak area in Germany and the Netherlands (Figure 2 b). However, a closer look reveals that this again is a lack of knowledge about the details of the disease that prevents a meaningful interpretation of these maps, i.e., how many days of  $R_0 > 1$  are actually needed for an USUV event to occur. When this threshold would be known, the average yearly number of days of  $R_0 > 1$  map can be converted to a categorized risk map showing whether there is a risk and how severe it is. Furthermore, it has to be questioned, if higher absolute  $R_0$  values during the transmission season would reduce the number of days of  $R_0 > 1$  days required for an USUV outbreak. Only when these primary questions are addressed, a more reasonable risk map can be generated.

### **Outlook**

Further efforts should strive towards the unification of the two streams of modeling. As shown in this study, the ecological niche model reflects spatial distribution better, while the epidemiological model has the advantage of capturing short term variabilities, as it uses daily temperature data. Ecological niche models are run with climate data which typically covers decades, and as a consequence, extreme weather events such as heat waves would not be captured. An integrated model could benefit from both models' advantages. For example, in a hierarchical approach, spatial distribution of risk could first be estimated by an environmental niche model, followed by a zoom in to a finer scale for the investigation of temporal risk patterns in high risk areas through an epidemiological model with well-updated parameters and variables. In this case, the finer temporal scale epidemiological model, using daily weather data or even weather forecast data, can work as a live early warning forecast. Instead of projecting where climate is suitable, ecological niche models can also be applied to exclude unsuitable regions. In addition, in an integrated approach, environmental niche models that estimate the abundance of vectors and hosts could be nested in an epidemiological model as well, in order to gain more precise information on the required vector-to-host ratio.

### **Conclusion**

In conclusion, this study highlights the necessity to consider different approaches to detect the current and future areas under risk of VBDs. environmental niche models and epidemiological models examine rather complementary aspects, especially in terms of short-term weather conditions versus long-term climatic conditions. Environmental niche models are typically built upon long-term climate data and thus can be used to gain a general overview of the areas at risk and estimate potential effects of climate change. Given enough spatially explicit occurrence records are available, these models are particularly useful for a rapid risk assessment of emerging VBDs, while more detailed data about the transmission mechanisms is gathered. Once this data is available, elaborate mechanistic models can offer more fine-grained insights on the progression of outbreaks, with the potential for short-term forecasts based on weather models. At this point, environmental niche models for host or vector populations can provide valuable input data for

advanced epidemiological models. Thus, using both approaches complementing each other is key for a comprehensive and effective risk evaluation.

Wide parts of Europe are currently at risk of USUV circulation, and its status of a mostly neglected emerging disease makes estimation of its potential future range difficult. Evidence suggests that USUV events may be more likely to occur in climatically favored regions within Europe such as the Po Valley in northern Italy (Pautasso et al., 2016) and the Rhine Valley (Jöst et al., 2011; Ziegler et al., 2015). At the same time, these areas have a high human population density and exhibit large urban areas and cities. Remnant wetland habitats along rivers serve as habitats for migratory bird stops resulting in a combined setting with humans being exposed to high risk. The detected spatial patterns can be used to indicate regions where surveillance activities should be focused and intensified.

### **Availability of data and material**

All climate data is publicly available from the sources mentioned in the manuscript. Occurrence records for Europe are publicly available from the sources listed in Additional File 1. Occurrence records from Germany were collected within a dead bird surveillance program of the Friedrich-Loeffler-Institut, Greifswald-Insel Riems, Germany and the Bernhard Nocht Institute for Tropical Medicine, Hamburg, Germany in cooperation with the German Mosquito Control Association (KABS), the Nature and Biodiversity Conservation Union (NABU), the local veterinary authorities and/or by the local state veterinary laboratories. These datasets were used under license for the current study, and so are not publicly available. They are however available from the authors upon reasonable request and with permission of the respective third parties involved.

### **Funding**

This work was partly funded by the Federal Ministry of Education and Research of Germany (BMBF), grants No. 01EI1702A and 01KL1601. Yanchao Cheng is funded by China Scholarship Council, No. 201506040059.

### **Authors' contributions**

YC, NT, ST, AJ, RL, and CB developed the concept of the study. YC, UZ and RL compiled the occurrence records. NT and YC processed the climate data and adapted the SEIR model. YC and NT ran the Maxent models. CB, ST and AJ supervised the modelling process. YC prepared the figures. All authors discussed the preliminary results and figures at various stages of the modelling process. YC and NT wrote the original draft of the manuscript. All authors discussed and revised the manuscript. All authors read and approved the final version of the manuscript.

## Acknowledgements

We would like to thank Reinhold Stahlmann for his work on the figures. We acknowledge the E-OBS dataset from the EU-FP6 project ENSEMBLES (<http://ensembles-eu.metoffice.com>) and the data providers in the ECA&D project (<http://www.ecad.eu>).

## References

- Allering, L., Jöst, H., Emmerich, P., Günther, S., Lattwein, E., Schmidt, M., ... Schmidt-Chanasit, J., 2012. Detection of Usutu virus infection in a healthy blood donor from south-west Germany, 2012. *Eurosurveillance*, 17, pii=20341.
- Ashraf, U., Ye, J., Ruan, X. D., Wan, S. F., Zhu, B. B., & Cao, S. B., 2015. Usutu virus: an emerging *flavivirus* in Europe. *Viruses*, 7, 219-238.
- Bakonyi, T., Erdélyi, K., Brunthaler, R., Dán, Á., Weissenböck, H., & Nowotny, N., 2017. Usutu virus, Austria and Hungary, 2010-2016. *Emerging Microbes & Infections*, 6, e85.
- Baldwin, R. A., 2009. Use of Maximum Entropy modeling in wildlife research. *Entropy*, 11, 854-866.
- Barbic, L., Vilibic-Cavlek, T., Listes, E., Stevanovic, V., Gjenero-Margan, I., Ljubin-Sternak, S., ... Savini, G., 2013. Demonstration of Usutu virus antibodies in horses, Croatia. *Vector-Borne and Zoonotic Diseases*, 13, 772-774.
- Barve, N., Barve, V., Jiménez-Valverde, A., Lira-Noriega, A., Maher, S. P., Peterson, A. T., ... Villalobos, F., 2011. The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecological Modelling*, 222, 1810-1819.
- Becker, N., Jöst, H., Ziegler, U., Eiden, M., Hoper, D., Emmerich, P., ... Schmidt-Chanasit, J., 2012. Epizootic Emergence of Usutu Virus in Wild and Captive Birds in Germany. *PLoS One*, 7, e32604.
- Bhatt, S., Gething, P. W., Brady, O. J., Messina, J. P., Farlow, A. W., Moyes, C. L., ... Hay, S. I., 2013. The global distribution and burden of Dengue. *Nature*, 496, 504-507.
- Börstler, J., Jöst, H., Garms, R., Krüger, A., Tannich, E., Becker, N., ... Lühken, R., 2016. Host-feeding patterns of mosquito species in Germany. *Parasites & Vectors*, 9, 318.
- Brown, J. L., 2014. SDMtoolbox: a python-based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses. *Methods in Ecology and Evolution*, 5, 694-700.
- Brugger, K., & Rubel, F., 2009. Simulation of climate-change scenarios to explain Usutu-virus dynamics in Austria. *Preventive Veterinary Medicine*, 88, 24-31.
- Cadar, D., Becker, N., Campos, R. D. M., Börstler, J., Jöst, H., & Schmidt-Chanasit, J., 2014. Usutu virus in bats, Germany, 2013. *Emerging Infectious Diseases*, 20, 1771-1773.
- Cadar, D., Lühken, R., van der Jeugd, H., Garigliany, M., Ziegler, U., Keller, M., ... Schmidt-Chanasit, J., 2017. Widespread activity of multiple lineages of Usutu virus, western Europe, 2016. *Eurosurveillance*, 22, pii=30452.
- Cadar, D., Maier, P., Müller, S., Kress, J., Chudy, M., Bialonski, A., ... Schmidt-Chanasit, J., 2017. Blood donor screening for West Nile virus (WNV) revealed acute Usutu virus (USUV) infection, Germany, September 2016. *Eurosurveillance*, 22, pii=30501.
- Calisher, C. H., Childs, J. E., Field, H. E., Holmes, K. V., & Schountz, T., 2006. Bats: Important reservoir hosts of emerging viruses. *Clinical Microbiology Reviews*, 19, 531-545.
- Calzolari, M., Gaibani, P., Bellini, R., Defilippo, F., Pierro, A., Albieri, A., ... Bonilauri, P., 2012. Mosquito, bird and human surveillance of West Nile and Usutu viruses in Emilia-Romagna region (Italy) in 2010. *PLoS One*, 7, e38058.
- Cavrini, F., Gaibani, P., Longo, G., Pierro, A. M., Rossini, G., Bonilauri, P., ... Sambri, V., 2009. Usutu virus infection in a patient who underwent orthotopic liver transplantation, Italy, August-September 2009. *Eurosurveillance*, 14, pii=19448.
- Cianci, D., Hartemink, N., & Ibanez-Justicia, A., 2015. Modelling the potential spatial distribution of mosquito species using three different techniques. *International Journal of Health Geographics*, 14, 10.
- Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., ... Böhner, J., 2015. System for automated geoscientific analyses (SAGA) v. 2.1.4. *Geoscientific Model Development*, 8, 1991-2007.

- Cramer, W., Holten, J. I., Kaczmarek, Z., Martens, P., Nicholls, R. J., Öquist, M., ... Szolgay, J., 2001. Europe. In J. J. McCarthy, O. F. Canziani, L. N. A., D. J. Dokken, & K. S. White (Eds.), *Climate Change 2001: Impacts, Adaptation, and Vulnerability - Contribution of Working Group II to the Third Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge: Cambridge University Press.
- Diekmann, O., Heesterbeek, J. A. P., & Metz, J. A. J., 1990. On the definition and the computation of the basic reproduction ratio  $R_0$  in models for infectious-diseases in heterogeneous populations. *Journal of Mathematical Biology*, 28, 365-382.
- Durand, B., Haskouri, H., Lowenski, S., Vachieri, N., Beck, C., & Lecollinet, S., 2016. Seroprevalence of West Nile and Usutu viruses in military working horses and dogs, Morocco, 2012: dog as an alternative WNV sentinel species? *Epidemiology and Infection*, 144, 1857-1864.
- Elith, J., Graham, C. H., Anderson, R. P., Dudik, M., Ferrier, S., Guisan, A., ... Zimmermann, N. E., 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 29, 129-151.
- Elith, J., Kearney, M., & Phillips, S., 2010. The art of modelling range-shifting species. *Methods in Ecology and Evolution*, 1, 330-342.
- Elith, J., Phillips, S. J., Hastie, T., Dudik, M., Chee, Y. E., & Yates, C. J., 2011. A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, 17, 43-57.
- Engel, D., Jöst, H., Wink, M., Börstler, J., Bosch, S., Garigliany, M. M., ... Schmidt-Chanasit, J., 2016. Reconstruction of the evolutionary history and dispersal of Usutu virus, a neglected emerging arbovirus in Europe and Africa. *Mbio*, 7, e01938-01915.
- Escrignano-Romero, E., Lupulović, D., Merino-Ramos, T., Blázquez, A. B., Lazić, G., Lazić, S., ... Petrović, T., 2015. West Nile virus serosurveillance in pigs, wild boars, and roe deer in Serbia. *Veterinary Microbiology*, 176, 365-369.
- Fourcade, Y., Engler, J. O., Rödder, D., & Secondi, J., 2014. Mapping species distributions with MAXENT Using a geographically biased sample of presence data: A performance assessment of methods for correcting sampling bias. *PLoS One*, 9, e97122.
- Gage, K. L., Burkot, T. R., Eisen, R. J., & Hayes, E. B., 2008. Climate and vectorborne diseases. *American Journal of Preventive Medicine*, 35, 436-450.
- Grottola, A., Marcacci, M., Tagliazucchi, S., Gennari, W., Di Gennaro, A., Orsini, M., ... Savini, G., 2017. Usutu virus infections in humans: a retrospective analysis in the municipality of Modena, Italy. *Clinical Microbiology and Infection*, 23, 33-37.
- Haylock, M. R., Hofstra, N., Klein Tank, A., Klok, E. J., Jones, P. D., & New, M., 2008. A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. *Journal of Geophysical Research*, 113, D20119.
- Hijmans, R. J., raster: Geographic Data Analysis and Modeling. R package version 2.5-8. 2016; Available from: <http://CRAN.R-project.org/package=raster>. In.
- Jöst, H., Bialonski, A., Maus, D., Sambri, V., Eiden, M., Groschup, M. H., ... Schmidt-Chanasit, J., 2011. Short Report: Isolation of Usutu virus in Germany. *American Journal of Tropical Medicine and Hygiene*, 85, 551-553.
- Kang, D. S., Tomas, R., & Sim, C., 2017. The effects of temperature and precipitation on *Culex quinquefasciatus* (Diptera: Culicidae) abundance: A case study in the Greater Waco city, Texas. *Vector Biology Journal*, 2, 1000116.
- Kilpatrick, A. M., & Randolph, S. E., 2012. Drivers, dynamics, and control of emerging vector-borne zoonotic diseases. *Lancet*, 380, 1946-1955.
- Kottek, M., Grieser, J., Beck, C., Rudolf, B., & Rubel, F., 2006. World map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift*, 15, 259-263.
- Kramer-Schadt, S., Niedballa, J., Pilgrim, J. D., Schroder, B., Lindenborn, J., Reinfelder, V., ... Wilting, A., 2013. The importance of correcting for sampling bias in Maxent species distribution models. *Diversity and Distributions*, 19, 1366-1379.
- Liu, C. R., Barry, P. M., Dawson, T. P., & Pearson, R. G., 2005. Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*, 28, 385-393.
- Liu, C. R., Newell, G., & White, M., 2015. On the selection of thresholds for predicting species occurrence with presence-only data. *Ecology and Evolution*, 6, 337-348.

- Liu, C. R., White, M., & Newell, G., 2013. Selecting thresholds for the prediction of species occurrence with presence-only data. *Journal of Biogeography*, *40*, 778-789.
- Lühken, R., Jöst, H., Cadar, D., Thomas, S. M., Bosch, S., Tannich, E., ... Schmidt-Chanasit, J., 2017. Distribution of Usutu virus in Germany and its effect on breeding bird populations. *Emerging Infectious Diseases*, *23*, 1991-1998.
- Mangili, A., & Gendreau, M. A., 2005. Transmission of infectious diseases during commercial air travel. *Lancet*, *365*, 989-996.
- Møller, A. P., Jokimäki, J., Skorka, P., & Tryjanowski, P., 2014. Loss of migration and urbanization in birds: a case study of the blackbird (*Turdus merula*). *Oecologia*, *175*, 1019-1027.
- Mordecai, E. A., Cohen, J. M., Evans, M. V., Gudapati, P., Johnson, L. R., Lippi, C. A., ... Weikel, D. P., 2017. Detecting the impact of temperature on transmission of Zika, dengue, and chikungunya using mechanistic models. *Plos Neglected Tropical Diseases*, *11*, e0005568.
- Morin, C. W., Comrie, A. C., & Ernst, K., 2013. Climate and Dengue transmission: Evidence and implications. *Environmental Health Perspectives*, *121*, 1264-1272.
- Muñoz, J., Eritja, R., Alcaide, M., Montalvo, T., Soriguer, R. C., & Figuerola, J., 2011. Host-feeding patterns of native *Culex pipiens* and invasive *Aedes albopictus* mosquitoes (Diptera: Culicidae) in urban zones from Barcelona, Spain. *Journal of Medical Entomology*, *48*, 956-960.
- Ng, V., Fazil, A., Gachon, P., Deuymes, G., Radojević, M., Mascarenhas, M., ... Ogden, N. H., 2017. Assessment of the probability of autochthonous transmission of chikungunya virus in Canada under recent and projected climate change. *Environmental Health Perspectives*, *125*, 067001.
- Nikolay, B., 2015. A review of West Nile and Usutu virus co-circulation in Europe: how much do transmission cycles overlap? *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *109*, 609-618.
- Nikolay, B., Diallo, M., Boye, C. S. B., & Sall, A. A., 2011. Usutu virus in Africa. *Vector-Borne and Zoonotic Diseases*, *11*, 1417-1423.
- Nsoesie, E. O., Kraemer, M. U., Golding, N., Pigott, D. M., Brady, O. J., Moyes, C. L., ... Brownstein, J. S., 2016. Global distribution and environmental suitability for Chikungunya virus, 1952 to 2015. *Eurosurveillance*, *21*, pii=30234.
- Ogden, N. H., Radojevic, M., Wu, X. T., Duvvuri, V. R., Leighton, P. A., & Wu, J. H., 2014. Estimated effects of projected climate change on the basic reproductive number of the lyme disease vector *Ixodes scapularis*. *Environmental Health Perspectives*, *122*, 631-638.
- Pautasso, A., Radaelli, M. C., Ballardini, M., Francese, D. R., Verna, F., Modesto, P., ... Casalone, C., 2016. Detection of West Nile and Usutu viruses in Italian free areas: Entomological surveillance in Piemonte and Liguria Regions, 2014. *Vector-Borne and Zoonotic Diseases*, *16*, 292-294.
- Pecorari, M., Longo, G., Gennari, W., Grottole, A., Sabbatini, A. M., Tagliazucchi, S., ... Rumpianesi, F., 2009. First human case of Usutu virus neuroinvasive infection, Italy, August-September 2009. *Eurosurveillance*, *14*, pii=19446.
- Peel, M. C., Finlayson, B. L., & McMahon, T. A., 2007. Updated world map of the Köppen-Geiger climate classification. *Hydrology and Earth System Sciences*, *11*, 1633-1644.
- Percivalle, E., Sasser, D., Rovida, F., Isernia, P., Fabbi, M., Baldanti, F., ... Marone, P., 2017. Usutu virus antibodies in blood donors and healthy forestry workers in the Lombardy region, northern Italy. *Vector-Borne and Zoonotic Diseases*, *17*, 658-661.
- Peterson, A. T., 2014. *Mapping disease transmission risk: enriching models using biogeography and ecology*. Baltimore: Johns Hopkins University Press.
- Phillips, S. J., Anderson, R. P., & Schapire, R. E., 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, *190*, 231-259.
- Poidinger, M., Hall, R. A., & Mackenzie, J. S., 1996. Molecular characterization of the Japanese encephalitis serocomplex of the *flavivirus* genus. *Virology*, *218*, 417-421.
- R Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2015; Available from: <http://www.R-project.org/>. In.
- Reiczigel, J., Brugger, K., Rubel, F., Solymosi, N., & Lang, Z., 2010. Bayesian analysis of a dynamical model for the spread of the Usutu virus. *Stochastic Environmental Research and Risk Assessment*, *24*, 455-462.

- Reisen, W. K., Fang, Y., & Martinez, V. M., 2006. Effects of temperature on the transmission of West Nile virus by *Culex tarsalis* (Diptera : Culicidae). *Journal of Medical Entomology*, *43*, 309-317.
- Rijks, J., Kik, M., Slaterus, R., Foppen, R., Stroo, A., Ijzer, J., ... Reusken, C., 2016. Widespread Usutu virus outbreak in birds in the Netherlands, 2016. *Eurosurveillance*, *21*, pii=30391.
- Rocklöv, J., Quam, M. B., Sudre, B., German, M., Kraemer, M. U. G., Brady, O., ... Khan, K., 2016. Assessing seasonal risks for the introduction and mosquito-borne spread of zika virus in Europe. *Ebiomedicine*, *9*, 250-256.
- Roiz, D., Ruiz, S., Soriguer, R., & Figuerola, J., 2014. Climatic effects on mosquito abundance in Mediterranean wetlands. *Parasites & Vectors*, *7*, 333.
- Rubel, F., Brugger, K., Hantel, M., Chvala-Mannsberger, S., Bakonyi, T., Weissenböck, H., ... Nowotny, N., 2008. Explaining Usutu virus dynamics in Austria: Model development and calibration. *Preventive Veterinary Medicine*, *85*, 166-186.
- Samy, A. M., Thomas, S. M., Abd El Wahed, A., Cohoon, K. P., & Peterson, A. T., 2016. Mapping the global geographic potential of Zika virus spread. *Memorias Do Instituto Oswaldo Cruz*, *111*, 559-560.
- Samy, A. M., van de Sande, W. W. J., Fahal, A. H., & Peterson, A. T., 2014. Mapping the potential risk of Mycetoma infection in Sudan and South Sudan using ecological niche modeling. *Plos Neglected Tropical Diseases*, *8*, e3250.
- Santini, M., Vilibic-Cavlek, T., Barsic, B., Barbic, L., Savic, V., Stevanovic, V., ... Savini, G., 2015. First cases of human Usutu virus neuroinvasive infection in Croatia, August-September 2013: clinical and laboratory features. *Journal of Neurovirology*, *21*, 92-97.
- Simonin, Y., Sillam, O., Carles, M. J., Gutierrez, S., Gil, P., Constant, O., ... Foulongne, V., 2018. Human Usutu virus infection with atypical neurologic presentation, Montpellier, France, 2016. *Emerging Infectious Diseases*, *24*, 875-878.
- Tatem, A. J., Rogers, D. J., & Hay, S. I., 2006. Global transport networks and infectious disease spread. *Advances in Parasitology*, *62*, 293-343.
- Tjaden, N. B., Caminade, C., Beierkuhnlein, C., & Thomas, S. M., 2018. Mosquito-borne diseases: Advances in modelling climate-change impacts. *Trends in Parasitology*, *34*, 227-245.
- Tjaden, N. B., Suk, J. E., Fischer, D., Thomas, S. M., Beierkuhnlein, C., & Semenza, J. C., 2017. Modelling the effects of global climate change on Chikungunya transmission in the 21st century. *Scientific Reports*, *7*, 3813.
- Tjaden, N. B., Thomas, S. M., Fischer, D., & Beierkuhnlein, C., 2013. Extrinsic incubation period of Dengue: Knowledge, backlog, and applications of temperature dependence. *Plos Neglected Tropical Diseases*, *7*, e2207.
- VanDerWal, J., Shoo, L. P., Graham, C., & William, S. E., 2009. Selecting pseudo-absence data for presence-only distribution modeling: How far should you stray from what you know? *Ecological Modelling*, *220*, 589-594.
- Vilibic-Cavlek, T., Kaic, B., Barbic, L., Pem-Novosel, I., Slavic-Vrzic, V., Lesnikar, V., ... Savini, G., 2014. First evidence of simultaneous occurrence of West Nile virus and Usutu virus neuroinvasive disease in humans in Croatia during the 2013 outbreak. *Infection*, *42*, 689-695.
- Vittecoq, M., Lecollinet, S., Jourdain, E., Thomas, F., Blanchon, T., Arnal, A., ... Gauthier-Clerc, M., 2013. Recent circulation of West Nile virus and potentially other closely related flaviviruses in southern France. *Vector-Borne and Zoonotic Diseases*, *13*, 610-613.
- Weissenböck, H., Bakonyi, T., Rossi, G., Mani, P., & Nowotny, N., 2013. Usutu virus, Italy, 1996. *Emerging Infectious Diseases*, *19*, 274-277.
- Weissenböck, H., Kolodziejek, J., Url, A., Lussy, H., Rebel-Bauder, B., & Nowotny, N., 2002. Emergence of Usutu virus, an African mosquito-borne flavivirus of the Japanese encephalitis virus group, central Europe. *Emerging Infectious Diseases*, *8*, 652-656.
- Williams, M. C., Knight, E. M., Haddow, A. J., & Simpson, D. I. H., 1964. Isolation of West Nile virus from man and of Usutu virus from the bird-biting mosquito *Mansonia aurites* (Theobald) in Entebbe area of Uganda. *Annals of Tropical Medicine and Parasitology*, *58*, 367-374.
- Wu, T., Perrings, C., Kinzig, A., Collins, J. P., Minter, B. A., & Daszak, P., 2017. Economic growth, urbanization, globalization, and the risks of emerging infectious diseases in China: A review. *Ambio*, *46*, 18-29.

- Ziegler, U., Fast, C., Eiden, M., Bock, S., Schulze, C., Hoeper, D., ... Groschup, M. H., 2016. Evidence for an independent third Usutu virus introduction into Germany. *Veterinary Microbiology*, 192, 60-66.
- Ziegler, U., Jöst, H., Müller, K., Fischer, D., Rinder, M., Tietze, D. T., ... Groschup, M. H., 2015. Epidemic spread of Usutu virus in southwest Germany in 2011 to 2013 and monitoring of wild birds for Usutu and West Nile viruses. *Vector-Borne and Zoonotic Diseases*, 15, 481-488.

**Additional File 1. Records of USUV-infected bird locations confirmed by PCR collected from the literature.**

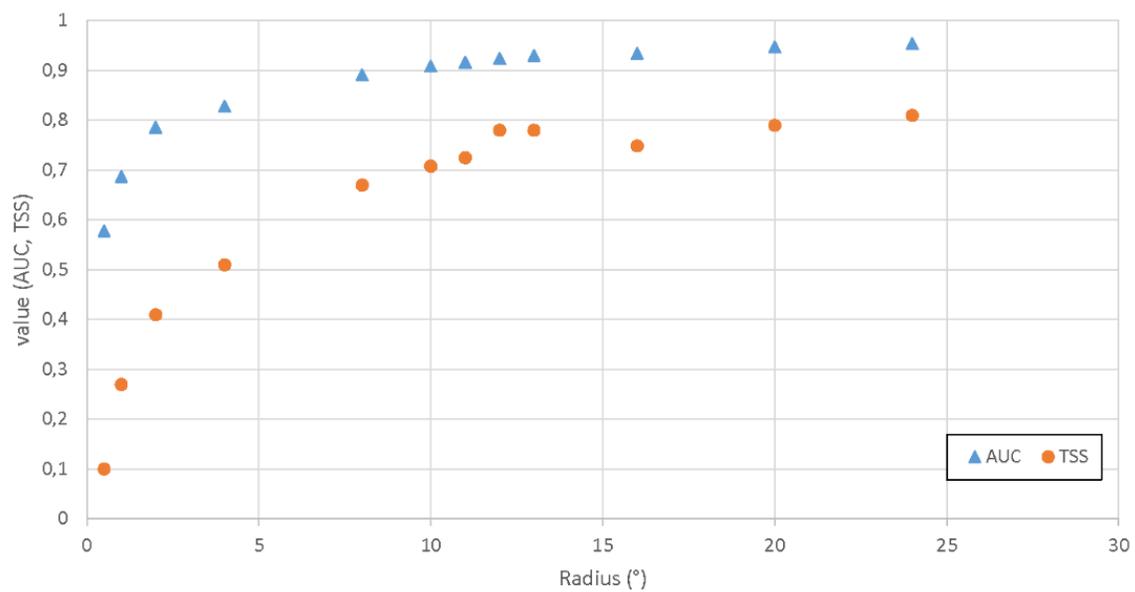
Countries	Outbreak years	Data type	Reference
Austria	2003–2005	Map	[1]
Hungary	2005–2006	Map	[2]
Italy	2009	Map	[3]
Italy	2009	Map	[4]
Italy	2010	Map	[5]
Austria and Hungary	2010–2016	Coordinates	[6]
Italy	2011	Map	[7]
Germany	2011, 2015	Site description	[8]
Czech Republic	2011–2012	Coordinates	[9]
Germany	2011–2013	Map	[10]
Belgium	2012	Coordinates	[11]
Italy	2012	Site description	[12]
Germany	2013	Coordinates	[13]
Italy	2013	Map	[14]
France	2015	Site description	[15]
Netherlands	2016	Map	[16]

**References**

1. Chvala S, Bakonyi T, Bukovsky C, Meister T, Brugger K, Rubel F, et al. Monitoring of Usutu virus activity and spread by using dead bird surveillance in Austria, 2003-2005. *Vet Microbiol.* 2007;122(3-4): 237-245. doi: 10.1016/j.vetmic.2007.01.029.
2. Bakonyi T, Erdélyi K, Ursu K, Ferenczi E, Csörgo T, Lussy H, et al. Emergence of Usutu virus in Hungary. *J Clin Microbiol.* 2007;45(12): 3870-3874. doi: 10.1128/jcm.01390-07.
3. Calzolari M, Bonilauri P, Bellini R, Albieri A, Defilippo F, Maioli G, et al. Evidence of simultaneous circulation of West Nile and Usutu viruses in mosquitoes sampled in Emilia-Romagna region (Italy) in 2009. *PLoS One.* 2010;5(12): e14324. doi: 10.1371/journal.pone.0014324.
4. Tamba M, Bonilauri P, Bellini R, Calzolari M, Albieri A, Sambri V, et al. Detection of Usutu virus within a West Nile virus surveillance program in northern Italy. *Vector Borne Zoonotic Dis.* 2011;11(5): 551-557. doi: 10.1089/vbz.2010.0055.
5. Calzolari M, Gaibani P, Bellini R, Defilippo F, Pierro A, Albieri A, et al. Mosquito, bird and human surveillance of West Nile and Usutu viruses in Emilia-Romagna region (Italy) in 2010. *PLoS One.* 2012;7(5): e38058. doi: 10.1371/journal.pone.0038058.
6. Bakonyi T, Erdélyi K, Brunthaler R, Dán Á, Weissenböck H, Nowotny N. Usutu virus, Austria and Hungary, 2010-2016. *Emerg Microbes Infect.* 2017;6(10): e85. doi: 10.1038/emi.2017.72.
7. Calzolari M, Bonilauri P, Bellini R, Albieri A, Defilippo F, Tamba M, et al. Usutu virus persistence and West Nile virus inactivity in the Emilia-Romagna region (Italy) in 2011. *PLoS One.* 2013;8(5): e63978. doi: 10.1371/journal.pone.0063978.
8. Ziegler U, Fast C, Eiden M, Bock S, Schulze C, Hoepfer D, et al. Evidence for an independent third Usutu virus introduction into Germany. *Vet Microbiol.* 2016;192: 60-66. doi: 10.1016/j.vetmic.2016.06.007.

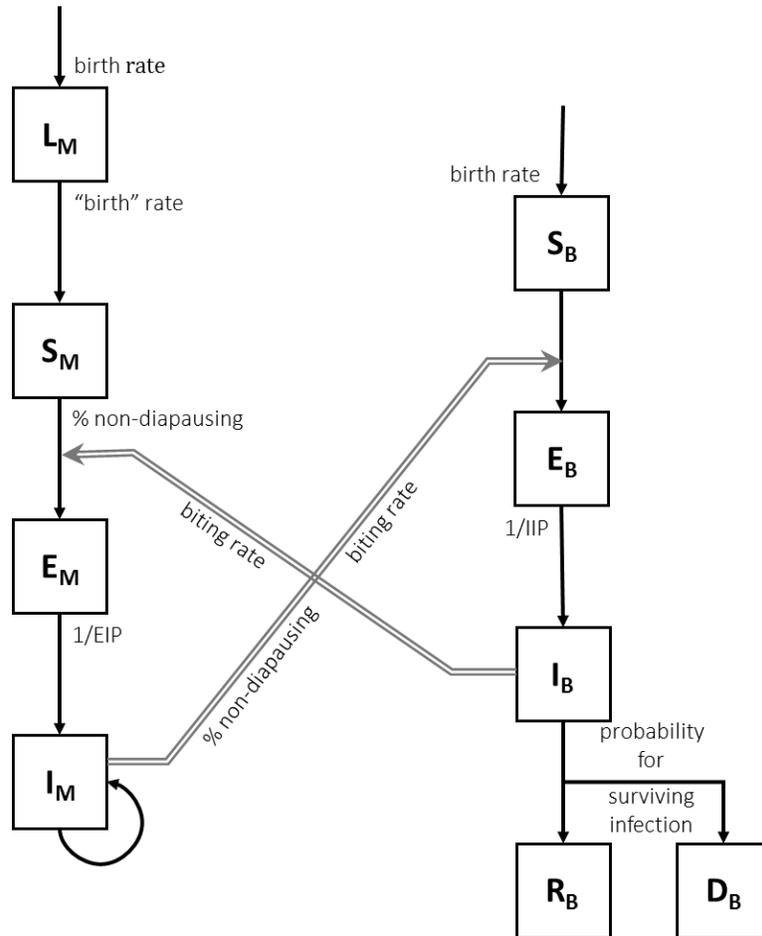
9. Hubálek Z, Rudolf I, Čapek M, Bakonyi T, Betášová L, Nowotny N. Usutu virus in blackbirds (*Turdus merula*), Czech Republic, 2011-2012. *Transbound Emerg Dis.* 2014;61(3): 273-276. doi: 10.1111/tbed.12025.
10. Ziegler U, Jöst H, Müller K, Fischer D, Rinder M, Tietze DT, et al. Epidemic spread of Usutu virus in southwest Germany in 2011 to 2013 and monitoring of wild birds for Usutu and West Nile viruses. *Vector Borne Zoonotic Dis.* 2015;15(8): 481-488. doi: 10.1089/vbz.2014.1746.
11. Garigliany MM, Marlier D, Tenner-Racz K, Eiden M, Cassart D, Gandar F, et al. Detection of Usutu virus in a bullfinch (*Pyrrhula pyrrhula*) and a great spotted woodpecker (*Dendrocopos major*) in north-west Europe. *Vet J.* 2014;199(1): 191-193. doi: 10.1016/j.tvjl.2013.10.017.
12. Grisenti M, Vázquez A, Herrero L, Cuevas L, Perez-Pastrana E, Arnoldi D, et al. Wide detection of Aedes flavivirus in north-eastern Italy - a European hotspot of emerging mosquito-borne diseases. *J Gen Virol.* 2015;96: 420-430. doi: 10.1099/vir.0.069625-0.
13. Cadar D, Becker N, Campos RDM, Börstler J, Jöst H, Schmidt-Chanasit J. Usutu virus in bats, Germany, 2013. *Emerg Infect Dis.* 2014;20(10): 1771-1773. doi: 10.3201/eid2010.140909.
14. Calzolari M, Pautasso A, Montarsi F, Albieri A, Bellini R, Bonilauri P, et al. West Nile virus surveillance in 2013 via mosquito screening in northern Italy and the influence of weather on virus circulation. *PLoS One.* 2015;10(10): e0140915. doi: 10.1371/journal.pone.0140915.
15. Lecollinet S, Blanchard Y, Manson C, Lowenski S, Laloy E, Quenault H, et al. Dual emergence of Usutu virus in common blackbirds, eastern France, 2015. *Emerg Infect Dis.* 2016;22(12): 2225-2227. doi: 10.3201/eid2212.161272.
16. Rijks J, Kik M, Slaterus R, Foppen R, Stroo A, Ijzer J, et al. Widespread Usutu virus outbreak in birds in the Netherlands, 2016. *Euro Surveill.* 2016;21(45): pii=30391. doi: 10.2807/1560-7917.es.2016.21.45.30391.

## Appendix 2. Buffer radii vs. model performance



### Appendix 3. Detailed description of the SEIR model.

A simplified diagram chart of Usutu virus (USUV) epidemiological model:



Each health state of mosquitoes / black birds can be described by Ordinary differential equations (ODEs).

Population growth of black birds:

$$\frac{dN_B}{dt} = r_B N_B = b_B N_B - m_B N_B$$

$N_B$  is the total number of black birds,  $r_B$  is the population growth rate,  $b_B$  is the birth rate and  $m_B$  is the mortality rate (B stands for black birds).

Follow logistic population growth (density dependent model):

$$\frac{dN_B}{dt} = r_B \left(1 - \frac{N_B}{K_B}\right) N_B$$

As

$$r_B = b_B - m_B$$

it can be written as:

$$\frac{dN_B}{dt} = \left( b_B - (b_B - m_B) \frac{N_B}{K_B} \right) N_B - m_B N_B$$

$K_B$  stands for environmental capacity. It can be understood as the maximum number of individuals that can be supported by the environment under ideal conditions.

The population of “larval” mosquitoes (includes all aquatic stages of *Culex* mosquitoes, only females taken into account) also follows logistic population growth:

$$\frac{dL_M}{dt} = (b_L N_M - m_L L_M) \left( 1 - \frac{L_M}{K_M} \right) - b_M L_M$$

$L_M$  is the total number of larvae,  $b_L$  is the birth rate of larvae,  $m_L$  is the mortality rate of larvae,  $b_M$  is the “birth rate” of mosquitoes (transformation from larvae to adult mosquitoes). Note here: although also following logistic population growth, mosquito growth is divided to aquatic and terrestrial stages, thus the equation looks different from black birds’.

Total density of terrestrial stages of *Culex* mosquitoes ( $N_M$ ):

$$\frac{dN_M}{dt} = b_M L_M - m_M N_M$$

Cross-infection between mosquitoes and black birds:

$$\lambda_B = \beta_B \frac{I_B}{K_B} = \kappa P_B \frac{I_B}{K_B}$$

$$\lambda_M = \beta_M \frac{I_M}{K_B} = \kappa P_M \frac{I_M}{K_B}$$

$\lambda_B$  denote the possible fraction of cross-transmission from birds to mosquitoes, and  $\lambda_M$  vice versa.  $\beta_B$  is the product of biting rate ( $\kappa$ ) and transmission possibility from birds to mosquitoes ( $P_B$ ), and  $\beta_M$  vice versa. Transmission possibility from mosquitoes to birds is  $P_M$ .

Then the different health states of birds can be described by following ODEs:

1. The susceptible black bird population ( $S_B$ )

$$\frac{dS_B}{dt} = \left( b_B - (b_B - m_B) \frac{N_B}{K_B} \right) N_B - m_B N_B - \delta_M \lambda_M S_B$$

It can be understood as:

(current total number of susceptible black birds) = (current total number birds) – (natural death of birds) – (birds moving to the next health state)

Here “natural death of birds” means deaths not due to Usutu virus (USUV) infection.

2. The exposed black bird population ( $E_B$ )

$$\frac{dE_B}{dt} = \delta_M \lambda_M S_B - m_B E_B - \gamma_B E_B$$

$\delta_M$  : Percentage of non-hibernating mosquitoes

$\gamma_B$  : The exposed – infected/infectious rate of birds

From this equation:

(current total number of exposed black birds) = (birds coming into this health state from the previous stage) – (natural death of birds) – (birds moving to the next health state)

3. The infected black bird population ( $I_B$ )

$$\frac{dI_B}{dt} = \gamma_B E_B - m_B I_B - \alpha_B I_B$$

$\alpha_B$ : The removal rate, removed from the previous health state, either get recovered (immunized) or dead

Similar as 2.

4. The black bird deaths ( $D_B$ ) due to USUV infection

$$\frac{dD_B}{dt} = \nu_B \alpha_B I_B$$

$\nu_B$ : the percentage of bird deaths due to USUV infection

5. The recovered black bird population ( $R_B$ )

$$\frac{dR_B}{dt} = (1 - \nu_B) \alpha_B I_B - m_B R_B$$

And

$$N_B = S_B + E_B + I_B + R_B$$

Note: In this model both horizontal and vertical virus transmission in birds are not taken into account, so the transmission is limited to through mosquitoes' blood meal.

Similarly, the different health states of mosquitoes are described as following:

6. The larval population of *Culex* mosquitoes:

$$\frac{dL_M}{dt} = (b_L N_M - m_L L_M) \left(1 - \frac{L_M}{K_M}\right) - b_M L_M$$

7. The susceptible mosquito population:

$$\frac{dS_M}{dt} = b_M L_M - m_M S_M - \delta_M \lambda_B S_M$$

From this equation, similar to bird equations:

(current total number of susceptible mosquitoes) = (mosquitoes entering this health state from the previous stage) – (natural death of mosquitoes) – (mosquitoes moving to the next health state).

8. The exposed mosquito population:

$$\frac{dE_M}{dt} = \delta_M \lambda_B S_M - m_M E_M - \gamma_M E_M$$

$\gamma_M$ : The exposed – infected/infectious rate of mosquitoes

9. The infected mosquito population:

$$\frac{dI_M}{dt} = \gamma_M E_M - m_M I_M$$

And

$$N_M = S_M + E_M + I_M$$

Note: Infectious mosquitoes remain in the infectious state and will not get recovered.

In addition,  $\delta_M$  is determined by the latitude and the calendar day of the year.

$$\delta_M = 1 - \frac{1}{1 + 1775.7 \exp[1.559(D - 18.177)]}$$

Of which  $D$  denotes “Daytime length”, and

$$D = 7.639 \arcsin \left[ \tan(\epsilon) \tan(\varphi) + \frac{0.0146}{\cos(\epsilon) \cos(\varphi)} \right] + 12$$

$$\epsilon = 0.409 \sin \left( \frac{2\pi(d - 80)}{365} \right)$$

$\varphi$ : Geographic latitude

$d$ : The calendar day

The final  $R_0$  equation:

$$R_0 = \sqrt{\left[ \frac{\delta_M \gamma_M \beta_M}{(\gamma_M + m_M) m_M K_B} \frac{S_B}{} \right] \left[ \frac{\delta_M \gamma_B \beta_B}{(\gamma_B + m_B) (\alpha_B + m_B) K_B} \frac{S_M}{} \right]}$$

**Additional Table 1:** Parameters for  $R_0$  equation

	parameter	value
the population growth rate	$r_B$	$r_B = b_B - m_B$
birth rate	$b_B$	$b_B(d) = 0.125 \frac{(x/\beta)^{\alpha-1} \exp(-x/\beta)}{\beta \Gamma(\alpha)}$
		$x = 0.1(d - 105)$ , $d$ is transformed Julian calendar day $\alpha=1.52$ , $\beta=1.93$ , $\Gamma(\alpha)=0.887$
mortality rate	$m_B$	0.0012
the birth rate of larvae	$b_L$	$b_L(T) = 2.325\kappa(T)$
		$T$ : Daily Mean Temperature
mortality rate of larvae	$m_L$	$m_L(T) = 0.0025T^2 - 0.094T + 1.0257$
“birth rate” of mosquitoes (transformation from larvae to adult mosquitoes).	$b_M$	$b_M(T) = 0.1b_L$
mortality rate of mosquitoes	$m_M$	$m_M(T) = 0.1m_L$
possible fraction of cross-transmission from birds to mosquitoes	$\lambda_B$	$\lambda_B(T) = \beta_B(T) \frac{I_B}{K_B} = \kappa(T) P_B \frac{I_B}{K_B}$
product of biting rate ( $\kappa$ ) and transmission possibility from birds to mosquitoes ( $P_B$ )	$\beta_B$	$\beta_B(T) = \kappa(T) P_B$ $P_B=0.125$
biting rate	$\kappa$	$\kappa(T) = \frac{0.344}{1 + 1.231 \exp(-0.184(T - 20))}$
possible fraction of cross-transmission from mosquitoes to birds	$\lambda_M$	$\lambda_M(T) = \beta_M(T) \frac{I_M}{K_B} = \kappa(T) P_M \frac{I_M}{K_B}$
product of biting rate ( $\kappa$ ) and transmission possibility from mosquitoes to birds ( $P_M$ )	$\beta_M$	$\beta_M(T) = P_M \kappa(T)$

	parameter	value
Percentage of non-hibernating mosquitoes	$\delta_M$	$\delta_M = 1 - \frac{1}{1 + 1775.7 \exp[1.559(D - 18.177)]}$ $D = 7.639 \arcsin \left[ \tan(\epsilon) \tan(\varphi) + \frac{0.0146}{\cos(\epsilon) \cos(\varphi)} \right] + 12$ $\epsilon = 0.409 \sin \left( \frac{2\pi(d - 80)}{365} \right)$
exposed – infected/infectious rate of birds	$\gamma_B$	0.667
removal rate, removed from the previous health state, either get recovered (immunized) or dead	$\alpha_B$	0.182
the percentage of bird deaths due to USUV infection	$\nu_B$	0.3
The exposed – infected/infectious rate of mosquitoes	$\gamma_M$	$\gamma_M(T) = 0.0093T - 0.1352, T \geq 15^\circ$ $\gamma_M(T) = 0, T < 15^\circ$

\* Note that highlighted parameters are also documented in Table 2.

# Appendix 1: An epidemiological model on West Nile virus

## Literature checking & what kind of model we need

In terms of modeling the temporal outbreak risk of West Nile virus (WNV), there are several epidemiological models available (21 in total, see Figure S1). Among these models, however, only few of them take into account of temperature related parameters of mosquitoes, e.g. (Kioutsioukis & Stilianakis, 2019; Laperriere et al., 2011; McMillan et al., 2020; Rubel et al., 2008). As WNV is transmitted by mosquitoes, temperature plays an important role in WNV transmission and circulation. Here comes the first requirement of the WNV model: (1) it should include temperature-dependent variables, especially for mosquitoes. (I will apply the temperature related parameters collected from these papers.) Another critical aspect that needs to be considered is that (2) not only local birds are contributing to the WNV circulation, but also migratory birds, especially for temporal risk dynamics of WNV transmission. To tackle this, Bergsman's model (Bergsman et al., 2016) included both local birds and migratory birds. However, this model is an **SIR** model instead of an **SEIR** model --The *exposed* state of bird hosts is discarded. Recently, there is a paper included both local birds and migratory birds (McMillan et al., 2020), and this model is an **SEIR** model. However, in this paper, the authors introduced clinical and sub-clinical states of the hosts. After all, it is not recommended to introduce new parameters if there is no obvious benefit: (3) a WNV model that is sophisticated and still neat is preferred in this case.

Up to date, there has been no published models that fit all our requirements. Consequently, herewith we construct one from the scratch. This model is an SEIR model, i.e. it includes health states such as susceptible, exposed, infected/infectious, and removed/recovered of both vectors and hosts. The framework of this model is similar to that of (Kioutsioukis & Stilianakis, 2019; Laperriere et al., 2011; Rubel et al., 2008).

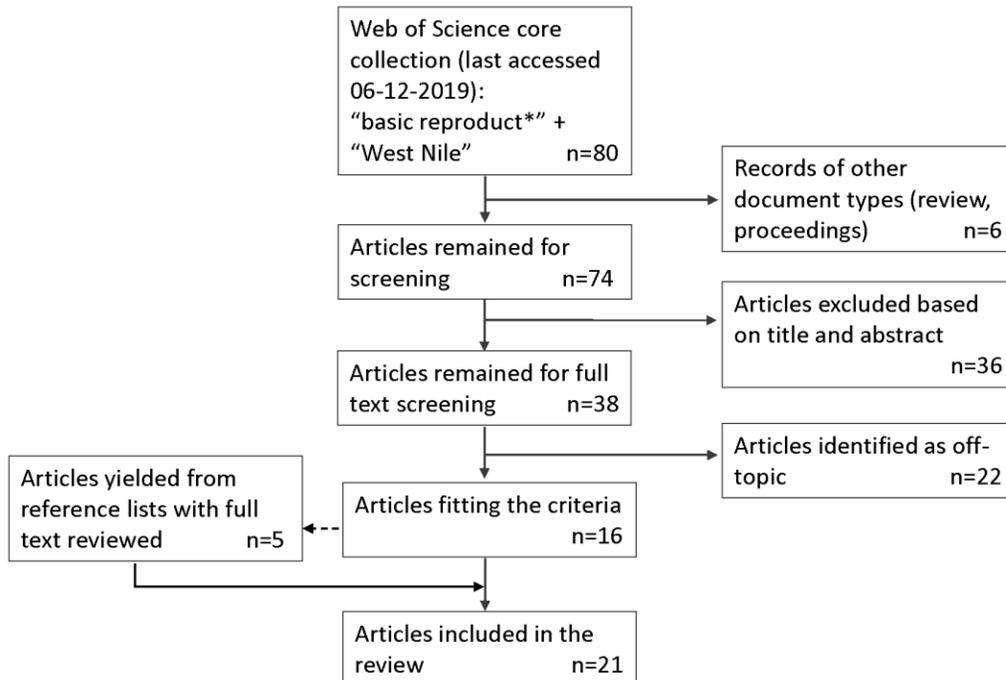


Figure S1. Literature review via Web of Science.

### Model construction

Each health state of mosquitoes / birds can be described by Ordinary differential equations (ODEs).

Population growth of local birds:

$$\frac{dN_{B_1}}{dt} = r_{B_1}N_{B_1} = b_{B_1}N_{B_1} - m_{B_1}N_{B_1}$$

$N_{B_1}$  is the total number of local birds,  $r_{B_1}$  is the population growth rate,  $b_{B_1}$  is the birth rate and  $m_{B_1}$  is the mortality rate ( $B_1$  stands for local birds).

Follow logistic population growth (density dependent model):

$$\frac{dN_{B_1}}{dt} = r_{B_1} \left( 1 - \frac{N_{B_1}}{K_{B_1}} \right) N_{B_1}$$

As

$$r_{B_1} = b_{B_1} - m_{B_1}$$

it can be written as:

$$\frac{dN_{B_1}}{dt} = \left( b_{B_1} - (b_{B_1} - m_{B_1}) \frac{N_{B_1}}{K_{B_1}} \right) N_{B_1} - m_{B_1}N_{B_1}$$

$K_{B_1}$  stands for environmental capacity. It can be understood as the maximum number of individuals that can be supported by the environment under ideal conditions.

The population of "larval" mosquitoes (includes all aquatic stages of mosquitoes, only females taken into account) also follows logistic population growth:

$$\frac{dL_M}{dt} = (b_L N_M - m_L L_M) \left( 1 - \frac{L_M}{K_M} \right) - b_M L_M$$

$L_M$  is the total number of larvae,  $b_L$  is the birth rate of larvae,  $m_L$  is the mortality rate of larvae,  $b_M$  is the "birth rate" of mosquitoes (transformation from larvae to adult mosquitoes).

Note here: although also following logistic population growth, mosquito growth is divided to aquatic and terrestrial stages, thus the equation looks different from local birds'.

Total density of terrestrial stages of mosquitoes ( $N_M$ ):

$$\frac{dN_M}{dt} = b_M L_M - m_M N_M$$

Cross-infection between mosquitoes and local birds:

$$\lambda_{B_1 M} S_M = C_1 \frac{I_{B_1}}{K_{B_1} + K_{B_2}} S_M$$

$$\lambda_{M B_1} S_{B_1} = \beta_1 \frac{I_M}{K_{B_1} + K_{B_2}} S_{B_1}$$

$\lambda_{B_1M}$  denote the possible fraction of cross-transmission from local birds to mosquitoes, and  $\lambda_{MB_1}$  vice versa.  $\beta_1$  is the product of biting rate and transmission possibility from birds to mosquitoes, and  $C_1$  vice versa.

Cross-infection between mosquitoes and migratory birds:

$$\lambda_{B_2M}S_M = C_1 \frac{I_{B_2}}{K_{B_1} + K_{B_2}} S_M$$

$$\lambda_{MB_2}S_{B_2} = \beta_1 \frac{I_M}{K_{B_1} + K_{B_2}} S_{B_2}$$

Then the different health states of birds can be described by following ODEs (for local birds we use script  $B_1$ , and for migratory birds use  $B_2$ ):

1. The susceptible local bird population ( $S_B$ )

$$\frac{dS_{B_1}}{dt} = \left( b_{B_1} - (b_{B_1} - m_{B_1}) \frac{N_{B_1}}{K_{B_1}} \right) N_{B_1} - m_{B_1} N_{B_1} - \frac{\beta_1 I_M}{K_{B_1} + K_{B_2}} S_{B_1}$$

It can be understood as:

(current total number of susceptible local birds) = (current total number birds) – (natural death of birds) – (birds moving to the next health state)

Here “natural death of birds” means deaths not due to West Nile virus (WNV) infection.

2. The exposed local bird population ( $E_B$ ), here is for local birds

$$\frac{dE_{B_1}}{dt} = \frac{\beta_1 I_M}{K_{B_1} + K_{B_2}} S_{B_1} - m_{B_1} E_{B_1} - \gamma_{B_1} E_{B_1}$$

$\delta_M$  : Percentage of non-hibernating mosquitoes

$\gamma_B$  : The exposed – infected/infectious rate of birds

From this equation:

(current total number of exposed local birds) = (birds coming into this health state from the previous stage) – (natural death of birds) – (birds moving to the next health state)

3. The infected local bird population ( $I_B$ )

$$\frac{dI_{B_1}}{dt} = \gamma_{B_1} E_{B_1} - m_{B_1} I_{B_1} - \alpha_{B_1} I_{B_1}$$

$\alpha_B$ : The removal rate, removed from the previous health state, either get recovered (immunized) or dead

Similar as 2.

4. The local bird deaths ( $D_B$ ) due to WNV infection

$$\frac{dD_{B_1}}{dt} = \nu_{B_1} \alpha_{B_1} I_{B_1}$$

$\nu_B$ : the percentage of bird deaths due to WNV infection

5. The recovered local bird population ( $R_B$ )

$$\frac{dR_{B_1}}{dt} = (1 - \nu_{B_1}) \alpha_{B_1} I_{B_1} - m_{B_1} R_{B_1}$$

And

$$N_B = S_B + E_B + I_B + R_B$$

Note: In this model both horizontal and vertical virus transmission in birds are not taken into account, so the transmission is limited to through mosquitoes' blood meal.

Similarly, the different health states of mosquitoes are described as following:

6. The larval population of mosquitoes:

$$\frac{dL_M}{dt} = (b_L N_M - m_L L_M) \left(1 - \frac{L_M}{K_M}\right) - b_M L_M$$

7. The susceptible mosquito population:

$$\frac{dS_M}{dt} = b_M L_M - m_M S_M - \frac{C_1 I_{B_1} + C_2 I_{B_2}}{K_{B_1} + K_{B_2}} S_M$$

From this equation, similar to bird equations:

(current total number of susceptible mosquitoes) = (mosquitoes entering this health state from the previous stage) – (natural death of mosquitoes) – (mosquitoes moving to the next health state).

8. The exposed mosquito population:

$$\frac{dE_M}{dt} = \frac{C_1 I_{B_1} + C_2 I_{B_2}}{K_{B_1} + K_{B_2}} S_M - m_M E_M - \gamma_M E_M$$

$\gamma_M$ : The exposed – infected/infectious rate of mosquitoes

9. The infected mosquito population:

$$\frac{dI_M}{dt} = \gamma_M E_M - m_M I_M$$

And

$$N_M = S_M + E_M + I_M$$

Note: Infectious mosquitoes remain in the infectious state and will not get recovered.

### The ODE matrix of WNV transmission

Including mosquitoes, local birds and migratory birds

$$d \begin{pmatrix} E_{B_1} \\ I_{B_1} \\ E_{B_2} \\ I_{B_2} \\ E_M \\ I_M \end{pmatrix} = \begin{pmatrix} \frac{\beta_1 I_M}{K_{B_1} + K_{B_2}} S_{B_1} \\ 0 \\ \frac{\beta_2 I_M}{K_{B_1} + K_{B_2}} S_{B_2} \\ 0 \\ \frac{c_1 I_{B_1} + c_2 I_{B_2}}{K_{B_1} + K_{B_2}} S_M \\ 0 \end{pmatrix} - \begin{pmatrix} \gamma_{B_1} E_{B_1} + m_{B_1} E_{B_1} \\ -\gamma_{B_1} E_{B_1} + \alpha_{B_1} I_{B_1} + m_{B_1} I_{B_1} \\ \gamma_{B_2} E_{B_2} + m_{B_2} E_{B_2} \\ -\gamma_{B_2} E_{B_2} + \alpha_{B_2} I_{B_2} + m_{B_2} I_{B_2} \\ \gamma_M E_M + m_M E_M \\ -\gamma_M E_M + m_M E_M \end{pmatrix}$$

then the Jacobian matrix:

$$T = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & \frac{\beta_1}{K_{B_1} + K_{B_2}} S_{B_1} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{\beta_2}{K_{B_1} + K_{B_2}} S_{B_2} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{c_1}{K_{B_1} + K_{B_2}} S_M & 0 & \frac{c_2}{K_{B_1} + K_{B_2}} S_M & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\Sigma = \begin{pmatrix} \gamma_{B_1} + m_{B_1} & 0 & 0 & 0 & 0 & 0 \\ -\gamma_{B_1} & \alpha_{B_1} + m_{B_1} & 0 & 0 & 0 & 0 \\ 0 & 0 & \gamma_{B_2} + m_{B_2} & 0 & 0 & 0 \\ 0 & 0 & -\gamma_{B_2} & \alpha_{B_2} + m_{B_2} & 0 & 0 \\ 0 & 0 & 0 & 0 & \gamma_M + m_M & 0 \\ 0 & 0 & 0 & 0 & -\gamma_M & m_M \end{pmatrix}$$

Find the inverse matrix of  $\Sigma$ , ( $\Sigma^{-1}$ )

To keep the deduction process simple, here we use the elements symbol instead of full ODE expressions

$$\Sigma|E_1 = \left( \begin{array}{cccccc|cccccc} a_{11} & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ a_{21} & a_{22} & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & a_{33} & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & a_{43} & a_{44} & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & a_{55} & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & a_{65} & a_{66} & 0 & 0 & 0 & 0 & 0 & 1 \end{array} \right)$$

$$E_1|\Sigma^{-1} = \left( \begin{array}{cccccc|cccccc} & & & & & & \frac{1}{a_{11}} & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & -\left(\frac{a_{21}}{a_{11}} \cdot \frac{1}{a_{22}}\right) & \frac{1}{a_{22}} & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{a_{33}} & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -\left(\frac{a_{43}}{a_{33}} \cdot \frac{1}{a_{44}}\right) & \frac{1}{a_{44}} & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{a_{55}} & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -\left(\frac{a_{65}}{a_{55}} \cdot \frac{1}{a_{66}}\right) & \frac{1}{a_{66}} \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right)$$

$$\Sigma^{-1} = \left( \begin{array}{cccccc} \frac{1}{a_{11}} & 0 & 0 & 0 & 0 & 0 \\ -\left(\frac{a_{21}}{a_{11}} \cdot \frac{1}{a_{22}}\right) & \frac{1}{a_{22}} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{a_{33}} & 0 & 0 & 0 \\ 0 & 0 & -\left(\frac{a_{43}}{a_{33}} \cdot \frac{1}{a_{44}}\right) & \frac{1}{a_{44}} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{a_{55}} & 0 \\ 0 & 0 & 0 & 0 & -\left(\frac{a_{65}}{a_{55}} \cdot \frac{1}{a_{66}}\right) & \frac{1}{a_{66}} \end{array} \right)$$

Now replace the element symbols with original ODE expressions

$$\Sigma^{-1} = \begin{pmatrix} \frac{1}{\gamma_{B_1} + m_{B_1}} & 0 & 0 & 0 & 0 & 0 \\ -\left(\frac{-\gamma_{B_1}}{\gamma_{B_1} + m_{B_1}} \cdot \frac{1}{\alpha_{B_1} + m_{B_1}}\right) & \frac{1}{\alpha_{B_1} + m_{B_1}} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\gamma_{B_2} + m_{B_2}} & 0 & 0 & 0 \\ 0 & 0 & -\left(\frac{-\gamma_{B_2}}{\gamma_{B_2} + m_{B_2}} \cdot \frac{1}{\alpha_{B_2} + m_{B_2}}\right) & \frac{1}{\alpha_{B_2} + m_{B_2}} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{\gamma_M + m_M} & 0 \\ 0 & 0 & 0 & 0 & -\left(\frac{-\gamma_M}{\gamma_M + m_M} \cdot \frac{1}{m_M}\right) & \frac{1}{m_M} \end{pmatrix}$$

$K_L = -T \times \Sigma^{-1}$ ; while in the  $T$  matrix, there are only 3 non-zeros rows, follow (Diekmann et al., 2010)

$$K = -E_2' T \Sigma^{-1} E_2$$

$$E_2 = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{vmatrix}; E_2' = \begin{vmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{vmatrix}$$

$$E_2' T = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & \frac{\beta_1}{K_{B_1} + K_{B_2}} S_{B_1} \\ 0 & 0 & 0 & 0 & 0 & \frac{\beta_2}{K_{B_1} + K_{B_2}} S_{B_2} \\ 0 & \frac{c_1}{K_{B_1} + K_{B_2}} S_M & 0 & \frac{c_2}{K_{B_1} + K_{B_2}} S_M & 0 & 0 \end{pmatrix}$$

$$\Sigma^{-1} E_2 = \begin{pmatrix} \frac{1}{\gamma_{B_1} + m_{B_1}} & 0 & 0 \\ -\left(\frac{-\gamma_{B_1}}{\gamma_{B_1} + m_{B_1}} \cdot \frac{1}{\alpha_{B_1} + m_{B_1}}\right) & 0 & 0 \\ 0 & \frac{1}{\gamma_{B_2} + m_{B_2}} & 0 \\ 0 & -\left(\frac{-\gamma_{B_2}}{\gamma_{B_2} + m_{B_2}} \cdot \frac{1}{\alpha_{B_2} + m_{B_2}}\right) & 0 \\ 0 & 0 & \frac{1}{\gamma_M + m_M} \\ 0 & 0 & -\left(\frac{-\gamma_M}{\gamma_M + m_M} \cdot \frac{1}{m_M}\right) \end{pmatrix}$$

$$K = -E_2' T \Sigma^{-1} E_2$$

=

$$- \begin{pmatrix} 0 & 0 & \frac{\beta_1}{K_{B_1}+K_{B_2}} S_{B_1} \left( \frac{\gamma_M}{\gamma_M+m_M} \cdot \frac{1}{m_M} \right) \\ 0 & 0 & \frac{\beta_2}{K_{B_1}+K_{B_2}} S_{B_2} \left( \frac{\gamma_M}{\gamma_M+m_M} \cdot \frac{1}{m_M} \right) \\ \frac{c_1}{K_{B_1}+K_{B_2}} S_M \left( \frac{\gamma_{B_1}}{\gamma_{B_1}+m_{B_1}} \cdot \frac{1}{\alpha_{B_1}+m_{B_1}} \right) & \frac{c_2}{K_{B_1}+K_{B_2}} S_M \left( \frac{\gamma_{B_2}}{\gamma_{B_2}+m_{B_2}} \cdot \frac{1}{\alpha_{B_2}+m_{B_2}} \right) & 0 \end{pmatrix}$$

$$E_3 = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix}$$

$\lambda E_3 -$

$$K = \begin{pmatrix} \lambda & 0 & \frac{\beta_1}{K_{B_1}+K_{B_2}} S_{B_1} \left( \frac{\gamma_M}{\gamma_M+m_M} \cdot \frac{1}{m_M} \right) \\ 0 & \lambda & \frac{\beta_2}{K_{B_1}+K_{B_2}} S_{B_2} \left( \frac{\gamma_M}{\gamma_M+m_M} \cdot \frac{1}{m_M} \right) \\ \frac{c_1}{K_{B_1}+K_{B_2}} S_M \left( \frac{\gamma_{B_1}}{\gamma_{B_1}+m_{B_1}} \cdot \frac{1}{\alpha_{B_1}+m_{B_1}} \right) & \frac{c_2}{K_{B_1}+K_{B_2}} S_M \left( \frac{\gamma_{B_2}}{\gamma_{B_2}+m_{B_2}} \cdot \frac{1}{\alpha_{B_2}+m_{B_2}} \right) & \lambda \end{pmatrix}$$

$$\begin{aligned} &= \lambda^3 - \frac{\beta_1}{K_{B_1}+K_{B_2}} S_{B_1} \left( \frac{\gamma_M}{\gamma_M+m_M} \cdot \frac{1}{m_M} \right) (\lambda) \frac{c_1}{K_{B_1}+K_{B_2}} S_M \left( \frac{\gamma_{B_1}}{\gamma_{B_1}+m_{B_1}} \cdot \frac{1}{\alpha_{B_1}+m_{B_1}} \right) - \\ & (\lambda) \frac{\beta_2}{K_{B_1}+K_{B_2}} S_{B_2} \left( \frac{\gamma_M}{\gamma_M+m_M} \cdot \frac{1}{m_M} \right) \frac{c_2}{K_{B_1}+K_{B_2}} S_M \left( \frac{\gamma_{B_2}}{\gamma_{B_2}+m_{B_2}} \cdot \frac{1}{\alpha_{B_2}+m_{B_2}} \right) \\ &= \lambda \left( \lambda^2 - \frac{\beta_1}{K_{B_1}+K_{B_2}} S_{B_1} \left( \frac{\gamma_M}{\gamma_M+m_M} \cdot \frac{1}{m_M} \right) \frac{c_1}{K_{B_1}+K_{B_2}} S_M \left( \frac{\gamma_{B_1}}{\gamma_{B_1}+m_{B_1}} \cdot \frac{1}{\alpha_{B_1}+m_{B_1}} \right) - \right. \\ & \left. \frac{\beta_2}{K_{B_1}+K_{B_2}} S_{B_2} \left( \frac{\gamma_M}{\gamma_M+m_M} \cdot \frac{1}{m_M} \right) \frac{c_2}{K_{B_1}+K_{B_2}} S_M \left( \frac{\gamma_{B_2}}{\gamma_{B_2}+m_{B_2}} \cdot \frac{1}{\alpha_{B_2}+m_{B_2}} \right) \right) \end{aligned}$$

The eigenvalue of this matrix:

$$\lambda_1 = 0$$

$$\lambda_2 =$$

$$\sqrt{\frac{\beta_1}{K_{B_1}+K_{B_2}} S_{B_1} \left( \frac{\gamma_M}{\gamma_M+m_M} \cdot \frac{1}{m_M} \right) \frac{c_1}{K_{B_1}+K_{B_2}} S_M \left( \frac{\gamma_{B_1}}{\gamma_{B_1}+m_{B_1}} \cdot \frac{1}{\alpha_{B_1}+m_{B_1}} \right) + \frac{\beta_2}{K_{B_1}+K_{B_2}} S_{B_2} \left( \frac{\gamma_M}{\gamma_M+m_M} \cdot \frac{1}{m_M} \right) \frac{c_2}{K_{B_1}+K_{B_2}} S_M \left( \frac{\gamma_{B_2}}{\gamma_{B_2}+m_{B_2}} \cdot \frac{1}{\alpha_{B_2}+m_{B_2}} \right)}$$

$$= \frac{1}{K_{B_1}+K_{B_2}} \sqrt{\frac{\gamma_M S_M}{m_M (\gamma_M+m_M)} \left( \frac{\gamma_{B_1} \beta_1 c_1 S_{B_1}}{(\gamma_{B_1}+m_{B_1})(\alpha_{B_1}+m_{B_1})} + \frac{\gamma_{B_2} \beta_2 c_2 S_{B_2}}{(\gamma_{B_2}+m_{B_2})(\alpha_{B_2}+m_{B_2})} \right)}$$

$$\lambda_3 = -\lambda_2$$

So the largest eigenvalue of this NGM is  $\lambda_2$ .

$$R_0 = \lambda_2$$

From this equation, the migratory bird compartments can be easily removed as well; then the equation would be the same as that of a simpler model with only local bird as host.

### References

- Bergsman, L. D., Hyman, J. H., & Manore, C. A., 2016. A mathematical model for the spread of West Nile virus in migratory and resident birds. *Mathematical Biosciences and Engineering*, 13, 401-424.
- Diekmann, O., Heesterbeek, J. A. P., & Roberts, M. G., 2010. The construction of next-generation matrices for compartmental epidemic models. *Journal of the Royal Society Interface*, 7, 873-885.
- Kioutsioukis, I., & Stilianakis, N. I., 2019. Assessment of West Nile virus transmission risk from a weather-dependent epidemiological model and a global sensitivity analysis framework. *Acta Tropica*, 193, 129-141.
- Laperriere, V., Brugger, K., & Rubel, F., 2011. Simulation of the seasonal cycles of bird, equine and human West Nile virus cases. *Preventive Veterinary Medicine*, 98, 99-110.
- McMillan, J. R., Armstrong, P. M., & Andreadis, T. G., 2020. Patterns of mosquito and arbovirus community composition and ecological indexes of arboviral risk in the northeast United States. *Plos Neglected Tropical Diseases*, 14.
- Rubel, F., Brugger, K., Hantel, M., Chvala-Mannsberger, S., Bakonyi, T., Weissenböck, H., Nowotny, N., 2008. Explaining Usutu virus dynamics in Austria: Model development and calibration. *Preventive Veterinary Medicine*, 85, 166-186.

## List of abbreviations

ASU: Administrative Spatial Unit

CDC: Centers for Disease Control and Prevention (US)

COS: Change-of-Support

COVID-19: Coronavirus disease 2019

ECDC: European Center for Disease Prevention and Control

EIP: Extrinsic incubation period

EM: Epidemiological model

ENM: Ecological/Environmental niche model

HPV: human papillomavirus

NGM: Next generation matrix

ODE: Ordinary differential equations

STD: Sexually transmitted disease

USUV: Usutu virus

VBD: Vector-borne disease

WHO: World Health Organization

WNV: West Nile virus

## Other academic activities

### Conferences and training

<b>2018</b> Bayreuth, Oct. 11 <sup>th</sup>	<b>Bayreuth Center of Ecology and Environmental Research Workshop</b> Talk: Evaluating the Risk for Usutu virus circulation in Europe: Comparison of Environmental Niche Models and Epidemiological Models
Canary Islands, Spain Mar. 6- 20 <sup>th</sup>	<b>La Palma Science School</b> <b>Application of Remote Sensing for Protected Areas and Biodiversity</b>
<b>2017</b> Berlin, Oct. 12-13 <sup>th</sup>	<b>National Symposium on Zoonoses Research 2017</b> Poster: Modelling spatial risk of Usutu disease in Europe: a comparison between a species distribution model and an epidemic model
Bayreuth, July 10 -15 <sup>th</sup>	<b>Bayreuth International Summer School - Public Health &amp; Environment</b> <b>Impact of Environmental Changes and Diseases</b>
<b>2016</b> Berlin, Oct. 13–14 <sup>th</sup>	<b>National Symposium on Zoonoses Research 2016</b> Poster: Human Vector-Borne and Zoonotic Diseases: State of the Art and Research Challenges in Face of Climate Change and Globalization.
Berlin, Oct. 10–12 <sup>th</sup>	<b>Workshop: Introduction to Interdisciplinary Communication, Presentation and Collaboration</b>
<b>2014</b> Sacramento, USA, Aug. 10–15 <sup>th</sup>	<b>99<sup>th</sup> Annual Meeting of the Ecology Society of America</b> Talk: Abundance and occupancy of leopard and their prey in Wangqing Leopard Reserve, China

### Manuscript in preparation

Tjaden, NB; **Cheng, Y**; Beierkuhnlein, C; Tomas, SM: Chikungunya beyond the tropics: Where and when do we expect disease transmission in Europe?

## Teaching

Winters 2019/20  
*In English*

**Global Change Ecology/Progress in Global Change Research**  
co-taught with Dr. Anja Jaeschke

**Academic Working Methods and Skills**

co-taught with Dr. Anja Jaeschke and Dr. Stephanie Thomas

Summer 2018  
*In English*

**Public Health and Global Change**

Part 2 – Vector-borne diseases and Global Change. Bayreuth International  
Summer School 2018

Teaching: A brief introduction of epidemiological models & A case study based  
on Usutu disease

## Reviewer activities

Biosafety and Health

International Journal of Health Geographics

Plos One

## Acknowledgements

I remember the day I came to Germany, alone, with a small luggage filled with a rice cooker and a big blanket, and an old laptop without battery (surprisingly it's still alive). Back then, I barely knew anyone around. Through the last five years, my life has become way more interesting because of Carl and my friends here. The rice cooker has fed me and my friends here; the blanket is now the one I share with my partner; even the old laptop serves as the sports equipment for playing YouTube workout videos. No need to say, I am more than grateful to what I have now and the people I enjoy being around with.

First of all, I would like to thank Carl for all the help and support. Carl has constantly encouraged and inspired me to work on the topics I am interested in; and more importantly, Carl guided me into a confident person. I came to Germany even before knowing Carl in person, and I would never regret about it!

I would like to thank Martin, my first friend here, for our small but nice "Wandergruppe" and your kind encouragement through all these years; Reinhold for being a supportive and wise friend and the wonderful person who you are; Vanessa for the nice time we have been sharing together and the friendship we have in each other; Eva for our wonderful "Frauenverein" and the nice Yoga time, Samuel for our most stable lunch group (which consists only of you, Nils and me) and the inspiring yet still funny conversations we had together!

Many thanks to people who have helped me in China as well. I would like to thank Prof. Ge for guiding me and inspiring me when I was an undergraduate, Prof. Mou for encouraging me and all the nice coffee (probably the best!), Prof. Han for helping me with the scholarship, and of course, the China Scholarship Council for granting me the scholarship! I would like to thank my family in northeastern China as well for being supportive through these years, Sean for encouraging me to study abroad in the first place; Zhiyu, for your kind words and generous help. Special thanks go to Xiuying Sun, my teacher during my bachelor studies, thank you for helping me through the hardest time in my life.

Furthermore, I would like to express my thanks to my family in Germany as well, for supporting me and making Sauerland pretty much my second home here; to Stephanie, who has got me involved in the vector-borne disease group; to Khishigee and Anna, who are always so supportive and considerate; to Sabrina, who has helped me so much with enormous paperwork ever since she works here!

I would like to especially thank Anja. You inspired me into teaching and to participate the FBZHL courses. More importantly, you've made our office a nice and cozy place, and you always get me! I miss the times of us sitting together, we just couldn't stop chatting with each other!

Lastly, I would like to thank my partner Nils. Thank you for being such a nice and kind person, for being patient with me especially when I am cranky, for making me a better person; and of course, also for being a nice colleague! We surely make a good team! Thank you for always being there for me!

## **(Eidesstattliche) Versicherungen und Erklärungen**

(§ 5 Nr. 4 PromO)

*Hiermit erkläre ich, dass keine Tatsachen vorliegen, die mich nach den gesetzlichen Bestimmungen über die Führung akademischer Grade zur Führung eines Doktorgrades unwürdig erscheinen lassen.*

(§ 8 S. 2 Nr. 5 PromO)

*Hiermit erkläre ich mich damit einverstanden, dass die elektronische Fassung meiner Dissertation unter Wahrung meiner Urheberrechte und des Datenschutzes einer gesonderten Überprüfung hinsichtlich der eigenständigen Anfertigung der Dissertation unterzogen werden kann.*

(§ 8 S. 2 Nr. 7 PromO)

*Hiermit erkläre ich eidesstattlich, dass ich die Dissertation selbständig verfasst und keine anderen als die von mir angegebenen Quellen und Hilfsmittel benutzt habe.*

(§ 8 S. 2 Nr. 8 PromO)

*Ich habe die Dissertation nicht bereits zur Erlangung eines akademischen Grades anderweitig eingereicht und habe auch nicht bereits diese oder eine gleichartige Doktorprüfung endgültig nicht bestanden.*

(§ 8 S. 2 Nr. 9 PromO)

*Hiermit erkläre ich, dass ich keine Hilfe von gewerblichen Promotionsberatern bzw. -vermittlern in Anspruch genommen habe und auch künftig nicht nehmen werde.*

Bayreuth, 14.04.2021 .....

Ort, Datum, Unterschrift