ARTHROPODS AND MEDICAL ENTOMOLOGY - ORIGINAL PAPER

Can data from native mosquitoes support determining invasive species habitats? Modelling the climatic niche of *Aedes japonicus japonicus* (Diptera, Culicidae) in Germany

Antje Kerkow [1,2,3] · Ralf Wieland [1] · Linus Früh [1] · Franz Hölker [2,3] · Jonathan M. Jeschke [2,3,4] · Doreen Werner [1] · Helge Kampen [5]

Abstract
Invasive mosquito species and the pathogens they transmit represent a serious health risk to both humans and animals. Thus, predictions on their potential geographic distribution are urgently needed. In the case of a recently invaded region, only a small number of occurrence data is typically available for analysis, and absence data are not reliable. To overcome this problem, we have tested whether it is possible to determine the climatic ecological niche of an invasive mosquito species by using both the occurrence data of other, native species and machine learning. The approach is based on a support vector machine and in this scenario applied to the Asian bush mosquito (*Aedes japonicus japonicus*) in Germany. Presence data for this species (recorded in the Germany since 2008) as well as for three native mosquito species were used to model the potential distribution of the invasive species. We trained the model with data collected from 2011 to 2014 and compared our predicted occurrence probabilities for 2015 with observations found in the field throughout 2015 to evaluate our approach. The prediction map showed a high degree of concordance with the field data. We applied the model to medium climate conditions at an early stage of the invasion (2011–2015), and developed an explanation for declining population densities in an area in northern Germany. In addition to the already known distribution areas, our model also indicates a possible spread to Saarland, southwestern Rhineland-Palatinate and in 2015 to southern Bavaria, where the species is now being increasingly detected. However, there is also evidence that the possible distribution area under the mean climate conditions was underestimated.

Keywords Citizen science · Invasive species distribution models · Machine learning · Occurrence probability · Support vector machine

Handling Editor: Una Ryan

✉ Antje Kerkow
antje.kerkow@zalf.de; kerkow@igb-berlin.de

[1] Leibniz Centre for Agricultural Landscape Research (ZALF), Eberswalder Str. 84, 15374 Müncheberg, Germany

[2] Department of Biology, Chemistry, Pharmacy, Institute of Biology, Freie Universität Berlin, Königin-Luise-Str. 1-3, 14195 Berlin, Germany

[3] Leibniz-Institute of Freshwater Ecology and Inland Fisheries (IGB), Müggelseedamm 310, 12587 Berlin, Germany

[4] Berlin-Brandenburg Institute of Advanced Biodiversity Research (BBIB), Altensteinstr. 34, 14195 Berlin, Germany

[5] Friedrich-Loeffler-Institut (FLI), Federal Research Institute for Animal Health, Südufer 10, 17493 Greifswald - Insel Riems, Germany

Introduction

Due to globalisation, facilitating long-distance traffic, mass tourism and worldwide trade, increasing numbers of invasive mosquitoes have recently arrived and subsequently established themselves in Germany and mainland Europe (Medlock et al. 2015). As they include potential vectors of a wide range of human and animal pathogens (Schaffner et al. 2013; Becker et al. 2014), they have become a major research issue. To estimate the risk and take protective measures against mosquito-borne disease outbreaks, it is of utmost importance to know the suitable habitats of the various vector species.

We think that the habitats of invasive species can be analysed most effectively in a stepwise procedure. The first step should be to identify the climatic niche within the invaded area. This is challenging because the spread of invasive

species (especially short-lived exothermic insects) is influenced by spontaneous weather events on the one hand, but also by long-term climate, which has impacted vegetation and mosquito populations in the past. In the second step, landscape elements such as land use and altitude need to be considered. However, some regions, although suitable for an invasive species, are unlikely to become populated by active migration due to dispersal barriers. In the case of mosquitoes, for example, these can be large cultivated areas. On the other hand, there are propagation paths that enable rapid passive spread, for example, along roads and waterways, because some mosquito species may be displaced by cars in their adult forms (Eritja et al. 2017) or by container ships in their egg stages (Eritja et al. 2005; Hofhuis et al. 2008; Reiter 2010). Therefore, in the third step, the consideration of propagation paths is useful, including a propagation simulation.

In this paper, we focus on the first step and try to identify climatically suitable areas for the Asian bush mosquito (*Aedes japonicus japonicus* (Theobald, 1901)) in Germany.

The species is one of eight non-indigenous culicid species recently registered in Germany, and apparently the most widespread of them (Kampen et al. 2017; Koban et al. 2019). It originates from East Asia (Miyagi 1971) and was first recorded in 1998 and 2000 in North America and Europe. Since then, it has expanded its distribution range rapidly on both continents (Peyton et al. 1999; Kampen and Werner 2014; Kampen et al. 2017). In its native range as well as in most invasion areas, the climate is temperate and characterised by winters with frost and snowfall. However, it was also found in subtropical and tropical climates such as Florida and Hawaii (Egizi and Fonseca 2015; Riles et al. 2017). The eggs of this species are desiccation and frost resistant (Reuss et al. 2018) and are laid by the females in rock pools of rivers, water-filled tree holes or various kinds of small artificial containers that are able to collect water such as flower pots or vases, buckets, ash trays and bird baths (Tanaka et al. 1979; Scott 2003; Kampen et al. 2012; Kaufman et al. 2012). Egg hatching and larval development in spring begin at 4–5 °C, and development time decreases significantly with temperatures rising up to 28 °C (Scott 2003; Burger and Davis 2008; Kampen et al. 2016a). Temperatures above 34 °C inhibit larval development (Scott 2003). The species usually overwinters in the egg stage, but in warmer regions, it is also possible and observed to hibernate in the larval stage (Reuss et al. 2018; Bova et al. 2019). Immature stages are usually found both sooner in spring and later in autumn than coexisting mosquito species (Iriarte et al. 1991; Burger and Davis 2008; Kaufman and Fonseca 2014).

For habitat analysis of both invasive and native species, data-driven machine learning approaches have been widely applied and proven successful (Drake et al. 2006; Jeschke and Strayer 2008; Früh et al. 2018). In general, species distribution models can be created using either presence-only or combined presence-absence data. Absence data would doubtlessly

improve the distribution predictions for an invasive species (Elith et al. 2006; Vaclavik and Meentemeyer 2009), and they can be obtained either via monitoring programmes or computer simulations. The simulated so-called background or pseudo-absence data are randomly distributed points in the entire model area or only in regions without evidence and with a certain spatial distance to the known occurrences (VanDerWal et al. 2009; Barbet-Massin et al. 2012).

When modelling invasive species that have only recently been established, the available dataset often poses a problem because usually only a small number of presence data are available. Furthermore, absence data cannot be used for modelling (Liu et al. 2013), as it is unknown whether the species is not present in a certain area because (i) the area does not offer appropriate habitat conditions, (ii) the area is not accessible due to dispersal barriers, or (iii) not enough time has passed for the species to arrive. Additionally (iv), it is possible that the species does occur in an area, even though it has not been found during field surveys. The latter may particularly be the case when very large areas are surveyed, as during the *Ae. j. japonicus* monitoring throughout Germany (Kampen et al. 2016a).

Although the use of background/pseudo-absence data leads to an artificially enforced high evaluation of a model and is therefore hotly debated (VanDerWal et al. 2009; Vaclavik and Meentemeyer 2009), the majority of invasive mosquito distribution models rely on them. Currently, the most frequently applied system of modelling is the maximum entropy (MaxEnt) modelling algorithm (Phillips et al. 2006; Fischer et al. 2011; Rochlin et al. 2013; Thomas et al. 2014; Melaun et al. 2015). Combinations of machine learning algorithms for presence and background data have also been used with the objective to obtain more robust predictions than with a single algorithm (Kraemer et al. 2015; Cunze et al. 2016). However, this procedure is also under discussion (Demertzis et al. 2017).

Predictive distribution maps based on either single or combined machine learning algorithms risk underestimating the future habitats of invasive species, as only a short time has passed since the arrival of the species. Furthermore, training data may not reflect all possible ranges under the selected climate variables due to lacking propagation paths between the suitable habitats. To circumvent these problems, scientists have used data from other countries and continents, where the mosquito species of interest is native or had already been present for a longer period of time (Fischer et al. 2011; Thomas et al. 2014). In this way, a large dataset can be provided. However, a given species may fill specific ecological niches in particular countries due to different environmental contexts (Jeschke and Strayer 2008). Therefore, using data of the native range may not always lead to accurate predictions for an invaded area.

Thanks to a nationwide monitoring programme established in Germany several years ago, which also includes the citizen

science project "Mueckenatlas" (Werner et al. 2014; Kampen et al. 2015; Walther and Kampen 2017), a large amount of occurrence data is available concerning the German mosquito fauna, which includes around 50 species. Based on the idea that each species probably occupies its individual and unique ecological niche, we have tested whether it is possible to distinguish the climatic niche of *Ae. j. japonicus* from the climatic niches of native species. Thus, contrary to modelling approaches relying on occurrence data outside the model region, as well as to a modelling approach based on life history data obtained from laboratory analyses (Wieser et al. 2019), we aim to determine the realised ecological niche (Jeschke and Strayer 2008). This was achieved by the application of a machine learning approach based on a support vector machine (Cortes and Vapnik 1995), presence data of *Ae. j. japonicu*s and three native mosquito species from the invaded region (Germany). The algorithm calculates a dividing hyperplane for the vector dataset of environmental parameters and for the four mosquito species, respectively. It is not easy to identify which weather data explain the distribution patterns of the species at best. We have therefore developed a procedure in the framework of this study, which has been published separately (Wieland et al. 2017).

The support vector machine (Cortes and Vapnik 1995) is a supervised learning algorithm and frequently used for data-driven species distribution models (e.g. Drake et al. 2006; Fukuda et al. 2013). It was developed in the 1990s along with a number of other machine learning techniques such as regression tree and random forest. The algorithm was primarily implemented as a linear binary classifier that maximises the margin between two sampling groups (Cortes and Vapnik 1995; Kampichler et al. 2010). When calculating the exact position and orientation of the hyperplane that separates the groups from each other, the algorithm does not consider all data points, but only those closest to the plane, the so-called supported vectors. Advanced forms of support vector machines use non-linear kernels (Nalepa and Kawulok 2019). By means of the "kernel-trick", the training data are mapped into a higher dimensional space which simplifies the computation of the separating hyperplanes for complicated patterns (Drake et al. 2006; Kampichler et al. 2010; Nalepa and Kawulok 2019). A kernel of the RBF (radial basis function) type was used in this approach.

Research by our working group (Früh et al. 2018) has shown that even with noisy data, the application of the support vector machine in combination with a low-dimensional space-kernel is relatively robust against over-fitting. The algorithm is thus a suitable method of machine learning for our relatively small data sets which are prone to this issue. Früh et al. (2018) found that the algorithm did not always achieve the highest accuracy in training, but often came out best in validation. The support vector machine as a classifier produces binary outputs, but probability values can be generated in a post-processing step (Platt 1999). These probabilities give a much more detailed picture of the spatial extent of the habitat than binary values.

In summary, the aims of this study are (i) to develop a model dependant on climate parameters that enables accurate predictions of future distributions of an invasive mosquito species based on a limited amount of presence-only data, and (ii) to apply this model to *Ae. j. japonicus* in Germany.

Materials and methods

Native species

For habitat separation, we have selected two native mosquito species that are widespread and common throughout Germany: *Aedes vexans* (Meigen, 1830) and *Aedes geniculatus* (Olivier, 1791). *Aedes vexans* is a floodwater species of the lowlands, where it develops in temporary water bodies. It is adapted to temperate climates (Becker et al. 2010) and develops in tremendous numbers after heavy rainfall and flooding. Larval hatching starts at temperatures above 9 °C, but development is optimal at 30 °C. As with *Ae. j. japonicus*, the eggs are resistant against drought and frost and can even sustain temperatures as low as − 20 °C (Becker et al. 2010).

Aedes geniculatus preferably breeds in water-filled tree holes, open tree stumps and branch axils that can be found on mature, deciduous trees (Dahl and Blackmore 2001). Eggs are laid on humid walls of the wood, and the larvae hatch after rainwater has filled the holes (Dahl and Blackmore 2001). Like the Asian bush mosquito, *Aedes geniculatus* overwinters either in the egg stage in cold regions or in the larval stage in warmer regions and the eggs can withstand freezing.

We additionally picked one species, *Anopheles daciae* Linton, Nicolescu and Harbach 2004, whose occurrence is particularly concentrated in the north-eastern part of Germany in order to take account of particular continental conditions in Germany. *Anopheles daciae* is a member of the *Anopheles maculipennis* complex, which has only recently been described (Nicolescu et al. 2004). Little is known yet about the ecology of this species. In the framework of a nationwide mosquito monitoring programme in Germany, *An. daciae* was found in the northern Upper Rhine Valley, in the eastern North German lowlands and in the lower river valleys in southern Germany (Kampen et al. 2016b). This indicates linkage of occurrence to both low altitude and relatively continental climate conditions.

Figure 1 shows the collection sites of each model species in Germany from 2011 to 2014, based on nationwide systematic active and passive mosquito monitoring. The passive monitoring data originate from the "Mueckenatlas" project and account for 68.9% of the data used for training and evaluation
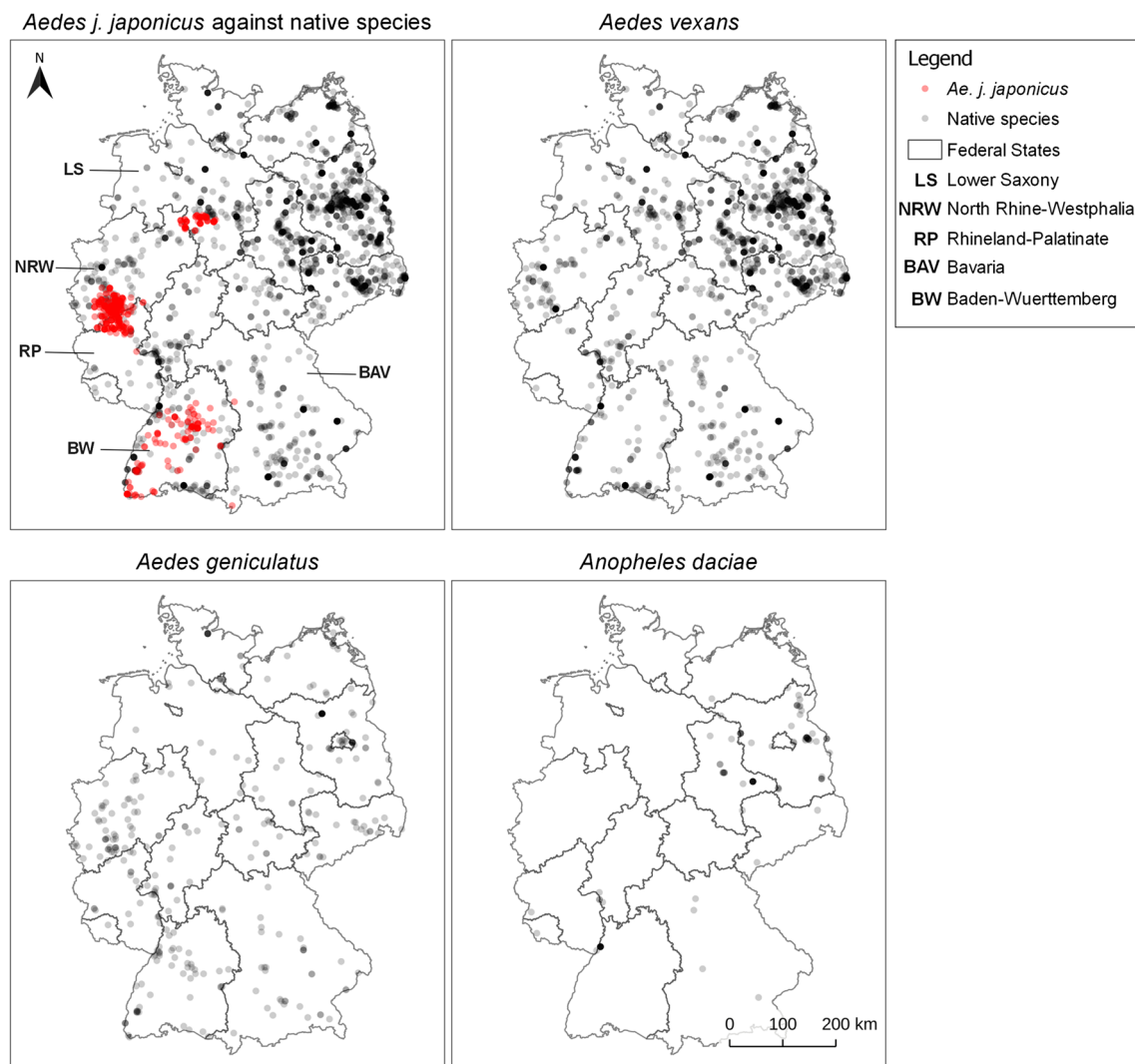
Fig. 1 Visualisation of species sampling data of the years 2011–2014 (training years for the model). The evidence points appear in high colour intensity when several of the same colour overlay each other. This image was created with QGIS 3.8.2

of this model. The active monitoring data are derived from on-site collections in regions from which invasive mosquito species were submitted, by examining possible breeding sites for larvae and installing mosquito traps.

Data access

Mosquito database

The mosquito occurrence data are derived from monitoring projects carried out in Germany since 2011 and stored inside the German national mosquito database (CULBASE, https:// culbase.fli.de/Public/Page/Info.aspx). Amongst others, the database contains collection sites (including coordinates), mode of collection (including dates) and methods of species determination. The CULBASE database can be accessed via an export interface that enables querying data of the species and the year of interest, and downloading these data as Excel

files, one file for each species. Data of *Ae. j. japonicus*, *Ae. vexans*, *Ae. geniculatus* and *An. daciae* from the years 2011–2014 were filtered for training the model, and data from 2015 to evaluate it. The Excel files contained the coordinates and the dates of collection. These data were entered into our object-oriented model using the PYTHON "PANDAS" module.

Weather and climate data

We used freely available data from the German Weather Service (Deutscher Wetterdienst 2017) which are gridded for Germany with a cell size of 1 km × 1 km and provide hourly, daily, monthly, seasonal, yearly and multi-annual resolutions. In the following, we define data up to an annual resolution as weather data and multi-year averages as climate data.

For each of the four mosquito species and their occurrence points in the training years 2011, 2012, 2013 and 2014, we

created Excel tables with the help of a PYTHON script. These tables were the input for training the model; they contained the weather data for each occurrence point. According to Wieland et al. (2017), we used the following eight weather variables: mean temperature in spring (March, April, May) [T13], September [T09], October [T10] and December [T12]; sum of precipitation in February [P02], April [P04] and June [P06]; and drought index of autumn (September, October, November) [D15]. Wieland et al. (2017) use the same model as presented here and describe a method for the analysis of weather data that explain the distribution of the mosquito species at best. Therefore, a set of 37 data was pre-selected that may play a role in a biological point of view. These were, for example, temperature and precipitation data during the breeding season and the temperature and number of frost days during the winter months. The data sets were reduced to the best combination using a genetic optimisation procedure. The optimisation parameter was the *f*1 score (see subsection "Model validation" for the definition).

Model training

As shown in Fig. 2, we used the species occurrence data and the weather data from 2011 to 2014, respectively, for training the model. The support vector machine only needed to distinguish between two classes, the invasive target species on the one hand, and the native species on the other hand. By combining the monitoring data of several native species to one class, we converted a multi-classification problem into a binary classification problem, which helps to correctly calculate a decision boundary for the target species in space (Garreta and Moncecchi 2013).

As a training algorithm, we used the support vector machine from the SCIKIT toolbox (Pedregosa et al. 2011). Our training dataset contained 508 occurrence points for *Ae. j. japonicus*, 2056 for *Ae. vexans*, 322 for *Ae. geniculatus* and 102 for *An. daciae*. These data and the stored weather conditions were read in from the Excel tables with the PYTHON "PANDAS" tool. To mitigate the unbalanced numbers of mosquito species data, a maximum of 1000 data points per species were randomly selected for training the model. We used a radial basis function (RBF) kernel and passed the following hyper-parameters as arguments to the constructor of the support vector machine: gamma = 0.0005, C = 1.0, tolerance = 1e-10, probability = true.

The support vector machine calculated the hyperplane for the separation of both classes. With the returned binary values for the training data (positive or negative), the confusion matrix was calculated. By re-training the parameters on a sigmoid function, the original outputs were mapped into probability values on a scale from 0 (very low probability of occurrence) to 1 (very high probability of occurrence) (Platt 1999; Pedregosa et al. 2011). This procedure is implemented by default in SCIKIT-LEARN. The latter training results were saved and used for creating distribution maps of *Ae. j. japonicus* under certain weather conditions.

Analysis of the training classes

To illustrate the different dependencies of the species on the eight weather variables, we presented the values for the two training classes ("*Ae. j. japonicus*" and "native species") at their respective sampling sites in violin plots. We additionally tested the data of both groups for significant differences ($p < 0.05$) with the Wilcoxon-Mann-Whitney test, using R and the

Fig. 2 Workflow: Model training and validation. This image was created with Inkscape 0.92
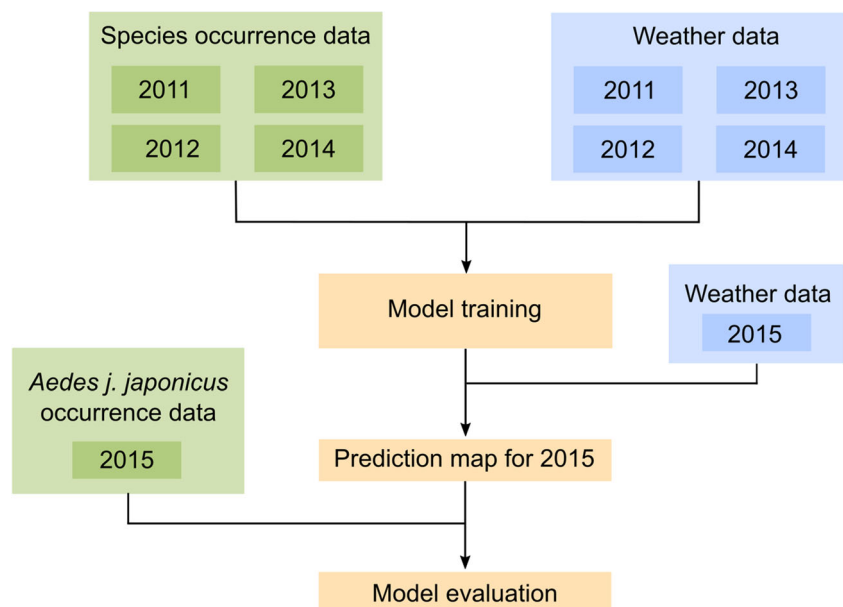
wilcox.test-function where we set the following parameters: paired = FALSE, alternative = "two.sided".

Model validation

Classification

The classification result was evaluated by calculating a confusion matrix and the values of precision, recall and $f1$ score. A confusion matrix summarises the reclassification results for the set of test data (Garreta and Moncecchi 2013). Based on this matrix, the values for precision and recall are calculated for the two classes ("Ae. j. japonicus" and "native species"). The precision computes the proportion of true positive values (TP) from the predicted positive values, including true positive and false positive (FP) values (Eq. 1). By contrast, the recall calculates what proportion of the positive observations was correctly evaluated. This calculation (Eq. 2) considers the false negative (FN) values.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \qquad (1)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \qquad (2)$$

The $f1$ score represents the harmonic mean of precision and recall (Eq. 3).

$$f1 = \frac{2(\text{Prec} \cdot \text{Rec})}{\text{Prec} + \text{Rec}} \qquad (3)$$

To calculate the training results for both classes, weighted mean values were calculated for precision, recall and $f1$ score (Eq. 4). When x is the validation variable and n the sum of observations, then applies:

$$x = \frac{x_{\text{japonicus}} \cdot n_{\text{japonicus}} + x_{\text{native}} \cdot n_{\text{native}}}{n_{\text{japonicus}} + n_{\text{native}}}. \qquad (4)$$

Predictive power

In order to validate the predictive power of our model, we used the Ae. j. japonicus sampling data of 2015, which had not been included in the training, and analysed the pixel values of the predictive distribution map, referring to the weather conditions of 2015. We drew a histogram and calculated the median and interquartile range to describe the pixel values for the validation dataset.

Colonisation potential between 2011 and 2015

The weather conditions of individual years can differ significantly from one another and with regionally varying intensity,

influencing the distribution potential of the mosquito species. We have therefore carried out a model application for the average climate values for the years 2011–2015 and produced a corresponding habitat map for Ae. j japonicus.

Results

Analysis of the training classes

All eight weather variables show overlapping areas for both training groups, i.e. the native species and Ae j. japonicus (Fig. 3). However, with respect to all weather variables, the distributions differ significantly from each other with $p < 0.02$.

Model validation

Classification

The combination of our training algorithm and the chosen weather variables yielded very good classification results. As can be seen in the confusion matrix (Table 1), 241 of 308 test points (78.2%) on the map, verified to be climatically suitable areas for Ae. j. japonicus in 2015, are correctly evaluated (true positives). Conversely, only 67 test points (21.8%) where Ae. j. japonicus was present lay in the area categorised to belong to the class of native species. From the target class of native species supported by 115 test points, only 20.9% were predicted to belong to the class of Ae. j. japonicus.

Regarding precision, recall and $f1$ score, the model classifies the test dataset of Ae. j. japonicus extremely well and only slightly less well for the test data of the native species. The $f1$ score in total (weighted average) was 79% (Table 2).

Predictive power

A comparison of the predicted occurrence probability for Ae. j. japonicus using the weather conditions and the species sampling data from 2015 indicates a high degree of accuracy of the model. This also becomes apparent in the prediction map for the year 2015 (Fig. 4). The occurrence probabilities for the validation points ($n = 308$) have a median of 0.78 and lie in the interquartile range between 0.52 and 0.81 (Fig. 5).

Colonisation potential between 2011 and 2015

According to the average climate conditions of the period 2011–2015 (Fig. 4, right), the highest colonisation potentials for Ae. j. japonicus (80–100%) lie in the West German federal states of North Rhine-Westphalia (NRW), Baden-Wuerttemberg (BW), Saarland (SL) and in some parts of Rhineland-Palatinate and Hesse. In addition, there is a highly
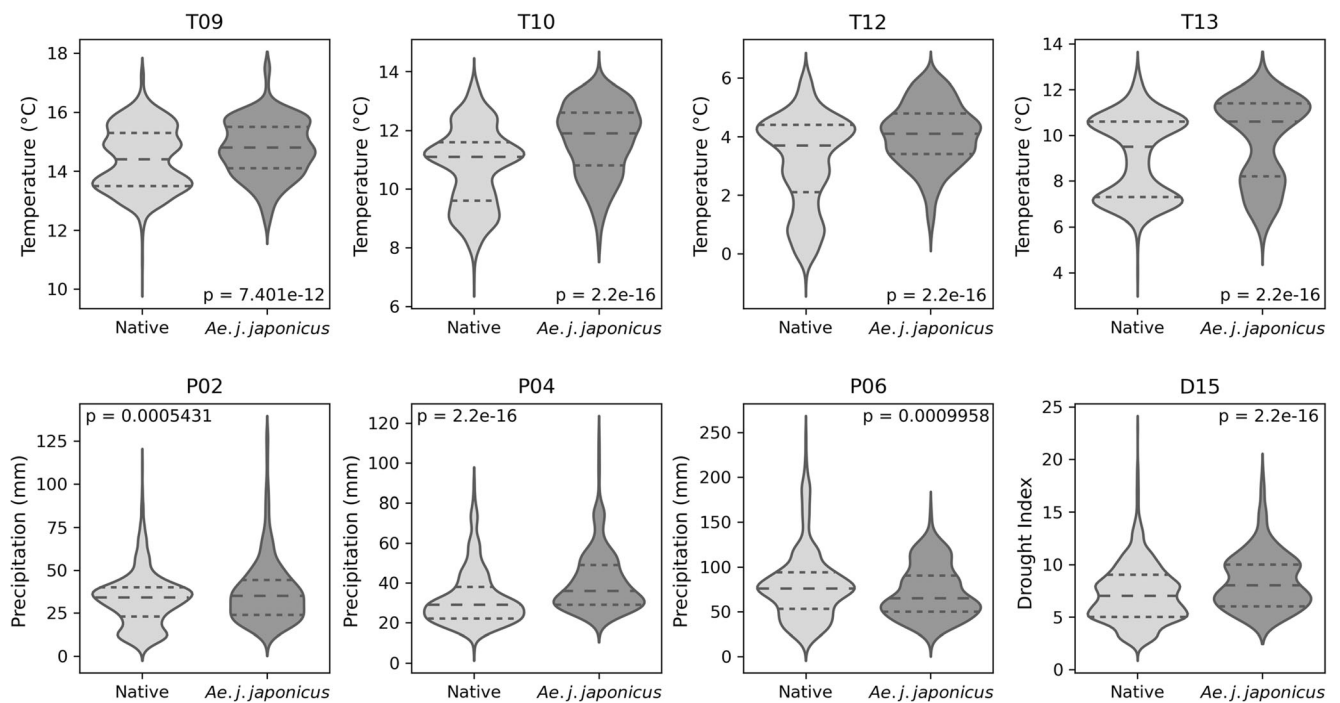
Fig. 3 Weather conditions at the mosquito collection sites in the corresponding year of monitoring (between 2011 and 2015) classified by the training group. *Aedes vexans*, *Aedes geniculatus* and *Anopheles daciae* are grouped under the term "Native". T09 = mean temperature in September, T10 = mean temperature in October, T12 = mean temperature in December, T13 = mean temperature in spring (average of March, April and May), P02 = sum of precipitation in February, P04 = sum of precipitation in April, P06 = sum of precipitation in June, D15 = drought index of autumn (average of September, October and November). This image was created under Python 3.7

probable distribution area (60–80%) in the southwest of Lower Saxony (LS).

Discussion

The comparison of the prediction map for the year 2015 with field collection data from 2015 (which have not been included in the model training) showed an extremely good correspondence. The concentration of predicted suitable regions for *Ae. j. japonicus* in western Germany (Baden-Wuerttemberg, Rhineland-Palatinate, Hesse, North Rhine-Westphalia) is not surprising, as we had a high amount of training data from these regions. However, the high occurrence probabilities in Bavaria (south-eastern Germany) under the weather conditions of the year 2015 is remarkable, as is the predicted spread

far into the most northern regions of Germany. Indeed, *Ae. j. japonicus* specimens were not only found in 2015 in an area of roughly 900 km² in south-eastern Bavaria on the border to Austria (Zielke et al. 2016), but this species was actually even more widespread in 2016 and 2017 in large areas of southern Bavaria (Kampen et al. 2017; Koban et al. 2019).

The model application to the average climate conditions of the period 2011–2015 shows where good colonisation conditions had been since an early invasion stage during the entire period. We refer to this short period because *Aedes j. japonicus* was discovered in southern Germany (Baden-Wuerttemberg) only in 2008 and in western and northern Germany in 2012. However, in order to make predictions for a mean climate based on a longer period, the model has to be re-calibrated. With the result of long-term averaged climate conditions, it is possible to predict future developments of *Ae. j. japonicus* in Germany.

Table 1 Confusion matrix for the trained model and validation data from 2015

		Predicted class		Sum
		Ae. j. japonicus	Native species	
Observed class	*Ae. j. japonicus*	241 (**TP**, TN)	67 (**FN**, FP)	308
	Native species	24 (**FP**, FN)	91 (**TN**, TP)	115
Sum		265	158	423

TP true positive, *TN* true negative, *FN* false negative, *FP* false positive (first place and bold referring to *Ae. j. japonicus*, in second place referring to the class of native species)

Table 2   Validation of the model training with test data from 2015

Class	Precision	Recall	*f*1
Ae. j. japonicus	0.91	0.78	0.84
Native species	0.58	0.79	0.67
Total	0.82	0.78	0.79

The calibration and forecast for the future period of 2021–2050 are presented by Kerkow et al. (2019).

According to the map for the average climate conditions from 2011 to 2015, the climate in north-eastern Germany was less suitable than in 2015 alone. The suitable area in Lower Saxony is clearly smaller and the prediction values were low (around 10–50%) where the population was located. This is consistent with the observation that the population seems to have decreased over the period 2011–2014, but has expanded again since 2015 (Koban et al. 2019).

In order to better assess the effectiveness of our modelling technique, we compared our results for the averaged climate conditions with those of two other models for Germany produced in completely different ways: (i) Melaun et al. (2015) used occurrence and weather data for 2011–2013 and seven different weather variables to train a MaxEnt model using simulated background data. The result is mapped to the climate of the period 1950–2000 and differs only slightly from ours. A completely different approach (ii) was carried out by Wieser et al. (2019). Using experimental life history data obtained from *Ae. j. japonicus* specimens from southern Germany (temperature-dependent development, mortality, reproduction and density-dependent larval mortality rates), the authors modelled
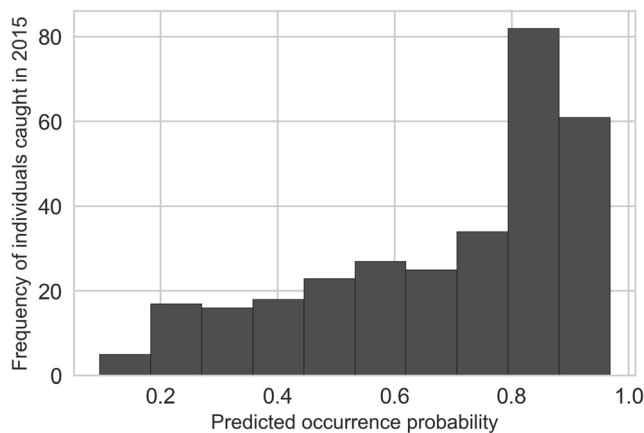

Fig. 5 Predicted occurrence probabilities of *Ae. j. japonicus* in Germany, related to weather conditions of 2015, compared with field collections in 2015. This image was created with Python 2.7

population dynamics for two specific locations and determined where the species can establish stable populations in Germany. The analysis, based on climate data from 1993 to 2013, reveals that the range may be larger than we have modelled. In detail, the habitats in southwestern and western Germany have a similar pattern, but the size of the suitable area in northern Germany is much larger, reaching to the far north and eastwards as far as Berlin.

This leads us to conclude that our model may underestimate the possible spread of *Ae. j. japonicus* in Germany. Although we have already used occurrence data of four years for the training, they may not yet be sufficient for all suitable climatic conditions. However, our result is remarkably
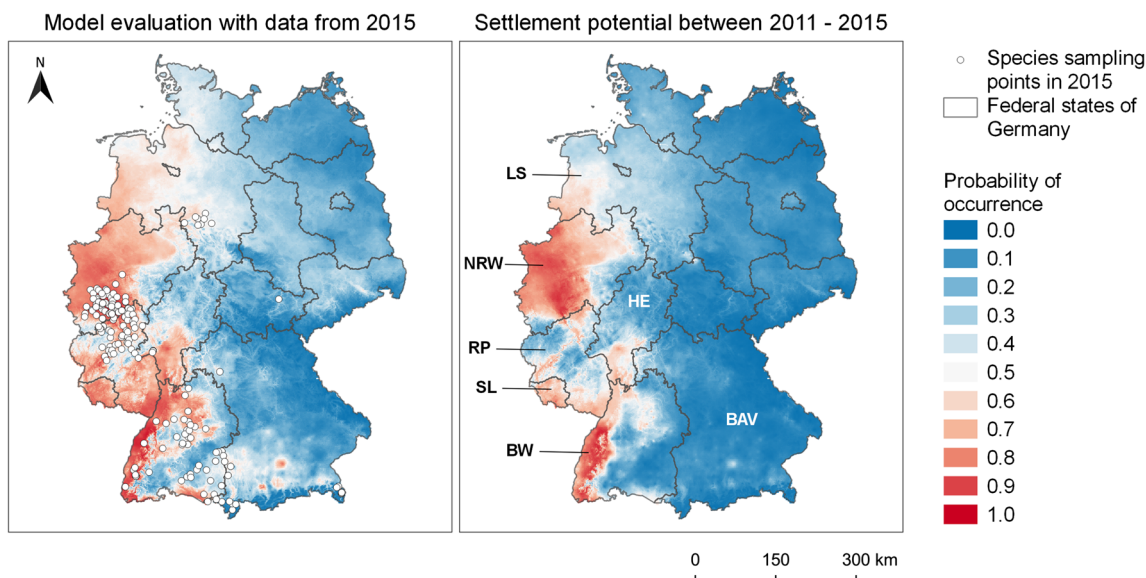

Fig. 4 Left: predicted occurrence areas of *Ae. j. japonicus* for the year 2015 as opposed to field samplings in 2015. Species data from 2015 were not included in the model training. Right: Average colonisation potential in the period 01/2011–12/2015 (labelled federal states: LS, Lower Saxony; NRW, North Rhine-Westphalia; HE, Hesse; RP, Rhineland-Palatinate; SL, Saarland; BW, Baden-Wuerttemberg; BAV, Bavaria). This image was created with QGIS 3.8.2

congruent with that of the MaxEnt model and thus able to compete with a well-established and widely used method.

A possible problem with data-driven species distribution models, apart from the possibility that not all environmental conditions are reflected, is the quality of the training data. In our case, 70% of our data originate from the citizen science project "Mueckenatlas". Thus, the modelling results may be slightly biased, as the majority of mosquitoes submitted by the participants are most likely from densely populated regions. However, about 30% of the training data originate from systematic field collections which particularly concentrated on the margins and surroundings of known distribution areas of *Ae. j. japonicus* (Kampen et al. 2016a).

Regarding the selection of mosquito species to replace absence data in machine learning, the question arises as to why we chose the species *Ae. vexans*, *Ae. geniculatus* and *An. daciae*. Our modelling approach is driven solely by climate data, and the aim was to determine the climatic-ecological niche of the Asian bush mosquito. We therefore searched for mosquito species, which, when considering some relevant climate parameters (mean annual and summer precipitation, drought indices and temperature data), seemed to differ in their occurrence spectrum from that of *Ae. j. japonicus*. We have tested different species combinations in the initial phase of modelling and found, firstly, that the use of more or less than three species tends to diminish the results measured by the total *f*1 score. Secondly, the combination of the two common species *Ae. vexans* and *Ae. geniculatus* together with the species *An. daciae*, which presumably prefers more continental conditions, achieved particularly good results. In a later modelling phase, we optimised the selection of climate variables by means of a deep learning method (Wieland et al. 2017). The existence of competitive conditions was not a criterion for the selection of the species. Competitive situations for mosquitoes result from oviposition and development of pre-imaginal stages in the same habitats. Amongst our species selected for modelling, only the dendrolimnobiotic species *Ae. geniculatus* (Becker et al. 2011) occasionally lays its eggs in the same water sources as *Ae. j. japonicus* (Damiens et al. 2014; Seidel et al. 2016).

Another interesting aspect of our result is that our model performs very well in validation, although the weather and climate data of the two training classes "native species" and "*Ae. j. japonicus*" are noisy and partly overlap (Fig. 4). Consequently, the support vector machine was able to handle this circumstance effectively. This is probably due to the use of the RBF kernel, which can identify complicated distribution patterns in the data. When comparing different algorithms and their combinations for our study (Früh et al. 2018), this algorithm also achieved the best result of all individually applied algorithms. However, if more mosquito data become available in the future via the citizen science project, we would consider using learning algorithms that are much more powerful in general, such as XG-Boost (Chen and Guestrin 2016;

Brownlee 2017). The main strength of the support vector machine lies in the handling of smaller amounts of data (Pedregosa et al. 2011; Nalepa and Kawulok 2019).

As mentioned before, mesoclimate data with a resolution of 1 km^2 can only give an approximate picture of the suitable habitats and thus a possible distribution of a mosquito species. At the local level, landscape forms and the resulting microhabitats also play a role. It is also possible that unfavourable climatic conditions may be compensated in some way. For example, in regions with low precipitation, people may regularly provide water-filled breeding habitats such as rain water barrels or flowerpots. Therefore, in a further step, the results of this model were combined with landscape and wind data, which led to a significant improvement of the results as measured by the hit rate for *Ae. j. japonicus* (Kerkow et al. 2019).

Conclusions

Our model approach appears to be suitable for predicting the distribution area of the Asian bush mosquito *Ae. j. japonicus* in newly invaded areas. The results were achieved by the exclusive use of presence data of this species and three native species. The model output matches extremely well with presence data from 2015, which was not included in the model training. Due to the use of presence-only data, the method presented here is well-suited to datasets that are based on passive monitoring programmes such as citizen science projects. Our result is consistent with that of another data-driven machine-learning approach that used generated absence data (Melaun et al. 2015). However, when applying the method, it is important to bear in mind that the dispersal potential of an invasive species is underestimated if the dataset is too small to represent the realised ecological niche of a species.

With the help of large-scale, long-term averaged climate data, only a rough impression of the possible further spread of invasive mosquito species in Germany can be obtained, independent of the modelling technique. On a smaller scale, land use forms correlating with the presence of certain breeding habitats and causing certain microclimates have an additional important impact.

Data availability Mosquito occurrence data used in this model are provided in the supplementary information files. We uploaded four Excel tables,

one for each mosquito species, containing the year of discovery, the geo-coordinates and the weather values relevant for both model training and validation. The CULBASE database is not yet publicly accessible, further mosquito data are available on reasonable request to DW or HK.

Compliance with ethical standards

Conflict of interest   The authors declare that they have no conflict of interest.

References

Barbet-Massin M, Jiguet F, Albert CH, Thuiller W (2012) Selecting pseudo-absences for species distribution models: how, where and how many? Methods Ecol Evol 3:327–338. https://doi.org/10.1111/j.2041-210X.2011.00172.x

Becker N, Petric D, Zgomba M, Boase C, Madon MB, Dahl C, Kaiser A (2010) Mosquitoes and their control, 2nd ed. Springer-Verlag, Berlin Heidelberg. https://doi.org/10.1007/978-3-540-92874-4

Becker N, Huber K, Pluskota B, Kaiser A (2011) *Ochlerotatus japonicus japonicus* - a newly established neozoan in Germany and a revised list of the German mosquito fauna. Euro Mosq Bull 88–102

Becker N, Krueger A, Kuhn C, Plenge-Boenig A, Thomas S, Schmidt-Chanasit J, Tannich E (2014) Mosquitoes as vectors for exotic pathogens in Germany. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 57:531–540. https://doi.org/10.1007/s00103-013-1918-8

Bova J, Soghigian J, Paulson S (2019) The prediapause stage of Aedes japonicus japonicus and the evolution of embryonic diapause in Aedini. Insects 10:222. https://doi.org/10.3390/insects10080222

Brownlee J (2017) XGBoost with Python: Gradient boosted trees with XGBoost and scikit-learn, v1.5. Machine Learning Mastery. Ebook. https://machinelearningmastery.com/xgboost-with-python/

Burger JF, Davis H (2008) Discovery of *Ochlerotatus japonicus japonicus* (Theobald) (Diptera: Culicidae) in Southern New Hampshire, USA and its subsequent increase in abundance in used tire casings. Entomol News 119:439–444. https://doi.org/10.3157/0013-872X-119.5.439

Chen T, Guestrin C (2016) XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD'16. ACM Press, San Francisco, California, USA, pp. 785–794. https://doi.org/10.1145/2939672.2939785

Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20:273–297. https://doi.org/10.1023/A:1022627411411

Cunze S, Koch LK, Kochmann J, Klimpel S (2016) *Aedes albopictus* and *Aedes japonicus* - two invasive mosquito species with different temperature niches in Europe. Parasit Vectors 9:573. https://doi.org/10.1186/s13071-016-1853-2

Dahl C, Blackmore M (2001) The distribution and status of *Ochlerotatus geniculatus* (Olivier) in Fennoscandia. Eur Mosq Bull 9:12–16

Damiens D, Ayrinhac A, Bortel WV, Versteirt V, Dekoninck W, Hance T (2014) Invasive process and repeated cross-sectional surveys of the mosquito *Aedes japonicus japonicus* establishment in Belgium. PLoS One 9:e89358. https://doi.org/10.1371/journal.pone.0089358

Demertzis K, Iliadis L, Anezakis V-D (2017) Commentary: *Aedes albopictus* and *Aedes japonicus* - two invasive mosquito species with different temperature niches in Europe. Front Environ Sci. https://doi.org/10.3389/fenvs.2017.00085

Deutscher Wetterdienst (2017) Wetter und Klima. In: CDC (Climate Data Centre). https://www.dwd.de/DE/klimaumwelt/cdc/cdc_node.html. Accessed 8 Nov 2017

Drake JM, Randin C, Guisan A (2006) Modelling ecological niches with support vector machines. J Appl Ecol 43:424–432. https://doi.org/10.1111/j.1365-2664.2006.01141.x

Egizi A, Fonseca DM (2015) Ecological limits can obscure expansion history: patterns of genetic diversity in a temperate mosquito in Hawaii. Biol Invasions 17:123–132. https://doi.org/10.1007/s10530-014-0710-5

Elith J, Graham CH, Anderson RP, Dudík M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick JR, Lehmann A, Li J, Lohmann LG, Loiselle BA, Manion G, Moritz C, Nakamura M, Nakazawa Y, Overton JMM, Peterson AT, Phillips SJ, Richardson K, Scachetti-Pereira R, Schapire RE, Soberón J, Williams S, Wisz MS, Zimmermann NE (2006) Novel methods improve prediction of species' distributions from occurrence data. Ecography 29:129–151. https://doi.org/10.1111/j.2006.0906-7590.04596.x

Eritja R, Escosa R, Lucientes J, Marquès E, Roiz D, Ruiz S (2005) Worldwide invasion of vector mosquitoes: present European distribution and challenges for Spain. Biol Invasions 7:87. https://doi.org/10.1007/s10530-004-9637-6

Eritja R, Palmer JRB, Roiz D, Sanpera-Calbet I, Bartumeus F (2017) Direct evidence of adult *Aedes albopictus* dispersal by car. Sci Rep 7:1–15. https://doi.org/10.1038/s41598-017-12652-5

Fischer D, Thomas SM, Niemitz F, Reineking B, Beierkuhnlein C (2011) Projection of climatic suitability for *Aedes albopictus* Skuse (Culicidae) in Europe under climate change conditions. Glob Planet Change 78:54–64. https://doi.org/10.1016/j.gloplacha.2011.05.008

Früh L, Kampen H, Kerkow A, Schaub GA, 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. https://doi.org/10.1016/j.ecolmodel.2018.08.011

Fukuda S, De Baets B, Waegeman W, Verwaeren J, Mouton AM (2013) Habitat prediction and knowledge extraction for spawning European grayling (*Thymallus thymallus* L.) using a broad range of species distribution models. Environ Model Softw 47:1–6. https://doi.org/10.1016/j.envsoft.2013.04.005

Garreta R, Moncecchi G (2013) Learning scikit-learn: machine learning in Python, 1st edn. Packt Publishing Ltd, Birmingham

Hofhuis A, Reimerink J, Reusken C, Scholte E-J, de Boer A, Takken W, Koopmans M (2008) The hidden passenger of lucky bamboo: do imported *Aedes albopictus* mosquitoes cause Dengue virus transmission in the Netherlands? Vector Borne Zoonotic Dis 9:217–220. https://doi.org/10.1089/vbz.2008.0071

Iriarte WLZ, Tsuda Y, Wada Y, Takagi M (1991) Distribution of mosquitoes on a hill of Nagasaki City, with emphasis to the distance from human dwellings. Trop Med 33:55–60

Jeschke JM, Strayer DL (2008) Usefulness of bioclimatic models for studying climate change and invasive species. Ann N Y Acad Sci 1134:1–24. https://doi.org/10.1196/annals.1439.002

Kampen H, Werner D (2014) Out of the bush: the Asian bush mosquito *Aedes japonicus japonicus* (Theobald, 1901) (Diptera, Culicidae) becomes invasive. Parasit Vectors 7:59. https://doi.org/10.1186/1756-3305-7-59

Kampen H, Zielke D, Werner D (2012) A new focus of *Aedes japonicus japonicus* (Theobald, 1901) (Diptera, Culicidae) distribution in Western Germany: rapid spread or a further introduction event? Parasit Vectors 5:284. https://doi.org/10.1186/1756-3305-5-284

Kampen H, Medlock JM, Vaux AG, Koenraadt CJ, van Vliet AJ, Bartumeus F, Oltra A, Sousa CA, Chouin S, Werner D (2015) Approaches to passive mosquito surveillance in the EU. Parasit Vectors 8:9. https://doi.org/10.1186/s13071-014-0604-5

Kampen H, Kuhlisch C, Fröhlich A, Scheuch DE, Walther D (2016a) Occurrence and spread of the invasive Asian bush mosquito *Aedes japonicus japonicus* (Diptera: Culicidae) in West and North Germany since detection in 2012 and 2013, respectively. PLoS One 11:e0167948. https://doi.org/10.1371/journal.pone.0167948

Kampen H, Schäfer M, Zielke DE, Walther D (2016b) The *Anopheles maculipennis* complex (Diptera: Culicidae) in Germany: an update following recent monitoring activities. Parasitol Res 115:3281–3294. https://doi.org/10.1007/s00436-016-5189-9

Kampen H, Schuhbauer A, Walther D (2017) Emerging mosquito species in Germany–a synopsis after 6 years of mosquito monitoring (2011–2016). Parasitol Res 116:3253–3263. https://doi.org/10.1007/s00436-017-5619-3

Kampichler C, Wieland R, Calmé S, Weissenberger H, Arriaga-Weiss S (2010) Classification in conservation biology: a comparison of five machine-learning methods. Ecol Inform 5:441–450. https://doi.org/10.1016/j.ecoinf.2010.06.003

Kaufman MG, Fonseca DM (2014) Invasion biology of *Aedes japonicus japonicus* (Diptera: Culicidae). Annu Rev Entomol 59:31–49. https://doi.org/10.1146/annurev-ento-011613-162012

Kaufman MG, Stanuszek WW, Brouhard EA, Knepper RG, Walker ED (2012) Establishment of *Aedes japonicus japonicus* and its colonization of container habitats in Michigan. J Med Entomol 49:1307–1317. https://doi.org/10.1603/ME12061

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

Koban MB, Kampen H, Scheuch DE, Früh L, Kuhlisch C, Janssen N, Steidle JLM, Schaub GA, Werner D (2019) The Asian bush mosquito *Aedes japonicus japonicus* (Diptera: Culicidae) in Europe, 17 years after its first detection, with a focus on monitoring methods. Parasit Vectors 12:109. https://doi.org/10.1186/s13071-0193349-3

Kraemer MU, Sinka ME, Duda KA, Mylne AQ, Shearer FM, Barker CM, Moore CG, Carvalho RG, Coelho GE, Bortel WV, Hendrickx G, Schaffner F, Elyazar IR, Teng H-J, Brady OJ, Messina JP, Pigott DM, Scott TW, Smith DL, Wint GW, Golding N, Hay SI (2015) The global distribution of the arbovirus vectors *Aedes aegypti* and *Ae. albopictus*. eLife. https://doi.org/10.7554/eLife.08347

Liu C, White M, Newell G (2013) Selecting thresholds for the prediction of species occurrence with presence-only data. J Biogeogr 40:778–789. https://doi.org/10.1111/jbi.12058

Medlock JM, Hansford KM, Versteirt V, Cull B, Kampen H, Fontenille D, Hendrickx G, Zeller H, Van Bortel W, Schaffner F (2015) An entomological review of invasive mosquitoes in Europe. Bull Entomol Res 105:637–663. https://doi.org/10.1017/S0007485315000103

Melaun C, Werblow A, Cunze S, Zotzmann S, Koch LK, Mehlhorn H, Dörge DD, Huber K, Tackenberg O, Klimpel S (2015) Modeling of the putative distribution of the arbovirus vector *Ochlerotatus japonicus japonicus* (Diptera: Culicidae) in Germany. Parasitol Res 114:1051–1061. https://doi.org/10.1007/s00436-014-4274-1

Miyagi I (1971) Notes on the *Aedes* (Finlaya) *chrysolineatus* subgroup in Japan and Korea (Diptera: Culicidae). Trop Med 13:141–151

Nalepa J, Kawulok M (2019) Selecting training sets for support vector machines: a review. Artif Intell Rev 52:857–900. https://doi.org/10.1007/s10462-017-9611-1

Nicolescu G, Linton Y-M, Vladimirescu A, Howard TM, Harbach RE (2004) Mosquitoes of the *Anopheles maculipennis* group (Diptera: Culicidae) in Romania, with the discovery and formal recognition of a new species based on molecular and morphological evidence. Bull Entomol Res 94:525–535. https://doi.org/10.1079/BER2004330

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E (2011) Scikit-learn: machine learning in Python. J Mach Learn Res 12:2825–2830

Peyton EL, Campbell SR, Candeletti TM, Romanowski M, Crans WJ (1999) *Aedes* (Finlaya) *japonicus japonicus* (Theobald), a new introduction into the United States. J Am Mosq Control Assoc 15:238–241

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. Ecol Model 190:231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

Platt J (1999) Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In: Advances in large margin classifiers. MIT Press, Cambridge, pp 61–74

Reiter P (2010) The standardised freight container: vector of vectors and vector-borne diseases. Rev Sci Tech 29:57–64. https://doi.org/10.20506/rst.29.1.1960

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 11:81. https://doi.org/10.1186/s13071-018-2659-1

Riles MT, Smith JP, Burkett-Cadena N, Connelly CR, Morse GW, Byrd BD (2017) First record of *Aedes japonicus* in Florida. J Am Mosq Control Assoc 33:340–344. https://doi.org/10.2987/17-6696.1

Rochlin I, Ninivaggi DV, Hutchinson ML, Farajollahi A (2013) Climate change and range expansion of the Asian tiger mosquito (*Aedes albopictus*) in Northeastern USA: implications for public health practitioners. PLoS One 8:e60874. https://doi.org/10.1371/journal.pone.0060874

Schaffner F, Medlock JM, Van Bortel W (2013) Public health significance of invasive mosquitoes in Europe. Clin Microbiol Infect 19:685–692. https://doi.org/10.1111/1469-0691.12189

Scott JJ (2003) The ecology of the exotic mosquito *Ochlerotatus* (Finlay) *japonicus japonicus* (Theobald 1901) (Diptera: Culicidae) and an examination of its role in the West Nile virus cycle in New Jersey. Rutgers University, New Jersey

Seidel B, Montarsi F, Huemer HP, Indra A, Capelli G, Allerberger F, Nowotny N (2016) First record of the Asian bush mosquito, *Aedes japonicus japonicus*, in Italy: invasion from an established Austrian population. Parasit Vectors 9:284. https://doi.org/10.1186/s13071-016-1566-6

Tanaka K, Mizusawa K, Saugstad ES (1979) A revision of the adult and larval mosquitoes of Japan (including the Ryukyu Archipelago and the Ogasawara Islands) and Korea (Diptera: Culicidae). Contributions of the American Entomological Institute, 16:1–987

Thomas SM, Tjaden NB, van den Bos S, Beierkuhnlein C (2014) Implementing cargo movement into climate based risk assessment of vector-borne diseases. Int J Environ Res Public Health. https://doi.org/10.3390/ijerph110303360

Vaclavik T, Meentemeyer RK (2009) Invasive species distribution modeling (iSDM): are absence data and dispersal constraints needed to predict actual distributions? Ecol Model 220:3248–3258. https://doi.org/10.1016/j.ecolmodel.2009.08.013

VanDerWal J, Shoo LP, Graham C, Williams SE (2009) Selecting pseudo-absence data for presence-only distribution modeling: how far should you stray from what you know? Ecol Model 220:589–594. https://doi.org/10.1016/j.ecolmodel.2008.11.010

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 54:1790–1794. https://doi.org/10.1093/jme/tjx166

Werner D, Hecker S, Luckas M, Kampen H (2014) The citizen science project "Mückenatlas" supports mosquito (Diptera, Culicidae) monitoring in Germany. Proc. 8th Int. Conf. Urban Pests 20–23 July 2014 Zurich Switz. 119–124

Wieland R, Kerkow A, Früh L, Kampen H, Walther D (2017) Automated feature selection for a machine learning approach toward modeling a mosquito distribution. Ecol Model 352:108–112. https://doi.org/10.1016/j.ecolmodel.2017.02.029

Wieser A, Reuss F, Niamir A, Müller R, O'Hara RB, Pfenninger M (2019) Modelling seasonal dynamics, population stability, and pest control in *Aedes japonicus japonicus* (Diptera: Culicidae). Parasit Vectors 12:142. https://doi.org/10.1186/s13071-019-3366-2

Zielke DE, Walther D, Kampen H (2016) Newly discovered population of *Aedes japonicus japonicus* (Diptera: Culicidae) in Upper Bavaria, Germany, and Salzburg, Austria, is closely related to the Austrian/Slovenian bush mosquito population. Parasit Vectors 9:1. https://doi.org/10.1186/s13071-016-1447-z