Ecological Modelling 452 (2021) 109594

Contents lists available at ScienceDirect

Ecological Modelling

journal homepage: www.elsevier.com/locate/ecolmodel

Combined climate and regional mosquito habitat model based on machine learning

Ralf Wieland [a,*], Katrin Kuhls [a], Hartmut H.K. Lentz [b], Franz Conraths [b], Helge Kampen [b], Doreen Werner [a]

[a] Leibniz Centre for Agricultural Landscape Research, Eberswalder Str. 84, 15374 Müncheberg, Germany
[b] Friedrich-Loeffler-Institut Federal Research Institute for Animal Health, Hauptsitz Insel Riems, Südufer 10, 17493 Greifswald-Insel Riems, Germany

ARTICLE INFO

Keywords:
Citizen science data
Mosquito habitat modelling
Machine learning
XGBoost
West Nile virus
Vector borne diseases

ABSTRACT

Besides invasive mosquito species also several native species are proven or suspected vectors of arboviruses as West Nile or Usutu virus in Western Europe. Habitat models of these native vectors can be a helpful tool for assessing the risk of autochthonous occurrence, outbreaks and spread of diseases caused by such arboviruses. Modelling native mosquitoes is complicated because of the perfect adaptation to the climatic and landscape conditions and their high abundance in contrast to invasive species. Here we present a new approach for such a habitat model for native mosquito species in Germany, which are considered as vectors of West Nile virus (WNV). Epizootic emergence of WNV was registered in Germany since 2018. The models are based on surveillance data of mosquitoes from the German citizen science project "Mückenatlas" complemented by data from systematic trap monitoring in Germany, and on data freely available from the Deutscher Wetterdienst (DWD) and OpenStreetMap (OSM). While climatic factors still play an important role, we could show that habitat suitability is predictable only by the combination of the climate model with a regional model. Both models were based on a machine-learning approach using XGBoost. Evaluation of the accuracy of the models was done by statistical analysis, determining among others feature importances using the SHAP-Library. Final output of the combined climatic and regional models are maps showing the superposed habitat suitability which are generated through a number of steps described in detail. These maps also include the registered cases of WNV infections in the selected region of Germany.

1. Introduction

Mosquitoes are vectors of a wide range of arboviruses (arthropod-borne viruses). Diseases caused by such viruses usually occur in tropical and subtropical regions. However, a considerable increase of disease cases caused by West Nile, Usutu, Dengue and Chikungunya viruses was registered recently, with an accumulation of outbreaks in Western and Southern Europe (Martinet et al., 2019; Vilibic-Cavlek et al., 2019). Reasons of this trend are the growing number of imported cases of such infections and the expansion of invasive mosquito vectors as *Ae. albopictus* due to globalisation. Climate change is another important cause since the expansion of invasive mosquito species as well as the transmission dynamics of the viruses can be further enhanced by global warming (Reuss et al., 2018; Ciota and Keyel, 2019; Metelmann et al., 2019). As several arboviral diseases arrived meanwhile in Western Europe an important question is whether local mosquito species, which are well adapted and occur at high abundances and densities, can drive the transmission of the respective viruses. Information about vector competence of local mosquito species for the specific arboviruses is, however, still limited (Brugman et al., 2018; Martinet et al., 2019).

West Nile virus (WNV) is a widespread zoonotic arbovirus involving several bird species as natural reservoirs. An increasing number of outbreaks of WNV infection is reported from Southeastern and Western Europe in birds, horses and humans. Also in Germany WNV has been recognised as a threat to animal and public health since its first autochthonous epizootic emergence in 2018 (12 birds, two horses) and 2019 (76 birds, 36 horses) (Ziegler et al., 2019, 2020). In 2019 the first five autochthonous human cases were recorded. These outbreaks correlated with two of the warmest summers and early autumns in Germany of the last decades. WNV has been isolated from several native as well as invasive mosquito species in Western Europe (Engler et al., 2013; Martinet et al., 2019). Especially the invasive *Ae. japonicus* is considered as a potential key bridge vector for WNV in this region due to its high vector competence (Wagner et al., 2018). The first demonstration of WNV in mosquitoes in Germany has recently been

* Corresponding author.
E-mail address: rwieland@zalf.de (R. Wieland).

https://doi.org/10.1016/j.ecolmodel.2021.109594
Received 16 October 2020; Received in revised form 27 April 2021; Accepted 29 April 2021
Available online 6 May 2021

reported from *Cx. pipiens* (Kampen et al., 2020). Of the mosquitoes indigenous to Germany, *Cx. pipiens* (biotypes *pipiens* and *molestus*), *Cx. torrentium* and *Ae. vexans* are considered as the most important vectors, with *Cx. pipiens* (biotype *pipiens*) and *Cx. torrentium* showing the highest transmission efficiencies in experimental transmission studies (Jansen et al., 2019; Holicki et al., 2020; Wöhnke et al., 2020).

Mosquitoes require a number of climatic and structural features of a landscape for the different stages of their development. Habitat models can help in understanding the geographic distribution range of the vectors in unsampled areas in order to assess the risk of occurrence of specific arboviruses and to apply targeted surveillance. In the simplest case, models are developed only on the basis of bio-climatic data (mainly temperature and precipitation) (Lee et al., 2016; Valdez et al., 2018; Jácome et al., 2019). Prerequisite for the use of such models is a high variance of the climate in the landscape of the study area (e.g. transitions from high mountains to lowlands, often covering several climate zones) (Roiz et al., 2014; Paz, 2015; Hahn et al., 2015; Jácome et al., 2019; Cunze et al., 2020). This climatic variance does not exist in the northeastern German lowlands. In the case of invasive species, modelling is often still possible, as these species are not yet fully adapted to the regional climate. Invasive mosquitoes prefer regions with a climate comparable to their regions of origin (Früh et al., 2018; Kerkow et al., 2019).

The present paper focuses on modelling the habitat requirements of native mosquito species. The specificity of native mosquitoes is their perfect adaptation to the climatic and landscape conditions as well as their wide distribution and high abundance. The development of habitat models is therefore challenging in comparison to invasive vector species. The perfect adaptation of native mosquitoes leads to considerable lower accuracies of the models exclusively based on climatic variables in comparison to invasive species. This is why other additional variables are required which are reflecting the specificity of the respective regions as for example topography, vegetation or land use. Such structural features of the landscape can either be combined with the bio-climatic data to a model (Myer and Johnston, 2019), or they can be developed separately and then combined. In Kerkow et al. (2020) landscape structures were linked to the climate model via a fuzzy model. This powerful and very flexible approach requires expert knowledge e.g. of entomologists, for modelling. On the other hand, the dependence on experts who are willing and able to contribute to modelling is disadvantageous. Even if an expert were available, we would like to show a way to develop a model that is generated only from the data. The prerequisite is that sufficient data are available for machine learning. In the present work we tested whether a clever combination of climate-based models and models that take regional characteristics into account can be used to create reliable habitat models of indigenous mosquitoes.

Traditional monitoring and surveillance of disease-carrying mosquitoes covering the whole country (e.g. by using specific traps) is limited by the available financial and labour resources. To overcome these problems, passive surveillance activities have been launched in several European countries (Bartumeus et al., 2018; Kampen et al., 2015). Such citizen science programmes already made a large contribution in cataloguing the biodiversity of native and invasive mosquito species, in monitoring their distribution, and even in understanding the mode of spread of invasive species (Walther and Kampen, 2017). Although the number of collected specimens per site is low in comparison to the number obtained by traps, the large geographical coverage gives a roughly realistic picture of the distribution of the respective species. The disadvantage of such citizen science data (CSD) is that the locations where the submitted mosquitoes were collected most probably are related to the place of residence of the collectors rather than to the specific breeding sites of the respective mosquitoes. Typical mosquito areas may remain unsampled because only few people are living there. On the other hand, some of the most abundant species have an anthropophilic feeding behaviour and are found predominantly in human environments as e.g. *Cx. pipiens* biotype *molestus*, while other anthropophilic species frequently migrate over long distances from the breeding site to the preferred host, as e.g. *Ae. vexans* (Gutsevich et al., 1970; Vinogradova, 2000; Becker et al., 2010; Hamer et al., 2014; Verdonschot and Besse-Lototskaya, 2014).

In 2012, such a citizen science project called "Mückenatlas" has been launched in Germany as part of a nation-wide mosquito monitoring programme supervised by the Leibniz Centre for Agricultural Landscape Research (ZALF) and the German Federal Research Institute for Animal Health - Friedrich-Loeffler-Institut (FLI) (Kampen et al., 2015). Data from both, the citizen science project and routine field collections are continuously submitted to the German mosquito database CULBASE. Although collection sites of "Mückenatlas" submissions often concentrate in densely colonised areas in and around larger cities, there is a good matching of passive and active monitoring so far, e.g. for invasive species, showing the potential of CSD to complement data from traditional monitoring (Walther and Kampen, 2017).

In the present study we used "Mückenatlas" data as well as data from field studies extracted from CULBASE to address the following questions:

- Is it possible to combine climatic and regional factors to achieve models with high accuracies specifically for native mosquitoes?
- Is it possible to implement the habitat model on the base of machine learning without using expert knowledge?
- Is it sufficient to use CSD for such models or is there a need to supplement these data by traditional targeted monitoring data?

In this work, we present a new approach for modelling highly adapted native mosquito species using data from a citizen science programme, at the same time showing the high potential of such data for the development of surveillance and control measures of vector-borne diseases.

2. Methods

A habitat model for native mosquitoes must take into account both, climatic conditions and regional specifics. Therefore, the final model consists of two parts, the weather model (DWD model) and the regional model. The DWD model models the dependence of the mosquitoes on the climate, and the regional model their habitat requirements. The workflow of this modelling approach is shown in Fig. 1

2.1. Data

For the development of the combined modelling approach we used data from the citizen science project "Mückenatlas" ('sent') and in addition data originating from systematic active monitoring ('sample') in Germany, or the combination of both ('all').

After determination of the mosquito species, all data linked to a mosquito collection, including date of capture and geographic coordinates of the collection site, are entered into the CULBASE. Information on confirmed infections with WNV in 2019 was provided by the FLI, where the national data are stored by the Animal Diseases Reporting System ("Tierseuchennachrichtensystem").

Out of the potential native vector species of WNV we selected two, each representing a specific type of mosquito with respect to their habitat and/or behaviour (Gutsevich et al., 1970; Vinogradova, 2000; Becker et al., 2010; Hamer et al., 2014; Verdonschot and Besse-Lototskaya, 2014): *Cx. pipiens* (house mosquito) and *Ae. vexans* (floodwater mosquito) to verify the modelling methodology (Table 1).

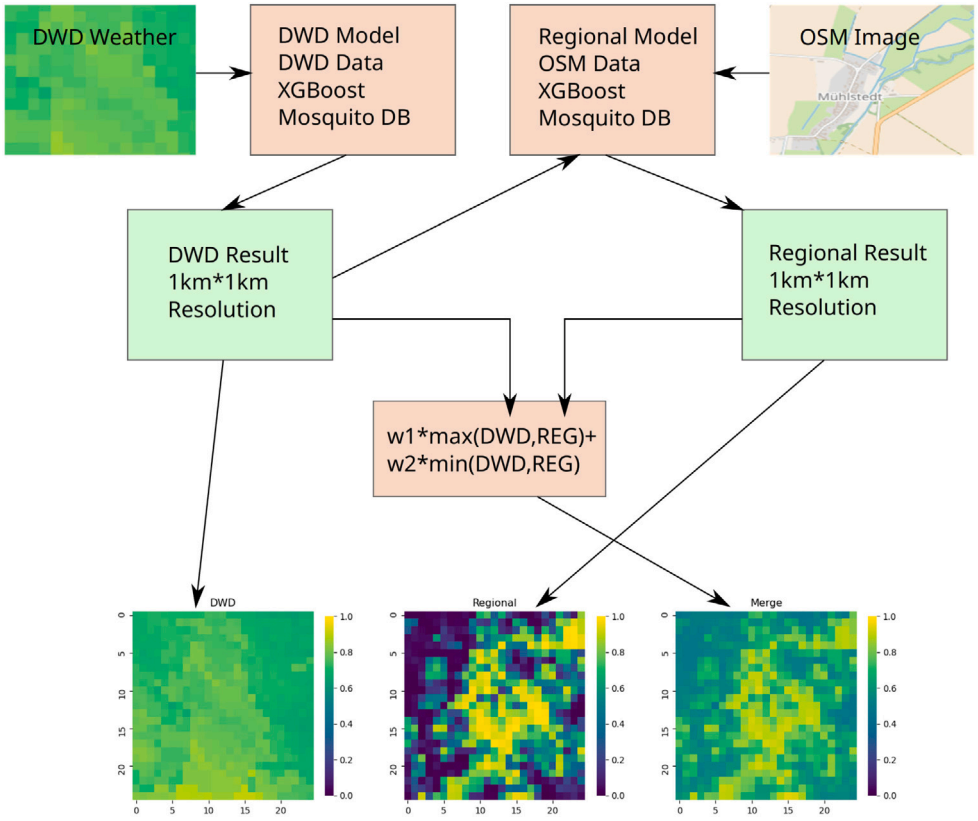

Fig. 1. Overview of the used modelling approach. The equation in the middle shows the calculation of a combination of the DWD model and the regional model.

Table 1
Mosquito species selected for modelling and number of occurrence data for the time period 2012–2019.

Type	Species	Sent	Sample	All
Floodwater mosquito	*Ae. vexans*	2210	1159	3369
House mosquito	*Cx. pipiens*	4084	1478	5562

2.2. Weather model

The DWD model is based on data freely available from the "Deutscher Wetterdienst" (German Weather Service: DWD). The DWD offers a variety of weather data consisting of temperature, precipitation, radiation data, etc. (e.g. Tmin, Tmax, Tmean) for the period from 1881 until today. These data are available as raster data with a resolution of 1 km × 1 km for Germany. The available temporal resolution is daily, monthly, seasonal and yearly. For modelling the mosquito habitat, monthly, seasonal and yearly data are especially interesting (Kerkow et al., 2020).

For the development of the DWD model, we used the mosquito trapping data (occurrence data 'sent' and 'all') of the respective mosquito species from CULBASE. The absence points (NP) were added by random sampling (Thuiller et al., 2020). According to the sampling strategy, the occurrence data (positive sites — PP) and the absence points (negative sites — NP) were assigned to the grids of the DWD. The grids were loaded and reprocessed using the open source software SAMT2 (see Section 2.5) (Wieland et al., 2015).

From the large amount of DWD data, specific climatic variables were selected according to the recommendations of BIOCLIM[1] based on ecological and environmental factors affecting the development, behaviour and activity of mosquitoes.

[1] http://www.worldclim.org/bioclim.

Table 2
Climatic variables selected for the DWD model.

Feature name	Short name	Description
Precq(3)	Pq3	Average precipitation March–May
TDiff(7)	TD7	Tmax–Tmin of July
annual_tmin	TMinA	Minimum annual temperature
Tempq(6)	Tq6	Average temperature June–August
Precq(6)	Pq6	Average precipitation June–August
TDiff(4)	TD4	Tmax–Tmin of April
GRq(4)	GQ4	Global radiation April–June
GRq(7)	GQ7	Global radiation July–September
annual_prec()	PA	Annual precipitation
annual_tmax()	TmaxA	Maximum annual temperature

For the transformation of the DWD data into bio-climatic data a Python module (bioclim.py) was developed, which is available as open source software. It can be extended and customised by the users. Eq. (1) is intended to illustrate the implementation of the transformations from DWD to bio-climatic variables:

$$V[1] = bio.Tempq(3) \quad (1)$$

The variable V[1] is assigned the average temperature of the months March to May. The function Tempq(t) combines three months, beginning with month 3 (March). Table (1) shows the climatic variables selected for the DWD model (see Table 2).

Despite the use of the bio-climate transformation, it is not certain that the most important features were chosen for the model. For the classification task, the modern algorithm XGBoost (Ma et al., 2020) was chosen as the modelling tool. XGBoost together with the statistically based analysis software SHAP (see Section 2.5) (Lundberg et al., 2017) allows the determination of the feature importance. An alternative method for determining feature importance based on a support vector machine was presented in Wieland et al. (2017).

The derived weather models always refer to a specific year. To balance the influence of the chosen year, an ensemble of three years (Y_{dry}, Y_{wet}, Y_{normal}) was used. This ensemble ($Y_{ensemble}$) was calculated according to Eq. (2)):

$$Y_{ensemble} = \frac{Y_{dry} + Y_{wet} + Y_{normal}}{3} \quad (2)$$

The DWD model consists of the following steps:

- **modelling step:** Training a classifier (XGBoost) using the occurrence data (positive sites — PP) and the absence data (negative sites — NP).
- **analysis step:** Assessment of the classifier accuracy, recall and precision of validation data; determination of feature importance (SHAP).
- **application step:** Generation of a map from the $Y_{ensemble}$ of the 'climatic habitat' of the selected mosquito species for a target year.

2.3. Regional model

For the regional model, maps are needed as a basis. The freely available OpenStreetMap (OSM) maps (OpenStreetMap contributors, 2017) were chosen. These maps are characterised by a high level of detailing and actuality. Compared to the often used Corine Land Cover (Bielecka and Jenerowicz, 2019) the OSM maps have a higher resolution (10 m ∗ 10 m). For mosquito species, which usually do not fly very far, the high resolution is essential. An OSM map of 4 km × 4 km was loaded for each positive (PP) and negative (NP) mosquito site in the centre.

In a next step, the OSM maps are analysed by means of a histogram with regard to land use (urban, forest, grassland, lakes, rivers, etc.). The distribution of land use data are the features used for training the XGBoost. This quite simple procedure does not consider neighbourhood relationships between land use types. For example, it would be important to know if a lake is surrounded by forest or not. But as shown below, this simplification is not problematic. The regional model consists of the following steps:

- **data provision:** Downloading maps from the OSM server.
- **modelling step:** Training a classifier (XGBoost) using the occurrence data (positive sites — PP) and the absence data (negative sites — NP).
- **analysis step:** Assessment of classifier accuracy, recall and precision of validation data; determination of feature importance (SHAP).

2.4. Combination DWD model and regional model

Both models are combined to a final outcome (y_{comb}) according to Eq. (3):

$$y_{comb} = w_1 \times max\{DWD, REG\} + w_2 \times min\{DWD, REG\} \quad (3)$$

with $DWD = p(mosquito|weather)$ and $REG = p(mosquito|region)$ (p = probability); the weighting factors $w_1 + w_2 = 1$ score the model outputs. If $w_1 > w_2$, then the dominant model is ranked higher than the subdominate model and vice versa. The idea to use this formula comes from fuzzy theory (Zadeh, 1965; Lin and Yang, 2020). The minimum corresponds to the rather pessimistic view that a model can only be as good as the worst part. The maximum implements the optimistic view that if at least one model is good, the mosquito is established. The mixture of both models according to Eq. (3) should come close to the truth.

In order to present the generated models in a more analysable and comprehensive form to the user, the combined model was subjected to a cubic spline interpolation and transferred to a map with a size of 25 km×25 km. The size of the map is configurable, but 25 km×25 km has proven to be useful, also with respect to the average flight distances of local mosquitoes (Gutsevich et al., 1970; Vinogradova, 2000; Becker et al., 2010; Hamer et al., 2014; Verdonschot and Besse-Lototskaya, 2014).

Thus, the final merging ('Merge') of the two models ('DWD' and 'Regional') consists of the following steps:

- **model combination:** According to Eq. (3) and optimisation of the parameters w_1 and w_2,
- **visualisation:** Projection of the model result into a high resolution map of 25 km × 25 km; adding the case of WNV infection.

2.5. Machine learning with XGBoost

XGBoost is one of the latest developments of boost algorithms. The idea behind "boost" is the combination of weak models into a powerful model (Schapire, 1990). Through a rich set of parameters, XGBoost allows to optimise the machine learning. However, the default values are often already sufficiently good, so that the time-consuming optimisation of the parameters can be limited to a few important ones. Especially important in our experience are the learning rate (eta), the maximum depth of weak learners (max_depth) and the regularisation parameters (lambda and alpha). XGBoost reads tables with X-values and a vector with y-values. These have to be prepared before training, which is often a major part of the work to be done. In the present case, the preparation of the histograms from the maps was the most time-consuming step. XGBoost itself can be used for classification and regression. The algorithm is extremely fast, so that the training of the image data was done in a few seconds on the PC. If this is not enough, XGBoost can also be accelerated with the help of a GPU. The use of XGBoost is described in detail in Brownlee (2015).

SHAP opens up a method to analyse and understand the training of XGBoost. SHAP is based on the cooperative game theory (Lundberg Sc and Lee, 2017). It uses the trained XGBoost model (explainer = shap.TreeExplainer(model)) and the training data (shap_values = explainer.shap_values(X)). SHAP's statistical approach also allows it to be applied to deep neural networks, which opens up the possibility of examining alternative model structures for the same problem. SHAP offers a variety of visualisations, of which only the shap.summary_plot(shap_values, X) was used here. SHAP has proven to be extremely important in the evaluation of (blackbox) models. For more information, Lundberg et al. (2017) is recommended.

The Spatial Analysis and Modelling Tool (SAMT) (Wieland et al., 2015) developed at ZALF is used here to read and write rasterised geographic DWD data. Since it is available as compiled Python code, it can also read and write large data sets in a few seconds. The possibility of binary storage of raster data increases the speed tenfold again.

Opencv2 (Villian, 2019) is a sophisticated image processing software originally developed by Intel. Today it is open source and available for C, C++, Python and Java. In addition to the image processing used here, machine learning algorithms are also implemented. Since its basis is an implementation in C, the modules for Python are very fast, which is also necessary for processing the many images.

The software applied (Python with the modules: numpy, pandas etc.) is open source and can be used freely.

3. Results

3.1. DWD model

The ensemble was formed from the models of the years 2016, 2017, and 2018. The results of the validation run for the mosquito species *Ae. vexans* and *Cx. pipiens* are shown in Tables 3 and 4. The data was divided into 70% for training and 30% for validation. This was done for both the DWD model and the OSM model.

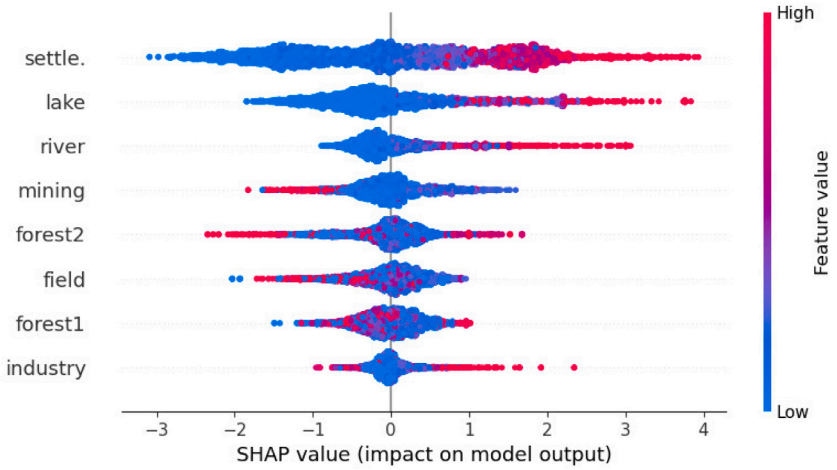

Fig. 2. Feature importance of *Ae. vexans* with: settle. = settlement, forest1 = mixed forest, forest2 = coniferous forest.

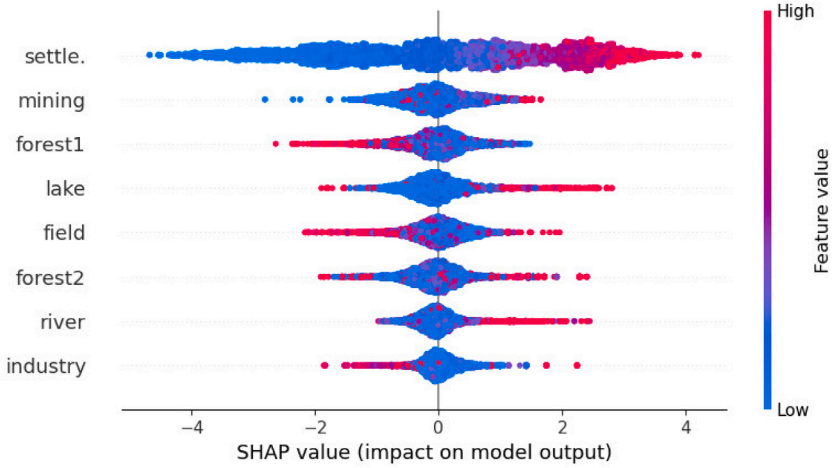

Fig. 3. Feature importance of *Cx. pipiens* with: settle. = settlement, forest1 = mixed forest, forest2 = coniferous forest.

Table 3
DWD model for *Ae. vexans*, N: size of data (number of occurrence points).

Data	Year	Accuracy	Recall	Precision	N
all	2016	0.831	0.832	0.832	613
all	2017	0.783	0.811	0.773	538
all	2018	0.74	0.692	0.782	82
sent	2016	0.750	0.728	0.773	547
sent	2017	0.805	0.864	0.777	410
sent	2018	0.704	0.700	0.667	72

Table 4
DWD model for *Cx. pipiens*, N: size of data (number of occurrence points).

Data	Year	Accuracy	Recall	Precision	N
all	2016	0.743	0.750	0.731	1688
all	2017	0.737	0.696	0.783	1440
all	2018	0.683	0.636	0.685	485
sent	2016	0.716	0.719	0.703	1332
sent	2017	0.705	0.705	0.702	987
sent	2018	0.697	0.714	0.685	472

It should be noted that in 2018 the models of *Cx. pipiens* (Table 4) but especially of *Ae. vexans* (Table 3) have only few data for the training. Therefore such a model should not be used independently. Nevertheless, this model was included in the ensemble because the year 2018 was extremely hot and dry. The 2018 mean annual temperature was the highest with 10.45 °C, and the annual precipitation of 586.3 mm was the lowest from 2011 to 2020 in Germany (DWD). The feature importance graphs for the DWD models are provided in the supplementary material.

The two DWD models generated with the 'all' and the 'sent' data sets showed for both tested species no significant differences in their validation results.

Table 5
OSM model for *Cx. pipiens* (pip) and *Ae. vexans* (vex), N: size of data (number of occurrence points).

Species	Data	Accuracy	Recall	Precision	N
pip	all	0.884	0.888	0.882	5562
pip	sent	0.884	0.886	0.878	4084
vex	all	0.878	0.902	0.860	3369
vex	sent	0.848	0.847	0.847	2210

3.2. OSM model

Table 5 summarises the validation data of the training of the regional models for the selected two species. In contrast to the annual data of the DWD model, the regional data are obtained from all years (2012–2019), which explains the higher number of training data.

The two regional models also show no significant differences between the results generated with the 'all' and 'sent' data. Consequently,

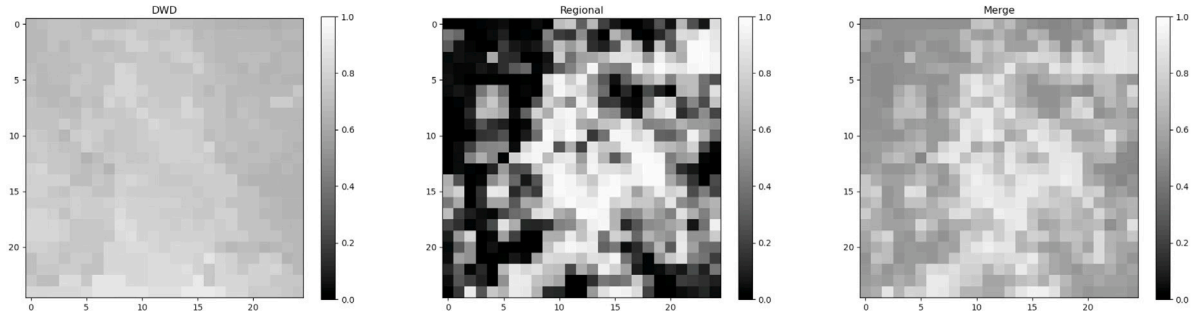

Fig. 4. Combination ('Merge') of the DWD model (target year 2019) with the regional model for *Ae. vexans* in the area of the city Bitterfeld-Wolfen in Saxony-Anhalt (25 km×25 km), where WNV infections occurred in 2019.

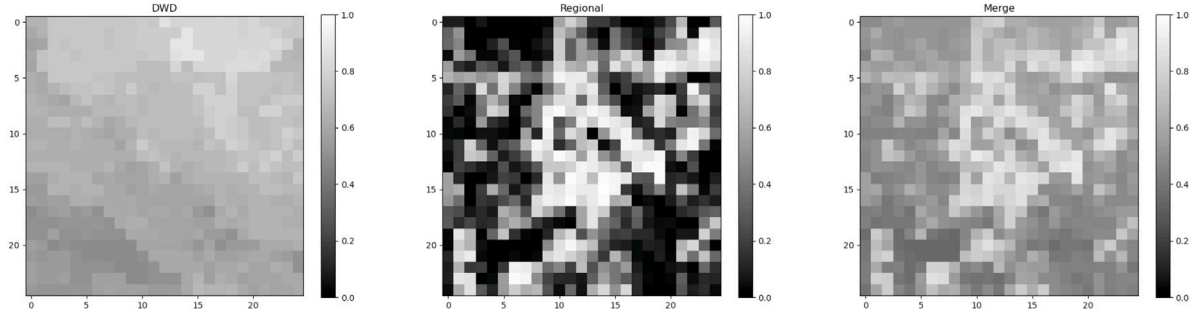

Fig. 5. Combination ('Merge') of the DWD model (target year 2019) with the regional model for *Cx. pipiens* in the area of the city Bitterfeld-Wolfen in Saxony-Anhalt (25 km×25 km), where WNV infections occurred in 2019.

the data from the "Mückenatlas" ('sent') alone would be sufficient for modelling. This conclusion is also valid for the DWD models.

In the following step, the SHAP library is used to calculate the feature importance for the two mosquito species. The data 'all' which combine 'sample' and 'sent' are used as a basis. Fig. 2 shows the feature importance of the floodwater mosquito *Ae. vexans* and Fig. 3 of the house mosquito *Cx. pipiens*.

Fig. 2 shows that, beside settlements, also lakes and rivers are most important for *Ae. vexans*, while agricultural structures are rather avoided. The *Ae. vexans* is a floodwater mosquito therefore lakes and rivers are important. In Fig. 2 they are at position two and three. Low availability of lakes and rivers (blue) leads to negative SHAP values and thus speaks against a suitable habitat. A high availability of lakes and rivers (red), on the other hand, leads to positive SHAP values and thus to a suitability as a habitat.

Fig. 3 shows also the preference of *Cx. pipiens* for settlements. Structures connected with mining and avoiding mixed forest are more important parameters than the presence of water, whereas lakes seem to be preferred over rivers.

3.3. Simulation results

To verify the combination of the DWD model and the regional model, both were first tested separately and then together according to Eq. (3). Figs. 4 and 5 show that both models contribute to habitat modelling. $w_1 * max\{DWD, REG\} + w_2 * min\{DWD, REG\}$ means that w_1 supports the dominant part (here mostly REG) and w_2 supports the subdominant part (here mostly DWD). We tried a range of $w_1 \in [0.2, 0.8]$ with $w_2 = 1 - w_1$ and visually evaluated the combination (w_1, w_2). It turned out that $(w_1 = 0.7, w_2 = 0.3)$ was reasonable. It was also reasonable for alternative locations (nearby Berlin and nearby Dresden). However, the number of occurrence points of mosquitoes in the regions (infected or not) was too low (<15) for numerical optimisation.

After completion of the development of the two models (DWD ensemble models of specific years and regional model) for the mosquito species *Ae. vexans* and *Cx. pipiens* they were applied to specific real regions in a simulation. In order to demonstrate the combination of the two models we have chosen a 25 km × 25 km area around the city Bitterfeld-Wolfen in Saxony-Anhalt, where in 2019 a number of cases of WNV infections were registered in birds and horses. In the centre of the chosen region there is a confirmed case of WNV infection (Figs. 4, 5).

Fig. 6 shows the simulation results for *Ae. vexans* for the selected Bitterfeld-Wolfen region displayed on a map and a heat map. On the left, the OSM map is shown. The heat map shows overlay of the OSM map and the habitat quality (DWD,OSM) which was interpolated using a bicubic interpolation from 1 km ∗ 1 km to 30 m ∗ 30 m. Dark areas mark high habitat suitability. Fig. 7 shows the simulation results for *Cx. pipiens*. All cases of WNV infection from the selected region are shown as red dots. We selected the area in such a way, that one of the cases is in the centre of the map.

Both results are similar at first sight. However, it is notable that the occurrence of *Ae. vexans* is more strongly associated with water bodies in comparison to *Cx. pipiens*. *Cx. pipiens* is more oriented towards urban regions. The models of both species agree with the occurrence of cases of WNV infections. An assignment of the cases of WNV infections to a specific mosquito species was not possible at this stage, since the number of cases is not sufficient for such conclusions and also other species considered as potential vectors of WNV should be tested.

4. Discussion

The central question of the present work was to find an approach for habitat modelling of native mosquito species, which are ubiquitous and perfectly adapted to their habitats. Some of them are considered as potential or even proven vectors for arboviruses. While models of invasive species, which are mostly based on climatic factors, revealed high accuracies, the case of native species is more complicated and needs a more complex modelling strategy. Prerequisite of a good model is a solid data base of surveyed mosquitoes. As active monitoring is often limited by the availability of human and financial resources, a

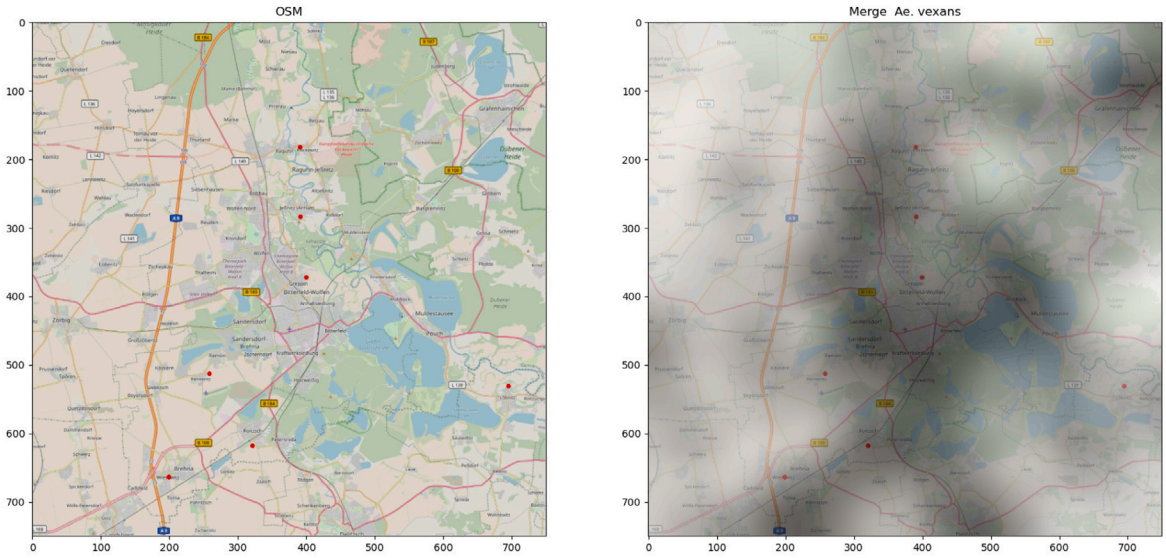

Fig. 6. Simulation result and heat map for *Ae. vexans* in the area of the city Bitterfeld-Wolfen (25 km × 25 km, resolution = 30 m). Dark areas mark high habitat suitability. Cases of WNV infections (2019) are marked by red dots.

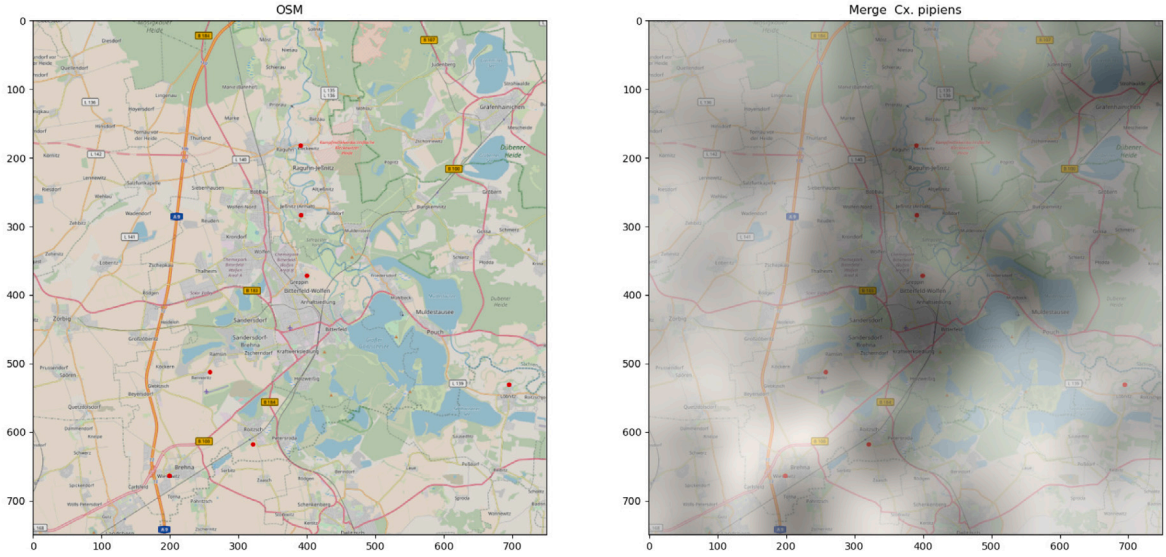

Fig. 7. Simulation (target year 2019) result and heat map for *Cx. pipiens* in the area of the city Bitterfeld-Wolfen (25 km × 25 km, resolution = 30 m). Dark areas mark high habitat suitability. Cases of WNV infections (2019) are marked by red dots.

countrywide sampling is not possible. Passive monitoring could be a solution for this problem since submissions are covering the whole country, however, also this kind of data collection is prone to certain biases (e.g. submissions mostly from densely inhabited regions). Therefore, we tested the applicability of citizen science data (CSD), in our case of the "Mückenatlas", and compared the quality of the models with a combined data set of the CSD and of data from active monitoring. We could show that in case of our modelling approach the CSD of the "Mückenatlas" are sufficient to develop models with high accuracies, which are comparable to those of the combined data set. In contrast, the application of data from active sampling alone was not sufficient because of the limited number of occurrence points (data not shown).

From the small number (10–12) of potential WNV vectors we selected two species, which (i) were represented by a sufficient number of collections (occurrence points) in the data base and (ii) which differ in their biological traits (e.g. habitat preferences of the distinct developmental stages, temperature optimum, diapause, breeding sites, flight distances, host feeding preferences etc.). In a first step the power and accuracy of the models based on either climatic or regional factors were tested separately.

The accuracy of the DWD model was partly poor (≈0.7). This was expected for native mosquitoes, because they are well adapted to the climate. The combination of three DWD models corresponding to different weather conditions (e.g. hot and dry summer in 2018) proved to be useful. Furthermore, it was tried to train the models with simple features like [T3...T10] or [P3...P10]. The resulting accuracy was only slightly worse than the accuracy based on the bio-climatic data. Nevertheless, such simple data have the advantage of being easy to interpret. This means that the biologist can interpret the models based on the temperature or precipitation series. This interpretability is a great advantage when using simple data series. Together with modern analytical methods such as SHAP, even blackbox models can be analysed. In the presented example we used selected bio-climatic variables, chosen according to the biology of mosquitoes. The climate models possibly could be further improved by adapting the variables to the biological and behavioural specificities of the selected mosquito species.

In contrast to the DWD models, the regional models have a rather high accuracy. This shows that at least for native mosquitoes regional characteristics play a greater role in the habitat than climatic conditions. However, the regional characteristics alone are not sufficient to calculate the habitat quality, as shown in Figs. 4 and 5. The climatic conditions lead to a large-scale habitat quality, while the regional conditions rather add the small-scale characteristics to the calculation. The weights (w_1) and (w_2) from Eq. (3) were estimated and then used in all simulations: ($w_1 = 0.7$) and ($w_2 = 0.3$). They can be adjusted if necessary.

The visualisation part of the simulation involves selecting the coordinates of a case of WNV infection and selecting a 25 km × 25 km map section with that case in the centre. Other registered cases, which occurred in the same area, are also displayed. The habitat suitability calculated with the model combination is displayed as a heat map. This heat map is blended with the map and forms the result of the entire habitat modelling. If there is an outbreak of WNV infections in a region, an assignment of the analysed various mosquito species considered as potential or proven vectors can be made via the result maps. This can help in the design of disease control measures.

The methodology presented is based exclusively on machine learning and therefore does not require expert knowledge. It can be efficiently applied for different mosquito species or even other arthropods transmitting diseases including also parasitic and bacterial diseases, provided that sufficient training data are available.

The innovation of the present work is the possibility of generating detailed map sections of regions with cases of WNV infections, a kind of zooming in on the map. The size of the map section (25 km ∗ 25 km) is adjustable (e.g. to the flight distance of the selected mosquito), but limited by the computing capacity. Based on the habitat suitability and the occurrence points of WNV infections the risk of spread of the virus into adjacent areas can be assessed, taking into account also the vector capacity of the respective species.

In the future, the presented method will be used in the development of a spatially distributed simulation of mosquito dispersal and in particular the spread of WNV. The simulation will then be performed on a central 25 km ∗ 25 km grid surrounded by 8 grids of equal size (D8 environment). Movements in space can be modelled by means of an agent model (Grimm et al., 2005; Lenfers et al., 2018).

The approach could be also supplemented by other factors as wind (dispersal of mosquitoes), or water levels and flooding dynamics (e.g. changing water levels of rivers due to intense rainfalls or other reasons in remote upstream regions) which were shown to be useful factors (Lončarić and Hackenberger, 2013; Verdonschot and Besse-Lototskaya, 2014; Kerkow et al., 2019). Also population density of mosquitoes could be added as a factor. In the present work we used only occurrence points, not the amount of specimens per occurrence point. The dynamics of mosquito populations can be modelled as described in Laperriere et al. (2011). The prediction of increased adult population densities is essential for the implementation of targeted control measures.

5. Conclusion

The work has shown that even difficult modelling tasks can be solved by combining different modelling approaches, in this case a climate model and a regional model. Even the exclusive use of citizen science data, if collected over many years and with a sufficient submission quantity as well as territory coverage, can be sufficient to solve scientific problems. It is essential, however, that even the most sophisticated machine learning methods can only provide the hoped-for boost in modelling through a sound selection of data.

CRediT authorship contribution statement

Ralf Wieland: Model development, Programming and writing. **Katrin Kuhls:** Simulation, Visualisation, Writing. **Hartmut H.K. Lentz:** Providing data of WNV infections, Review and editing. **Franz Conraths:** Responsible for hosting the database CULBASE, Responsible for the Animal Disease Reporting System. **Helge Kampen:** Collecting and providing mosquito data from active monitoring. **Doreen Werner:** Supervision, Collecting and providing mosquito data from active monitoring, Responsible for the "Mückenatlas".

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Special thanks to the OSM community, which made it possible to load many tens of thousands of maps for modelling from their servers. The authors thank also all citizen scientists and active collectors of mosquitoes, who made the "Mückenatlas" possible. Special thanks to Mrs. Jutta Falland from the ZALF for her diligent data management of the "Mückenatlas".

The present work was funded by the Federal Ministry of Food and Agriculture of Germany (BMEL) through the Federal Office for Agriculture and Food (BLE), grant number 2819113519.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ecolmodel.2021.109594. The supplemented material shows the feature importanze of the DWD model and visualises the simulation results.

References

Bartumeus, F., Oltra, A., Palmer, J., 2018. Citizen science: A gateway for innovation in disease-carrying mosquito management? Trends Parasitol. 34 (9), 727–729. http://dx.doi.org/10.1016/j.pt.2018.04.010.

Becker, N., Petrić, D., Zgomba, M., Boase, C., Madon, M., Dahl, C., Kaiser, A., 2010. Mosquitoes and their Control. Springer, Heidelberg, Dordrecht, New York, p. 577.

Bielecka, E., Jenerowicz, A., 2019. Intellectual structure of CORINE land cover research applications in web of science: a europe-wide review. Remote Sens. 11 (17), 1–22. http://dx.doi.org/10.3390/rs11172017.

Brownlee, J., 2015. XGBoost with Python. Machine Learning Mastery, Pty. Ltd., Vermont Victoria 3133, Australia, p. 115.

Brugman, V.A., Hernández-Triana, L.M., Medlock, J.M., Fooks, A.R., Carpenter, S., Johnson, N., 2018. The role of *culex pipiens* L. (Diptera: Culicidae) in virus transmission in Europe. Int. J. Environ. Res. Public Health 15 (2), 389. http://dx.doi.org/10.3390/ijerph15020389.

Ciota, A.T., Keyel, A.C., 2019. The role of temperature in transmission of zoonotic arboviruses. Viruses 11 (11), 1013. http://dx.doi.org/10.3390/v11111013.

Cunze, S., Kochmann, J., Klimpel, S., 2020. Global occurrence data improve potential distribution models for *Aedes japonicus japonicus* in non-native regions. Pest Manage. Sci. 76 (5), 1814–1822. http://dx.doi.org/10.1002/ps.5710.

Engler, O., Savini, G., Papa, A., Figuerola, J., Groschup, M.H., Kampen, H., Medlock, J., Vaux, A., Wilson, A.J., Werner, D., Jöst, H., Goffredo, M., Capelli, G., Federici, V., Tonolla, M., Patocchi, N., Flacio, E., Portmann, J., Rossi-Pedruzzi, A., Mourelatos, S., Johnson, N., 2013. European surveillance for west nile virus in mosquito populations. Int. J. Environ. Res. Public Health 10 (10), 4869–4895. http://dx.doi.org/10.3390/ijerph10104869.

Früh, L., Kampen, H., Kerkow, A., Schaub, G.A., Walther, D., Wieland, R., 2018. Modelling the potential distribution of an invasive mosquito species: comparative evaluation of four machine learning methods and their combinations. Ecol. Model 388, 136–144.

Grimm, V., Revilla, R.E., Berger, U., Jeltsch, F., Mooij, W.M., Railsback, St.F., Thulke, H.H., Weiner, J., Wiegand, Th., DeAngelis, D.L., 2005. Pattern-oriented modeling of agent based complex systems: lessons from ecology. Science 310 (5750), 987–991. http://dx.doi.org/10.1126/science.1116681.

Gutsevich, A.V., Monchadsky, A.S., Shtakelberg, A.A., 1970. Fauna of the USsr. Diptera. In: Mosquitoes of the Family Culicidae, Vol. 3. p. 384, (in Russian).

Hahn, M.B., Monaghan, A.J., Hayden, M.H., Eisen, R.J., Delorey, M.J., Lindsey, N.P., Nasci, R.S., Fischer, M., 2015. Meteorological conditions associated with increased incidence of west nile virus disease in the United States, 2004-2012. Am. J. Trop. Med. Hyg. 92 (5), 1013–1022. http://dx.doi.org/10.4269/ajtmh.14-0737.

Hamer, G.L., Anderson, T.K., Donovan, D.J., Brawn, J.D., Krebs, B.L., Gardner, A.M., Ruiz, M.O., Brown, W.M., Kitron, U.D., Newman, C.M., Goldberg, T.L., Walker, E.D., 2014. Dispersal of adult *Culex* mosquitoes in an urban west nile virus hotspot: a mark-capture study incorporating stable isotope enrichment of natural larval habitats. PLoS NTD 8 (3), e2768. http://dx.doi.org/10.1371/journal.pntd.0002768.

Holicki, C.M., Ziegler, U., Răileanu, C., Kampen, H., Werner, D., Schulz, J., Silaghi, C., Groschup, M.H., Vasić, A., 2020. West nile virus lineage 2 vector competence of indigenous *Culex* and *Aedes* mosquitoes from Germany at temperate climate conditions. Viruses 12 (5), 561. http://dx.doi.org/10.3390/v12050561.

Jácome, G., Vilela, P., Yoo, ChKy, 2019. Present and future incidence of dengue fever in Ecuador nationwide and coast region scale using species distribution modeling for climate variability's effect. Ecol. Model 400, 60–72. http://dx.doi.org/10.1016/j.ecolmodel.2019.03.014.

Jansen, S., Heitmann, A., Lühken, R., Leggewie, M., Helms, M., Badusche, M., Rossini, G., Schmidt-Chanasit, J., Tannich, E., 2019. Culex torrentium: A potent vector for the transmission of west nile virus in central europe. Viruses 11 (6), 492. http://dx.doi.org/10.3390/v11060492.

Kampen, H., Holicki, C.M., Ziegler, U., Groschup, M.H., Tews, B.A., Werner, D., 2020. West nile virus mosquito vectors (Diptera: Culicidae) in Germany. Viruses 12 (5), 493. http://dx.doi.org/10.3390/v12050493.

Kampen, H., Medlock, J.M., Vaux, A.G.C., Koenraadt, C.J.M., van Vliet, A.J.H., Bartumeus, F., Oltra, A., Sousa, C.A., Chouin, S., Werner, D., 2015. Approaches to passive mosquito surveillance in the EU. Parasit Vectors 8 (1), 9. http://dx.doi.org/10.1186/s13071-014-0604-5.

Kerkow, A., Wieland, R., Früh, L., Hölker, F., Jeschke, J.M., Werner, D., Kampen, H., 2020. Can data from native mosquitoes support determining invasive species habitats? Modelling the climatic niche of *Aedes japonicus japonicus* (Diptera, Culicidae) in Germany. Parasitol. Res. 119 (1), 31–42. http://dx.doi.org/10.1007/s00436-019-06513-5.

Kerkow, A., Wieland, R., Koban, M.B., Hölker, F., Jeschke, J.M., Werner, D., Kampen, H., 2019. What makes the Asian bush mosquito *Aedes japonicus japonicus* feel comfortable in Germany? A fuzzy modelling approach. Parasit Vectors 12 (1), 106. http://dx.doi.org/10.1186/s13071-019-3368-0.

Laperriere, V., Brugger, K., Rubel, F., 2011. Simulation of the seasonal cycles of bird, equine and human West Nile virus cases. Prev. Vet. Med. 98 (2, 3), 99–110.

Lee, K.J., Chung, N., Hwang, S., 2016. Application of an artificial neural network (ANN) model for predicting mosquito abundances in urban areas. Ecol. Inform. 36, 72–180. http://dx.doi.org/10.1016/j.ecoinf.2015.08.011.

Lenfers, U., Weyl, U.A., Clemen, Th., 2018. Firewood collection in South Africa: adaptive behavior in social-ecological models. Land 7 (3), 97. http://dx.doi.org/10.3390/l{and}7030097.

Lin, H., Yang, X., 2020. Dichotomy algorithm for solving weighted min–max programming problem with addition-min fuzzy relation inequalities constraint. Comput. Indust. Eng. 146, 1–7.

Lončarić, Ž., Hackenberger, B.K., 2013. Stage and age structured *Aedes vexans* and *Culex pipiens* (Diptera: Culicidae) climate-dependent matrix population model. Theor. Popul. Biol. 83, 82–94. http://dx.doi.org/10.1016/j.tpb.2012.08.002.

Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, Su-In, 2017. From local explanations to global understanding with explainable AI for trees. Nat. Mach. Intell. 2 (1), 56–67.

Lundberg Sc, M., Lee, Su-In, 2017. A Unified Approach to Interpreting Model Predictions. In: 31st Conference on Neural Information Processing Systems (NIPS 2017) Long Beach, CA, USA. 1-10.

Ma, J., Cheng, J., Xu, Z., Chen, K., Lin, C., Jiang, F., 2020. Identification of the most influential areas for air pollution control using XGBoost and Grid Importance Rank. J. Clean. Prod. 274, 1–12.

Martinet, J.P., Ferté, H., Failloux, A.B., Schaffner, F., Depaquit, J., 2019. Mosquitoes of North-Western Europe as potential vectors of arboviruses: a review. Viruses 11 (11), 1059. http://dx.doi.org/10.3390/v11111059.

Metelmann, S., Caminade, C., Jones, A.E., Medlock, J.M., Baylis, M., Morse, A.P., 2019. The UK's suitability for *Aedes albopictus* in current and future climates. J. R. Soc. Interface 16 (152), 20180761. http://dx.doi.org/10.1098/rsif.2018.0761.

Myer, M.H., Johnston, J.M., 2019. Spatiotemporal Bayesian modeling of West Nile virus: identifying risk of infection in mosquitoes with local-scale predictors. Sci. Total Environ. 650 (Pt 2), 2818–2829. http://dx.doi.org/10.1016/j.scitotenv.2018.09.397.

OpenStreetMap contributors, 2017. Planet dump retrieved from. https://planet.osm.org, https://www.openstreetmap.org.

Paz, S., 2015. Climate change impacts on west nile virus transmission in a global context. Philos. Trans. R. Soc. Lond. B Biol. Sci. 370 (1665), 20130561. http://dx.doi.org/10.1098/rstb.2013.0561.

Reuss, F., Wieser, A., Niamir, A., Bálint, M., Kuch, U., Pfenninger, M., Müller, R., 2018. Thermal experiments with the Asian bush mosquito (*Aedes japonicus japonicus*) (Diptera: Culicidae) and implications for its distribution in Germany. Parasit Vectors 11 (1), 81. http://dx.doi.org/10.1186/s13071-018-2659-1.

Roiz, D., Ruiz, S., Soriguer, R., Figuerola, J., 2014. Climatic effects on mosquito abundance in Mediterranean wetlands. Parasit Vectors 7, 333. http://dx.doi.org/10.1186/1756-3305-7-333.

Schapire, R.E., 1990. The strength of weak learnability. Mach. Learn. 5, 197–227.

Thuiller, W., Georges, D., Engler, R., Breiner, F., 2020. Biomod2: ensemble platform for species distribution modeling, R package version: 3.4.6.

Valdez, L.D., Sibona, G.J., Condat, C.A., 2018. Impact of rainfall on *Aedes aegypti* populations. Ecol. Modell. 385, 96–105. http://dx.doi.org/10.1016/j.ecolmodel.2018.07.003.

Verdonschot, P.F.M., Besse-Lototskaya, A.A., 2014. Flight distance of mosquitoes (Culicidae): a metadata analysis to support management of barrier zones around rewetted and newly constructed wetlands. Limnologica 45, 69–79.

Vilibic-Cavlek, T., Savic, V., Petrovic, T., Toplak, I., Barbic, L., Petric, D., Tabain, I., Hrnjakovic-Cvjetkovic, I., Bogdanic, M., Klobucar, A., Mrzljak, A., Stevanovic, V., Dinjar-Kujundzic, P., Radmanic, L., Monaco, F., Listes, E., Savini, G., 2019. Emerging trends in the epidemiology of West Nile and usutu virus infections in Southern Europe. Front. Vet. Sci. 6 (437), http://dx.doi.org/10.3389/fvets.2019.00437.

Villian, A.F., 2019. Mastering OpenCV4 with Python. Packt Publishing Ltd., Birmingham, p. 513.

Vinogradova, E.B., 2000. Culex Pipiens Pipiens Mosquitoes: Taxonomy, Distribution, Ecology, Physiology, Genetics, Applied Importance and Control. Pensoft, Sofia-Moscow, p. 250.

Wagner, S., Mathis, A., Schönenberger, A.C., Becker, S., Schmidt-Chanasit, J., Silaghi, C., Veronesi, E., 2018. Vector competence of field populations of the mosquito species *Aedes japonicus japonicus* and *Culex pipiens* from Switzerland for two West Nile virus strains. Med. Vet. Entomol. 32 (1), 121–124. http://dx.doi.org/10.1111/mve.12273.

Walther, D., Kampen, H., 2017. The citizen science project 'Mueckenatlas' helps monitor the distribution and spread of invasive mosquito species in Germany. J. Med. Entomol. 54 (6), 1790–1794. http://dx.doi.org/10.1093/jme/tjx166.

Wieland, R., Groth, K., Linde, F., Mirschel, W., 2015. Spatial analysis and modeling tool version 2 (SAMT2) a spatial modeling tool kit written in python. Ecol. Inform. 30, 1–5. http://dx.doi.org/10.1016/j.ecoinf.2015.08.002.

Wieland, R., Kerkow, A., Früh, A., Kampen, H., Walther, D., 2017. Automated feature selection for a machine learning approach toward modeling a mosquito distribution. Ecol. Modell. 352, 108–112.

Wöhnke, E., Vasić, A., Răileanu, C., Holicki, C.M., Tews, B.A., Silaghi, C., 2020. Comparison of vector competence of *Aedes vexans* Green River and *Culex pipiens* biotype *pipiens* for West Nile virus lineages 1 and 2. Zoonoses Public Health 67 (4), 416–424. http://dx.doi.org/10.1111/zph.12700.

Zadeh, L.A., 1965. Fuzzy sets. Inf. Control 8 (3), 338–353.

Ziegler, U., Lühken, R., Keller, M., Cadar, D., van der Grinten, E., Michel, F., Albrecht, K., Eiden, M., Rinder, M., Lachmann, L., Höper, D., Vina-Rodriguez, A., Gaede, W., Pohl, A., Schmidt-Chanasit, J., Groschup, M.H., 2019. West nile virus epizootic in Germany, 2018. Antiviral Res. 162, 39–43. http://dx.doi.org/10.1016/j.antiviral.2018.12.005.

Ziegler, U., Santos, P.D., Groschup, M.H., Hattendorf, C., Eiden, M., Höper, D., Eisermann, P., Keller, M., Michel, F., Klopfleisch, R., Müller, K., Werner, D., Kampen, H., Beer, M., Frank, C., Lachmann, R., Tews, B.A., Wylezich, C., Rinder, M., Lachmann, L., Lühken, R., 2020. West nile virus epidemic in Germany triggered by epizootic emergence, 2019. Viruses 12 (4), 448. http://dx.doi.org/10.3390/v12040448.